



INSTITUTO
SUPERIOR
TÉCNICO

UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Narrative Interfaces for Personal Document Retrieval

Daniel Jorge Viegas Gonçalves
(Master)

Dissertation for the degree of Doctor of Philosophy in
Information Systems and Computer Engineering

Adviser: Doctor Joaquim Armando Pires Jorge

Jury

Chairman: Dean of the Technical University of Lisbon

Members: Doctor Alan Dix

Doctor Nuno Manuel Carvalho Ferreira Guimarães

Doctor Arlindo Manuel Limede de Oliveira

Doctor Joaquim Armando Pires Jorge

Doctor Manuel João Caneira Monteiro da Fonseca

June 2007



INSTITUTO
SUPERIOR
TÉCNICO

UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Narrative Interfaces for Personal Document Retrieval

Daniel Jorge Viegas Gonçalves
(Master)

Dissertation for the degree of Doctor of Philosophy in
Information Systems and Computer Engineering

Adviser: Doctor Joaquim Armando Pires Jorge

Jury

Chairman: Dean of the Technical University of Lisbon

Members: Doctor Alan Dix

Doctor Nuno Manuel Carvalho Ferreira Guimarães

Doctor Arlindo Manuel Limede de Oliveira

Doctor Joaquim Armando Pires Jorge

Doctor Manuel João Caneira Monteiro da Fonseca

June 2007

Resumo

As formas tradicionais de organizar o crescente número de documentos pessoais com que temos que lidar, baseadas na sua classificação em hierarquias, são cada vez menos eficazes. Efectivamente, as soluções actuais para a sua recuperação são limitadas, consistindo preferencialmente em pesquisas de palavras-chave. Todo um conjunto de informação autobiográfica, proveniente do contexto criado pelas interacções entre os utilizadores e os seus documentos, é ignorado. Essa informação é, no entanto, mais relevante e fácil de recordar pelos utilizadores do que os indícios fornecidos pelo sistema operativo.

Apresentamos um novo paradigma de interacção, Interfaces Baseadas em Narrativas, que permite a recuperação de documentos pessoais através de histórias contadas pelos utilizadores descrevendo-os. As narrativas permitem estruturar e recordar informação autobiográfica. Um conjunto de entrevistas permitiu-nos conhecer qual a sua forma e conteúdos típicos. Estudos com utilizadores mostraram ser possível contar histórias ao computador como se de um ouvinte humano se tratasse, que a informação sobre os documentos nelas contida é verídica, e que possuem um poder discriminativo suficiente para identificar documentos concretos. Adicionalmente, verificou-se ser possível a eficaz recuperação de documentos textuais e não textuais usando esta abordagem, demonstrando assim a sua validade.

Abstract

Traditional ways to organise the documents we must daily deal with are becoming increasingly ineffective. Indeed, hierarchic classification of all those documents is cumbersome, and current retrieval solutions, based on keyword search, are limited. They disregard a wealth of autobiographic information about the users and their documents, arising from the personal context of past transactions between them. However, that information is more relevant and easier to recall than are the mechanisms provided by operating systems.

We present a new interaction paradigm, Narrative-Based Interfaces to support users retrieving personal documents by telling descriptive stories. Such narratives make it easier to remember relevant information, by structuring it in meaningful ways. A set of interviews yielded intimate knowledge of the contents and structures of these document-describing stories. User studies show how it is possible to tell stories to the computer as if it were a human listener, how stories are trustworthy, faithfully describing documents, and how they are discriminative enough to allow specific documents to be identified. Furthermore, we verified that it is possible to effectively retrieve textual and non-textual documents using our approach, thus demonstrating its validity.

Palavras Chave

Interfaces Baseadas em Narrativas, Recuperação de Documentos Pessoais, Informação Autobiográfica, Gestão de Informação Pessoal, Desenho Centrado no Utilizador, Interfaces Pessoa-Máquina

Keywords

Narrative-Based Interfaces, Personal Document Retrieval, Autobiographic Information, Personal Information Management, User-Centred Design, Computer-Human Interaction

Acknowledgements

My most heartfelt thanks to my loving wife, Sonia. Without her continuous and unwavering support, it would not have been possible to get where I am today.

While still unable to read these words, my thanks to my son, Artur, whose timely arrival provided me with renewed motivation, when it began to falter.

My parents and sister (you're next!) were responsible for being the beacon showing me the way that led to this moment. Your help won't be forgotten.

To my in-laws, Fátima, António, and Marco, my thanks for embracing me into your family and supporting me at all times.

Everyone at INESC-ID's Intelligent Multimodal Interfaces Group contributed to this work with their insights. While everyone helped where they could, I extend my special thanks to Manuel João Fonseca, André Martins, Filipe Dias, and Alfredo Ferreira Jr.

Tiago Guerreiro worked tirelessly on the BIRD project, while at the same time helping me create the prototypes that allowed me to complete my research. I could not have done it without you!

Luis Rodrigues, Carlos Rodrigues, and Renata Marin also contributed to the development of the prototype. My thanks to all of you and, in particular, to Luis.

Finally, my most special thanks to my adviser, Prof. Joaquim Jorge. Throughout these last few years, he steered me through the tortuous and difficult path that now ends. He helped me find my way when I was lost, provided me with his most valued and deep insight, and helped remove many obstacles found in this journey while becoming, more than a supervisor, a friend.

This research has been funded in part by Project BIRD, FCT POSC/EIA/59022/2004.

Contents

1	Introduction	1
1.1	Research Goals	4
1.2	Contributions	5
1.3	Document Structure	6
1.4	Projects	6
1.5	Publications	6
1.5.1	International Conferences	7
1.5.2	National Conferences	8
1.5.3	Journal Papers	8
2	Related Work	9
2.1	Existing Approaches	11
2.1.1	Spatial Organisation	12
2.1.2	Chronological Organisation	14
2.1.3	Use of Meta-Data	15
2.1.4	Tagging	25
2.1.5	Beyond the Desktop	29
2.2	Discussion	32
3	Stories About Documents	37
3.1	Procedure	39
3.1.1	Interview Analysis	40
3.2	Results	44
3.2.1	Subject Characterisation	45
3.2.2	Overall Story Properties	45
3.2.3	Story Elements	52
3.2.4	Story Structure	61
3.2.5	Other Results	68
3.2.6	Archetypical Stories	68
3.3	Guidelines	72
4	Crafting the Interface	77
4.1	Pre-Requisites	77

4.1.1	Universality	78
4.1.2	The Importance of Time	78
4.1.3	The Need for Dialogues	78
4.1.4	Other Documents	78
4.1.5	Archetypical Story Structures	79
4.1.6	Adaptability	79
4.1.7	Understanding the User's World	79
4.1.8	Inaccuracy / Uncertainty	79
4.1.9	Reflexive Transitions	80
4.1.10	Visual Memory	80
4.1.11	Don't Rely Solely on Text	80
4.2	The Prototypes	81
4.2.1	Prototype A — "Dialogues and Direct Manipulation"	81
4.2.2	Prototype B — "Fill In the Blanks"	83
4.2.3	The Dialogues	85
4.2.4	Use Scenarios	86
4.3	Procedure	90
4.3.1	Interview Script: Observer	91
4.3.2	Interview Script: Wizard	92
4.4	Results	93
4.4.1	Subject Characterisation	93
4.4.2	Stories	94
4.4.3	Comparing the Prototypes	103
4.4.4	User Reactions and Comments	106
4.5	Discussion	115
4.5.1	Telling Stories told to Computers	115
4.5.2	The Best Prototype	116
4.5.3	Some Improvements to the Interface	117
4.5.4	Conclusion	118
5	The Quill System	119
5.1	The Interface	119
5.2	The Quill Architecture	122
5.3	The Monitoring System	123
5.4	Quill	127
5.4.1	User Adaptability	127
5.4.2	Natural Language Understanding	127
5.4.3	World Knowledge	129
5.4.4	Searching for a Document	129
5.5	The Knowledge Base	130
5.5.1	Representing Autobiographic Knowledge	131
5.5.2	Choosing a Representation Formalism	132
6	Results and Discussion	149
6.1	Story Accuracy and Quality	151
6.1.1	Procedure	151
6.1.2	Storytelling	153
6.1.3	Story Quality	154
6.1.4	Story Accuracy	158
6.2	Discriminative Power and Retrieval Rate	165

6.2.1	Procedure	166
6.2.2	User Profile	170
6.2.3	Discriminative Power of Stories	175
6.2.4	Retrieval Rate	178
6.2.5	User Reactions	183
6.3	Discussion	185
7	Conclusions and Future Work	189
7.1	Dissertation Summary	189
7.2	Final Conclusions and Discussion	192
7.2.1	Benefits	192
7.2.2	Limitations	193
7.3	Ramifications	194
7.4	Future Research	195
	Bibliography	197
	A1 Guideing the Interviews	207
	A2 LFP Dialogues	211
	A3 User Comments about the LFPs	215
	A4 RDF Schemas used by Scroll	223
	A4.1 The Quill RDFS	223
	A4.2 The iQuill RDFS	234
	A5 Interview Transcripts	243

List of Figures

1.1	Gilgamesh Tablet	3
2.1	Pile Metaphor	12
2.2	Data Mountain	12
2.3	Cartographic Visualisation	13
2.4	BumpTop Prototype	13
2.5	Lifestreams	14
2.6	TimeScape	15
2.7	Placeless Documents' Presto	17
2.8	Haystack	20
2.9	Personal Chronicling Tools	22
2.10	MyLifeBits' Timeline View	23
2.11	Stuff I've Seen	24
2.12	Remembrance Agent	29
2.13	Roamware	30
2.14	Satchel	31
3.1	Interviewed Profiles	45
3.2	Story Length	47
3.3	Story Trains per Story Histograms	50
3.4	Element Frequencies for All Document Kinds	54
3.5	Induced Element Percentages by Document Kind	56
3.6	Overall Percentages of Induced Elements	57
3.7	Overall Frequency by Story Element	58
3.8	Frequencies vs. Occurrences	61
3.9	Occurrences / Frequencies Ratios	62
3.10	Element Transitions	63
3.11	Element Transitions (Normalised)	65
4.1	Prototype A: Overall View	82
4.2	Prototype B: Overall View	84
4.3	Prototype A: Storyboard	87
4.4	Prototype B: Storyboard	88

4.5	Using the Prototypes	94
4.6	Subject Age	94
4.7	Story Length Comparison	96
4.8	Element Frequencies Comparison Chart (Overall)	99
4.9	Story Element Occurrence Frequencies in the Previous Study	101
5.1	The Quill Interface	120
5.2	The Quill context menu	122
5.3	The Quill architecture	123
5.4	Inference Rule	131
6.1	Time Spent on Stories	154
6.2	Comparison of the Overall Element Frequencies	157
6.3	Story Accuracy Summary	161
6.4	Corrected Accuracy Values	162
6.5	Sorted Element Accuracy	163
6.6	KB Analyser	169
6.7	Average time per story type	172
6.8	Number of documents per user	173
6.9	Average number of documents per folder	173
6.10	Users per document type	174
6.11	Average number of documents by document type	175
6.12	Discriminative power of stories (overall)	177
6.13	Discriminative power by document type	178
6.14	Story length per document type	181
6.15	Target Document position in suggestion list	182
6.16	Questionnaire Answers	184
A2.1	Author, Purpose, Time, Co-Authors, Place, Other Documents	212
A2.2	Subject, Type, Exchanges, Tasks, Contents, Storage	213
A2.3	Versions, Name, World, Personal Life, Events	214

List of Tables

2.1	Related Work Comparison Table	33
3.1	Interviewed Backgrounds and Professions	46
3.2	Story Lengths	46
3.3	Story Lengths	47
3.4	Story Transitions	48
3.5	Story Train Numbers	49
3.6	Cumulative Story Percents for Story Train Numbers	49
3.7	Story Trains Lengths	51
3.8	Story Transitions in the First and Last Trains	51
3.9	Changes In Subject	52
3.10	Story Elements Frequencies	53
3.11	Induced Elements Percentages	55
3.12	Transitions Between Story Elements	62
3.13	Normalised Transitions Between Story Elements	64
3.14	Transition Probabilities	66
3.15	Symmetry Estimates	67
4.1	Story Lengths	86
4.2	Interviewed Subjects' Background, Age and Profession	95
4.3	Story Length Averages from Element Occurrence	96
4.4	Story Length Ratios	97
4.5	Explicit Order Changes	98
4.6	Comparison Table of Element Frequencies (% of stories)	99
4.7	Binomial test significance values comparing values in both studies	100
4.8	Element Ordering: Recent and Old Documents	102
4.9	Element Ordering: Other Documents and Overall	103
4.10	Questionnaire Results	104
4.11	Prototype Functionalities Usage	105
5.1	Ontology representation languages by paradigm	139
5.2	Concepts	143
5.3	Taxonomies	144

5.4	Relations/Functions	144
5.5	Instances	144
5.6	Axioms	145
5.7	Production Rules	145
5.8	Reasoning	146
5.9	Ontology Editors	146
6.1	Interview Location	151
6.2	Interviewed Subjects' Profiles	152
6.3	Position of Stories in the Interviews by Document Type	153
6.4	Story Length Comparison	155
6.5	Story Length Difference Confidence Intervals	155
6.6	Explicit Order Changes	155
6.7	Element Frequency and How it Differs from Previous Results	156
6.8	Relative Importance of Elements	158
6.9	Story Accuracy Summary	161
6.10	Overall Numbers of Verified Elements	162
6.11	Corrected Accuracy Values	162
6.12	Elements Sorted by Accuracy (Overall)	164
6.13	User Profiles	170
6.14	Interview Location	171
6.15	Story order in the interviews	171
6.16	Information Sources	172
6.17	Indexed email statistics	175
6.18	Indexed people statistics	176
6.19	Discriminative power of stories	177
6.20	Discriminative power by document type	178
6.21	Story classification regarding document retrieval	179
6.22	Textual and Non-Textual Retrieval Rate	180
6.23	Story elements that correctly identified documents	183
6.24	Story elements that correctly identified documents, by document type	183
A3.1	Prototype A: User Comments	217
A3.2	Prototype B: User Comments	217
A3.3	Prototype A: Notes	219
A3.4	Prototype B: Notes	222

A good story cannot be devised; it has to be distilled.

— *Raymond Chandler, 1947*

1

Introduction

Two tasks that no computer user can circumvent are to organise and later retrieve electronic documents. What is more, the numbers and types of documents users must deal with have become larger and more diversified. Not long ago, due in part to technological limitations, the average user had to deal with hundreds of documents, that were mainly text-based. With the advent of more powerful computers, larger storage media, and broadband Internet access, personal documents now number in the thousands, of all kinds of multimedia formats, such as images, video, and songs. Compounding on the problem, computers now pervade our everyday lives. New mobile devices made it possible for users to create and edit documents on hitherto impossible times and places.

Unfortunately, existing tools and techniques to manage electronic documents have not evolved at a similar rate, remaining essentially identical to what existed decades ago. They are based in the classification of all documents into a user-defined hierarchy, the filesystem, where each document is stored as a file, that must be given a filename. So deep-rooted is this organisation paradigm that it can be found in most facets of computer use, from World-Wide-Web bookmarks to email messages. Despite its widespread use, such an organisation approach is fraught with problems. It is often unclear when storing a document in which category to place it. Frequently, more than one appears adequate. Other times, none of the existing ones seems to apply. Deciding what to do in those cases is hard. It unduly increases the users' cognitive loads as they know that a bad decision at storage time might compromise a document's retrieval later on.

New approaches to solve the problem of organising and retrieving personal documents are, thus, sorely needed. However, many of the research that has been taking place does not properly address these problems. Indeed, while it would seem that works in the field of Information Retrieval might provide answers to the users' current needs, this is not the case. While helpful, those works deal primarily with the retrieval of data from non-descript document collections, such as electronic libraries or the World-Wide-Web. They focus on helping users to find documents on a certain subject or with certain contents. The problem that users nowadays face is different in that they want to retrieve their *personal documents*. Those, by their own nature, possess specific characteristics, not taken into account by "traditional" retrieval approaches.

Users have previous knowledge of personal documents, having interacted with them in the past. Thus, they can make use of a wealth of information besides the documents themselves. This autobiographic information stems from the shared context in which both the users and their documents are immersed. It might include important elements such as why a document was created or read, easier to remember than those thrust upon users by operating systems and applications. Retrieval strategies in which only a document's contents or location in the hierarchy are used do not fully realise the untapped potential of the autobiographic information users might recall. Even recent desktop search systems, such as Google Desktop, focus on document search based on keywords they might contain. This limits it to documents from which text can be extracted. Historically, this is understandable, as Google was initially designed for Web search. However, we must go beyond this state-of-the-art in order to fully address and solve the problem of personal document retrieval.

In this document we show how narratives are an effective tool for document retrieval applications. We describe the requirements of the interface to allow users to tell their stories unhindered and what is necessary for machines to understand such stories. Indeed, results garnered while evaluating a prototype narrative-based retrieval application strongly support the usefulness and adequacy of narratives to find personal documents where current approaches might prove inadequate.

A distinctive aspect of personal document retrieval is that most traditional evaluation strategies for information retrieval do not apply. Those that do become extremely difficult to assess. Traditionally, document test sets are used for evaluation of such techniques. The documents in those sets are pre-classified, being known beforehand which would satisfy certain search criteria. This makes possible the calculation of measures such as precision and recall, defined as the percentage of relevant documents returned by the retrieval tool that are relevant, and the ratio between the number of relevant documents returned and the total number of relevant documents in the test set. When dealing with personal documents, it is not possible to, beforehand, prepare a document test set. The only non-biased way to evaluate a retrieval strategy is to resort to the users' own documents. Only about those documents will they possess additional autobiographic in-

formation. Consequently, recall is impossible to measure, since the researchers will never know how many documents would be relevant. Precision, on the other hand, makes no sense. The users are not looking for any document that satisfies some criteria. They want to find specific documents. Thus, for personal document retrieval, precision is equivalent to retrieval success.

All considerations above lead to the conclusion that personal document retrieval with the help of autobiographic information is a different problem than those addressed by most previous studies. New techniques and approaches needed to be studied and developed. As part of the emerging field of Personal Information Management, the development of those techniques had to take into account the users' real needs, involving them at all moments in the research, so that relevant autobiographic information might be used effectively.

Some recent works, described in Chapter 2 try to make use of autobiographic information. However, it appears not as the central organising principle of the proposed solutions, but as one more aspect to consider. Often, only some information elements are used, and in no case was a study performed to identify what relevant information can be used. Finally, existing works lack an adequate interface that allows users to make use of relevant autobiographic in an effective and straightforward way. We created a new interaction paradigm, *Narrative-Based Interfaces for Document Retrieval*, that makes it possible.

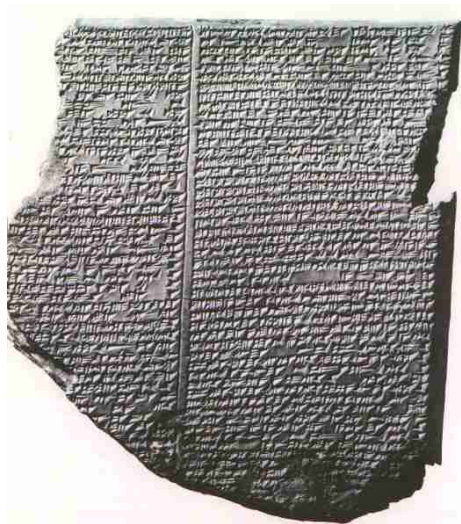


Figure 1.1: Gilgamesh Tablet

Using narrative-based interfaces, the users are able to tell stories about their documents to mention all autobiographic information that might help the system find those documents. Narratives provide a way to interrelate all relevant autobiographic information into a coherent whole. Instead of having to recall unconnected information tidbits, telling stories is a simple and natural way to convey that information to the computer. Indeed, all humans are natural born storytellers. Anthropologist Donald Brown found storytelling to exist across all races and cultures, being as basic to humans as childcare and fear of death [Brown 1991]. We tell stories from childhood to old age. So important are stories that

one of the first written texts produced by mankind, a Sumerian clay tablet, tells a story: the epic of Gilgamesh (Figure 1). It is even not uncommon for us to tell stories about our documents when searching them and someone sits at our side. An interface that is able to understand those stories and make use of the information in them is both natural and easy to use. As stories are a coherent whole, they capitalise on the users' associative memories and help them remember more and better autobiographic information.

1.1. Research Goals

Our main research goal was to *study narratives as the basis for the creation of interfaces that allow users to effectively retrieve their personal documents*. We expected to show that, using narratives, the users can convey enough information to find the personal documents they seek.

In order to achieve our primary goal, it was necessary to study different aspects of narratives about personal documents. Thus, we also expected *to be able to study narratives about documents in order to obtain a thorough characterisation of those narratives*, that might serve as the basis for the design of a narrative-based interface for document retrieval. Based on that description of narratives, we *find an adequate shape for narrative-based interfaces for personal document retrieval*, pointing to the main issues and constraints such interfaces must follow in order to successfully allow users to tell their stories.

Evidently, even if the interface was properly designed we needed to answer more questions regarding the inherent merit of narratives themselves. Stories might be told with little or no problems using the interface but might not be suited for document retrieval if they did not fulfil our expectations in terms of quality. Thus, another of our research goals was to *verify to what extent are stories about personal documents trustworthy*. Only by conveying truthful and accurate information can they be used to identify documents. If they had been found not to be completely trustworthy, we would have had to see to what extent it is possible to deal with the incorrect information in them.

Even if stories prove to be accurate and truthful, they might still be of little use to identify documents. This might happen if they fail to discriminate between documents. Consequently, we had to *determine the discriminative power of stories about documents*, verifying if stories are specific enough to pinpoint documents or if, on the other hand, they do not convey enough information to tell them apart from similar ones.

All the aforementioned goals must be achieved in order for stories to be considered an adequate interaction paradigm for document retrieval. However, there is a practical, engineering, aspect that must also be studied. Even if narratives, in theory, can be used by users to describe their documents, in practice it might be impossible to build a system to do so. For example, it might be impractical to collect all relevant autobiographic data, or it might be impossible to understand key elements in stories. Thus, we had a final research goal: *show that a narrative-based application for personal document retrieval can be built*.

To reach this final goal we needed to create a prototype of such an application and measure the actual success rate of document retrieval using it to find real documents of real users.

1.2. Contributions

This dissertation shows that stories are an effective device for retrieving personal documents. The work that led to that result also yielded the following contributions:

- *A new interaction paradigm, narrative-based interfaces*, applied to document retrieval. While many works in the area of automated storytelling exist, they are primarily concerned with the creation of stories by the computer. We show how the computer can understand stories told by users in order to perform a specific task: retrieve personal electronic documents.
- *An in-depth analysis of stories about personal documents*, in which not only their contents but also their structure were studied. This analysis can be used as the basis for the creation of narrative-based interfaces for document retrieval, but it can also be read just as a study of relevant autobiographic information users recall about their documents.
- *A set of guidelines for designing narrative-based personal document retrieval interfaces*, based on the users' stories, that points to relevant issues to be taken into account when designing such an interface.
- *A methodology for the analysis of autobiographic stories*, to understand what information users remember about them and in which ways. The procedure employed to perform the analysis of narratives about documents to use them for their retrieval can be replicated for other domains, beyond personal documents, as long as there is a story to be told.
- *A methodology for the evaluation of narrative-based interfaces* in particular, and personal information management solutions in general. Specific challenges such as having to deal with the users own documents make traditional evaluation methods inadequate. We show how the evaluation of a narrative-based interface can be successfully performed.
- *A prototype interface and system* for document retrieval, showing how the interface design guidelines can be instantiated. Different rounds of user evaluation in which stories were told to the interface to retrieve documents allowed to demonstrate the validity of our approach. Furthermore, this prototype proves that the related practical engineering challenges can be met.

1.3. Document Structure

To prove that narratives can be a basis for personal document retrieval, it is necessary to fully understand stories about such documents. In Chapter 3 we will describe a study in which this was done, culminating in a set of guidelines for the design of narrative-based document retrieval interfaces. Chapter 4 described our choice of an interface embodying those guidelines, with the help of the evaluation of low-fidelity prototypes. The best prototype led to the creation of the Quill system, described in Chapter 5, necessary to perform more user studies. We discuss them in Chapter 6, allowing us to prove that stories can, indeed, be told by users as a means to find their documents. Finally, in Chapter 7, we conclude by pointing the major results of this research, its strong and weak points, and indicating possible directions for future work in the area.

1.4. Projects

The following projects stemmed from the research described in this dissertation. Some of the implementation required for the creation of a prototype is a direct result of those projects:

- “Biógrafo Pessoal”, final course project of Luis Rodrigues and Carlos Rodrigues, where issues regarding the automatic gathering of autobiographic information were studied. September 2004 - September 2005
- “Documentos Virtualmente Reais”, final course project of Renata Marin, addressing the gathering and integration into the set of autobiographic data of information about paper documents using RFID technology.
- “BIRD - Biographic Information for the Retrieval of Documents”, Foundation for Science and Technology Project POSC/EIA/59022/2004, April 2005 - September 2007.

1.5. Publications

This is a list of all peer-reviewed publications describing the research that led to this dissertation.

1.5.1. International Conferences

- Daniel Gonçalves, Joaquim A. Jorge, Ubiquitous Access to Documents: Using Storytelling to Alleviate Cognitive Problems, in Proceedings HCI International 2003 - 10th International Conference on Human - Computer Interaction, vol. 4, pp. 374-378, 22-27 de June 2003, Crete, Greece.
- Daniel Gonçalves, Joaquim A. Jorge. Analyzing Personal Document Spaces. In Adjunct Proceedings HCI International 2003, pp.161-162, 22-27 June 2003. Crete, Greece.
- Daniel Gonçalves, Joaquim A. Jorge, An Empirical Study of Personal Document Spaces. In Proceedings DSV-IS'03. Lecture Notes on Computer Science, Springer-Verlag vol. 2844, pp. 47-60, 6-9 June 2003, Funchal, Portugal.
- Daniel Gonçalves, Joaquim Jorge, Describing Documents: What Can Users Tell Us? In Proceedings International Conference on Intelligent User Interfaces (IUI'2004), ACM Press, 13-16 January 2004, Funchal, Portugal.
- Daniel Gonçalves, Joaquim Jorge, Telling Stories to Computers. In Proceedings CHI2004, ACM Press, 27-29 April 2004, Vienna, Austria.
- Daniel Gonçalves, Joaquim Jorge, Telling Stories With Dialogue Boxes to Retrieve Documents. In Proc. TIDSE2004, Lecture Notes on Computer Science, Springer-Verlag vol.3105, pp.195-206, June 2004, Darmstadt, Germany.
- Daniel Gonçalves, Joaquim Jorge, "Tell Me a Story": Issues on the Design of Document Retrieval Systems. In Proceedings DSV-IS'04, Lecture Notes on Computer Science, Springer-Verlag, July 2004, Hamburg, Germany.
- Daniel Gonçalves, Joaquim Jorge, Evaluating Stories in Narrative-Based Interfaces In Proceedings International Conference on Intelligent User Interfaces (IUI'2006), ACM Press, 29 Janeiro - 1 February 2006, Sydney, Australia.
- Daniel Gonçalves, Joaquim Jorge, Quill: A Narrative-Based Interface for Personal Document Retrieval, In Proceedings CHI 2006, ACM Press, 22-27 April 2006, Montreal, Canadá.
- Daniel Gonçalves, Tiago Guerreiro, Renata Marin, Joaquim Jorge, Using Autobiographic Information to Retrieve Real and Electronic Documents. HCI International 2007, July 2007, Beijing, China.

1.5.2. National Conferences

- Daniel Gonçalves, Joaquim Jorge, Avaliação de Protótipos para Recuperação de Documentos usando Narrativas. In Proceedings Interacção 2004, July 2004, Lisbon, Portugal.
- Daniel Gonçalves, Joaquim Jorge. Avaliando a Veracidade de Histórias sobre Documentos, In Proceedings 13º Encontro Português de Computação Gráfica, October 2005, Vila Real.
- Renata Marin, Daniel Gonçalves, Tiago Guerreiro, Joaquim Jorge, Bridging the Gap Between Real and Electronic Documents Interacção 2006 - Segunda Conferência Nacional em Interacção Pessoa-Máquina, October 2006, Braga, Portugal.
- Pedro Bilo, Daniel Portásio Gonçalves, Manuel João Fonseca, Daniel Gonçalves e Joaquim Jorge , Above and Beyond: Outer Space as a Metaphor for Document Visualization and Browsing, Interacção 2006 - Segunda Conferência Nacional em Interacção Pessoa-Máquina, October 2006, Braga, Portugal.

1.5.3. Journal Papers

- Daniel Gonçalves, Joaquim Jorge, Evaluating the Accuracy of Document-Describing Stories, In Special Issue of VIRTUAL Journal: "Advances in Computer Graphics in Portugal 2005". (ISSN: 0873-1837)

2

Related Work

In some ways, it is nowadays harder for users to search a document on their own computers desktops than it is to search the Internet. Until recently, most research was focused on the general problem of finding relevant information in large shared collections, instead of helping users find their own documents. Mainly, this occurred because the amount of information created and directly handled by a single user was comparatively small. As such, it was within the abilities of the average person to adequately manage and retrieve their documents. However, with the recent dissemination of computers and applications, this is no longer the case. Not only do our daily activities lead to the creation of large numbers of documents, we have started to record things speculatively, on the off chance that they might later be needed. While not long ago 1TB of storage was estimated to be enough to store all information generated by a user throughout his or her life, this is no longer the case [Gemmell et al. 2006]. With large quantities of information and a lack of proper support to manage it, it is often the case when it appears that “stuff goes into the computer and doesn’t come out” [Boardman and Sasse 2004].

As this trend progressed, it soon became clear that studies on how to help users organise and retrieve personal information were necessary. One of the first works to perform such a study was made by Thomas Malone [Malone 1983], looking at paper, rather than electronic, documents. He identified two main organisation schemes: files and piles. Files contain several documents that share some feature. They possess a name or tag that reflects the common subject of all documents in them. On the other hand, piles are just

unstructured collections of unsorted documents, with no apparent relation between each other. Notwithstanding their apparent lack of organisation, piles are often used. Many users resort to them due to the classification problems posed by files. Other cues besides a document's subject, such as their visual look and position in the piles are then used to retrieve them. As time went by, other works reinforced these conclusions for other kinds of documents.

One of Deborah Barreau and Bonnie Nardi's [Barreau and Nardi 1995] user study found the same problems and made clear that a document's spacial location is an important factor in its retrieval. This corroborates Malone's findings: just by looking at a pile users can remember the documents therein.

Rodden's work [Rodden 1999] [Rodden and Wood 2003] shows similar patterns to apply for digital photographs, making clear the importance of user annotations of their photos as a source of semantic information that can be used to retrieve them. The visual aspect of the photos and the time at which they were taken were also considered to be important.

Peters [Peters 2001] [Peters 2002] studied the organisation of personal electronic information and found existing tools and applications provide insufficient support for the users' preferred organisation tasks.

Despite the fact that Malone's work is now over two decades old, only recently has the matter been consistently and repeatedly addressed, in the emerging field of *Personal Information Management (PIM)*. In it, the main concern is to study how users handle their information, and to support their tasks in doing so across all applications. Indeed, classification and retrieval problems become evident not only for documents but also for other domains where users manage their personal information. The organisation of Web bookmarks is done much in the same way in common web browsers. Hierarchies are used to store them, leading to the same problems found for documents [Abrams et al. 1998]. The management of email messages also suffers from this problem. It is typical even for average users to receive dozens of email messages each day [Bälter and Sidner 2002]. Those messages must be somehow organised, in order to remain useful. However, many users tend to keep large numbers of messages in their inboxes due to classification difficulties. Again, this is due to the fact that traditional email clients favour an hierarchic organisation. Furthermore, this multitude of different unrelated applications in which different kinds of personal information must be handled is one of the major problems faced by PIM researchers: the fragmentation problem [Bergman et al. 2006][Karger and Jones 2006]. While there is relevant information about a project or activity in different applications (filesystem, email client, etc.), there is no way to organise it all in one place, leading to parallel classifications, inconsistencies, and an inevitable duplication of work by the user.

While hindered by classification problems, email has an useful property bookmarks and

documents do not: it naturally associates each message to contextual autobiographic information, that makes messages easier to recall and find. For example, all messages are associated to a sending date and time, a sender and a subject. This information allows users to find their messages without having to classify them. In fact, so efficient is the usage of such information that some studies [Whittaker and Sidner 1996] show users resort to their email clients as a replacement for other tools. An inspection of the users' inboxes revealed not only email messages but also to-do lists or unread documents. Even though email clients don't directly support these tasks, their use of autobiographic information makes them better suited to organise and retrieve information, other than email messages, than special-purpose systems applications. The importance of information such as when a document was created and when it took place has been repeatedly verified [Dumais et al. 2003]. It can serve to help users find their documents by having a far richer information content than that which can be extracted from keywords entered in small text boxes, the traditional and currently preferred way to retrieve documents [Dumais 2007]. It can also help the users understand the information they are handling and move step-by-step toward a desired document [Teevan et al. 2004], instead of relying solely on search.

The widespread dissemination of mobile computing devices and ubiquitous computing will further compound on these problems. More and more people use Personal Digital Assistants (PDAs) and smartphones. Those devices now have the computing capabilities desktop PCs had just a few years ago. Furthermore, UMTS and Wi-Fi technologies have made it possible for such devices to be online at all times. Wearable computers are also slowly becoming a reality. We are, thus, on the verge of the genesis of a new computational era, with different applications [Abowd and Mynatt 2000] and interaction paradigms [Norman 1999]. One of its features will be the capability of accessing and editing documents at any time and location. This will increase the number of documents each user will handle, but also impose more serious constraints on the effort spent storing and retrieving them. On the other hand, the variety of contexts in which users will handle their documents will provide a wealth of context-based information that should be taken advantage of to help the users with their organisation tasks.

2.1. Existing Approaches

We will now describe the state-of-the art for personal document retrieval, presenting the best known, influential and recent works. Those works can be divided into four classes. Firstly, we will present those that take into account the importance of human spatial skills and try to help users find their documents resorting to their spatial location. In second place, we will describe approaches for whose organising principle is time. Next, we will discuss more solutions that go further in the use of different meta-data elements

pertaining documents. Finally, we will study a set of works that try to help users retrieve their documents using contextual information gathered not from the users' desktops but from the real world. At the end of the chapter the different merits and limitations of those works will be discussed.

2.1.1. Spatial Organisation

Works centred around the spatial organisation of documents try to make those documents recognisable to the user, from their position in space and visual aspect. Their major limitation is the screen real-estate available to do so.

Pile Metaphor

This is one of the first works in which documents are spatially organised [Mander et al. 1992], recognising the importance of piles and spatial memory as found by Malone [Malone 1983]. The authors performed an in-depth study of the usage of piles as an organising principle. Four usage patterns were identified, when trying to find a document: *edge* (a pile's recognising contents by looking at its edge); *restack* (removing documents from a pile and re-stacking them until the target document is found); *hinge* (open the pile in different places to inspect its contents); and *spread*, in which a pile's content is spread around a surface to allow all documents therein to be seen simultaneously. Based on that study, they implemented a pile widget for MacOS's desktop. Several tests showed that, while interested, users felt the approach lacked support for tasks a computer makes possible, such as filtering the information. The metaphor was taken to far.



Figure 2.1: Pile Metaphor

Data Mountain



Figure 2.2: Data Mountain

Data Mountain [Robertson et al. 1998] presents documents (in this case, web page bookmarks) on a perspective 3D plane where they can be positioned by the user. It is supposed that their position will help users recall the contents of documents and the relationship between them. A study helped understand what elements would better help users recognise their documents (abstract, title, page thumbnails, etc.). It was verified that the users' visual memory, taking the document's location into account, helps users find their documents. However, this approach

is limited by the number of documents that can be visualised before cluttering renders it unpractical.

Cartographic Organisation

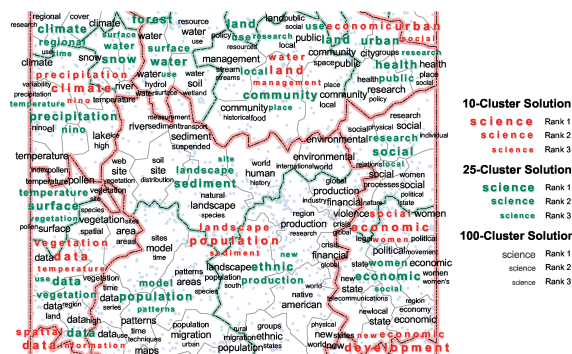


Figure 2.3: Cartographic Visualisation

being part of Kohonen maps, can be spatially organised in a way in which “similar” subjects are close by. Based on that organisation, the author infers polygons that act as borders of “regions” of a given subject. It is then possible to create a hierarchy of similar subjects. “Countries” delimit more generic subjects, while “districts” correspond to more specific subjects in them, and so on. Labels for each region are inferred from the data. While the prototype was created to visualise papers from a digital library, it might conceivably be applied to personal document. While this might not be a problem when searching any document on a given subject, it would make it unpractical to find individual documents, as those are not represented on the map.

BumpTop

Agarawala created a desktop replacement [Agarawala and Balakrishnan 2006] where documents can be positioned as if on a real desktop. A physics simulation engine was used to allow those documents to interact with each other and their environment in believable ways. To capitalise on our visual memory, all documents are presented as thumbnails, whenever possible. They can be tossed around and piled. Piles can be created by “lassoing” a group of documents and can be neater or less organised, as the user wishes. All documents in a pile are

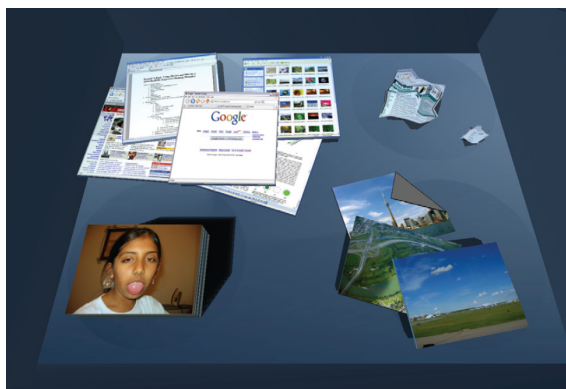


Figure 2.4: BumpTop Prototype

connected to each other by invisible springs, so that when a pile is moved, they remain together. Different interaction techniques and commands were implemented, allowing, for instance, for the user to nudge some documents off a pile to look at them. The interface was optimised for pen-based interaction.

2.1.2. Chronological Organisation

Time-based information can also help users retrieve their documents. That is the underlying hypothesis of works presented in this section, that organise documents chronologically.

Lifestreams

In Lifestreams, documents appear in a sequence or *stream*, partially overlapping their neighbours, with more recent documents closer to the user [Freeman and Fertig 1995] [Freeman and Gelernter 1996] and also in [Fertig et al. 1996]. A Lifestreams version for PDAs was also created [Freeman 1995]. The stream is supposed to contain all of a user's documents, starting with a birth certificate. Sub-streams can be created by filtering documents using different criteria. The contents of these sub-streams are

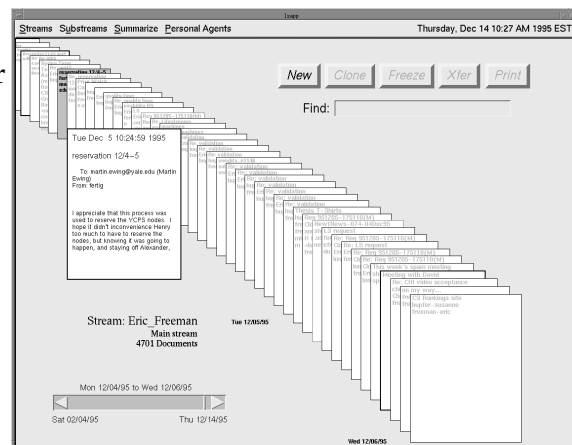


Figure 2.5: Lifestreams

automatically updated to reflect changes in documents. Documents can be created in the future, only appearing when that point in time is reached. This might be useful for the creation of reminders or delayed sending of email messages to other users. In fact, Lifestreams explicitly supports the sending of documents to other users. Those documents will be automatically placed in the recipient's streams appearing only when the appropriate time instant is reached. It is possible to summarise a stream. According to the type of the documents in that stream, a new document of a relevant kind (table, graphic, etc.) will be created to give an overall idea of the stream's contents. It is also possible to associate small programs to different system events. These will automatically be executed when a document is added to the stream or the first time a document is read, for instance. The main limitation of this work is that it is hard to handle streams with large document numbers.

TimeScope

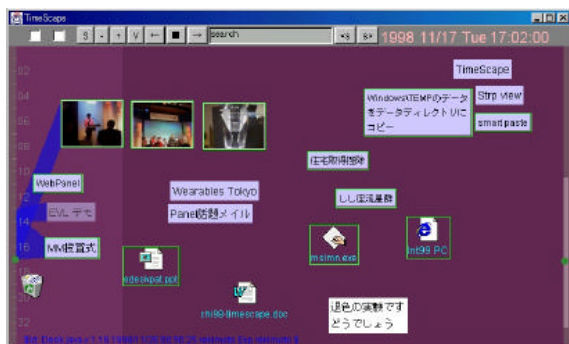


Figure 2.6: TimeScope

TimeScope consists of a desktop where all documents are associated to a given point in time [Rekimoto 1999a][Rekimoto 1999b]. Different documents are shown in the desktop at the date of their creation and progressively fade away as time goes by. Their spatial location in the desktop can be controlled by the user. TimeScope allows travels to the future or past, where users can see their desktop as it presented itself (or

will present itself) at that time. Trips to the future can be used to leave reminders and prepare documents that will only become visible at that time. It is possible to visualise a timeline that depicts the desktop's evolution over time. Keyword search is allowed. The user can seek the time instants in which the keywords appear in documents in the desktop at that moment. User reaction to the system was good. However, it was pointed out that the desktop is not large enough to encompass large numbers of documents.

2.1.3. Use of Meta-Data

Most works centred around the spatial and chronological organisation of documents are somewhat old. Indeed, it soon became evident that relying solely on one organising principle would not be enough. Documents are associated to many different meta-data elements, all of which might help to organise and retrieve them. Thus, most recent studies of personal document retrieval solutions, described in this section, employ meta-data to that end.

Semantic File Systems

One of the first works to use meta-data to organise documents and help users retrieve them was the Semantic File System [Sheldon et al. October 1991]. It assumes documents possess several properties. The values of those properties are automatically gathered by specialised modules (transducers) from different sources (email headers, for instance). Instead of looking for documents in a hierarchic file system, users navigate virtual directories that, actually, are just filter criteria for the different documents. For instance, the `/sfs/owner:/smith` 'directory' corresponds to all documents whose owner is Smith. This work was still based on command-line interfaces and fails to address recent problems such as documents in different locations. However, the concepts it uses are valid,

making it the conceptual basis for other research developed since, and described below.

Other File Systems

Other file systems exist in which it is possible to associate properties to files. It is the case of the filesystem used by the now unused BeOS [Giampaolo 1999] or Microsoft Windows' NTFS [Richter and Cabrera 1998]. Another example is the *Synopsis File System* [Bowman and John 1996]. On this system, each file has associated to it a synopsis (typified attributes). Files can be accessed in a way similar to what was possible for Jouvelot's Semantic File System: document types are used as different document classes and different operations can be performed on files depending on their classes. The idea is to avoid problems when a specific application to deal with some kind of document is not available, by providing default ways to handle those documents. For instance, the display method is responsible by visually rendering the file, regardless of its class, as HTML. Unfortunately, while there are filesystems that support meta-data, it is still necessary for the user to enter it for every file. As few are willing to tackle such a daunting task, this capability is seldom used.

The recent Windows Vista operating system also allows properties to be associated to files and virtual folders to be created based on some criteria, similarly to what was possible using the Semantic File System. It integrates a search tool able to find documents based on those properties.

Tapestry

Another of the more ancient works is Tapestry [Goldberg et al. 1992], a system that can supposedly deal with all kinds of documents but in practice tested only with email and usenet news messages. Documents are associated to properties extracted from the messages' headers, and those properties can be used as the basis for searching and filtering documents. Besides that, it is a collaborative system, in which users can annotate documents and share those annotations with others. This allows trustworthiness or interest ratings of documents. While showing collaborative work can be useful, it performs these tasks in a simple way.

Harvest

Harvest automatically gathers data about documents [Bowman et al. 1995]. It tries to integrate documents from different sources (WWW, FTP or email, for instance), thus solving the Fragmentation Problem. Although relevant for the heterogeneous nature

of the different sources it considers, it does not gather meta-data from the documents themselves and is dependent of auxiliary applications to retrieve those documents, such as Glimpse [Manber and Wu 1994], a system that allows text-based documents to be retrieved based on their contents.

Placeless Documents

A more recent and complete approach is Paul Dourish's Placeless Documents system [Dourish et al. 1999b][Dourish et al. 2000b]. Its main concern is to completely eliminate the need to deal with hierarchic document organisations. It tries to allow users to concern themselves with a document's contents rather than its location (hence the name "Placeless Documents"). A prototype application embodying these principles, Presto, was implemented [Dourish et al. 1999a]. In Presto, documents from different sources such as email messages and other documents produced by the user are managed based on the meta-data gathered about them. For each document type, there exists a specialised service capable of analysing and gathering meta-data (from email headers, for instance). That meta-data is organised as a set of named properties and their values and directly associated to the documents. Properties can also be arbitrarily created (and given values to) by users. No properties have special status. For instance, a file's name is just another property, which may even be absent in some cases. Presto does not store the documents themselves but, rather, a reference to their actual location in the filesystem. This abstraction layer makes it possible to hide that location from users.

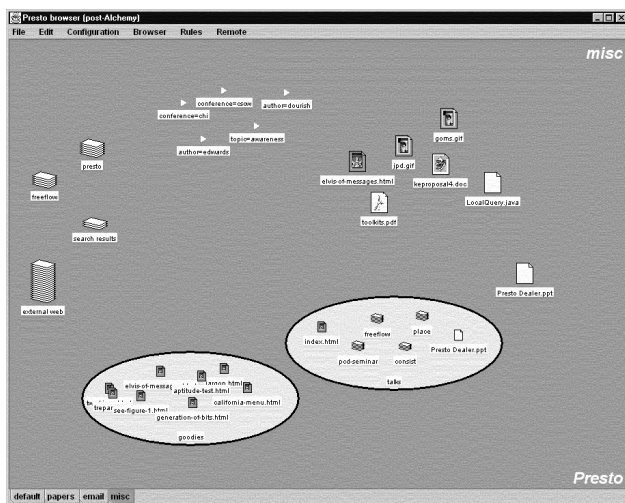


Figure 2.7: Placeless Documents' Presto

Different users can employ different properties and, within limits, access those created by others ("What documents did John flag as interesting?"). Document retrieval is done with the help of boolean search for property values. Also, the interface allows the creation of dynamic document collections. A document collection is nothing more than a persistent search, whose results are continuously visible and updated whenever the document set changes. Furthermore, it is possible to define inclusion lists containing documents

that will always be a part of a collection. Reciprocally, exclusion lists allow documents that should never be part of a collection to be enumerated.

In terms of compatibility, Presto provides a Java API in which the `stream` class is sub-

classified. This new class, when used by applications, automatically registers all relevant data in Presto. To provide compatibility with existing applications, a special-purpose NFS server was implemented. It serves as a mediator between those applications and Presto. Several issues are raised, such as the need to cache the documents. Those questions were considered in related studies [de Lara et al. 1999].

Still within the Placeless Documents approach, it is possible to use active properties [Dourish et al. 2000a]. To each property can be associated computer code that will be triggered in certain circumstances, such as the creation, reading or deletion of documents. Three kinds of code are considered: Validators, that analyse a document and have the power of cancelling the applicability of the property; Performers, that somehow change documents; and Notifiers, that don't change the document or its properties but perform some additional work after those changes occur. There are still the Delegates. An application can ask a document to implement a certain interface (in the Java meaning of the word). If the document itself is unable to do so, a Delegator is created to simulate the desired capabilities. The behaviour ceases to be performed solely by applications, becoming spread throughout the system ("exploded applications").

Presto has already been used in several applications, such as Flatland (an electronic board that registers everything written on it as Presto documents) [Dourish et al. 1999b] and a system to manage documents who have originally been hierarchically classified by a California state agency [Dourish et al. 1999c]. In those cases, the Placeless Documents approach seems promising. Its main problem is the relatively simple way in which document properties are handled. There is no synergy between them, and they are just arbitrary pairs of property names and respective values, that can change from document to document and, thus be hard to remember at retrieval time.

PACO

A system similar to Presto are Baeza-Yates' Persistent Attribute-Centric Objects (PACO) [Baeza-Yates and Terry Jones 1999]. Again, these objects can be retrieved resorting to automatically gathered properties associated to them. It is organised in three layers. The Storage layer contains the document management tasks. Documents can be stored in different ways (in the filesystem, databases, etc.) and places. This layer handles that diversity ensuring it is transparent to the user. The Representation Layer generates the attributes that are associated to objects. The Relationship layer allows relationships between objects to be established. Searches and other high-level tasks are also performed in this layer. Objects are nothing more than property sets, that may or not have contents. PACO does not handle content in any special way, but rather as another property. A same object can even have different content properties (different versions of a document, for instance).

This system was used to create Terry Jones' Attribute Value Systems [Jones 1998]. This system specifically handles documents. Each document is, as expected, associated a property set and its retrieval is achieved by performing queries on the values of those properties. Users don't deal directly with objects but, rather, with properties. It is possible to grant different permission levels to properties, making them public or private to other users. Since retrieval is possible only through the manipulation of properties, if an object has all its properties removed it will become invisible.

Two other applications that stemmed from PACO are DomainView and Known Space. DomainView [Baeza-Yates and Mecoli 1999] is a browser that allows users to visualise documents according to their domain. Each domain is defined from actual property values. Known Space [KnownSpace Group 2002] is, in practice, an application core written in Java that provides an integrated mechanism for the retrieval of document properties (including email messages and web pages). Small applets dubbed "simpletons" can be added to the system to provide specific behaviours (deal with email messages, suggest relevant web pages, etc.).

PACO has essentially the same problems found for Placeless Documents, by treating properties in isolation, without an organising principle that makes their use easier.

Haystack

Another relevant system is Haystack [David Huynh and Quan 2002]. While originally a system written in Perl to index documents [Adar 1996], quickly it grew to address concerns with the lack of gathered information, and allow the specification of inter-relations between objects. It aims to provide an uniform way to represent all different personal information data sources, allowing them to be grouped, annotated and linked regardless of their nature or source.

In a first state, its implementation was based on the construction of an internal representation of a user's document space, in which different objects were interconnected ("THIS document is of THIS type and written by THAT author"). [Karger and Stein 1997] [Adar et al. 1999]. This internal representation was based on a graph. Each node (or "straw") contained an information element, typically belonging to some primitive data type (text, numbers, etc.). There are some differentiated nodes ("Bales"), representing relations among straws (all properties of a document, for instance). The graphs arcs or "Ties" are responsible for the relation itself, and can be labelled according to the attribute they represent (the value of the attribute is the "straw" the "tie" points to).

The information managed by Haystack can have three different origins. The first is information already in the system. For instance, when a new URL is entered by the user, the corresponding document is automatically analysed and relevant properties extracted, if

possible (author, creation date, etc.). The second information source is the monitoring of the users' actions, which are recorded for future reference. This is done by special purpose proxies that stand between the users' applications and Haystack. One example of such information are the times at which web pages are accessed and the links that are followed. The third information source are annotations performed by the users themselves. The retrieval and manipulation of this information can be achieved by different applications that can be modularly coupled to the system.

Haystack follows the "one person, one machine" credo, implying that it runs on each user's machine, rather than online or on some remote sever. This limits Haystack's scope, as it cannot deal with document in multiple locations.

The system's first version was implemented in Java. The information handled by it was represented directly by Java class instances. This was found to be inefficient and inflexible. More recently, and using Semantic Web technology [Berners-Lee et al. 2001b], Haystack was transformed to represent all relevant information using RDF [W3C 2002]. This is a faster and more generic representation. Information is now gathered by specialised agents that can be modularly added to the system.

Furthermore, an interface called Ozone was created. It is called a "semantic interface" as every element (pixel, even) in the interface has some associated semantics. One of the interface's premises is that it does not handle documents in different ways depending on their type. It handles them in an unified way, resorting to their properties and inter-relations to understand which behaviours are appropriate. The interface is automatically created from the objects that should be visualised and the views associated to them. Thus, as already mentioned, all on-screen pixels are generated from some object and have semantics that depend on that object. Context menus display all available options for different object types. For instance, texts can be cut, pasted or spell-checked, Email messages can be replied to, forwarded, etc.

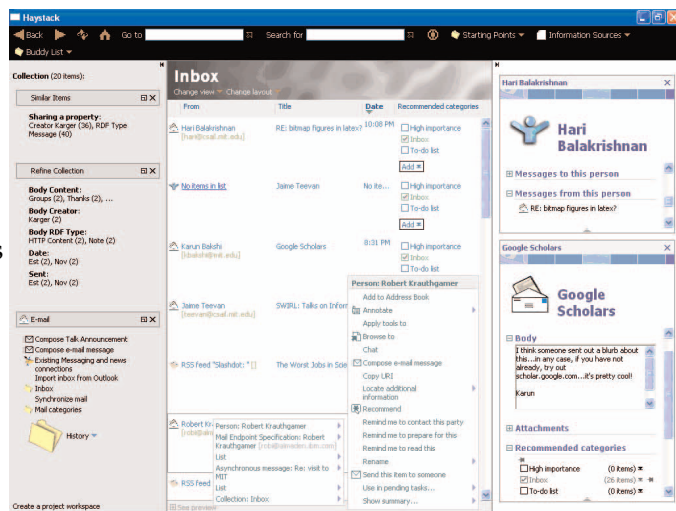


Figure 2.8: Haystack

The interface is adaptable, as it is defined mostly in RDF (mainly, the definition of each object type's views), using a special-purpose ontology. To make it easier to use RDF, a language called Adenine [Quan et al. 2002] was created that allows high-level code to be "compiled" into RDF triples.

Connections

Soule's Connections search tool [Soules and Ganger 2005] combines traditional content-based retrieval with information about the users' activities. The system continuously monitors file system calls on the users' machines. This allows it to understand which files were used at what times. Documents used within a given time-window, are considered to be related to each other. The resulting relation graph allows the results of a content-based search to be refined. For each result of such a search, the graph is traversed in a breadth-first manner, until a certain distance from the original file has been reached. All files found in the traversal are considered to be possible results for the search. The rationale for this is that even if the target document does not contain the keywords used as search criteria, a related document will. This approach's limitation is that it deals only with files. If a user accesses an email message, Connections will only be able to detect that the entire mailbox the message resides in has been used. This leads to the generation of incorrect relation graph edges and the polluting of search results.

Personal Chronicling Tools

Recognising that the users' daily activities can provide rich benefits when trying to retrieve personal information in general and personal documents in particular, Kim developed a prototype framework for the automated creation and maintenance of Personal Chronicles [Kim et al. 2004]. Personal chronicles are sequences of events resulting from the users' activities. An extensible monitoring system continuously records events such as opening and modifying documents, placing content into the clipboard, sentences entered using the keyboard, applications used, instant messages sent and received, etc. The user has the ability to manually annotate all documents, messages, applications or events, resorting to an ubiquitous "T-Button" integrated into the operating system's windows. Both the text automatically collected by the monitoring system for each event and the users' annotations can be used to retrieve those events (some of which pertaining documents) through keyword search (Figure 2.9). The results of that search can then be filtered by specifying values for the different meta-data attributes events might have (email direction, date, etc.). This work lacks proper user evaluation.

FotoFile

FotoFile [Kuchinsky et al. 1999] strives to help users organise and retrieve their personal digital photographs and other digital media. As photographs, audio, and video lack text-based information, they are prime candidates for retrieval techniques that employ different kinds of meta-data. FotoFile supports manual annotation of photographs but

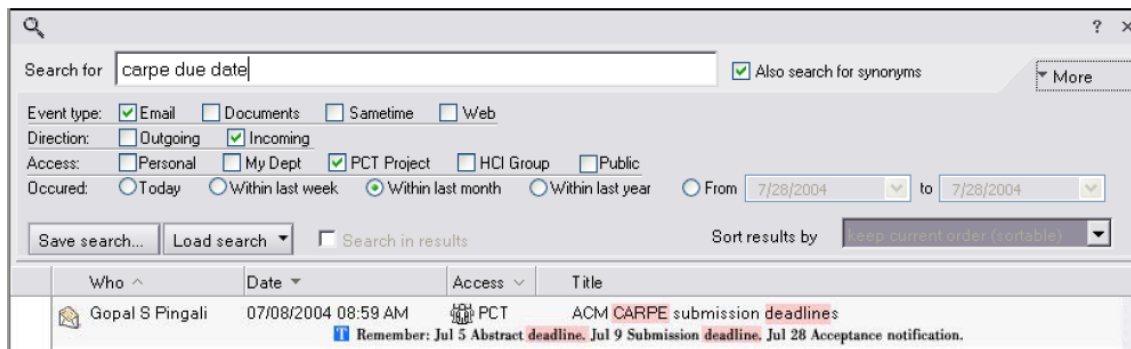


Figure 2.9: Personal Chronicing Tools

also tries to automatically extract features from media files. Face recognition is used to try to identify who is depicted on a photograph. Keyframes are extracted from videos. Once all content has been indexed, it is interactively displayed with the help of an Hyperbolic Tree.

This work shows how content analysis techniques can be used to gather information about documents. However, it lacks an uniform framework to deal with different kinds of documents, and disregards other meta-data that might be available from the context in which documents are used.

PhotoTOC

As was the case for FotoFile, PhotoToc [John C. Platt and Field 2003] also tries to help users retrieve digital media, namely, digital photographs. The photographs are clustered according to when they were created, and other attributes such as their colour histograms. A representative photograph for each cluster is chosen, and a overview+detail view allows users to quickly navigate the different clusters and visualise each one's contents. PhotoToc is thus able to manage large numbers of photographs. While limited to photographs, the evaluation of PhotoTOC shows that the interface plays an important role in document retrieval, as a good interface was necessary to allow users to adequately handle and retrieve their photographs.

MyLifeBits

MyLifeBits [Gemmell et al. 2002] [Gemmell et al. 2006], developed at Microsoft Research, tries to realise Vannevar Bush's vision for Memex, a device in which all of a users' books, records and communications can be automatically stored, allowing for their expedient retrieval. Similarly, MyLifeBits aims to be able to automatically record all information relevant for any given user. It purposefully avoids hierarchies to store documents and

other relevant content, thus trying to overcome the problem that classifying all this information would pose to the user. It uses a database to store content of different types, such as contacts, documents, email messages, events, photos, music and video, each with its own meta-data properties. Items can be linked together if they are somehow related (a photo to the contacts of the persons shown in it, for instance). The resulting inter-relations graph can help users navigate their “bits” of information in search of a specific one. This is a direct consequence of the realisation that keyword search is not enough, and that non-textual information is important. To better handle such information, users are allowed to annotate or tag all information bits as they see fit.

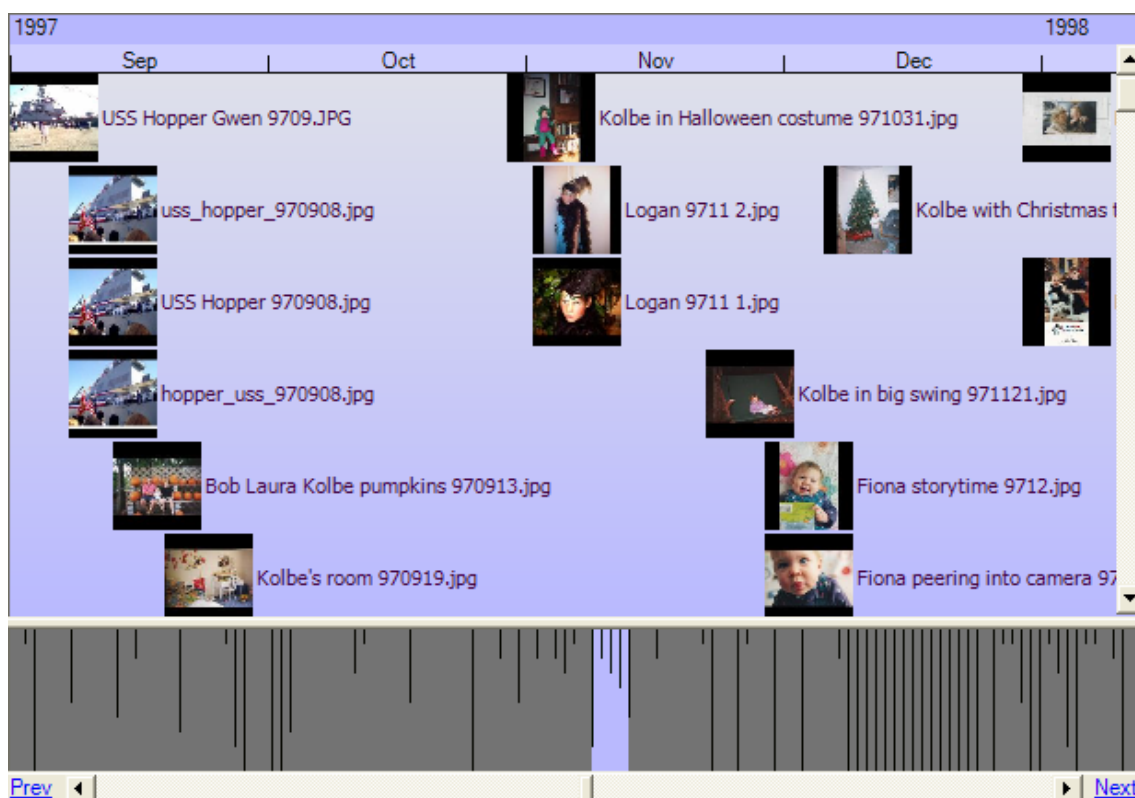


Figure 2.10: MyLifeBits' Timeline View

Users might prefer to reach their documents in different ways. Consequently, MyLifeBits supports several views of a users' “bits”. Four such views are the Thumbnail View (miniature images of all resources shown in a grid), Detail View (a list of resources and their properties), Timeline View (all resources shown on a time scale), and Clustered Time View, in which resources handled around a given time are clustered together on the same place of a timeline. Time plays an important role in the different views, being recognised as an important attribute of all resources.

Lookup by time is one of the most effective ways to look for a specific resource using LifeBits. The different views and underlying data model directly support it in efficient ways. It is also possible to follow relation links from any resource that possesses them, to reach other relevant “bits”. Keyword search is also possible. Finally, the user can filter

the resources displayed on a view by specifying criteria involving their meta-data.

MyLifeBits is extensible. Since the ultimate goal of the approach is to capture every bit of information relevant to a user, it allows for information from new sources to be inserted into the database. As our lives are not spent in front of a computer, external information sources are contemplated. It is the case of SenseCam [Gemmell et al. 2004]. SenseCam is a device worn at all times by the user that captures a digital photograph of whatever the user is looking at every second. Audio is also captured, as well as the users' physical location, with the help of GPS data. All this information is entered into MyLifeBits. From it, meaningful relationships with other information bits can be established.

Stuff-I've-Seen

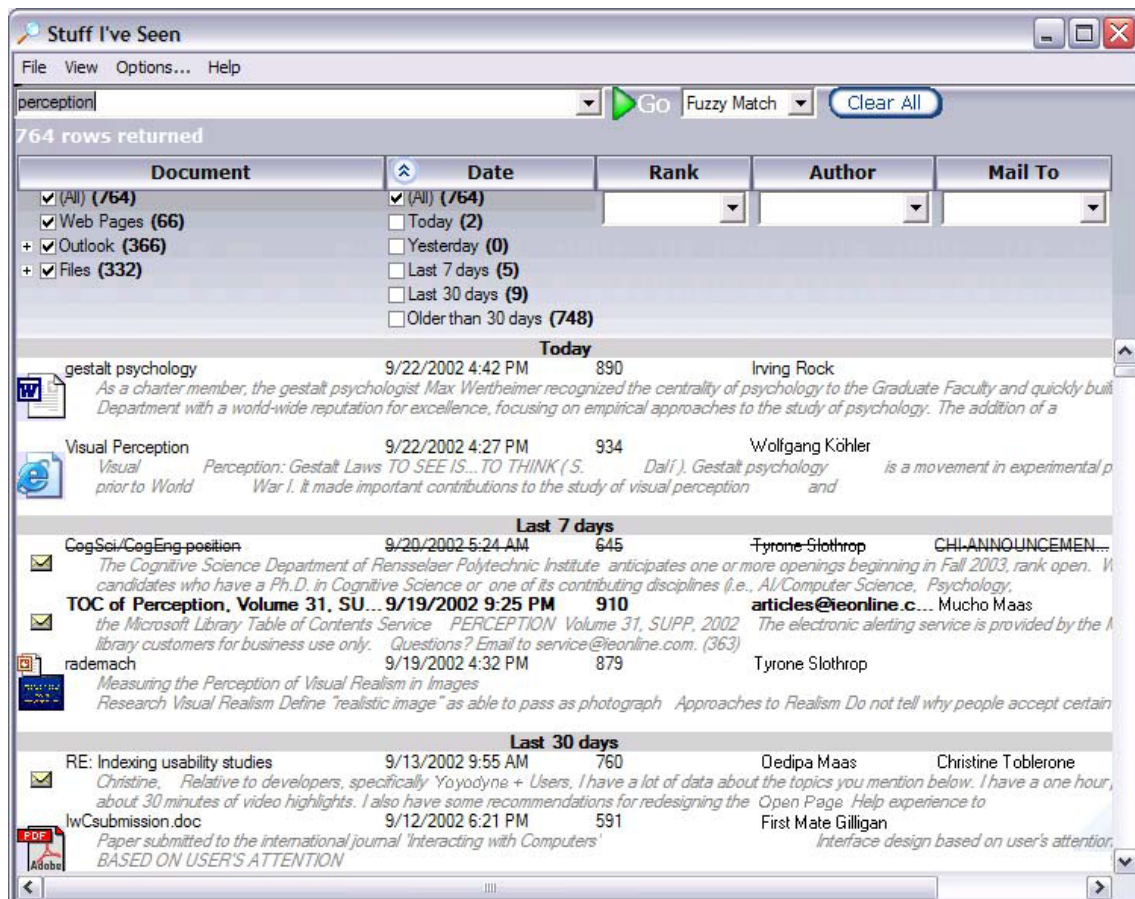


Figure 2.11: Stuff I've Seen

Rather than focusing on information retrieval, Stuff-I've-Seen is concerned primarily with *information re-use* [Dumais et al. 2003] [Cutrell et al. 2006a]. The authors recognise the existence of rich contextual cues for information the user has handled sometime in the past, and try to make use of those cues on an interface that allows it to be found again. This is accomplished by building an integrated index of the different information

in a users' computer: documents, web pages, email messages, etc. Meta-data is collected for all those elements.

Information can be retrieved with the help of an interface that allows users to search the unified index. The retrieval is done iteratively. First, users can perform a keyword search to find information elements that contain those keywords. Different properties can be shown about each element, as different columns. By default, elements are sorted by date, although clicking on each column's header can sort results by other criteria. Search results can be refined by filtering the possible values of the different meta-data columns. This filtering is performed with the help of combo-boxes or checkboxes created automatically from the possible values of those properties. Using those interface controls, users are able to properly restrict the values of those properties, thus narrowing down the search results.

More recent versions of *Stuff-I've-Seen* include a Landmarks Interface, that present users with a timeline view of all their information elements [Merrie Ringel and Horvitz 2003]. This was done after verifying that time is an important element often recognised by users. However, it was witnessed that users might not remember the actual date a document was handled on, but might recall instead a episode around that time, such as a child's birthday or the 9/11 attacks. Thus, both public and personal landmarks are used to annotate the timeline, hoping to make it more relevant to users. Public landmarks include holidays and news headlines, and personal landmarks are inferred by the users' calendar appointments and digital photographs.

The system's major problem resides on the limited number of meta-data properties that can be displayed simultaneously for each element. Furthermore, the users are required to begin their retrieval process by making a keyword search. Other meta-data might be present in the system that is better remembered by a user regarding a particular document or email message, but it is only possible to filter meta-data to refine previous results, instead of searching with its help.

2.1.4. Tagging

Recently, tagging has become a popular way to annotate information. It is used online on many Web2.0 applications, such as gmail (email messages), del.icio.us (bookmarks), CiteULike (scholarly papers), Flickr (digital photographs), YouTube (video), Last.fm (music), Odeo (podcasts), and Technorati (blogs). Users tag their content, and that of others, thus providing semantics that might otherwise be hard to discover. The different tags can then be the basis for keyword searching or browsing, allowing users to retrieve their own content and resources similarly tagged by others.

While tagging is increasingly used, there are still few peer-reviewed papers that analyse

and study it in a formal way. Most resources about tagging are available online, as blog posts and similar articles. Nevertheless some works have begun to study the cognitive processes involved in tagging, and its actual effectiveness.

The popularity of tagging might be explained by the relative ease with which it is done. Tagging is easier than classifying resources into a hierarchy. Hierarchic classification is exclusive, meaning that typically a resource can only be stored in one location of the hierarchy. When tagging, users can associate several tags to the same resource. It is no longer necessary to choose from the different possible classifications the one better suited to describe it [Golder and Huberman 2006]. This alleviates the cognitive load felt by users when managing their resources, as it removes the fear of choosing a bad classification that could result in a resource being lost forever [Sinha 2005].

Marlow [Marlow et al. 2006] states that tagging can be used for different reasons: future retrieval, contribution and sharing (adding to other resources on the same subject available to a community), attract attention (to get people to look at one's resource), play and competition (in games, often collaborative, whose goal is to produce tags according to some set of rules), self presentation (using tags to make oneself known to others), and opinion expression (sharing judgement values about a resource with others). Of all those reasons, only future retrieval is relevant for Personal Information Management. Tagging is useful mainly in collaborative, shared, environments, rather than personal use. Even so, it could be the basis for the design of a personal document retrieval application. However, there are several problems that make it unsuited to that task.

The three more serious problems with tagging are *polysemy*, *synonymy*, and *basic level variation* [Golder and Huberman 2006]. They all stem from the unrestricted way in which tags can be applied.

Polysemy occurs when tags contain words that might have more than one meaning. For instance, "mouse" might be referring to a computer peripheral or a mammal. Using the same word to tag resources about its different meanings will result in irrelevant results at retrieval time.

Synonymy originates from the use of different words with the same or similar meanings. Different resources about the same piece of hardware might be tagged with "pipe" or "tube". This will lead to incomplete query results.

Finally, basic level variation is a problem arising from the fact that most resources can be classified in a continuum semantic space. For instance, my aunt's cat could be tagged as "Mr.Whiskers", "cat", "feline", or "animal". All tags are correct, but the heterogeneous choice of tags on different levels will difficult the retrieval of related resources. In short, tagging exacerbates the problems associated with the fuzziness of linguistic and cognitive boundaries.

These and related problems explain why studies of tagging found a very low tag re-use. One such study found that each tag is employed, on average, by only 1.73 [Sen et al. 2006] users. Even so, this occurred only given the right usage context (values as low 1.10 as can be found otherwise). This shows how it is hard for different users to agree on tags, as becomes evident in collaborative tagging games such as ESP [von Ahn and Dabbish 2004] and Peekaboom [von Ahn et al. 2006]. Furthermore, tag use varies greatly from user to user. "Power Users" can contribute with thousands of tags, while others resort to only a few to annotate their resources [Golder and Huberman 2006]. When looking at tag re-use by the same user, the same study found that a tag might be re-used 14.9 times if it has some kind of personal significance to the user, but only as few as 2.6 times otherwise. As a single user must classify thousands of personal documents, such values show it will be hard to maintain consistency if they are to be tagged. This will, of course, make retrieval harder.

Another problem with tagging is the inherent difficulty of re-tagging. This becomes a problem if the users desire to employ new tags to replace that which was previously classified with some other tag, especially if there are thousands (or more) such resources, as is often the case.

A more serious issue is that tagging systems have problems coping with scale, as it is time-consuming and tedious to tag large numbers of resources. This, and the aforementioned problems can be slightly controlled in shared environments. The burden of tagging is spread by different users, and shared vocabularies might arise, under the form of folksonomies (sets of commonly used tags). In a PIM setting, however, all the tagging must be done by a single user.

The tagging of large numbers of resources, such as the ones automatically garnered by indexing tools in personal information management contexts by a single user can prove to be a daunting, if not impossible, task. Users no longer have to classify all resources into a hierarchy, but are still asked to act as filing clerks [Gemmell et al. 2006].

Finally, tagging in its purest form ignores additional contextual information resulting from user actions or background knowledge [Chirita et al. 2005]. For instance, relying solely on tagging to organise and retrieve email messages disregards meta-data in their headers.

For all these reasons, in the context of Personal Information Management in general and personal document retrieval in particular, tagging is seldom used as the main source of information and principal way to retrieve documents. Many of the property-based approaches described above resort to tags in some way, but always as an accessory (and optional) way to enter information into the system. Even so, for some approaches to personal document retrieval, tagging plays a central role. We will describe them below.

Phlat

Phlat [Cutrell and Dumais 2006][Cutrell et al. 2006b] supports a unified tagging scheme for personal content across storage systems (documents, email, etc.). Having been developed by the same team that created Stuff-I've-Seen, described above, is based on the same assumptions of that system. It was created using the Microsoft Desktop Search index, augmenting it by allowing users to tag all their files and email messages. Those tags are stored not in the index but with the resources themselves, as NTFS properties for files and MAPI properties for email messages.

Querying is performed by entering keywords and also by specifying filters for several meta-data properties. A limited number of properties is considered: Date, Tags, Path, People and Type. The results area is similar to that of Stuff-I've-Seen, with different properties in different columns, allowing results to be sorted according to each of those properties. Tagging can be performed in Phlat by selecting several results and then choosing which tags to apply to them.

Phlat's major problem is that tags can only be entered using it, requiring users to continuously access it for classification, and not only retrieval. However, tagging is only possible *after* retrieval, since only search results can be tagged. Also, only email messages and files can be tagged, narrowing the systems' scope.

A View on Views

Central to Ariel Shamir's approach is the concept of *View* [Shamir 2004]. A View is a workspace where all information about a specific project or activity is stored. By changing views, users would be able to change their work context, solving two problems: scattering (information required for a single activity is often scattered across locations and applications) and cluttering (irrelevant information is often found, distracting the user). Thus, when wanting to retrieve a document related to some project, the user needs only to invoke its respective View. Views can be created at will by users. Each view is associated a set of tags. All elements in a View are automatically tagged with those tags. This provides them with some semantic information. That semantic information is used to compute a semantic distance between elements. Based on that metric, a View displays not only the elements that belong to it (THIS DATA), but also RELEVANT DATA, other objects not explicitly placed in the View but deemed relevant to the activity it pertains.

Unfortunately, items must be manually placed into views, and the tags entered by users are essential to ensure their proper handling.

2.1.5. Beyond the Desktop

The works we will now describe are concerned not only with information gathered from the documents themselves and the users' computer, but with a wider context. They show how that information can help the user find personal information, and provide the means for the computer to better understand which documents might be relevant to the user at any given time.

Factoid

Factoid [Mayo 2001] is not directly concerned with document retrieval. More than an actual system, it is a proof-of-concept prototype. However, its nature makes it relevant in the solving of that problem.

It was conceived as a small passive device that users should carry with them at all times. It registers factoids about the users' activities, their location, etc. This will provide the users with a journal of their activities that other applications can take advantage of. In particular, such contextual autobiographic information can be used as the basis for document retrieval, by helping users recall their documents based on the context that surrounded their interaction with those documents.

Remembrance Agent

The Remembrance Agent is part of Bradley Rhodes' PhD research [Rhodes 2000a] and has suffered a meaningful evolution since its inception. We will describe three different relevant stages, as each embodies a relevant principle.

```

Jimminy-header <e15-384d|josh weaver|Fri, 2 Jan 1998 15:02:28 EST>
Wearables group meeting notes:
|
***Emacs: wearfolk Jan 7, 3:02pm | e15-384d(2) | josh weaver | | (Tex
1: 0.49 thad | 19 Dec 97 | ar paper for mobil networking and app j
2: 0.36 josh weaver | 16 Dec 97 | finger-tracking
3: 0.20 rhodes | 12 Dec 97 | vtt electronics talk
*remem-display*
remembrance agent started.

```

Figure 2.12: Remembrance Agent

In a first stage, [Rhodes and Starner 1996], it consisted of a program associated to the Emacs text editor that ran in the background and monitored all the user read or wrote. From that information, it continuously provided suggestions about relevant documents, based on that which the user was working on. That similarity is verified with the help of a relatively ancient system, SMART, that all rights automatically indexes all documents in the users' workspaces. One of the system's goals was not to be intrusive and preventing unduly distracting the user.

In the system's next version, [Rhodes 1997], context became more than just what a user was reading or writing. It now takes into account the surrounding environment. For instance, at a meeting the system will recognise the participants and suggest information related to them (previous meetings, the notes of those meetings, the participant's birthdays, etc.). The system will also try to suggest documents relevant to the ongoing discussion. This prototype version is based on a *hat-top-computer*, running Emacs, that can be used to enter text and see the suggestions and documents it provides, with the help of a heads-up-display. It uses the tf-idf algorithm [Salton and M.J.McGill 1983] to compare documents and determine their relevance. Different sensors it has access to include a Locator, that provides the users' location with the help of multiple emitters placed throughout the campus. Time sensors allow date and time to be recorded. People the user comes in contact with are identified with the help of Active Badges.

More recently, the different concepts embodied by the Remembrance Agent were applied to the World-Wide-Web, in the form of the Margin Notes system [Rhodes 2000b]. This system alters the WWW browser to allow it to record the user's actions (more specifically, the pages viewed by the user). At any moment, if it finds other documents (other pages, email messages or other documents) it thinks the user might find relevant based on the currently viewed page, it opens a side frame with links to those documents. It was tried to reduce the users' cognitive load when using the system. However, users don't find the suggested documents to be relevant enough to warrant the intrusion.

RoamWare

Roamware [Wiberg 2001] is a PDA-based application that allows sporadic information exchanges between participants in casual meetings (in hallways or in the elevator, for instance). In those unplanned meetings, it is difficult, after some time has elapsed, to remember who was actually present, the matter that was discussed, etc. RoamWare allows this context to be recorded, by communicating with RoamWare devices carried by other participants in the meeting. It allows the exchange of documents or other information between them, and remembers every participant. Privacy problems were taken into account, giving users the possibility to do not record information about specific meetings. Documents can be retrieved at a later time using a desktop application based on the contextual information gathered by RoamWare. While interest-



Figure 2.13: Roamware

ing, this work does not fully address the complexities inherent to the actual retrieval of documents at the meeting's time.

Satchel

Satchel has its roots on the Forget-Me-Not system, a memory helper, consisting on a device similar to current day PDAs, that registered the different events its user was involved in [Lamming et al. 1994] [Lamming and Flynn 1994]. Information such as phone calls, meetings, meeting places, documents used in those meetings or sent to other users was recorded. Forget-Me-Not was able to present that information chronologically, resorting to icons. It was, then, possible to filter the events shown by creating strings of icons representing the different conditions, and find the documents associated to those events.

Satchel [Flynn et al. 2000] went beyond its predecessor, motivated by the desire manifested by users of being able to access their documents at any time, at any place. Its latest version, developed based on different user studies [Eldridge et al. 2000] is implemented on a Nokia Communicator 9000 and consists on a wireless browser that allows users to navigate the web or their documents. The documents are provided by a dedicated HTTP server that has access to the users working areas. It does not resort to the HTML language, but, instead, to Halibut, a more succinct version of that language that allows bandwidth savings. Satchel is not capable of displaying the documents. Instead, it serves as a mediator between devices. Once a relevant document is located using Satchel, it is able to communicate with other Satchel-enabled devices, sending them the document's location. Those devices will then perform actions over those documents. For instance, a user might point its smartphone to a Satchel-enabled printer and state that the document should be printed out. A management of these transfers is completely transparent to the users. From their standpoint, Satchel appears to contain the document itself, instead of just a reference to their actual location. Some usability studies have shown the system's usefulness [Lamming et al. 2000].



Figure 2.14: Satchel

Again, while Satchel allows documents to be handled in a mobile setting, the retrieval mechanisms are basic consisting mainly on browsing the filesystem hierarchies. Specific and more efficient ways to retrieve documents in mobile settings were not studied.

IM³

The IM³ system, developed at the Ricoh California Research Center [Hull and Hart 2001], tries to integrate the management of paper and electronic documents. It is based on the premise that the cost of capturing and filing all personal documents should be nearly zero. This is accomplished by the use of special printers, fax machines and photocopiers, altered to store copies of the pages of all paper documents they handle on a central server. OCR is used to extract keywords from documents, whenever possible. This creates a searchable database of a users' documents. A web-based interface allows users to retrieve their documents using keyword search or by browsing a calendar depicting all documents handled at any given time. It lacks support for different relevant data, as it does for documents stored in the users' computers but not processed by one of the instrumented devices.

2.2. Discussion

Although interesting, none of the works we just discussed properly addresses all problems that concern us, regarding document organisation and retrieval. Table 2.2 shows a summary of the classification of those works according to different criteria:

Meta-Data : Assess if the solution gathers and uses meta-data other than keywords about documents in meaningful ways.

Large Numbers: Is the approach able to deal with all documents users possess, or can it only be of use for a subset of those?

Different Types: Is it possible to handle different kinds of documents (audio, video, etc.) in an integrated and uniform way, or is only a specific document kind supported?

Indexing: Does the system automatically gather information about the users' documents, email messages, etc., or does it require user intervention to classify or annotate them?

Individual Documents: Is it possible to visualise and access individual documents in a straightforward and immediate way?

External Information: Does the approach resort to external information sources such as the world-wide-web or room sensors?

Remote Access: Is it possible for users to access their documents remotely?

The major problem with all spatially-based approaches is the limited screen space available to display and position documents. They might be useful as desktop replacements,

	Approach	Meta-Data	Large Numbers	Different Types	Indexing	Individual Docs.	External Information	Remote Access
Space	Piles	×	×	✓	×	✓	×	×
	Data Mountain	×	×	×	×	✓	×	×
	Cartographic	×	×	×	✓	×	×	×
	BumpTop	×	×	✓	×	✓	×	×
Time	TimeScape	×	×	✓	×	✓	×	×
	Lifestreams	✓	×	✓	✓	✓	×	×
Meta-Data	SemanticFS	✓	✓	✓	✓	✓	×	×
	Tapestry	✓	✓	×	✓	✓	✓	×
	Harvest	✓	n.a.	✓	✓	✓	✓	×
	Placeless Docs.	✓	✓	✓	✓	✓	×	×
	PACO	✓	✓	✓	✓	✓	×	×
	Haystack	✓	✓	✓	✓	✓	✓	×
	Connections	✓	✓	×	✓	✓	×	×
	Pers. Chronicles	✓	✓	✓	✓	✓	×	×
	FotoFile	✓	✓	×	×	✓	×	×
	PhotoToc	✓	✓	×	×	✓	×	×
	MyLifeBits	✓	✓	✓	✓	✓	✓	×
	Stuff I've Seen	✓	✓	✓	✓	✓	✓	×
	Tags	Phlat	✓	✓	✓	✓	✓	×
View on Views		×	✓	×	×	✓	×	×
Beyond	Factoid	✓	n.a.	n.a.	✓	n.a.	✓	×
	Rem.Agent	✓	✓	×	✓	✓	✓	✓
	RoamWare	✓	×	✓	×	✓	✓	×
	Satchel	×	✓	✓	✓	✓	×	✓
	IM ³	×	✓	×	✓	✓	✓	✓

Table 2.1: Related Work Comparison Table

helping users organise a relatively low number of documents pertaining recent tasks. In the general case, however, they are unable to cope with the large numbers of personal documents that users now have at their fingertips. Furthermore, meta-data use is low. Many attributes that might help the users find specific documents are not displayed. If displayed, they would further clutter the workspace, making it possible to deal with even less documents. Finally, a lot of emphasis is placed on user organisation of documents. Retrieval becomes secondary as it is expected to be straightforward after all documents have been placed somewhere by the user. While easier than classifying all documents into a hierarchy, some sort of classification is still necessary, even implicitly, when choosing a document's placement.

Regarding TimeScape and Lifestreams, the systems that organise documents chronologically, it is possible to find similar limitations in terms of the number of documents that can be adequately handle, especially for TimeScape. In that system, only documents in the users' desktops are organised, rather than all of their documents. Lifestreams, on the other hand, presents document streams in a graphic way, again limiting the number of documents that can be successfully displayed. Furthermore, from all possibly relevant information about documents, the user, and the contexts that surround them, time is given a dominant role. Other kinds of information might be more natural or useful when retrieving certain documents but it is still necessary to recall when a document might have been handled before any other cues present themselves to the user.

It is not surprising, thus, that most recent developments in the area provide a better balance between time and other meta-data properties. They are not without problems, however. The Semantic File System is clearly unsuited to today's needs. The document retrieval mechanisms it supports are fairly basic, revolving around a command-line interface. Tapestry and Harvest are relatively simple works, more concerned with gathering information about documents than in using that information to help users retrieve them.

More recent works such as Placeless Documents and PACO are more versatile. However, document retrieval is performed with help of little more than keyword search. Haystack is more concerned with organising and displaying bits of information from separate sources than helping users retrieve actual documents. Furthermore, in all these works (and, to some extent, all property-based works), an additional problem exists: even while they might prevent the need for the user to classify their documents, the retrieval is based on the need for users to recall arbitrary properties and possible values, with no underlying mechanism to make this easier. They shift the cognitive load suffered by users from organisation time to retrieval time.

Approaches such as PhotoTOC, FotoFile and Connections are relatively simple, solving very specific retrieval problems for very specific document types. As such, they do not apply to the general case of personal document retrieval.

More mature solutions such as *Stuff-I've-Seen* and *MyLifeBits* are the ones that come closer to solving the personal document retrieval problem. However, they assume users will want to start looking for their documents with keyword searches. This might not be the case, if users better remember some other feature (whom they sent the document to, or when they wrote it, for instance). Only after some results have been returned can other meta-data properties be used to narrow down those results. If the document is not text-based and has no text-based meta-data associated to it, keyword search might not be possible at all.

A problem shared by all existing meta-data-based solutions is the lack of a formal study to discover which are the meta-data properties that users find most relevant when looking for their documents. All works assume that some information elements, such as time, are important, but often do not formally verify it. Other properties might be as important but remain unused. Thus, before using meta-data or autobiographic information to help users retrieve their documents, some effort should be spent discovering exactly what information is important and in which ways.

The merits and disadvantages of tagging systems have already been discussed earlier in this chapter. They are useful mainly in collaborative shared environments, and suffer from problems such as low tag re-use and difficulty dealing with large numbers of elements to be tagged. Also, resorting solely to tagging would ignore a wealth of information about documents that can be automatically garnered by a computing system. Tagging may be used as a way to enrich a more complete system as is the case in *Phlat*.

Works that try to go beyond the desktop and gather information from the users' environment or work in a mobile setting, such as the *Remembrance Agent*, *Satchel*, *RoamWare*, or *IM³* suffer from a serious problem: they require special hardware in order to function. Until (if ever) technologies such as active-badges and head mounted displays become commonplace, such approaches can be used only in restricted settings. This renders them impractical as a general solution for the personal document retrieval problem. Also, those approaches tend to be intrusive.

In short, there is a lack of a way to help users retrieve their personal documents in effective ways, with the help of different autobiographic information elements. A way to help users remember the elements they consider the most relevant, enabling the retrieval of all kinds of documents, even non-textual ones, must be devised. This is accomplished by our new interaction paradigm, *Narrative-Based Interfaces*, described in this document.

3

Stories About Documents

As discussed in Chapter 1, narratives about documents can be the basis for the creation of an interface that allows users to retrieve personal documents. Such an interface will need to be able to allow users to tell stories about their documents to the computer as they would to a human listener. To properly design it, we need first of all to understand what to expect from document-describing stories. Only in that way will the interface be able to properly address the issues involved in correctly capturing and understanding stories.

As a story is not just a set of unrelated information tidbits but, rather, a coherent whole, it is important to consider stories not only in terms of their *contents*, but also of their *structure*. This will allow the design of an interface that not only takes into account the most common and expected story elements in a narrative, but also how they inter-relate. This will enable it to anticipate the user's story elements, build expectations as to what elements will be referred to in the future, and about what factors might the user remember easily even if they were not volunteered in the first place, resulting in a dialogue that is both informative and not awkward.

To obtain a thorough characterisation of document-describing stories, we performed a study in which computer users were interviewed and asked to tell stories about their

documents. The subsequent analysis of those narratives yielded the much needed description of stories required for the design of the retrieval interface.

When designing the study, we had to keep in mind some relevant concerns. First of all, our goal was to design an interface that allows *all users* to retrieve their documents. As such, we strove to interview a diversified set of users, in terms of age, profession, background and computer expertise. While specific user profiles might give rise to different stories, using those as the basis for a general-use system would be inadequate and could yield poor results.

Also, we wanted to obtain stories as told by users with as little external interference as possible. We ensured this by keeping the interviewer's interference to a minimum (as described in Section 3.1). However, to prevent biasing the results, it was not feasible to select one of the users' documents and asking them to describe it. This would instantly provide them with hints about the documents, polluting our results. Indeed, just by looking at the document the users might be able to identify its visual aspect, authors, subject, title, etc. Instead, we asked the users to recall the documents they would be willing to describe, resorting only on their memories. This closely mimics a real-world usage scenario in which users for some reason remember a document they need to find.

As the users were asked to recall the documents they described, they could be tempted to mention the first that came to mind. Most likely, those would be documents they had recently worked on or read. Since we wanted the interface to be able to help users find *any of their documents*, we prevented this from happening. To that end, the users were asked to describe not just any of their documents but, rather, documents of different kinds: a Recent document, that they had created up to two weeks ago, and an Old document, which they had created at least six months ago. This allowed us not only to solve the problem of document variety, but also provided data that gave us important insights on the effect of time on the users' memories about their documents. Another aspect that might greatly influence the users' stories is whether they are describing a document they themselves created or helped create, or just one that they read. Different degrees of intimate knowledge about the documents could exist and shape the stories differently. To be able to evaluate any such differences (and further diversify the document sample), all users were also asked to describe a document of Other authors.

Finally, privacy concerns were paramount not only in this but in all user studies performed during our research. As users will be describing personal documents in terms of autobiographic information, privacy must be ensured at all times to prevent them from withholding information and distorting the data.

In the following sections, we will describe how the study was conducted and analyse the results it produced, culminating in a set of guidelines for the design of narrative-based interfaces for personal document retrieval.

3.1. Procedure

We decided to perform twenty interviews, in order to cover a wide range of users and also to get enough information to perform some statistical analysis. If, after twenty interviews, it was found that no clear trend had emerged from the data, more interviews would be undertaken. As it turned out, no further interviews were needed.

Each interview was divided into two parts. In the first, our work was described, motivating users for the rest of the interview and letting them know what was expected from them. An interview sheet was filled in with relevant user related information (age, job description, etc.). Also, we asked for consent for recording the interview, after ensuring the complete confidentiality of whatever information was collected. After their privacy had been guaranteed, all users agreed to the recording.

In the second part, the documents themselves were described. The participants were asked, in turn, to remember a Recent, Old or Other document. After they remembered a document, they were asked to “tell whatever they could remember about that document; tell us that document’s story”. We performed semi-structured interviews: rather than asking users a fixed set of questions (resulting not in stories but in separated pieces of information), the users were allowed to tell their stories with little interference. There were only two exceptions to this rule: when the user started to digress from the description of the document, or when the user seemed at a loss of what to say next. In those cases (properly handled in the result analysis as described below), the interviewer prompted the user to continue describing the document. This was done in a pre-defined and systematic way, using a set of sentences about the different story elements that might appear (see Appendix A1). Even so, we refrained from asking those questions, only doing it when absolutely necessary for the interview to proceed. This prevented dialogues between the users and interviewers, and kept their interference to a controlled and known minimum.

We tested the interview script in two test interviews, on different settings (office and coffee shop). This allowed us to perfect it before conducting the bulk of the study. In those test interviews we found the setting itself not to be too relevant, providing that the interviews took place away from the user’s computer (or with the computer turned off). If in front of their computers, the temptation to use them to search for the documents being described would hinder the storytelling process and distort the stories. Thus, in subsequent interviews, we contacted the users beforehand to arrange a time and a place for the interviews away from such distractions.

The two test interviews also gave us an initial idea of what to expect in stories. While we expected to find information such as a document’s subject and title (among others), only with the help of those test interviews were able to confirm this and fine-tune the

list of possible story elements and corresponding sentences to be used when the need to prompt the user to continue arose. Even so, this list remained open to any other elements that might appear during the course of the analysis of the interviews.

3.1.1. Interview Analysis

In order to collect the elements that constitute the users' stories, we produced transcripts of all interviews (they can be found in Appendix A5) and performed *content analysis* [Huberman and Miles 1991] [Yin 1989] on them. We classified story phrases as belonging to several expected story elements. While the list of elements had been completed from information gathered on the test interviews, we were prepared to collect other elements that would present themselves during the course of the study. However, that was not necessary, as no new elements came up.

Content analysis is often performed automatically with the help of specific applications. Alas, this was not done in our case. Automatic contents analysis requires the definition of a dictionary with the coding of different words or sequences of words. For instance, we could specify that the word "hour" indicates a mention to "Time". However, this is not adequate for our study. We were not looking for specific words or expressions. Rather, we were searching for story elements that might span entire sentences (or even more than one sentence). It is the high-level semantics of those sentences that interested us, and not those of their constituent elements. Furthermore, given the expected diversity of stories we hoped to find, it would be impossible to perform the analysis and collect story elements just by looking for specific pre-determined words. Hence, we performed the analysis of each interview by hand, using the set of guidelines presented below. Good sense was paramount in deciding how to code some phrases, but for the large majority of sentences, no doubt existed on how they should be coded.

Part of the story elements used to encode the interviews and analyse them were pre-established by the researchers, including obvious choices like Time and Subject. This initial list was validated and completed based on the stories collected in the aforementioned two test interviews. The complete element set used to code the other interviews was:



Time: Any temporal reference regarding when the document was produced or read, including references to specific holidays (Christmas, Easter, etc). Other calendar references should be classified either as Personal Information (birthdays, days when the user travelled somewhere, etc.) or World Information (Election Day, etc.)



Place: The physical place where a document was worked on or read. This element could be a reference to the users' home or working place but also, given the current capabilities of mobile devices, a location such as a train or coffee shop.



Co-Authors: Someone that wrote parts of the document or that otherwise helped produce it. Alternatively, someone that reviewed or proof-read a document, somehow annotating or contributing to its final version, could also be accepted as its co-author.



Purpose: The reason why the document was produced (to present in a conference, to prepare a class, to analyse different software products). Specific information regarding the contents of the document should be classified as its Subject. For instance, saying "the document contains an analysis of several database systems, made after some tests" is its subject. Saying it contains "an analysis of software to be bought by the department" relates to its purpose.



Author: The person who wrote the document. This could be the user that is telling the story, but also anyone else, if the document hasn't been created by the user (a newspaper article or conference paper, for instance)



Other Documents: References to other documents that were either produced or read at the same time that the document being described, documents that served as references for that document or documents it was used as a reference for.



Document Exchanges: References to the document being somehow exchanged with someone else, by email or other means. For instance, stating that a document was sent by email to a colleague, or that it was uploaded into a department's web site for the manager to see.



Document Type: A reference to the type of the primary content of a document (text, image, sound, etc.). Often, this element reflects the tool with which the document was produced (Microsoft Word, Microsoft Excel, etc.), as for many users a text file is exactly the same than a Word file (since they use no other tool to write text).



Subject: What the document describes. See Purpose.



Name: Explicit references to actual file or folder names, with regard to the document being described.



Tasks: Some task the user had to perform in relation with the document being described. For instance, having to go to a library to find an important reference, having to perform a set of interviews, or the need for the use of an image editing program to prepare a picture to be inserted into the document.



Storage: Explicit references to where a document is or was stored. This could mean references to a specific computer, disk, folder or other 'generic' storage place. Named references ('on a folder called FOO') should be classified in the Name category.



Versions: References do different versions. This includes working versions and documents with different versions for different audiences. For instance, a presentation that was created for a meeting whose copy was edited for another, or the same document translated to a different language.



Contents: Some reference to the contents or structure of the document. Both specific references (a specific phrase or image) and generic references ('it has images', 'I remember I included a pie chart') should be taken into account.



Personal Life Events: Events on a user's personal life that are somehow related to a document. This includes events that are marked on a user's agenda, such as a birthday, but also other noteworthy events (the user being sick when he produced the document, for instance). In short, they must directly relate to the user.



World Events: Events that occurred on the world at large when the document was being read or produced. It is the case of important world news or social events (elections, etc.) but also of something that happened around the user but not directly relating to him (a friend bought a new car, or a neighbour moved away).



Events: Some event that happened while the user was interacting with the document (someone entered the office, the phone rang, the lights went out, etc.). Only references of events occurring when the user was actually working on or reading the document should be considered.

Below is the excerpt of a transcript of one of the stories told to us by a user (Interview 14, in Appendix A5), describing a Recent Document. The original story was in Portuguese, the users' native language. Here we have translated it to English. In the excerpt, the manual encoding of phrases into the different story elements can be seen.

[Interviewed] Ok. This is the last document I worked on, [*about a week ago*]_{Time}. It was a document [*about a market survey project*]_{Subject}. So, This is an important document I created [*a week ago*]_{Time}. It was [*for the market surveys course I was enrolled in, the exploratory phase*]_{Purp}. It was a [*group work*]_{Co-Auth}, so it was hard to [*bring everything together*]_{Tasks} as parts of it were created [*in different computers*]_{Stor}. Namely, [*my colleague's*]_{Stor}. I called it... tried to call it [*'trabem1'*]_{Name} to know it was the [*first market studies project*]_{Purp}...

[Researcher] Besides your colleague, did anyone else helped you write the document, or read it to correct some things?

[I] Yes, [*my boyfriend*]_{Co-Auth}. He usually reads my works to help me with some language issues, some mistakes we don't detect. [*My school colleagues*]_{Co-Auth}, also.

[R] In this particular case, did you send the document to someone, by email, perhaps?

[I] In this case, [*I did not*]_{Exch}. The teachers wanted the works on paper, as they don't want to have to print them out. Sometimes its a lot of paper sheets. Other works I have turned in by email.

(...)

We gathered information not only about the occurrence of specific elements, but rather, on the frequency with which those elements appear in the stories. This allowed us to estimate what elements are more readily remembered or considered more important by users. In the story above, for instance, the Co-Authors element has an occurrence of one (it is there), and a frequency of three (it appears three times).

While performing the analysis, we distinguished between two kinds of elements: *spontaneous* and *induced*. Spontaneous elements are those users refer to on the course of their stories. Induced elements, on the other hand, are those users remember and mention after a question made by the interviewer. In the sample story, the mention to the document's name, 'trabem1', is spontaneous, as it occurred naturally in the middle of the user's speech. The Co-Author reference to the boyfriend, on the other hand, is induced, as it was prompted by the researcher's question.

Independently of how users remember a specific element (spontaneous or induced), negative answers were also recorded. For instance, if a user refers that a document had "no

co-authors”, we note that a reference to co-authors was made: there weren’t any. Not knowing if the document had any co-authors is not the same as knowing it had none and this allows us to maintain that difference in subsequent stages of the analysis. In the story above, this occurred regarding eventual email exchanges: the user knows not to have sent the document by email.

Apart from the contents analysis, we performed a *relational analysis* of the collected data. This kind of analysis enabled us to discover how the several story elements relate to each other. This analysis is equivalent to the production of a labelled oriented graph whose nodes are story elements and whose arcs represent the relations between those elements. The *strength* of a relation, recorded in the arcs’ labels, is the number of times an element was mentioned after another. All individual relations were considered to have the same strength. The order in which the elements were mentioned provides the *direction* of the relation and, hence, that of the corresponding arc. In practice, we recorded the graph information on adjacency tables, as it makes the statistical analysis of the results easier to perform.

Important for the relational analysis is the concept of *element transition*. Between two consecutive story elements, a transition is said to have occurred if they follow each other in the story. No transition between two elements is considered if the second one is induced, since it wasn’t the user to relate the two elements. This happened only due to the intervention of the interviewer. Also, no transition was considered between elements mentioned consecutively but too far apart in the story, when the user digressed in between. Finally, whenever the interviewer asked a question and the element mentioned as an answer does not match the question, it is not considered to be induced. The user just happened to remember some more information while the interviewer was asking the question. Nevertheless, to err on the side of caution, no transition was recorded when that situation arose.

In the sample story, above, we can find nine transitions in the first paragraph: from Time to Subject, Subject to Time, Time to Purpose, Purpose to Co-Authors, etc. No transition is recorded between the Purpose element at the end of the first paragraph and the Co-Author at the start of the third (the reference to the boyfriend) as the later is induced.

3.2. Results

In the following sections we will present the main results that were extracted from the analysis of all sixty stories collected on the twenty interviews. Whenever statistical tests are mentioned, confidence levels of 95% were considered.

3.2.1. Subject Characterisation

We interviewed 22 users. Of those, only twenty interviews were thoroughly analysed, since in the remaining two (number 21 and 22 in this document) specific documents weren't described as such. Instead, the interviewed chose to describe how they remember, organise and describe their documents in general. Rather than dismissing those interviews, we found they provide interesting insights into the interactions between users and their documents. We'll discuss them later in this chapter, in Section 3.2.5.

Of the twenty interviewed, eleven (55%) are female and nine (45%) male (Figure 3.1a). Their ages ranged from 24 to 56 years old (Figure 3.1b), and their academic backgrounds from high-school undergraduates to PhDs. Only four have some kind of specific computer science education. Their professions are as diverse as law students, secretaries and statisticians (Table 3.1). Subjects whose profession is listed as "Information Higher Technician" work in different departments of a large (about 1,000 employees) corporation, with different competencies. Their jobs range from database administration to web design and pure management tasks. For the most part, they require no more computer skills than daily usage of common applications, such as word processors and spreadsheets. We verified this on-site. We are, thus, confident we captured a fairly diverse user sample.

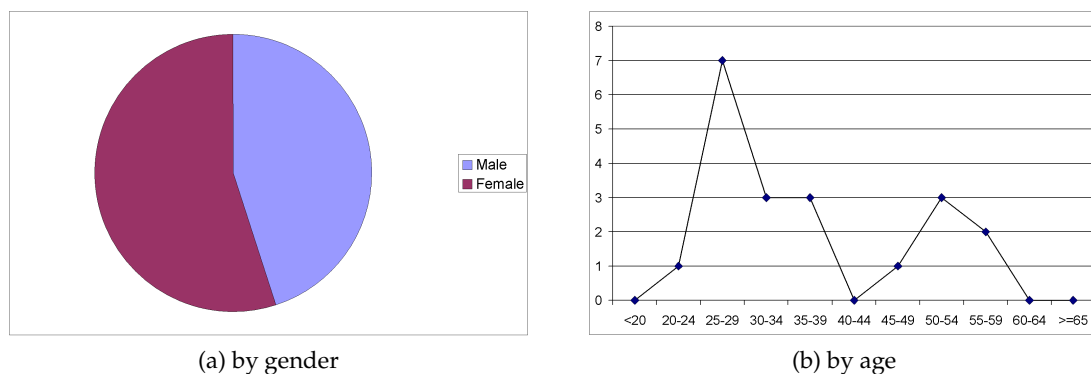


Figure 3.1: Interviewed Profiles

3.2.2. Overall Story Properties

In regard to overall story structure, there are four main factors to take into account: length, number of transitions between story elements, number (and length) of uninterrupted story trains, and number of changes in subject.

Num	Age	Profession	Academic Background
1	28	Teaching Assistant	Computer Science MSc
2	37	Administrative Secretary	12th Grade
3	55	High School Board President	Organisation and Management MSc
4	27	Statistics Higher Technician (SHT)	Economics Degree
5	27	Project Manager	Electronics and Computers Degree
6	27	Economist (Social Security)	Economics Degree
7	25	Law Student	12th Grade
8	52	Information Higher Technician (IHT)	Mathematics Degree
9	37	IHT	Computer Science for Management Degree
10	51	Education Sciences College Teacher	Educative Supervision MSc
11	49	IHT	11th Grade
12	56	Education Sciences College Teacher	Educational Sciences PhD
13	37	College Auxiliary Professor	Silviculture PhD
14	25	High-School Mathematics Teacher	Mathematics Teaching Degree
15	24	Biological Engineering PhD student	Biological Engineering Degree
16	50	SHT	Enterprise Organisation and Management Degree
17	27	IHT	Statistics and Information Management MSc
18	31	IHT	12th Grade
19	31	IHT	Computer Science Degree
20	34	IHT	Statistics and Information Management MSc
21	28	<i>Operations Director</i>	<i>Social Anthropology Degree</i>
22	34	<i>IHT</i>	<i>Physics Degree</i>

Table 3.1: Interviewed Backgrounds and Professions

Story Length

The average story length was 15.85 elements (stdev = 5.97). We found, however, differences between the lengths of stories pertaining to different kinds of documents. The longest are those describing Recent documents, with 18.5 elements on average (stdev=5.87), followed by 16.9 elements for stories about Old documents and 12.15 (stdev=6.76) elements for stories about Other documents (stdev=4.72), as displayed on Table 3.2.

	Recent	Old	Other	Total
Average	18.5	16.9	12.15	15.85
St.Dev	5.87	5.57	4.72	5.97
Shortest	10	8	7	7
Longest	35	27	27	35
Number of stories	20	20	20	60

Table 3.2: Story Lengths

This seems to make sense: it is easier to remember a recent document than an older one, and even more than a document not written by the users themselves. The use of Mann-Whitney U tests (given the non-normality of the sample data for Other documents, that prevents t-tests from being used) shows there is a statistically significant difference (with 95% confidence) between the means for Recent and Old documents and that of documents by Other authors. However, no significant difference between the means for

Recent and Old documents was found (a t-test was used in this case as both data samples are normal). The only conclusion that can be inferred, then, is that users told longer stories about their own documents than about those of others.

Looking at a possible dependency between story length and age, we found no real relation between those values (Figure 3.2a). There is a significant statistic correlation between the two (the Pearson Correlation value has a significance of 0.723). However the value of the correlation coefficient is so low (-0.084) that it is almost irrelevant. This was a surprising result, given the known cognitive problems that arise with age. It was to be expected that younger users would find it easier to remember some facts about their documents. Different factors might account for this surprising result. The number of interviewed persons might have been too small for such a trend to establish itself, or the ages considered were not high enough for cognitive problems to manifest themselves, or no subject was old enough for such problems to manifest themselves. However, looking at the standard deviations within each age class (displayed as error bars in the graphic), we clearly see individuals in nearly all classes fall within the error bars of other classes, reinforcing the conclusion that no significant differences exist between them. It seems that, when remembering documents, users of all ages (up to 60 years old) have the same abilities.

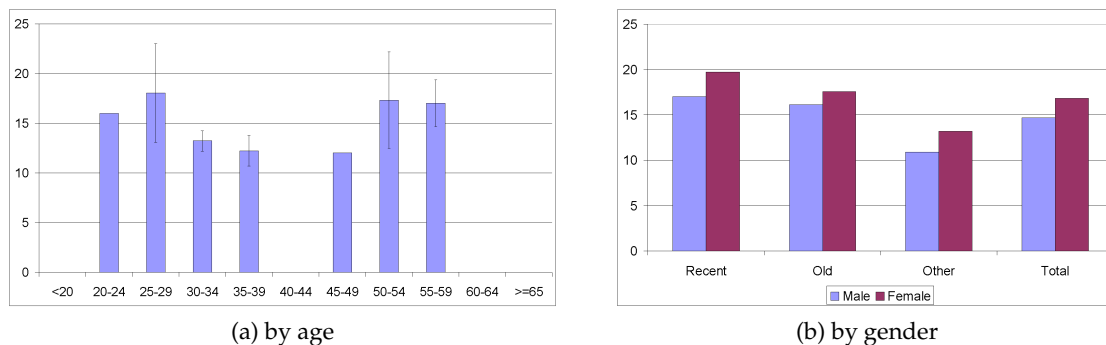


Figure 3.2: Story Length

		Recent	Old	Other	Total
Male	Average	17	16,11	10,89	14,67
	St.Dev	3,3	5,58	3,45	4,91
	Shortest	13	9	7	7
	Longest	23	27	17	27
Male	Average	19,73	17,55	13,18	16,81
	St.Dev	7,28	5,75	5,49	6,63
	Shortest	10	8	9	8
	Longest	35	26	27	35

Table 3.3: Story Lengths

Trying to find a relation between story length and gender, however, had positive results. As Table 3.3 and Figure 3.2b show, women tend to tell longer stories than men, for all

document kinds. Furthermore, the longest story told by a female is 35 elements long and the longest told by a male is merely 27. A t-test confirms this difference to be statistically significant for all document kinds. It seems that either verbal or cognitive abilities are more developed in female users. We'll discuss this matter again later on, when we discuss the numbers of transitions and the length of story trains.

Number of Transitions

Moving on to the number of transitions per story, this number tells us, when compared to the story length, how spontaneous a story was. As no transitions were recorded when an element occurred due to external intervention, fewer transitions mean that most elements were told in answer to questions asked by the interviewer.

As can be seen in Table 3.4, despite some remarkably large transition numbers (around 20), the averages are considerably smaller. Taking a look at the transitions/length ratios, we find that they don't vary much from document type to document type. In fact, it would overall seem that as many as 46.7% of story elements arise spontaneously. Several t-tests confirm that the mean values between document classes are not significantly different. Regardless of document type, on average, nearly half the narratives were completely controlled by the interviewed.

	Recent	Old	Other	Total
Average	9.3	7,55	5.53	7.4
St.Dev	6.06	6,96	4.18	5.97
Shortest	2	0	1	0
Longest	27	21	19	27
Transitions / Length	50.27%	44.67%	43.86%	46.69%

Table 3.4: Story Transitions

There is, as already mentioned, a rather large oscillation in this value from individual to individual. Despite the average values, there are registered cases for nearly all possible ratios. Nevertheless, Kolmogorov-Smirnov tests show the statistical distributions for these values to be normal, allowing us to trust the averages mentioned above.

Looking for possible explanations for the variations in this ratio, we looked for a correlation between the values for all document types and age and gender. Regarding gender, the average ratio for male subjects is 40.77% (stdev= 18.96%) and 45.43% (stdev=16.37%) for female subjects. A t-test shows these values not to be significantly different: the transition/size ratios do not depend on the subject's gender. As for age, we calculated the Pearson's correlation value between age and ratio and found a positive significant but weak correlation (0.221). Thus, it seems there is a small tendency for older users to be more in control of their stories.

Story Trains

The number and length of story trains (uninterrupted transition sequences) found in the narratives give us an idea of how stories are structured, regardless of transition number. In our sample story (page 43), we find three story trains, corresponding to each of the subject's paragraphs (only the beginning of the third is present). Each intervention by the interviewer breaks a story train. The first train, corresponding to the first paragraph, has a length of eight, as it is composed of eight transitions between nine different elements.

Story Train Numbers

	Recent	Old	Other	Total
Average Number	3.7	2.6	2.3	2.87
St.Dev	1.81	1.7	1.3	1.7

Table 3.5: Story Train Numbers

Looking at Table 3.5, we see that stories are, on average, composed by 2.87 different trains. However, the fairly high standard deviation lets us foresee large divergences from this number. The histograms in Figure 3.3 confirm this. They also show us that, despite this dispersion, most stories have three story trains or less (Table 3.6). With the notable exception of stories about Recent documents, over 85% of stories share this property (the cumulative percentage of documents with three or less story trains is 85% for Old and Other documents and 55% for Recent ones).

	Recent	Old	Other	Total
0	0	5	0	1.7
1	5	30	35	23.3
2	30	45	60	45
3	55	85	85	75
4	70	85	90	83.3
5	80	85	100	90
6	95	95	-	96.7
7	95	100	-	98.3
8	100	-	-	100

Table 3.6: Cumulative Story Percents for Story Train Numbers

In order to establish if the different mean values for the different document kinds are significantly different, we performed several Mann-Whitney U tests (the samples don't follow the normal distribution invalidating the use of t-tests). We found no significant difference between the distributions for Old and Other documents ($\alpha=0.585$). However,

there was a marked difference between Recent and Other documents ($\alpha=0.009$). Between Recent and Old documents, the distributions are barely similar ($\alpha=0.052$). In short, we can conclude that the larger differences in story train numbers exist between stories about Recent documents (with more story trains) and those about documents of other kinds.

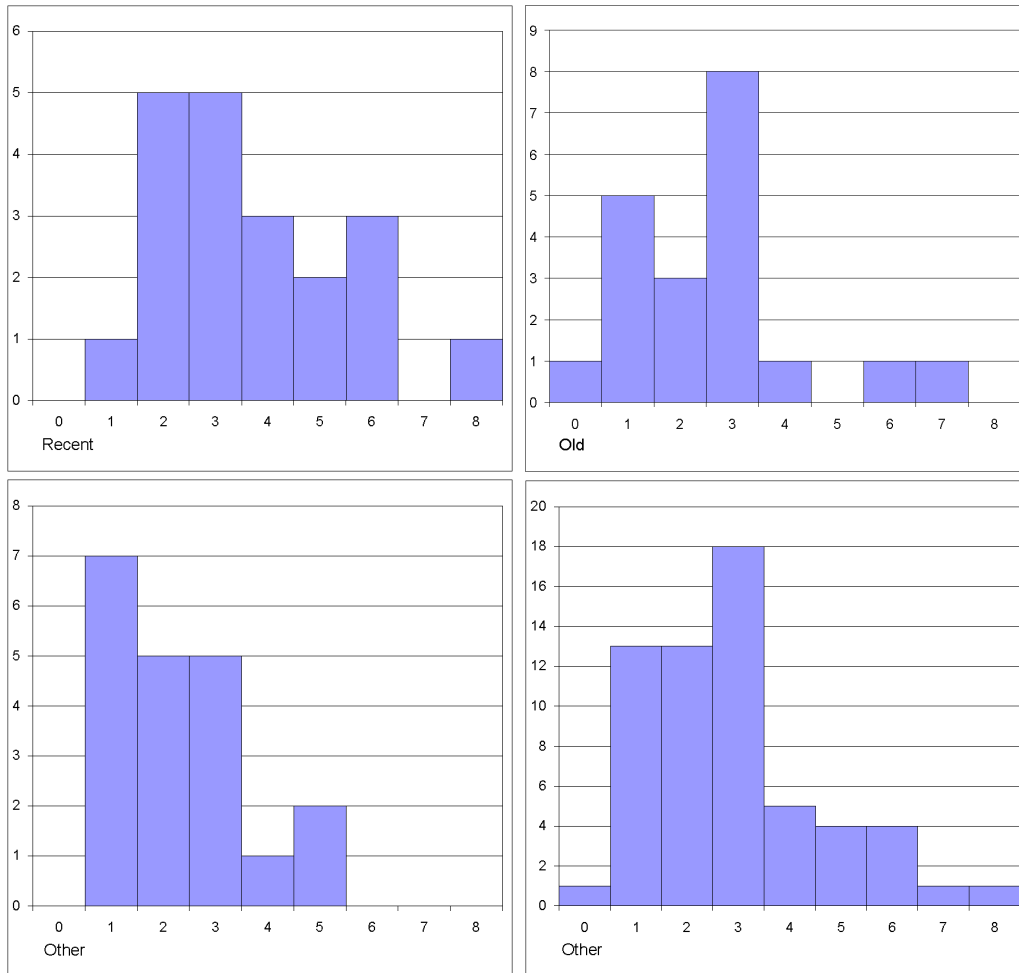


Figure 3.3: Story Trains per Story Histograms

Looking for relations between the numbers of story trains and the age of the interviewed, a Pearson Correlation value of 0.492 had a significance value of only 0.027, indicating no statistically significant dependency. Regarding gender, however, a t-test revealed the mean values of 2.53 (stdev=1.07) and 2.51 (stdev=0.84) for males and females, respectively, to be significantly different. However, they are so close to each other that, in practice, no relevant conclusion can be reached.

Story Train Lengths

More interesting than the number of story trains by itself is their size (the number of transitions in them) and how they are distributed along the narratives. Again, the average

values for the different document kinds are similar, between 2.51 and 2.90 (Table 3.7). However, the high standard deviation reveals great asymmetries between the sizes of different trains.

	Recent	Old	Other	Total
Average Length	2.51	2.9	2.33	2.58
St.Dev	2.26	3.47	1.89	2.6
Shortest	1	1	1	1
Longest	12	20	8	20

Table 3.7: Story Trains Lengths

In fact, a quick inspection of the data shows us that the first two trains of nearly every story are by far the longest. The remaining are fairly shorter (often only one transition long), with the possible exception of the last, that sometimes is again a bit longer. This could be directly related to how the interviews were conducted. In the end of each one, the interviewed were asked if they had anything else to say about the document they were describing, resulting in a final ‘burst’ of information. Confirming this, Table 3.8 shows the percentage of the total story transitions in the first two and final trains for stories with more than three trains (for the comparison to make sense). As many as 75% of all transitions in a story can be contained in the first and last trains. Around 50% are in the first two trains, suggesting special care should be taken when looking at the information therein.

	Recent	Old	Other	Total
First Two (%)	44,55	62,39	54,75	50,16
Std. Dev. (%)	16,09	9,67	18,78	16,38
Last (%)	19,87	16,14	18,79	18,91
Std. Dev. (%)	16,09	12,27	12,64	13,94
Total (first + last)	60.64	74,66	73.54	69.07

Table 3.8: Story Transitions in the First and Last Trains

Changes In Subject

A fourth and final relevant property of narratives is how often does a change in subject spontaneously occur. A change in subject takes place when the user, in the middle of his description, starts talking about something completely unrelated to what he was describing up to that point, with no external factor to prompt it, before returning to the description of the document. We also considered a change in subject to have occurred when the user answers with something completely unrelated to the question the interviewer has just made.

As can be seen in Table 3.9, changes in subject don't often occur. On average, they show up only once in each story. They do take place from time to time and might need to be taken into consideration.

	Recent	Old	Other	Total
Average	1,25	1,00	1,05	1,10
St.Dev	1,07	0,97	1,15	1,05
Shortest	0	0	0	0
Longest	3	3	4	4

Table 3.9: Changes In Subject

No significant difference was found between the means for the different document kinds. We did find a correlation between the subjects' age and how often changes in subject occur: a significant ($\alpha=0.682$) Pearson Correlation value of 0.098 was found. This shows a minute, nearly negligible, positive correlation between the two variables. In practice, it is irrelevant. A Mann-Whitney test showed no significant difference regarding gender.

3.2.3. Story Elements

We'll now discuss what story elements were present in the several narratives we considered. In this analysis took into account not only the number of times each element showed up in a story (frequency) but also if it was used at all in that story (occurrence). With this analysis, we distinguish between elements that users consider more important or harder to describe (mentioned more than once in a story) and those that are part of every story, even if users, consciously, don't consider them as important. Also, we address the problem of some bias that might occur if considering only frequencies, when users repeated the same element in a story due to speech cadence or narrative rhythm.

We'll also consider separately the spontaneous and induced story elements.

Frequency

In Figure 3.4 and Table 3.10, a comparison between the numbers of elements that appeared in all stories for the different document kinds can be found.

As can easily be seen, there is an apparent similarity between the elements that occurred in stories about Recent and Old documents. Despite some quantitative differences, the graphics are shaped in almost the same way. Statistically (with help of the Mann-Whitman test), we found significant differences only for the **Subject** element. This difference shows how, when a document is recent, users tend to reiterate its subject on

Element		Recent	Old	Other	Sub-Total	Total
Time	Spontaneous	28	20	11	59	100
	Induced	10	14	17	41	
Place	Spontaneous	11	12	1	24	60
	Induced	16	15	5	36	
Co-Author	Spontaneous	13	14	3	30	68
	Induced	17	19	2	38	
Purpose	Spontaneous	29	23	11	63	80
	Induced	5	9	3	17	
Author	Spontaneous	2	1	10	13	23
	Induced	2	3	5	10	
Subject	Spontaneous	19	9	19	47	70
	Induced	9	8	6	23	
Other Docs.	Spontaneous	14	14	8	36	74
	Induced	10	15	13	38	
Personal	Spontaneous	7	10	6	23	36
	Induced	5	3	5	13	
World	Spontaneous	5	4	1	10	15
	Induced	3	1	1	5	
Exchanged	Spontaneous	12	5	12	29	66
	Induced	13	13	11	37	
Type	Spontaneous	18	15	14	47	67
	Induced	9	5	6	20	
Tasks	Spontaneous	19	15	12	46	77
	Induced	12	9	10	31	
Storage	Spontaneous	19	15	15	49	83
	Induced	12	12	10	34	
Version	Spontaneous	5	7	0	12	31
	Induced	10	9	0	19	
Contents	Spontaneous	20	20	18	58	76
	Induced	6	9	3	18	
Events	Spontaneous	1	3	0	4	7
	Induced	1	2	0	3	
Name	Spontaneous	8	3	4	15	18
	Induced	0	2	1	3	

Table 3.10: Story Elements Frequencies

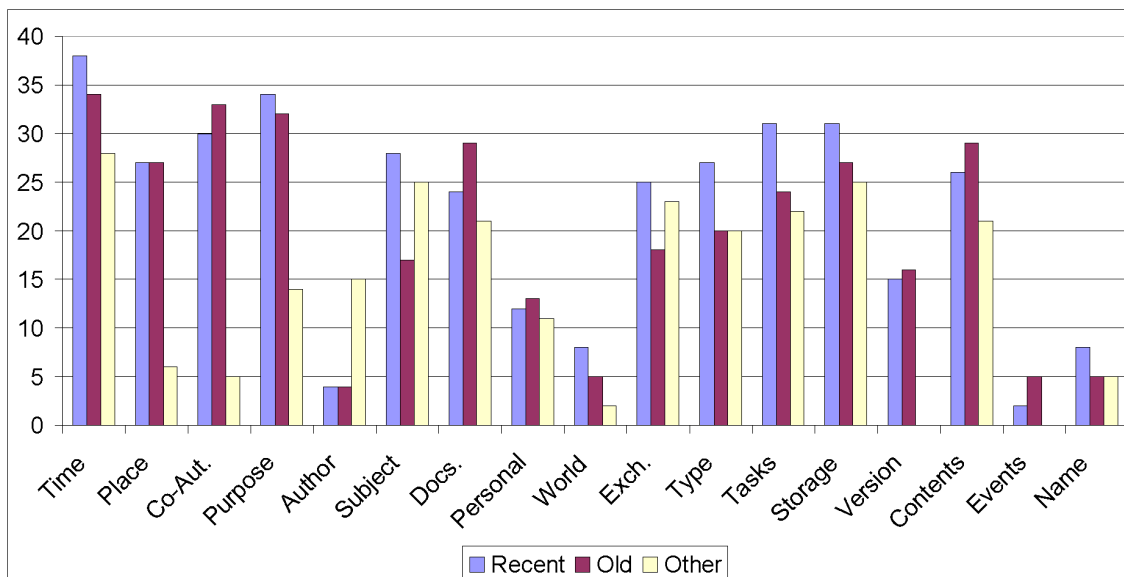


Figure 3.4: Element Frequencies for All Document Kinds

their narratives. For instance, in Interview 16, close to the start of the story, the user tells us the document is about a “(...) *dispute with the phone company* (...)” (its subject). However, after mentioning the document’s Purpose the user goes on by telling what the dispute is about and thus giving more information about the Subject: “(...) *the document I produced was a description of the whole situation I had to complain about [to the phone company]* (...)”. This does not mean that the Subject doesn’t show up in most stories about Old documents. It only shows users tend to mention Subject more in each story when it is still fresh in their memories.

Elements that appear in stories about Other documents, not written by the user, have a clearly different distribution. The most noteworthy differences are related to the frequencies of **Place**, **Co-Authors**, **Purpose**, **Author**, and **Version**. Statistical tests (Mann-Whitman) confirm these values to be statistically different from those corresponding to other document kinds. The differences in Author and Version are easy to explain. When the user itself is the author of a document, he will seldom mention it, taking that fact for granted. Also, it is hard if not impossible for a person to know if a document someone else wrote has different versions.

Co-Authors are seldom remembered when a user reads a document. Only the author, if something, is remembered. In Interview 6, the subject, when asked about the document’s authors, states that “(...) *I cannot remember! There were four or five authors, but I cannot remember.*”

The place where a document was read is also harder to remember. We think this is due to the act of reading being more passive and, thus, less prone to memorable interactions than actually writing the document. Furthermore, the place where the interactions took

place is not as important as why and how they occurred. Hence, it is hard for users to remember where they interacted with a document. We noticed that, sometimes, the subjects took for granted where the interaction took place, assuming it in their discourse, but never voicing it. This occurred mainly when the interaction happened where the interview was taking place, or when it could only have happened in one place. If the document arrived by email and the subject only has email access at the office, for instance.

Regarding a document's Purpose, the reason for the difference seems to be similar to that for the difference in references to Subject in stories about Recent and Old documents: for those documents, the users tend to reiterate what the document was written for, since it is easier for them to remember it.

Spontaneous vs. Induced Elements

As can be seen in Table 3.11 and Figure 3.5, as many as 66% of the occurrences of a story element are induced. This was to be expected given the number of element transitions discussed in Section 3.2.2. There is no clear overall trend regarding differences between the several document kinds. No statistically significant difference was found between the values for stories describing Recent and Old documents.

	Time	Place	Co-Aut.	Purp.	Auth.	Subj.	Docs.	Pers.	World
Recent	35,00	65,42	55,17	16,67	5,00	35,00	40,00	17,50	12,50
Old	54,17	60,00	66,25	23,33	15,00	40,00	52,50	12,50	5,00
Other	59,58	22,50	10,00	11,25	22,50	21,67	57,50	21,67	5,00
Total	49,58	49,31	43,81	17,08	14,17	32,22	50,00	17,22	7,50
	Exch.	Type	Tasks	Stor.	Vers.	Cont.	Events	Name	
Recent	47,67	29,58	45,83	47,92	46,25	23,75	2,50	0,00	
Old	56,25	25,00	45,00	43,33	35,00	37,50	7,50	10,00	
Other	34,17	22,50	34,17	40,00	0,00	5,83	0,00	1,67	
Total	46,03	25,69	41,67	43,75	27,08	22,36	3,33	3,89	

Table 3.11: Induced Elements Percentages

Between documents created by the user and Other documents, statistically significant differences were found only for **Place**, **Co-Author** and **Version**, as could be expected after noticing that it is for those elements that the average values differ the most. As already mentioned, no reference to a document's Version was ever made for documents created by others, explaining this difference (there are no induced elements because there are no elements at all). As for Place and Co-Author, they are seldom mentioned. It is no surprise that when they are remembered at all, hinting at a special importance, they might appear spontaneously.

Given that there are almost no significant differences between the several document

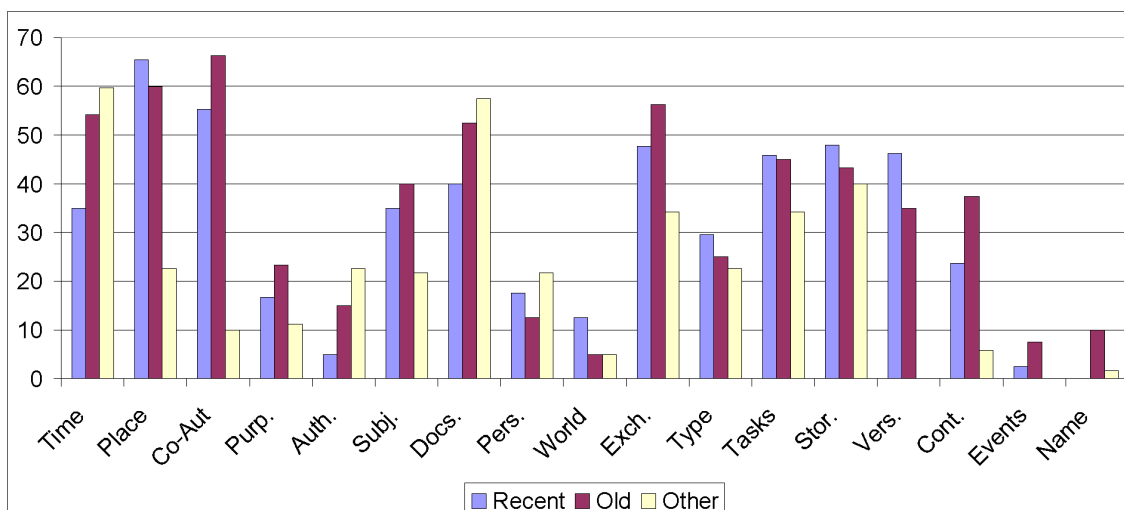


Figure 3.5: Induced Element Percentages by Document Kind

kinds, we can look at the average values in Figure 3.6 to understand what elements are more often induced, and why. The less often induced elements are **Purpose, Author, Personal Events, World Events, Events** and **Name**. As we'll see in the following section, with the exception of Purpose, all these categories are exactly those that occur less frequently. Keeping in mind that induced elements are those subjects remembered after being asked about them, the fact that these elements were rarely mentioned and, when they were, they appeared spontaneously, means that either they are so important they are remembered without need for external aid, or no amount of suggestion can make the users remember them. Purpose's case is different. It is an element that is seldom induced but that appears fairly often in the narratives. This shows it to be something users consider important and easy to remember: mentioned often spontaneously.

Some elements had to be induced an average number of times: **Subject, Type, Version, and Contents**. They are more important than those mentioned in the previous paragraph, but not the most important (described below). It is noteworthy that the most often induced elements pertain, primarily, the context in which a document was handled, while the more spontaneous ones pertain the documents themselves. It would seem that data about the documents more easily comes to mind when describing them. This might be due to the data itself, or to the fact that traditional retrieval tools are unable to use that information, thus conditioning users not to recall it.

The more often induced elements are **Time, Place, Co-Author, Other Documents, Exchanges, Tasks** and **Storage**. All of these appear fairly often in stories (Figure 3.7), at least once per story, on average. They are important elements, but hard to remember: mentioned often but only after something triggered the subject's memories about them. Even so, no element is, on average, induced more than 50% of the times it appears, showing that, even if it is hard to remember, there is a fair chance it might come up spontaneously.

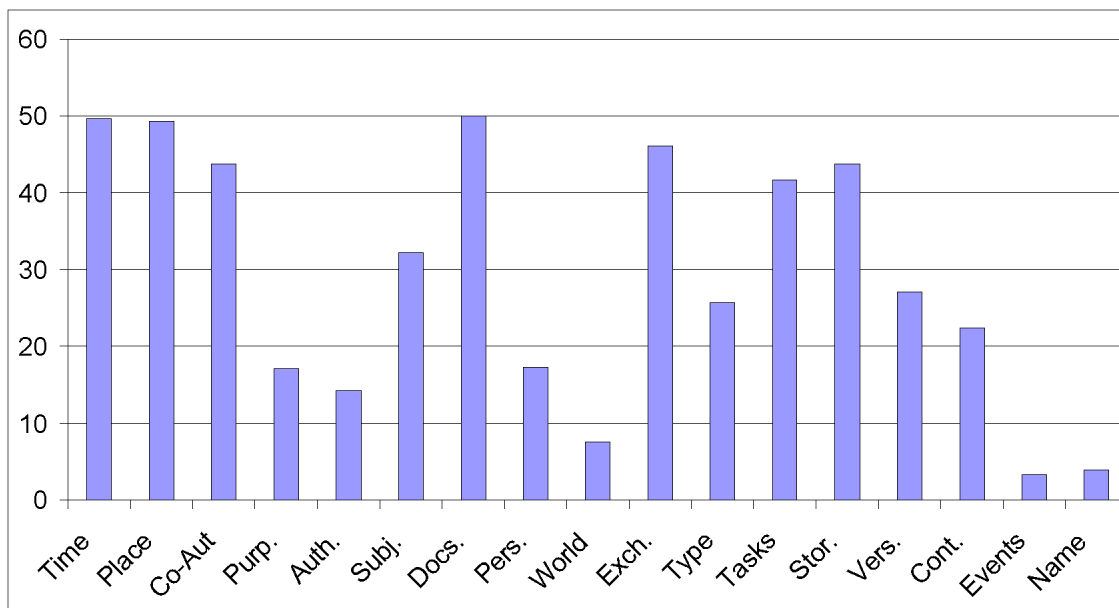


Figure 3.6: Overall Percentages of Induced Elements

Despite the average values, there are large differences between individuals. In fact, given that on each story each element is mentioned only a few times (between one and three, and up to six in some specific cases), it is not surprising that the individual percentages vary a lot. No clear pattern can be discerned for each individual. The average values should be taken into account, then, as they reveal clear overall trends. Even so, the large variability of individual values suggests that if an element is not spontaneously mentioned, it is important to ask about it because even elements that often are spontaneous might not be readily remembered in the narrative for a specific document.

Overall Frequency

Overall, the most frequently mentioned elements, those users consider more important, were **Time, Place, Co-Author, Purpose, Subject, Other Documents, Exchanges, Type, Tasks, Storage and Contents**. Taking into account that we analysed 60 stories, on average each of those elements appears at least once in each story. As we already stated, this shows how users tend to repeat some information to reinforce and clarify it.

The most important element seems to be **Time**, mentioned 100 times in 60 stories. The level of accuracy tends to vary. For Recent documents it is fairly specific: “(...) about one hour and a half ago (...)” (I2); “(...) I worked on it today.” (I5); “(...) in the beginning of the week” (I16). For Old documents, on the other hand, the time of their conception is only roughly remembered: “(...) about three years ago” (I7); “(...) a year ago, I delivered it around April (...)” (I17). In stories about Other documents, the references to Time vary in accuracy: “(...) around July.” (I3); “(...) a month, a month and a half ago.” (I7); “(...)

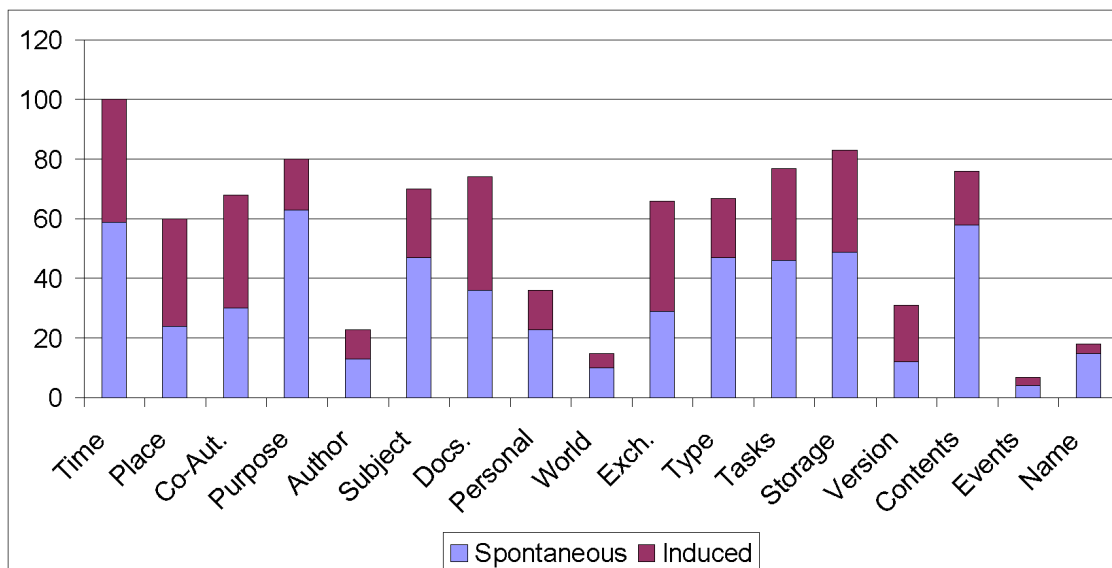


Figure 3.7: Overall Frequency by Story Element

last Tuesday. Four days ago." (I12). The accuracy depends solely on how long ago the document was handled.

There were 60 references to the **Place** where a document was read or written. However, they are very accurate: *"At home"* (I7); *"It was updated here"* (I9); *"Here. It was created here."* (I16)

References to **Co-Authors** also occurred fairly often, but seldom were actual names mentioned. More often than not, what the subjects could remember was if the document had co-authors or not: *"No."* (I5, when asked if there were any co-authors); *"No, no. Just me."* (I11); *"(...) some school colleagues."* (I14); *"Yes, it had co-authors"* (I13)

As to **Purpose**, the references also tend to be specific, including not only where but also for what the document will be used: *"(...) it will be used in the school's newspaper (...)"* (I3); *"(...) as part of the mission report (...)"* (I8); *"(...) for the Aerospace 2000 conference in Barcelona"* (I13); *"(...) used as a decision support tool."* (I4); *"(...) to manage all my projects (...)"* (I5)

A document's **Subject** also appeared regularly. The mentioned subjects were, as expected, of very diverse natures: *"(...) the subscription to a magazine (...)"* (I9); *"(...) the weekly results of my work"* (I15); *"... an analysis of the company's communications infrastructure"* (I18)

The **Other Documents** that were mentioned sometimes include actual paper documents, and not electronic ones. Also, it was common for users to mention the existence of other documents, without actually specifying what documents they were talking about (but apparently remembering them). Finally, sometimes the reference to another document

was enough to cause a ‘short story’ about that document to be told: “(...) a model form (...)” (I2); “(...) I consult documents about similar projects (...)” (I5); “A magazine about information technologies (...)” (I9); “(...) other files in the same folder” (I10); “(...) the documents in the survey (...)” (I10); “(...) another document she had already given me, about her thesis proposal (...)” (I12); “(...) a horrible book, we really had to read, because the author was the conference’s chairman (...)” (I20)

Information about the document **Exchanges** usually described email exchanges, but also other forms, such as posting it on a web site: “Yes, it was sent by email.” (I7); “(...) we deliver it posting it in that web page.” (I17) In most cases, the users were able to remember who they had exchanged the documents with.

References to a document’s **Type**, included not only the mention of specific formats (“text”, “spreadsheet”, “LaTeX”), but also to applications commonly used to handle documents of a given kind (“Word”, “Excel”, “Powerpoint”).

Moving on to the **Tasks** the subjects remember associated with a document, we find editing tasks in the computer and tasks done ‘in the real world’: “(...) went to the library to find some references (...)” (I1); “(...) downloaded and selected the photos.” (I3); “(...) I gather information from several files and perform some calculations (...)” (I4); “Gather some information (...)” (I9); “(...) I printed the document (...)” (I12); “(...) copy [it] from paper into the computer” (I14); “I made some research on the Internet” (I15); “(...) interpret the graphics (...)” (I20)

References to where the document was **Stored** often mention entire computers, but also removable media and specific (unnamed) locations in a hard drive or local networks. In the case of online documents, the site is often mentioned: “(...) the MIT site (...)” (I1); “(...) stored in the network.” (I4); “(...) in a folder for that subject.” (I5); “(...) in a cell-disk! One of those dongles (...)” (I6); “(...) stored in a CD (...)” (I14);

As to the **Contents**, it was common to find mentions to some specific information about the document’s structure. Occasions when specific contents were mentioned were rare: “(...) I included several conditional formattings to (...)” (I4); “(...) I make several different sheets (...)” (I5); “It had a sentence that started by ‘And to those persons that...’ ” (I7); “(...) with images (...)” (I9); “(...) it was divided into tables (...) in the second column were the accurate mathematical definitions. (...) It had lots of graphics (...)” (I14)

Less mentioned were references to **Authors**, **Personal Events**, **World Events**, **Versions**, **Events**, and **Names**. This shows how those elements are harder to remember and considered not as important by the users. The small amount of references to **Authors** results from the already mentioned fact that the subjects tend to take that information for granted, since most of the times the authors were themselves. When an author is mentioned, especially for documents not created by the user, it is not always possible to remember the actual name, mainly for foreign authors.

References to **Personal Events** are also fairly uncommon, ranging from events that happened to the subjects themselves or to someone directly related to them. Often it is something that could be found on someone's agenda, but not always: *"It was the day my car's battery went dead."* (I3); *"(...) I had to work on Saturdays (...)"* (I4); *"(...) we moved to our new home."* (I6); *"(...) [the night] I went to the bar (...)"* (I7); *"(...) I finished it before my vacations."* (I9); *"(...) my last trip to Brussels (...)"* (I13); *"(...) my son had a serious asthma crisis (...)"* (I20).

Almost completely absent were references to **World Events**. In the few occurrences, most times they referred to events not directly associated to the users but still directly related to their jobs or co-workers. Only once was some important news event mentioned: *"(...) the beginning of the Lages Summit (...)"* (I6); *"(...) after Easter, [those schools] don't have classes. We had to change the schedule (...)"* (I10); *"(...) it was made at the same time than that 'statistical operation' [such as a national census] (...)"* (I11); *"(...) that day, the company's email server wasn't working (...)"* (I19); *"It was when [person] left the association (...)"* (I7)

When **Versions** were mentioned, it was often to state that they didn't exist, for a particular document. Only rarely did some explicit references to actual versions occur. No really useful information could be found about for this element. *"I [had to choose] between the final version and the previous versions [of the document]"* (I8); *"(...) I don't know if [that] is the last version"* (I1)

The least mentioned story element was the **Events** that might have occurred when the subject was interacting with the document. In fact, there are only seven such references, for five different stories. Often, the events describe actions by the user and not related to the document, and not events that happened outside the control of the user. It would seem such incidents are without importance and, hence, quickly forgotten: *"(...) I prepared instant soups (...)"* (I1); *"Someone arrived at my home (...)"* (I7); *"(...) I went to prepare dinner (...)"* (I7); *"I was tired and had to go to the supermarket (...)"* (I12)

Finally, there were some references to **Names**, either of the document files themselves or of folders where those files are stored. In some cases, no specific names were uttered, but it was clear the user had a specific well identified folder in mind: *"(...) [the file] is called 'análises'."* (I4); *"(...) in a folder called 'Market Studies' (...)"* (I14)

Occurrence

A legitimate concern regarding the analysis of the frequencies of the several story elements is that those that appear more frequently could appear many times in some stories and not at all in others. Hence, it is important to verify what elements actually appear in most stories.

In Figure 3.8, we can directly compare the frequency of each element with the number of

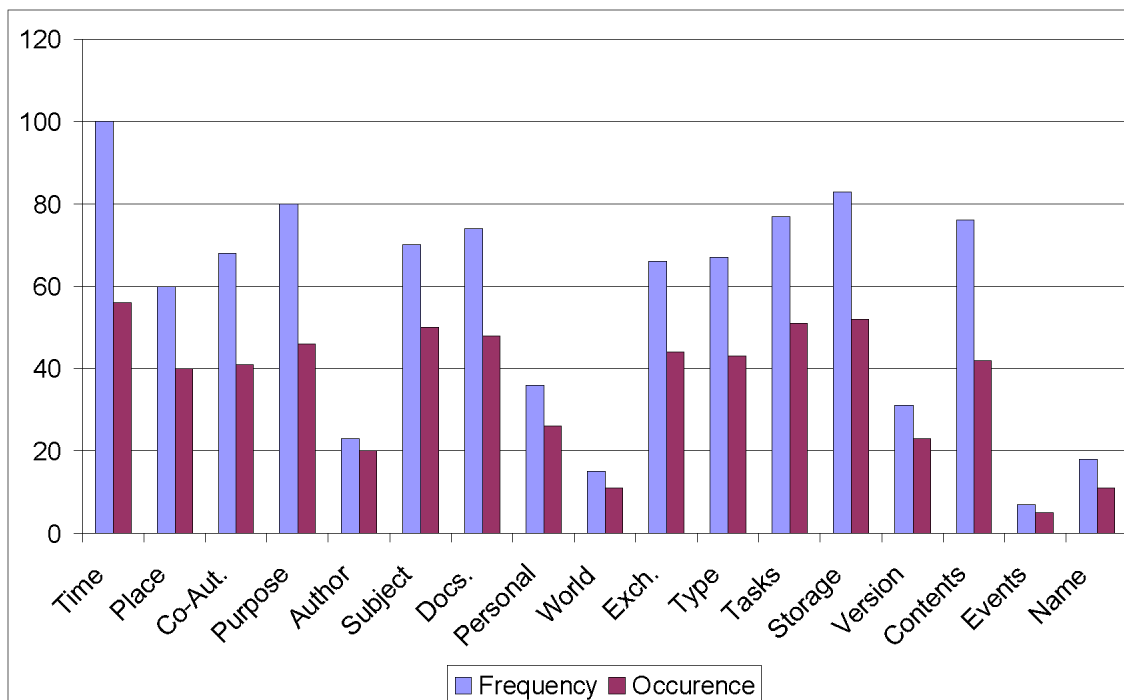


Figure 3.8: Frequencies vs. Occurrences

stories where they are mentioned. At first sight, the different series in the graphic seem to have a similar shape, albeit on a different scale. Confirming this, we see in Figure 3.9 that the ratio between the frequencies and occurrences is fairly constant, around 0.66. The biggest (but not too large) differences can be found for **Time**, **Purpose**, **Author** and **Contents**. References to a document’s author appear only once in the stories they occur. References to Time, Purpose and Contents, on the other hand, tend to appear more often in each story. This suggests they are remembered more inaccurately by the subjects, who feel the need to repeat and clarify the information about such elements. The Contents element, in contrast with Purpose and Time who have only one value, can be mentioned several times actually providing new information. This was found to happen: “*It contains an introduction, goals, contents, (...) In general, it contains the course syllabus (...) It has no tables.*” (I10); “*(...) it was a large document (...) it had maps (...)*” (I4)

Overall, no large bias was introduced by the repeating of elements within a story.

3.2.4. Story Structure

In a narrative, not only the constituting elements themselves are important. It is also crucial to understand how those elements relate to each other. To that end, we performed a relational analysis, registering all transitions between story elements. Only transitions that naturally occurred were considered, and not those resulting in induced story elements, as explained in Section 3.1.1.

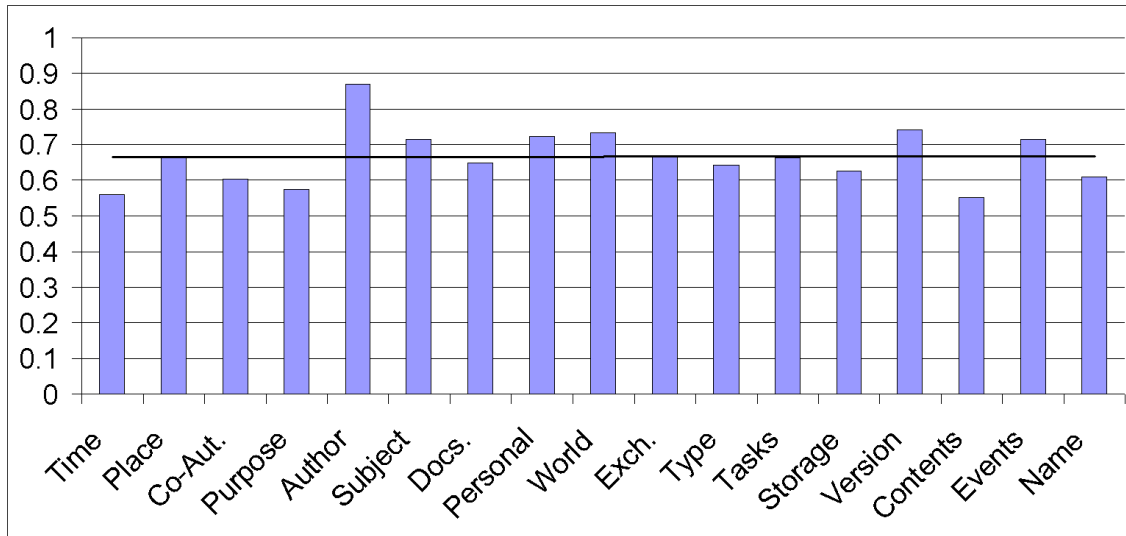


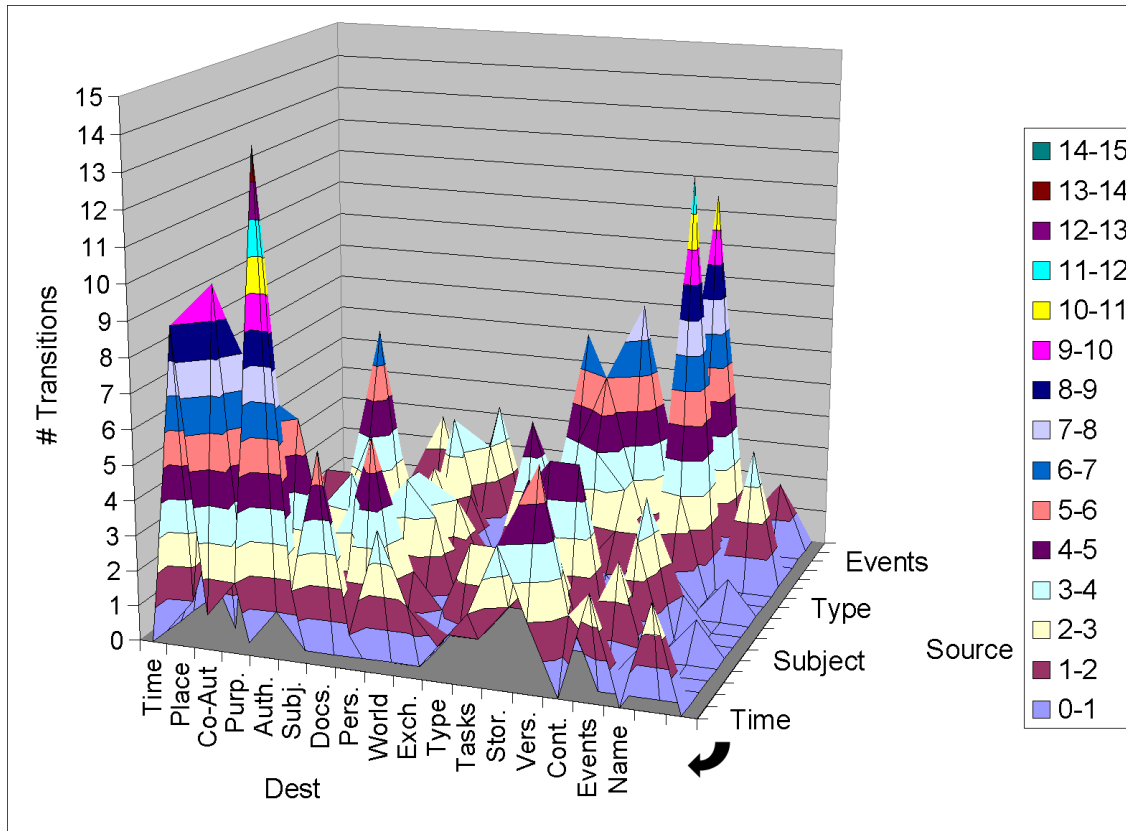
Figure 3.9: Occurrences / Frequencies Ratios

	Time	Place	CAut	Purp	Auth	Subj	Docs	Pers	Wrld	Exch	Type	Task	Stor	Vers	Cont	Evts	Name
Time	9	1	2	14	2	6	1	4	2	1	2	4	2	0	3	0	3
Place	1	10	0	1	0	0	0	0	0	1	1	2	6	2	0	0	0
Co-Aut.	2	0	7	1	2	1	6	0	0	1	1	2	2	1	0	1	0
Purpose	6	5	3	6	2	4	2	2	1	3	3	5	3	0	3	0	0
Author	1	0	3	0	0	3	0	1	0	2	1	1	0	0	0	0	0
Subject	7	0	0	3	4	2	4	3	0	2	2	5	5	0	3	0	1
Docs.	1	1	1	0	0	3	4	0	1	2	4	4	1	0	4	0	0
Personal	3	0	2	0	0	1	1	3	1	0	0	1	0	0	2	1	0
World	1	0	0	1	0	0	0	2	1	0	1	0	1	0	0	0	0
Exch.	0	1	2	4	1	2	1	0	0	5	2	1	2	2	1	0	1
Type	1	1	2	7	0	3	1	0	0	2	2	4	3	0	3	0	0
Tasks	0	1	2	4	0	2	2	0	0	2	4	6	3	0	12	0	0
Storage	2	2	1	0	0	4	1	0	0	2	7	2	8	2	4	0	4
Version	0	1	0	0	0	0	3	0	1	1	0	0	2	3	2	0	0
Contents	1	0	1	1	0	1	4	0	0	0	4	4	4	0	11	1	0
Events	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0
Name	0	0	0	3	0	1	0	0	0	0	0	0	2	0	0	0	2

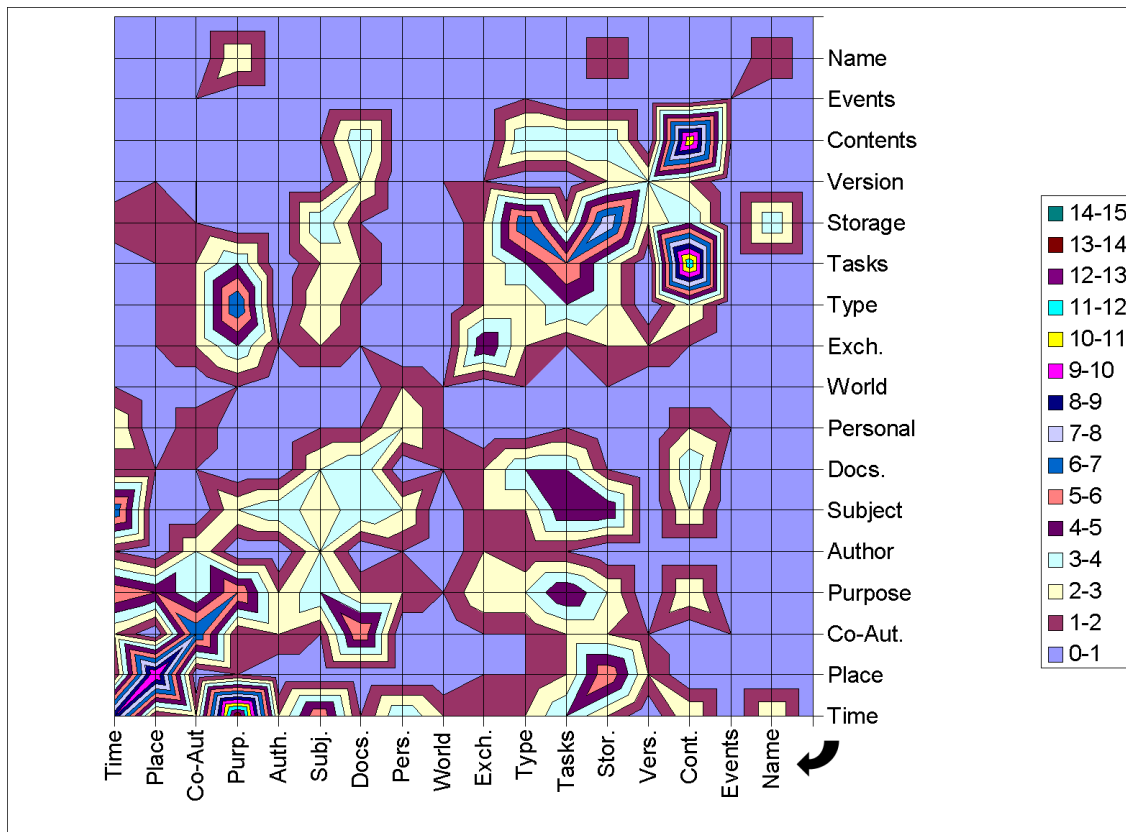
Table 3.12: Transitions Between Story Elements

We found, as expected, that for a large number of element pairs, no transitions were registered at all in any of the sixty stories (Table 3.12 and Figure 3.10, the most frequent transitions are highlighted). Even taking into account the associative nature of human memory, they were never mentioned together, such as Subject-Place or Time-Version. Overall, only 36.7% of possible element pairs transitions occurred more than once. This supports the idea that those elements are unrelated in the subjects' minds.

There are some pairs whose relation is relatively stronger, however. It is easy to notice three very strong transitions: **Time-Purpose**, **Tasks-Contents**, and **Contents-Contents**.



(a) 3D map



(b) surface map

Figure 3.10: Element Transitions

Not as strong but also important are **Place-Place**, **Time-Time**, **Storage-Storage**, **Subject-Time**, **Co-Author-Co-Author**, **Type-Purpose**, and **Storage-Type** (ordered by decreasing importance).

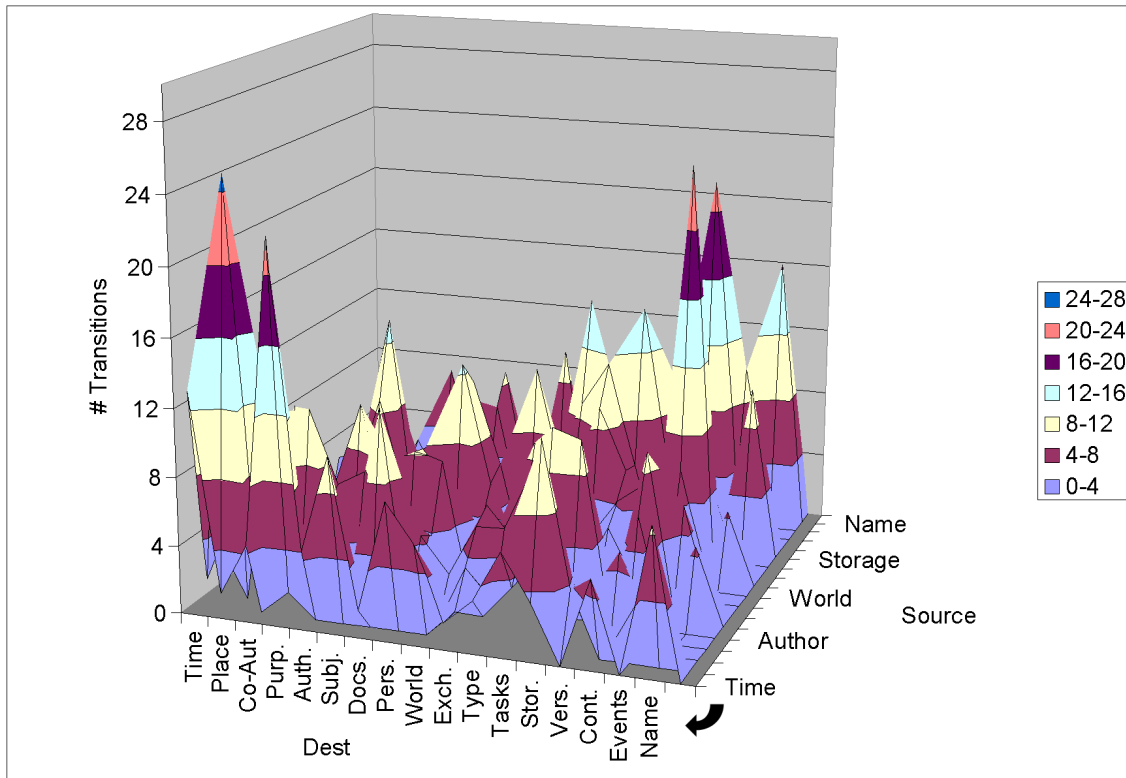
The first thing that can be noticed from those transitions is that it is often common for transitions between equal elements to appear. We can see this looking at the lower-left/upper-right diagonal. This again shows how users tend to reinforce elements they think are most important, harder to explain, or about which they remember more.

We were concerned that some transitions might have appeared a large number of times because the elements that constituted them were frequent, and not because they were actually associated to each other. For instance, if elements A and B occurred only three times each and always consecutively, the transition value, three, would be less than if A and B occurred ten times each and four of those times they appeared together. However, in the first case A and B are clearly more strongly related to each other (B follows A 100% of the time). Trying to solve this problem, we normalised the transition numbers, dividing them by the sum of the frequencies of the elements involved. The rationale for this is that we must have witnessed a larger number of transitions between elements with higher frequencies to be confident that some significant relation exists.

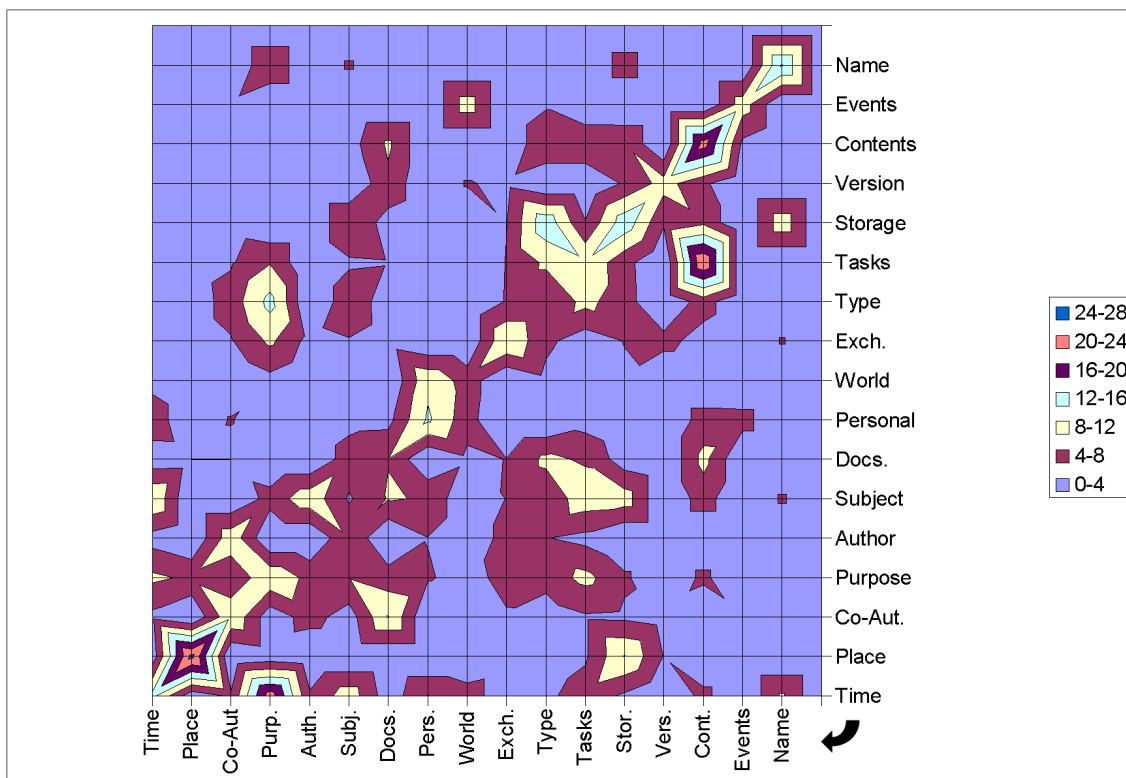
	Time	Place	CAut	Purp	Auth	Subj	Docs	Pers	Wrld	Exch	Type	Task	Stor	Vers.	Cont.	Events	Name
Time	13.09	1.54	4.53	22.19	4.65	10.28	1.61	8.13	5.90	1.69	3.08	6.07	3.34	0	5.19	0	8.63
Place	1.54	25.00	0	1.39	0	0	0	0	0	1.59	1.54	3.41	12.08	3.77	0	0	0
Co-Aut.	2.94	0	11.21	1.54	5.64	2	12.15	0	0	3.57	1.75	3.51	3.28	2.04	0	3.13	0
Purpose	8.84	7.25	4.64	11.11	6.90	7.83	5.71	4.44	2.38	5.39	5.49	9.18	4.77	0	4.97	0	0
Author	2.33	0	10.64	0	0	7.50	0	3.85	0	6.08	4.17	2.70	0	0	0	0	0
Subject	11.79	0	0	5.70	10.63	3.57	8.19	7.78	0	4.17	4.93	10.52	9.97	0	5.88	0	4.55
Docs.	1.61	1.79	1.85	0	0	6.27	7.91	0	3.13	4.08	8.96	7.41	1.82	0	9.14	0	0
Personal	6.69	0	4.35	0	0	2.50	2.78	12.56	5.56	0	0	2.33	0	0	5.76	5.56	0
World	2.56	0	0	2.70	0	0	0	11.11	6.25	0	2.86	0	3.13	0	0	0	0
Exch.	0	1.59	3.64	8.89	2.63	3.77	2.04	0	0	11.73	4.96	1.79	4.17	5.44	2.13	0	4.35
Type	1.85	1.54	5.89	13.34	0	6.67	1.96	0	0	4.25	4.35	9.20	5.58	0	5.97	0	0
Tasks	0	1.96	3.51	7.64	0	3.39	3.71	0	0	4.17	8.87	11.47	5.35	0	23.73	0	0
Storage	3.09	3.70	1.64	0	0	7.97	2.17	0	0	3.87	14.77	4.26	14.54	4.50	7.11	0	10.26
Version	0	1.89	0	0	0	0	7.01	0	4.35	2.94	0	0	4.50	—	4.88	—	0
Contents	1.59	0	1.79	1.67	0	2.17	8.76	0	0	0	7.70	7.85	7.53	0	21.79	2.94	0
Events	0	0	3.13	0	0	0	0	0	10.00	0	3.45	0	0	—	0	—	0
Name	0	0	0	7.47	0	4.55	0	0	0	0	0	0	5.90	0	0	0	16.25

Table 3.13: Normalised Transitions Between Story Elements

We can see the normalised values in Table 3.13 and Figure 3.11 (a scale factor of 100 was applied to each cell to make values more perceptible). It is easy to confirm looking at the table, that the stronger transitions are **Place-Place**, **Time-Purpose**, **Tasks-Contents**, and **Contents-Contents**. Those are exactly the four element pairs we found to be more important while disregarding the element frequencies. Only the order changes: Place-Place moves from the fourth place to the first, showing that, indeed, it is more important that it would seem at first.



(a) 3D map



(b) surface map

Figure 3.11: Element Transitions (Normalised)

The next most important element pairs, with similar values, are **Storage-Type**, **Storage-Storage**, **Type-Purpose**, **Time-Time**, **Personal-Personal**, and **Co-Authors-Other Docs**. Again we find several coincidences with the pairs found before. Furthermore, if we look at the tables we'll see that the pairs that don't coincide have the next higher values. This confirms that, independently of the number of times they were mentioned, there are indeed several relevant element pairs. Noteworthy is the fact that the normalised values show that the most common transition pairs are, if anything, more important than it would at first appear: their values are significantly larger in relation to those of other transitions.

Another interesting aspect to consider is, given a certain story element, finding what is the most likely to follow. To estimate this we calculated the transition probabilities from each element to another (Table 3.14).

	Time	Place	CAut	Purp.	Auth.	Subj.	Docs.	Pers.	World	Exch.	Type	Tasks	Stor.	Vers.	Cont.	Events	Name
Time	0.161	0.018	0.036	0.25	0.036	0.107	0.018	0.071	0.036	0.018	0.036	0.071	0.036	0	0.054	0	0.054
Place	0.042	0.417	0	0.042	0	0	0	0	0	0.042	0.042	0.083	0.25	0.083	0	0	0
Co-Aut.	0.074	0	0.259	0.037	0.074	0.037	0.222	0	0	0.037	0.037	0.074	0.074	0.037	0	0.037	0
Purpose	0.125	0.104	0.063	0.125	0.042	0.083	0.042	0.042	0.021	0.063	0.063	0.104	0.063	0	0.063	0	0
Author	0.083	0	0.25	0	0	0.25	0	0.083	0	0.167	0.083	0.083	0	0	0	0	0
Subject	0.171	0	0	0.073	0.098	0.049	0.098	0.073	0	0.049	0.049	0.122	0.122	0	0.073	0	0.024
Docs.	0.038	0.038	0.038	0	0	0.115	0.154	0	0.038	0.077	0.154	0.154	0.038	0	0.154	0	0
Personal	0.2	0	0.133	0	0	0.067	0.067	0.2	0.067	0	0	0.067	0	0	0.133	0.067	0
World	0.143	0	0	0.143	0	0	0.286	0.143	0	0.143	0	0.143	0	0	0	0	0
Exch.	0	0.04	0.08	0.16	0.04	0.08	0.04	0	0	0.2	0.08	0.04	0.08	0.08	0.04	0	0.04
Type	0.034	0.034	0.069	0.241	0	0.103	0.034	0	0	0.069	0.069	0.138	0.103	0	0.103	0	0
Tasks	0	0.026	0.053	0.105	0	0.053	0.053	0	0	0.053	0.105	0.158	0.079	0	0.316	0	0
Storage	0.051	0.051	0.026	0	0	0.103	0.026	0	0	0.051	0.179	0.051	0.205	0.051	0.103	0	0.103
Version	0	0.077	0	0	0	0	0.231	0	0.077	0.077	0	0	0.154	0.231	0.154	0	0
Contents	0.031	0	0.031	0.031	0	0.031	0.125	0	0	0.125	0.125	0.125	0	0	0.344	0.031	0
Events	0	0	0.25	0	0	0	0	0	0.25	0	0.25	0	0	0	0	0.25	0
Name	0	0	0	0.375	0	0.125	0	0	0	0	0	0	0.25	0	0	0	0.25

Table 3.14: Transition Probabilities

As can be seen, there are no almost certain transitions to be expected. The most probable are **Place-Place** (0.417), **Name-Purpose** (0.375), **Contents-Contents** (0.344), and **Tasks-Contents** (0.316). These are, with the exception of Name-Purpose (not really significant given how rarely the Name element was mentioned), the most frequent transitions. The other frequent transition noted above, **Time-Purpose**, appears with a probability of only 0.25 (even so, a rather high value, in comparison with most transitions). Also with a fairly high transition probability we find **Co-Author-Co-Author** (0.259), **Author-Co-Author** (0.25), **Author-Subject** (0.25), **Place-Storage** (0.25), **Type-Purpose** (0.241), **Version-Other Documents** (0.231), **Version-Version** (0.231), and **Co-Author-Other Documents** (0.222). We did not consider transitions involving World Events, Name or Events, since those elements occurred so infrequently that no reasonable conclusion can be inferred from the data.

Several frequent transitions have high probabilities, but that is not always the case. For instance, Time-Time, the fifth most frequent transition, only has a probability of 0.161. This difference is due to the fact that there are many transitions that start with time, making the respective element pairs very frequent. However, since they are all frequent, the probabilities of each when compared to the others (with origin in time) are not very high. In other cases we find the reverse: the transitions have low frequency overall values, but when compared with others with the same origins are fairly probable. It is the case of Author-Co-Author, for instance.

Finally, we would like to discover if some elements are associated to one another only in a given order (as suggested by a close inspection of Table 3.14), or if that order is not important. To verify how symmetrical are the transitions, we computed, for each transition, a symmetry estimate, given, for each element pair (A, B) by $(NTrans(A, B) - NTrans(B, A)) / (NTrans(A, B) + NTrans(B, A))$. The result is a number between -1 and 1. Those values represent the most extreme asymmetries. The closer to 0, the more symmetric is the relation.

	Time	Place	CoAut	Purp.	Auth.	Subj.	Docs.	Pers.	World	Exch.	Type	Tasks	Stor.	Vers.	Cont.	Events	Name
Time	0																
Place	0	0															
Co-Aut.	0	-	0														
Purpose	-0.4	0.67	0.5	0													
Author	-0.33	-	0.2	-1	-												
Subject	0.08	-	-1	-0.14	0.14	0											
Docs.	0	1	-0.71	-1	-	-0.14	0										
Personal	-0.14	-	1	-1	-1	-0.5	1	0									
World	-0.33	-	-	0	-	-	-1	0.33	0								
Exch.	-1	0	0.33	0.14	-0.33	0	-0.33	-	-	0							
Type	-0.33	0	0.33	0.4	-1	0.2	-0.6	-	-1	0	0						
Tasks	-1	-0.33	0	-0.11	-1	-0.43	-0.33	-1	-	0.33	0	0					
Storage	0	-0.5	-0.33	-1	-	-0.11	0	-	-1	0	0.4	-0.2	0				
Version	-	-0.33	-1	-	-	-	1	-	1	-0.33	-	-	0	0			
Contents	-0.5	-	1	-0.5	-	-0.5	0	-1	-	-1	0.14	-0.5	0	-1	0		
Events	-	-	0	-	-	-	-	-1	1	-	1	-	-	-	-1	0	
Name	-1	-	-	1	-	0	-	-	-	-1	-	-	-0.3	-	-	-	0

Table 3.15: Symmetry Estimates

In Table 3.15, we show the results for all element pairs. Cells marked with a dash ('-') correspond to element pairs for which no transition was verified on either direction. There are several absolutely symmetric transition pairs apart from the diagonal (highlighted), but more often we find pairs where transitions were found in only one direction.

We are, of course, more interested in understanding what is the symmetry of the most frequent transitions (apart from those in the diagonal). Those transitions are **Time-Purpose**, **Tasks-Contents**, **Subject-Time**, **Type-Purpose**, and **Storage-Type**. Of all those element pairs, only Subject-Time is almost perfectly symmetric (0.08). The others are asymmetric but not entirely so: they have measures with absolute values of 0.4 (Time-Purpose,

Type-Purpose and Storage-Type) or 0.5 (Tasks-Contents). This is not surprising: if they were symmetric, their mirror images should also appear in the list of the most frequent transitions.

3.2.5. Other Results

Apart from the twenty interviews we just analysed in depth, there were two more that didn't follow the pre-arranged interview script, making impossible their comparison to the others and a similar analysis. However, they are not without interest and the insights they contain are described in this section.

The subject of Interview 21 is an Operations Director of a large media corporation. He deals with several products and departments and organises his documents in a large but well defined folder hierarchy. Rather than remembering all documents he sees, for all subjects or areas, he *"[goes] to talk with someone of [the document's] area, and she'll tell me where the document is"*. The name of each individual document is also subject to a rigid nomenclature, that allows an immediate identification of its subject, purpose, kind and who is responsible for it.

Also important seems to be his mailbox. He never deletes a message and has a fairly elaborated message folder hierarchy where all emails are classified by hand. Those folders are organised not only by product or area, but also by person. All mails he sends to someone are stored in the appropriate folder. About the messages themselves, *"the subject and size are fundamental to identify and locate a document"*.

In the other interview, Interview 22, the subject reports how he usually resorts to search tools to look for specific words in documents, in order to find that which he is seeking. He tries not to resort to system dates, since he works in different machines and the dates are not always synchronised. This is positive evidence supporting the need for applications and interfaces to seamlessly handle different computers or locations [Nielsen 2002]. When looking for a document, he is often able to identify the correct one from a birds-eye-view of its contents.

3.2.6. Archetypical Stories

As expected and verified in the interview analysis, stories about document share several properties. In particular, their structure follows a certain pattern, as revealed by the existence of preferred transitions between story elements, described above. Those results, while valid, are difficult to use in a straightforward way in the design of narrative-based document retrieval interfaces. As they stand, it is hard to get a clear picture of the sto-

ries' overall structure. Thus, it was decided to use the data to infer archetypical story structures, easier to understand and analyse.

In order to discover what we could consider to be 'typical' or 'average' stories, we had to take into account not only the transitions between story elements, but also the probability of a particular element occurring in a given story position, or showing up in the story at all. Considering only the transitions, we might know that a purpose element often follows time references, but not if that occurs in most stories or whether it appears near the beginning or the end of the narrative.

A way to model stories that takes all those factors into account are Hidden Markov Models (HMM), trained with the element sequences of the stories we collected. All story elements were taken into account, both spontaneous and induced. It could be argued that induced elements should not be considered, as they could have been mentioned in different orders if the questions asked during the course of the interview were different or presented in other combinations. However, since we are trying to find typical stories, it would make no sense to consider only story parts where no induced elements appear. Important information could be lost, either regarding the position of elements in the story or regarding the overall story structure.

We trained the HMM considering the states to be the several story elements, taking advantage of our existing computation of element transition probabilities. As outputs, we considered the several time instants, calculating the probabilities of the model being on a given instant when some story element was seen. As we already had element transition probabilities, we did not need to train the HMM. Rather, we initialised it with the values we had previously computed. The initial state probabilities were calculated by looking at the first element in all stories and determining their relative frequency. The element transition probabilities were used to initialise the hidden state transition probabilities matrix. As we considered 18 different elements, there are 18 hidden states in our model. In possession of those values, the Viterbi Algorithm was used to discover what state sequence (in this case, story elements) would have most probably occurred when we witnessed the 'outputs' that correspond to several different ordered time instants.

We faced an additional difficulty: not all stories have the same length. This makes it harder for the model to estimate what element should appear on a certain position: if most stories fall short of that position, a small number of narratives could have a large influence on what the model thinks should be present there. This accounts for some 'weird' results towards the end of the inferred stories, where low-frequency elements show up. We tried to minimise that effect without unnecessarily limiting the stories that were produced. We generated stories with the average length of the story type they pretend to represent. We then discarded the final third of the resulting story, roughly corresponding to the standard deviation of story length (marked in *italic* below).

We trained the HMM with five different story sets, corresponding to the three different document types, all documents produced by the user (Recent and Old), and all stories.

Recent Documents

The typical story told about recent documents is nineteen elements long (the average length of such stories) and is structured as follows:

Time, Purpose, Time, Purpose, Subject, Author, Co-Authors, Co-Authors, Exchanges, Exchanges, Versions, World, World, World, World, Tasks, Contents, Other Documents, Name

The story begins with some short sentences revolving around Time and Purpose of the document, reiterating a little to clarify those issues. Then it moves out to other elements, such as the Subject, Author, Co-Authors and Exchanges, that appear naturally after references to Co-Authors, with whom the document could have been exchanged. It is rather surprising to find that World Events shows up in a fairly prominent position, given the low frequency in which that element shows up in actual stories. We attribute this to the fact that in those story positions, no other element clearly occurs more frequently than another, compounding on the trustworthiness issue discussed above. It is our opinion that, in practice, only one reference to world events should be taken into account.

Old Documents

This time, we considered a story length of seventeen elements. The typical story for these documents is the following:

Purpose, Time, Purpose, Place, Storage, Author, Co-Authors, Co-Authors, Exchanges, Exchanges, Type, Contents, Storage, Tasks, World, Personal, World

The story is fairly similar to that of a recent document. The major differences are the mentions to Place and Storage early in the story. It seems those elements are important to properly situate the document. Storage is, of course, related to Place. It appears document Type and Contents are also fairly important, from their position in the stories.

Documents Produced By the User

This story was inferred from both the stories about recent and old documents, with the average length of eighteen:

Time, Purpose, Time, Place, Storage, Author, Co-Author, Co-Author, Exchanges, Exchanges, Type, Contents, *Storage, Co-Authors, Exchanges, Type, Contents, Versions*

This story, as expected, shares most features of the stories for Recent and Old documents. In fact, nearly identical to the story for older documents, suggesting those stories follow less flexible patterns than the ones about recent documents. The differences appear towards the end, below the confidence threshold. The fact that the trends for Older documents are predominant also means they are more trustworthy. The fainter pattern for Recent documents reflects the more diverse nature of such stories, resulting in difficulties in predicting a typical narrative.

Other Documents

With regard to documents not created by the user, the average length is only twelve, accounting for a shorter story. This was to be expected since users are bound to remember more details about documents they themselves produced.

Time, Purpose, Author, Exchange, Tasks, Type, Co-Authors, Other Documents, *Other Documents, Personal, Contents, Contents*

This story's structure is fairly different from those relative to other documents. Time and Purpose remain important in the beginning of the story. Then the user moves on to the document's Author, Exchanges (given that most documents are sent to them by someone or sent by them to someone else), Tasks, Type (tasks undertaken to visualise documents of a given type, for instance). Other Documents appear next, often those that used the document being described as a reference.

Overall Story

Finally, we produced a typical narrative corresponding to all sixty stories told by the participants. The average length of all stories, sixteen, was used:

Time, Purpose, Time, Place, Storage, Co-Authors, Co-Authors, Co-Authors, Exchange, Exchange, Type, *Contents, Other Documents, Personal, World, Personal*

A Closer Look...

Not all elements appear in all inferred stories. This is due to the fact that, especially for less frequent and important elements, the missing elements didn't appear enough times in stories to establish a trend. This was to be expected, taking into consideration the transition probabilities discussed in the previous section. However, an important fact asserts itself: in all stories the Time when a document was produced or read and its Purpose are the first to be mentioned. Other frequently mentioned elements (as shown in Figure 3.4) such as Place, Storage, Type, Exchanges and Co-Authors also seem to be important, appearing in the first two thirds of the inferred stories. This shows that these stories reflect our previous findings.

3.3. Guidelines

From the results presented in the previous section, some interesting trends emerge. In this section, we will distill them into a set of principles for the design of narrative-based interfaces for the retrieval of personal documents.

G1 – Customisation

We found little relevance of personal factors such as gender and age to the way stories are told. They seemed to influence none of the story measures we collected, such as story lengths, transition/size ratios (an estimate of how in control of their stories the subjects were), sizes of story trains, or frequencies of changes in subject. The only exceptions were that women tend to tell longer stories than man, and that older persons are marginally more in control of their stories than younger ones.

In general, *no user customisation will be necessary in relation to what to expect from a story*. This does not preclude other customisations, such as adapting the interface to the particular subjects users usually work on, or to better visualise a particular Personal Document Space. The only relevant user-dependent feature in the story capturing process is to try to get more information from female users, since it is easier for them to provide it.

G2 – Memory

We expected to find that a user's memory about a document fades with time. Indeed, we noticed that a document's Subject is mentioned more often in stories about Recent documents than in those about Old ones. However, no significant difference was found for the remaining elements, story length, or transition numbers. Likewise, no differences were recorded in the percentages of induced elements stories: nearly half the narratives were spontaneously told by the subjects.

Where time seems to be of importance is for the numbers of story trains. Stories about recent documents are divided into more story trains than others. This shows that those stories are more fragmented. Another difference seems to relate to how accurate the story elements are: users still remember them for older documents, but with less detail and confidence.

What does seem to affect the knowledge a user could remember about a document is its origin. Stories about documents created by the user, regardless of when, are longer. Some elements such as Place or Purpose are mentioned more often, suggesting they are easier to remember.

In short, some differences in the story structures and accuracy can be expected according to the age of the document being described. However, the biggest differences derive from the document's origin. It is important to determine it early in the narrative, to correctly form expectations about what can be found ahead in the story.

G3 – The Importance of Dialogues

For some story elements, a significant number of occurrences were induced by some question posed by the interviewer. Elements such as time, place, and other documents are among them. They are also some of the most frequent elements, suggesting that users consider them important and can actually remember them, if asked.

It is important to establish dialogues with users in order to obtain all information they can actually remember. Some care should be taken about changes in subject. However, they are fairly rare and should pose no significant problem.

The dialogues should not waste time and resources trying to discover certain elements, such as Author, Personal Events, World Events, Events and Names. They are rarely mentioned but generally spontaneously, showing that if they are remembered at all, they will most likely be volunteered with no need for inducement.

G4 – Context-Dependent Information

It is common for stories to include indirect references to elements that are taken for granted by the storyteller. For instance, references to the Place where a document was produced and its Author are based on assumptions or contextual information. No specific places or names are mentioned because they seem obvious to the person telling the story. This happens, for instance, if a document was produced in the place the story is being told, or if it arrived by email and the user only has email access at work. *It is important to take the context in which the story is told into consideration, comparing it to a model of the users' world and of users themselves* (more on this below).

G5 – Ambiguity

Some level of ambiguity is common in stories. For instance, references to time become more inaccurate for older documents. Instead of specific dates or times, entire months are mentioned. Something similar occurs when trying to remember names of authors or co-authors. The user can remember what the name sounded like, or that it had some co-authors, but not their actual names. *Some level of ambiguity must be tolerated by narrative-based interfaces.* Techniques to automatically disambiguate stories with the help of context and user and world models are to be considered. Users themselves often try to help, providing information about the same element more than once in the same story. That willingness to help should be encouraged and used.

G6 – World and User Models

When referring to such elements as Purpose, World Events or Personal Events, a wide range of information can be conveyed. It is probably impossible to simply use keywords extracted from the stories to effectively gain some insight on what document is being talked about. Trying to understand those elements just by looking at what was said is also insufficient, due to great numbers of things that would be important to understand them but are taken for granted and not explicitly mentioned. To aid in that understanding, *a model of the world around the users and of the users themselves (including typical activities, co-workers, etc.) should be used.* Important information can also be found on the user's agenda, and also in that of his friends or co-workers. Some facts from the 'wider world', such as important news could also be helpful, albeit more rarely.

G7 – Overall Document Structure

Users remember more easily overall document structures than actual keywords or phrases in that document. *Some technique that identifies the overall structure or visual appearance of a document and can use that information to differentiate among several documents would be useful.* This suggests that most existing approaches based on keyword search, such as Google Desktop, might not provide all mechanisms users might consider relevant to help them retrieve their documents

G8 – Events During the Interaction With the Document

In short, *they are not relevant.* It was extremely rare for any such events (someone entering the office, a phone call, etc) to be remembered.

G9 – Recursive Stories

When describing related documents, it is common for several information elements pertaining those documents to be told. They can constitute small recursive stories (stories within a story). *Special care should be taken to capture those elements, which provide important information, while keeping in mind they relate to a document different than the one the story is about.* Also, those stories should somehow be controlled in order to prevent the storyteller from losing himself in them, sidetracking from the document he really wants to find.

G10 – Expected Elements and Structure

The stories we analysed share, up to a point, similar structures. Designers of narrative-based interfaces should take advantage of those similarities. They will allow the system to know what to expect from the stories, help guide the user towards providing useful information, and collect that information.

Most stories are composed by several trains, often no more than three or four. Of all trains, *special attention should be given to the first two, where half the information provided in the whole story will be concentrated.* The last can be useful, too, but not as much.

Some story elements are more frequent than others, and should be expected more often (see Section 3.2.3). Several will be mentioned only if prompted by some external factor (Section 3.3). This information is useful, helping decide if some more information should

be expected (if some frequent elements weren't yet mentioned) or not. It will help decide whether it's worthy to invest some time and effort to discover more elements or not.

G11 – Probable Transitions

Of all possible transitions between different story elements, only 37% have some credible probability of showing up. Of those, five are to be expected fairly often (see section 3.2.4). Combining this information with the probabilities of what will be the next element, given the current point in the narrative, it will be possible to build expectations of what the next element in the story will be. This will help recognise it and extract all relevant information, facilitating disambiguations.

4

Crafting the Interface

In Chapter 3 we described the analysis of document-describing narratives. That analysis gave us an insight on what to expect from those stories, resulting in a set of guidelines for the design of narrative-based interfaces for personal document retrieval. However, while those guidelines provide a starting point, they can be instantiated in different ways. In this chapter we will describe how the actual interface was designed, with the help of user evaluation of low-fidelity prototypes.

4.1. Pre-Requisites

A properly designed interface should, besides the usual usability concerns, take into account several requisites. Overall, the interface should allow users to tell their stories to the computer in a way that is similar to what takes place when telling stories to humans. This can be accomplished by closely following the previous chapter's design guidelines and other insights gained from the analysis of the stories. Also, our own research goals impose several *a priori* restrictions on the interface. We will now describe the main considerations taken into account when shaping the interface.

4.1.1. Universality

It is our research goal to show how narratives can be used to retrieve any kind of personal document. Traditional retrieval interfaces either focus on text-based documents (and keyword search) or are otherwise specially tailored for other kinds of documents (images, songs, etc.). Our narrative-based interface should be able to properly handle and display those documents. Also, it should be expressive enough to allow stories about different document kinds to be told. This includes allowing references to relevant aspects of those documents, as found in stories.

4.1.2. The Importance of Time

We verified that time plays a decisive role when the users describe their documents. One of the first things all users try to establish is *when* the document was created or read. Time is so important that we found that element to appear, on average, nearly twice in each story. Thus, it stands to reason that it should be given special importance in the interface, to allow the users to specify it without problems.

4.1.3. The Need for Dialogues

One of the major results of our previous study was that dialogues are essential in helping users remember all the information they possess about a document. In fact, it was common for some story elements to be mentioned only after some question or sentence reminded the users of them. Thus, the interface should somehow mimic this process, reminding users of other potentially relevant information they could tell about their documents.

4.1.4. Other Documents

We also witnessed that users often described other documents when trying to remember the one they are looking for. Those other documents could have been created for the same purpose as the one being sought, could have been read as a reference for it, or some other reason altogether. The interface should take this into account and allow users not only to mention other documents, but also to specify in which way they relate to the target document. However, this should be done in some constrained, well defined way, to prevent confusing the users by over-complicating the interface or the storytelling process.

4.1.5. Archetypical Story Structures

In the previous study it became evident that some story elements are more common than others, and so are some element transitions. With the help of that information, it was possible to train a Hidden Markov Chain to infer archetypical story structures. Those structures should be used to help both the users and the interface by guiding the storytelling process.

4.1.6. Adaptability

While archetypical story structures should be used, it must not be forgotten that a somewhat large variability was found in the stories of different users. Consequently, the interface should not be rigid, allowing the users to tell the stories as they see fit and, if possible, adapting to each users' needs.

4.1.7. Understanding the User's World

We found that information about the user's life (deadlines, birthdays, meetings, conferences, etc.) can be important, either by itself, as a document's Purpose, for instance, or to provide some context to the story. That information is likely to be found on the user's agenda, or in his email messages. The interface should help the users access that information in an effective way. Also, the underlying system should use that information to better understand the users' stories.

4.1.8. Inaccuracy / Uncertainty

The elements in the users' stories have variable precision. For instance, while in one story the Time element can be told as "last Thursday in the afternoon", a fairly precise statement, in another it can be referred to as "around summer...". As the users' memories are bound not to be perfect, this might lead to inaccuracies, especially when users try to be very precise and/or when describing older documents (this issue is further studied in Chapter 6).

The interface must, thus, be able to deal with inaccurate information. It should provide mechanisms for the users to correct information already in stories, and to add further information to better explain something they had previously stated. It should adhere to the *fail-soft* principle, which states that, even if an error is committed, it should be as inconsequential and easy to recover from as possible. In our case, even if some story

element is not very precise or accurate it is hoped that the target-document can still be found, and if not, maybe simply correcting a story element (as opposed to re-telling the whole story) is necessary.

4.1.9. Reflexive Transitions

Reflexive transitions were often common in document-describing stories (when the same element was mentioned two or more times in a row). Usually, they occurred when the users felt the need to complete some information they had just mentioned. To make things simpler, the interface should make it easy for the user to reiterate story elements.

4.1.10. Visual Memory

It was verified that in many stories the users mention the overall visual aspect of a document, even for those that are not images. The interface should take advantage of the users' visual memories, especially when displaying possible search results to them. By displaying their appearance, the interface can make the target-document easier to identify, disrupting the flow of the story as little as possible.

4.1.11. Don't Rely Solely on Text

As stories seem inherently text-based, it is tempting to design an interface that is based solely on the entry of free-form text, either in written or spoken form. Although this might have some advantages, mirroring more closely the telling of a story in a more informal context (to another human, for instance), it is an approach that should be avoided as not to compromise the interface's usefulness.

First and foremost, there are problems inherent to the use of natural language in computing systems. While the discourse domain are the users' documents, the high variability of what can be found in them and the context that surrounds them makes it impossible to constrain beforehand what the users can mention. Thus, trying to understand natural language with no restrictions, and given the diversity of things that can be mentioned in a story, would probably be an insurmountable.

Also, just allowing users to tell their stories, with no kind of structure at all, will allow them to digress, as found in the study described in Chapter 3. The spurious information thus introduced in the story will make finding the document even more difficult.

Furthermore, using a purely text-based interface would make it strictly dependant on the

language of the user. It might not be easy to port it across language and cultural barriers.

Finally, relying solely on text entry would make the interface harder to use. It would require the free-form introduction of a fairly large amount of text, which can be a time-consuming process. Also, not all users have the same writing skills or patience and time to do so. Thus, resorting solely to natural language would alienate some users.

4.2. The Prototypes

To correctly assess what interface would better embody the guidelines and restrictions at our disposal, we decided to build and evaluate low-fidelity prototypes. A low-fidelity prototype is created using pencil and paper. It is not a computer application. Instead, the designer draws a sketch of the shape the interface should take. This sketch does not contain any detailed text or images, nor is it created with aesthetic concerns in mind. Instead, the major areas of the interface are identified, together with just enough information to understand it and simulate some pre-defined tasks. Low-fidelity prototypes are a cheap and quick way to test different interface proposals, as they require no coding or debugging. Furthermore, they allow the designers to notice major design flaws early in the interface creation process, and correct them with ease.

In this study, we analysed two different Low Fidelity Prototypes of interfaces that allow document-describing stories to be told. The first of those prototypes is based on direct manipulation of on-screen graphic elements, following the conventions of common GUI-based applications. The second actually tries to construct and display the story of each document as a small (interactive) text narrative. These options represent two radically different look-and-feels and interaction paradigms.

4.2.1. Prototype A — “ Dialogues and Direct Manipulation”

This prototype is based on the sequential introduction of the several story elements with the help of specialised dialogues, one for each different element, that are graphically represented as individual entities that can be directly manipulated by the user (Figure 4.1).

The screen is divided into three different areas. In the dialogue area (1), the several dialogues are presented to the users. Each allows users to enter a different story element. They are presented in succession to the user. The order in which this is done is that of the archetypical stories inferred in Chapter 3. Different orders are considered, accounting for different document types.

If the element being suggested to the user is not what he would like to mention at that

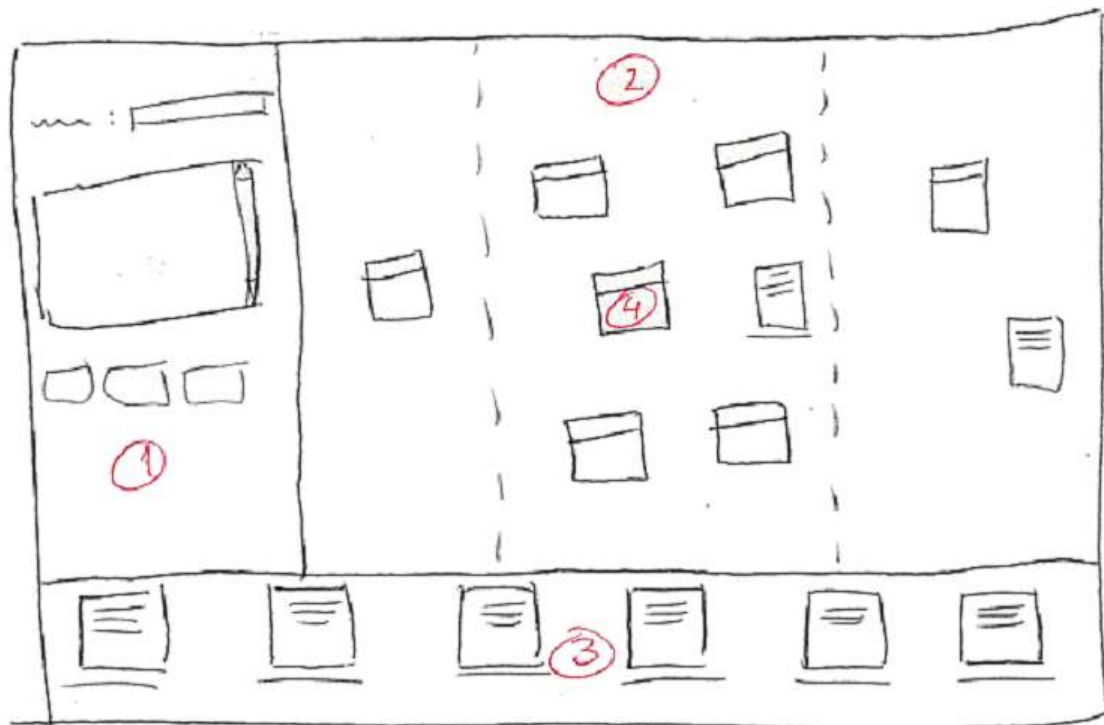


Figure 4.1: Prototype A: Overall View

time, the element's name can be pressed as a button allowing the choice of another element from a drop-down list. The dialogue will promptly reflect that choice. These dialogue boxes help the users remember the elements they can mention, and can play the role of an listener dialoguing with the user to find out more.

Each time an element is told using the dialogues, a small rectangular box (4) will appear in the central area (2). Those boxes contain the story elements the user has mentioned, helping to maintain an overall idea of the story. They can be directly manipulated, allowing the corresponding elements to be changed (with the help of the appropriate dialogue). The position of those boxes in the central area is not irrelevant. In fact, the leftmost and rightmost parts of that area (separated from the rest by a dashed line) represent the past and the future of the target-document, respectively. So, it is possible for the user to move boxes there to establish a chronological order among the story elements. For instance, an Other Document about a document that was written before the target-document should be moved to the past area. This will allow temporal information to be conveyed not only in absolute terms (the dialogue), but also by establishing time relations among story elements ("the document was written BEFORE the conference", etc.)

Finally, the third area, at the bottom (3), will contain several document suggestions made by the interface from the story that has been told so far. It is expected that the target-

document will, eventually, appear there. Those documents, if recognised as having some relation to the target-document, can be dragged to the central area, after specifying how they are related to the document being searched. This will provide additional hints as to where that document might be. This suggestion area accommodates several documents at once, increasing the chances of displaying the target-document and helping cope with story inaccuracy. Also, since the suggestions continuously change as the story is told, they can help shape the story and keep the user on track.

One of the dialogues, the one for the Other Documents element, provides several fields where a small amount of free-form text can be entered to describe those documents and how they relate to the target document. The ability to drag documents from the suggestions bar to the central area is an alternative way to mention Other Documents in the narratives.

Also, wherever relevant the dialogues used in the interface to elicit story elements from the users already allow several different kinds of information to be mentioned at once. This addresses the problem of reflexive transitions, that become redundant. Thus, they are not explicitly suggested by either prototype (although users have the option of creating them by manually selecting the appropriate story element from the list).

4.2.2. Prototype B — “Fill In the Blanks”

This prototype tries to help the users to tell their stories by presenting those stories in text form, as they are told. Instead of just presenting dialogues to the users, incomplete natural language sentences are shown. Each of those sentences corresponds to a story element, and can be completed with the help of dialogues. After some time, the document's story begins to take shape, providing a context for the next story elements. With this approach we can evaluate if this representation for a story is more natural than separated boxes containing the different story elements.

The screen is divided into three different areas (Figure 4.2). The first are the dialogues that allow the introduction of story elements, the same used for Prototype A (1). Secondly, we find a central area where the sentences that make up the story are located (2). Finally, on the bottom of the screen there is an area where several documents are suggested to the users (3).

There are three buttons, marked 4, 5, and 6 on Figure 4.2. These allow the users to control the order in which they tell the story elements. The button marked as 4 is labelled “I can't remember”. It can be pressed by the user when he doesn't remember the story element being suggested to him by the interface. The sentence and dialogue corresponding to that element will disappear, being replaced by those of the next element in the story. Button 5 is labelled “It didn't happen”. When the interface suggests that the user describes



Figure 4.2: Prototype B: Overall View

something he knows not to have happened (sending the document to another user, for instance), he can press this button. Finally, the third button is labelled “I want another!” and allows the users to select what element they want to describe next from a list.

The sentences and respective dialogues are shown at the same time. This helps the user to understand what is being asked, and will hopefully result in a more effective interaction.

As an example, the interface could start by presenting the following sentence to the user:

This is a document I read .

At the same time, the dialogue where information about Time can be entered will be visible. If, for instance, the user specifies July 2006 as the relevant time period, the sentence would change accordingly. At the same time, the next element would be requested.

This is a document I read . Its author is .

Again, the sentence will change after the information is provided to the interface:

This is a document I read . Its author is .
I created the document for .

If the user can't remember why he created the document, he can press the “I can't remember” button. The next element will then be suggested:

This is a document I read . Its author is .
It is a document.

This process will repeat itself until the target-document is found. The user can, at any point, click on an element that is already in the story to correct it, if he changes his mind or remembers additional information.

The several sentences used in this prototype can be found in Table 4.1. When more than one sentence is present, or there are alternative words inside a sentence, this means that there were two different versions of that sentence: one for documents created by the user, and another for those of other authors, which the user just read.

4.2.3. The Dialogues

The several dialogues (which can be found in (Appendix A2) used in both prototypes were designed taking into consideration the nature of story elements discovered in Chap-

Time	I worked on this document TIME. This is a document I read TIME.
Place	I worked on/read this document at PLACE.
Co-Author	While working on the document, I had the help of CO-AUTHORS. The document had CO-AUTHORS.
Purpose	I created/read the document for PURPOSE.
Author	Its author is AUTHOR.
Subject	The document is about SUBJECT.
Other Doc.	It reminds me of OTHER DOCUMENT.
Personal	I worked on/read the document at a time when, in my PERSONAL LIFE.
World	I worked on/read the document when WORLD.
Exchanged	I received or sent this document to EXCHANGES.
Type	It is a FORMAT document.
Tasks	To create/read the document, I had to TASKS.
Storage	The document is stored in STORAGE PLACE.
Version	There are VERSIONS of the document.
Contents	I remember the document contained CONTENTS.
Events	While I was working on/reading the document, it occurred that EVENT.
Name	The document's name is NAME.

Table 4.1: Story Lengths

ter 3. They allow the introduction of the most common kinds of elements found in that study. Some dialogues allow free-form text entry. This happens for those elements that can have so many different value kinds that prevent a more structured approach (at least without some study of its own). All dialogues have a 'Help' button on the upper-right corner that, when pressed, results in an explanation of that dialogue (provided orally by the researcher).

The present study allowed several results on the expressiveness and usability of the dialogues to be reached. They will be presented in Section 4.4.

4.2.4. Use Scenarios

To better understand how the interfaces could be used to retrieve documents, we'll now present a use scenario for each prototype. The numbers in parenthesis refer to the appropriate step on the storyboards presented after each scenario.

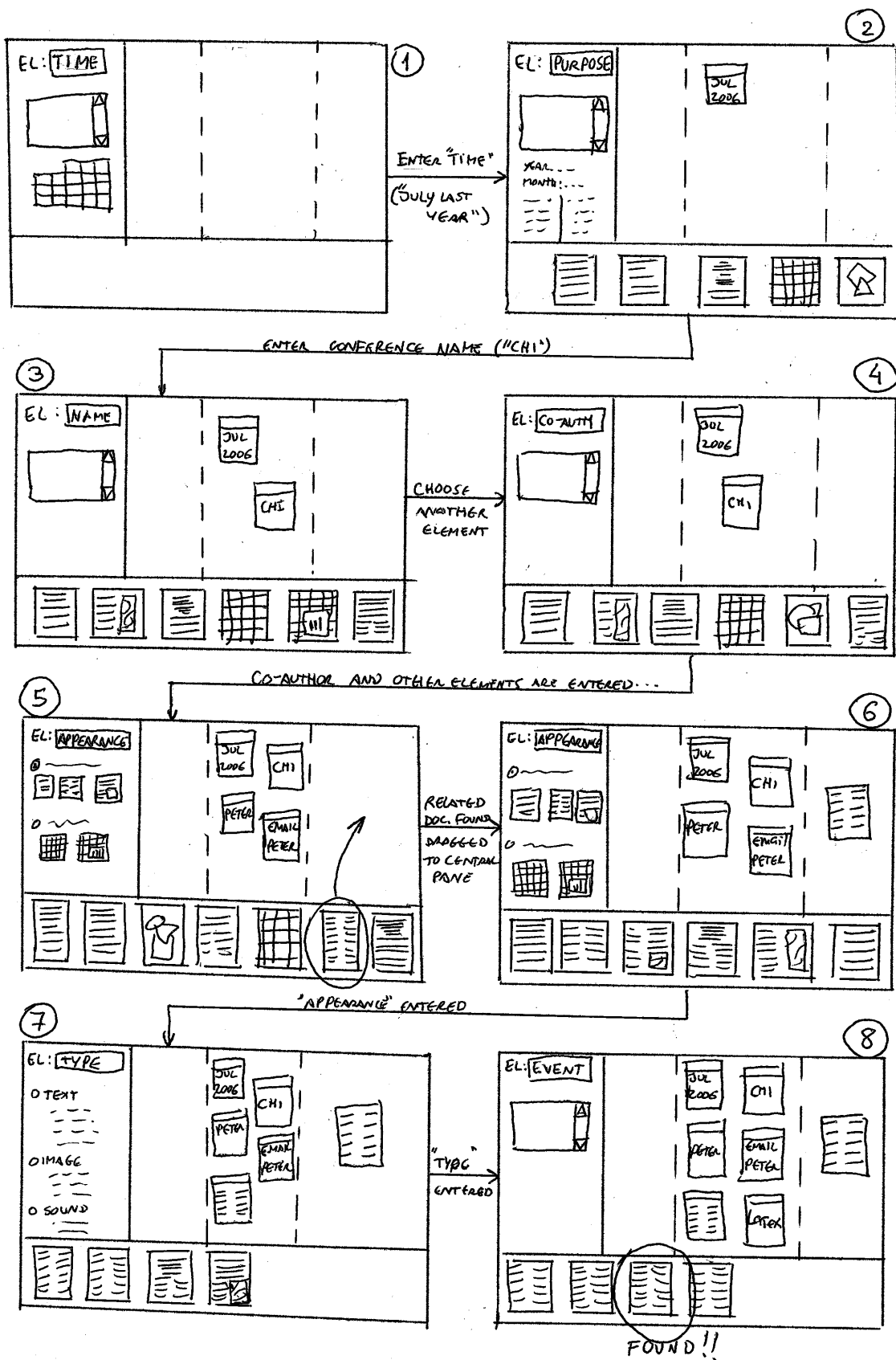


Figure 4.3: Prototype A: Storyboard

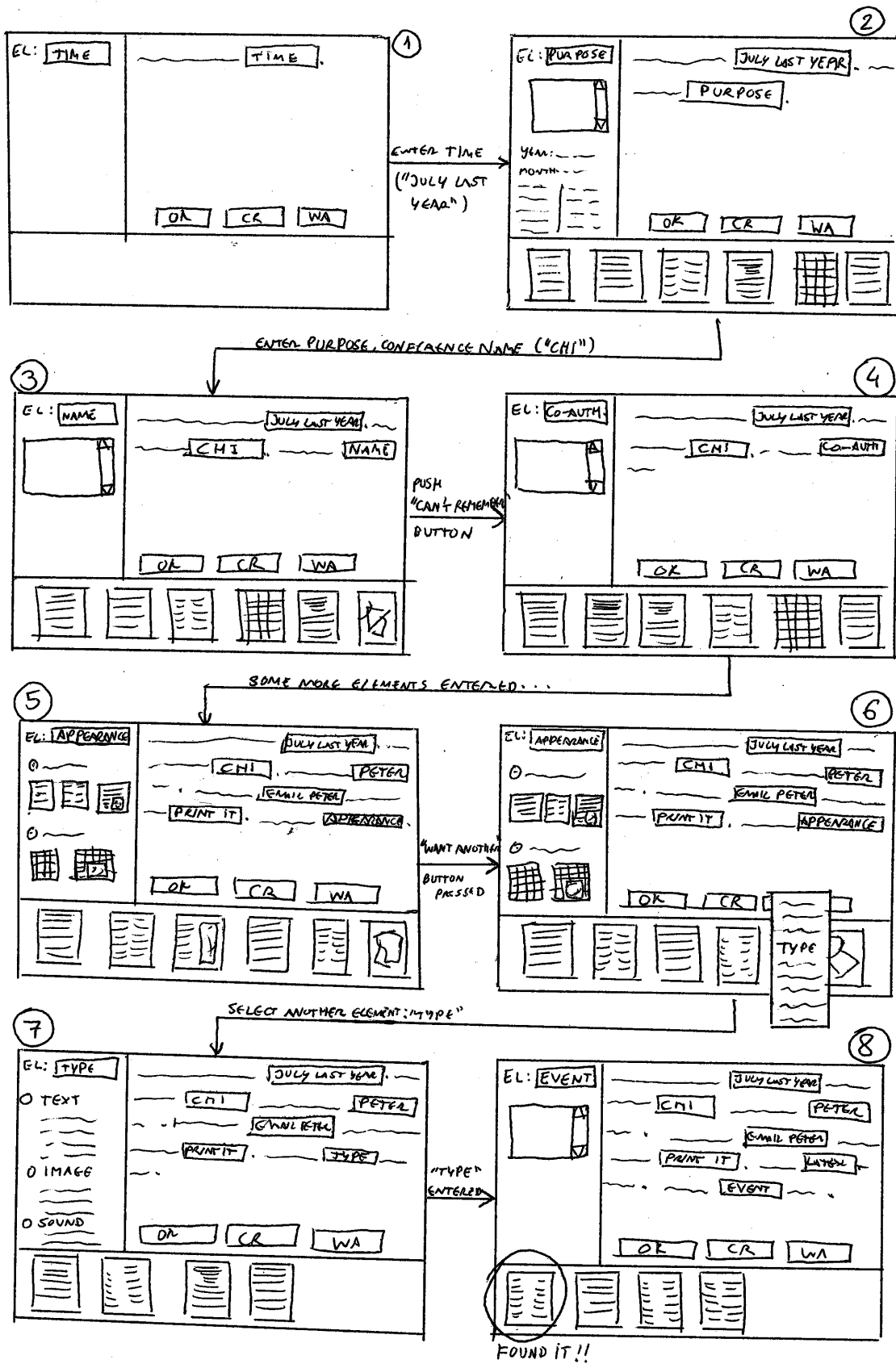


Figure 4.4: Prototype B: Storyboard

Prototype A

John is working with his computer when his cell phone rings. It is a call from Peter, his colleague and friend. He asks John for a copy of the paper they wrote together for the CHI conference some time ago. John has no idea of where the document might be located, so he decides to help his friend using the new narrative-based document retrieval system he recently installed.

The system starts by suggesting that John tells when the document was created (1). This seems reasonable, since he knows it was created “around June last year”. After he gives the system that information, it now suggests [the most likely story element to come after time, in John’s stories] that the purpose of the document should be told (2). “Now, we wrote it for the CHI conference, of course!”, thinks John, and inputs that information into the system. Meanwhile, at the bottom of the screen, some document suggestions start to appear. None seems to be the target-document. The interface now asks about the file’s name (3). John can’t remember it, so he selects a new story element he does remember: Co-Authors. He then proceeds in telling the story: “it was something I wrote with Peter’s help... I sent it to him several times by email, while we were working on it...” New boxes appear in the interface’s central area reflecting the story elements John is creating (5). Looking at the suggested documents at the bottom of the screen, John recognises one of them. “Aha!”, he thinks, “This one I wrote for the conference that came after CHI! The one I’m looking was certainly written before this one!”. He promptly drags that document to the ‘future’ area of the interface (5 and 6). The suggestions change again, but still the target-document doesn’t show up... “Well, I remember it had two columns and a chart in the first page!” (6). This is still not enough! The interface tries to extract more information from John, asking him about the document’s type. “It was a LaTeX document, of course!”(7). That information was all the system needed. One of the suggestions is now the target-document which John sends to Peter.

Prototype B

This scenario corresponds to the undertaking of the same task as in the scenario that was described for Prototype A, but using Prototype B instead.

John is working with his computer when his cell phone rings. It is a call from Peter, his colleague and friend. He asks John for a copy of the paper they wrote together for the CHI conference some time ago. John has no idea of where the document might be located, so he decides to help his friend using the new narrative-based document retrieval system he recently installed.

When John starts telling his story, the system shows him a sentence about when the doc-

ument was created (1). This seems reasonable, since he knows it was created “*around June last year*”. He gives the system that information, using the dialogue that appeared when the sentence did. The interface now shows John a sentence in which it suggests [the most likely story element to come after time, in John’s stories] that the purpose of the document should be told (2). “*We wrote it for the CHI conference, of course!*”, thinks John, and inputs that information into the system. Meanwhile, at the bottom of the screen, some document suggestions appeared. None seems to be the target-document. The interface now asks about the file’s name (3). John can’t remember it, so he presses the “I can’t remember’ button”. This causes the sentence and dialogue pertaining the document’s name to disappear, being replaced by the ones about the document’s co-authors (4). John proceeds with telling the story: “*it was something I wrote with Peter’s help... I sent it to him several times by email, while we were working on it...*” The sentences appear in the interface’s main area (5). The target-document still hasn’t shown up... “*Well, I remember it had two columns and a chart in the first page!*”. This is still not enough! The next sentence is about some event that might have happened while John was editing the document. He can’t remember any, but he does remember it was a LaTeX document! He presses the “I want another!” button, selects “Type” (5 and 6) and enters that information (7). It was all the system needed. One of the suggestions is now the target-document which John sends to Peter (8).

4.3. Procedure

The two prototypes were evaluated by different sets of users. In the evaluation, the researcher simulates the functional part of the prototype. Rather than performing a formal heuristic evaluation with the help of usability experts, the test subjects were everyday users. They were asked simply to use the system as if they were trying to recover a real document. In fact, the goal of this study is, rather than perfecting the interface, to understand if the stories told to a computer can be similar to those told to humans, and to see which prototype seems more promising to capture those stories.

Each prototype was evaluated by a set of ten users. Each user was explained the goal and procedure of the interview, and then asked to think of three actual documents and, in turn, to tell their stories using the interface presented to him. The three documents were: a **Recent Document** (created by the user no longer than two weeks ago), an **Old Document** (created by the user at least six months ago) and an **Other Document** (not created by the user). These document types were chosen to match the types used in the previous study in which stories were told, free-form, directly to the interviewer.

Two researchers were present during the course of the interviews: the **observer** and the **wizard**. The observer was responsible for taking note of which story elements were men-

tioned by the users, and also of all comments, doubts, and other interesting facts that might occur during the interviews. The wizard's job was to make the prototype react according to the users' desires.

4.3.1. Interview Script: Observer

The observer's main tool was an interview. There, he started by filling in a header section containing personal information about the user (name, age, gender, etc.). Then, while the user told the stories, the elements told in those stories are marked in a grid (one for each story), in chronological order. If the element is mentioned in answer to a suggestion made by the interface, it should be marked with a cross ('X'). If the user explicitly selected the element, it must be marked with a dot or star ('*'). This will allow an analysis of how comfortable do users perceive the element order suggested by the interface to be.

After all stories are told, the observer handed the interview form to the users in order for them to fill in a small questionnaire. Apart from a write-in section where users could provide their comments and suggestions about the interface, they are asked to rate six sentences from '1-Strongly Disagree', to '4-Strongly Agree'. An even number of rating levels was chosen to prevent the users from not committing themselves with an answer, by choosing the middle value. The sentences were:

Q1: It was easy to tell stories about documents, using this interface: With this sentence, we could estimate the simplicity of use of the interface, and, to an extent, how intrusive it is.

Q2: The retrieval of documents was satisfying: This sentence directly related to the subjective satisfaction felt by the users when using the interface.

Q3: I was able to tell everything I remembered about the documents: Allowed us to estimate the interface's completeness, regarding its ability to collect the information from the users and to elicit all relevant information.

Q4: The interface is different from document searching mechanisms I'm familiar with, because it allows me to tell stories about documents: With this sentence we were able to evaluate if the users feel they really were telling stories about their documents, or rather if they thought the interface to be just another 'find tool' such as the ones they already know.

Q5: The interface allows me to tell the stories the way I want to: The flexibility of the interface was estimated with the help of this sentence's ratings.

Q6: The interface was easy to understand and use: From the ratings of this sentence it was possible to understand how simple and natural the users found the interface.

4.3.2. Interview Script: Wizard

As stated before, the wizard was responsible for simulating the functional part of the Low Fidelity Prototypes. He reacted to the users' actions and changed the state of the low-fidelity prototype accordingly.

One of the most important aspects to take into account when conducting the evaluation was the order in which the different story elements were suggested to the users. This order was not random. Rather, it derived directly from the results of the previous study in which stories were analysed, corresponding to archetypical story structures inferred from those results. This allowed the current study to validate that order, and to, hopefully, present the several elements in an order that made telling a story as natural as possible. The only exception was related to the Author and Time elements. In fact, it was shown on the previous study that the structure of stories can vary according to those two elements. Hence, they were the first two to be suggested to the user, in that order. Afterwards, the elements were suggested as follows:

Recent/Old: Purpose, Place, Co-Authors, Subject, Other Documents, Exchanges, Type, Contents, Storage, Tasks, Version, Name, World, Personal, Events

Other: Purpose, Exchange, Tasks, Type, Co-Authors, Other Documents, Subject, Storage, Contents, Personal, Place, World, Version, Events, Name

Another difference from the archetypical structure inferred in the previous study that was taken into account was the lack of reflexive transitions, as discussed before (Section 3.2.4).

Apart from what was just mentioned, there were some guidelines the wizard followed which were specific for each of the two prototypes. They will now be described.

Regarding how long stories must be, the wizard tried to extract as much information from the users as possible. Thus, it only presented them the target documents when all possible story elements had been suggested. Also, it was tried to be keep those suggestions as coherent as possible with the story the user is telling. For instance, if a user mentioned that the target document has two columns, it would have made no sense to present to him document suggestions with just one.

Prototype A

Every time a box containing a story element was placed on the interface's central area, a small paper in which that element is written (albeit in an abbreviated way) was placed inside it, using the paper flap provided for that end.

Prototype B

The wizard took special care to present only the sentences which corresponded to the document being described by the user. For instance, if a document created by the user was being described, it was sensible to use the sentence “I worked on the document at ...” rather than “I read that document at...”

Also, although on a functional interface the sentences should change after the user tells the respective story element, in this study the original sentence was left on the interface, for practical reasons.

4.4. Results

In this section, we’ll present all the results the data allowed us to reach. There are two main result areas: the comparison of the prototypes and the analysis of the story structures, when compared with the stories told to a human listener.

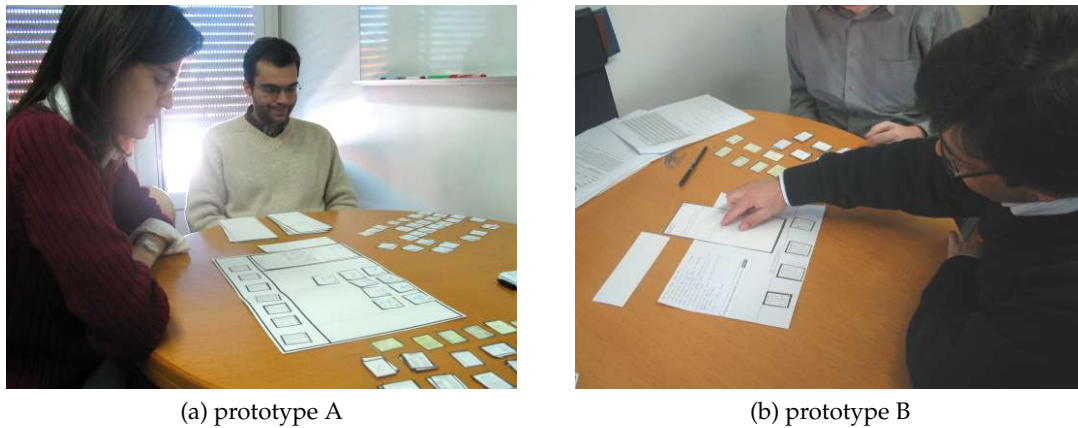
We’ll start by presenting a description of the participants in the study. Then, we’ll look at the stories, their structure and contents, and how they differ from what was previously found. The prototype that allows users to tell stories that are more similar to those told to humans is better, as it interferes less in the storytelling process. Next we’ll discuss the prototypes and their relative merit in capturing the users’ stories. Finally, we’ll take a look at the comments made by users while evaluating the prototypes, to gain further insight of the storytelling process using the computer.

4.4.1. Subject Characterisation

We interviewed twenty persons in total, ten of which evaluated Prototype A and the remaining ten Prototype B. Overall, thirteen users (65%) were males, and seven (35%) females. Their ages ranged from 22 to 56 years old (Figure 4.6), and their professions from Translator and Economist to Computer Science Student. For the most part, they had College degrees or were college students (Table 4.2).

Although the age distribution isn’t uniform (difficult for a random sample of small size), this was not considered a problem, since in the previous story-analysing study few age-related differences were found.

Looking at the users for each prototype we find that, for Prototype A, three of them (30%) were females and seven (70%) males. For Prototype B, we have four females (40%) and six males (60%). The users that evaluated Prototype B were, on average, younger than



(a) prototype A

(b) prototype B

Figure 4.5: Using the Prototypes

those that evaluated Prototype A. Again, this is not problematic, for the aforementioned reasons.

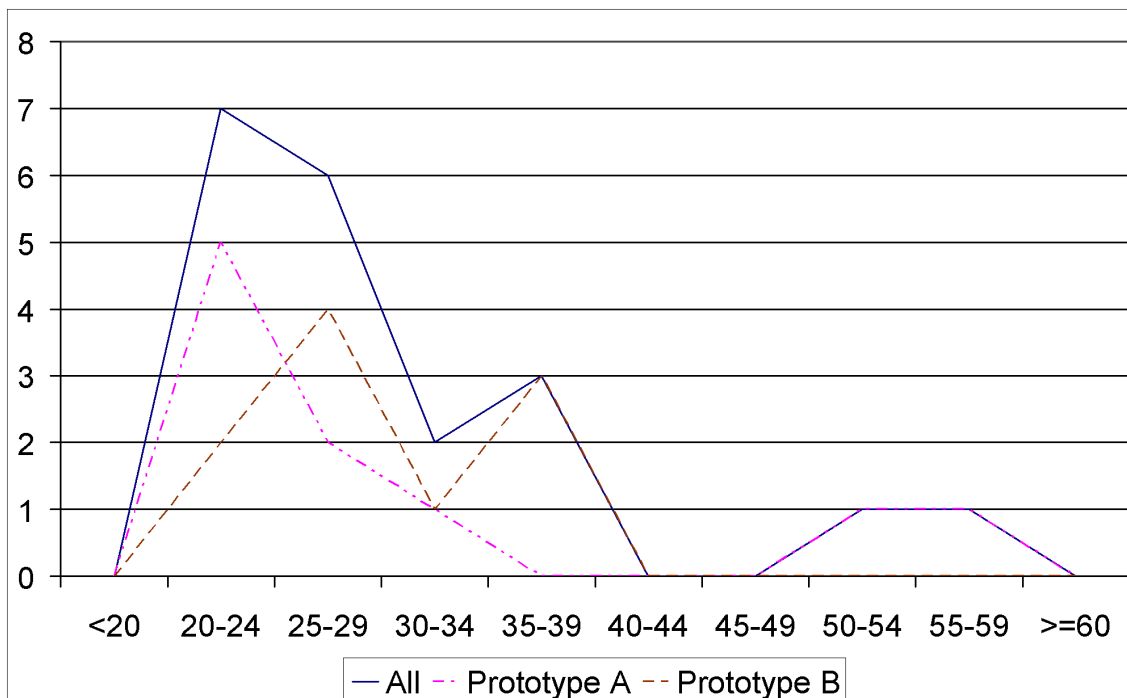


Figure 4.6: Subject Age

4.4.2. Stories

One of the main research questions to which we wanted to provide an answer with this study was if stories told to a computer, on a structured environment, are similar to those told without restrictions, to a human listener. Furthermore, the interface that produces more similar stories is undoubtedly the best of the two. Two different aspects of stories had to be verified to reach that answer: structure and contents.

Prototype	Gender	Age	Academic Background	Profession
A	M	25	Comp. Sci. College Student	Student
A	M	25	Comp. Sci. College Student	Student
A	M	23	Comp. Sci. College Student	Student
A	M	23	Comp. Sci. College Student	Student
A	F	22	Comp. Sci. College Student	Student
A	F	31	Comp. Sci. College Student	Student
A	M	56	Education Sciences PhD	College Professor
A	M	51	Educational Supervision MSc	College Professor
A	M	24	Comp. Sci. College Student	Student
A	F	24	Bioengineering PhD Student	PhD Student
B	M	35	Comp. Sci. PhD Student	Teaching Assistant
B	M	24	Comp. Sci. College Student	Student
B	M	31	Comp. Sci. College Student	Comp. Sci. Engineer
B	M	24	Comp. Sci. MSc Student	Researcher
B	F	38	College Degree	Translator
B	F	35	Statistics and Management MSc	Information Higher Technician
B	M	29	Social Anthropology Degree	Operations Manager
B	F	27	Economics Degree	Economist
B	M	28	Comp. Sci. College Student	Comp. Sci. Engineer
B	F	29	Comp. Sci. College Student	Information Higher Technician

Table 4.2: Interviewed Subjects' Background, Age and Profession

Story Structure

We can evaluate story structure in two ways. First, we'll take a look at the length of stories. Secondly, the story elements were suggested to the users in an order inferred from stories told to humans. Deviations from that order will give us an estimate of its adequacy and naturality in this new context (although few strong or highly probable element transitions were found in the previous study).

Story Length

With regard to story length, we found the values in Table 4.3, depicted in Figure 4.7. One thing is evident: Prototype B seems to be able to elicit longer stories from the users. To see for which prototype the stories better approach those told in unstructured environments, compared the length values with the ones obtained in the previous study. However, the values to be compared are those pertaining to the occurrence of story elements, rather than their frequency. For instance, if Time was referred twice in a story, it would have a frequency of two and an occurrence of one. Occurrence must be used because reflexive transitions were eliminated from our prototypes, as the dialogues for element entry allowed several bits of information to be told in just one interaction. Thus, the values for story length, in this study, should be close to the occurrence values of the previous one.

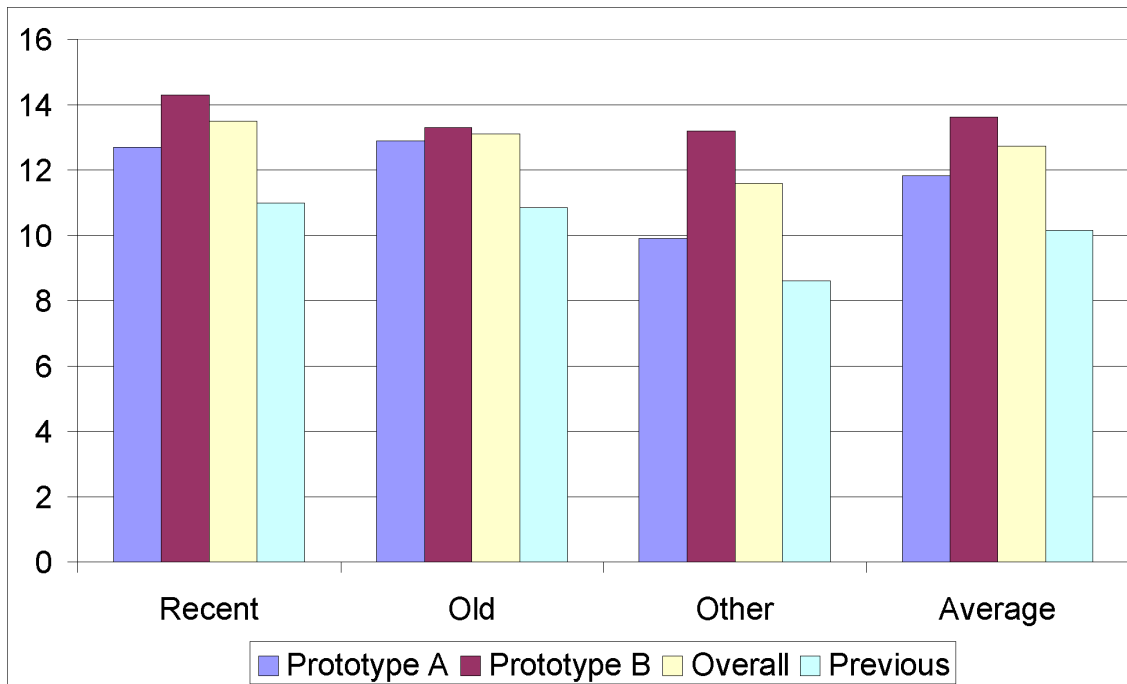


Figure 4.7: Story Length Comparison

	Recent	Old	Other	Overall	Overall StDev
Prototype A	12.7	12.9	9.9	11.83	1.68
Prototype B	14.3	13.3	13.2	13.63	0.58
Overall (A&B)	13.5	13.1	11.6	12.73	1
Previous Study	11	10.85	8.6	10.15	2.1
Number of stories for each document type in each prototype					10
Number of stories for each document type in the previous study					20

Table 4.3: Story Length Averages from Element Occurrence

As can readily be seen, the values found in this study are a bit larger than those previously found. It would seem, then, that a more structured environment helps users to tell longer stories! In fact, t-tests with a 95% confidence level confirm there is a statistically significant difference between story lengths in both studies for all cases except that of stories about Other Documents told using Prototype A.

There appears to be some further relation between the values in both studies. To verify this hypothesis, the ratios between the current and previous values were calculated (Table 4.4). The prototype that seems to present a more uniform relation with the data from the previous study is Prototype A, in which stories are 15%-19% longer than previously found. That range rises to 23%-53% for Prototype B. Overall, the ratios are similar for Recent and Old documents, and a bit larger, but not too much for those of Other authors. In short, stories told to computers are around 20% longer for documents written by the users themselves, and 35% longer for those of other authors.

	Recent	Old	Other	Overall
Current A / Previous	1.15	1.19	1.15	1.17
Current B / Previous	1.3	1.23	1.53	1.34
Current A&B / Previous	1.23	1.21	1.35	1.25

Table 4.4: Story Length Ratios

We feel that the reason for this length difference, from the users' reactions, might be the fact that the dialogues and sentences presented to the users by the interface help jog their memories and cause them to remember information they wouldn't otherwise. This is consistent with the fact that it is for Other Documents that the difference is greater: those were the documents for which users had more trouble remembering data. Also, Prototype B provides longer stories than Prototype A (a t-test with 95% confidence confirms this for the overall case). This confirms that, by presenting the story in a textual, sequential form, displaying all elements in context, that prototype helps the users to remember more information about their documents. Also, the fact that it shows the users more information about possible story elements (the sentences that make up the story) might also be a factor in helping them remember relevant information.

Element Order

When using the prototypes, the users had the choice of not accepting the story element proposed to them at a given time, and either let the interface choose another one, or select one from a list. The times in which they opted to choose a new element reveal that the order being suggested was not the one considered more natural by the users. Since this order directly reflects the findings about story structure from the previous study, the number of times it was broken is an estimate of eventual differences between preferred element orders in both studies

We didn't consider the times users just let the interface choose a new element as disruptive of story structure. In fact, that often occurred as a result of our attempt to extract as much information as possible from the users, and, consequently, asking them about all possible story elements, even those we knew beforehand the users would have trouble remembering. From the previous study, we know that typical stories are shorter than those we tried to get the users to tell. Rather, when asked about all elements, in the previous study, it was common for users not to remember some of them. The elements that the users most often didn't remember now are the same they had trouble remembering before, since they continue to be the most infrequent, as we'll see below. For all those reasons, we feel that not taking into account, in this section, unremembered elements better mimics the previous study and facilitates its comparison with the current results. Even if we did consider those elements as order disruptive, they would not invalidate our con-

clusions, since they occurred, on average, around 3.4 times per story for each document class and for both prototypes, showing that the prototype in use doesn't make a difference, and accounting only for 20% of the total element order (there are 17 elements).

The relevant values for this section can be found in Table 4.5. There, we can see the average number of times the users chose a new element from the list. While, at first sight, it would seem that, overall, users frequently chose new elements (30% of users in 23% of stories), appearances can be deceiving. As can be readily verified, there is a large difference between the behaviours of users using different prototypes. While those using Prototype A often chose a new element (half the users did it, on average, 1.27 times per story on 43% of stories), those evaluating Prototype B seldom did so (only 10% of users, in 3% of stories, with an average of 0.07 times per story). It would seem that, in establishing an order preference, using a computer or not is not as important as the interface in use. A good interface can act in a way in which users don't feel differently than they would when telling stories in an unstructured environment, with respect to element order.

It is evident, then, that the suggested order seemed perfectly natural for users of Prototype B. For those using Prototype A, the 1.27 times per story a new element was selected from the list amount only to 7.4 percent of all possible elements. Hence, we can conclude that the order discovered in the previous study is, for the most part, respected in the current one, and felt natural to users when telling their stories.

	Recent	Old	Other	Overall	Ov. Stdev	%Stories	%Users
Prototype A	1.4	1.3	1.1	1.27	0.15	43%	50%
Prototype B	0	0	0.2	0.07	0.12	3%	10%
A + B	0.7	0.65	0.65	0.67	0.03	23%	30%

Table 4.5: Explicit Order Changes

Story Contents

In this section, we'll study the frequencies of the several story elements found in the stories, comparing them with those found in the previous study, in which stories were told to human listeners. We'll also try to understand if the relative importance of story elements has changed from one study to the other. This will allow us to discover to what extent the information told to a computer differs from those, unrestricted, stories.

Element Frequencies

Comparing the values found in this study with those of element occurrence in the previous one, we find that most frequencies are somewhat similar, but some differences do

occur. This can be seen in Table 4.6 and, more easily, in Figure 4.8, that display the percentage of stories in which the different elements were found.

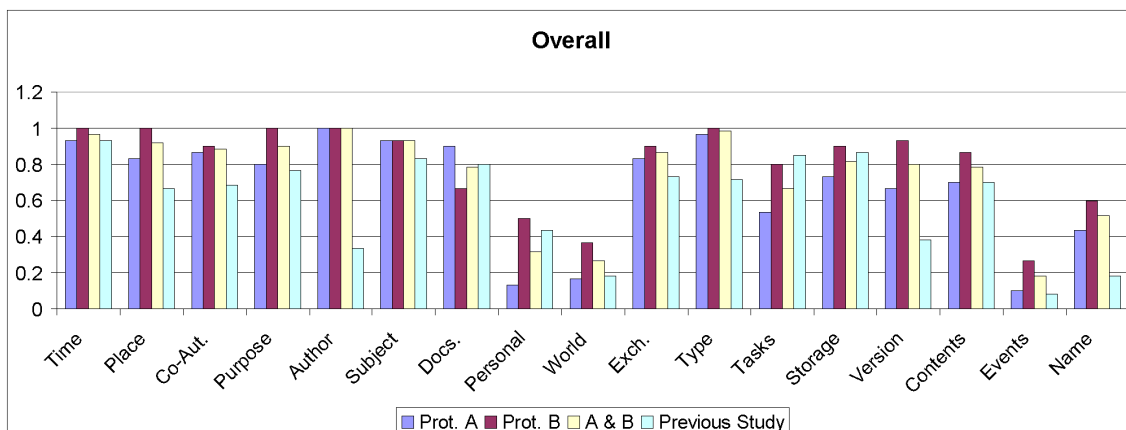


Figure 4.8: Element Frequencies Comparison Chart (Overall)

	Recent				Old				Other				Overall			
	PrA	PrB	A&B	Prev	PrA	PrB	A&B	Prev	PrA	PrB	A&B	Prev	PrA	PrB	A&B	Prev
Time	0.9	1	0.95	0.95	1	1	1	0.95	0.9	1	0.95	0.9	0.93	1	0.97	0.93
Place	1	1	1	0.95	0.9	1	0.95	0.8	0.6	1	0.8	0.25	0.83	1	0.92	0.67
Co-Aut.	0.9	1	0.95	0.85	1.1	1	1.05	1	0.6	0.7	0.65	0.2	0.87	0.9	0.88	0.68
Purpose	0.9	1	0.95	0.85	0.9	1	0.95	0.95	0.6	1	0.8	0.5	0.8	1	0.9	0.77
Author	1	1	1	0.1	1	1	1	0.2	1	1	1	0.7	1	1	1	0.33
Subject	1	0.9	0.95	0.85	0.8	1	0.9	0.75	1	0.9	0.95	0.9	0.93	0.93	0.93	0.83
Docs.	1	0.7	0.85	0.75	1.1	0.6	0.85	0.85	0.6	0.7	0.65	0.8	0.9	0.67	0.78	0.8
Personal	0.1	0.5	0.3	0.4	0.1	0.5	0.3	0.45	0.2	0.5	0.35	0.45	0.13	0.5	0.32	0.43
World	0.2	0.5	0.35	0.25	0.2	0.4	0.3	0.2	0.1	0.2	0.15	0.1	0.17	0.37	0.27	0.18
Exch.	0.7	0.9	0.8	0.7	1	0.8	0.9	0.7	0.8	1	0.9	0.8	0.83	0.9	0.87	0.73
Type	1	1	1	0.8	1	1	1	0.65	0.9	1	0.95	0.7	0.97	1	0.98	0.72
Tasks	0.5	0.8	0.65	0.95	0.5	0.8	0.65	0.9	0.6	0.8	0.7	0.7	0.53	0.8	0.67	0.85
Storage	0.9	1	0.95	0.95	0.7	0.9	0.8	0.8	0.6	0.8	0.7	0.85	0.73	0.9	0.82	0.87
Version	0.6	1	0.8	0.6	1	0.9	0.95	0.55	0.4	0.9	0.65	0	0.67	0.93	0.8	0.38
Contents	0.8	0.9	0.85	0.75	0.7	0.8	0.75	0.75	0.6	0.9	0.75	0.6	0.7	0.87	0.78	0.7
Events	0.2	0.4	0.3	0.05	0	0.1	0.05	0.2	0.1	0.3	0.2	0	0.1	0.27	0.18	0.08
Name	0.6	0.7	0.65	0.25	0.6	0.5	0.55	0.15	0.1	0.6	0.35	0.15	0.43	0.6	0.52	0.18

Table 4.6: Comparison Table of Element Frequencies (% of stories)

To discover which differences in element frequencies are statistically significant, we performed several binomial tests (two-tailed), in which the observed element frequencies for the different prototypes and stories was compared to a set of expected frequencies: the values found for the occurrence of story elements in the previous study. The binomial test allows us to verify if the differences between those values (observed and expected) are statistically significant. Although a Chi-squared test is more often used to compare observed and expected frequencies, it could not be applied in this case because some of its pre-conditions were not met (notably, more than 20% of the cells had frequencies less than 5, for most story elements). The binomial test suffers no such restrictions, and is applicable without problems when the variables are dichotomous, as is the case. We chose

a confidence level of 99% for our tests. Hence, if the significance value returned by the test is less than 0.01, we must reject the null hypothesis and conclude there are significant differences between the values we already had and those now observed. We use a stricter confidence value than that used elsewhere in this document (usually, 95%) since a large number of tests was performed (204). A significance level of 95% might cause around ten such tests to be erroneous in their conclusion. With a significance level of 99% only two or three will, thus limiting the number of eventually spurious conclusions.

The results of the several binomial tests can be found in Table 4.7, where the significance values that allow us to reject the null hypothesis (indicating there are significant value differences) are highlighted.

	Prototype A				Prototype B				Prototypes A+B			
	Rec.	Old	Other	All	Rec.	Old	Other	All	Rec.	Old	Other	All
Time	0.53	0.72	1.22	1.15	0.72	0.72	0.24	0.03	1.28	0.72	0.78	0.46
Place	0.72	0.41	0	0.01	0.72	0.02	0	0	0.72	0.14	0	0
Co-Aut.	0.81	2	0	0	0.08	2	0	0	0.35	2	0	0
Purpose	0.81	0.53	0.5	0.66	0.08	0.72	0	0	0.35	1.28	0.01	0.01
Author	0	0	0	0	0	0	0	0	0	0	0	0
Subject	0.08	0.83	0.24	0.04	0.81	0.01	1.22	0.04	0.35	0.18	0.78	0.04
Docs.	0.01	0.08	0.06	0.06	0.77	0.01	0.39	0.02	0.45	1.19	0.17	0.85
Personal	0.01	0	0.04	0	0.49	0.82	0.82	0.36	0.5	0.26	0.5	0.09
World	0.83	1.26	1.35	0.9	0.03	0.06	0.27	0	0.43	0.39	0.65	0.14
Exch.	1.17	0	1.18	0.1	0.07	0.48	0.02	0	0.48	0.07	0.41	0.02
Type	0.02	0	0.07	0	0.02	0	0	0	0.02	0	0.02	0
Tasks	0	0	0.46	0	0.03	0.27	0.48	0.36	0	0	1.17	0
Storage	0.53	0.39	0.01	0.01	0.72	0.41	0.7	0.59	1.28	1.18	0.13	0.34
Version	1.17	0	0	0	0	0	0	0	0.1	0	0	0
Contents	0.83	0.77	1.17	1.1	0.18	0.83	0.01	0	0.45	1.17	0.25	0.2
Events	0.03	0.02	0	0.77	0	0.41	0	0	0	0.14	0	0.02
Name	0	0	0.81	0	0	0	0	0	0	0	0.04	0

entries with '0' mean the actual value was < 0.01)

Table 4.7: Binomial test significance values comparing values in both studies

From the table, it is evident that there are several elements for which statistically significant differences occur (36.2% of all performed tests). Taking a closer look at the data (Table 4.6), we verify that, with a few exceptions, the differences reflect the fact that the relevant story elements appear more frequently in the present study, showing that slight increase in frequency to be significant. The only cases in which there are significant differences but where the elements are actually *less* frequent in stories told using the prototypes are:

- In Prototype A, Personal Life Events, Tasks, and Events (Old Documents)
- Overall, for Tasks.

The most atypical story element seems to be Tasks. Users mention them less often when telling stories to a computer than they do when telling those stories to a human. Proto-

type B seems to more easily allow users to mention Personal Life information and document Storage information than Prototype A. We'll discuss these differences in greater detail ahead.

It is important to mention, at this point, that the differences found regarding the Author element were induced by the prototypes themselves: the first two story elements were always Author and Time. That is the reason why Author was mentioned in all stories, unlike what occurred in the previous study, in which Author information was often taken for granted when describing documents the users themselves had written.

Relative Importance

Finally, it was verified that the relative order of the different story elements, according to their frequency, had changed from the previous study. This is hardly surprising, since we're dealing with fairly low story numbers (only 20 of each type) and many elements (especially the most infrequent) had occurrence values very close to each other, which, again with the small numbers involved, made the ordering very sensitive to small changes.

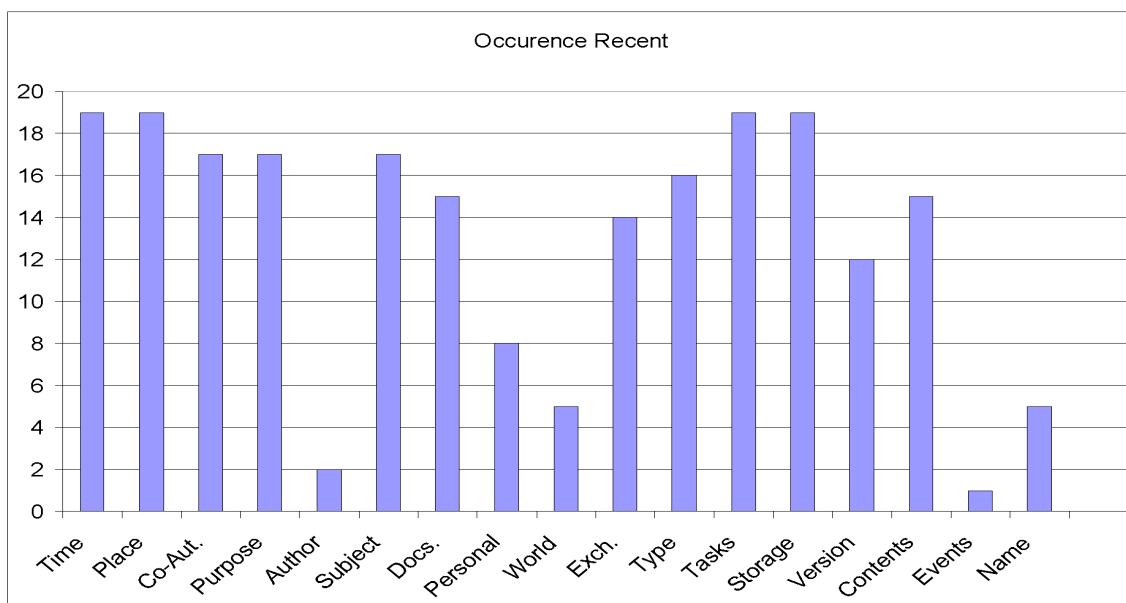


Figure 4.9: Story Element Occurrence Frequencies in the Previous Study

Since the order was very sensitive, the fact that two elements don't maintain their relative positions might not be relevant, since their actual importance in the stories might remain unchanged. To try to ascertain to what extent the element's relative importance had changed, we took advantage of a result of the previous study which recognises that there are, in practice, two story elements classes: those that are very common, and those that are seldom mentioned. In fact, for all document classes there is, at some point, a gap in the

frequency values of the elements. Figure 4.9, resulting from the previous study illustrates this point for Recent documents: most elements are fairly common, with frequencies up to 19, but there is a 20% gap between Version (12) and the next element, Personal Life Events (8), in which no story element falls. We divided the elements for all story classes using this criterion (looking for a gap of 20% or more), and, after ordering the elements from the data obtained in this study, we verified if the groups maintained their ordering. The results can be found in Table 4.8 and Table 4.9, in ascending frequency order. Highlighted elements are those in the “unimportant” category. Information about Author is not present since, as discussed, it was induced to appear in all stories as its first element.

As can be seen, for the most part the groups maintain their relative order, with a few notable exceptions. For Recent Documents, Name seems to gain some importance, as it does for Old Documents. In those documents, it would seem that Personal Life Information became even more infrequent. For Other Documents, we find more radical changes. Those changes all stem from the fact that several elements occur more frequently. This was to be expected, as we have already seen that stories are longer. So, several elements gain importance and move from the “unimportant” to the “important” class. No element lost importance, though. For the most part element importance remains unaltered.

Recent				Old			
PrA	PrB	A&B	Prev	PrA	PrB	A&B	Prev
Personal	Events	Events	Events	Events	Events	Events	Name
Events	Personal	Personal	Name	Personal	World	Personal	Events
World	World	World	World	World	Name	World	World
Tasks	Docs.	Name	Personal	Tasks	Personal	Name	Personal
Name	Name	Tasks	Version	Name	Docs.	Tasks	Version
Version	Tasks	Exch.	Exch.	Contents	Contents	Contents	Type
Exch.	Contents	Version	Contents	Storage	Exch.	Storage	Exch.
Contents	Exch.	Contents	Docs.	Subject	Tasks	Docs.	Contents
Co-Aut.	Subject	Docs.	Type	Place	Storage	Exch.	Subject
Purpose	Co-Aut.	Co-Aut.	Co-Aut.	Purpose	Version	Subject	Place
Storage	Place	Purpose	Purpose	Exch.	Co-Aut.	Place	Storage
Time	Purpose	Storage	Subject	Time	Place	Purpose	Docs.
Docs.	Storage	Subject	Place	Type	Purpose	Version	Tasks
Place	Time	Time	Storage	Version	Subject	Time	Purpose
Subject	Type	Place	Tasks	Co-Aut.	Time	Type	Time
Type	Version	Type	Time	Docs.	Type	Co-Aut.	Co-Aut.

Table 4.8: Element Ordering: Recent and Old Documents

Other				Overall			
PrA	PrB	A&B	Prev	PrA	PrB	A&B	Prev.
Events	World	World	Events	Events	Events	Events	Events
Name	Events	Events	Version	Personal	World	World	Name
World	Personal	Name	World	World	Personal	Personal	World
Personal	Name	Personal	Name	Name	Name	Name	Version
Version	Co-Aut.	Co-Aut.	Co-Aut.	Tasks	Docs.	Tasks	Personal
Co-Aut.	Docs.	Docs.	Place	Version	Tasks	Contents	Place
Contents	Storage	Version	Personal	Contents	Contents	Docs.	Co-Aut.
Docs.	Tasks	Storage	Purpose	Storage	Co-Aut.	Version	Contents
Place	Contents	Tasks	Contents	Purpose	Exch.	Storage	Type
Purpose	Subject	Contents	Tasks	Exch.	Storage	Exch.	Exch.
Storage	Version	Place	Type	Place	Subject	Co-Aut.	Purpose
Tasks	Exch.	Purpose	Docs.	Co-Aut.	Version	Purpose	Docs.
Exch.	Place	Exch.	Exch.	Docs.	Place	Place	Subject
Time	Purpose	Subject	Storage	Subject	Purpose	Subject	Tasks
Type	Time	Time	Subject	Time	Time	Time	Storage
Subject	Type	Type	Time	Type	Type	Type	Time

Table 4.9: Element Ordering: Other Documents and Overall

4.4.3. Comparing the Prototypes

So far, we've seen what differences in story contents and structure could be found between stories told using the two prototypes. We'll now discuss the users' reactions to them.

Questionnaire

All users were asked to fill in a small questionnaire after using the prototypes. In them, they had to rate six sentences with a value ranging from "1 – Completely Disagree" to "4 – Completely agree" (further details are discussed in section 4.3.1 of this document). Those sentences directly related to different aspects of the user experience. Table 4.10 summarizes the results collected from the questionnaires. Regretfully, given that only ten users per prototype were involved, a set of pre-conditions for the application of statistical tests must be verified and, in this case, they were not met. However, some values allow a direct comparison, which we'll now present.

Overall, the users liked both prototypes, giving them good scores on all criteria, reinforcing our assumption that using stories to find documents is a good approach. Taking a look at the individual prototypes, it would seem that Prototype B is the preferred one since it got better scores for all criteria except Satisfaction (but even for that sentence the average value is close to the one attributed to Prototype A).

	Average		Mode		StDev.	
	Prot. A	Prot. B	Prot. A	Prot. B	Prot. A	Prot. B
S1 - Simplicity	3.3	3.7	3	4	0.48	0.48
S2 - Satisfaction	3.4	3.3	4	3	0.7	0.48
S3 - Completeness	2.9	3.2	3	4	0.57	1.03
S4 - Novelty	3.6	3.9	4	4	0.52	0.32
S5 - Flexibility	3.2	3.2	3	3	0.63	0.63
S6 - Understandability	3.3	3.7	4	4	0.82	0.48

Table 4.10: Questionnaire Results

It is clear that Prototype B is considered more simple and easy to use than Prototype A (3.7 vs. 3.3). Not only is the average value greater, but so is the mode. Having to deal with a set of apparently unrelated story elements in graphic form might have been the decisive factor contributing to this result, confusing users. The text-based solution of Prototype B is more natural and, thus, simpler.

The users' subjective satisfaction was slightly greater for Prototype A (3.3 vs. 3.4). However, taking a look at the value of the standard deviation, we find that its value for Prototype A is fairly large, showing more variability in the results than for Prototype B. This, allied to the very close average values, does not allow us to conclude there is any significant difference between the prototypes (indeed, a 95% confidence interval for the difference of the means of $[-0.43, 0.63]$ shows that difference can be zero).

As to completeness (the ability to capture all the information the users remembered about their documents), again Prototype B seems to win (3.2 vs. 2.9). It possesses a rather high standard deviation, though, which compromises this conclusion. Again, a confidence interval for the difference of the means of $[-1.03, 0.43]$ proves the difference can be zero. So, both prototypes allow, in fact, the same kind of information to be told. This is not unexpected, since both use the same dialogues to collect the story elements.

Regarding novelty, users manifested a clear preference for Prototype B (3.9 vs. 3.6). Prototype A looks more like wizard-based solutions they are used to in other applications. Prototype B, on the other hand, is "unlike anything they have ever seen", as a user told when filling in the questionnaire. That prototype helps the users to feel they are telling a story, rather than just filling in property values. This is important for the success of our document-retrieving approach.

There isn't much to say about flexibility: both prototypes seem equally flexible in the eyes of the users. Again, this might be directly related to the fact that the same dialogues were used in both prototypes.

Finally, Prototype B seems easier to understand than Prototype A (3.7 vs. 3.3). This leads us to conclude that presenting the story using natural language sentences is less

demanding, cognitively-wise, than separating the story into its elements on a graphic way.

The Prototypes in Use

We'll now discuss in which ways the different functionalities provided by the prototypes were used. This will allow us to find out which are more relevant and should be improved and reused in future prototypes, and which are not important.

In both prototypes users could, at any point in a story, choose a different story element from a list, just move on when they didn't remember an element allowing the prototype to choose the next one, and correct the information already present in the story. In Prototype A the users could, additionally, drag documents from the suggestion area to the central part of the interface, and move story elements to the past or future areas of the interface. The numbers of times each of those capabilities was used can be found in Table 4.11.

		Recent	Old	Other	Avg	Stdev	%Stories	%Users
Prototype A	Chose	1.4	1.3	1.1	1.27	0.15	43.33	50
	Moved On	2.44	2.8	5.2	3.48	1.5	76.67	80
	Correct Element	0	0.1	0	0.03	0.06	3.33	10
	Drag Document	0.6	0.4	0.2	0.4	0.2	23.33	50
	Drag Past & Future	0.27	0.18	0.18	0.21	0.05	23.33	40
Prototype B	Chose	0	0	0.2	0.07	0.12	3.33	10
	Moved On	2.7	3.5	3.44	3.21	0.45	93.33	100
	Correct Element	0	0	0	0	0	0	0
A + B	Chose	0.7	0.65	0.65	0.67	0.03	23.33	30
	Moved On	2.57	3.15	4.32	3.35	0.89	85	90
	Correct	0	0	0	0	0	0	0

Table 4.11: Prototype Functionalities Usage

As was already discussed in Section 4.4.2 , it was common for users of Prototype A to choose new elements from a list (1.27 times per story), and users of Prototype B seldom did so (0.07 times per story). This reflects the fact that the interface of Prototype A is more confusing, and leads the users to feel the need to choose another element due to some misunderstanding as to what a given element might be, or how it fits into the story. Prototype B has the upper edge, in this aspect. This explains why stories captured by Prototype B are, on average, longer than those captured by Prototype A: the users of Prototype A sometimes refused to provide the element being suggested by the interface and chose another instead.

As to the number of times in which users moved on and requested a different element suggestion, we find average values of 3.48 and 3.21 times per story for Prototype A and Prototype B, respectively. It would seem that only the user's memory about certain elements is relevant for the usage of this functionality, and not the prototype in use. Overall,

we find that users forget, on average, 3.35 story elements per story, which is consistent with the registered story lengths and, as already stated when discussing them, indicates that a structured environment might help users remember more information about their documents.

Still comparing both prototypes, we verify that, for either one of them, virtually no one used the correction functionality provided to correct element values already entered in the story. It seems users have great confidence in what they remember. In a real system, it is conceivable that, when not finding a document after stating all they can remember, the users might activate this feature.

Relating only to Prototype A, dragging documents from the suggestion area to the central area of the interface, establishing a relation between them and the target-document is used, but not very often. In fact, only half the users, for 23% of stories, did so. This doesn't even amount to one dragging per story, on average, the real value being 0.4 times. However, in the prototype, the only thing that might be recognised as related to the target-document is a document's appearance. Again, in a real application, perhaps other factors could lead to further use of this capability.

Finally, and still for Prototype A, moving story elements to the areas of the interface representing the past and future was not very common either. Only 40% of users ever did so, for 23% of stories. Unlike dragging documents from the suggestions area, moving elements to the past or future is something that could be fully taken advantage of in the prototype. Hence, it seems likely that it is not something that would be very advantageous in a real system. A user even took the time to explicitly mention that the different temporal areas of the interface might be confusing, since some story elements are atemporal and it makes no sense to drag them around. Some users did resort to it, though, showing that some consider it a desirable feature. It might be worthwhile to implement an improved, less confusing, version of it.

4.4.4. User Reactions and Comments

In the course of the interviews, all comments made by users when telling the stories were noted by the observer, together with relevant insights the researcher could gather from their behaviour. The notes were classified into four kinds: interesting element values, limitations of the prototypes, confusions (misunderstanding what a given element represents, for instance), and general observations about the interface. A complete list of those notes can be found in A3.3 and A3.4, in Appendix A3. Also, most users took advantage of a write-in section on the questionnaire to comment on the interface (see A3.3 and A3.4 in Appendix A3). We'll now discuss the main trends and conclusions that those notes and comments allowed us to find.

Prototype Limitations

Most comments and notes reflect some limitations of the prototypes in some aspect of capturing the users' stories. Here, we'll discuss those related to the overall interaction with the prototypes. Problems with specific dialogues will be considered separately, below. All these are to be taken into account when developing the actual functional interface.

Other Documents

Although it was possible to tell the interfaces about related documents, in the appropriate dialogue, most users felt this was not enough. In fact, four users complained about the lack of better support for recursive stories about other documents, in Prototype A, and one did so for Prototype B. The lack of that support lead to confusion in some users that, after mentioning another document (or dragging it from the suggestion area to the central area in Prototype A), assumed that the next story element should be about that document, and not the target-document. This occurred once for Prototype A and twice for Prototype B.

The fact that documents could be dragged from the suggestions area in Prototype A might account for the larger number of complaining users of that prototype. In fact, User 1 complained about the insufficiency of just dragging a document, and the need of a way to convey more information about it. User 2 commented that *"A chronological relation between documents is useful in telling the story, but sometimes a small story about other documents would also be of use, instead of having to point at actual documents"*.

Events

It was widely mentioned that the several Personal Life Events, World Events, and Events story elements are confusing and even useless. In fact, we verified that at least once for each prototype the several event types had to be thoroughly explained to the user before he could proceed with telling the story. Even so, some confusion kept showing up. For instance, User 13 confused Personal Life with Events, and User 7 did it with Time.

Those elements were often hard to remember (as discussed in Section 4.4.2), annoying some users. They were considered such nuisances that some users took the time to explicitly mention it in their comments. User 9 stated that *"The dialogues about Personal Life are completely useless."*, and User 10 that *"Information about Events, World Events and Personal Life are unnecessary and useless"*. Likewise, User 15 complained that *"The World Events part seems to be superfluous"* and, less radically, User 19 mentioned that *"The question about*

Events is not very useful when describing older documents”.

Although this aspect of the interfaces got most of the negative comments, some users did like it, as can be verified from the frequencies of the relevant story elements (mainly, Personal Life Events, as seen on). One user even mentioned that *“Associating the reading or writing of a document with World Events helped me tell the story to find the document”* (User 20). Hence, some kind of compromise must be reached when trying to collect this information. The solution to this problem, perhaps, is something User 12 said: *“I think that the questions about Personal Life and World Events could be merged into just one”*.

Element Order

Surprisingly, only users of Prototype A complained about the order in which they were asked to tell the story. Users 1 and 2 mentioned they *“(know information about) an element, but would rather mention something else at that point”*. User 3 mentioned that *“The dialogue sequence seemed a bit random and they appeared to be unrelated to each other”*. User 6 shared this opinion when stating that *“The order of story elements wasn’t the one that naturally came to my mind, causing me to expend some effort not to forget something while I had to state other elements, mainly for the Recent Document”*. User 2 provided a suggestion that might help alleviate this problem: *“Given that, often, the suggested element is not the one we’d want, maybe a set of two or three alternate elements could be suggested instead of just one?”*.

This lack of satisfaction with element order explains why, for prototype A, it was common for users to select a different element from the list (as discussed in Section 4.4.2). This didn’t happen for users of Prototype B, who never complained about it. So, the complaining must be directly related to Prototype A not mimicking the telling of stories adequately. Even if the reason for the complaints resides in the fact that Prototype A made it clearer to users that they could choose different elements at all times, the fact remains that this flexibility had the pernicious effect of distracting them from the story-telling process resulting in fewer and worse story elements.

Elements List

There were some reactions to the elements list, from which a new element can be selected. User 1 found it useful: *“When I didn’t remember anything else to tell, I read the element list in search of inspiration”*. However, he noted that *“The element order can be influenced by the element list”*. Maybe a way to prevent this, and at the same time increase the list’s usability is to *“(sort it) by the order in which they are expected in the stories”* (User 12). Not everyone agrees that the elements list has no problems. User 17 noted that *“Since there are a limited number of story elements, this can limit what we can remember about an older document”*.

Another request made by User 1 is that *“The system could show, on the elements list, which elements were already mentioned. This would be useful when several elements are already in the story and we can’t remember what we have said”*. Apart from its obvious implications regarding the elements list, this sentence shows that Prototype A makes it hard to understand what has already been said and makes it difficult for users to feel in the presence of a story.

Redundancy

Some users complained that there was some needless redundancy in the interface. This often happened when some information users conveyed for an element coincided with what should be mentioned elsewhere. For instance, User 19 considered Purpose and Subject to be the same thing (as we’ll see below, he wasn’t the only one to consider this). That user had, also, mentioned some keywords on the document when describing its Purpose, again introducing some redundancy. This led him to comment that *“It would be an important improvement if the questions asked about the different story elements depended on what had already been told”*. User 3 also feels some things are needlessly repeated, when stating that *“There is some redundancy between some dialogues”*. User 16 reiterated this point: *“Some questions are redundant in some contexts”*.

Another time where redundancy was often found was when telling the interface about the documents’ Authors and Co-Authors. Sometimes, no specific author was given more importance over the others, and all of them (Authors AND Co-Authors) were specified at the same time. This happened to User 2 and User 3, who mentioned that *“(we) already had told it the document’s co-authors, and it asked for them again....”* and *“It is unnecessary to have different dialogues for author and co-author.”*, respectively.

Supporting Other Documents

It was not very common, but two users noticed that they felt that the interface only supported text-based documents. User 18 mentioned that *“Databases unsupported”*, and User 13 wrote that *“The system should handle multimedia files (Movies, Audio, Photos, etc.) better.”*

Time-Related Interface Areas

Users of Prototype A had the choice of dragging story elements from the central area of the interface to the areas representing the past and future of the target-document. We’ve already seen (Section 4.4.3) that this was rarely done. Taking a look at the users’ comments, we start to understand why this happens: those areas are confusing. User 1 stated, right after starting to use the prototype, that *“(the) Past/Present/Future areas confusing”*. He

had a hard time understanding what could or should be moved there, since there is nothing that allows users to know if a given story element is atemporal or not. For instance, it makes no sense to say that the document's subject *was* something, before it was created. This led him to write in the questionnaire that *"The elements in the interface should be visually distinct according to their atemporality, to better understand which can be moved to the past or future interface areas."*

Not to Know is Different from Knowing Not...

Something that users of Prototype B mentioned was that it is indeed different to know something not to have happened and not remembering something. Users 11 and 12 noted this, and User 11 further mentioned that the "It didn't happen" button that allows that information to be mentioned is good because he *"(doesn't) want to need to look at the dialogue to say that something didn't happen"*.

Other Comments

Apart from what was just mentioned, there were some more noteworthy comments and suggestions made by users with the goal of improving the interface as a whole. User 4 noted that *"There should be some sound or visual signal when the suggested documents change"*, because it might be hard to acknowledge that change. This reflects that an important aspect of the interface was perhaps too successful: the document suggestions should be readily accessible but not intrusive and distracting.

User 9 noted that it would be easier to recognise relevant documents in the suggestion area if the documents' names appeared next to their thumbnails. User 1 said that some 'intelligence' might help both the system and the user in finding the document. For instance, transitivity can help gather some new information (*"if this document was written before that one and that one's Purpose was the same conference than the target-document, then..."*)

The Dialogues

A comment made by User 2 probably reflects the opinion of several users about the dialogues used in both prototypes: *"The dialogues for introducing story elements need some improvement"*. Users 9 and 13 also took the time to write something similar. Due to the difficulty in understanding some dialogues (we've seen an example of that above when discussing the several Event-related dialogues), it is a good idea to include additional information in them (descriptions, examples, etc.). This was explicitly requested by users

2 and 9: *“Display some help text directly on the dialogue boxes, instead of having to press a button, could help to better understand what is expected”*.

We will now take a look at the most noteworthy difficulties met by users when using the dialogues. This data was can the basis for improvements in future versions of the interface.

Time

Several users manifested the desire for different time periods to be directly selectable in the Time dialogue, showing that those already present were not enough. User 13 wrote it explicitly: *“Other Time intervals should be considered in that dialogue”*. User 8 also mentioned this when using the prototype.

Another aspect that causes some confusion in this dialogue is the agenda where specific events could be selected to identify when the document was written. User 4 had a hard time understanding it, and User 1 expected it to be fully functional, instead of just allowing the selection of an event. The agenda was also one of the main problems with the Purpose dialogue, as we’ll see next.

Purpose

All users except User 19 had a hard time understanding why is the user’s agenda present in the Purpose dialogue and it had to be explained to them. They found it difficult to understand how this could relate to Purpose. Mostly, this happened because it was hard to recognise the agenda for what it was, since no real appointments or events could be presented. It remains to be verified if, using real information, this confusion persists. User 2 commented that perhaps better than the agenda would be some kind of access to one’s email messages. User 4 would also like to see the agenda provide more functionality.

Moving on, the other major problem with Purpose was differentiating it from Subject. In fact, five users found this difficult, because sometimes, the document’s Purpose is stated on its Subject.

Co-Author

User 1 complained about not being able to say that, apart from the co-authors he remembered, there were some others. Also, as we’ve already seen, some users felt this dialogue to be redundant since they had already mentioned all co-authors in the Author dialogue.

Other Documents

The main comment about this dialogue was the lack of better support for stories about Other Documents, as previously discussed.

Exchanged

Some users were misled by the name of this story element. Often, they had *sent* or *received* the document by email (for instance), but got nothing in return, so no real *exchange* had occurred... This was the opinion of users 11, 18, and 20. Also, it was felt that it would be useful to separate the information about documents sent to someone from that about documents received. Users 6, 11, and 12 shared this desire.

Type

It was not evident, for all users, what file type they should mention in this dialogue, especially when the document had parts that had been produced in different applications (a Microsoft Word document in which there are some charts built using Excel, as exemplified by User 17). There was, also, some confusion with different Versions of documents, since they might have different file types (a Word document and its PDF version, for instance). Finally, some file types might not be explicit to some users. For instance, User 8 confused the Microsoft Word format with Plain Text.

Tasks

Some users were confused by this dialogue, in which only a text box was present. Without any further support, it became difficult for them to understand how they should proceed. User 8, 12, and 14 verbally manifested this.

Storage

User 19 suggested that "*it would be interesting to try to collect some information about the actual location of the file (directory, etc.)*", something the dialogue, in its present form, doesn't allow. Also, new kinds of storage not contemplated by the interface were mentioned, such as CVS repositories and PDAs.

Contents

Due to the large variability of possible document contents, several suggestions on how to improve this dialogue were made by users. Firstly, some users felt there was some redundancy between Contents and Purpose, since often the keywords to be entered were already part of Purpose. User 4 stated that *“The introduction of words in the document could be merged with stating its purpose”*, and his opinion was shared by users 5 and 8.

Also useful, according to users 1 and 2, would be to allow information to be specified in the negative form, such as telling that a document DIDN'T contain some keywords. Some kind of support in describing images in the document, rather than just specifying keywords in the text, was felt by users 1 and 11.

User 9 had doubts when choosing the document visual look since its first page had a different appearance than the remaining. In fact, some users tended to only fill in the keywords text box, or select a visual appearance. Those two things seem not to be well together, suggesting that a new story element, Structure, could be created.

Finally, users 7, 8, and 13 mentioned some information the interface was unable to capture: the document's overall structure (how many chapters it had, etc.).

Name

The largest problem with this dialogue was the confusion some users made between a document's file name and its title. It was the case of users 16 and 20.

Events

Apart from the aforementioned problems with event-related dialogues, it should be noted that the Events dialogue is sometimes particularly confusing, since it can be understood as 'anything to do with the document'. For instance, User 18 mentioned, as an Event, the proximity to Christmas. That is a Time reference or a World Event, at best. User 2 confused events with exchanges. Perhaps, as User 11 mentioned, the problem is that *“The word 'Events' doesn't make very clear what is to be said (perhaps use 'interruptions' instead?)”*.

Something New About Stories

The several interviews allowed us not only to study the different reactions to the interfaces and how stories are told, but also to find some new possible values for story elements that hadn't yet been mentioned by users in the previous study. We'll briefly

mention the most relevant ones.

Place was one of the elements that got interesting values. Some users revealed that they had worked on the document at their parent's house, or at a friend's. A user that works as a consultant said that a document had been written in the workplace, but since he moves from office to office, 'workplace' represents several actual physical places. User 14 went so far as to mention he had worked on the document 'everywhere', since he had it on his laptop he carries with him at all times.

We've witnessed that it is possible to remember Co-Authors without knowing their names: User 6 didn't remember their email address. This could also happen for the Author.

About Other Documents, one user mentioned some email messages as relating to the target-document, and another an on-line document (with an URL). More interestingly, User 15 referred a document that was created a year later than the target-document, but was used for a similar Purpose (a yearly project report).

As expected, Personal Life was a fairly diverse topic. Family-related information showed up more than once ("*I was at my parent's*", "*Both my husband and son were ill*", "*niece's birthday*"). Work-related data was also common ("*I worked with persons of another department*", "*I finished my PhD*"). Finally, there were references to time-based information ("*Right before going on vacations*", "*my birthday the previous week*", "*went out for dinner and went to sleep at 4 a.m*"). On a lighter note, a user mentioned she had written a document while she "*(...) was annoyed*". It is evident that, although some information can be found on the user's agenda or email, that isn't the case for all of it.

World Events, unlike Personal Life, were revealed as something that would be rather easy to find, just by looking at news archives. Some examples are "*9/11*", "*Lunar Eclipse*", "*Portugal won't host the America's Cup (sailing)*", "*revolution in Georgia*", and the "*suicide bombing of Italian barracks in Iraq*". Only a user mentioned something that could be difficult to verify in a straightforward way: "*a member of MSc thesis juror appeared on TV*".

Information about document Exchanges included placing the document on a shared drive on a local network, storing it in a CVS repository, making it available on the World Wide Web, giving a printed version to someone, or even sending it to another person using an Instant Messaging tool's file-sharing ability. Storage was sometimes related with Exchanges, and similar places were mentioned: CVS repository, Web site, local network, but also a company's intranet and removable media.

One user mentioned as a Version a new document that the old one had been the basis for. Finally, the few Events that occurred during the editing or reading of the document were as diverse as getting a telephone call, someone showing up at the office, or the office's ceiling falling down!

4.5. Discussion

From the wealth of information that could be extracted from the interviews and just discussed in Section 4.4, we can infer the answers we seek for the research question that motivated this study: which interface is better suited to allow users to tell stories about their computers. In the following sections we will discuss those answers and other major results that the data allowed us to reach.

4.5.1. Telling Stories told to Computers

In short, we verified that *users can, indeed, tell stories to computers in a way similar to how they do it to other persons*. There are some differences between those kinds of stories but, if anything, they are advantageous when considering the ultimate goal of building narrative-based document retrieval interfaces.

Stories Become Longer

Stories told to computers are usually longer than those told to humans. It would seem that *a structured interface can help users both to organise their thoughts and remember more information*. Reinforcing this conclusion is the fact that Prototype A, that users thought more confusing, was able to elicit shorter stories than Prototype B. Also, we've discussed that the Tasks dialogue, lacking some structuring and presenting just a text box to the users, was difficult to use. This conclusion is also confirming a result from the previous study, which stated that dialoguing interfaces were essential to help users tell their stories. We've just witnessed this to be the case.

Element Order Can Remain the Same

The order in which the users tell their stories is the one that had been found in the previous study. This is true especially for Prototype B, in which the users almost never chose a different element than the one suggested by the interface. This is not the case for Prototype A, showing that, if enough care is put into the construction of the interface, it can nearly perfectly mimic the experience of telling stories to a human listener, as far as structure is concerned.

Some Elements Become more Frequent...

Allied to the larger story lengths we find that *many elements appear more frequently in the stories*. This is true especially for Prototype B. Again, this shows how it is easier for users to remember information about their documents. Although the absolute frequencies have changed, *the several story elements maintain, for the most part, their relative importance in stories*. The relative order of some elements was very sensitive to small fluctuations, since their frequency values were close to each other. However, the order between groups of elements with similar relative importance is maintained.

...And Some Should be Considered with Care...

Several users manifested their dislike for Personal Life Events, World Events, and Events. Those elements were often confused with each other, and were hard to remember. Events, especially, was seldom mentioned, and when at all, it was often confused with something else. Some persons did use Personal Life and World Events with success. *These elements, especially Events, should be rethought and, probably, merged, to prevent confusing the users.*

It also *makes sense to divide the Contents element into two: Contents and Appearance.*

Stories told to Computers are Better...

All things considered, it becomes apparent that not only stories told to computers are similar to those told to humans, but they are, in fact, *better* if a document retrieving interface is to be built. The fact that they are longer and that some elements appear more frequently will allow those interfaces to have access to more information with which to identify the documents being sought.

4.5.2. The Best Prototype

The prototype that better embodies a good story-capturing interface is, without a shadow of a doubt, *Prototype B*. Not only is it able to elicit longer stories from the users, but also persons using that prototype find it less confusing and tell their stories without feeling the need to change the order in which the several elements are suggested to them. The way in which the several story elements are presented to the users, resorting to natural language sentences, was also shown to be important in helping them to get an overall feeling of what they had already told. Also, in Prototype B stories are presented in a linear fashion. Some psychology studies shows linearly presented information is more

easily understood by many people [Cudicio 1986]. All these factors help the users to feel more like they are telling a story, when using this prototype.

Unlike Prototype B, Prototype A was problem-ridden, starting with the different past / present / future areas that some users found confusing. The fact that several story elements were presented independently of each other, in separate boxes made it difficult for users to know what they had already told the interface, and completely broke the illusion of telling a story composed of interconnected, related, elements.

The questionnaire results confirm that, although all users liked both prototypes, Prototype B was considered, unarguably, simpler, easier to understand, and innovative.

4.5.3. Some Improvements to the Interface

Something else this study made evident is the need to improve some aspects of the interface. Some dialogues need to be changed, to better answer to the users' needs. The interface as a whole could also be improved, as we will see in the next chapter.

Other Documents and Tasks

Overall, users felt the *need for better support for telling information about Other Documents*. This confirms the findings of the first study about this necessity. Better support for telling the interface about the Tasks that were necessary to write or read the documents is also lacking .

Redundancy

Another important aspect to be considered is that users sometimes ended up repeating the same information, when two different story elements coincide (Subject and Purpose, or Author and Co-Authors, for instance). This increased the users' frustration in using the prototypes, and should be prevented whenever possible. This can be done by *providing better information about the different story elements, by building some 'intelligence' into the interface that allows it to change its behaviour to take into account information already told by the user, or by merging similar dialogues, such as Author and Co-Authors or the event-related dialogues*.

Smart Elements List

The elements list from which the users can select the element they want to mention at that time can be improved in several simple ways, such as sorting the elements according to their expected order in the story and marking those elements that have already been mentioned.

4.5.4. Conclusion

In order for the existing results to be trusted, it was necessary to verify if the patterns found in stories told to users were maintained when similar stories were told to computers, in a more structured environment. That was the goal of the present study, where we tried not only to ascertain the differences between stories told to computers and those told to humans, but also what interface is best suited to capture those stories. To that end, two different low-fidelity prototypes were evaluated by two sets of users.

We found that, indeed, stories about documents can be told to computers in much the same way in which they are told to human listeners. There are differences, but those differences actually work in favour of the construction of document retrieving interfaces. Stories tend to be longer, and some elements appear more frequently, as it appears that the structured environment in which stories are told helps to jog the users' memories. The order in which elements are mentioned in the stories and their relative importance are, essentially, the same as the ones found in the previous study. The results of that study were, thus, validated.

Of the two prototypes evaluated by users in this study, the one that better allowed them to feel like they were, indeed, telling stories, and which was able to capture more (and better!) information was the one based on natural language sentences that can be completed by users. This occurred because that way of telling stories feels more natural and less confusing to the users, but also because it gives them a feeling of 'wholeness' of the story, by presenting all its elements in context with each other.

Furthermore, we were able to extract other results regarding story elements and possible improvements to the interface.

5

The Quill System

We obtained a complete characterisation of document-describing stories and used it to create a set of guidelines for the design of narrative-based interfaces for document retrieval. Those guidelines were the basis for the design of two low-fidelity prototypes of possible interfaces. Their evaluation showed that it is possible for users to tell stories to such interfaces in a way that is similar to telling them to human listeners. However, this is not enough to validate stories as a basis for personal document retrieval. Several questions remain unanswered. We still don't know if it is possible to tell stories to a computer with no human intervention whatsoever. Also, several practical aspects, such as story trustworthiness and discriminative power remain unknown.

To properly address these issues, it is necessary to use stories to retrieve documents in a situation as close to real as possible. Only by actually allowing users to try to retrieve their documents will answers to the missing research questions be reached. To accomplish this, we created a prototype of a personal document retrieval system, which we called Quill.

5.1. The Interface

Quill's interface closely mimics the low-fidelity prototype with the best performance, described in Chapter 4 (Prototype B). The interface evolved iteratively from that prototype, successively incorporating changes motivated by user reactions throughout the differ-

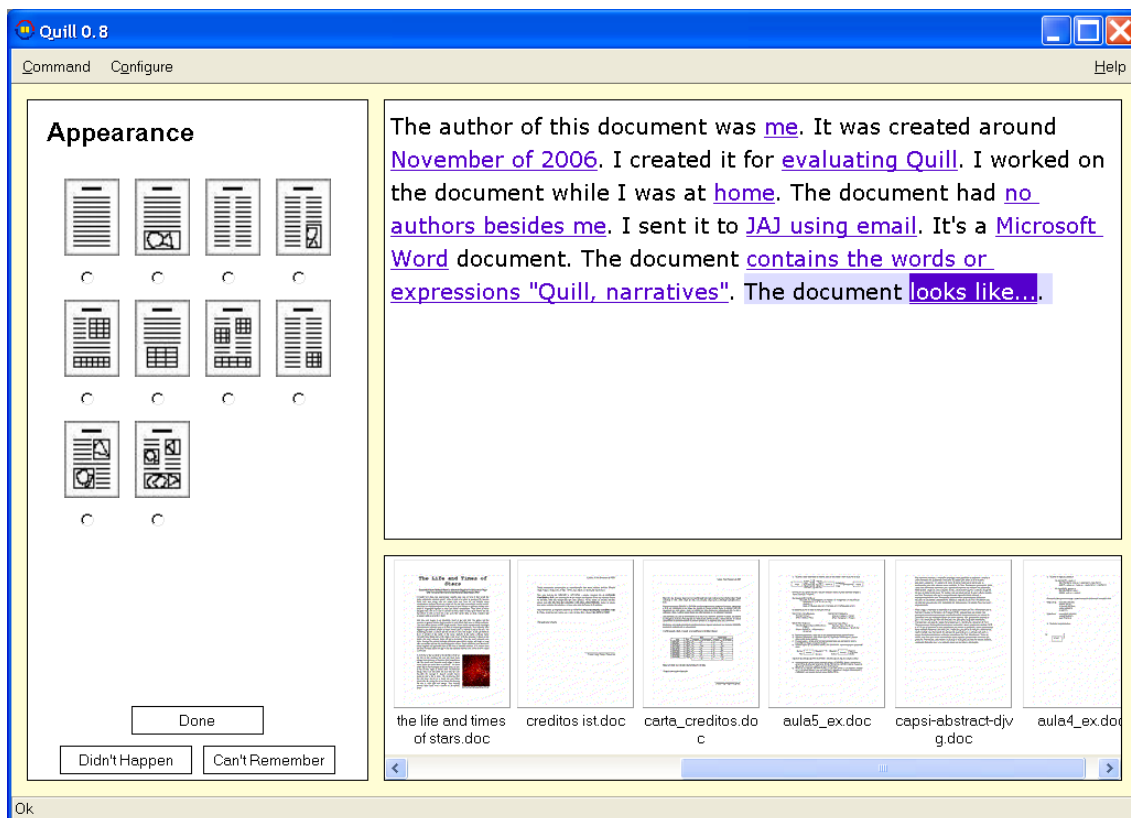


Figure 5.1: The Quill Interface

ent studies in which it was used. We will now describe the interface's (and underlying system's) final version. There are only two relevant changes between the low-fidelity prototype and it. The first are the positions of the story control buttons ("I Want Another", "I Don't Remember" and "Done"). Those were placed below the narrative in its written form and are now below the story element dialogues. The second is related with the different story elements proposed to the users. In the different tests (described in Chapter 6), we concluded that the "World Events" and "Events" elements were rarely mentioned and often mistaken for one another. They were merged into just one element, "Events". On the other hand, the "Contents" element was divided into two. The first, "Contents", refers to text or other information in the document. The second, "Appearance", allows users to mention a document's visual aspect. This was done to prevent any confusion arising when describing documents without a visual appearance, such as song files.

The application window is divided into three different areas (Figure 5.1). The larger area at the top-right corner of the interface is the Story Area. There, the story is incrementally written whenever the user enters new information. Each element is represented by a sentence, initially with incomplete information, but that will change to reflect the information entered by the users. Part of the sentence will be a hotspot that can be clicked (underlined, in the figure), allowing the correction of the information therein.

Some changes might occur to the sentences when the user enters new information, to account for plurals, etc. We strove to produce a human-readable story, while taking care not to make too many or too evident changes. If the sentences changed significantly, it might be hard for users to establish a relation between both versions of the sentences, unduly distracting and confusing them.

The information is entered into the story with the help of specialised dialogues, one for each story element. These are placed to the left of the story area (the one for the Appearance is shown in Figure 5.1). The different elements are suggested to the users in the order inferred from stories told to human interviewers, as it was done for the Low-Fidelity Prototypes in Chapter 4. The Contents and Appearance elements appear in succession. The first element asked about in each story is the document's Author. This allows the different orders found for the user's own documents and those of other authors to be followed from the start.

Some dialogues provide text boxes in which the users can write their description of a story element. The fail-soft principle was used when implementing Quill. Even if the system is unable to fully understand what was entered by the user in a given dialogue, it will carry on, instead of issuing error messages or otherwise distracting the users. We found this to be important in maintaining the feeling of telling a story, and not disrupting a narrative's flow. Also, it prevents time-consuming trial and error iterations. In some dialogues, where some data can be expected as an answer as found in the interviews, those choices are presented to users in the form of combo boxes, radio buttons and other adequate interface widgets. This helps prevent errors and makes the storytelling process quicker, by preventing users from having to write everything.

At all times, visible at the bottom of the element entry dialogues, there are three buttons that allow the user to enter additional information into the story. The "Done" button just commits whatever information the user entered in the dialogue. The "Didn't Happen" button can be used to state that something didn't occur (an Event during the handling of the document, for instance). Finally, the "Can't Remember" button should be pressed if the user cannot remember if a given element took place or not (sending the document to someone by email, for instance). A button that allowed the user to arbitrarily choose another element was removed from the early versions of Quill, since few users ever do so. Now, the dialogue's title can be clicked causing a list to pop up, from which the users can choose any of the elements.

The system is continuously looking for the target document, based on the story told so far. The most likely candidates are displayed at the bottom of the interface window. For each, a thumbnail is generated, whenever possible. This allows the users to identify the document they are looking for without getting distracted from the story. Indeed, as already discussed, one of the things the users remember well about documents are their visual aspects.



Figure 5.2: The Quill context menu

Double-clicking on a thumbnail opens the document with the application registered in the system to handle it. A right-click causes a context menu to appear (Figure 5.2). In that menu, the users can see a larger version of the thumbnail and may opt to open the document, open its containing folder, or view the document’s properties (creation and modification dates, full pathname, etc.). Furthermore, users can use the context menu to state that a document, while not the one being sought, is somehow related to it. This will introduce a “Other Document” element into the story, and present the appropriate dialogue, with some of the fields automatically filled in.

While the interface in Figure 5.1 uses the English language, the interface was implemented in a language-independent way. A Portuguese version also exists, allowing Quill’s use by non English-speakers. This allows us to extend our user base for the different user tests.

5.2. The Quill Architecture

The interface we just described is able to capture stories in a natural and straightforward way. However, in order for those stories to be understood, some infrastructure is required. The different components in the system are depicted in Figure 5.3. Central to the system is the Scroll Knowledge Base (KB). All autobiographical information required to understand the stories and find the documents is stored therein. This includes not only indexes of the users’ documents, but also of their actions while at the computer (emails sent, web pages visited, etc.). That information was deemed necessary to understand the stories’ likely contents, as seen in Chapter 3. Also, we decided to use a KB rather than a simple database because it also became obvious from the analysis of stories that additional knowledge, about the user and the world at large, would be necessary to understand them. By using a KB we can represent all sorts of common-sense-related knowledge that we’ll be able to interrelate with the autobiographic information in mean-

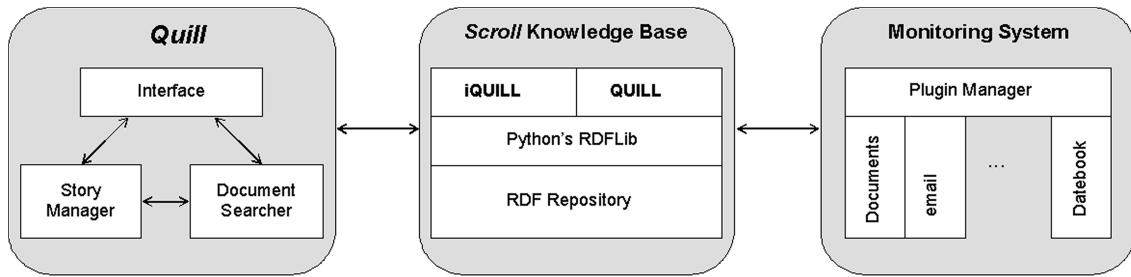


Figure 5.3: The Quill architecture

ingful ways. Indeed, our use of a KB in this way is one key difference between Quill and other desktop search programs: rather than just indexing the documents and trying to retrieve them based on their properties, we use more information, in a wider context.

All knowledge is fed into the KB by an automatic monitoring system that gathers it from different sources. This monitoring system is plugin-based. Each plugin deals with an information source. This separation between components ensures the system's extensibility, by allowing more sources to be added at a later time with little effort. The users' privacy is guaranteed given that the KB is stored locally. Only an intrusion on the users' own computer could compromise personal information. However, this would be the truth regardless of Quill's presence.

The KB makes the autobiographic information available to the Quill interface, where the users will tell their stories. A Story Manager collects all information told by users, adapting the story text accordingly and choosing what element is to be suggested to the user next. It uses a "Document Searcher" algorithm to look for promising documents, that are then displayed in the document suggestion area of the interface. In the following sections we will describe each of these components in more detail.

5.3. The Monitoring System

In order for Quill to be able to use stories to find a users' documents, it must access a wide range of knowledge. An index of the users' documents is undoubtedly necessary, but also additional autobiographic information, that can be used to understand the stories. However, relying on data explicitly provided by the users would undoubtedly fail, since no one would be willing to consistently provide it. As this would pose serious limitations to our approach, we solved this problem by preventing the need for such manual annotations. We created a monitoring system that continuously observes what is going on in the users' computers, selects the relevant information, and updates the knowledge base accordingly. This information can originate from several sources, such as the users' emails, datebook, web pages visited, etc. As such, the monitoring system was designed

modularly, with a plugin-based architecture. Each plugin is responsible for the monitoring of a different data source.

The different plugins allow not only for different data sources to be added at a later time without needing to alter the existing system, but also make it possible to allow for different configurations and applications users might have, by selectively activating different plugins. This was one of the most difficult practical aspects of Quill's creation. As Quill was intended for user testing, it required a certain degree of adaptability and robustness. It was necessary to account for texts in different languages, incorrectly formatted email messages, different applications, and different operating system configurations, just to name a few. While to do so required a fairly large and time-consuming effort, it was deemed necessary as to be able to test Quill with a wide range of users. It would not have been correct to account for just a limited set of configurations, to be found in a controlled test environment. Since it is our goal to allow users to tell stories about *their own documents* to find them, all meaningful tests must be done in the users' machines.

The implemented plugins will now be succinctly described.

Documents

Two different plugins were implemented to monitor the users' documents. The first is a one-time plugin, to be run once when Quill is first installed. It is responsible for indexing all documents already residing in the user's computer. It traverses the users' disks and indexes all relevant documents. The second plugin is a real-time monitoring plugin. It continuously monitors all changes in the filesystem, to maintain the information about the users documents up to date. To do so, it intercepts low-level operating system events, such as "File Renamed" or "File Created", and uses a rule-based system to abstract those events into meaningful high-level events to be stored in the knowledge base. This knowledge allows Quill to understand users when they mention creating or modifying a document in their stories. All information about the document itself is, of course, also updated in the KB.

Both plugins are prepared to produce a consistent result even when activated simultaneously, so that a continuous monitoring of the users' documents can be performed even when the one-time indexing is underway. A whitelist/blacklist system is in place to filter directories in which the users know no documents to exist. System directories such as "Windows" and "Program Files" are filtered out by default. Also, since we are interested in the users' documents and not any file, both consider only known document types (determined by a file's extension), such as Microsoft Office documents, images, plain text files, etc. The list of document types can be adjusted by editing a configuration file. These filtering mechanisms prevent irrelevant data to be stored in the knowledge base, while remaining flexible enough to allow correct results.

For each document, a wealth of information is stored in the KB. This includes all data that can be gleaned from the filesystem, such as file and path names, creation date, size and document type. However, a more thorough processing of every document is performed, to extract from it all information possible.

Text-based documents are converted to plain text and tokenized. Then, the Porter stemming algorithm [Porter 1980] is used to find the stems of the different words in the text. Stemming is important since it will make matching the elements in the users' stories to the keywords easier, as tense, plurals, and other inflections cease to be a problem. Finally, we use the tfidf algorithm [Salton and M.J.McGill 1983] to select the keywords that best represent the document. The 20 most relevant are stored in the KB. As with the Quill interface, the monitoring subsystem is language-independent. The Porter algorithm produces good results for English language texts and is used by default, but a module containing a different algorithm can be easily incorporated into the program to handle other languages.

Apart from the documents' contents, if any meta-data is somehow found to be associated to them, as is the case for Microsoft Office applications or meta tags in HTML pages, that meta-data will also be collected. Meta-data specific for non-textual documents is also collected whenever possible. For instance, mp3 and ogg files will be inspected for the meta information in their ID3 tags. Again, this was implemented modularly, so that new file types can be handled simply by making a module that understands them available to Quill. For instance, it would be easy to start extracting the data in the digital photos' EXIF headers.

email

A very important source of autobiographic information are a user's emails. By indexing email messages, the system knows what documents were sent or received by email, but also the subjects they were related to, the people the user knows, and when a document was handled or a subject considered.

As for the users' documents, we have two different kinds of plugins to deal with the users' emails: one to be run when the system is first installed and that collects information about all messages already in the users' systems, and another to run continuously to keep the information about the users' emails up to date. As users might resort to different email clients to manage their emails, we had to implement three different one-time email plugins, for three popular clients: Mozilla Thunderbird, Microsoft Outlook and Microsoft Outlook Express. Each of those plugins automatically finds the users' mailboxes and indexes them. Apart from the information in the messages' headers, keywords are also generated, using tfidf, for each message.

There are two real-time plugins, one for handling incoming messages and another to deal with outgoing ones. They work as proxy POP3 and SMTP servers, that stand between the users' client applications and their ISP's actual servers and are able to intercept and process all messages sent and received by the users.

All documents attached to email messages are also indexed as personal documents. While working independently from the documents plugins, the email plugins strive not to index the same document more than once. Whenever a document is discovered as being attached to a message, the plugin consults the KB to see if that document is already present in the KB. For instance, it can be a document somewhere in the filesystem that was sent as an attachment by the user. In that case, instead of creating KB entries for a new document, the email plugin simply annotates the existing document with the information that it was sent by email as an attachment. This matching is done by comparing md5 hash keys, computed for each document.

Datebook

Also important to understand stories about documents are the users' datebooks, as they provide a glimpse of the wider context that surrounds them. Plugins were created to index two kinds of datebooks: Palm Desktop (used by Palm PDAs) and Microsoft Outlook Calendar. All events in the datebooks are analysed and stored in the KB. A plugin to index all contacts in a Microsoft Outlook Address Book was also created. These plugins are cyclically executed to update the KB.

Web

To understand the subjects the users were interested on, the news they were exposed to, and get a glimpse of what was happening in the real world while they handled their documents, all web pages visited by users and documents they downloaded are also inspected by a special-purpose plugin. All meta-data available for a page is stored, as are keywords generated using tfidf. As for the email plugins, some effort is taken in ensuring that no duplicate document entries are created. This plugin works as web proxy, intercepting all requests made by the users' browsers.

Applications

A plugin that monitors the applications being run by users was also implemented. This plugin continuously monitors the processes being run in the users' machines, and resorts to a list of relevant processes to filter those that might be relevant (Office applications, for instance), storing that information in the KB. Whenever an unknown process is found

(not known to be either relevant or irrelevant), it asks the user how that process should be classified.

Knowing which applications were used can help understand when certain tasks described by users in their stories took place, and also when documents of certain types might have been handled.

Printer

By intercepting operating system events produced when a document is added to the print queue, it is possible, using this plugin, to know when a document was printed.

5.4. Quill

While the interface was already described earlier in this chapter, some underlying aspects were not explored. We will now describe some of the main features used by the interface to help users tell their stories.

5.4.1. User Adaptability

As mentioned before, the order in which the different story elements are suggested to users is inferred from that found to be most common in stories told to human listeners. This was done with the help of Hidden Markov Models that indicate what element to suggest next based on those already told by the user. While this order is, on average, a good one, particular users might have idiosyncratic ways of telling their stories. To accommodate those users, Quill monitors any order changes they might request and adapts the weights of the Markov Model to reflect this. With time, the order in which story elements are suggested becomes the one users prefer.

5.4.2. Natural Language Understanding

We constrained what the users can mention in their stories as much as possible, within reasonable limits based on the contents of stories gathered in the interviews. Even so, in some of the story element dialogues in Quill's interface, free-form text is allowed. As discussed before (Section 4.1.11), unrestricted text entry would not be adequate for Quill. However, it can be used as a way for users to better explain certain story elements, instead of entering the story as a whole. Understanding that text becomes easier as the

dialogue in which the text is entered provides the first clue to what its meaning might be. For instance, in the Time dialogue any text entered by the user is likely to describe an instant in time. The parsing of natural language (NL) sentences is, thus, performed by the different story element dialogues.

Two different parsing methods are used. Sentences are first parsed using context-free grammars. A specific grammar is used for each dialogue. A chart parser performs the actual parsing, since it provides adequate performance [Russell and Norvig 2003]. Also, we use augmented grammars to automatically derive the phrases' semantics during the parsing process. Assuming compositional semantics (the meaning of each component can be derived solely from those of its sub-components), each rule in the grammar is associated with a lambda calculus formula that is evaluated when the rule is applied to generate the semantics of the resulting phrase element. One such rule is given below:

$$S(\text{rel}(\text{obj})) \rightarrow \text{NP}(\text{obj}) \text{ VP}(\text{rel})$$

If parsing the phrase "John eats", and the semantics of "John" and "eats" are found to be John and $\lambda x. \text{Eats}(x)$, then the rule will allow the semantics of the whole phrase, $\text{Eats}(\text{John})$, to emerge automatically.

While not important in all cases, this automatic extraction of semantics can be very helpful in some cases. It is the case of the dialogue for the Time element, which tries to parse text entered in it as a reference to a time instant. In this case, the semantics generated by the parser are timestamps identifying the time period in question. For instance, the grammar for the Time dialogue generates, as the semantics of sentences such as "around Christmas a couple of years ago", a timestamp for December 25, 2005. This timestamp can then be directly compared with a document's creation or modification date.

Parsing using context-free grammars, although powerful, is rigid and very sensitive to user input. Typos or sentence structures not predicted by the grammar's designer cause the parsing process to fail. Furthermore, it is an all-or-nothing process: even if a phrase was correctly parsed, no result will be returned if the parsing algorithm fails somewhere in the sentence. A different parser was used when the chart parser failed: a Chunk parser [Abney 1991]. Chunk parsers try to decipher sentences one piece or "chunk" at a time. Instead of building hierarchic parse trees, a chunk parser uses its rules to classify different non-overlapping sentence chunks. For instance, a chunk parser might be used to identify all noun phrases and verbs in a sentence. The entire sentence does not need to be correctly parsed, which results in some words not being placed in any chunk. This makes them more tolerant to mistakes and does not require all possible sentence structures to be predicted in a grammar, but also leaves to the rest of the program the responsibility to extract some meaning from the different chunks.

When the chart parser fails to parse a sentence, the chunk parser is used to identify rel-

evant chunks. Each dialogue then tries to interpret the information in those chunks to the best of its ability, according to the domain suggested by that specific story element. Again using the Time element's dialogue as an example, it looks for chunks classified as noun phrases, looks for numbers in them and tries to find out if they represent years, months or days (for instance, "1999" cannot be a month or day, and "2nd" is most likely a day). Explicit mentions to month names in noun phrases are also detected, as are holiday names. All this data is then used to compute a timestamp that can be used in the same way as those automatically generated by the chart parser.

5.4.3. World Knowledge

The analysis of stories told by users in their interview (Chapter 3) led us to conclude that it would be important to know the context that surrounds the users in order to properly understand their stories. The monitoring system accomplishes this by storing in the KB information gleaned from the users' actions and different information sources such as datebooks and web pages. A different, more general kind of knowledge is also needed, though: common-sense knowledge about the world. While limited in scope and usage, at least some such knowledge is necessary.

For instance, if a user mentions a document was created around New Year's Eve, Quill must know that "New Year's Eve" is a holiday that occurs on January 1st of every year. This kind of knowledge was stored in the KB and used by the different story element dialogues whenever needed. Continuing the example above, if a chunk parser produced a noun phrase with the expression "New Year's Eve", it would look in the KB for some indication of what it could mean, and discover the date it refers to and its periodicity. In this way, instead of having to hard-code every such detail, Quill's expressiveness and understanding power can be enhanced just by providing it with more knowledge in the KB.

5.4.4. Searching for a Document

So far, we've seen how users can enter story elements into a story using special-purpose dialogues, and how natural language can be used in a limited context, with the support of world knowledge, to understand those story elements. We will now explain how that information is used to actually find documents.

Whenever the user enters a new element into the story, a new set of inference rules is created by its corresponding dialogue. Those rules result from the understanding the dialogue was able to perform of the data entered, resorting to the KB if necessary. Information entered into the story before can also be used to disambiguate the information

and generate the rules.

The different inference rules are then passed to the Document Searcher sub-module of Quill. It evaluates each of those rules in the KB. For instance, a rule such as

$$(Document(x) \wedge CreationDate(x,y) \wedge 1-Jan-2007 >= y \wedge 31-Jan-2007 <= y)$$

would identify, as bindings of the x variable, all documents created in January 2007. The Document Searcher module then assigns a score (positive or negative) to each document thus identified. Those scores are passed into the Searcher together with the inference rule, by the dialogues. The sum of all scores from all inference rules provides a ranking for all documents. Those better ranked (with higher scores) are suggested to the user in the document suggestion area of the interface as probable matches.

5.5. The Knowledge Base

Both Quill and the Monitoring System access the same KB (Figure 5.3). This KB uses RDF and RDF Schema as knowledge representation formalisms. RDF allows knowledge to be efficiently stored, and provides an extensible framework with which to work [W3C 2002]. There is a continuum of increasingly expressive languages in the RDF family (RDF, RDF-Schema, and three flavours of OWL) allowing us, if necessary, to upgrade the expressiveness of Quill at a later time, with little effort. It is a W3C standard that aims to bring semantic information to the World Wide Web. When this becomes a reality, such information will help us understand the users' actions when on- and off-line.

All knowledge in RDF is stored as a set of triples, in the shape ($\langle subject \rangle, \langle predicate \rangle, \langle object \rangle$). This is flexible, as it imposes little structure to what can be represented. However, it becomes cumbersome to deal with those triples directly. So, we created an abstraction layer on top of RDF which we called *Scroll*. Scroll allows more complex constructs than RDF triples to be handled with ease. The usual methods for KB interaction, *Tell* and *Ask* are available, but classes and their attributes can also be directly handled. Furthermore, Scroll is implemented as a semantic network. We implemented path- and node-based inference, and designed a schema called *iQuill* (short for Quill Inference Package), that defines a series of case-frames for the representation of first-order logic (FOL) like formulae in RDF. Scroll can use those case frames to perform inference. The expressivity of *iQuill* is similar to FOL, without the ability to represent negation and existential quantifiers (for computational efficiency reasons). On the other hand, procedural attachment and functions are allowed. It is possible, for instance, to compare two strings, check the inequality of two numbers or perform a multiplication. The following inference rule can be represented in Scroll as depicted in Figure 5.4:

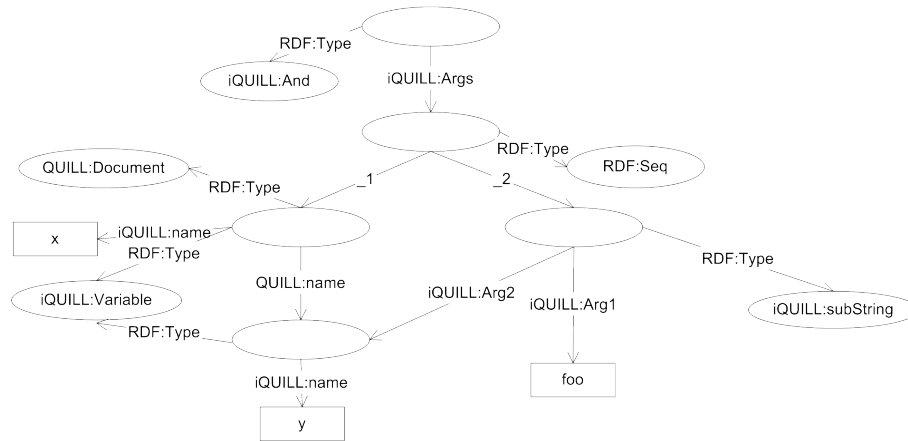


Figure 5.4: Inference Rule

$$\forall x \forall y \text{Document}(x) \wedge \text{Name}(x, y) \wedge \text{substring}(\text{"foo"}, y)$$

It returns as possible bindings for the x and y variables all documents whose names contain the string 'foo'. A complete description of Scroll and its capabilities can be found in the Scroll User Manual, at <http://narrative.shorturl.com/scroll.pdf>.

5.5.1. Representing Autobiographic Knowledge

All knowledge gathered by the different plugins of the monitoring system is stored in a Scroll knowledge base, using the Quill RDFS Schema. Two main classes are the basis for this schema. The schema's expressiveness derives from the elements found in stories in the interviews. The entire schema can be found in Appendix A4.

The "Document" Class

The *Document* class contains all fields required to store information about a document. The most straightforward of those fields allow the representation of data collected directly from the filesystem, such as a document's filename, size or extension. Other information, such as the document's creator, keywords or title can also be represented, as can references to different versions of the same document.

Instead of creating subclasses for the different document kinds, we chose to store knowledge about all documents in an uniform way, using just the *Document* class. This prevents the need to treat some documents as special cases, and makes it easier for the different monitoring plugins and Quill to handle them.

The “Event” Class

The second major class, *Event*, allows the storage of every relevant action detected by the system: sending or receiving an email, accessing a web page, meeting a co-worker, etc. Each event has a start and finish time, a description, a set of participants and of documents involved in it.

Similarly to what was done for the users’ documents, the single *Event* class is used for every possible event, allowing them to be uniformly handled by Quill. Two fields, `eventType` and `eventDirection` allow the differentiation of several kinds of events, if necessary.

For instance, if user Jack sent an email to user Jill arranging for a meeting, with an attached figure (the map with directions to Jack’s office), an *Event* of `eventType` `email` and `eventDirection` `OUT` would be recorded. Its participants would be Jack and Jill, it would point to a *Document* instance with information about the email body, and have recorded as other related documents the map.

This class has proven to be expressive enough to represent all knowledge produced by the different existing monitoring plugins.

Other Classes

Apart from the *Document* and *Event* classes mentioned above, two other auxiliary classes are defined in the Quill schema. The *Person* class represents a person, which can have multiple aliases (“Daniel Gonçalves”, “Daniel J. Gonçalves”, etc.) and multiple email addresses.

The remaining class, *Locus* is used to represent one of the users’ machines, where documents might reside. While still not in use, this class opens the way for the design of a distributed version of Quill, able to index documents across machines, and identifying their whereabouts with their locii.

5.5.2. Choosing a Representation Formalism

When creating Scroll, it was necessary to choose which knowledge representation formalism would better suit our needs, based on the nature of stories and the autobiographic knowledge needed to understand them. The choices for a formalism to use in that representation are many, from Frame-based system to logic-based approaches. Of these, we chose the Resource Description Framework (RDF), but not before carefully evaluating the different alternatives, as we will now describe.

Requisites

The characterisation of document-describing stories obtained in Chapter 3 and our own research goals impose a set of pre-requisites to what should be possible to represent in the KB and in which ways it should be manipulated.

(Most) Information is Document-Related...

In stories, most of what is told is information describing a document. So, *a way to clearly associate properties to (the representation of) a document is required*. However, not all information is directly related to the document. For instance, in a story a user mentions that a certain document was written for a conference that “happens every year”. The latter phrase pertains not the document but the conference. Hence, *the formalism should not impose the direct relation of every information element to the documents*. It should be flexible enough to allow the representation of knowledge from different sources and about different entities to be represented in an effective and uniform way. However, by thus dispersing the information, *some good inference capabilities must be supported by it*.

Different Accuracy Levels

Not all elements are equally accurate. Even for the same element this might be true, according to the document being described. For instance, a reference to Time can be, for a document, that it was written “last Thursday after lunch”, and for another “read last year around summer time”. So, *a single level of granularity for the autobiographic information cannot be imposed by the formalism*. *A pre-determined, non-extensible list of possible values is also out of the question*.

Distributed Personal Document Spaces

The documents being sought are often not stored in a single computer. Nowadays, it is common for a single user to access several computers (at home, in the workplace, etc.). This is a problem that has been recognized with recurring frequency in recent years [Nielsen 2002]. So, *a formalism that supports either a distributed KB or the easy transfer of relevant information across several computers is a necessity*.

Incremental Construction

Unlike other domains, the KB will continuously grow larger with the addition of new autobiographic information about the users and their documents. Thus, the formalism we choose should be *flexible enough to allow for that incremental information gathering and easily correcting already present information (and propagating the consequences of those changes)*.

Thousands of Documents

Previous studies [Gonçalves and Jorge 2003] also showed that the total number of documents among which is the one the user is searching for might be in the thousands or tens of thousands (but hardly more). Each element will have associated to it lots of information, as previously discussed. The formalism should be *efficient and scalable* to deal with all that amount of information.

World Knowledge

While the stories themselves convey lots of information about a user's documents, a large amount of knowledge is assumed to be known. It is the case of a document's Author, when referring to documents created by the users themselves. Things like a relative's birthday and the like are also taken for granted, along with the information that birthdays occur every year on the same day. In short, *knowledge representing facts about the world and the user are required. The formalism of choice should allow such knowledge to be represented seamlessly with other autobiographic information.*

Ontologies

An **ontology** can be defined as an explicit, machine-readable, specification of a conceptualization, for sharing and reuse. It is often related to a specific domain and provides a commonly agreed upon and shared understanding of that domain, making explicit which are the objects of the domain, that we can talk about, what are the relations linking them together, and which are the axioms governing their behaviour. Ontologies can be shared using appropriate formalisms, allowing researchers to take advantage of already existing work. For instance, if we need an ontology describing books (each book has an author, a publishing date, is organized in chapters, etc.) chances are that someone already built one and we can reuse it. The existence of ontologies and other repositories of world knowledge should be taken into account when choosing the formalism.

Possible Representation Languages Paradigms

As already discussed, the knowledge in the KB can be represented using a wide range of approaches. The following is not an exhaustive survey of knowledge representation paradigms. Indeed, while we could represent knowledge using a large number of languages and formalisms, most are not viable. Our main concerns when making that decision were:

- **The language must allow the explicit representation of knowledge.** Some approaches usable to store or, rather, *encode* knowledge do so in an implicit way. It is the case of Neural Networks, Decision Trees, and Procedural Knowledge. Since we'll need to handle specific pieces of information, in discrete instants of time, those representation paradigms weren't considered. They simply aren't suited for the storage of autobiographic information.
- **Need for an "universally-known" language:** We should look at existing, established languages, that might be based on those paradigms but possess well known properties. Those languages are often well understood and stable. This is important when considering ontology reuse, and ties directly with the next topic.
- **Availability of resources.** For existing languages, there is a large amount of documentation available, to help use and understand them. Also, several tools, from programming libraries to ontology editors support those languages. This leads to the existence of ontologies that might, eventually, be reused.

Based on the restrictions above, we considered the knowledge representation paradigms succinctly described below as viable choices.

First Order Logic

Also called First Order Predicate Calculus, it is perhaps one of the better-known knowledge representation (KR) languages. In it, we can find **constant symbols** that represent actual objects in the considered domain, **predicates** that represent relations between objects, **variables** that stand for unknown objects and **functions** that stand for functions, defining objects from other objects. The variables can be existentially or universally quantified. The predicates can be grouped into formulas using **logical connectives**. Constants, variables and functions are called **terms**, since they all, in a way or another, represent objects. Predicates can only be applied to terms.

The major advantage of FOL is that it has well-defined semantics, allowing entities to be tied to actual objects in the discourse domain and to assign truth values to formulas. Also,

FOL is **sound**, meaning that any argument that is provable with its deductive system (using Generalized Modus-Ponens or Resolution, for instance) is valid according to its semantics. It is also **complete**: any argument valid according to its semantics is provable in its deductive system (this is rarely true for more elaborate logics).

Although it is fairly simple to use, FOL has some serious disadvantages. It is **semi-decidible**: we'll always be able to find a proof that a theorem derives from a set of axioms if it exists. However, if there is no such proof, the theorem prover might never terminate! It lacks ways of structuring knowledge. Rather, all must be represented as apparently unrelated formulas. This includes no explicit support for concept taxonomies, inheritance, etc. Also, it is impossible to represent exceptions in FOL (in a feasible manner). This prevents default reasoning. Everything must be explicitly represented and derived using inference, making the representation of large quantities of information cumbersome and inefficient. Finally, it is **monotonic**, meaning that new information never invalidates existing conclusions. This might be a problem in dynamic domains where information changes over time.

Frames

Frames were first introduced by Marvin Minsky [Minsky 1981], but have been frequently used since. A frame can be viewed as the skeleton of a concept. When defining a frame, we specify the typical expectations about a given concept class. For instance, if defining a frame to represent computers we'll characterize it with **attributes** such as the processor type, display size, etc. Whenever we represent a specific computer, we'll automatically know that it has all those characteristics.

Everything about a specific concept is contained in the frame that represents it. Frames can, thus, be thought of as object-oriented. Inheritance is directly supported in frame-based systems. For instance, we could create a new frame for Macintosh Computers by deriving it from the more general Computer frame and restricting the values of some attributes.

Attributes can, themselves, be characterised by a set of properties, called **facets**. An example of a facet can be the attribute's type (eventually, another frame). Default values are another recurrent facet.

Reasoning in frames can be conducted in different ways. The *inheritance* process can derive characteristics of subclasses and instances by changing related characteristics on superclasses. *Procedural attachment* running procedures in specific situations that, for instance change attribute values when other attributes have changed. Finally, *rule-based reasoning* allows production systems to be used. In those systems, a set of rules exists. Those rules are structured in a conditional way, with an antecedent and a consequent.

Whenever the antecedent is true, the consequent will take place, eventually asserting new knowledge that allows other rules to fire.

The main disadvantage of frames is the lack of well-defined semantics and a relatively small expressive power (we can only represent what the frame language allows us to). Also, we are often limited to a single hierarchy, in which all frames must fit. This can lead to problems when more than one taxonomy applies, requiring the creation of fictitious frames to separate the existing possibilities.

Semantic Networks

Semantic Networks were first proposed by Quillian [Quillian 1967]. Systems in the Semantic Networks' family such as SNePS [Shapiro 1971][Shapiro and Martins 1990] and were still receiving lots of attention in a not too distant past [Sowa 1991].

Semantic networks represent knowledge essentially by looking at individual concepts and the relations between those concepts. Although there are many different kinds of semantic networks, they can be viewed as graphs where the nodes represent concepts and the arcs or links between the nodes represent relations between those concepts.

Although each link represents a relation, some aspects must be considered when making a choice of which links will be used. Since only nodes are concepts, we can only talk about information in nodes. So, while we could represent the knowledge that "John loves Mary" with two nodes, one representing John and the other Mary, and a link labeled "love" between them, a better representation might be to use links labeled "agent", "object" and "action" point at nodes representing John, Mary and Love, respectively, and departing from a node representing the entire sentence. In this way, we can now talk about the properties of Love (it is now a concept) and of the entire sentence (representing something like "Paul thinks that John loves Mary").

A wide range of logical connectives can also be used, depending on the actual formalism. This is also the case for taxonomies and similar relations. It all boils down to the choice of case frames (sets of links working together with a specific meaning). Even if translations to FOL are sometimes possible, the economy in the representation and understandability make semantic networks easier to use.

Inference in semantic networks is reminiscent of graph-handling algorithms. The two main inference mechanisms are **node-** and **path-based** inference. In node-based inference, we can look for similar patterns in the network, by matching nodes and their interconnections. Path-based inference, on the other hand, looks just at the path between two nodes and infers another path from it. For instance, we can say that if we spot an arc labeled `father` from node x to node y , and an arc labeled `mother` from z to x , we can infer a new arc, `grandmother`, between z and y .

The main advantage of semantic networks is that it represents knowledge in a structured way, representing the relations between concepts explicitly. Its main problem is that, for the most part, the links are arbitrarily defined by the knowledge engineer. The lack of more strict case frames with special meaning in the network (to define facets, for instance) makes them extremely versatile (we are not bound to a fixed set of possible representations) but can be a major pitfall if great care is not taken. This can lead to inconsistent representations and invalid inferences.

Description Logics

Description Logics (DLs), based on Structural Inheritance Networks, bridge the gap between semantic networks and frames, by possessing formal semantics and, simultaneously, providing ways to structure knowledge [Brachman 1977]. One of the paradigmatic systems of this kind is KL-ONE [Brachman and Schmolze 1985]. DLs describe knowledge in terms of concepts and relations that are used to *automatically* derive classification taxonomies. A major characteristic of a DL is that concepts are defined in terms of descriptions using other roles and concepts.

For instance, we can (simplistically...) define the `Car` concept as being a `Vehicle` with an `Engine` and at least four `Tires`. From this, it would be possible to automatically know that `Car` is a subclass of `Vehicle`. The KB is build from such small descriptions rather than by explicitly defining all taxonomies. Also, unlike other approaches, we have strict inheritance. It is not possible to have a `WeirdCar` concept that is a `Car` with three `Tires`. Only restrictions of existing domains are allowed. This was not the case with other approaches (notably `Frames`).

In DLs, there is a separation between the Terminological and Assertional definitions (TBox and ABox). In the TBox, concepts are defined in terms of other concepts (just like our definition of `Car`, above). It defines the *terminology* of the domain. In the ABox we represent information about actual objects in the world. It is used to make *assertions* about the world, such as specifying that `C001` is a `Car`. Concepts in the ABox are connected to the definitions in the TBox. Hence, the reasoner automatically knows that `C001` has an `Engine` and four `Tires`.

Inference in DLs can take several forms. **Subsumption** allows the verification of whether a category is a subset of another category, by comparing their definitions. **Classification** checks if an object belongs to a specific category (using subsumption if available). **Satisfiability** checks if the definition of a concept makes sense (for instance, saying that a car has more than three `tires` and less than two `tires` doesn't). Finally, we find all the usual **inheritance** properties.

Ontology Specification Languages

Several languages for ontology construction are currently in use. We'll now consider them. Older languages include Ontolingua/KIF, OKBC, OCML, LOOM, and FLogic. Those have been used in several projects to build ontologies. More recently, as a part of the Semantic Web effort that aims to bring semantic information to the World Wide Web in a machine-readable way, new languages have arisen: RDFS, DAML+OIL, and OWL. We are aware of the existence of other languages, such as OIL, CycL, SHOE, etc. However, either they are used mainly for specific projects, or they have been superseded by more recent ones, discussed in this document. Thus, they will not be mentioned here.

In Table 5.1 we can see how each of those languages is positioned in relation to the representation paradigms that we just discussed. As we can see, Frames and Description Logics are the most important influences.

	FOL	Frames	Sem. N.	DL
Ontolingua	x	x		
OKBC		x		
OCML	x	x		
LOOM				x
FLOGIC	x	x		
RDFS		x	x	
DAML+OIL		x		x
OWL		x		x

Table 5.1: Ontology representation languages by paradigm

We'll start by succinctly describing the languages and then move on to a detailed comparison.

Ontolingua

Ontolingua is the language used by the Ontolingua Server, a distributed, collaborative project for the creation of ontologies [Farquhar et al. 1996]. It is based on KIF (Knowledge Interchange Format). KIF is fairly simple, based on first order logic (and, thus, monotonic) with minor extensions to reason about relations. Ontolingua further extends KIF by including a Frame Ontology, bringing the power of frames to the language (taxonomies, etc.).

OKBC

Rather than an actual language, OKBC (Open Knowledge Base Connectivity) is a protocol for accessing knowledge stored in knowledge representation systems [Chaudhri et al. 1997]. It can be used as an uniform API to access KBs in other languages (such as Ontolingua), providing those languages share some properties (being frame-based, for instance).

While, in itself, not a KR language, its API can be used as one, since it defines an implicit, frame-based, representation formalism, complete with basic inference mechanisms.

OCML

OCML (Operational Conceptual Modeling Language) was originally developed in the context of the VITAL project, to provide operational modelling capabilities for the VITAL workbench. It is frame-based and is able to represent relations, functions and rules. OCML is more concerned with operational issues than Ontolingua. To this end it includes interpreters for functional and control terms, as well as a proof system which integrates inheritance with backward chaining, function evaluation and procedural attachments. It is, nevertheless, Ontolingua-compatible.

LOOM

The LOOM language [MacGregor 1991] is a descendant of KL-ONE knowledge representation formalism [Brachman and Schmolze 1985] and, thus, based in Description Logics. It was developed by researchers in the Artificial Intelligence research group at the University of Southern California's Information Sciences Institute.

Declarative knowledge in Loom consists of definitions, rules, facts, and default rules. As expected from a DL-based language, two separate Terminological and Assertional sub-languages can be identified. A deductive engine called a classifier utilizes forward-chaining, semantic unification and object-oriented truth maintenance technologies in order to perform inference.

FLOGIC

FLogic or Frame-Logic [M. Kifer and Wu 1995] is an attempt to integrate the Frame-based paradigm and FOL. All structural aspects of frames (inheritance, fields, etc.) can be declaratively represented in a FOL-like manner. It possesses a sound and complete resolution-based inference mechanism, and well-defined semantics.

Semantic Web (RDFS, DAML+OIL, OWL)

The Semantic Web [Berners-Lee et al. 2001a] is an ongoing project to bring semantic information to the World Wide Web. It provides a framework for semantic data to be shared and reused online, enriching the existing Web with machine-readable and understandable semantic information. It is a joint effort of several research and industrial partners led by the W3C.

The base language for the Semantic Web is the **Resource Description Framework (RDF)** [W3C 2002]. Although other languages such as XOL, OML and SHOE have been defined and used in the past, RDF is the official W3C recommendation for the semantic web and the one really in use nowadays. RDF is fairly simple, and based on semantic networks. Information is stored in triples in which *resources* (concepts) are linked together with *properties* (predicates). We can also use *statements*, to represent actual constant objects. All resources and properties are identified by unique URIs (Uniform Resource Identifiers), and namespaces are used to prevent name clashes between domains. Several types of containers, and basic tipification and taxonomy are supported. Although it can be represented in other ways (using N3, for instance), RDF is often represented in XML, taking advantage of all the existing tools for that language.

RDF alone is enough to represent assertional information about a domain. To define class hierarchies, properties, domains and ranges, we must resort to **RDF Schema (RDFS)** [Brickley and Guha 2004], a frame-based extension to RDF.

With RDFS we can define which classes exist, what are their properties, and represent information about actual objects in the KB. While sufficient for many applications, it is still not enough to build more elaborate ontologies. For instance, no cardinality constraints or transitive, reflexive or transitive relations can be modeled. Thus, other extensions aim at bringing richer descriptive power to the Semantic Web. One of the better known is **DAML+OIL** [F. van Harmelen and Horrocks 2001]. Initially two different languages, DAML (DARPA Agent Markup Language) and OIL (Ontology Interchange Language), they were merged together by a joint effort of EU and US researchers in the area. DAML+OIL is based on description logics. It has well-defined semantics.

When submitted to the W3C for standardization, it fell under the wing of the Web-Ontology Working Group that created the **Ontology Web Language (OWL)** based in DAML+OIL, fully integrating it with RDFS [Brachman 2004]. It introduces a few changes (for instance, the ability to directly state that properties can be symmetric or the removal of qualified number restrictions) but is, for the most part, similar in expressive power to DAML+OIL. OWL comes in three flavors, with growing expressive capacities: OWL Lite, OWL DL, and OWL Full.

Comparing the Languages

Oscar Corcho and Asunción Gómez-Pérez [Corcho and Gómez-Pérez 2000] defined a framework for the evaluation of Ontology Specification Languages . In this framework, the languages are compared according to seven different criteria:

- **Concepts:** This criterion is concerned with the expressiveness in which the different concepts can be represented in the language. This includes the ability to represent:
 - *Metaclasses*
 - *Partitions* (sets of disjoint concepts)
 - *Documentation* in natural language (to improve readability)
 - *Instance attributes* (can have different values for each instance); *Class attributes* (the same value for all instances of a class); *Local Scope* (avoiding clashes among attributes of different classes); and *Polymorph Attributes* (attributes with the same name and different behaviour for different concepts).
 - *Facets*, or attribute properties, including the specification of *default values*, *type*, *cardinality constraints*, *documentation*, *creation of new facets*, and *operational definitions* (invoking a procedure, rule, etc. when determining a property's value)
- **Taxonomies:** The ability to define hierarchical dependencies between the concepts, including:
 - *subclassOf*
 - *Disjoint Decomposition* (instances cannot belong to two subclasses at the same time)
 - *Exhaustive subclass decomposition* (like disjoint decomposition, but the specified subclasses include all possible instances)
 - *No subclass* (state that a concept is not a specialization of another concept).
- **Relations and Functions:** the main concerns in this criterion were:
 - *Representation of n-ary relations/functions*
 - *Use of type constraints in the arguments*
 - *Use of integrity constraints* to check the correctness of an argument's value
 - *Use of operational definitions* to infer attribute values automatically using procedures, formulas or rules.
- **Axioms:** The capability to assert axioms (always true propositions) to use as the basis for inference processes.
- **Instances:** The concerns here are threefold:

- Possibility of the definition of instances of concepts (“John is an instance of Teacher”)
- Possibility of the definition of facts (instances of relations)
- Definition of claims (assertions of a fact by an instance, such as “Paul thinks John is a Teacher”)
- **Production Rules:** usage of production rules, and their ability to have:
 - *Disjunctive and conjunctive premises*
 - *Chaining mechanisms define declaratively*
 - *Possibility and truth values attached to the rules*
 - *Procedures included in the consequent (daemons)*
 - *Updates of the KB*
- **Reasoning Capabilities:** the kinds of reasoning supported by the language.

The several languages were used by the authors to model a specific domain, allowing the comparison of their expressive powers. Tables 5.2 through 5.8, were created based on that work, adapted to the domain of narratives about documents, and including assessments of OWL and DAML+OIL, not originally addressed by the authors.

CONCEPTS	Onto	OKBC	OCML	LOOM	FLogic	RDFS	DAML+OIL	OWL
Metaclasses	+	+	+	+	+	+	+	+
Partitions	+	-	-	+	-	-	+	+
ATTRIBUTES								
Instance	+	+	+	+	+	+	+	+
Class	+	+	+	+	+	-	+	+
Polymorphic	+	+	+	+	+	+	+	+
Local scope	+	+	+	+	+	+	+	+
FACETS								
Default value	-	+	+	+	+	-	-	-
Type constraint	+	+	+	+	+	+	+	+
Cardinality const.	+	+	+	+	+/-	-	+	+
Documentation	+	+	+	+	-	+	+	+
Proc. knowledge	-	-	+	+	-	-	-	-
New Facets	+	+	-	+	-	+	+	+

(+ supported; - unsupported; +/- there is a workaround)

Table 5.2: Concepts

In terms of the ability to represent Concepts, we find differences mostly in terms of the representation of Partitions, Default Values and Procedural Knowledge. Although RDFS does not allow partitions *per se*, it is namespace-based. Since several namespaces can be used in a single ontology, this might provide a workaround in some cases.

Basic Taxonomies are supported by all languages. More complex features, however, aren't supported at all in OKBC and RDFS, and not supported in a straightforward way in OCML and FLOGIC.

TAXONOMIES	Onto	OKBC	OCML	LOOM	FLogic	RDFS	DAML+OIL	OWL
Subclass of	+	+	+	+	+	+	+	+
Exhaustive partitions	+	-	+/-	+	+/-	-	+	+
Disjoint Decompositions	+	-	+/-	+	+/-	-	+	+
Not subclass of	+/-	-	-	+/-	-	-	+	+

(+ supported; - unsupported; +/- there is a workaround)

Table 5.3: Taxonomies

RELATIONS/FUNCTIONS	Onto	OKBC	OCML	LOOM	FLogic	RDFS	DAML+OIL	OWL
Functions as relations	+	+	-	+	+	-	+	+
Concepts: unary rels.	+	+	+	+	-	-	+	+
Slots: binary rels.	+	+	+	+	-	+	+	+
n-ary rels./functs.	+	+/-	+	+	+/-	+/-	+/-	+/-
Type constraints	+	+	+	+	+	+	+	+
Integrity constraints	+	+	+	+	+	-	-	-
Operational defs.	-	-	+	+	+	-	-	-

(+ supported; - unsupported; +/- there is a workaround)

Table 5.4: Relations/Functions

With regard to the use of relations/functions, all traditional languages support most features fairly well. Most problems occur when looking at integrity and operational constraints. These often require some kind of programming code to be inserted into the ontology, which is easy to do on older languages that are LISP-based, but not so as simple in more recent ones (RDFS, DAML+OIL, OWL). We'll find this trend throughout this analysis: there is no procedural binding in those languages. Even so, DAML+OIL and OWL do allow constraints to be defined with regard to the cardinality of relations.

INSTANCES	Onto	OKBC	OCML	LOOM	FLogic	RDFS	DAML+OIL	OWL
Instances of concepts	+	+	+	+	+	+	+	+
Facts	+	+	+	+	+	+	+	+
Claims	-	-	-	-	-	+/-	+/-	+/-

(+ supported; - unsupported; +/- there is a workaround)

Table 5.5: Instances

The main difference in terms of the creation of instances is the ability to represent claims. RDFS, sharing some properties of semantic networks (where this is easy to do if facts are reified) allows this to happen (and, consequently, so can the languages built based on RDFS).

Representation of axioms in FOL is commonplace, and few languages allow SOL and named axioms to exist. DAML+OIL and OWL allow the representation of axioms about the properties of relations (reflexivity, etc.)

Production rules are seldom supported so should be a decisive factor unless really needed. Extensions to allow rules to be represented in OWL are under consideration.

AXIOMS	Onto	OKBC	OCML	LOOM	FLogic	RDFS	DAML+OIL	OWL
1st-order.logic	+	+/-	+	+	+	+/-	+	+
2nd-order.logic	+	+/-	-	-	-	-	+	+
Named axioms	+	+	+	-	-	-	-	-

(+ supported; - unsupported; +/- there is a workaround)

Table 5.6: Axioms

PROD.RULES	Onto	OKBC	OCML	LOOM	FLogic	RDFS	DAML+OIL	OWL
								1
PREMISES								
Conjunctive	-	-	+	+	-	-	-	-
Disjunctive	-	-	+	+	-	-	-	-
CONSEQUENT								
Truth values	-	-	-	-	-	-	-	-
Execution of procedures	-	-	+/-	+	-	-	-	-
Updating the KB	-	-	+	+	-	-	-	-

(1) Extensions to OWL to handle rules are under consideration (OWL Rules Language, SWRL)

(+ supported; - unsupported; +/- there is a workaround)

Table 5.7: Production Rules

In terms of reasoning capabilities, DAML+OIL and OWL are both sound and complete, a feat hard to achieve. Exception handling is seldom used. Again, we see that the lack of procedural binding is the Achilles' Heel of Semantic Web-related languages. In terms of Chaining, although RDF/DAML+OIL/OWL don't seem to support it directly, there is nothing to prevent specific applications of implementing it, as it is done for semantic nets. The formal properties of such an approach remain to be studied.

Tools

Another important factor to consider when deciding what ontology specification language to use is what tools and other resources are available. The fact is that most languages exist only in relation to a specific tool. For instance, while Ontolingua can be used to build ontologies in the general case, the main support for their use is the Ontolingua Server. Also, besides the ontology-building applications themselves, it is of capital importance in our case to analyse the existence of programming resources to handle the ontologies away from such applications.

There are several ontology editing applications. We won't exhaustively mention them all (not an easy, if possible, task). Rather, we considered the most used to have an idea of the support of each language. The following table displays those applications, and the languages they support.

As can be seen, even if they are more recent, RDF-based languages are the ones with more support. This bodes well for their future and hints at a large community of users

REASONING	Onto	OKBC	OCML	LOOM	FLogic	RDFS	DAML+OIL	OWL
INFERENCE.ENG.								
Sound	-	-	+	+	+	-	+	+
Complete	-	-	-	-	+	-	+	+
CLASSIFICATION								
Automatic classif.	-	-	-	+	-	-	+	+
EXCEPTIONS								
Exception handling	-	-	-	-	+	-	-	-
INHERITANCE								
Monotonic	+	+	+	+	+	ND	+	+
Non-monotonic	+/-	+	+/-	+	+	ND	-	-
Single Inheritance	+	+	+	+	+	+	+	+
Multiple inheritance	+	+	+	+	+	+	+	+
PROCEDURES								
Execution of procedures	+	+	+	+	-	-	-	-
CONSTRAINTS								
Constraint checking	+	+	+	+	+	-	-	-
CHAINING								
Forward	-	-	+	+	+	+/-	+/-	+/-
Backward	-	-	+	+	+	+/-	+/-	+/-

(+ supported; - unsupported; +/- there is a workaround; ND not determined)

Table 5.8: Reasoning

EDITORS	Onto	OKBC	OCML	LOOM	FLogic	RDFS	DAML+OIL	OWL
OL server	x	x						No
Protégé 2K	x	x				x	x	Yes x
Ontosaurus				x				No
WebOnto			x					Yes
WebODE						x	x	No x
OntoEdit					x	x	x (OIL)	No
OilEd							x	Yes
RDFEdt						x	x	No
KAON						x		Yes x
DOE						x	x	Yes

Table 5.9: Ontology Editors

and developers. It is noteworthy that many of the applications don't have full-fledged inference engines. So, despite the fact that certain kinds of inference are allowed by the languages, it is hard, in practice, to use them. Even the systems that handle reasoning often do so in incomplete ways. For instance, neither of them handles exceptions.

Some applications provide APIs that allow their capabilities to be used in third-party programs. It is the case of Protégé 2000, WebODE and KAON.

In terms of stand-alone libraries to build applications using the languages, without compromising with a specific programming language or application API, for the most part little is found.

For OKBC, there are four full implementations in Java and one in LISP (<http://www-ksl-svc.stanford.edu:5915/doc/release/okbc/index.html>). For OCML there is the OCML

Web Kit, written in LISP (<http://neuron-ai.tuke.sk/~hatala/ocml-webkit/webkit.htm>), and another in Java (<http://www.rizniz.com/code/ocml/download.html>). None seem to exist or to be widely known, for the remaining “traditional” languages.

For RDF, on the other hand, there are lots of libraries, for several programming languages. This includes:

- **Python:** RDFLib, 4RDF, CWM, Pырple, Redfoot
- **Perl:** RDF::Core, RDF::Notation3
- **C:** Redland (that has APIs for many programming languages), Raptor
- **Java:** Jena, Sesame,

There are libraries for other languages (PHP, Prolog, LISP, C#, etc.) as well, but nothing is gained in their systematic enumeration (a more complete list can be found at <http://www.w3.org/2001/sw/RDFCore/20030331-advance.html>). Most are simple RDF parsers, but can be used as the basis for the construction of more elaborate systems. Libraries for the use of OWL can be found at <http://www.w3.org/2001/sw/WebOnt/impls>.

Discussion

In terms of representation capabilities, there is no clear winner in terms of expressiveness. RDFS is notably less expressive than other languages. However, it does allow claims to be represented, something that might be useful in our domain, since we are dealing with a wide range of information that can be inaccurate or ambiguous. With the exception of claims, the language of choice would depend on the need for specific KR capabilities, none of which seems, at the moment, crucial in our domain.

With regard to the languages’ reasoning capabilities, DAML+OIL and OWL are the only ones that are sound and complete. On the other hand, they lack procedural binding. Non-monotonic reasoning might also be important, and isn’t supported by those languages.

In terms of editing and programming tools, RDFS, DAML+OIL and OWL are clearly better. Since they are W3C standards, there is a wide (and growing) community supporting them. Also, the fact that they didn’t appear tied to particular applications to begin with means that several stand-alone implementations and libraries are available. This is particularly true for RDF and RDFS, and not so much for DAML+OIL and OWL, but several new tools are under development. The existence of those libraries allows us to remain unconcerned with problems of portability, programming language and overall efficiency.

The structure of RDF-based languages is simpler than that of others, making the implementation of missing features easier. Also, since it is based on W3C standards such as

XML, we can take advantage of all the XML-related technology already available, including parsers, validators, etc.

It is important to notice that, despite the fact that more powerful inference mechanisms are possible with other languages, they are often not implemented in the tools available to handle them. This brings them closer to DAML+OIL in terms of available inference packages.

RDF-based languages have another advantage, that we will call “modeling scalability”: we can begin by trying to use less expressive variants, such as RDFS. If, later on, they seem insufficient for our needs, we’ll be able to switch to a more complete language (OWL-Lite, OWL-DL or OWL-Full) at little cost.

Since the Semantic Web family of languages is becoming widespread and has better support for application building, was, thus, our choice for the representation of the autobiographic knowledge required for Quill to understand stories and retrieve documents.

6

Results and Discussion

Based on user studies, we were able to understand stories about personal documents. This allowed us, with the help of the evaluation of low-fidelity prototypes to design the Quill narrative-based interface for document retrieval.

While Quill was designed based on real stories told to humans, this does not suffice to prove that narratives are, indeed, a good way to allow users to retrieve their documents. Quill might be unable to properly capture and understand those stories, or the information therein might not be enough to correctly identify a document. Thus, in order to prove our assumption, further evaluation is required .

There are four research questions that need to be satisfactorily answered before we can declare our approach a success:

Are stories told using Quill similar to those told to humans?

While evaluating the low-fidelity prototypes (Chapter 4), we were able to verify that stories told using the prototypes were similar to those told to human listeners. However, nothing guarantees that Quill is able to elicit such stories from users. Some aspects of its implementation might have failed. More importantly, while users told their stories to

the low-fidelity-prototypes there was still a human researcher present. This could have unwillingly influenced the results. It is thus important to see if stories told using an actual functional prototype share the properties of those told using the LFP.

Are stories trustworthy?

Stories are a good way to convey fairly large amounts of information about users' documents. However, while the *quantity* of information is adequate, its *quality* might not be. Indeed, it might happen that stories contain lots of incorrect or inaccurate information. If this is the case, it will be impossible to use them to find documents. The correct documents would not be identified by such stories, frustrating users and compromising the retrieval process. It is thus important to understand to what extent this happens, and verify if any accuracies that might arise are manageable and don't unduly hinder the retrieval process.

What is the discriminative power of stories?

Even if stories are told without problems and are trustworthy, it might still be impossible to use them successfully to retrieve documents if they are not discriminative enough. If a story is only able to discriminate between tens or hundreds of personal documents, this won't be enough to help users find a specific document, as it will still force them to browse through unmanageable document listings in search of the one they need.

What retrieval rate can be achieved using stories?

Even if, in theory, a truthful story can be told without problems and identifies a reduced number of documents, it might still not be possible to successfully retrieve a specific document. The knowledge base might not contain enough information to facilitate it, not all documents could have been properly indexed, etc. Consequently, it is of capital importance to estimate the overall retrieval success rate and verify if it would be satisfactory.

To answer all the above research questions, we performed two different user studies, at two different moments in time. The first study allowed us to answer the first two questions, and the second provided replies to the latter two. As will be shown, all four questions were satisfactorily answered, proving that narratives about personal documents can provide the basis for the design of correct and efficient personal document retrieval interfaces and systems.

6.1. Story Accuracy and Quality

This first study was performed with a previous version of Quill. Its interface was similar to that of the final version described in Chapter 5, but the monitoring subsystem had yet to be perfected. This didn't pose a problem as it was not necessary to fully index all of the users' autobiographic information. Indeed, this study was designed to answer the first two research questions enunciated in the previous section (story quality and accuracy). To do so, actual documents needed not be found. It were the stories themselves that were important.

6.1.1. Procedure

In order to collect data with the prototype, we needed access to the users' computers. Only thus would we be able to compare information in their stories to actual verifiable facts about real documents. This made the study more difficult and raised privacy problems. Also, average interview took between 45 minutes to 1 hour, due to the need to index (at least partially) the users' documents, and to manually retrieve the documents and compare them (and the surrounding context) with the information told in the stories. In the studies described in Chapter 3 and Chapter 4, a coffee break would suffice to interview the users, often in an informal setting. In the present study, due to the need to access the users' computers, many of the interviews were performed in the users' workplace (Table 6.1), where it became disruptive of their work and their colleague's, even if we tried to prevent this as much as possible. All these factors limited the number of subjects willing to participate in the study. In the end, we chose to interview ten subjects, the same number that had evaluated the low-fidelity prototype that led to Quill's creation. Six were male and four female, with ages ranging from 26 to 56 (Table 6.2).

Workplace	6
Home	3
Other	1

Table 6.1: Interview Location

We could have interviewed colleagues and students, but since all are experts in computer science, this would surely distort the results. Instead, we strove to deal with a broader range of backgrounds, as was also the concern in previous enquiries. The academic backgrounds and professions of the interviewed subjects can be found in Table 6.2. There are still four persons whose jobs are in the informatics area. However, since we found no bias related to either profession or academic background in previous studies, we are confident that this won't unduly distort our results.

Profession	Academic background	Gender	Age
Prep. School Teacher	High School (incomplete Physical Ed. deg.)	M	56
High-School Teacher	Electrotechnic Eng. degree (MSc student)	M	33
Monitor	Comp Sci Engineering degree	M	26
PhD Student	Biological Engineering degree	F	26
Informatics Higher Tech.	Information Systems organisation degree	M	37
Comp. Sci Consultant	High School (incomplete Comp.Sci deg.)	M	32
Informatics Higher Tech.	Comp Sci Engineering degree	F	32
Informatics Higher Tech.	Comp Sci Engineering degree	F	31
Social Security Unit Director	Economics Degree (MSc student)	F	29
Lawyer	Law school	M	27

Table 6.2: Interviewed Subjects' Profiles

After meeting the users in the place their computer is located (or one of them, if they had access to more than one), we explained the interview's goal and how it was going to be conducted. Then, the prototype was installed in the users' machines. Ensuring that all the files required for the prototype to run were placed in a well identified directory and guaranteeing that it could later be deleted without a trace was very important in securing the users' collaboration.

While the program indexed the users' documents, a quick tutorial on how the interface works was provided to the users. We also filled in the interview forms during that time. After a reasonable amount of time (around 15 minutes), the indexing process would be interrupted, if it hadn't still finished by then. This led most interviews to be conducted with only a partial index of the user's documents, preventing us from tackling the question of the stories discriminative power in this study, but it was nonetheless a positive trade-off, since more complete tests showed that a complete indexing of a medium-small sized set of documents (around 10000) could take, in the version of Quill used, up to 30 minutes, depending on the documents' types and sizes.¹

The users were then asked to tell three stories about documents of the usual three different document types: a Recent document, created by the user up to a couple of weeks ago; an Old one, created by the user at least a year ago; and an "Other" document, not created by the user at all. These are the same three kinds of documents for which stories had been previously collected in other studies. To prevent a bias due to the users' increasing familiarity with the interface, the order in which the stories about different document types were requested from different users varied (Table 6.3). All stories were timed.

After each story was told, the users were requested to actually find the document they had just described using traditional mechanisms (except in those cases where Quill actually found the document). Although no quantitative data on this was collected, it was often the case that they retold the story out loud while searching, showing it to be a natural way to organise the information. When a document was found, actual facts per-

¹At the time of the second study described in this chapter, a great performance boost had been achieved. For instance, PDF files that took around 3 minutes to index were then dealt with in 5 seconds.

	1st	2nd	3rd
Recent	5	3	2
Old	3	7	0
Other	2	0	8

Table 6.3: Position of Stories in the Interviews by Document Type

taining it would be compared to those in the story, and saved for future analysis. We will describe this process in more detail in the next chapter.

6.1.2. Storytelling

No user required more detailed explanations about how to use the interface apart from the initial tutorial. With few exceptions, all interface features were correctly used and understood. Below is one of the stories told by one of the users, copied directly from the application.

The author of this document was me. It was created around 10 of May of 2004. I created the it for PCM Report. I worked on the document while I was at home and the workplace and At my colleague’s home, in college. André Martins worked with me on the document. The document is about CGEMS Advanced Search Engine. This document reminds me of no other. I sent it to André Martins using email and LAN (shared folders, etc.). It’s a PDF document. The document contains the words or expressions “Search Engines, CGEMS, Java, SIGGRAPH” and looks like a two-column with lots of images and a little text. The document is stored in Laptop and Other computer. To write it, I had to developed a prototype for PCM, Search the Web, Read many related papers. It had different versions. Its filename was something like “pcm final”.

As it is easy to notice, there are some cases in which the English isn’t the best, but for the most part the sentence adaptation to the users’ inputs was enough to produce a good human-readable text that makes sense.

All but two users chose to tell their stories using the English version of Quill (those two did it in Portuguese). They were all fairly proficient in the language. Nevertheless, the phrases they used to tell the stories were sometimes in Portuguese. It is easier to read and understand a foreign language than to write in it.

While the users had no trouble interacting with the interface, a noticeable learning curve was observed. This was reflected in the time it took the users to tell a story, as depicted

in Figure 6.1. The time it took them to tell the third story was on average only 60% of the time spent on the first one, regardless of the type of document being described.

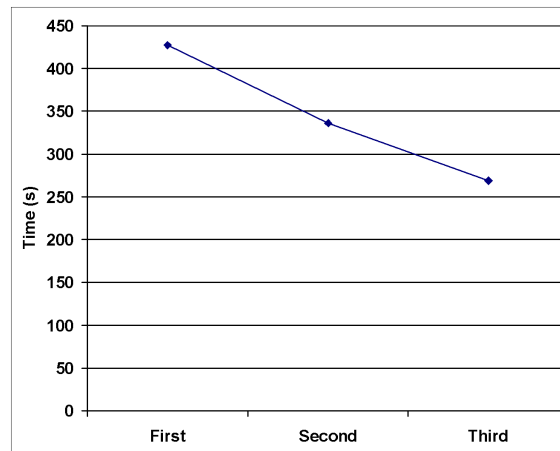


Figure 6.1: Time Spent on Stories

6.1.3. Story Quality

As stated before, one of the main goals of this study is to verify if stories told to computers are, indeed, similar to those told in similar conditions but using a low-fidelity prototype. To do so, we will compare both sets of stories in terms of structure and contents.

Story Structure

The structure of stories is reflected on two different features: story length and the order of the several story elements that constitute them. Length can be compared directly. As for the elements' order, again, as mentioned in Chapter 5, the order in which the interface suggests story elements to the users is the one inferred from interviews told to humans. More importantly, it is the same used with the low-fidelity prototype. So, by looking at how often the users chose different elements using the "I Want Another" button, we'll have an estimate of whether the element order is the same or not.

Story Length

In Table 6.4 we present the lengths of the stories collected in the current study and those told using the low-fidelity prototype.

It can readily be seen that the values are very similar, differing only slightly. The ratios between the lengths of current and previous stories are 98%, 101.5% and 100.7%, for Re-

	Current		Previous	
	Avg	StDev	Avg	StDev
Recent	14	1.054093	14.3	2.057507
Old	13.5	1.080123	13.3	1.251666
Other	13.4	1.429841	13.3	2.057507

Table 6.4: Story Length Comparison

cent, Old and Other documents, respectively. All seems to indicate that, regarding the story lengths, the two sets of stories are, indeed, equal. T-tests do not show any statistically significant difference to exist. Calculating confidence intervals for the difference of the means (Table 6.5) reinforces this result, as it shows that, with 95% confidence, that the difference can be zero or, at the most, 1.73 (for Recent documents).

	Recent	Old	Other
σ_{diff}	0.731	0.523	0.792
$1.96 \times \sigma_{diff}$	1.433	1.025	1.553
C.I.	[-1.73, 1.13]	[-0.82, 1.22]	[-1.45, 1.65]

Table 6.5: Story Length Difference Confidence Intervals

Element Order

Trying to keep the results as directly comparable as possible, the order in which the interface suggested the different story elements to the users is the same that was followed for the low-fidelity prototypes. That order was inferred from the very first stories collected, told directly to humans in an unstructured environment with no intervening system (Chapter 3). As such, we can assume that given that users were able to request another element at any time they so desired, the number of times they did so reflects differences between the presented order and another that would be more natural to them. The fewer the changes, the more convenient is the element ordering.

The number of times that another element was chosen is presented in Table 6.6.

	Recent	Old	Other	Avg	StDev	% Users	% Stories
Current	0.1	0.1	0.1	0.1	0	10%	10%
Previous	0	0	0.2	0.07	0.12	10%	3%

Table 6.6: Explicit Order Changes

Again, as for the previous study, only one user ever requested different elements than the ones suggested by the interface. Furthermore, this happened only 0.1 times per story, regardless of document type. In absolute terms, the fact that the number of stories in

which it occurred was now 3 instead of one is not as relevant as noticing that all three were told by the same user. We're dealing with her individual usage patterns, rather than an overall tendency. In short, we conclude that the element order remains, for the most part, unchanged from that found using the low-fidelity-prototype and in stories told to humans.

Story Contents

We will now analyse to what extent are the stories' actual contents similar to those told using the low-fidelity-prototype. From the results of the previous section, we can expect at least some degree of similarity to exist, since the transitions between the different story elements occurred in the same order. What remains to be seen is whether their relative importance remains unchanged. Thus, in the following lines we will examine the element's occurrence frequencies in the different stories, comparing them to the ones previously taken as typical.

Table 6.7 summarises both the frequencies found and their differences to those in the previous study, highlighted in orange. A similar comparison, for the overall values, can be seen graphically represented in Figure 6.2.

	Recent		Old		Other		Overall	
	Current	≠ prev.	Current	≠ prev.	Current	≠ prev.	Current	≠ prev.
Time	10	0	10	0	10	0	30	0 (0%)
Place	10	0	10	0	9	-1	29	-1 (-3%)
Co-Author	10	0	9	-1	9	2	28	1 (4%)
Purpose	10	0	10	0	10	0	30	0 (0%)
Author	10	0	10	0	10	0	30	0 (0%)
Subject	10	1	10	0	10	1	30	2 (7%)
Other Doc.	9	2	6	0	7	0	22	2 (10%)
Personal	3	-2	3	-2	2	-3	8	-7 (-47%)
World	2	-3	2	-2	2	0	6	-5 (-45%)
Exchanged	9	0	8	0	10	0	27	0 (0%)
Type	10	0	10	0	10	0	30	0 (0%)
Tasks	8	0	7	-1	6	-2	21	-3 (-13%)
Storage	10	0	10	1	10	2	30	3 (11%)
Version	10	0	9	0	10	1	29	1 (4%)
Contents	10	1	9	1	10	1	29	3 (12%)
Events	1	-3	2	1	1	-2	4	-4 (-50%)
Name	8	1	9	4	8	2	25	7 (39%)

Table 6.7: Element Frequency and How it Differs from Previous Results

In the rightmost of the table's columns, we can see not only the absolute value of the difference between the frequencies, but also the percentage of the original values that those differences represent. It is immediately apparent that, with few exceptions, the frequencies are very similar. For many cases there are no differences at all, and where they are found they are of little importance. The larger differences seem to occur for four

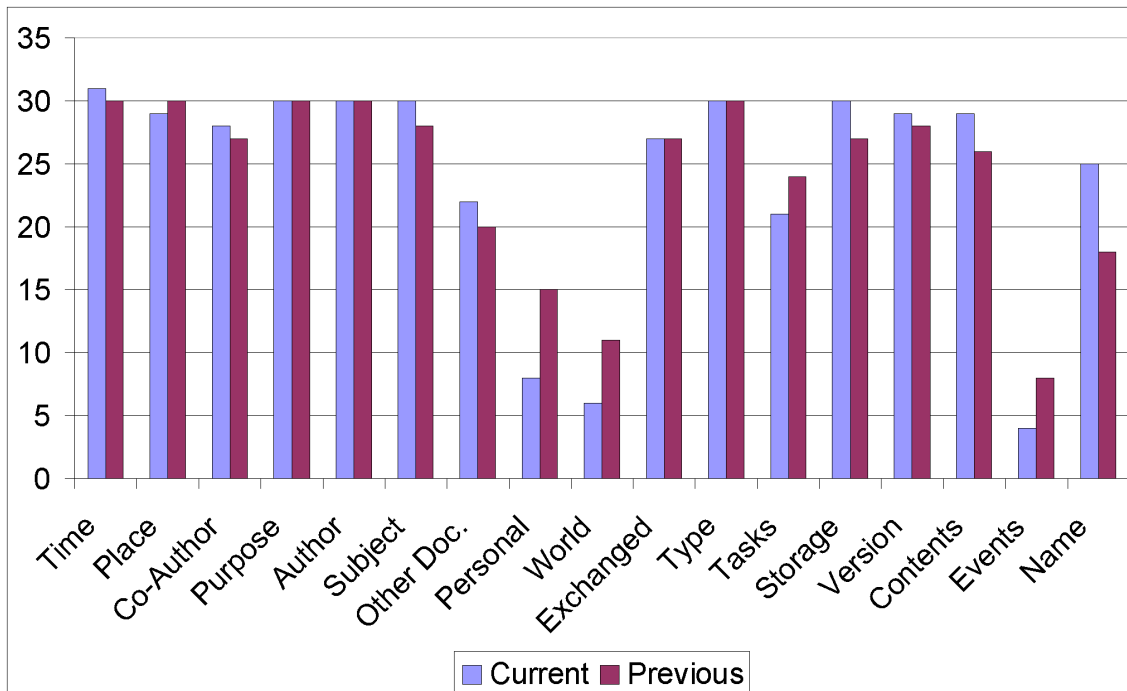


Figure 6.2: Comparison of the Overall Element Frequencies

elements. Document **Names** were mentioned 39% more often in stories told to the prototype, while **Personal Life** information, **World Events** and **Events** occur less frequently (47, 45 and 50 percent, respectively). The difference regarding Names might be explainable by a conjunction of two facts. Firstly, all users were working on their computers prior to our arrival, which may have made them more conscious of their directory structures and naming conventions. Many were at work, and two of them mentioned explicit naming guidelines for professional documents. In second place, the simple fact that they were sitting at a computer might have placed them in a more “appropriate” mindset than in a more informal environment, as was the case of the previous study. Regardless of why, more information is always welcome... As to the remaining elements, the results are hardly surprising. Those are the three elements that have proven to be more unreliable and harder to remember in all of our studies. Most users simply don’t associate them with documents. Given that we are comparing fairly small numbers, when dealing with them, even small changes are reflected in large percentages.

Looking now at the relative importance of story elements, regardless of absolute value, we find that the results found so far are reinforced. Directly comparing the absolute order of the different story elements would not be correct, as already discussed in Chapter 4. With the small numbers we’re dealing with, small changes could lead to element order swaps. Instead, we decided to consider the elements’ relative importance to each other in the following way: looking at the element frequencies for each study, we can easily find that there are two different types of elements: those that are mentioned in nearly all stories and those far less important, mentioned much rarely. A quick look at Figure 6.2

reveals this. We divided the element set according to this criterion. More specifically, all elements mentioned in at least 70% of stories went into the “Common” group, and the remaining went into the “Rare” group, highlighted in Table 6.8.

The elements in all columns were sorted according to their frequency for the particular study and document type they relate to. “Rare” elements are highlighted by a darker background. For Recent documents, no changes were found. For Old and Other documents and overall, only the Name element ceases to be “Rare” and becomes “Common”. As already stated, the fact that Personal Life, Events and World Events have different frequencies in absolute value isn’t very relevant because they were always unimportant and remain so.

In conclusion, the contents of stories remain largely unchanged from those collected in the previous study, with the notable exception of the name element that gained some importance.

Recent		Old		Other		Overall	
Current	Previous	Current	Previous	Current	Previous	Current	Previous
<i>Events</i>	<i>Events</i>	<i>Events</i>	<i>Events</i>	<i>Events</i>	<i>World</i>	<i>Events</i>	<i>Events</i>
<i>World</i>	<i>Personal</i>	<i>World</i>	<i>World</i>	<i>Personal</i>	<i>Events</i>	<i>World</i>	<i>World</i>
<i>Personal</i>	<i>World</i>	<i>Personal</i>	<i>Name</i>	<i>World</i>	<i>Personal</i>	<i>Personal</i>	<i>Personal</i>
Name	Name	<i>Other Doc.</i>	<i>Personal</i>	Tasks	<i>Name</i>	Tasks	<i>Name</i>
Tasks	Other Doc.	Tasks	<i>Other Doc.</i>	Other Doc.	Co-Author	Other Doc.	Other Doc.
Exchanged	Tasks	Exchanged	Contents	Name	Other Doc.	Name	Tasks
Other Doc.	Contents	Co-Author	Exchanged	Co-Author	Storage	Exchanged	Contents
Author	Exchanged	Contents	Tasks	Place	Tasks	Co-Author	Co-Author
Co-Author	Subject	Name	Storage	Author	Contents	Contents	Exchanged
Contents	Author	Version	Version	Contents	Subject	Place	Storage
Place	Co-Author	Author	Author	Exchanged	Version	Version	Subject
Purpose	Place	Place	Co-Author	Purpose	Author	Author	Version
Storage	Purpose	Purpose	Place	Storage	Exchanged	Purpose	Author
Subject	Storage	Storage	Purpose	Subject	Place	Storage	Place
Time	Time	Subject	Subject	Time	Purpose	Subject	Purpose
Type	Type	Time	Time	Type	Time	Type	Time
Version	Version	Type	Type	Version	Type	Time	Type

Table 6.8: Relative Importance of Elements

6.1.4. Story Accuracy

Together with the stories themselves, we collected data on whether the information they contain is accurate or not. We did this by verifying each of the stories’ elements to the best of our ability once the users had finished telling them. In this section we will describe how this was done and discuss the stories overall accuracy levels.

Methodology

Not all elements are amenable to the same degree of verification. For instance, a document's filename can be easily checked against the real filename, making its confirmation a trivial matter. If the document had exchanges or not is not as easy to verify, since it would entail checking every email message, every file-transportation medium, every interaction with other users. This is, of course, impossible to do, and even if it was done, 100% certainty could never be ensured. For those elements, the users had to make a case for their choices and were questioned about them until the interviewer was satisfied there was nothing more to learn. If it seemed reasonable enough, given other hints directly collected from the users' computers and the documents themselves, we considered the information as accurate. This wasn't just asking the users "are you sure". The elements were thoroughly explained, examples of meanings that might have eluded the users were given, and "no stone was left unturned" when questioning the users.

Even so, and to make things more correct, we distinguish between two different "accurate" statements: those we managed to verify "beyond all reasonable doubt", mostly from "hard data", and those we just had no way of verifying but seemed to be correct. There are, therefore, "Accurate Unverified" and "Accurate Verified" elements. We will present our results based on that distinction. In some cases, it was just impossible to mark something "beyond all reasonable doubt". For instance, if a user told us no Event occurred, we have no way of verifying it.

Below are the methods used to verify the elements' validity (apart from the dialogues with the users we just described):

- **Time:** we checked the actual file's system dates, either for creation or modification. We had some degree of tolerance when deciding if a data was correct or not: 1 day for documents written "one" or "two days ago"; 5 days for documents written "one" or "two weeks ago"; 10 days for documents written "a month ago"; 90 days for documents written "a year ago" and a month for documents written around any specific date specified by the users.
- **Place:** if dealing with a desktop system and the document was supposedly worked on at its location, we considered it correct. Otherwise, we questioned the users based on the documents' contents and apparent purpose.
- **Co-Author:** we checked both the filenames and contents looking for the author's names. Documents of a clear personal nature were considered as having just one author, as told by the users. Otherwise, we questioned the users based on the documents' contents and apparent purpose.
- **Purpose:** we questioned the users, in cases where purpose was not evident from the contents.

- **Author:** it was usually possible to tell without a doubt who the document's main author had been, either since it appeared explicitly in the content or because of its subject or purpose.
- **Other Documents:** we requested to see the documents whenever possible (often in the same folders), confronting the users with them. In cases where bits of documents were used in other documents, the verification was immediate.
- **Personal Life:** if it was something in the users' agenda or known personally by the researcher (some of the interviewed are old acquaintances), we considered it ok.
- **World Events:** the main way to verify this was to resort to the users' own knowledge of world events (verifying it when necessary) and checking that they happened around the Time of the document.
- **Exchanges:** we tried to understand if this information was accurate or not from the document's contents and apparent purpose, and by dialoguing with the users.
- **Type:** document type was easy to directly verify
- **Tasks:** if the tasks described by the users reflected on the document's contents (inserting images, preparing graphics, etc.), we would verify it directly.
- **Storage:** directly verified
- **Versions:** we looked in the directory where the document is stored in search of other versions of it.
- **Contents:** directly verified
- **Events:** there really was no way to verify this apart from dialoguing with the users.
- **Name:** directly verified. We considered it correct if only a partial name was given (for instance, "test" for "test2003.doc")

Overall Accuracy

We were able to verify beyond any doubt 81% of the elements, on average, which is fairly good considering the limitations described in the previous section. This allowed us to detect some elements that were not accurate, and to verify most that we considered to be so. A summary of the accuracy results can be found in Table 6.9 and Figure 6.3. There, we can see the number of elements in all stories ("Total"), the number of elements that were considered to be accurate ("Acc."), which of those we were able to verify "beyond any doubt" ("Verified"), and for those for which we were unable to do so ("Unverified"). The percentages of the number of elements those values represent are also in the table, and depicted in the graphic.

	Total	Acc.	Unverified	Verified	% Acc.	% Unverified	% Verified
Recent	162	151	27	124	93%	17%	77%
Old	148	137	26	111	93%	18%	75%
Other	149	132	30	102	89%	20%	68%
Overall	459	420	83	337	92%	18%	73%

Table 6.9: Story Accuracy Summary

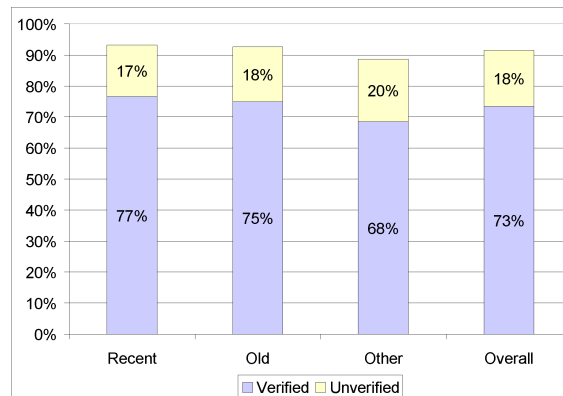


Figure 6.3: Story Accuracy Summary

There doesn't appear to be relevant differences between stories about Recent or Old documents! It would seem that users are equally good remembering them. T-tests didn't confirm the existence of any statistically significant difference. Furthermore, the 95% confidence interval of the difference for the means is $[-0.8, 2.8]$, showing that the difference can be, indeed, zero. Also, stories for documents of other authors seem more inaccurate than those of the user. This is something that could already be expected. T-tests establish they have, indeed, different accuracies (with 95% confidence).

Moving on to the overall accuracy of stories we find it to be, while not perfect, more than adequate. Looking at all the elements we considered accurate, accuracy values are around 90%. In the conservative worst-case scenario in which we consider only verified elements, the numbers are around 70%. On average, between 73% and 92% of what users tell in their stories is, indeed accurate. Since there are 17 possible elements in a story, it means that between 5 and 1 will be wrong.

If we look at Table 6.10, we see that most of the difference between the elements considered accurate and those that were verified to be so beyond doubt is due, mostly, to three hard-to-verify elements: Personal Life Events, Events and World Events. These are also the rarest in stories. Hence, they could not be checked most of the time because what the users told was that nothing had happened, and this is very hard, if not impossible, to do. If we ignore them, the corrected values for story accuracy, displayed in Table 6.11, bring closer together the verified and unverified values.

Overall				
Element	Verified?	Acc.	Acc. Verified	% Acc. Verified
Author	30	28	28	100%
Co-Authors	29	26	26	100%
Contents	29	27	26	96%
Events	3	20	3	15%
Exchanges	21	28	19	68%
Name	22	16	15	94%
Other Documents	18	20	14	70%
Personal Life	8	18	7	39%
Place	25	29	25	86%
Purpose	29	30	29	97%
Storage	27	29	26	90%
Subject	29	30	29	97%
Tasks	15	28	15	54%
Time	29	23	22	96%
Type	29	26	25	96%
Versions	26	26	24	92%
World Events	4	16	3	19%

Table 6.10: Overall Numbers of Verified Elements

	Acc.	Unverified	Verified	% Acc.	%Unverified	%Verified
Recent	129	11	118	92%	8%	84%
Old	122	14	108	93%	11%	82%
Other	115	16	99	88%	12%	76%
Overall	366	41	325	91%	10%	81%

Table 6.11: Corrected Accuracy Values

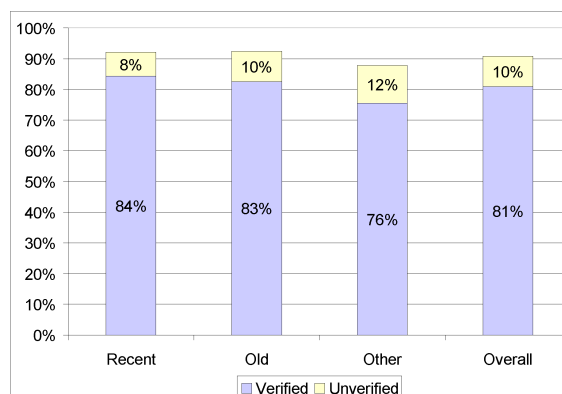


Figure 6.4: Corrected Accuracy Values

In Figure 6.4 we see the graphic for the corrected values, which gives us a more immediate idea of the improvement. Considering these new values, between 81% and 91% of elements can be expected to be accurate, overall (a bit less for documents of other authors). This will correspond to 1 to 3 story elements. Is it safe to assume the new values that disregard some elements? Given that we're dealing not with the users providing incorrect information but not providing any at all, yes. In fact, we hope to extract from the users information they remember in association with a document. For those three elements, if they don't remember anything it is as good as if it hadn't, indeed, happened. This motivated the merging of those elements into just two, Events and Personal Life, as described in Chapter 5.

Element Accuracy

It is important to know what elements are more often wrong, to better cope with their imprecision. The graphic in Figure 6.5 shows the accuracy of each separate element, for all document types. The same information, in table form, can be found in Table 6.12.

Some elements were notoriously difficult to verify. It is the case of the three least accurate elements are Personal Life, World Events and Events that we have discussed in the previous section: they are seldom mentioned by users that just say "nothing happened", which is impossible to verify. Also in this situation are Exchanges and Tasks. We were convinced they were correct in most cases, but unable to get "hard data" to verify them.

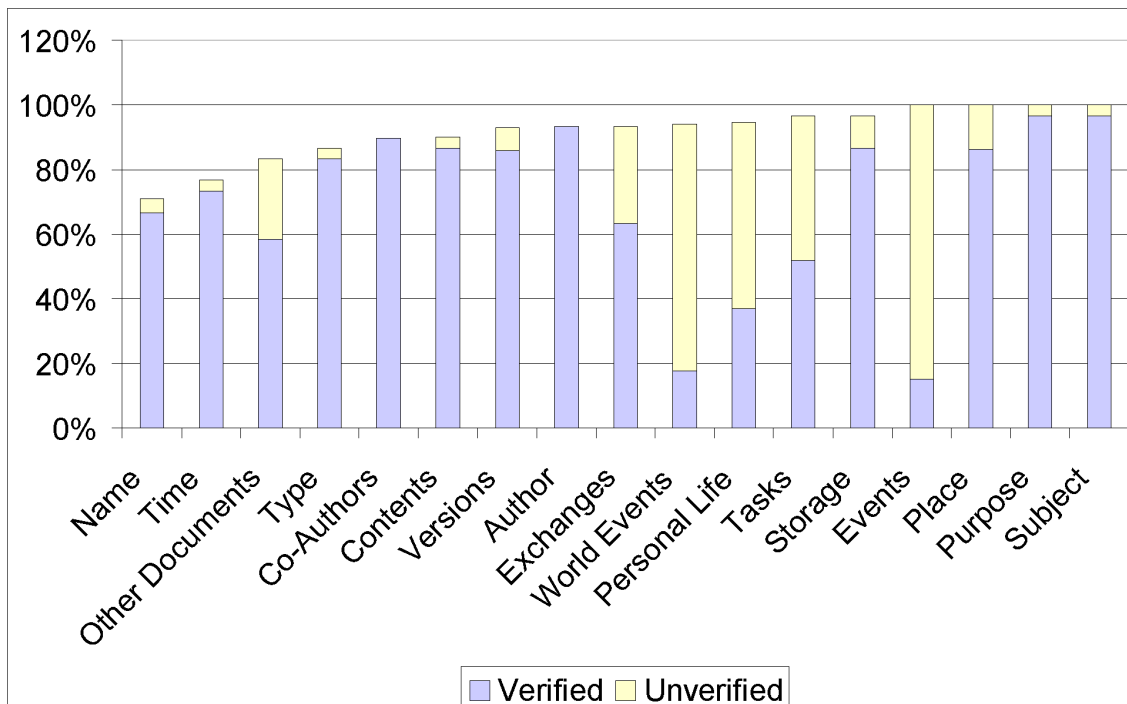


Figure 6.5: Sorted Element Accuracy

Element	Verified	Total Acc.	Unverified
Name	67%	71%	4%
Time	73%	77%	3%
Other Documents	58%	83%	25%
Type	83%	87%	3%
Co-Authors	90%	90%	0%
Contents	87%	90%	3%
Versions	86%	93%	7%
Author	93%	93%	0%
Exchanges	63%	93%	30%
World Events	18%	94%	76%
Personal Life	37%	95%	58%
Tasks	52%	97%	45%
Storage	87%	97%	10%
Events	15%	100%	85%
Place	86%	100%	14%
Purpose	97%	100%	3%
Subject	97%	100%	3%

Table 6.12: Elements Sorted by Accuracy (Overall)

Name seems to be the least frequently well remembered element. We witnessed cases that were altogether wrong, but also cases in which the users had some idea of the real name and swapped parts of the name (“*janeiro2005*” for “*2005janeiro*”, U7), the actual name was abbreviated (“*Exame Nacional*” for “*EN*”, U2), or where part of the name suggested by the users was part of the real name (“*estudo produtividade bolseiros*” for “*produtividade bolseiros – calculos*”, U4).

Next, we find **Time**. For the most part, the wrong elements were “near misses”, falling just outside the predefined tolerance intervals. They are already fairly generous but perhaps they should be adjusted.

The third less accurate element are **Other Documents**. More often than not, the users got something right, but part of the information would be wrong. It was the case of U9, that correctly stated the document had the same subject but not its name, or U4, that mentioned the correct authors but, again, not the name. It seems the same elements cause problems for either the target document or the Other Documents.

Type mix-ups were due, mainly, to confusions between formats of the same kind: plain text for Word and Powerpoint for image (U1), or PDF for Word (U7), for instance. Online documents also caused some trouble. For instance, U6 was unsure of whether a Microsoft Access database file it had, usually accessed through his personal web site was a “Web Document” or “database”. This might motivate some interface changes.

The other elements all have accuracies above 90%. No further relevant error trends could be identified.

Regarding eventual differences in accuracy for the different document types, the only noteworthy aspect is that Author is far less well remembered for Other Documents than for documents of the user (80% vs 100%).

6.2. Discriminative Power and Retrieval Rate

Having verified that stories about documents can be told to Quill in a way similar to that in which they are told to humans, and asserted that the information in those stories is trustworthy, we performed a second study. This study's goal is twofold: verify the stories' discriminative power and measure the retrieval success rate. The first goal is straightforward to accomplish: after each user tells a story, it is necessary to verify how many documents it describes. The second goal, however, is harder to attain, due to both theoretical and practical limitations.

The traditional measures for the evaluation of a retrieval system are its precision and recall, that measure how many of the documents returned by the system are relevant, and how many of the existing relevant documents are identified. These measures would be nearly impossible to measure in our case. Furthermore, they are not adequate to evaluate a system such as ours, dealing with personal documents. Precision does not make sense, as the users are looking for specific documents, rather than any documents that might fit some general criteria, as is the case, for instance, of web search. Precision, thus, is equivalent to retrieval success. Recall, on the other hand, requires the number of documents that would satisfy the user to be known. But, since we are dealing with real collections of personal documents rather than with a pre-existing test set, that number is unknown. It would require the users to inspect all their documents and state which would be relevant, something they would most certainly not be willing to do, even if it was practical. And, again, most times only a document would be relevant, so there would be no point in measuring recall anyway. Alternatively, the retrieval success rate and discriminative power of stories are better measures to evaluate the approach's quality.

Another problem that had to be dealt with when conducting the study is that each interview might literally last for several hours. Firstly, it will be necessary to index all of the users' documents, email messages, agendas, etc. Empirically, this can take between a few minutes, for users with just hundreds of documents, to several hours if they number in the tens of thousands. Then, the interview itself will spend another 30 to 45 minutes of the users' time. Overall, this amounts to a time span that makes it impractical to visit the users at their workplace (where most users have their documents) and performing the entire interview in one session, as it would be damaging to their work.

A similar concern arises when considering an eventual direct comparison of our approach with other systems. The current state of the art for desktop search is Google Desktop. As such, it would be the prime candidate for comparison. However, to set up a second system would further compound on the problem, causing most users to refuse to participate. Also, Google Desktop follows a different approach from ours, as it relies primarily on text-based information inside documents, or their file names, and its interaction paradigm is keyword entry. Our system, Quill, is very different from Google's as it allows a large amount of diverse autobiographic information to be entered and used to retrieve documents, of which keywords are just one aspect. While Google Desktop is based on entering as little information as possible, Quill expects the users to tell a full story about their documents. Ultimately, Google Desktop could be used as a small part of Quill, with the help of its API. In short, while both systems strive to accomplish the same goal (document retrieval) their underlying methodologies and interaction paradigms are different.

Another problem underlying our tests is that stories about documents might be of particular relevance to retrieve documents that are hard to find by other means, namely, navigating the hierarchy, when the users have some idea of where to find them. This occurs, mainly, for older documents, harder to remember and retrieve. However, the autobiographic information that we will be able to collect regarding those documents will be limited. While a continuous gathering of such information is possible, it would require the system to have been installed on the users' computers for months or years, which clearly is not the case. This is a limitation we must cope with. The problem will be minimised by trying to correlate as much of the available information as possible (emails, etc.) but will nevertheless have some impact on the results.

Finally, it must be kept in mind that the Quill system is just a proof-of-concept prototype and, as such, it might not realise all the approach's potential.

The methodology we now describe takes into account all these problems and limitations.

6.2.1. Procedure

Given the aforementioned, we evaluated the approach's intrinsic merits rather than those aspects that might depend on the prototype's quality or other extraneous factors.

Quill was installed in the machines of 20 users, and instrumented to create a log of all stories told, as well as the interface features used. In a first interview, the users were explained the study's goal and Quill was installed and configured. It then started indexing all of the users' documents and also the email messages and agenda, whenever possible. This setup phase was done by the researcher, rather than, for instance, making Quill available online for the users to install. This was important as many

users lacked the expertise to do so themselves. Furthermore, it was vital to correctly configure Quill. As it had to cope with a wide range of applications and configurations for the different users' machines, it required a simple but careful setup.

Since the indexing of all relevant information took some hours, the users were instructed to leave their computers running during breaks or during the night, so that the indexing might proceed with as little impact as possible.

In the following day, on a second interview after the indexing process has finished, the system was evaluated. In those cases where it became apparent that the indexing process would not take no more than a few minutes, and the users were willing to wait, both sessions were merged into one, thus making the entire process quicker and less disturbing for the different participants involved in it.

The evaluation proceeded as follows:

- Indexing statistics were collected using a special-purpose script.
- The Quill interface and its features were explained to the users.
- The users were requested to remember, in turn, three different documents (as was done in all previous studies): a Recent document (created by them up to two weeks ago), an Old document (created by them over six months ago) and a document of some Other author.
- Using Quill, they tried to find those documents, while the interviewer took note of relevant behaviours, comments, etc. in an interview form.
- We recorded on which position in the document suggestion area the target document appeared, and after how many story elements.
- The result of each search was analysed, as belonging to one of the following categories:
 - The document was found
 - The document was found “one click away” (the user found a document in the same folder, recognised the fact, and opened it using Quill)
 - The document wasn't found because of some mistake in the story (although Quill can cope with some amount of mistakes, there is a limit beyond which this is no longer possible).
 - The document wasn't found.
- Whenever relevant (especially if the document wasn't found), the reason for the failure was be noted, if possible.

- The information in each story was entered in a helper tool, the KB Analyser. This tool allows the discriminative power of stories to be evaluated, as it looks in the KB for all documents that match the story.
- It was registered, for each document, if keyword-based information in the story would have identified it, as well as non-keyword information and filename information. This allowed us to verify if some documents found using Quill would *not* be found using keyword-based approaches such as Google Desktop.
- The user filled a satisfaction questionnaire and be able to offer comments and suggestions about the system.

Complete privacy was ensured to all users, an important factor giving that we were dealing with their personal and often sensitive documents. To that end, all data collected is anonymized. Also, at the end the users' systems were restored to the state they were before the interview. We found this to be an important factor when ensuring user compliance to previous studies.

Success Criteria

It is important to define beforehand the values to be reached in order for the approach embodied in Quill to be a success. Those values were:

- A discriminative power of at most five: the stories identify, on average, five documents at the most. This is important as five is the number of documents that can be shown comfortably as search results in Quill using standard computer monitors. Furthermore, if more documents were to be displayed it would take longer for the user to scan the document suggestion list when looking for a possible match. This might distract them from the story being told and interfere with the storytelling process, which should be avoided.
- The retrieval success rate for textual documents should be of at least 75%.
- The retrieval success rate for non-textual documents should be of at least 50%.

Since we were dealing with a proof-of-concept prototype, these values are good enough to show the approach's validity and confidently show that an industrial-grade product would yield much better results. Also, the fact that more autobiographic information might have been collected in a real usage situation when the system has been running for months would further improve the results. This is especially true for non-textual documents, hence the lower expected success rate. In short, these values reflect some practical problems with the prototype and tests themselves, rather than any problems intrinsic to the approach.

Software Tools

For this study, Quill v0.8, as described in Chapter 5 was used. Another tool we used during these tests was the KB Analyser. This tool gives the researcher the means to perform arbitrary queries on Quill's KB and presents the documents identified by those queries in a simple way. Each query can act as a document filter and can be turned on and off at will. For instance, in Figure 6.6, the 69 documents displayed in the list on the right side of the window are those that satisfy the rules in the third set ("Time") but not the first two (unchecked in the queries list).

Although a set of pre-defined queries found to be useful exists, new queries can be entered and saved for future use in other interviews. Also, all necessary tools to convert human-readable information to the internal format with which they are stored in the KB are provided. It is the case of the date-conversion and text tokenization fields on the bottom-left corner.

This tool allows researchers to explore different scenarios and identify how many documents are described by a story.

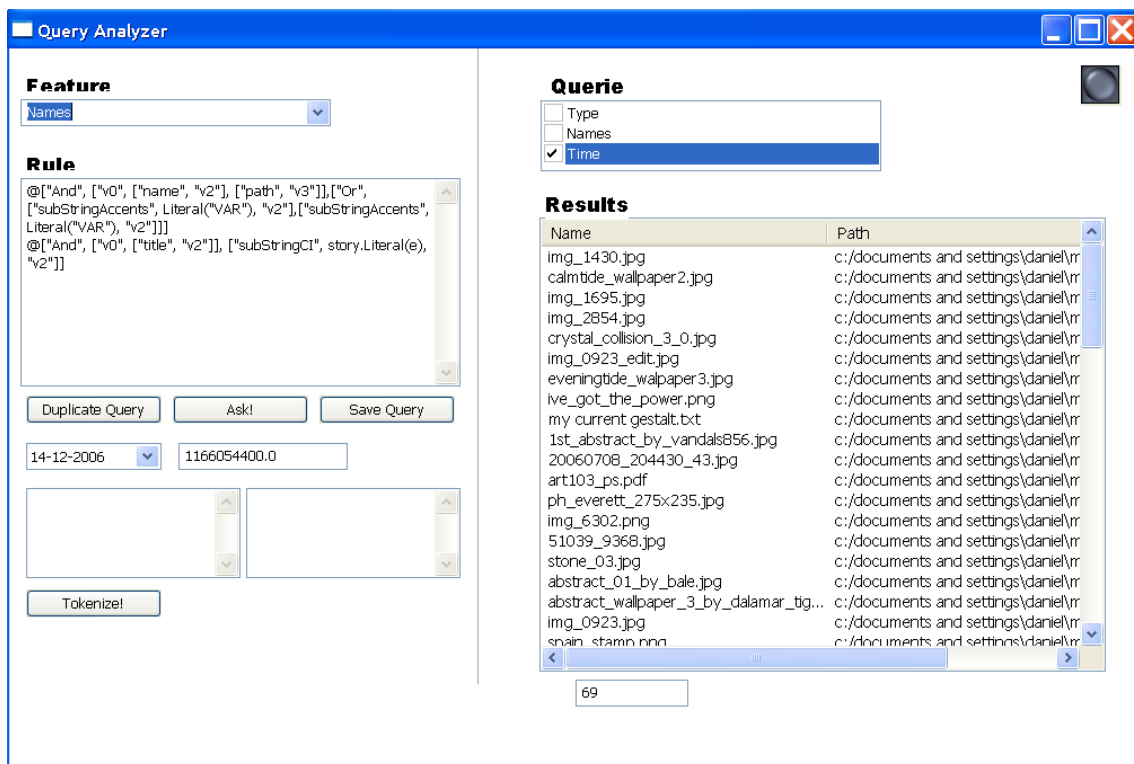


Figure 6.6: KB Analyser

6.2.2. User Profile

Overall, 21 users were interviewed. The time spent with each interview varied from one to five hours, depending not only on the interview itself but on the time it took to index the users' documents. Disregarding the indexing process, telling the stories, analysing their outcome and determining their discriminative power took between 30 minutes and one hour. Fifteen users (71%) were interviewed in just one session, while the documents of 6 others (29%) were indexed in a different session than the interview itself. Their ages ranged between 20 and 60 years old, with a mean of 36 years, and 11 were male and 10 female (52% vs 48%). As can be seen in Table 6.13, their backgrounds also varied greatly, from medical doctors to college students, passing through administrative staff and lawyers.

#	Profession	Schooling	Gender	Age
1	Retired Professor	Sci.Ed. PhD	M	60
2	Retired Professor	Educational Supervision MSc	F	33
3	PhD Student	Lic. Eng. Bio	F	27
4	Lawyer	Law School	M	28
5	Comp.Sci. Engineer	GIS MSc	F	34
6	Teacher	College Degree (Natural Sci.)	F	50
7	Teacher	College Degree (Maths.)	F	38
8	Administrative Staff	High School	F	37
9	Administrative Staff	High School	F	32
10	Pres. Exec. Council of EB2,3 School	MSc Management	F	58
11	Admin/Accountant	9th school year	M	54
12	Teacher (Phys.Ed.)	High School	M	57
13	Student/Call Center Operator	College Student	M	25
14	Teacher	College Degree	M	52
15	College Student (Comp. Sci.)	College Student	M	26
16	College Student (Comp. Sci.)	College Student	M	25
17	College Student (Comp. Sci.)	College Student	M	23
18	Medical Doctor	College Degree (Medicine)	F	26
19	College Student (Comp. Sci.)	College Student	M	20
20	College Student (Comp. Sci.)	College Student	F	23
21	Comp. Sci. Engineer	College Degree	M	35

Table 6.13: User Profiles

We are, thus, fairly confident that user profile does not influence our results, although there is a slight bias towards people with a computer science background. However, as these are, if anything, the harder test subjects (they use their computers more and have more documents), any such bias won't unduly benefit the results.

Table 6.14 shows that most interviews were performed at the user's workplace (52%), followed by those that occurred in the users' homes (37%). Two interviews (10%) took place elsewhere, namely a coffee shop and the subject's parents' home.

Workplace	11
Home	8
Other	2

Table 6.14: Interview Location

Conducting the Interviews

In the several interviews we asked the users to tell three different stories (for Recent, Old and documents of Other authors). Only four of the 21 users had any kind of previous contact with Quill, and even then this happened nearly two years ago for an early and somewhat different version of the prototype. Thus, no user was proficient with Quill, allowing for some learning factor to manifest itself. To prevent any undue bias to the data derived from that learning process, the order in which the users had to tell their stories varied, as summarised in Table 6.15.

	1st story	2nd story	3rd story
Recent	6	9	6
Old	9	7	5
Other	6	5	10

Table 6.15: Story order in the interviews

Even so, we did witness a learning curve, as depicted in Figure 6.7. At first, it would seem that the second stories took longer than the first and the third ones took less time, but given the high standard deviations (139.99, 137.93 and 82.91 for Recent, Old and Other documents, respectively), this might not be the case. Indeed, t-tests (with 95% significance) confirmed this, showing that the means cannot be unequivocally stated to be different. Therefore, while it seems that some learning could be in place, it is too slow to clearly reflect on the data. Maybe a fourth story would continue the descending trend. It should be mentioned that document-rich users might have taken longer to find their documents because of a small but noticeable performance difference when Quill handles large document collections.

While the times we just discussed were those required to find the target documents, the users were asked to tell complete stories, mentioning everything they remembered about the documents even if they had already been found. They were unable to inspect the documents to do so, continuing to rely on their memories. This allowed us to collect more complete sets of information regarding documents to see, in the worst case scenario, what could be used to find them.

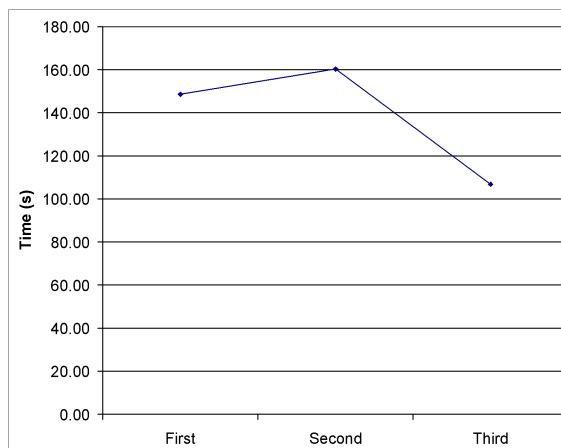


Figure 6.7: Average time per story type

Document Statistics

Despite the fact that the monitoring system was able to index the users' documents, emails and agenda, this didn't occur for every user. Table 6.16 shows that only five users (23%) had their email indexed. The remaining relied on on-line email services such as gmail, hotmail or sapo, that are not indexable by our software without downloading them through a POP3 connection. This would be impractical either because the interview took place in a location without internet access (laptop users) or because of time constraints (users at the workplace). Regarding the agendas, it seems they are not popular among regular users: only one used her Microsoft Outlook agenda.

The lack of email and agenda information might have had some impact on the retrieval rate of stories, but as we will discuss later, not a sufficiently large one to compromise the results.

Email		Agenda	
Yes	5	Yes	1
No	16	No	20

Table 6.16: Information Sources

We also collected some statistics on documents found on the users' computers. The number of documents varied greatly from user to user, as depicted in Figure 6.8. The mean is 7741 documents but the standard deviation is predictably large: 10762. Even so, the graphic easily shows that most users have only a few thousand documents to concern themselves with. As reported in a previous study [Gonçalves 2002], there seem to be three kinds of users: "document-rich", with tens of thousands of documents, "document-poor" with just a few hundreds, up to four or five thousand, and those in between.

It is interesting to note that of the three file-rich users, two had a computer science back-

ground but the other (user four) is a lawyer.

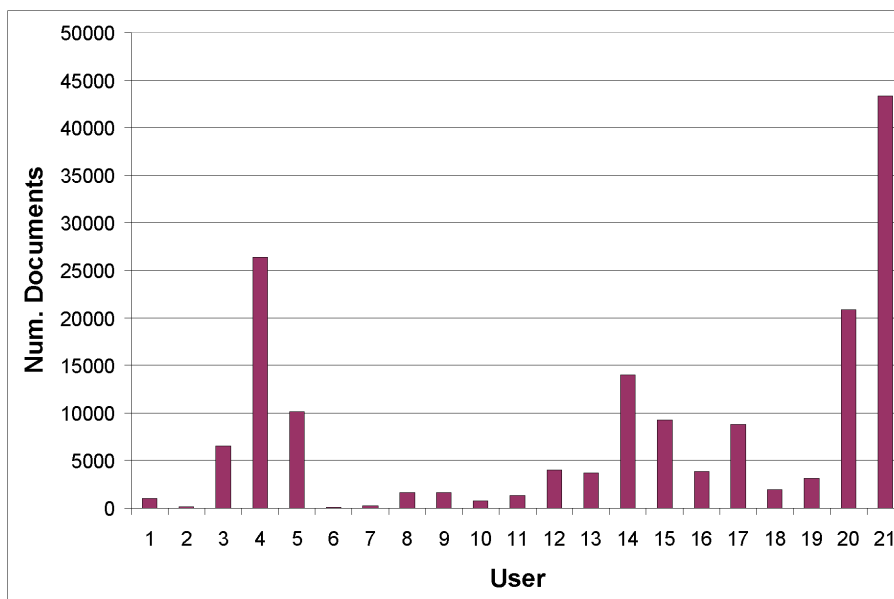


Figure 6.8: Number of documents per user

The number of folders or directories each user had to organise their documents in also varied greatly, from 14 to 8819. The average is 893 (std. dev. 1947) and the median 188, showing that most users rely on a fairly low number of folders. Furthermore, there is little correlation between the number of folders and that of documents (a correlation factor of 0.34). This hints at different degrees of “neatness” when storing documents. This is confirmed by the average number of files per folder, depicted in Figure 6.9.

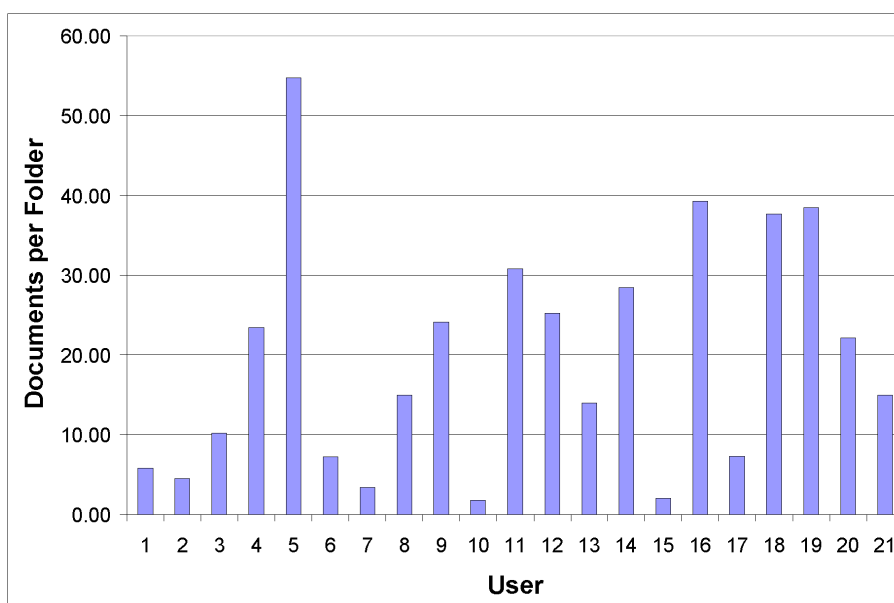


Figure 6.9: Average number of documents per folder

Regarding the types of documents found in the users’ computers, we see that all users

possess text, image and spreadsheet documents. Presentation, video and sound (mainly song files) documents were also popular. For nearly all users we found files of unknown types. This is due, for the most part, to application-related data files that are stored in folders where users also store their documents (e.g.: “My Documents”).

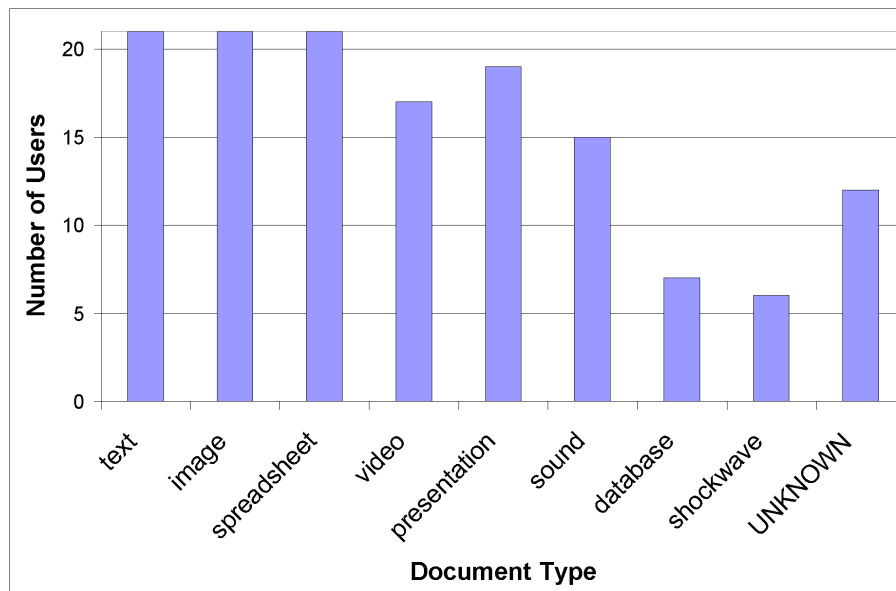


Figure 6.10: Users per document type

Also interesting is to note that text documents (mostly Microsoft Word and PDF files) and images are still the most important ones (Figure 6.11). However, unlike we found in a previous 2002 study [Gonçalves 2002], images are now the most common type of document. Indeed, the advent of digital photography has caused even casual users to produce large numbers of images and store them on their computers. There was some variability from user to user, of course, but the most important aspect to note is that 14 of the 21 users had more images than text documents. Sound files have also become more preponderant in recent years, now amounting for 43% of text documents. In 2002, sound documents were just 1,7% of text documents. Furthermore, now 71% of users have them, instead of the 55% in 2002. Video documents are not as frequent, but nevertheless 81% of users now possess at least one such document, whereas in 2002 only 55% did.

For the five users that had their emails indexed, we found the numbers shown in Table 6.17. User 1 had only 75 mails on his email client application, and User 15 had 8289. In all cases we found mostly incoming email, as was to be expected given that average users receive far more email messages than they write.

We also collected statistics on the persons known by the users, referred to in email messages and as authors of documents. The monitoring system uses a set of heuristics to associate the several emails and aliases a person has so that, for instance, both dan@gmail.com and djvg@gmail.com can be recognised as the same person, Daniel Gonçalves. The number of people with which users correspond varies from 42 to 1932 (Table 6.18). More

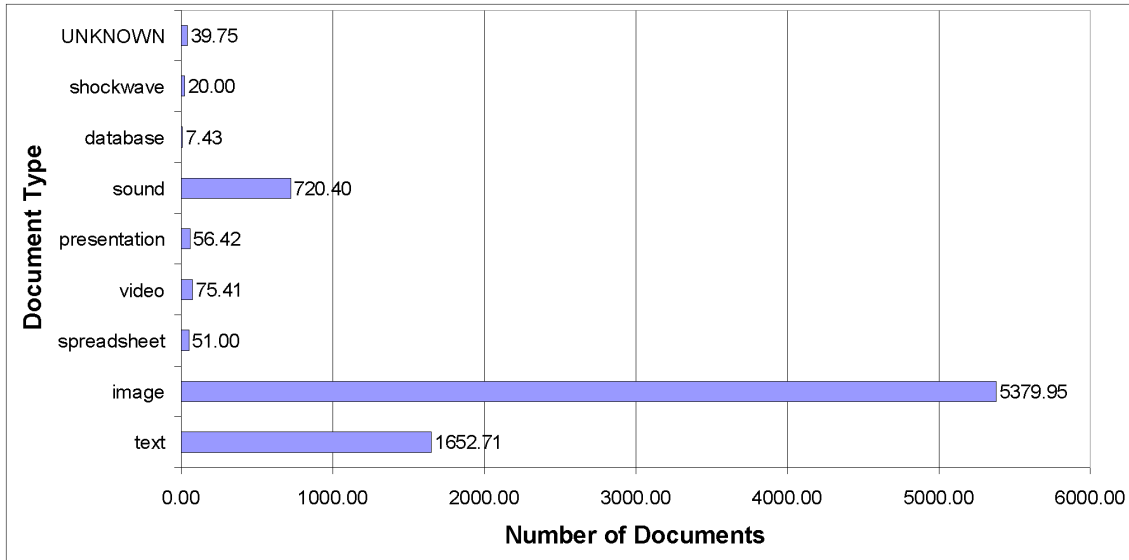


Figure 6.11: Average number of documents by document type

User	Total	Incoming	Outgoing
1	75	75	0
3	4323	4202	121
10	762	753	9
15	8289	8209	80
21	1671	1670	1
Average	3024	2941.8	82.2
Std. Dev.	3356	3297.5	137.6

Table 6.17: Indexed email statistics

interestingly, on average each person has close to 2 different email addresses (1.77) and can be referred to by 1.68 names per each address, in email headers (“Daniel Gonçalves” and “Daniel J. Gonçalves” can both appear as names for djvg@gmail.com, for instance).

6.2.3. Discriminative Power of Stories

To measure the discriminative power of stories, after each story was told, we used the KB Analyzer tool to find exactly how many documents indexed by Quill the story described. For instance, a story might state that “the name of the document was ‘foo’ ” and, while the target document was, indeed, ‘foo.doc’, others like ‘foo.xls’ and ‘foo_bar.doc’ could also exist (this element would have a discriminative power of three).

This task could become quite complex, but it was not impossible to do. In the first place, we used KB Analyzer to find documents that satisfied concrete information in the story. That is the case of a document’s filename, the date in which it was created or read, and

User	# Participants	Mail/Participant	Name/Participant
1	42	2.45	3
3	92	2.78	1.67
10	1932	1.02	1.03
15	1103	1.22	1.28
21	507	1.38	1.43
Average	735.2	1.77	1.68
Std. Dev.	792.77	0.79	0.77

Table 6.18: Indexed people statistics

the document type. Those restrictions were often enough to reduce the size of the set of possible documents to a manageable number, in the dozens instead of hundreds or thousands. Then, with the help of the users, the list of candidate documents would be further reduced. For instance, it would be possible, for some users, to look at some documents and immediately know their subject wasn't the same of the document described in the story. Sometimes, it would be necessary to open documents to verify their contents. Erring on the side of caution, this was done whenever the researcher needed some clarification, or even when the user seemed confident about eliminating some document from the list but failed to properly explain why.

There were cases in which the information in stories was not correct. This was to be expected, as it has been showed that between 9% and 19% of the information in a story could be wrong. We dealt with this using KB Analyzer's capability to study different scenarios, by turning on and off different constraints at will (as described in Section 6.2.1). If a given constraint proved to be incorrect, it would be turned off. Sometimes, given that for most cases the target document had been found, the researcher judiciously avoided entering obvious mistakes into KB Analyzer.

There were also a few cases when the users described documents not present in their computers. When it happened, and no other document was correctly described by the story, we considered the discriminative power to be zero. This happened for six of the 63 stories. Those values were not considered in the analysis that will ensue.

Finally, there were cases (five times) where the user was not looking for a single specific document, but rather a set of documents: a photo album. For instance, User 4 was looking for his "summer vacation photos". As any of those photos would satisfy the user, we considered the entire album (the folder where the photos were stored) to be a single document, while noting how many photos were in it (176, 14, 52, 118 and 15, in the five cases we found).

Overall, we found document-describing stories to have, on average, a *discriminative power of 2.51*. Although the standard deviation is fairly large, at 4.75, calculating the

confidence interval shows the mean to be between 2.13 and 2.88 with 95% confidence. *This value is well below the pre-established goal of 5.* It shows how stories can be used to identify, if not univocally then nearly so, personal documents. Three documents can be easily handled and told apart by users. Even if that wasn't the case opening a maximum of three documents to find the correct one is still reasonable. Finally, this shows that a document-retrieval interface need not display large numbers of candidate documents to the users, if it correctly uses the information in stories to find them.

Furthermore, it should be noted that, while the mean for the discriminative power is 2.51, the median and mode are 1.0, showing that, for the most part, stories identify single documents. In fact, 35 stories (55.5% of all stories told) did so. Figure 6.12 gives a better idea of this.

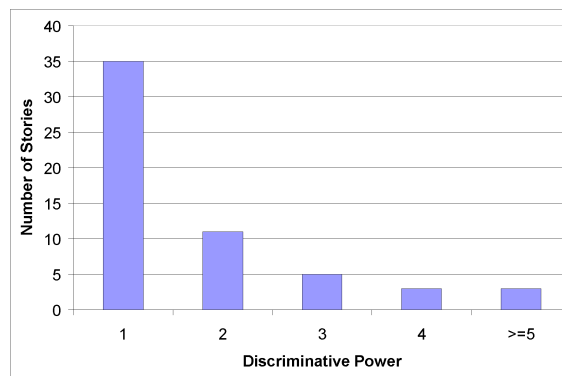


Figure 6.12: Discriminative power of stories (overall)

	Recent		Old		Other		Overall	
1	15	75%	11	61%	9	47%	35	61%
2	3	15%	3	17%	5	26%	11	19%
3	1	5%	2	11%	2	11%	5	9%
4	1	5%	0	0%	2	11%	3	5%
5	0	0%	0	0%	0	0%	0	0%
6-7	0	0%	1	6%	0	0%	1	2%
8-10	0	0%	0	0%	0	0%	0	0%
> 10	0	0%	1	6%	1	5%	2	4%
Total	20	100%	18	100%	19	100%	57	100%

Table 6.19: Discriminative power of stories

Table 6.19 shows how, for the most part, stories identify small numbers of documents. An overall 4% of stories identifies 10 or more documents. Those two stories described 21 and 31 documents, a text document (of several similar ones the user created at regular intervals) and a specific photograph (the story described all photographs taken at the same time).

It is also interesting to note that the discriminative power seems lower for Old docu-

ments when compared to Recent ones, and lower still if we consider Other documents (Table 6.20). Again, t-tests did not find the differences to be statistically significant. However, such differences were to be expected, as previous studies showed that while not necessarily more incorrect, stories for Old and Other documents tend to be more imprecise than those for Recent documents. For instance, while for a Recent document a user might state it was made last week, for an Old document he might only remembered it was created around summer 2004. Confirming that is the fact that the stories' discriminative powers vary far more for Old and Other documents than for Recent ones. This indicates that while the large majority of stories about Recent documents have discriminative powers of 1 or two, stories for Old and Other documents, being more imprecise, can sometimes discriminate far less between documents. Figure 6.13 clearly reveals this.

	Recent	Old	Other
Average	1.4	2.83	3.37
Std. Dev.	0.82	4.77	6.77
Number of Stories	21	21	21

Table 6.20: Discriminative power by document type

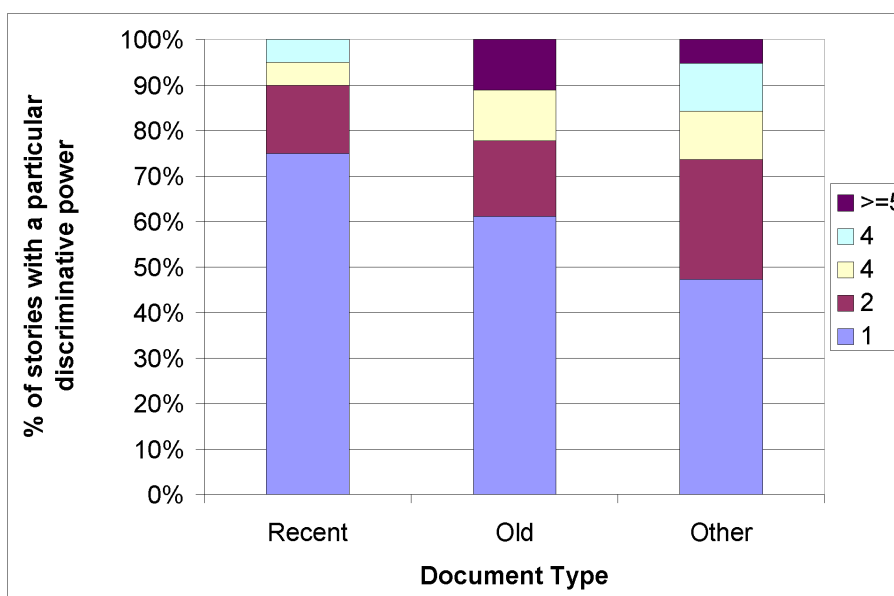


Figure 6.13: Discriminative power by document type

6.2.4. Retrieval Rate

To measure the retrieval rate of stories, we started by looking at the retrieval success for each individual narrative. As stated before, four outcomes were considered:

- The document was found

- The document was found “one click away”
- The document wasn’t found due to some mistake in the story
- The story was correct but nevertheless the document wasn’t found

Of these four outcomes, we considered the first two a success and the former two a failure, as Quill should be good enough to deal with some imprecision in stories. Note that although 21 stories were told for each document type, there were cases (5 in total) where users described documents that would never be found, as they were not actually present in their computers. They had been deleted sometime in the past or were stored somewhere else. This is the price to pay for dealing with real document collections that change over time, over which we don’t have control. We did not consider those stories successes or failures, simply removing them from the study that follows.

Retrieval Success Rates

As Table 6.21 shows, most documents were found with no problems. In fact, *the overall retrieval success rate was 87.9%*. Of 58 documents, 51 were found with no problems. While 81.0% of documents appeared in Quill’s document suggestion area, 6.9% (4 documents) were found “one click away”: the users recognized a document they remembered to be stored in the same folder as the one they were seeking, and opened that folder using Quill’s context menu, promptly finding the target document. It is arguable that as the story became longer, that document would eventually appear in Quill.

	Recent		Old		Other		Overall	
Total Present	20		19		19		58	
Found	19	95.00%	13	68.40%	15	79.00%	47	81.00%
1-Click Away	0	0.00%	3	15.80%	1	5.30%	4	6.90%
Story Mistake	1	5.00%	2	10.50%	2	10.50%	5	8.60%
Not Found	0	0.00%	1	5.30%	1	5.30%	2	3.50%
Not there...	1	5.00%	2	10.50%	2	10.50%	5	8.60%

Table 6.21: Story classification regarding document retrieval

The five documents that were not found due to mistakes in the stories that described them were non-textual documents. In Quill’s current version, the amount of information that can be used to identify non-textual documents is limited if no email or agenda data has been indexed. This was the case for Users 4, 13, 17, 18 and 19, who told those mistaken stories. In all cases, information that Quill could use, such as when the document was handled or its file or pathname were told incorrectly. This made it impossible for Quill to find the document, resorting to what little correct information was left. Wrong Time information seems to have played an important role here. As one of the most important elements, it was given a slight preponderance over the others. It would not cause

any problems when other elements were correct, but in this case it might have proven counterproductive.

Finally, two documents were not found even when the information in the story was correct. Both were text-based documents and, after close inspection, two major factors contributed to their non-discovery. Firstly, Quill does not perform full text indexing on the document. Instead, it uses the tf-idf algorithm [Salton 1988] to extract relevant keywords. One of the documents that were not found was described by User 8 as a recipe for a carrot cake. However, after close inspection, we found that recipe to be one page of a 23-page document where the other pages were official letters of the institution the user worked in. As such, keywords in the recipe itself, only a small part of the document, were not deemed important enough by tf-idf. In the second place, the lack of other information sources besides the documents themselves and email might have compromised the retrieval success. User 10 described the second document Quill was unable to find, her curriculum vitae. There was information in the story that Quill was unable to use to identify the document but was, nevertheless, correct. A future version of Quill, that can make better use of the information in stories, or a usage situation in which Quill has been running for an extended period of time on the users' computer would probably have no trouble with that document. The approach in itself remains unscathed, as both documents were not found due to limitations of the prototype.

	Recent	Old	Other	Overall
Overall Success Rate:	95.00%	84.20%	84.20%	87.90%
Retrieval (Text):	100.00%	92.30%	93.30%	95.20%
Retrieval (Non-Text):	83.30%	66.70%	50.00%	68.80%

Table 6.22: Textual and Non-Textual Retrieval Rate

Data was also collected regarding the retrieval success rates for text-based and non-text-based documents. We considered text-based documents all those from which keywords could be extracted (Microsoft Word, Excel, Powerpoint, PDF, plain text, etc.). As Table 6.22 shows, *Quill was able to retrieve 95.2% of all text-based documents*, for a total of 40 out of 42. The two that weren't found are those discussed above, not found even when the stories were correct. As for non text-based documents, the *retrieval success rate drops to 68.8%* (11 out of 16) that, while not as good, is still perfectly acceptable, considering that the amount of autobiographic information available to Quill in these tests was limited. Also, those are documents (mostly images and songs) that traditional retrieval tools have trouble finding, unless they are specialised tools for specific document types. The five non text-based documents that were classified as unsuccessful were those discussed above: information in the stories describing them was wrong, and the remaining correct information was not sufficient to overshadow the mistake.

It can also be noted that Old documents are harder to find than Recent ones, and Other

documents are even harder. This is not unexpected and again related to the inexactitude and inaccuracy of stories for those kinds of documents, as previously discussed. Even so, we find overall success rates of 84.2% for those documents, with non-textual documents suffering more with only 66.7% and 50% rates for Old and Other documents respectively. Even so, at no time did the values fall below our pre-defined goals of 75% for text-based documents and 50% for non-text based documents.

Story Length and Suggestion Display

While the users were told to continue telling their stories even after the target document was found, we noted when this took place. This allows us to estimate, in a real usage situation, how long would the users' stories have to be before their documents are found.

We found that, on average, stories have to be 5.8 elements long (with a 4.46 standard deviation) before a document is found. A 95% confidence interval places the universe's mean between 5.60 and 6.01. As users become more knowledgeable about Quill and, with continued use, it adapts to their preferred element orders, those can be the first six story elements to be suggested to the users, greatly increasing the retrieval performance. Note that the target document might actually be visible in the document suggestion area before what the values reveal (indeed, in some cases, it was), as we recorded the moment when the users noticed it there.

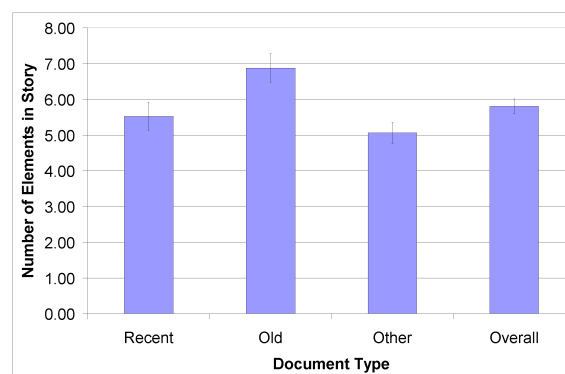


Figure 6.14: Story length per document type

Figure 6.14 summarizes the results (the error bars depict the 95% confidence interval). While it seems that stories for Old documents are around one element longer than those for other document types, no statistically significant difference was found ($\alpha=0.05$).

Finally, we also recorded in which position of Quill's document suggestion area did the target documents appear. In Figure 6.15 we can see that, on average, all documents appeared among the four first suggestions. This is an excellent result because, as we have already discussed, because it does not unduly distract users from the stories and the interface can easily accommodate five documents at once in the document suggestion area

(due to screen resolution limitations). This, coupled with the discriminative power of 2.5, gives full confidence in terms of making the target documents visible and recognizable to the users.

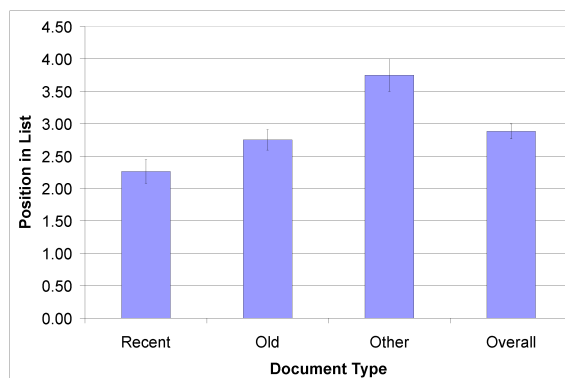


Figure 6.15: Target Document position in suggestion list

The Need for Autobiographic Information

We also collected data to show how effective was non-textual information entered in the stories in identifying promising documents. We noted, for every document that was found using Quill, if text-based information told by the users in their stories (keywords or expressions that might occur in a document) would correctly identify the document. Quill doesn't ask directly for keywords. We considered keywords to be everything the user entered as the Subject, Purpose or Contents of a document. We also noted for documents for which no keywords were mentioned (or mentioned erroneously) if they would have been correctly identified by their file or pathname, or by any other information in the stories.

Note that a document being identified using keywords doesn't necessarily mean that keywords would suffice to find it. It means that some relevant keywords were correctly remembered by the user. It might be the case that all documents with those keywords in them might still be too many to correctly identify the one being sought. Again erring on the side of caution, we'll consider that a document that *might* have been correctly identified using keywords *could* have been found without resorting to further information.

Table 6.23 summarizes our findings. It can be readily seen that, while 64.7% of all documents described by users could have been identified by the keywords mentioned in them, this wasn't true for the remaining 28 documents. However, for *all* documents the stories describing them contained some other correct information, besides keywords (dates, tasks, types, etc.). This means that *traditional approaches based on the search of keywords in documents would fail to find up to 35.29% of documents found by Quill!*

Of the 35.29% of documents that would have not been found using keyword search,

Total number of documents found	51	
Identified by keywords	33	64.71%
Identified by other info	51	100.00%
Identified by keywords ONLY	0	0.00%
Not Identified by kws but identified by name	14	27.45%
Identified using other info (incl. name)	18	35.29%
Identified using other info ONLY	4	7.84%

Table 6.23: Story elements that correctly identified documents

27.45% would have been identified by information regarding their file or pathnames. Even then, the remaining 7.84% of documents would not be identified.

In short, even approaches that make use of a document's content and its filename to find (such as Google Desktop) *would not find 7.84% of the documents found by Quill*. This, of course, considering that just keywords or pathname would have a high enough discriminative power to be able to bring the right document to the user's attention, which might not be the case.

Looking at the results by document type (Table 6.24), we don't find significant differences between them in terms of being correctly identified by keywords or pathnames.

	Recent	Old	Other
Total number of documents found	19	16	16
Identified by keywords	11	10	12
Identified by other info	19	16	16
Identified by keywords ONLY	0	0	0
Not Identified by kws but identified by name	6	5	3
Identified using other info (incl. name)	8	6	4
Identified using other info ONLY	2	1	1

Table 6.24: Story elements that correctly identified documents, by document type

6.2.5. User Reactions

Apart from all the quantitative data collected that will allow us to answer our research questions, a fair amount of qualitative data was also gathered by the interviewer from the direct observation of the users' actions to their comments. Furthermore, the users were asked, at the end of the interviews, to answer a small satisfaction questionnaire, where they could give us their suggestions on how to improve Quill. We'll now discuss that information.

The questionnaire results can be found in Figure 6.16. The error bars depict the confi-

dence intervals for the mean (for 95% confidence). Each of the six statements could be rated on a scale ranging from 1 – Completely Disagree, to 4 – Completely agree. An even number of values was used to prevent the users from not compromising. The six statements were:

- Q1 – It is easy to tell stories about documents using the system
- Q2 – The retrieval of documents was satisfactory
- Q3 – I managed to tell everything I remembered about the documents
- Q4 – The system is different from other search mechanisms
- Q5 – The system lets me tell stories how I like to tell them
- Q6 – The system was easy to understand and use

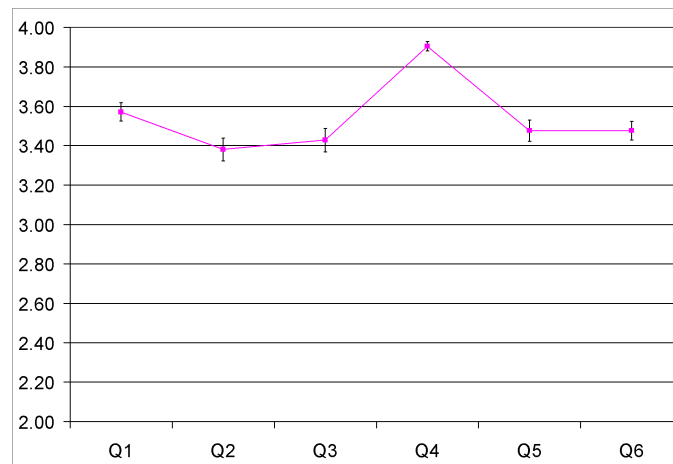


Figure 6.16: Questionnaire Answers

As the graphic shows, the results were well above 3, showing the users had a very favorable opinion of Quill. This is particularly noticeable in terms of Quill's novelty, that merited an average of 3.9. In all cases the mode was 4.

In fact, the users were so pleased with the system that some actually used it to find documents they needed around the time of the interview and had eluded them for a couple of days before. Three users requested Quill to remain installed in their machines, as they greatly appreciated its value. As of the writing of this document (five months later), at least one user still uses Quill regularly to retrieve her documents.

As for user comments, apart from praise for Quill, the most relevant were:

- Complaining that the Appearance story element should allow for other possible visual appearances of documents (two users).

- The need for the ability to specify time intervals, instead of time instants, although, unbeknownst to the users, Quill already deals with intervals internally and does not reveal it to prevent undue confusion on novice users.
- Other information regarding image documents, such as size, resolution, etc. A further study would be necessary to verify to what extent that information would be useful, especially for non-tech savvy users.
- The ability to emphasize some story elements, giving them a certainty rating.
- Some users complained that they had to tell repeated information in their stories. This was to be expected, and results from the fact that some people view a document's Purpose and Subject as being the same. Nevertheless, this is not true for all users. For instance, a document's subject can be "narrative-based interfaces" and its purpose a conference to which it is submitted. This problem can easily be minimized by Quill's ability to learn the users' preferred story structures, by moving one of those elements to the end of the story, where it might end up not being suggested to the user.

Besides these comments, the researchers also took notice of some other relevant things during the course of the interviews. First of all, it was witnessed that the dialogue for the Other Documents story element still causes some confusion on some users. Also, many users were unable to tell the difference between different types of text documents. For them, text documents are Microsoft Word documents and vice-versa. This, of course, happened with less experienced users. Also, we sometimes verified that the target document was already visible in the document suggestion area for one or two story elements before the users noticed it. The usual *modus operandi* was to tell the story without interruptions and only when a story train had been completed would they look at the suggestion area to see if any result was there. This shows that that area does not unduly distract users, but also that some other visual cue might be required to show the users that something has changed.

6.3. Discussion

With the studies described in this chapter we aimed at providing answers to four research questions and proving narratives can be an adequate way for users to convey autobiographic information to computers to retrieve personal documents.

The first question addressed the most basic requisite of narrative-based interfaces: *can users tell stories to the computer in a natural way?* If the storytelling process is not simple or straightforward, a narrative-based interface could become awkward or difficult to use. To

answer this question we collected stories told to computers, using the Quill prototype, and compared them to those told using a low-fidelity interface. We verified that the stories' structure remains the same: the stories have the same length, and their elements appear in the same order. Regarding their contents, all elements have similar frequencies, except Name that appears somewhat more frequently. In short, *there is no significant difference between the stories told to the low-fidelity prototype (with a human's help), and those told to an actual functioning system..* This allows us to conclude that the first question can be positively answered.

Our second concern was *story trustworthiness*. While users can tell their stories to Quill, the information therein must be accurate and correct, in order for it to be usable to find the desired document. We found that, *for the most part, we can believe in the users*. We strove to verify beyond any (reasonable) doubt the accuracy of story elements. We managed to do so for 81% of elements, and verified that between 73% and 92% of story elements can be trusted to be accurate. This means that 1 to 5 elements of any given story might be inaccurate. If some adjustments are made, taking into account the nature of some story elements, we can expect as much as 81% to 91% to be correct (1 to 3 inaccurate elements per story). Also, we verified that inaccurate elements are not necessarily wrong, often being quite close to the exact values they should take. As Quill takes no information element told by the user as an absolute truth but, rather, it just assigns it a certain merit value in relation to the other information in the story, such a low number of inaccurate elements won't unduly hinder the retrieval process (as shown in the second study described above).

Having shown that stories can be told to computers and have in them information that can be used to retrieve personal documents, we set out to establish if stories were discriminative enough to identify the target documents in a way in which a manageable amount of candidates could be presented to the users. We verified that, indeed, document-describing *stories have a discriminative power of 2.51*, meaning that between two and three documents at most are identified by them. Actually, *in most cases stories identify only one document*. Also, stories about older documents and those of other authors seem to be less discriminative, as was to be expected from the slightly less precision of those stories. Even so, in the worst case, stories will have a discriminative power of 3.37.

All that remained was to see if in practice, even with the quality, trustworthiness and discrimination criteria satisfied, a narrative-based interface can be successfully used to retrieve personal documents. To validate the approach as a whole, Quill's retrieval success rate was measured and *found to be 87.9%*. In fact, *text documents were successfully retrieved 95.2% of times, and non-textual documents (images, songs, etc.) 68.8%*. These are good results, and well above the pre-established goals of 75% and 50% for textual and non-textual documents, respectively.

We were also able to verify that stories need not be longer than six or seven elements, and

that target documents usually show in the first four positions of the document suggestion area. As the documents therein are those best ranked by Quill in light of the story being told, this shows that the ranking process is well adjusted. Finally, users qualitatively showed their appreciation of Quill.

A doubt remained on whether Quill and narrative-based document retrieval would perform better than traditional keyword-based approaches and applications such as Google Desktop. We verified that of the documents *Quill was able to retrieve 35.3% would not have been found using keyword search alone*. Also, even if filenames were used as criteria, *still 7.8% of documents found by Quill would remain unfound* by other systems. Quill, by combining a wide range of autobiographic information, performs better than most existing systems.

In short, we were able to provide positive answers to the four research questions. Not only were the intrinsic qualities of narratives about documents (trustworthiness and discriminative power) proven, but we also showed that it is possible to successfully use them to retrieve personal documents.

7

Conclusions and Future Work

In this chapter, we will conclude this dissertation. We will start by summarising the major contributions it describes. As no work is perfect, we will mention our research's weaknesses that, while not invalidating any of the results we reached, could merit some attention as future work. A description of other works that had their origin in the one here described will follow. We will then conclude by pointing at possible avenues to follow in future work.

7.1. Dissertation Summary

Our motivation for the undertaking of our research was to help users organise and retrieve the growing numbers of documents users must daily contend with. Furthermore, retrieving personal documents is not the same as retrieving documents for the World-Wide-Web or digital libraries. Users and their documents share a context, as those documents have been handled before, in specific situations. Therefore, information users might recall about that context would undoubtedly help them retrieve documents, providing the computer is able to understand it. This raised another issue: how to structure that information so that users are able to convey it in a natural and effective way? We

proposed that narratives might provide such a way. By resorting to our ability to tell stories, and taking into account how they inter-relate the information therein, it would become easier for users to recall and express relevant autobiographic data about their documents and the context in which they were handled.

Based on those considerations, we defined our main research goal: to study narratives as the basis for the creation of interfaces that allow users to effectively retrieve their personal documents. In order for this goal to be reached, we needed to establish what narratives about documents are like and then find the appropriate shape an interface should have to allow users to tell them to their computers. Once such an interface was designed, we were able to study the intrinsic qualities of narratives regarding document retrieval: their trustworthiness, discriminative power, and the ability to actually retrieve documents using a narrative-based interface.

Those goals were achieved, allowing our approach to go beyond the state-of-the-art, as it overcomes most of the limitations found in current alternative systems. Unlike spatially-based approaches, it is not limited in the numbers of documents it can handle. It makes use of a wide range of autobiographic information about the users and their documents, considering all equally important. This allows different users to retrieve their documents resorting to whatever information they more easily recall, instead of having to rely primarily on one type of information, as is the case of Time-based solutions. Finally, unlike property-based and tagging systems, narrative interfaces provide an underlying organising principle for the relevant autobiographic information, by allowing users to convey it as stories. This obviates the problem of having to remember arbitrary properties and possible values. Unlike tagging, no effort is required from the users when organising their documents, as their retrieval does not depend on the entry of arbitrary keywords that are difficult to recall and use consistently in the classifications of different documents.

To arrive at the aforementioned results, it was necessary to understand what can be found in them. Thus, we interviewed over 20 users with different backgrounds and collected 60 stories about documents. Three different document kinds were considered: documents created recently by the users, documents they created over six months ago, and documents not created by them. Stories were subject to content and relational analysis, which provided us with a thorough characterisation of those narratives. We identified the different elements that might appear in stories about documents (Time, Place, Co-Authors, Purpose, Author, Subject, Other Documents, Personal Life Events, World Events, Exchanges, Type, Tasks, Storage, Versions, Contents, Events, and Name), noting which were most likely to appear and which were easier to remember. We verified that stories can possess between 12 and 19 such elements. Also, the probabilities in which the elements might succeed each other in stories were computed.

Based on this characterisation, we trained Hidden Markov Models to infer archetypical story structures, and devised a set of guidelines for the creation of narrative-based docu-

ment retrieval interfaces.

Those guidelines, while useful, were not enough to determine the actual shape of an interface. To find it, we performed a set of user studies evaluating two different low-fidelity prototypes created based on the guidelines. A new set of 60 stories was collected compared to the ones gathered in the previous study. This allowed us to conclude that the prototype in which the stories are written textually was superior to the one based on the direct manipulation of graphically-represented story elements. Stories told using the first prototype were almost identical to those told to humans, which was not the case for the second one, showing the first one is less intrusive.

Having chosen an interface, we had to implement it in order to perform further user studies. The resulting system, Quill was implemented based on the better low-fidelity prototype, and suffered some modifications based on user reactions throughout the different testing rounds. Underlying the interface, there is a Knowledge Base (KB) in which all autobiographic information can be stored. This KB is based on the RDF formalism, chosen after a close scrutiny of possible alternatives. However, as RDF can be cumbersome to deal with, we created an abstraction layer, dubbed Scroll, that implements high-level semantic network-like primitives that allow us to interact with RDF in effective but meaningful ways. The information is fed into the KB by an auxiliary monitoring system, that continuously monitors different information sources and updates the KB. Its implementation was challenging, having to cope with the huge variety of applications and configurations we might find in the users' computers, but as it is plugin-based, it can easily accommodate different information sources.

In the possession of Quill, we were finally able to perform user tests to confirm narratives as an organising principle for the conveyance of autobiographic data to the computer. We conducted two studies in which we were able to answer all outstanding questions. We verified that stories told using Quill are similar to those told to humans, as expected given that Quill directly derives from a low-fidelity prototype for which this held true as well. We established that between 81% and 91% of all information in stories is correct. Furthermore, stories about documents have an average discriminative power of 2.51, well within the number of elements that can be suggested to a user to choose from without incurring in high cognitive loads and distractions. Furthermore, we verified that the users were able to retrieve 95.2% of all their text documents and 68.8% of non-textual documents. Even more impressive, 35.29% of all documents found would not have been found resorting to keyword search alone, and 7.84% would not have been found even looking at a document's filename. Given the fact that Quill is just a prototype, these results are more than enough to show the approach it embodies, narrative-based document retrieval, to be a success.

7.2. Final Conclusions and Discussion

The results of our research help users manage and retrieve their documents in natural and effective ways. This was accomplished by recognising that narratives about documents can be a way to help users remember autobiographic information about those documents, and convey that information to the computer in meaningful ways. As those stories refer to previous interactions between the users and their documents, narrative-based interfaces solve a problem that is related but different to that of information retrieval in its purest form. We were not only concerned with the actual retrieval process but, more importantly, with the users' interaction with the retrieval system. Thus, it was necessary to involve the users in all steps of the research and develop techniques to allow them to tell their stories as naturally as possible. By doing so we were able to create a prototype of an interface that allows this to happen. That interface, Quill, was used by real users to retrieve real documents allowing us to reach the most important results of our thesis, proving that stories are accurate and have a high discriminative power. Overall, the document retrieval rates measured using our approach were more than enough to confirm its validity, satisfactorily proving the concepts it involves.

7.2.1. Benefits

Most existing document retrieval approaches are limited to indicia provided by the filesystem, such as a document's file and path-names. Our approach looks at documents in the same way users regard them, instead of forcing them to use conventions existing for historical reasons. For users, a personal document is much more than a file. It is something that was created in a certain context and that is intimately known.

Our approach overcomes the problems found when trying to retrieve a personal document by allowing users to resort to all autobiographic information they can remember. Furthermore, by using stories as an organising principle, it is able to elicit large amounts of relevant information from users at little cost. Not only are people natural-born storytellers, stories inter-relate their constituent elements. This helps stimulate the users' associative memories when trying to remember their documents. Furthermore, this idea was well received by users, that praised not only its usefulness and novelty, but, in some cases, asked for the prototype to remain installed on their machines, for personal use.

Narrative-based interfaces for document retrieval are superior to keyword-based search since they resort to non-textual information not amenable to keyword indexing. Also, while keyword-based search engines can be repeatedly queried in search of a document, each query is independent of others. The user must start anew every time, and any continuity must be provided by appropriately choosing search terms. Narrative-based

interfaces offer a way to convey all relevant information in a single coherent story, inter-relating it, and using it as a whole to identify promising documents.

7.2.2. Limitations

While we were able to prove that narrative-based interfaces for personal document retrieval are a valid way to recover documents, they are not without limitations.

Maybe the biggest limitation of our approach is that it applies solely to personal documents. As it assumes the users will tell stories about those documents, there must be a story to tell. This is true only for documents with which they have interacted in the past, preferably their own. While stories can be told about other documents, they can never be as rich and complete as stories about personal documents.

Another limitation is a practical one: in order to understand the users' stories and match them against possible documents, it is necessary to gather large amounts of information regarding the users and their activities. This might prove difficult in some contexts. In particular, it is hard to produce a system that is able to cope with the myriad of systems and applications that might be relevant. However, this is not unsurmountable, as we were able to show with our prototype.

Also, in order to understand events and other information that might be needed to understand the stories, it is necessary for the users' activities to have been monitored at the time they took place. Thus, only some time after the system has been installed might enough information be available to fully comprehend narratives. However, we've shown it is possible to circumvent this limitation, up to a point, by analysing data already residing in the users' hard drives.

A disadvantage of narrative-based interfaces when compared with traditional approaches is that it is more language-dependant. As stories are constructed, they should be displayed in the users' native language. Also, all information will probably be entered using it. This hinders the creation of general-purpose, universal, applications. More than a matter of translating the interface components, some of the underlying mechanisms must be changed. For instance, stemming and parsing should adapt to the users' language. We were able to show how to circumvent this limitation by making language a module of Quill that can be replaced at will. Still, localisation efforts will be harder than for traditional point-and-click solutions.

Finally, our approach is clearly more useful for documents that are hard to find. Many documents can be easily found using traditional approaches and, in those cases, our solution can be slower. Narrative-based interfaces excel for documents the user has lost track of.

7.3. Ramifications

Our research has served as a starting point for other works. Four different projects arose from it. They will be succinctly described in the next paragraphs.

The most important of the projects was *BIRD - Biographic Information for the Retrieval of Documents* (FCT POSC/EIA/59022/2004). BIRD is a National Science Foundation project spanning two-and-a-half years. Its goals coincide, to some extent, with those of our research: prove that autobiographic information can be the basis for the retrieval of documents. However, it goes further by exploring other ways to display and organise that information, besides narratives. Preliminary results show narratives to yield better results. Currently, work on the BIRD project is studying how to consider in an integrated and interrelated way personal documents residing in different computers, and retrieved seamlessly on any one of them (or even elsewhere!) with an extended version of Quill.

As a part of BIRD, Renata Marin performed her final course project on which tried to *bridge the gap between electronic and paper documents*. She tried to extend our ideas presented in this dissertation and include new sources of autobiographic information arising not only from the users' computers but from the real world. In particular, by using RFID technology, she was able to integrate paper documents in to the approach. Her work made it possible to record information about such documents in the Scroll Knowledge base, dealing with them in the same way as any other document. Furthermore, the establishment of a relationship between electronic and paper versions of the same document became possible. This enriched the retrieval process by allowing each of those versions to be retrieved from the other. Also, the existence or not of paper versions of documents, and the act of printing them became part of the autobiographic information that can be mentioned by users to retrieve their documents. Overall, Renata's work serves as a proof-of-concept for the extensibility of our approach, showing how different, heterogeneous sources of information collected from the environment surrounding the user can be used to enrich the narratives told to find documents.

Two other students, Luis and Carlos Rodrigues, also did their final course work, *Personal Biographer*, under the auspices of our research. Their project dealt with practical engineering issues regarding the automated monitoring of the users' actions and other available data sources. Their work was able to demonstrate that automated data collection can be performed, and that, while difficult, it can be achieved with appropriate development methodologies, creativity and with reasonable effort.

Finally, Daniel Gonçalves (an undergraduate student) and Luis Bilo, embraced our idea of using autobiographic information to retrieve documents in a marginal but related way. The same underlying architecture and knowledge base is used, but they tried to display the information therein in meaningful ways, constructing a browsing tool, *Magoo*, in

which documents can be found navigating through autobiographic dimensions mapped onto the screen using an “Outer Space” metaphor. For instance, differently detailed views appear as galaxies, solar systems, etc.

All these works are evidence not only of the interest generated by our research, but of its versatility and extensibility.

7.4. Future Research

The results of our research pave the way for further work in different areas, both extending our results and helping solve problems that, being marginal to our work, we did not address.

Multimodality

Storytelling is easily amenable to modalities other than text. While allowing the unconstrained telling of entire stories by voice, for instance, would be unpractical, given the current state of the art of the technologies involved, combining text entry with spoken words might make the storytelling process easier. The reverse, reading the story out loud as it is written, might also help users focus on the story and remember more information (or even recognise some mistake they might have committed). Another modality that might be useful are calligraphic interfaces. When describing the visual aspect of a document, a way to jot down a rough sketch of the remembered appearance and then matching it to existing documents would be interesting. Other modalities might likewise enrich the interaction experience.

Other Objects

Our work was focused on the retrieval of personal documents. However, our approach might easily translate for the retrieval of personal objects. In order for the user to be able to tell stories about documents, a previous interaction with that document, in a shared context, must have occurred some time in the past. This is true for other personal objects other than documents. While the kinds of stories told by users would be different, the overall approach remains valid. Furthermore, we have, in this dissertation, shown how to study narratives describing objects of a given domain (documents) and built an interface for the retrieval of those objects. The same methodology can be used for other domains.

Ubiquitous Computing

As proven by the “Bridging the Gap” work, it is possible to incorporate real-world information into the knowledge base and use that information to retrieve documents. As ubiquitous computing technology matures, it should be possible to resort to new information sources about the users and their environments, allowing for a better understanding of stories. Those new information sources would also help with the retrieval of real-world objects other than documents, as suggested in the previous paragraph.

Certainty Degree

We discovered, as described in Chapter 6, that information in stories is, for the most part, truthful and accurate. However, it is also true that, in concrete cases, users have an informal ideal of how certain they are about a given story element. A way to help the users convey that information to the interface in a simple way that does not distract them from the storytelling process might be of interest. This certainty levels could then be used to weigh the different documents identified by the information in each story element, producing better results.

Bibliography

- [Abney 1991] Steven P. Abney. Parsing by chunks. In Steven P. Robert C. Berwick and Carol Tenny, editors, *Principle-Based Parsing: Computation and Psycholinguistics*, pages 257–278. Kluwer Academic Publishers, 1991.
- [Abowd and Mynatt 2000] Gregory D. Abowd and Elizabeth D. Mynatt. Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1):29–58, 2000.
- [Abrams et al. 1998] David Abrams, Ronald Baecker and Mark H. Chignell. Information archiving with bookmarks: Personal web space construction and organization. In *Proceedings CHI'98*, pages 41–48. 1998.
- [Adar 1996] Eytan Adar. Haystack: A personal, intelligent, indexing system. 1996.
- [Adar et al. 1999] Eytan Adar, David Karger and Lynn Andrea Stein. Haystack: Per-user information environments. In *Proceedings CIKM 1999*. 1999.
- [Agarawala and Balakrishnan 2006] Anand Agarawala and Ravin Balakrishnan. Keepin' it real: pushing the desktop metaphor with physics, piles and the pen. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1283–1292. ACM Press, New York, NY, USA, 2006. ISBN 1-59593-372-7.
- [Baeza-Yates and Mecoli 1999] Ricardo Baeza-Yates and Claudio Mecoli. *DomainView: A Desktop Metaphor based on User Defined Domains*. Master's thesis, Dept. of Computer Science, Univ. of Chile, 1999.
- [Baeza-Yates and Terry Jones 1999] Ricardo A. Baeza-Yates and et al. Terry Jones. A new data model: Persistent attribute-centric objects. 1999.

- [Barreau and Nardi 1995] Deborah Barreau and Bonnie A. Nardi. Finding and reminding: file organization from the desktop. *ACM SIGCHI Bulletin*, 27(3):39–43, 1995.
- [Bergman et al. 2006] Ofer Bergman, Ruth Beyth-Marom and Rafi Nachmias. The project fragmentation problem in personal information management. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 271–274. ACM Press, New York, NY, USA, 2006. ISBN 1-59593-372-7.
- [Berners-Lee et al. 2001a] Tim Berners-Lee, J. Hendler and Ora Lassila. The semantic web. *Scientific American*, 2001a.
- [Berners-Lee et al. 2001b] Tim Berners-Lee, James Hendler and Ora Lassila. The semantic web. *Scientific American*, 2001b.
- [Bälter and Sidner 2002] Olle Bälter and Candace L Sidner. Bifrost inbox organizer: giving users control over the inbox. In *Proceedings of the second Nordic conference on Human-computer interaction*, pages 111–118. ACM Press, 2002. ISBN 1-58113-616-1.
- [Boardman and Sasse 2004] Richard Boardman and M. Angela Sasse. “stuff goes into the computer and doesn’t come out”: a cross-tool study of personal information management. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 583–590. ACM Press, New York, NY, USA, 2004. ISBN 1-58113-702-8.
- [Bowman et al. 1995] C. Mic Bowman, Peter B. Danzig, Darren R. Hardy, Udi Manber and Michael F. Schwartz. The Harvest information discovery and access system. *Computer Networks and ISDN Systems*, 28(1–2):119–125, 1995.
- [Bowman and John 1996] C. Mic Bowman and Ranjit John. The synopsis file system: From files to file objects. In *Proceedings of the Joint W3C/OMG Workshop on Distributed Objects and Mobile Code*. 1996.
- [Brachman 1977] R. J. Brachman. *A Structural Paradigm for Representing Knowledge*. Ph.D. thesis, Harvard University, 1977.
- [Brachman 2004] R. J. Brachman. Owl web ontology language overview. 2004. <http://www.w3.org/TR/owl-features/>.
- [Brachman and Schmolze 1985] R. J. Brachman and J. G. Schmolze. An overview of the kl-one knowledge representation system. *Cognitive Science*, 9(2):171–216, 1985.
- [Brickley and Guha 2004] D. Brickley and R.V. Guha. Resource description framework (rdf) schema specification, w3c proposed recommendation. 2004. <http://www.w3.org/TR/PR-rdf-schema>.
- [Brown 1991] Donald E. Brown. *Human Universals*. McGraw-Hill, New York, 1991.

- [Chaudhri et al. 1997] V. Chaudhri, A. Farquhar and R. Fikes et al. The generic frame protocol 2.0. 1997.
- [Chirita et al. 2005] P. Chirita, R. Gavriloaie, S. Ghita, W. Nejdl and R. Paiu. Activity based metadata for semantic desktop search. In *Proceedings of 2nd ESWC*. 2005.
- [Corcho and Gómez-Pérez 2000] O. Corcho and A. Gómez-Pérez. A roadmap to ontology specification languages. In *12th Int'l Conf. Knowledge Eng. and Knowledge Management, Lecture Notes in Artificial Intelligence*, pages 80–96. Springer-Verlag, Berlin, 2000.
- [Cudicio 1986] Catherine Cudicio. *Comprendre la PNL: La programmation neurolinguiste outil de communication*. Les éditions d'organisation, 1986.
- [Cutrell and Dumais 2006] Edward Cutrell and Susan T. Dumais. Exploring personal information. *Communications of the ACM*, 49(4):50–51, 2006.
- [Cutrell et al. 2006a] Edward Cutrell, Susan T. Dumais and Jaime Teevan. Searching to eliminate personal information management. *Commun. ACM*, 49(1):58–64, 2006a.
- [Cutrell et al. 2006b] Edward Cutrell, Daniel Robbins, Susan Dumais and Raman Sarin. Fast, flexible filtering with phlat. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 261–270. ACM Press, New York, NY, USA, 2006b. ISBN 1-59593-372-7.
- [David Huynh and Quan 2002] David Karger David Huynh and Dennis Quan. Haystack: A platform for creating, organizing and visualizing information using RDF. In *Semantic Web Workshop, The Eleventh World Wide Web Conference 2002 (WWW2002)*. 2002.
- [de Lara et al. 1999] Eyal de Lara, Karin Petersen, Douglas B. Terry, Anthony LaMarca, James Thornton, Michael Salisbury, Paul Dourish, W. Keith Edwards and John Lamping. Caching documents with active properties. In *Workshop on Hot Topics in Operating Systems*, pages 8–13. 1999.
- [Dourish et al. 2000a] Paul Dourish, W. Keith Edwards, Jon Howell, Anthony LaMarca, John Lamping, Karin Petersen, Michael Salisbury, Doug Terry and Jim Thornton. A programming model for active documents. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 41–50. ACM Press, 2000a. ISBN 1-58113-212-3.
- [Dourish et al. 2000b] Paul Dourish, W. Keith Edwards, Anthony LaMarca, John Lamping, Karin Petersen, Michael Salisbury, Douglas B. Terry and James Thornton. Extending document management systems with user-specific active properties. *ACM Transactions on Information Systems (TOIS)*, 18(2):140–170, 2000b.

- [Dourish et al. 1999a] Paul Dourish, W. Keith Edwards, Anthony LaMarca and Michael Salisbury. Presto: an experimental architecture for fluid interactive document spaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 6(2):133–161, 1999a.
- [Dourish et al. 1999b] Paul Dourish, W. Keith Edwards, Anthony LaMarca and Michael Salisbury. Using properties for uniform interaction in the presto document system. In *Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 55–64. ACM Press, 1999b. ISBN 1-58113-075-9.
- [Dourish et al. 1999c] Paul Dourish, John Lamping and Tom Rodden. Building bridges: customisation and mutual intelligibility in shared category management. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, pages 11–20. ACM Press, 1999c. ISBN 1-58113-065-1.
- [Dumais 2007] Susan Dumais. Information retrieval in context. In *IUI '07: Proceedings of the 12th international conference on Intelligent user interfaces*, pages 2–2. ACM Press, New York, NY, USA, 2007. ISBN 1-59593-481-2.
- [Dumais et al. 2003] Susan Dumais, Edward Cutrell, JJ Cadiz, Gavin Jancke, Raman Sarin and Daniel C. Robbins. Stuff i've seen: a system for personal information retrieval and re-use. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 72–79. ACM Press, New York, NY, USA, 2003. ISBN 1-58113-646-3.
- [Eldridge et al. 2000] Marge Eldridge, Mik Lamming, Mike Flynn, Chris Jones and David Pendlebury. Studies of mobile document work and their contributions to the satchel project. *Personal Technologies, Special Issue on HCI and Mobile Devices*, pages 102–112, 2000.
- [F. van Harmelen and Horrocks 2001] P. Patel-Schneider F. van Harmelen and I. Horrocks. Reference description of the daml+oil ontology markup language. Technical report, 2001.
- [Farquhar et al. 1996] A. Farquhar, R. Fikes and J. Rice. The ontolingua server: A tool for collaborative ontology construction. In *Proceedings of KAW96*. 1996.
- [Fertig et al. 1996] S. Fertig, E. Freeman and D. Gelernter. Lifestreams: An alternative to the desktop metaphor. In *Proceedings ACM SIGCHI Conference on Human Factors in Computing Systems Conference Companion (CHI '96)*, pages 410–411. ACM Press, 1996.
- [Flynn et al. 2000] Mike Flynn, David Pendlebury, Chris Jones, Marge Eldridge and Mik Lamming. The satchel system architecture: mobile access to documents and services. *Mobile Networks and Applications*, 5(4):243–258, 2000.

- [Freeman and Fertig 1995] E. Freeman and S. Fertig. Lifestreams: Organizing your electronic life. In *Proceedings AAAI Fall Symposium: AI Applications in Knowledge Navigation and Retrieval*. 1995.
- [Freeman and Gelernter 1996] E. Freeman and D. Gelernter. Lifestreams: A storage model for personal data. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 25(1):80–86, 1996.
- [Freeman 1995] Eric Freeman. Lifestreams for the newton. *PDA Developer*, 3(4):42–45, 1995.
- [Gemmell et al. 2006] Jim Gemmell, Gordon Bell and Roger Lueder. Mylifebits: a personal database for everything. *Commun. ACM*, 49(1):88–95, 2006.
- [Gemmell et al. 2002] Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker and Curtis Wong. Mylifebits: fulfilling the memex vision. In *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, pages 235–238. ACM Press, New York, NY, USA, 2002. ISBN 1-58113-620-X.
- [Gemmell et al. 2004] Jim Gemmell, Lyndsay Williams, Ken Wood, Roger Lueder and Gordon Bell. Passive capture and ensuing issues for a personal lifetime store. In *CARPE'04: Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 48–55. ACM Press, New York, NY, USA, 2004. ISBN 1-58113-932-2.
- [Giampaolo 1999] Dominic Giampaolo. *Practical File System Design with the Be File System*. Morgan Kaufmann, 1999.
- [Goldberg et al. 1992] David Goldberg, David Nichols, Brian M. Oki and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.
- [Golder and Huberman 2006] Scott Golder and Bernardo A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2006.
- [Gonçalves 2002] Daniel Gonçalves. Os espaços de documentos pessoais. Technical report, Instituto Superior Técnico, 2002.
- [Gonçalves and Jorge 2003] Daniel Gonçalves and Joaquim Jorge. An empirical study of personal document spaces. In *Proceedings DSV-IS 2003*, volume 2844 of *Lecture Notes on Computer Science*, pages 47–60. Springer-Verlag, 2003.
- [Huberman and Miles 1991] M. Huberman and M. Miles. *Analyse des données qualitatives. Recueil de nouvelles méthodes*. De Boeck, 1991.
- [Hull and Hart 2001] Jonathan Hull and Peter Hart. Toward zero-effort personal document management. *Computer*, 34(3):30–35, 2001.

- [John C. Platt and Field 2003] Mary Czerwinski John C. Platt and Brent A. Field. Photoc: automatic clustering for browsing personal photographs. In *Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia*, volume 1, pages 6–10. IEEE, 2003.
- [Jones 1998] Terry Jones. Attribute value systems: an overview. 1998.
- [Karger and Stein 1997] D. Karger and L. Stein. Haystack: Per-user information environments. 1997.
- [Karger and Jones 2006] David R. Karger and William Jones. Data unification in personal information management. *Commun. ACM*, 49(1):77–82, 2006.
- [Kim et al. 2004] Pilho Kim, Mark Podlaseck and Gopal Pingali. Personal chronicling tools for enhancing information archival and collaboration in enterprises. In *CARPE'04: Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 56–65. ACM Press, New York, NY, USA, 2004. ISBN 1-58113-932-2.
- [KnownSpace Group 2002] KnownSpace Group. <http://www.knownspace.org>. 2002.
- [Kuchinsky et al. 1999] Allan Kuchinsky, Celine Pering, Michael L. Creech, Dennis Freeze, Bill Serra and Jacek Gwizdka. Fotofile: a consumer multimedia organization and retrieval system. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 496–503. ACM Press, New York, NY, USA, 1999. ISBN 0-201-48559-1.
- [Lamming and Flynn 1994] M. Lamming and M. Flynn. Forget-me-not: intimate computing in support of human memory. In *Proceedings FRIEND21 Symposium on Next Generation Human Interfaces*. 1994.
- [Lamming et al. 1994] M. G. Lamming, P. J. Brown et al. The design of a human memory prosthesis. *Computer Journal*, 37(3):153–163, 1994.
- [Lamming et al. 2000] Mik Lamming, Marge Eldridge, Mike Flynn, Chris Jones and David Pendlebury. Satchel: providing access to any document, any time, anywhere. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(3):322–352, 2000.
- [M. Kifer and Wu 1995] G. Lausen. M. Kifer and J. Wu. Logical foundations of object-oriented and frame-based languages. *Journal of the ACM*, 1995.
- [MacGregor 1991] R. MacGregor. Inside the loom classifier. *SIGART Bulletin*, 2(3):70–76, 1991.

- [Malone 1983] Thomas W. Malone. How do people organize their desks?: Implications for the design of office information systems. *ACM Transactions on Information Systems (TOIS)*, 1(1):99–112, 1983.
- [Manber and Wu 1994] U. Manber and S. Wu. GLIMPSE: A tool to search through entire file systems. In *Proceedings of the USENIX Winter 1994 Technical Conference*, pages 23–32. San Fransisco, CA, USA, 1994.
- [Mander et al. 1992] Richard Mander, Gitta Salomon and Yin Yin Wong. A “pile” metaphor for supporting casual organization of information. In *Conference proceedings on Human factors in computing systems*, pages 627–634. ACM Press, 1992. ISBN 0-89791-513-5.
- [Marlow et al. 2006] Cameron Marlow, Mor Naaman, danah boyd and Marc Davis. Position paper, tagging, taxonomy, flickr, article, toread. In *Collaborative Web Tagging Workshop of the WWW2006 conference*. 2006.
- [Mayo 2001] Bob Mayo. Factoid. 2001.
- [Merrie Ringel and Horvitz 2003] Susan Dumais Merrie Ringel, Edward Cutrell and Eric Horvitz. Milestones in time: The value of landmarks in retrieving information from personal stores. In *Proceedings INTERACT 2003*. 2003.
- [Minsky 1981] Marvin Minsky. A framework for representing knowledge. In J. Hauge-land, editor, *Mind Design*, pages 95–128. MIT Press, 1981.
- [Nielsen 2002] Jakob Nielsen. Supporting multiple-location users. 2002.
- [Norman 1999] Donald Norman. *The Invisible Computer*. MIT Press, 1999.
- [Peters 2001] Rodney E. Peters. Exploring the design space for personal information management tools. In *Proceedings CHI’01 (poster)*. ACM Press, 2001.
- [Peters 2002] Rodney E. Peters. Organize this! investigating personal information management practices. In *Proceedings Conference proceedings on Designing interactive systems 2002*. ACM Press, 2002.
- [Porter 1980] M.F. Porter. An algorithm for suffix stripping. *Program*, (14):130–137, 1980.
- [Quan et al. 2002] Dennis Quan, David Huynh, Vineet Sinha and David Karger. Adenine: A metadata programming language. 2002. Student Oxygen Workshop.
- [Quillian 1967] M.R. Quillian. Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, (12):410–430, 1967.
- [Rekimoto 1999a] Jun Rekimoto. Time-machine computing: a time-centric approach for the information environment. In *Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 45–54. ACM Press, 1999a. ISBN 1-58113-075-9.

- [Rekimoto 1999b] Jun Rekimoto. Timescape: A time machine for the desktop environment. In *Proceedings of ACM CHI'99 Extended Abstracts*, pages 180–181. ACM Press, 1999b.
- [Rhodes 1997] Bradley J. Rhodes. The wearable remembrance agent: A system for augmented memory. *Personal Technologies Journal Special Issue on Wearable Computing*, (1):218–224, 1997.
- [Rhodes 2000a] Bradley J. Rhodes. *Just-In-Time Information Retrieval*. Ph.D. thesis, MIT Media Laboratory, Cambridge, MA, 2000a.
- [Rhodes 2000b] Bradley J. Rhodes. Margin notes: Building a contextually aware associative memory. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI '00)*. 2000b.
- [Rhodes and Starner 1996] Bradley J. Rhodes and Thad Starner. The remembrance agent: A continuously running automated information retrieval system. In *Proceedings of The First International Conference on The Practical Application of Intelligent Agents and Multi Agent Technology (PAAM '96)*, pages 487–495. 1996.
- [Richter and Cabrera 1998] J. Richter and L.F. Cabrera. A file system for the 21st century: Previewing the windows nt 5.0 file system. *Microsoft Systems Journal*, 1998.
- [Robertson et al. 1998] George Robertson, Mary Czerwinski, Kevin Larson, Daniel C. Robbins, David Thiel and Maarten van Dantzich. Data mountain: using spatial memory for document management. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 153–162. ACM Press, 1998. ISBN 1-58113-034-1.
- [Rodden 1999] Kerry Rodden. How do people organise their photographs. In *Proceedings of the BCS IRSG 21st Annual Colloquium on Information Retrieval Research*. 1999.
- [Rodden and Wood 2003] Kerry Rodden and Kenneth R. Wood. How do people manage their digital photographs? In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 409–416. ACM Press, New York, NY, USA, 2003. ISBN 1-58113-630-7.
- [Russell and Norvig 2003] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, second edition, 2003.
- [Salton and M.J.McGill 1983] G. Salton and M.J.McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., 1983.
- [Salton 1988] Gerald Salton. *Automatic Text Processing*. Addison-Wesley Longman Publishing Co., Inc., 1988.

- [Sen et al. 2006] Shilad Sen, Shyong K. Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F. Maxwell Harper and John Riedl. tagging, communities, vocabulary, evolution. In *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 181–190. ACM Press, New York, NY, USA, 2006. ISBN 1-59593-249-6.
- [Shamir 2004] Ariel Shamir. A view on views. In *Lecture Notes in Computer Science*, number 3031, pages 90–10. Springer-Verlag, Germany, 2004.
- [Shapiro and Martins 1990] S. C. Shapiro and J. P. Martins. Recent advances and developments - the sneps 2.1 report. In Deepak Kumar, editor, *Current Trends in SNePS-Semantic Network Processing System: Proceedings of the First Annual SNePS Workshop*, pages 1–13. Springer-Verlag, 1990.
- [Shapiro 1971] S.C. Shapiro. A net structure for semantic information storage, deduction and retrieval. In *Proceedings of the Second International Joint Conference on Artificial Intelligence*, pages 512–523. Morgan Kaufmann, 1971.
- [Sheldon et al. October 1991] Mark A. Sheldon, David K. Gifford, Pierre Jouvelot and James W. O’Toole Jr. Semantic file systems. *Proceedings of the 13th ACM Symposium on Operating Systems Principles*, pages 16–25, October 1991.
- [Sinha 2005] Rashmi Sinha. A cognitive analysis of tagging (or how the lower cognitive cost of tagging makes it popular). available online at http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html, 2005. Verified at April 11, 2007.
- [Skupin 2002] André Skupin. A cartographic approach to visualizing conference abstracts. *IEEE Computer Graphics and Applications*, 22(1):50–58, 2002.
- [Soules and Ganger 2005] Craig A. N. Soules and Gregory R. Ganger. Connections: using context to enhance file search. In *SOSP '05: Proceedings of the twentieth ACM symposium on Operating systems principles*, pages 119–132. ACM Press, New York, NY, USA, 2005. ISBN 1-59593-079-5.
- [Sowa 1991] J. Sowa. *Principles of Semantic Networks*. Morgan-Kaufman, San Mateo, CA, 1991.
- [Teevan et al. 2004] Jaime Teevan, Christine Alvarado, Mark S. Ackerman and David R. Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422. ACM Press, New York, NY, USA, 2004. ISBN 1-58113-702-8.
- [von Ahn and Dabbish 2004] Luis von Ahn and Laura Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326. ACM Press, New York, NY, USA, 2004. ISBN 1-58113-702-8.

- [von Ahn et al. 2006] Luis von Ahn, Ruoran Liu and Manuel Blum. Peekaboom: a game for locating objects in images. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 55–64. ACM Press, New York, NY, USA, 2006. ISBN 1-59593-372-7.
- [W3C 2002] W3C. Resource description framework (RDF) homepage. 2002. <http://www.w3.org/RDF/>.
- [Whittaker and Sidner 1996] Steve Whittaker and Candace Sidner. Email overload exploring personal information management of email. In *Conference proceedings on Human factors in computing systems*, pages 276–283. ACM Press, 1996. ISBN 0-89791-777-4.
- [Wiberg 2001] Mikael Wiberg. Roamware: an integrated architecture for seamless interaction in between mobile meetings. In *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*, pages 288–297. ACM Press, 2001. ISBN 1-58113-294-8.
- [Yin 1989] R. Yin. *Case Study. Design and Methods*. Sage Publications, 1989.

A1

Guiding the Interviews

Recent	Old	Other	Element	Sample Questions
			Time	<p>"Do you still remember when that was?"</p> <p>"Do you still know when that document was written?"</p>
			Place	"And where did you work on that document? At home? At the office?"
			Co-Authors	<p>"Did someone else read the document?"</p> <p>"Did someone else help you write the document?"</p>
			Purpose	<p>"What did you write the document for?"</p> <p>"What did you use that document for?"</p>
			Author	<p>"Do you still remember the author's name?"</p> <p>"Do you still remember something about the author?"</p>
			Subject	<p>"What was the document's subject?"</p> <p>"What was the document about?"</p>
			Other Docs.	<p>"Is there any other document you used or read while preparing this one?"</p> <p>"Did you write anything else at the time?"</p> <p>"Did you use that document in preparing another one?"</p>
			Personal Life	<p>"Do you remember something that happened on your personal life around the time you were working on that document?"</p> <p>"Do you remember a birthday, or something important that happened to you when you were writing that document?"</p>
			World in general	<p>"Did something happen to someone you know at the time?"</p> <p>"Do you remember any news or something important that might have happened to someone you know around that time?"</p>
			Exchanged	"Did you send the document to someone?"
			Type	"What is the kind of that document? Is it a Word document? Excel?"
			Task	"Was there a particularly difficult task in preparing that document?"
			Storage	<p>"Where is the document stored now?"</p> <p>"If you wanted to find the document now, do you remember where it is?"</p>
			Version	"Did the document have several versions?"
			Contents	"Do you remember if the document has a particularly large table, or perhaps a colourful graphic?"
			Event	"When you were working on the document, did someone come into the office? Did you receive an important phone call?"

A2

LFP Dialogues

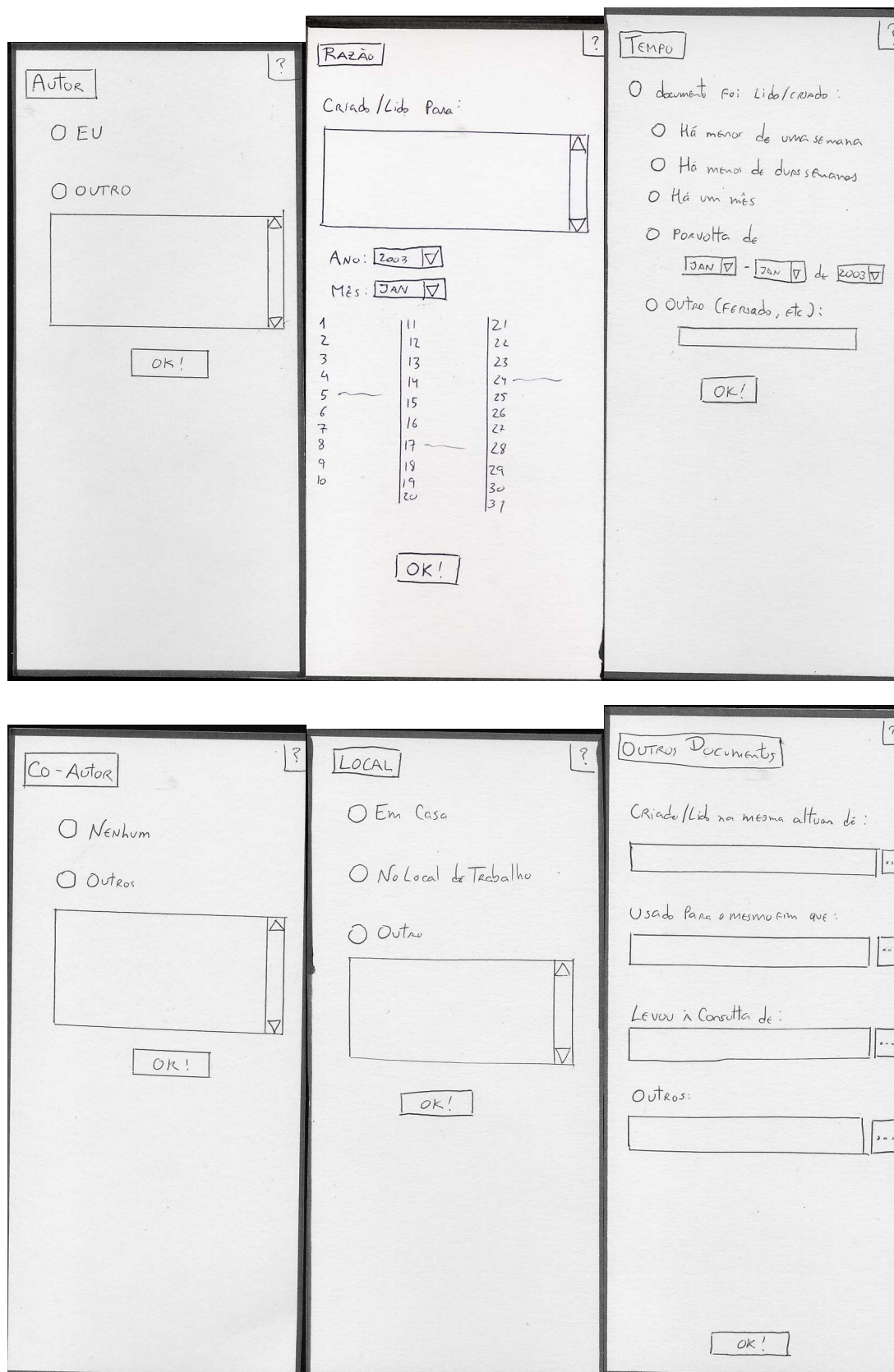


Figure A2.1: Author, Purpose, Time, Co-Authors, Place, Other Documents

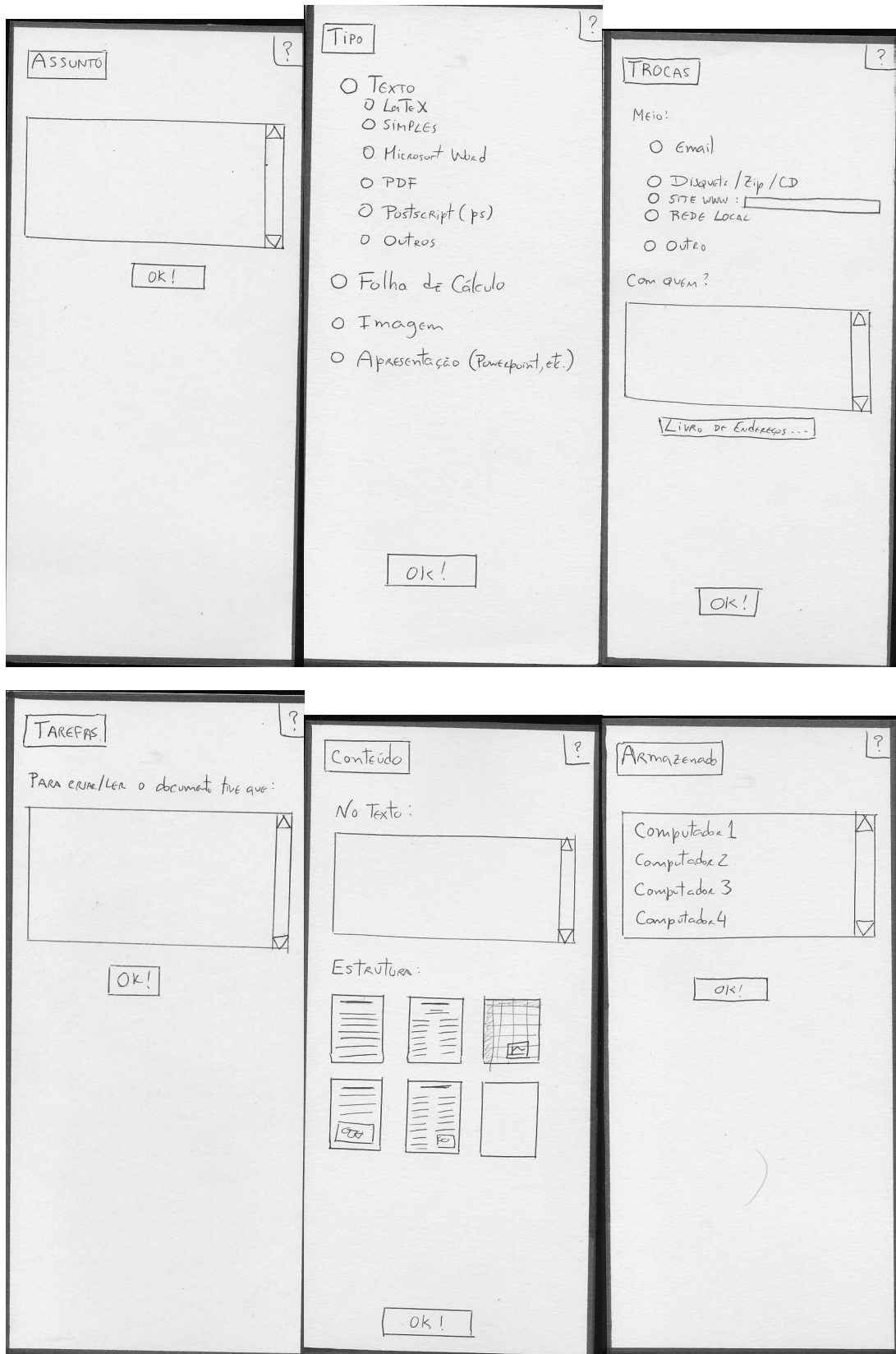


Figure A2.2: Subject, Type, Exchanges, Tasks, Contents, Storage

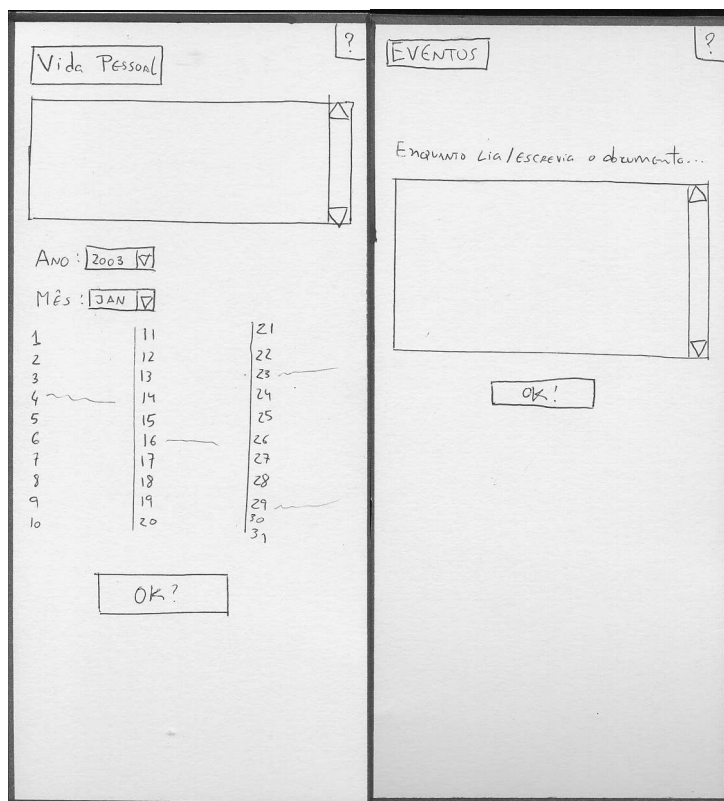
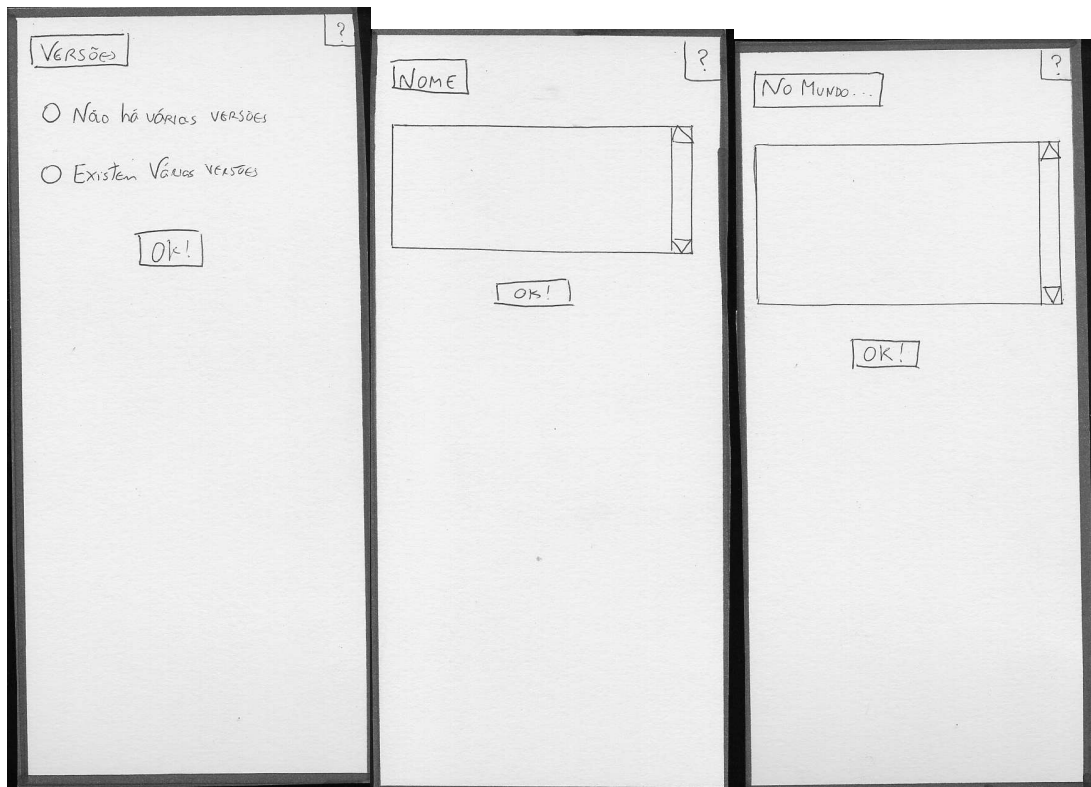


Figure A2.3: Versions, Name, World, Personal Life, Events

A3

User Comments about the LFPs

User											Comment
1	2	3	4	5	6	7	8	9	10	Total	
x										1	The system could show, on the elements list, which elements were already mentioned. This would be useful when several elements are already in the story and we can't remember what we have said
x										1	The probability of inserting a repeated element into a story is low, and should only happen when the user explicitly requests it
x										1	The elements in the interface should be visually distinct according to their atemporality, to better understand which can be moved to the past or future interface areas.
x										1	When I didn't remember anything else to tell, I read the element list in search of inspiration
x										1	I think that stories are told by following the order suggested by the interface, until faced with a hard to remember element or something that is felt as more important, and at that point another element is chosen.
	x									1	The most complicated dialog was Purpose: it shows only a calendar, but it should also allow us to select email messages.
	x									1	I already had told it the document's co-authors, and it asked me for them again...
	x									1	The dialogues for introducing story elements need some improvement.
	x									1	Given that, often, the suggested element is not the one we'd want, maybe a set of two or three alternate elements could be suggested instead of just one?
	x							x		2	Display some help text directly on the dialogue boxes, instead of having to press a button, could help to better understand what is expected
	x									1	A chronologic relation between documents is useful in telling the story, but sometimes a small story about other documents would also be of use, instead of having to point at actual documents.
		x								1	There is some redundancy between some dialogues
		x								1	It is unnecessary to have different dialogues for author and co-author.
		x								1	I think that a document's structure is more related to file type than to its contents
		x								1	The dialogue sequence seemed a bit random and they appeared to be unrelated
			x							1	There should be some way to specify time ranges in the displayed agenda
			x							1	The introduction of words in the document could be merged with stating its purpose

Continued on next page...

User											Comment
1	2	3	4	5	6	7	8	9	10	Total	
			x							1	There should be some sound or visual signal when the suggested documents change
					x					1	The order of story elements wasn't the one that naturally came to my mind, causing me to expend some effort not to forget something while I had to state other elements, mainly for the Recent Document.
					x					1	Showing related documents is good.
								x		1	The menus (sic.) aren't very clear. IT is hard to fully understand some of the functionalities therein.
								x		1	The dialogues about Personal Life are completely useless.
									x	1	Information about Events, World Events and Personal Life are unnecessary and useless.

Table A3.1: Prototype A: User Comments

User											Comment	
11	12	13	14	15	16	17	18	19	20	Total		
	x									1	I think that the questions about Personal Life and World Events could be merged into just one.	
		x								1	There are some ambiguous / confusing dialogues, that should be corrected.	
		x								1	Other Time intervals should be considered in that dialogue	
		x								1	The system should handle multimedia files (Movies, Audio, Photos, etc.) better.	
				x						1	The World Events part seems to be superfluous	
					x					1	Some questions are redundant in some contexts	
						x				1	Since there is a limited number of story elements, this can limit what we can remember about a older document.	
							x			1	I think it would be worthwhile to actually implement this prototype.	
								x		1	It would be an important improvement if the questions asked about the different story elements depended on what had already been told.	
									x	1	I think it would be interesting to try to collect some information about the actual location of the file (directory, etc.)	
									x	1	The question about Events is not very useful when describing older documents.	
										x	1	Associating the reading or writing of a document with World Events helped me tell the story to find the document.

Table A3.2: Prototype B: User Comments

User											Total	Comment	Kind
1	2	3	4	5	6	7	8	9	10				
x	x	x	x	x	x	x	x	x	x	x	10	Agenda in Purpose is confusing	Conf
x											1	Past/Present/Future areas confusing	Conf
	x										1	Confusion: Events vs. Exchanges	Conf
	x										1	Confusion: Element told after Other Documents related to the Other Document described (and it shouldn't)	Conf
		x	x				x				3	Confusion: Contents vs. Subject	Conf
			x								1	The agenda, in Time, is confusing	Conf
						x					1	Confusion: Personal Life vs. Time	Conf
							x				1	Confusion: "Plain Text" vs. MSWord	Conf
							x				1	Confusion: "words in the text"	Conf
								x			1	The several event types can be confused	Conf
								x			1	Confusion: Exchanges vs. Other Documents	Conf
							x				1	Co-Authors: Me	Conf
								x			1	Exchanges: Placed On-Line (later mentioned as storage)	Conf
					x						1	Co-Authors: doesn't remember the name, but it does the email	El
x											1	Exchanges: Instant Messaging	El
						x	x				2	Exchanges: printed version given to someone	El
	x				x						2	Other Documents: Emails exchanged about the target	El
					x						1	Other Documents: on-line (URL)	El
					x						1	Other Documents: received at the same time and used to build the target	El
						x					1	Personal Life: an important event (finishing the PhD)	El
								x			1	Storage: at a web site	El
	x										1	Storage: Removable Media	El
					x						1	Storage: shared network drive	El
						x	x				2	Tasks: not about the document, but preparing FOR the document	El
					x						1	Tasks: things done during and after interacting with target	El
					x						1	There was a deadline for the document in the agenda	El
x											1	Time: "A long time ago..."	El
			x								1	Time: Other intervals...	El
			x								1	Type: HTML	El
						x					1	Versions: the target was the base for another, full-fledged, document	El
						x	x				2	Content: organizational structure (chapters, etc.)	Lim

Conf–User confusion El–Element value Lim–Interface limitation Obs–Misc. observation

Continued on next page...

User											Total	Comment	Kind
1	2	3	4	5	6	7	8	9	10				
		x									1	Take two documents (images and text, for instance) and relate them (ex: it had an image like this one and a text like that one)	Lim
x											1	Only images in the content (but not an image file!)	Lim
x											1	Time is fuzzy...	Lim
x											1	Extrapolate data from elements already in the story (using transitivity, for instance)	Lim
x											1	Since an agenda is displayed, it should be fully-functional	Lim
x											1	Atemporal elements should be distinguishable from the others	Lim
x											1	The explicit way to tell in which way a document reminds of another is important	Lim
x	x										2	Ability to specify Contents in the negative ("It DIDN'T contain...")	Lim
	x				x	x	x				4	Use recursive stories about Other Documents	Lim
		x									1	Other Documents dragged to the central area should visually resemble other story elements	Lim
			x								1	Signal in some way (visually or with sound) when doc. suggestions changed	Lim
					x						1	Exchanges: separate sending and receiving	Lim
							x				1	Time should contemplate different intervals	Lim
								x			1	The document's name should appear with the thumbnails	Lim
x											1	Co-Authors: there can be more than the ones mentioned, which we can't remember.	Lim
									x		1	Document Structure (in contents): separate first page from the others	Lim
x	x										2	The user knows an element, but would rather mention something else at that point	Obs
x											1	The element order can be influenced by the element list	Obs
	x										1	Co-Author mentioned in the Author dialogue, resulting in redundancy	Obs
					x						1	Recent Document was received from someone else with some content, and the user was its co-author.	Obs
							x				1	Tasks: difficulty in stating them. Something more structured than a text box needed.	Obs

Conf–User confusion

El–Element value

Lim–Interface limitation

Obs–Misc. observation

Table A3.3: Prototype A: Notes

User											Total	Comment	Kind
11	12	13	14	15	16	17	18	19	20				
x	x	x	x	x	x	x	x		x		9	The Agenda in Purpose is confusing	Conf
x											1	Confusion: Exchanges	Conf
x											1	The several event types can be confused	Conf
x											1	The word 'Events' doesn't make very clear what is to be said (perhaps use 'interruptions' instead?)	Conf
	x										1	Confusion: Version vs. Name	Conf
		x		x	x			x	x		5	Confusion: Purpose vs. Subject	Conf
		x									1	Confusion: Events vs. Personal Life	Conf
				x							1	"I want another" used instead of "I can't remember" by mistake	Conf
			x								1	What should be said about Tasks is not clear	Conf
			x								1	Personal Life: New Year (but this is Time or World, not PL)	Conf
			x								1	Confusion: Storage vs. Place	Conf
					x	x					2	Confusion: Element told after Other Documents related to the Other Document described (and it shouldn't)	Conf
					x						1	Events needed to be explained	Conf
					x						1	Confusion: Contents	Conf
					x				x		2	Confusion: Document Title instead of its name	Conf
							x		x		2	Confusion 'Exchanges' (nothing was given in return)	Conf
							x				1	Two formats, Word and PDF (actually, two versions)	Conf
							x				1	Event: 'Christmas'	Conf
									x		1	Instead of other documents, the user mentioned "other applications"	Conf
				x							1	Purpose: "to print it" (!?)	Conf
						x					1	Type: Word and Excel (?)	Conf
									x		1	"Placed several versions inside the same document"	Conf
x											1	Author using short story ("that guy whose...")	El
							x				1	Contents: a little text with lots of tables	El
									x		1	Events: "I got a telephone call"	El
									x		1	Events: "I had to stop reading when the train reached its destination"	El
							x				1	Events: "The crazy man who works in the next office showed up..."	El
		x									1	Events: The ceiling fell down!	El
						x					1	Exchanged: Local Network	El
		x									1	Exchanges: CVS	El
									x		1	Exchanges: Email, web site	El
								x			1	Exchanges: received it and there were co-recipients	El
				x							1	Name: "it starts with an 'm'"	El
				x							1	Other Documents: "used for the same purpose that..." a year later...	El

Conf-User confusion

El-Element value

Lim-Interface limitation

Obs-Misc. observation

Continued on next page...

User											Total	Comment	Kind
11	12	13	14	15	16	17	18	19	20				
		x									1	Other Documents: some docs. produced by his work-team	El
									x		1	Personal Life: "I was annoyed"	El
							x				1	Personal Life: "I was at my parent's"	El
									x		1	Personal Life: "I worked with persons of another department"	El
					x						1	Personal Life: Both husband and son were ill	El
				x							1	Personal Life: niece's birthday	El
				x							1	Personal Life: Project Audit (also the Purpose)	El
								x			1	Personal Life: Right before going on vacations	El
								x			1	Personal Life: user's birthday the previous week	El
								x			1	Personal Life: went out for dinner and went to sleep at 4a.m.	El
x											1	Place remembered from printing type (laser vs. inkjet)	El
			x								1	Place: All over (edited on a laptop)	El
	x										1	Place: at a friend's house	El
							x				1	Place: at home and at parent's home	El
x		x		x							3	Place: at home and at work	El
		x									1	Place: workplace, but more than one (consultant)	El
									x		1	Purpose: "to update myself on that subject"	El
									x		1	Storage: company's intranet	El
		x									1	Storage: CVS repository	El
								x			1	Storage: in the desktop!	El
						x					1	Storage: Local Network	El
		x									1	Tasks: surfing the web for information	El
							x				1	Time: a day ago	El
								x			1	Time: a year ago	El
x											1	Time: more than one month ago	El
		x									1	Time: more than three years ago	El
									x		1	Time: two or three years	El
						x					1	Versions: several, adapted for different things	El
									x		1	World: "The thing that was supposed to happen in Portugal" (America's Cup...)	El
					x						1	World: 9/11	El
									x		1	World: Carlos Cruz went to jail	El
							x				1	World: Christmas	El
									x		1	World: Lunar Eclipse	El
					x						1	World: member of MSc thesis juror appeared on TV	El
								x			1	World: Portugal won't host the America's Cup (sailing)	El
		x									1	World: revolution in Georgia	El

Conf-User confusion

El-Element value

Lim-Interface limitation

Obs-Misc. observation

Continued on next page...

User											Total	Comment	Kind
11	12	13	14	15	16	17	18	19	20				
						x					1	World: Suicide bombing of Italian barracks in Iraq	El
		x									1	World: Y2K bug	El
		x							x		2	Content: organizational structure (chapters, etc.)	Lim
x	x										2	I don't remember different from It didn't happen	Lim
x											1	I don't want to need to look at the dialogue to say something didn't happen	Lim
x	x										2	Exchanges: separate sending and receiving	Lim
x											1	Contents: Describe Images	Lim
	x										1	Use recursive stories about Other Documents	Lim
	x										1	Element list should be sorted by the order in which they are expected in the stories	Lim
							x				1	Other Document: Created at the Same time and for the same end	Lim
								x			1	Storage: PDA	Lim
									x		1	Place: On the train (where printed version was read)	Lim
							x				1	Databases unsupported	Lim
x											1	Personal Life: things I've done vs. family-related events	Obs
x											1	Show the dialogues at the same time than the sentences is important	Obs
	x										1	Tasks: some structured way of telling them is required	Obs
	x										1	Exchanges of previous versions, rather than the final document	Obs
								x			1	Content already mentioned (redundancy!)	Obs

Conf–User confusion

El–Element value

Lim–Interface limitation

Obs–Misc. observation

Table A3.4: Prototype B: Notes

A4

RDF Schemas used by Scroll

A4.1. The Quill RDFS

```
<?xml version="1.0"?>
```

```
<!--
```

```
    This is the Quill RDF Schema for use with Quill and Scroll
```

```
    Copyright (C) 2004 Daniel Goncalves
```

```
    QUILL Namespace defined by this document located at  
    http://narrative.shorturl.com/quill
```

```
    Revision History:
```

3/10/2004 - v0.1 - Basic document description

20/10/2004 - v0.2 - Some new classes added, and some changes to others especially for Event.

4/11/2004 - v0.3 - Version-related fields added to the Document class

4/11/2004 - v0.4 - StoryElement class added.

21/11/2004 - v0.4.1 - New mimeType and eventStartTime and eventEndTime properties

28/07/2005 - v0.4.2 - New eventOtherEvents property in the Event class

8/08/2005 - v0.4.3 - New deleted property in the Document class

13/12/2005 - v0.5 - no attributes of Document class are now seqs or bags

-->

```
<!DOCTYPE rdf:RDF [  
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">  
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">  
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">  
  <!ENTITY quill "http://narrative.shorturl.com/quill#">  
>
```

```
<rdf:RDF xml:lang="en"  
  xmlns:rdf="&rdf;"  
  xmlns:rdfs="&rdfs;"  
  xmlns:quill="&quill;#"  
  xml:base="&quill;">
```

```
<!-- Document Class -->
```



```
<rdfs:Class rdf:ID="Document">
  <rdfs:comment>All relevant information about a document</rdfs:comment>
</rdfs:Class>

<rdf:Property rdf:ID="name">
  <rdfs:comment>
    The file's name. It can be an URL for online docs.
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="#rdfs:Literal"/>
</rdf:Property>

<rdf:Property rdf:ID="extension">
  <rdfs:comment>The file's extension</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="#rdfs:Literal"/>
</rdf:Property>

<rdf:Property rdf:ID="mimeType">
  <rdfs:comment>The file's mime-type</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="#rdfs:Literal"/>
</rdf:Property>

<rdf:Property rdf:ID="class">
  <rdfs:comment>
    The document class. Valid classes are:
      text,image,spreadsheet,database,presentation,pim,shockwave,
      web script,bookmark,video,sound,executable,system,
      source,palm,archive,Backup,SymLink,NoExt,UNKNOWN
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="#rdfs:Literal"/>
</rdf:Property>

<rdf:Property rdf:ID="path">
  <rdfs:comment>The document's location on the hard drive</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="#rdfs:Literal"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="locus">
  <rdfs:comment>The locus in which a document is located</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="#Locus"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="RFID_ID">
  <rdfs:comment>The ID of the RFID tag that identifies the document's printed counter</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="created">
  <rdfs:comment>
    When it was created, in seconds since the epoch, as returned
    by the python function time.gmtime()
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="modified">
  <rdfs:comment>
    When it was modified, in seconds since the epoch, as returned
    by the python function time.gmtime(). For each modification a
    modified arc should point to the respective timestamp.
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="accessed">
  <rdfs:comment>
    When it was accessed, in seconds since the epoch, as returned
    by the python function time.gmtime().
    One arc for each</rdfs:comment>
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="size">
  <rdfs:comment>The file size in bytes</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&xsd;integer"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="keywords">
  <rdfs:comment>The set of all relevant keywords.
  One arc for each</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="creator">
  <rdfs:comment>The author(s) of the document.
  One arc for each</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="title">
  <rdfs:comment>The document's title</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&rdf;Literal"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="prevVersion">
  <rdfs:comment>The document that was the previous version of this one.</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="#Document"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="nextVersion">
  <rdfs:comment>The document that is the next version of this one.</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="#Document"/>
</rdf:Property>
```

```
<!-- This one might be a CRC or MD5 key or might end up not being used at all...
```

(other methods of finding versions might be better? Shouldn't we register a new version of a document only when its name changes, and just update the information in the existing record otherwise?)

-->

```
<rdf:Property rdf:ID="hash">
  <rdfs:comment>A hash value identifying the document</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&xsd;integer"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="physicalLocation">
  <rdfs:comment>The physical location of the document</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&rdf;Literal"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="deleted">
  <rdfs:comment>Flag that should be 1 if the doc. has probably been deleted
                or 0, otherwise
</rdfs:comment>
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="&xsd;integer"/>
</rdf:Property>
```

<!-- EFFICIENCY TWEAKS

All time instants should be the object of a triple of the form:

(quill:time-instant, quill:value, *instant*)

All keywords should be also registered in the KW namespace in triples like:

(KW:*keyword*, KW:hasdoc, *the document*)

-->

```
<!-- Lucus Class -->
```

```
<rdfs:Class rdf:ID="Locus">  
  <rdfs:comment>Identifying a locus on the users's PDFs</rdfs:comment>  
</rdfs:Class>
```

```
<rdf:Property rdf:ID="description">  
  <rdfs:comment>A string describing the Locus</rdfs:comment>  
  <rdfs:domain rdf:resource="#Locus" />  
  <rdfs:range rdf:resource="&rdfs;Literal"/>  
</rdf:Property>
```

```
<rdf:Property rdf:ID="address">  
  <rdfs:comment>The Locus IP address, if available</rdfs:comment>  
  <rdfs:domain rdf:resource="#Locus" />  
  <rdfs:range rdf:resource="&rdfs;Literal"/>  
</rdf:Property>
```

```
<!-- DocumentInteraction Class -->
```

```
<rdfs:Class rdf:ID="DocumentInteraction">  
  <rdfs:comment>An interaction with the document</rdfs:comment>  
</rdfs:Class>
```

```
<rdf:Property rdf:ID="interactionTime">  
  <rdfs:comment>When the interaction took place, in seconds since  
    the epoch, as returned by the python function time.gmtime().  
</rdfs:comment>  
  <rdfs:domain rdf:resource="#DocumentInteraction" />  
  <rdfs:range rdf:resource="&xsd;integer;"/>  
</rdf:Property>
```

```
<rdf:Property rdf:ID="interactionDuration">  
  <rdfs:comment>How long the interaction took, in seconds.</rdfs:comment>  
  <rdfs:domain rdf:resource="#DocumentInteraction" />
```

```
    <rdfs:range rdf:resource="&xsd;integer;" />
  </rdf:Property>
```

```
<!-- Event Class -->
```

```
<rdfs:Class rdf:ID="Event">
  <rdfs:comment>
    An event involving the user and/or his documents
  </rdfs:comment>
</rdfs:Class>
```

```
<rdf:Property rdf:ID="eventStartTime">
  <rdfs:comment>
    When an event occurred, in seconds since the epoch, as returned
    by the python function time.gmtime()
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Event" />
  <rdfs:range rdf:resource="&rdfs;Literal" />
</rdf:Property>
```

```
<rdf:Property rdf:ID="eventEndTime">
  <rdfs:comment>
    When an event occurred, in seconds since the epoch, as returned
    by the python function time.gmtime()
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Event" />
  <rdfs:range rdf:resource="&rdfs;Literal" />
</rdf:Property>
```

```
<rdf:Property rdf:ID="eventType">
  <rdfs:comment>
    Of what type was the event?? Possible types (more to add
    later, eventually with subclassing):
    email, www, appointment, transfer
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Event" />
  <rdfs:range rdf:resource="&rdfs;Literal" />
</rdf:Property>
```

```
<rdf:Property rdf:ID="eventDescription">
  <rdfs:comment>A textual description of the event, if relevant</rdfs:comment>
  <rdfs:domain rdf:resource="#Event" />
  <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>

<rdf:Property rdf:ID="eventDocument">
  <rdfs:comment>
    The document to which this event directly pertains (if any)
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Event" />
  <rdfs:range rdf:resource="&rdf;Bag"/>
</rdf:Property>

<rdf:Property rdf:ID="eventOtherDocuments">
  <rdfs:comment>
    Other documents involved in the event. It is expected
    that each element of the Bag is an instance of quill:Document
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Event" />
  <rdfs:range rdf:resource="&rdf;Bag"/>
</rdf:Property>

<rdf:Property rdf:ID="eventOtherEvents">
  <rdfs:comment>
    Other events related in the event. It is expected
    that each element of the Bag is an instance of quill:Event
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Event" />
  <rdfs:range rdf:resource="&rdf;Bag"/>
</rdf:Property>

<rdf:Property rdf:ID="eventParticipants">
  <rdfs:comment>
    All involved participants. It is expected that each element of
    the Bag is an instance of quill:Person</rdfs:comment>
  <rdfs:domain rdf:resource="#Event" />
  <rdfs:range rdf:resource="&quill;Bag"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="eventDirection">
  <rdfs:comment>
    When relevant, either 'in' or 'out'. It marks, for instance,
    incoming and outgoing mail, downloading/uploading, etc.
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Event" />
  <rdfs:range rdf:resource="&quill;Literal"/>
</rdf:Property>
```

```
<!-- Person Class -->
```

```
<rdfs:Class rdf:ID="Person">
  <rdfs:comment>
    A person the user must interact with
  </rdfs:comment>
</rdfs:Class>
```

```
<rdf:Property rdf:ID="personNames">
  <rdfs:comment>
    The name(s) it is know to have
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Person" />
  <rdfs:range rdf:resource="&rdf;Bag"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="personEmails">
  <rdfs:comment>
    The email(s) it is know to have
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Person" />
  <rdfs:range rdf:resource="&rdf;Bag"/>
</rdf:Property>
```

```
<rdf:Property rdf:ID="personBirthdate">
  <rdfs:comment>
    The person's birthdate, in seconds since the epoch, as returned
    by the python function time.gmtime()
  </rdfs:comment>
  <rdfs:domain rdf:resource="#Person" />
  <rdfs:range rdf:resource="&xsd;integer"/>
```



```
</rdf:Property>
```

```
<!-- Story Element Class -->
```

```
<rdfs:Class rdf:ID="StoryElement">
```

```
  <rdfs:comment>
```

```
    An element in a document-describing story
```

```
  </rdfs:comment>
```

```
</rdfs:Class>
```

```
<rdf:Property rdf:ID="seType">
```

```
  <rdfs:comment>
```

```
    What kind of element?
```

```
  </rdfs:comment>
```

```
  <rdfs:domain rdf:resource="#storyElement" />
```

```
  <rdfs:range rdf:resource="&rdf;Literal"/>
```

```
</rdf:Property>
```

```
<rdf:Property rdf:ID="seValue">
```

```
  <rdfs:comment>
```

```
    A high-level representation of the value, flag or other
    information to be read by the program. Easier than wadding
    through the constraints.
```

```
  </rdfs:comment>
```

```
  <rdfs:domain rdf:resource="#storyElement" />
```

```
  <rdfs:range rdf:resource="&rdf;Literal"/>
```

```
</rdf:Property>
```

```
<rdf:Property rdf:ID="seConstraints">
```

```
  <rdfs:comment>
```

```
    The constraints this story element imposes on the kb. Each element
    of the Bag should be a iQuill:Connective
```

```
  </rdfs:comment>
```

```
  <rdfs:domain rdf:resource="#storyElement" />
```

```
  <rdfs:range rdf:resource="&rdf;Bag"/>
```

```
</rdf:Property>
```

```
</rdf:RDF>
```

A4.2. The iQuill RDFS

```
<?xml version="1.0"?>
```

```
<!--
```

```
    This is the iQuill RDF Schema for use writing inference rule patterns in
    Quill and Scroll
```

```
    Copyright (C) 2004 Daniel Goncalves
```

```
    iQuill Namespace defined by this document located at
    http://narrative.shorturl.com/iquill
```

```
    Revision History:
```

```
        4/11/2004 - v1.0 - First version
```

```
-->
```

```
<!DOCTYPE rdf:RDF [
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">
  <!ENTITY iquill "http://narrative.shorturl.com/iquill#">
]>
```

```
<rdf:RDF xml:lang="en"
  xmlns:rdf="&rdf;"
```

```

xmlns:rdfs="&rdfs;"
xmlns:iquill="&iquill;"
xml:base="&iquill;">

```

```
<!--
```

PLEASE NOTE!!

We'll implement FOL-like inference capabilities, EXCEPT:

- all variables are universally quantified (what makes sense in most queries anyway)
 - NO negation (major efficiency hog...)
- ```
-->
```

```
<!-- Variable Class-->
```

```

<rdfs:Class rdf:ID="Variable">
 <rdfs:comment>A variable node</rdfs:comment>
</rdfs:Class>

```

```

<rdf:Property rdf:ID="name">
 <rdfs:comment>The variable's name</rdfs:comment>
 <rdfs:domain rdf:resource="#Variable" />
 <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>

```

```
<!-- Efficiency Matters -->
```

```

<rdf:Property rdf:ID="IOrder">
 <rdfs:comment>
 Specify the order in which the arcs leading from the subject node

```

```
 of this property should be evaluated.
 </rdfs:comment>
 <rdfs:range rdf:resource="#&rdf;Seq"/>
</rdf:Property>

<!-- Connective Class and sub-classes -->

<rdfs:Class rdf:ID="Connective">
 <rdfs:comment>A logic connective</rdfs:comment>
</rdfs:Class>

<rdf:Property rdf:ID="Args">
 <rdfs:comment>The Connective's arguments</rdfs:comment>
 <rdfs:domain rdf:resource="#Connective" />
 <rdfs:range rdf:resource="#&rdf;Seq"/>
</rdf:Property>

<rdfs:Class rdf:ID="And">
 <rdfs:subClassOf rdf:resource="#Connective" />
</rdfs:Class>

<rdfs:Class rdf:ID="Or">
 <rdfs:subClassOf rdf:resource="#Connective" />
</rdfs:Class>

<rdfs:Class rdf:ID="Imply">
 <rdfs:subClassOf rdf:resource="#Connective" />
</rdfs:Class>

<rdf:Property rdf:ID="Cq">
 <rdfs:comment>The Consequent</rdfs:comment>
 <rdfs:domain rdf:resource="#Imply" />
 <rdfs:range rdf:resource="#&rdfs;Bag"/>
</rdf:Property>
```

```
<!-- Constraints to use in the rules... -->

<rdfs:Class rdf:ID="Constraint">
 <rdfs:comment>A comparison Constraint</rdfs:comment>
</rdfs:Class>

<rdf:Property rdf:ID="Arg1">
 <rdfs:comment>The first argument of a Constraint</rdfs:comment>
 <rdfs:domain rdf:resource="#Constraint" />
</rdf:Property>

<rdf:Property rdf:ID="Arg2">
 <rdfs:comment>The second argument of a Constraint</rdfs:comment>
 <rdfs:domain rdf:resource="#Constraint" />
</rdf:Property>

<!-- Meta Constraints... -->

<rdfs:Class rdf:ID="MetaConstraint">
 <rdfs:comment>Meta Constraints</rdfs:comment>
 <rdfs:subClassOf rdf:resource="#Constraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="equal">
 <rdfs:subClassOf rdf:resource="#MetaConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="diff">
 <rdfs:subClassOf rdf:resource="#MetaConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="inContainer">
 <rdfs:subClassOf rdf:resource="#MetaConstraint" />
</rdfs:Class>
```

```
<!-- Math Constraints... -->
```

```
<rdfs:Class rdf:ID="MathConstraint">
 <rdfs:comment>Math Constraints</rdfs:comment>
 <rdfs:subClassOf rdf:resource="#Constraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="greater">
 <rdfs:subClassOf rdf:resource="#MathConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="greaterOrEqual">
 <rdfs:subClassOf rdf:resource="#MathConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="lesser">
 <rdfs:subClassOf rdf:resource="#MathConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="lesserOrEqual">
 <rdfs:subClassOf rdf:resource="#MathConstraint" />
</rdfs:Class>
```

```
<!-- String Constraints... -->
```

```
<rdfs:Class rdf:ID="StringConstraint">
 <rdfs:comment>String Constraints</rdfs:comment>
 <rdfs:subClassOf rdf:resource="#Constraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="subString">
 <rdfs:subClassOf rdf:resource="#StringConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="startsWith">
 <rdfs:subClassOf rdf:resource="#StringConstraint" />
</rdfs:Class>
```

```

<rdfs:Class rdf:ID="endsWith">
 <rdfs:subClassOf rdf:resource="#StringConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="subStringCI">
 <rdfs:subClassOf rdf:resource="#StringConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="subStringAccents">
 <rdfs:subClassOf rdf:resource="#StringConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="startsWithCI">
 <rdfs:subClassOf rdf:resource="#StringConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="endsWithCI">
 <rdfs:subClassOf rdf:resource="#StringConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="equalCI">
 <rdfs:subClassOf rdf:resource="#StringConstraint" />
</rdfs:Class>

<rdfs:Class rdf:ID="diffCI">
 <rdfs:subClassOf rdf:resource="#StringConstraint" />
</rdfs:Class>

<!-- Functions to use in the rules... -->

<rdfs:Class rdf:ID="Function">
 <rdfs:comment>A function</rdfs:comment>
</rdfs:Class>

<rdf:Property rdf:ID="Op1">
 <rdfs:comment>The first argument of a Function</rdfs:comment>
 <rdfs:domain rdf:resource="#Function" />
</rdf:Property>

```

```
<rdf:Property rdf:ID="Op2">
 <rdfs:comment>The second argument of a Function</rdfs:comment>
 <rdfs:domain rdf:resource="#Function" />
</rdf:Property>

<rdf:Property rdf:ID="Res">
 <rdfs:comment>The result of a Function</rdfs:comment>
 <rdfs:domain rdf:resource="#Function" />
 <rdfs:range rdf:resource="&quill;Variable" />
</rdf:Property>

<!-- Meta Functions... -->

<rdfs:Class rdf:ID="MetaFunction">
 <rdfs:comment>Meta Functions</rdfs:comment>
 <rdfs:subClassOf rdf:resource="#Function" />
</rdfs:Class>

<rdfs:Class rdf:ID="containerLen">
 <rdfs:subClassOf rdf:resource="#MetaFunction" />
</rdfs:Class>

<!-- String Functions... -->

<rdfs:Class rdf:ID="StringFunction">
 <rdfs:comment>String Functions</rdfs:comment>
 <rdfs:subClassOf rdf:resource="#Function" />
</rdfs:Class>

<rdfs:Class rdf:ID="concatenate">
 <rdfs:subClassOf rdf:resource="#StringFunction" />
</rdfs:Class>

<rdfs:Class rdf:ID="upper">
 <rdfs:subClassOf rdf:resource="#StringFunction" />
</rdfs:Class>
```



```
<rdfs:Class rdf:ID="lower">
 <rdfs:subClassOf rdf:resource="#StringFunction" />
</rdfs:Class>

<!-- Math Functions... -->

<rdfs:Class rdf:ID="MathFunction">
 <rdfs:comment>String Functions</rdfs:comment>
 <rdfs:subClassOf rdf:resource="#Function" />
</rdfs:Class>

<rdfs:Class rdf:ID="absolute">
 <rdfs:subClassOf rdf:resource="#MathFunction" />
</rdfs:Class>

<rdfs:Class rdf:ID="multiply">
 <rdfs:subClassOf rdf:resource="#MathFunction" />
</rdfs:Class>

<rdfs:Class rdf:ID="add">
 <rdfs:subClassOf rdf:resource="#MathFunction" />
</rdfs:Class>

<rdfs:Class rdf:ID="subtract">
 <rdfs:subClassOf rdf:resource="#MathFunction" />
</rdfs:Class>

<rdfs:Class rdf:ID="divide">
 <rdfs:subClassOf rdf:resource="#MathFunction" />
</rdfs:Class>

<rdfs:Class rdf:ID="intDivide">
 <rdfs:subClassOf rdf:resource="#MathFunction" />
</rdfs:Class>

<rdfs:Class rdf:ID="remainder">
 <rdfs:subClassOf rdf:resource="#MathFunction" />
</rdfs:Class>
```

```
<rdfs:Class rdf:ID="truncate">
 <rdfs:subClassOf rdf:resource="#MathFunction" />
</rdfs:Class>

<rdfs:Class rdf:ID="nearestInt">
 <rdfs:subClassOf rdf:resource="#MathFunction" />
</rdfs:Class>

</rdf:RDF>
```

# A5

## Interview Transcripts

---

### Interview 1

[Researcher] A primeira coisa que te peço que faças é para pensares num documento qualquer, concreto, que tenhas escrito tu, em que tenhas trabalhado, nos últimos dias ou...

[Interviewed] Nos últimos dias?

[R] Sim. Ou no último par de semanas, no máximo.

[I] Pode ser aquele em que estou a trabalhar agora?

[R] Pode ser no que estás a trabalhar agora. Um documento em que estejas a trabalhar agora. Pensa num, em concreto...

[I] Está bem. Ok.

[R] E agora, diz-me do que é que te lembras sobre ele. Não só em termos de interacção com o computador mas tudo o que achas relevante...

[I] Pronto. É um artigo que tinha enviado para o meu orientador. Tínhamos enviado para uma conferência já há um tempo. Não foi aceite... entretanto por acaso tinha o documento no Linux... no Unix, na área do [grupo de trabalho]. Por acaso eu não tinha... ou seja... tinha enviado isso já há uns meses, há um tempo. E agora...

[R] Lembras-te, desculpa interromper-te, lembras-te para onde é que foi enviado?

[I] A conferência?

[R] Sim. Ainda te lembras de qual foi?

[I] Acho que era a LPAR 2003. Pronto. E agora nós recebemos um CFP de outra conferência e vamos experimentar mandar esse mesmo artigo, com umas alterações, adequadas mais a esta conferência. Então agora fui tentar recuperá-lo... portanto... eu tinha aqui... antes trabalhava mais na área do [Grupo de Trabalho]. Agora trabalho no Windows. não trabalho na área, trabalho no próprio disco, e fui recuperá-lo à área. Por acaso enviei-o ao meu orientador para fazer a contagem do número de palavras, ou de o número de caracteres, já não me lembro, e por acaso eu não sabia se tinha mesmo a última versão. Ainda hei de lhe perguntar se era mesmo a última versão, se ela teria uma mais recente, mas pronto foi esse o documento. Queres que diga mais alguma coisa do documento. Queres que diga mais alguma coisa do documento?

[R] Se te lembrares de mais alguma coisa... Em relação por exemplo ao tema do documento, ou se...

[I] O tema do documento é o tema da minha tese de mestrado, [tema].

[R] Ok. Lembras-te de mais alguma coisa? Quando estavas a editar o documento houve alguma, algum percalço, alguma coisa em particular de que te lembres?

[I] Portanto eles pediram para... (pausa) Eles pediram para... percalço... não...

[R] Nada em particular?

[I] Ele foi feito em LaTeX. Algumas coisas estavam em Word, e eu lembro de ter aproveitado algum texto. Tinha feito artigos anteriores em Word e posto lá. Ele foi sempre trocado entre mim e o meu orientador. Ela foi escrevendo umas coisas e eu fui escrevendo outras... Isto também já foi há uns meses e por isso não me lembro assim...

[R] Ok. E trabalhaste nele mais no técnico, em casa? Tenso noção disso ou não é muito relevante?

[I] trabalhei mais... acho que trabalhei nos dois lados. Ah! Não! Como era em LaTeX sou capaz de ter trabalhado mais aqui no técnico mas devo ter trabalhado nos dois sítios. Imprimia e depois lia em casa para corrigir e depois chegava aqui e alterava.

[R] E só para terminar sobre este documento. Associas algum evento do mundo em geral... mesmo que não seja pessoal teu. Por exemplo, foi na altura em que começou a guerra não sei aonde... alguma coisa que associes assim imediatamente ou nem por isso?

[I] Se calhar é um mau exemplo porque é um documento que já foi alterado muitas vezes e...

[R] Exacto, mas nalguma das vezes em que trabalhaste nele há alguma coisa assim que te venha à memória imediatamente? Se vier, óptimo. Se não também não há problema nenhum, obviamente.

[I] Lembro-me... lembro-me que para o escrever houve uma altura que eu tinha pegado numa revisão, portanto, um outro artigo que tinha sido enviado com algum (...) e tinham enviado críticas portanto... mas nem sequer foi um artigo que eu escrevi, é um anterior, do meu orientador. E lembro-me de ter ido à biblioteca aqui para ler as referências que diziam lá nessas críticas para, para ver... pronto... para corrigir, eventualmente.

[R] Então se calhar agora vamos pensar... Vamos fazer o mesmo exercício mas desta vez vamos pensar num documento que tenhas escrito mas já há mais tempo, dois anos, uma coisa assim. Uma coisa no início do mestrado. Pensa num documento, para estarmos a pensar num em concreto e depois voltamos a fazer...

[I] Mas algum em especial???

[R] Um de que te lembres facilmente, para ter alguma coisa que contar. Pode ser o primeiro que te vier à cabeça, que tenhas feito nos últimos anos do curso, no princípio do mestrado, um relatório de qualquer coisa...

[I] Uma página, pode ser uma página?

[R] Sim, sim.

[I] Em HTML?

[R] Sim.

[I] Ok. Pode ser o projecto de IHM.

[R] Então, o que é que te lembras dele?

[I] Lembro-me... para já que antes de fazermos o documento, quer dizer a página, o HTML, tivemos que fazer umas entrevistas. Fazer as entrevistas. Tivemos que ir à associação... eu e o meu colega de grupo, tivemos que ir à ACAPO... Pronto. Mas isso não é muito relacionado com o documento mesmo... É mais o que aconteceu antes.

[R] Exacto. Mais com o documento em si. Se me quisesses dizer 'olha, preciso que me encontres o documento que foi assim ou assado...'

[I] Lembro-me muito bem que quando estávamos a fazer aquilo, o documento, estávamos ali no laboratório da LEIC. Eu lembro-me de estar a fazer os botões. Fazer imagens, por efeitos. A fazer os botões para o documento. Isso foi num PC da LEIC ali no 2º andar. Lembro-me de ter editado lá... Lembro porque acho piada, nos projectos, normalmente, não fico por ali a desenhar. Dessa vez lembro-me. Lembro-me também que quando foi para entregar esse projecto lembro-me que tivemos que por aquilo numa área especial do... naquela altura acho que ainda se chamava Camões... numa área especial para os professores de IHM poderem aceder facilmente aos... para não ter que se andar a fazer submissões nem mandar por mail estava num sitio especial. Algures no camões, na máquina. Ou então era mesmo na nossa área com permissões para eles lá irem...

[R] Em relação a...

[I] Lembro-me que editei em vários sítios. Na altura estava numa residência de estudantes e editei lá, eu e o meu colega, editamos aqui no Técnico.

[R] Em relação a... Lembras-te mais ou menos da altura em que isso foi? Da data em que isso foi?

[I] Sim, foi prái há uns... Foi IHM, portanto foi no 4a. ano. 4o? Sim, acho que é 4o. ano.

[R] E em termos de... de outras pessoas que tenham trabalhado nisso contigo, provavelmente o teu colega de grupo...

[I] (interrompendo) O meu colega!

[R] Há mais alguém que possa estar envolvido ou foram só vocês os dois...

[I] Não, fomos só nos os dois. Não houve cópias de projecto nem nada disso (risos). Fomos só nos os dois...

[R] E sobre o que é que era o projecto, já agora?

[I] O nosso era sobre terminais Braille. Para cegos, portanto. Aquilo tinha uns pontinhos e os cegos, supostamente, através dos dedos sabiam o que devia aparecer no display.

[R] E outras coisas que tenhas escrito na altura. Lembras-te assim de mais algum documento?

[I] Sim... Lembro-me que na mesma altura estava a escrever o projecto de aprendizagem, que deu-me muito mais trabalho... quer dizer, foi dos projectos que mais deu trabalho no meu curso, pelo menos no meu ano, estava a escreve o projecto de aprendizagem eu e o meu colega e mais outros dois colegas. (...) Portanto lembro-me de estar nós dois mais outros dois éramos o grupo de aprendizagem, éramos quatro. E fizemos em casa de um colega, o projecto, e ao mesmo tempo quer eles quer nós estávamos também a fazer IHM. Portanto andávamos, se não me engano, eu andava ou era com uma Zip ou era com disquetes mas tenho a impressão que andava com uma Zip, com os dois projectos lá dentro, foi em simultâneo que estavam a fazer os projectos. Projectos e relatório e não sei o quê. Lembro-me até de uma altura que... acho que foi no mesmo dia, tínhamos que entregar no mesmo dia ou quase no mesmo dia e eles, os outros dois que faziam grupo de IHM também estavam mais atrasados do que nós e tiveram que parar o projecto de aprendizagem e voltar para o de IHM e nos continuámos no de aprendizagem.

[R] E assim do mundo em geral, houve acontecimentos que tenham... Associas alguma coisa, neste caso ao projecto?

[I] Associo, mas... Sim associo. Estava, por exemplo, quando estávamos a fazer ambos os projectos e íamos alterar, na LEIC, lembro-me de um colega nosso levar sopas instantâneas, tinha uma resistência e uma caneca uma resistência que se ligava à tomada, ia à casa de banho e enchia aquilo de água, aquecia, punha as sopas instantâneas, eram Maggi ou Cup'a'Soup e o pessoal alimentava-se assim... Íamos ao pavilhão de Civil e bebíamos café, porque aquilo eram muitas horas seguidas...

[R] Então só para terminar sobre este documento, se tivesses que o encontrar agora...

[I] (interrompendo) Já agora! Lembro-me que o Cup'A'Soup era uma coisa mais ou menos recente. Quer dizer, não era recente mas estava a ter um bocadinho de fama nessa altura.

[R] Ok. Estava eu a dizer, se agora alguém te pedisse: 'olha, mostra-me aí o teu projecto de IHM', neste caso, fazes ideia de onde é que ele está ou nem por isso?

[I] Eu há pouco tempo tentei aceder e não sei se acedi à última versão. Eu tinha na minha área do Camões que agora é o Mega, eu tinha lá esse projecto, mas não sei se aquilo é a última versão. E por acaso eu fui lá e achei piada porque não estava a funcionar porque as maiúsculas tinham passado todas para minúsculas ou uma coisa assim qualquer. Não sei o que foi isso, se foi por ter mudado do camões... Não sei! Por qualquer razão, os ficheiros chamavam outros ficheiros com minúsculas e não correspondiam as maiúsculas com as minúsculas, então... Eu acho que aquilo

nem sequer era a última versão. A última versão nem sei se a guardei no disco lá em casa. Ou está no disco lá em casa ou está na área. A última versão.

[R] Ok. Então para terminarmos isto vamos só pensar num terceiro ficheiro, e este é mais rápido de descrever. Desta vez é um ficheiro não feito por ti. Uma coisa que recebeste de outra pessoa ou que foste buscar a qualquer lado. Por exemplo um artigo, uma página que tenhas consultado...

[I] Recente, antigo?

[R] Pode ser relativamente recente...

[I] Tem que estar em formato electrónico, obviamente...

[R] Sim...

[I] Sei lá...

[R] Um artigo qualquer que tenhas lido para preparar...

[I] Por exemplo... Foi numa cadeira do quinto ano, um artigo... Quer dizer, houve vários. Por acaso não me estou a lembrar de nenhum em especial. Tenho que me lembrar de um em especial...

[R] Era melhor, mas senão, se te lembrares de uma classe de artigos que no fundo, na tua cabeça sejam todos semelhantes, vem dar ao mesmo, portanto... Pensar num em concreto é só para focar a descrição que vais fazer.

[I] Ah, sim... Ok, estou-me a lembrar de um. Numa cadeira do 5o ano, se não me engano... Um artigo... Portanto, estava a fazer um survey para a cadeira. Tinha que procurar artigos e ver sobre um determinado tema. Queres o tema?

[R] Sim, já agora.

[I] Era Wearable Computers. E tive que procurar na net vários artigos e lembro-me de encontrar um que era sobre... Estava até numa forma esquisita. Se não me engano estava em LaTeX. Eu não conhecia bem o LaTeX na altura... Portanto, eu vi o texto e tinha uns caracteres, umas barras esquisitas, que eu julgo agora que era LaTeX. Aquilo era sobre interfaces neuronais para Wearable Computers, mas era a ideia, não estava nada concretizado.

[R] E de onde é que o tiraste? quem era o autor?...

[I] Eu tirei-o de um site no MIT, de algures no MIT e lembro-me que a pessoa que o escreveu, estava lá escrito, era um aluno de licenciatura se não me engano, devia estar no último ano, e teve 'A' naquele artigo. Lembro-me disso, não me lembro do nome do gajo. Teve 'A' e era do MIT... Não me lembro do nome dele.

[R] Portanto, mais uma vez, a pergunta do costume. Assim do mundo em geral associas alguma coisa a esse artigo ou não há assim nada em particular, na altura em que o estiveste a ler ou que o estiveste a tirar aconteceu alguma coisa?

[I] Mas do mundo... Quando me fazes essa pergunta não sei se estás a... algo que interaja com o artigo ou algo que...

[R] Não, algo que aconteça na mesma altura. Quando eu estava a ler este artigo, começou a

Guerra do Golfo. Por exemplo, uma coisa assim, mais longe...

[I] Lembro-me, por exemplo, que um colega meu... Era individual, esse trabalho... Um colega meu estava a escrever um survey... Supostamente cada um escrevia um survey sobre um tema diferente, e descobri falando com ele que o tema dele estava muito próximo do meu. Ou seja ele estava a ter uma abordagem muito próxima da minha, e eu da dele... Ou seja, o tipo de artigos que estávamos a ir buscar era muito comum. Já tínhamos ido buscar artigos muito comuns. E pronto! Isso não me agradou muito nem a mim nem a ele porque íamos estar a fazer uma apresentação de coisas semelhantes. O dele era sobre computação afectiva mas ele, no fundo, foi buscar a computação afectiva ligada quase só a Wearable Computers. E eu estava a ver Wearable Computers ligados com computação afectiva e a ver cenários com utilizadores. E pronto. Decidi fazer uma abordagem mais... Isto no survey no qual usei esse artigo. Lembro-me também, por exemplo, não sei se é relevante, mas já agora... Lembro-me que tinha feito há pouco tempo a viagem de finalistas, porque lembro-me quando eu descobri isso com ele, ele estava com a sua recente namorada que tinha conhecido na viagem de finalistas (risos)... Pronto... Eu tinha feito a viagem de finalistas mesmo há pouco tempo...

[R] Só para terminar... Já tinha dito que era para terminar mas passou-me isto por alto há pouco... Há mais algum artigo no survey de que te lembres? Dentro dessa survey em particular. Às vezes há aquele artigo que toda a gente da área lê... Lembras-te de algum? No fundo, alguma coisa que tenhas lido mais ou menos nesta altura em que estamos a falar...

[I] Lembro-me de um que encontrei na Net cuja fotografia que lá vinha, a imagem, já tinha visto nas aulas teóricas dessa cadeira, que... onde essa fotografia era um homem de barbas com uma série de coisas de Wearable. Tinha um teclado aqui algures. Tinha umas coisas também nos óculos... Portanto, essa fotografia vinha lá e eu descobri isso na Net e depois é que vi que a fotografia era a mesma. Sinceramente não tinha lido bem o artigo nas aulas teóricas...

[R] Mas era sobre o mesmo tema, portanto...

[I] Era. Era sobre wearables e usei também... Essa fotografia não a usei, porque já tínhamos usado nas aulas teóricas... Mas também usei no survey.

[R] Não te lembras do nome do autor?

[I] Não. Eu para nomes sou muito mau...

[R] Ok! Ótimo! Não sei se queres acrescentar alguma coisa?

[I] Não.

## Interview 2

[Researcher] A ideia era pensarmos em primeiro lugar num documento qualquer em que esteja a trabalhar recentemente, uma carta, um documento qualquer e, no fundo, dizer-me tudo aquilo de que se lembra sobre o documento, não só em termos do documento em si, da interacção que teve com ele no computador, mas, eventualmente, outras coisas que associe, outros documentos que



tenha usado para fazer aquele, pessoas a quem tenha enviado esse documento. Assim informação mais abrangente...

A primeira coisa talvez seja mesmo pensar num documento concreto, para focar a história e depois descrevermos o documento.

[Interviewed] um documento em concreto!/? Isso é um bocado complicado... move-se tanta coisa aí...

[R] Um qualquer... Uma carta qualquer que tenha escrito para o [local de trabalho], uma coisa qualquer em que tenha trabalhado recentemente. Não é preciso que seja hoje ou ontem. Nas últimas semanas... uma coisa qualquer. Não importa muito qual é o documento em si. É preciso é escolher um para focar...

[I] (interrompendo) E informação de dois anos, isso conta? Como é que se recupera?

[R] Não. Preferia mesmo um documento muito concretamente.

[I] Olhe, eu acabo de pedir um... Tive que recorrer a outra pessoa para recuperar o documento, mesmo...

[R] Então pronto, pensando nesse documento, do que é que se lembra do documento? Quando é que foi a última vez que lhe mexeu?

[I] Concebi-o há cerca de hora e meia. Actualizei-o em Fevereiro de 2000 e agora precisava de actualizá-lo porque tem que ser actualizado todos os anos e precisei de recorrer a outra pessoa... perdi o disco e para ver se o recuperava tive que recorrer a outra pessoa. Por acaso tive a sorte de essa pessoa guardar a documentação também.

[R] Portanto, havia outra pessoa à qual o documento tinha sido enviado e que...

[I] (interrompendo) E por acaso é um daqueles documentos que são partilhados, que é dos recursos humanos, por acaso. Só que quem actualiza sou eu. O responsável dos recursos confirma ou não uma ou outra informação. E há outros documentos que são produzidos por ele e eu simplesmente confirmo mas também não movo a informação. Portanto, cada qual tem a sua organização e funciona assim. Por acaso ele tinha o meu mapa... Agora vou ter é que actualizar o meu.

[R] E... Portanto, pelo que percebi foi um documento que foi feito só por uma pessoa.

[I] sim, sim. Foi feito só por mim.

[R] Não houve mais ninguém... E houve outras coisas que tivessem sido consultadas na altura para fazer esse ou...

[I] Fui consultar ali a secção de pessoal docente quase toda. É a vossa informação docente.

[R] Portanto houve mais documentação que eles enviaram e que foi reunida num único documento. A informação, se voltasse a pedir à secção de pessoal, eles voltavam a disponibilizar. São coisas que...

[I] Há documentação que eles disponibilizam sim, todos os anos. A pedido.

[R] Então e depois esse documento foi usado para quê? Foi enviado só para o responsável do

pessoal docente? Foi...

[I] Só, só isso... É um instrumento mesmo de trabalho.

[R] E trabalhou nele aqui?

[I] Exacto. Aqui mesmo.

[R] Não houve às vezes um dia em que tenha levado também para casa para acabar qualquer coisa?

[I] Não... Essa política não existe. (risos)

[R] Ora bem... O tema do documento era a distribuição do serviço docente?

[I] Não... No fundo, era a situação dos docentes.

[R] Os créditos, a categoria profissional?

[I] Tempo na categoria, contactos, dispensas.

[R] E assim da sua vida pessoal, no sentido de uma coisa que até podia ter na agenda. Foi trabalhado na véspera do aniversário de alguém, ou no dia em que foi a uma loja qualquer? Pronto, alguma coisa de que se lembre que tenha feito mais ou menos na mesma altura do que o documento?

[I] (pausa) Não... não me é possível lembrar...

[R] E do mundo em geral? Foi no dia em que começou a guerra no golfo? Houve assim algum evento... Foi no dia em que houveram as eleições, talvez? Em que...

[I] Não... Por acaso está próximo do meu aniversário mas nenhum acontecimento especial...

[R] E qual era o tipo de documento?

[I] Tipo? Formato?

[R] Sim, foi feito no Word...

[I] Excel.

[R] Excel. E houve alguma tarefa especial... pronto, alguma coisa que tenha tido que utilizar antes de fazer esse mapa no Excel teve só que fazer algumas contas? Teve mesmo que telefonar a alguém para esclarecer alguma situação...

[I] Houve que recolher toda a informação sobre os docentes... todinha.

[R] Portanto isso implicou contactos com as pessoas, contactos com a secção de pessoal...

[I] Recolha de muitas listagens, mesmo, de informação. Quase a situação dos docentes desde que entraram para o técnico.

[R] Assim em concreto houve alguma coisa mais difícil?

[I] Foi lidar com aquelas listagens enormes... Há docentes com mais de 20 anos de carreira e é

uma pilha de folhas...

[R] E está guardado aonde? Está no tal disco que...

[I] Que estou a ver se recupero!

[R] Mas agora já voltou a ...

[I] Sim, quer dizer... São dois anos de informação perdida...

[R] ... que agora é preciso actualizar outra vez (assentimento). Então pronto. Quanto a este acho que estamos conversados. Agora gostava que fizéssemos o mesmo exercício que também tenha sido a [entrevistada] a fazer mas mais recente. Uma coisa que tenha sido criada mesmo nos últimos tempos. Nos últimos dias, na última semana.

[I] Ok.

[R] E era mais uma vez, pronto. Pensar num documento em concreto e depois descrever assim por alto do que é que se lembra dele. Mais ou menos nestas vertentes que estivemos a falar...

[I] O apoio ao ensino é aquilo em que tenho estado a trabalhar agora.

[R] E então o que é que...

[I] No fundo esse foi mais simples porque no fundo foi logo deferido pelo responsável o tipo de informação que ele pretendia e foi simplesmente criar o documento. Montá-lo, escrevê-lo.

[R] E mais uma vez foi preciso ir buscar informação a outros documentos?

[I] Não, essa informação vai vir é depois. Estabelecidos os parâmetros que se deseja, não é, para a gente depois recolher a informação toda. Neste caso é diferente.

[R] E depois quando isso tiver terminado será... É para entregar a uma pessoa em concreto? É para enviar para todo o departamento?

[I] É para distribuir aos alunos que pretendam ser candidatos ao apoio ao ensino.

[R] E mais alguém está a trabalhar nisso?

[I] Só eu e o [responsável].

[R] E mais uma vez está a ser feito aqui...

[I] Concebido aqui.

[R] E este ainda não está perdido, está ali no disco daquela máquina...

(risos)

[I] Exactamente.

[R] E mais uma vez, houve assim alguma coisa em concreto de que se lembre... Algum percalço a fazer o documento e que foi preciso...

[I] Não...

[R] Foi tudo sem problemas de maior... E já foi trocado com alguém para ler, para validar alguma coisa?

[I] Sim. Ele já o validou.

[R] Portanto foi enviado por correio electrónico, ele leu, devolveu...

[I] (assentimento)

[R] E o documento tem várias versões ou...

[I] Uma só versão. Só houve uma versão, mesmo.

[R] Ok. Acho que estamos conversados. Para terminar falta voltar a fazer este exercício mas desta vez uma coisa ligeiramente diferente, que é um documento que não tenha sido o [entrevistado] a fazer. Uma coisa que tenha vindo lá de cima da secção de pessoal, uma parte do Diário da República. Uma coisa qualquer que tenha recebido já há algum tempo, nos últimos meses, talvez e que tenha eventualmente sido utilizada para qualquer coisa. Imagine que agora precisava outra vez 'daquela carta que veio da secção de pessoal e avisava de qualquer coisa', por exemplo. O que é que se lembra de... Mais uma vez vamos pensar num em concreto e repetir o exercício. Dizer do que é que se lembra, se lembra ainda quando é que foi enviado, como é que foi enviado, quem é que o enviou... Dizer que informação é que lhe vem à cabeça...

[I] Uma situação em concreto? Houve uma ainda por acaso a semana passada. Posso-lhe dar a situação? O [professor] fez a agregação há quase quatro anos e de facto andámos aqui à procura da documentação e chegámos à conclusão que essa documentação até foi formulada por outro departamento e tivemos que recorrer realmente aos serviços do Técnico para recuperar a tal documentação. Podia até estar no DR... devia estar no DR mas tivemos que recorrer realmente ao Técnico para avisar a outra unidade para nos enviar essa informação.

[R] E mais ou menos em que altura é que foi isso? Quando é que foi preciso andar à procura disso?

[I] Foi a semana passada.

[R] E depois essa informação quando veio... Está guardada aonde, agora?

[I] Agora está no meu arquivo sobre os docentes.

[R] E mais uma vez aquelas perguntas um bocado mais abstractas mas que eu tenho que fazer. Há alguma coisa do mundo em geral de que se lembre quando pensa dessa situação em que estava aqui a trabalhar. Sei lá. Se calhar quando estava à procura da informação viu uma notícia qualquer sobre a guerra no Iraque? Sei lá, qualquer coisa assim de que se lembre...

[I] Não, por acaso não houve nenhuma notícia em particular...

[R] E houve alguma dificuldade em especial na produção desse documento, ao juntar essa informação toda?

[I] Não eu... é na pesquisa da mesma, que ainda está a ser definido. A estrutura do mesmo já está definido. Agora é montar o júri. Definir o próprio júri, porque o documento é um documento standard. O que houve foi a dificuldade da pesquisa, mesmo.

[R] É um documento standard quer dizer que existe um modelo do documento no qual pegou e agora está a...

[I] Exactamente. Um modelo reitoria e...

[R] Ok. Pronto, já enviou o documento a alguém ou ainda...

[I] Já, já. Já o enviei. O que pretendia mesmo era saber a composição desse júri. Só, simplesmente isso.

[R] Pronto. Acho que estamos conversados. Não sei se se lembra de mais alguma coisa assim na linha do que temos estado a falar que ache que seja interessante dizer?

[I] Não. Não me lembro de nada.

[R] Estamos conversados. Muito obrigado.

## Interview 3

[Researcher] Para começar a ideia era identificar, pensar num documento em concreto em que tenha trabalhado nos últimos dias, nas últimas semanas. Uma coisa qualquer que tenha escrito no computador agora recentemente. Começar por pensar mesmo numa coisa em concreto, AQUELE ficheiro, AQUELA carta, e agora dizer, no fundo, tudo aquilo que nos lembramos sobre esse documento. Sobre o que é que tratava, porque é que foi feito, a quem é que foi enviado. Qualquer coisa assim que venha à memória.

[Interviewed] Dos últimos documentos que fiz no computador... (pausa)

[R] Não precisa de ser uma carta, pode ser qualquer coisa que tenha sido feita no computador...

[I] Por acaso uma coisa... Não precisa de ser uma carta, não é? A última coisa foi um arquivo que foi guardado no computador de fotografias de uma actividade da escola. Foi a última coisa que eu fiz no computador.

[R] E por exemplo, que actividade em concreto é que foi.

[I] Foi metê-las... Seleccioná-las, essas fotografias... Bem, primeiro foi descarregar a máquina para lá. Depois foi seleccionar as que queria, ordená-las e depois guardá-las numa pasta de Fotos de Actividades.

[R] E esta actividade em concreto o que é que é?

[I] Esta actividade em concreto?

[R] Sim. O que é que foi fotografado?

[I] Foi o 'Jogo do 24'.

[R] Ok. E quando é que isso foi.

[I] Foi a semana passada.

[R] E foi lá na escola ou...

[I] Foi lá na escola.

[R] E isso depois já foi mostrado a alguém, foi enviado a alguém? impresso e posto em exposição? Ou foi só guardado e...

[I] Por enquanto foi... Não. Já foi impresso e posto em exposição no Polivalente.

[R] Ok. E houve alguém mais que tivesse ajudado a fazer isso?

[I] Duas pessoas pelo menos.

[R] E sabemos quais são...

[I] O [Pessoa 1] e o [Pessoa 2]

[R] E que é que tirou as fotografias mesmo?

[I] O [Pessoa 2]

[R] Houve alguma outra coisa que fosse preciso consultar, que fosse preciso... Quando se tiveram a tirar as fotografias foi preciso, por exemplo, consultar algum manual, para saber como se tiravam as fotografias? Ou para imprimir? As fotografias colocaram-se noutra documento já ou foram só impressas as fotografias em si?

[I] Não. Porque foi com a ajuda deles e não foi preciso consultar nada.

[R] E agora, vão ficar simplesmente guardadas onde estão ou vão ser utilizadas...

[I] Algumas vão ser utilizadas numa edição do jornal da escola, vão ser usadas também no álbum de actividades da escola... algumas, seleccionadas. Pronto, para já é isso, para além das que foram expostas no Polivalente em que os miúdos escolhem depois as que querem e quando é possível dá-se-lhas.

[R] E estão guardadas no computador?

[I] Estão guardadas no computador.

[R] E assim algum percalço que tenha havido... Algum acontecimento... Não necessariamente das coisas que foram feitas lá no computador mas outra coisa. Entrou alguém no gabinete na altura, ou foi naquele dia em que foi preciso ir buscar frango não sei aonde ou em que se foi meter gasolina. Há assim alguma coisa que venha à memória, assim de repente?

[I] Quer dizer, relacioná-las com...

[R] Sim. 'O dia em que estava a guardar as fotografias foi o dia em que entrou alguém' ou 'foi o dia em que depois fui carregar a bateria do telemóvel...'

[I] Ah! Por acaso foi o dia em que o meu carro ficou sem bateria!

[R] E assim mais do mundo em geral, também? Foi o dia em que chegou uma circular importante,

foi o dia em que...

[I] Foi o dia em que nós decidimos não participar no jogo do Sunny Delight por falta de verbas para o transporte e que criou uma certa confusão com os telefonemas para Portalegre. Não me lembro assim de mais nada. De resto foi um dia tranquilo.

[R] Ok. Pronto. Então agora que já despachámos este documento, estas fotografias, que era assim mais recente, vamos tentar repetir este exercício mas agora para uma coisa assim mais... Que tenha sido feita há um ano, há dois anos. Mais uma vez, começar por identificar um documento em concreto, que tenha sido produzido nessa altura, e depois ver do que é que ainda nos lembramos dele. Uma carta que tenha sido enviada, algum documento que tenha sido preparado porque fazia falta naquela altura na gestão da escola. Assim alguma coisa há um ano, dois anos.

[I] Há um ano...

[R] Ou mais tempo, eventualmente, se for mais fácil lembrar, mas já mais distante.

[I] (pensando) Um caso concreto... o preenchimento dos mapas da inspecção. É o que me lembro de ter feito.

[R] E foi há quanto tempo, mais ou menos?

[I] Para aí há 4, 5 meses... Pode ser a elaboração da Tesina?

[R] Foi há quanto tempo, mais ou menos?

[I] Foi há um ano e pico. Para aí há 14 meses... 15 meses!

[R] Portanto... E era, o documento da Tesina do doutoramento em Sevilha.

[I] Em Sevilha. Educação e Sociedade.

[R] Ok. E foi produzido aonde? Trabalhou-se aqui em casa e depois foi levado para lá? Foi todo lá feito?

[I] Não. Foi feito lá.

[R] E está guardado aonde agora? Em que computador é que isso está? Ou em que disquete?

[I] Está guardado em... (hesitando) em disquetes. Em disquete.

[R] E teve sempre lá guardado ou foi copiado depois para lá?

[I] Esteve sempre lá guardado. E guardado em disquete.

[R] Pronto... O autor é evidente quem foi. E não houve assim algumas partes ou...

[I] Houve, houve várias pessoas que ajudaram. A [Pessoa 3], a [Pessoa 4], tu... (risos)

[R] e foi enviado por correio electrónico, por algum mecanismo desses ou só foi depois entregue...

[I] Não. Só foi enviada uma parte por correio electrónico para a [Pessoa 5], a parte da apresentação, que mo devolveu em Powerpoint. Foi enviado por email, sim. Portanto, na primeira fase.

[R] Portanto houve várias versões?

[I] Houve uma primeira parte que foi enviado (ao orientador) para ser corrigido, para ser retificado no que ele entendesse. Depois foi devolvido. Depois voltou a ser enviado outra parte, já com outras correções. E depois, já na fase final de apresentação foi enviado para a [Pessoa 5], e foi devolvido em Powerpoint.

[R] Houve assim necessidade de consultar outros documentos para fazer esse? Ou consultar outros ou depois desse foram feitos outros...

[I] Para fazer esse documento? Houve necessidade de consultar a bibliografia.

[R] Ok. Há assim alguma coisa que, em particular, tenha sido mais fulcral. Às vezes aquelas referências obrigatórias neste género de trabalhos, ou um trabalho que fosse mesmo importante, ou foram todos assim mais ou menos...

[I] Foram todos. Nenhum se destacava...

[R] Houve assim, mais uma vez, algum percalço com a construção do documento?

[I] Houve toda a questão que se relacionava com a língua em que estava a ser elaborado. Portanto, ele começou por ser feito em português porque havia a garantia de que poderia ser aceite. Depois houve que modificá-lo na parte de apresentação que passo... teve que ser em espanhol. Houve alguma dificuldade em alterar tudo isso. E depois acabou por sair metade... portanto... o conteúdo propriamente em português mas toda a parte... tudo o que era de... índice, e de introdução teve que ser feito em espanhol.

[R] era um documento Word, ou tinha também tabelas de Excel, ou...

[I] Era um documento Word. Lembro-me de algumas imagens que foram introduzidas mas também não sei explicar o que foi que aconteceu. Sei que depois não deu para... Os espaço ocupado por essas imagens. Tive uma pessoa que me ajudou, foi o [Pessoa 2], o sistema, a maneira que usou para tirar essas imagens acabou por ocupar demasiado espaço e depois foi muito difícil inserir no trabalho. Isso lembro-me. Agora mais percalços não me lembro.

[R] Pronto. E assim da vida pessoal? Sei lá... foi na altura do aniversário de alguém? Há alguma coisa que venha agora à memória?

[I] Novembro... Dezembro... Foi próximo do Natal... Foi o... o tempo... portanto... foi encurtado. Tornou-se muito exíguo, a dada altura. Havia a possibilidade de ser mais tarde mas não me lembro já porquê foi quase que um ultimato, para ser até aquela data. E houve também um incidente em relação a um dos professores. Um que fazia parte do tribunal de tese que primeiro... Que era concretamente o presidente do tribunal, se não me engano, que primeiro aceitou mas depois à última da hora recusou-se a fazer parte do tribunal porque não quis aceitar o trabalho em língua portuguesa e houve que modificar algumas coisas.

[R] Pronto. Então quanto a este estamos despachados. Falta só um, pensar num documento. Desta vez para variar é um documento não que tenha sido criado por nós mas uma coisa que tenha sido recebida de outro lado qualquer. Um ofício que tenha sido recebido, uma carta importante, um artigo que tenha tido que ser lido...



[I] Pode ser a candidatura do PRODEP?

[R] Pode ser... Desde que tenha sido um documento recebido de outra fonte... portanto...

[I] Pode ser essa questão da inspecção... Os mapas da inspecção que tínhamos lá e tinham que ser preenchidos...

[R] E quando é que vieram?

[I] Vieram, portanto, a primeira versão chegou à escola em Julho.

[R] E vieram de onde?

[I] Da inspecção geral de educação.

[R] E agora está armazenado aonde? Está guardado aonde?

[I] Ficou no disco do computador e era suposto que aqueles mapas que vinham com aquele documento fossem preenchidos com os dados de elaboração de horários e do lançamento do ano lectivo e exportados. Só que acabou por não ser.. Aquilo era um sistema... trabalhava-se em Excel. Ele tirava o próprio... O programa fazia lá as contas que sabe fazer, só que havia lá uma deficiência qualquer no processo de exportação do documento que acabou por ser efectivamente preenchido mas a parte que... aquilo que se pretendia que fosse inovação que era os dados chegarem em tempo real à inspecção de modo a serem trabalhados por eles antes... Essa parte acabou por não resultar. Nos preenchamos com um suporte de papel que acabaram por ficar para serem examinados quando a inspecção viesse à escola ver os horários.

[R] Portanto eles foram enviados por correio electrónico?

[I] Não. Era um sistema diferente. Não era propriamente correio electrónico. Eles eram feitos e exportados no momento, tipo on-line, sei lá... Era um sistema diferente do correio electrónico.

[R] Portanto era no computador mas...

[I] Eram feitos e à media que estavam a ser feitos os dados eram vistos lá.

[R] Quem é que mais trabalhou nisso?

[I] A [Pessoa 6], a [Pessoa 7] e eu.

[R] Ok. Portanto o objectivo eventual é que a inspecção veja aquilo.

[I] Sim. Que tivesse acesso a eles no momento em que estavam a ser lançados os dados...

[R] Pronto. E mais uma vez, para terminar, aquela pergunta do costume. Assim do resto da vida em geral, do mundo, há alguma coisa que fique associada a esses dias, em que esse documento chegou? Algum evento em particular? Alguma coisa que tenha acontecido lá na escola? Ou não há nada assim que venha à memória?

[I] Não. Que me recorde não.

[R] Então, não sei se há mais alguma coisa que queira dizer, que tenha vindo à cabeça entretanto? Senão estamos conversados...

[I] Não... Mais nada.

[R] Ok. Terminámos. Obrigado.

## Interview 4

[Researcher] Para descobrirmos quais são os tais elementos que as pessoas usam o que eu vou pedir é para pensar num documento concreto, mesmo num ficheiro concreto, numa coisa que tenha sido feita nos últimos dias, na última semana, talvez. Assim uma coisa mesmo recente. Começar por identificar um documento em concreto e depois, no fundo, Dizer tudo o que vem à memória sobre esse documento. Não só sobre o documento em si, mas coisas que tenham acontecido enquanto foi feito, quando ele foi feito. Portanto, começar por identificar um documento e a partir daí identificar a história.

[Interviewed] Ok. Está bem. Já tenho um documento identificado.

[R] Pronto. Então o que é que se passa com esse documento?

[I] Esse documento é um documento em Excel. Pretende fazer um resumo e fazer cálculos de informação que está dispersa noutros ficheiros e vai servir de análise... O conteúdo dos quadros feitos nesse ficheiro irá ser para retirar algumas conclusões. Ou seja, nesse ficheiro a única coisa que eu faço são cálculos, reuno informação de vários ficheiros, faço alguns cálculos e ele não tem um fim definido...

[R] Vai ficar disponível para quem precisar eventualmente...

[I] À partida só quem vai precisar sou eu, mas ele vai-me servir de base para as decisões. Eu quero saber, esta decisão eu tomei porquê? Então eu vou pegar nesse ficheiro e vou ver: utilizei este ficheiro para calcular isto, calculei desta maneira, a hipótese é esta. Por exemplo, fiz determinados conditional formatting para saber situações que são mais...

[R] Exacto. Chamar logo a atenção quando ocorrerem.

[I] Exactamente. Chamar logo a atenção mais facilmente, enfim.

[R] Ok. Portanto, como já foi dito, foi preciso recorrer a outros ficheiros, outros documentos, na realidade. Coisas que já tinham sido feitas anteriormente pelo [Entrevistado], ou foi material que veio de outras fontes, que foi preciso pedir a outras pessoas...

[I] O material estava disponível. Não tinha sido feito por mim. Portanto, ele estava na rede, estava disponível para todos e eu fui buscar esses ficheiros, alguns deles bastante antigos, e fui buscar os ficheiros que já estavam disponíveis. Nenhum dos outros tinha sido feito por mim.

[R] E houve assim algum que se destacasse? Que fosse no fundo o padrão a partir do qual este novo documento foi feito, ou de onde grande parte da informação acabou por ser...

[I] Há três. Três que são bastante importantes.

[R] Muito bem. Houve mais alguma pessoa a trabalhar no documento, ou eventualmente um

revisor, alguém que tenha lido?

[I] Sim, sim. Já houve o meu chefe directo, já estive comigo a olhar para o documento.

[R] Portanto, trocaram o documento por email, provavelmente, ou foi mesmo...

[I] não, para podermos estar a ver aquele e os outros que serviram de base estivemos ali mesmo no meu local... nem sequer lhe enviei. Ele não vai utilizar. Foi no sentido de ver se aquilo prestava, se estava...

[R] Portanto, depreende-se daí que o documento foi feito no local de trabalho. E está neste momento armazenado no disco do computador no local de trabalho. Existe alguma cópia que tenha sido feita, alguma cópia de segurança numa disquete ou noutra máquina?

[I] Como é tão recente... ainda não. Eu normalmente faço todos os meses uma cópia dos meus ficheiros de trabalho. Como isto está a ser feito durante estas duas últimas semanas, é o documento mais recente com que estou a trabalhar ainda não...

[R] Ainda não foi... mas eventualmente será feita uma cópia.

[I] Sim. Eu normalmente faço backup dos meus ficheiros.

[R] Não sofreu ainda muitas alterações que mereçam versões? Não existem ainda versões do documento. É evidente que foi trabalhado e sofreu alterações, mas não existem mesmo versões, do tipo: 'a versão do ano passado, a versão deste ano...'

[I] Neste documento não. Nem tenciono fazer isso. O que eu penso fazer é dentro deste mesmo documento posso ter várias sheets, como tenho. Uma sheet por onde comecei, uma outra que tem um determinado trabalho e por aí fora, mas estou a deixar tudo lá dentro.

[R] Ok. Então só faltam algumas perguntas que são um bocado mais genéricas mas que são também importantes. A primeira é em relação às tarefas que foi preciso fazer para criar o documento, a recolha da informação, tudo isso. Houve assim algum percalço, alguma coisa que venha à memória, ou decorreu mais ou menos de forma linear e não houve assim nada de...

[I] Creio que decorreu de uma forma linear. O próprio nome do ficheiro é um nome muito... chama-se 'análises'. Portanto, serve exactamente as análises que depois vão sustentar a tal decisão que eu vou tomar.

[R] E em termos de ocorrências da vida pessoal, coisas que eventualmente até pudessem estar marcadas na agenda. Foi feito por exemplo no dia do aniversário de alguém, ou no dia em que se acabou a bateria do carro. Há assim alguma...

[I] Na data de criação desse ficheiro?

[R] Sim, ou nos dias em que o ficheiro tem sido trabalhado. Não necessariamente no dia da criação mas... Alguma que coisa que venha à memória. 'Este documento foi realmente aquele em que eu estive a trabalhar naquele dia em que cheguei atrasada porque tive um furo', por exemplo. Assim alguma coisa... vem alguma coisa à memória assim desse tipo?

[I] Não. Houve situações do género: estou a trabalhar no documento e ter que interromper. Por exemplo tive que interromper porque vim para as aulas. Estava a trabalhar no documento e deixei

uma nota, um comentário, a dizer: 'falta-me fazer isto e aquilo...'

[R] O ponto da situação...

[I] ... para saber amanhã como é que eu... por onde é que eu vou começar.

[R] E assim mais do mundo em geral. Há alguma coisa neste últimos dias que... alguma notícia até, na televisão. uma coisa que tenha acontecido ao vizinho? No geral há assim mais alguma coisa que se saliente?

[I] Histórias destes últimos dias? Desde aniversários, tenho tido. Outro tipo de trabalhos que também tenho feito. Isto é uma tarefa que tem um determinado calendário, e tenho que estar a intercalar com outras o que significa que embora este seja um trabalho de fundo há outras tarefas que vou ter que ir desenvolvendo. Mas assim um acontecimento que tenha marcado os últimos dias penso que não.

[R] Quanto a este documento estamos conversados. Agora vamos repetir este exercício mas desta vez para um documento que tenha sido criado já há mais tempo... Um ano, dois anos, talvez. Mais uma vez começar por identificar um documento em concreto e depois dizer aquilo que vem à memória...

[I] Ok. Também já me lembrei de um.

[R] E quando é que foi...

[I] Talvez... Em Março... Fevereiro-Março do ano passado.

[R] E então, o que é que...

[I] Ocorreu-me esse ficheiro porque foi um trabalho que foi a primeira vez que fiz um trabalho numa equipa tão alargada de pessoas. Enquanto que o ficheiro de que eu falava há pouco, o 'análises' é um ficheiro que fui eu que fiz, e está todo a ser pensado por mim, realizado por mim da forma que eu quero porque vai ser utilizado por mim. Esse outro éramos uma equipa de trabalho que tínhamos uma determinada informação para divulgar. O formato do ficheiro não foi apenas decidido por mim. A informação que consta nele também vem de outros ficheiros não só no Excel mas também no Access e toda essa montagem que foi feita não foi pensada só por mim. Já tinha havido antes algo parecido, foi reestruturado e também foi um ficheiro de análises. Mas nesse aí pretendia resumir, numa folha ainda que bastante vasta, uma folha com bastante informação, mas que pretendesse resumir de uma forma mais um menos concisa determinado tipo de informação e em sério, portanto, vários anos, como é que aquilo evoluiu.

[R] Portanto, foi feito mesmo no local de trabalho e teve outros co-autores, mesmo a trabalhar no documento em si...

[I] Mesmo a trabalhar no documento em si houve várias pessoas.

[R] Ainda é possível recordar quem foram essas pessoas. Se fosse preciso dizer: 'preciso de falar com uma pessoa que tenha trabalhado...'

[I] Sim, sim. Claro, claro. Todas elas.

[R] Eventualmente o documento em si foi trocado por correio electrónico entre as pessoas.

[I] Como era um ficheiro muito pesado, era bastante grande e tinha... tinha mapas, por exemplo, foi colocado na rede e só as pessoas que estavam a trabalhar naquele trabalho tinham permissões para aceder ao ficheiro.

[R] E neste momento está armazenado aonde?

[I] Está armazenado na rede. Uma vez que é um trabalho que já ficou fechado, ele está disponível para todas as pessoas que o quiserem ler. Ele está apenas com permissões de ler para todos, quer seja uma pessoa que tenha participado na equipa ou não, mas existe também um backup dele em CD para caso de acontecer alguma coisa.

[R] Já vimos que foi uma ferramenta de análise que foi usada para tomar várias decisões... E qual era o tema?

[I] Era... agregados macro-económicos.

[R] Ora... Esse provavelmente já teve várias versões...

[I] É assim... teve várias versões se bem que elas não foram guardadas. Ou seja, à media que fora sendo guardados os dados, as tais macros, os tais relatórios, aquilo foi sendo tudo alterado e foram sendo sobrepostas.

[R] Guardou-se apenas a última...

[I] Não era necessário ficar com o historial naquele ficheiro porque como aquilo vinha de vários sítios, os outros de onde eles vinham é que, esses sim...

[R] Já tinham esse historial e ....

[I] Exactamente.

[R] Portanto... mais uma vez aquelas perguntas mais genéricas. Quando se estava a fazer houve assim alguma tarefa particularmente difícil ou...

[I] Houve. Houve porque como aquilo vinha do Access, como eu disse há bocadinho, houve determinados... quando se fazia os External Data houve determinados procedimentos que não correram bem e a dada altura nós achávamos que estávamos com informação correcta, actual, e não estávamos. Porque o ficheiro que entretanto também já foi alterado exigia que fossem feitas tantas tarefas que se por acaso se tivesse esquecido de uma invalidava tudo. Portanto o ficheiro não estava muito usual. Não estava muito funcional. Então aconteceu isso. Aconteceu que pensávamos que tínhamos informação actualizada e quando começamos a perceber que havia algumas incoerências e fomos... bem, vamos começar por corrigir aqui o processo informático e ver se isto está mesmo tudo bem. E depois de já estarmos a analisar, de já termos algum trabalho feito verificámos que não estava bem.

[R] E mais uma vez, também da vida pessoal, ou do mundo em geral houve assim algum...

[I] Nessa altura fiquei a trabalhar muitas vezes à noite.

[R] Pois, isso nós lembramo-nos... (risos)

[I] Nomeadamente ao Sábado, também. Lembro-me perfeitamente de estar a trabalhar.

[R] Muito bem. Do mundo em geral houve alguma coisa que venha assim à memória? Nessa época aconteceu alguma coisa importante?

[I] Como aquilo era informação para ser utilizada por pessoas de outros serviços, do mesmo departamento mas de outros serviços, foi uma época que se caracterizou por muita interacção com muitas pessoas. Portanto, à medida que as coisas iam sendo terminadas nós íamos dizendo 'nesta altura esta informação pode ser consultada por vós', tínhamos muita interactividade.

[R] Portanto, trocas de telefonemas, trocas de correio electrónico com pessoas que estavam a utilizar...

[I] Exactamente.

[R] Pronto. Não sei se há assim mais alguma coisa sobre este documento que... que venha à memória...

[I] A não ser que ele entretanto foi refeito... Embora a ideia do documento seja muito boa, ele seja realmente... o formato em si e a visualização... muitas vezes nós não conseguimos, ao mesmo tempo que temos o ficheiro funcional, ter com muita informação por trás. O ficheiro é realmente muito visível, percebe-se muito facilmente. Uma pessoa que não perceba nada de economia nem que perceba o que é que aquilo... como é que se chegou até ali, sabe dizer: 'não, isto evoluiu desta forma, o que contribuiu mais foi isto...' É tudo muito perceptível, mas a forma como está construída é que é complicada, se calhar porque não somos de informática, não é, e quem construiu fomos nós, com os nossos conhecimentos... Exporta para Access, trás para cá, muda aqui, muda ali. Se calhar não somos muito... não temos se calhar uma componente técnica tão grande que nos permita ser mais rápidos e mais eficientes. É um ficheiro que tem bastante importância

[R] Para terminar, e agora já provavelmente mais rápido, voltar a contar tudo que vem à memória mas desta vez sobre um documento feito por outra pessoa. Uma coisa que tenha aparecido... Eventualmente um artigo, uma carta, uma circular... Qualquer coisa que tenha sido recebida há um mês, dois meses, ou há mais tempo...

[I] Neste trabalho que estou a fazer actualmente um dos ficheiros que lhe serve de apoio foi construído já há alguns anos por uma pessoa que nem sequer está já no [Local de Trabalho] e portanto peguei nele e na informação que ele tinha para tentar perceber-la e para tentar usar agora.

[R] E, portanto... há alguns anos. Dois, três? Tem alguma ideia mais concreta ou...

[I] Sim. Já há uns 3 ou 4 anos.

[R] E esse documento está... Bem, agora deve estar armazenado no disco local. Mas antes estava guardado aonde?

[I] Antes? Não sei se estava guardado no disco de alguém porque eu pedi-o. Ele não estava na rede e pedi esse ficheiro, portanto... quando comecei a fazer esta tarefa e fui à procura de informação que existia sobre este trabalho lembrando-me que havia uma pessoa que tinha feito determinado tipo de informação. Fui ter com essa pessoa e foi essa pessoa que me enviou o ficheiro. Portanto, se ela tinha guardado... Como ela me fez um forward podia até estar na caixa de correio...

[R] Na caixa de correio dela?

[I] Sim na caixa de correio dela, portanto não faço ideia de onde é que está.

[R] Portanto foi enviado por correio electrónico e é isso que nos sabemos da sua origem... É um ficheiro Word,...

[I] É Excel.

[R] É um Excel também... Foi utilizado como fonte de dados para o outro documento.

[I] Como fonte de dados, foi.

[R] Muito bem. Houve alguma coisa, alguma dificuldade na sua utilização? Foi preciso converter porque era um formato antigo, ou foi usado como outra fonte de dados normal?

[I] Foi preciso converter. Foi preciso converter porque o desenho que tinha o ficheiro... O propósito para o qual ele foi criado não era o mesmo que eu estava agora a querer usar, não é. Portanto, o formato não era igual. E não era igual também a desagregação. Tinha muito mais desagregação do que o que precisava. Tinha informação complementar de que eu também não precisei. Tinha arredondamentos como aquilo tinha as fórmulas que lhe fui retirar fiz algumas alterações a esse ficheiro inicial e tive também que começar por perceber era que informação é que estava lá, não é? Basicamente ler os quadros que estavam ali por detrás, não é? E alguns deles tinham ainda comentários e ver os comentários que tinham, alguma informação adicional,

[R] Perceber os dados que lá foram colocados...

[I] Exactamente.

[R] Pronto, então para terminar, outra vez aquela pergunta assim mais abstracta. Do resto do mundo e da vida há assim alguma... Foi preciso encontrar pessoalmente essa pessoal, num sitio fora do local de trabalho?

[I] A pessoa que fez o ficheiro não. A pessoa que fez o ficheiro não teve qualquer contacto com ela. Mas para aceder ao ficheiro, tive que... eu não sabia que existia ESTE ficheiro. Sabia que existia ALGUMA coisa feita naquela área. Não sabia da existência daquele. Nem sequer sabia que era um formato em Excel. Não sabia nada. Sabia que existia um trabalho feito naquela área, e fui ter com a pessoa que normalmente faz divulgações para o [Projecto] e perguntei-lhe se sabia onde é que estava, se havia informações sobre aquela área e disse-lhe a pessoa que tinha feito, porque eu sabia quem é que tinha feito. E que altura é que tinha sido feito nessa altura. E então a pessoa disse-me: 'tenho este ficheiro, tenho este, tenho este, tenho este.' Enviou-me um conjunto de ficheiros. Desse conjunto de ficheiros, eu peguei apenas num, que era aquele que interessava.

[R] Pronto. Mais uma vez, não sei se há mais alguma coisa assim que... que... venha à memória, que seja interessante dizer?

[I] Relativamente a este ficheiro não. Ele também pode ser... foi usado por outras pessoas porque no tal 'análises' eu usei-o... construí-o a partir de informação que constava nesse. Significa que o meu chefe por exemplo, quando estivemos a analisar esse também foi um dos ficheiros que olhamos. Não significa que tenhamos sido só os dois que ultimamente olhámos para ele. Não. Pode haver outras pessoas até porque ele está na posse de...

[R] ... de alguém que pode ter enviado a outras pessoas. Exactamente.

[I] ...a outras pessoas.

[R] Mas dentro aqui desta conversa penso que terminamos.

[I] Está bem.

## Interview 5

[Researcher] Para percebermos então quais são os tais elementos que nos vão ajudar a recuperar o documento aquilo que vamos fazer é identificar um documento concreto. Eventualmente será até só um ficheiro no computador. E um documento que se tenha trabalhado nos... Até hoje, talvez... Nos últimos dias, no último par de semanas, no máximo. Portanto um documento muito recente trabalhado agora recentemente. Uma vez identificado um documento, para ajudar a centrar a descrição, o que eu pedia era que descrevesse tudo o que vem à memória sobre esse documento. Eventualmente onde é que ele foi trabalhado, para que é que serve... tudo o que vier à memória sobre ele.

[Interviewed] Ok... Por exemplo eu tenho... Como faço gestão de projectos e trabalho com muitos projectos ao mesmo tempo, muitas obras ao mesmo tempo, o que eu costumo fazer é, tenho um ficheiro no Excel que me ajuda a saber qual é que é o tempo e o espaço em que estão os meus projectos. E o que é que eu faço é... basicamente ponho toda a informação que acho que é relevante mínima, mas acho que é relevante para todos os projectos, para todas as obras relacionadas com um determinado projecto e quando tenho necessidade vou lá actualizar, e vou lá buscar informação.

[R] Portanto, é um documento no fundo que ajuda à gestão do tempo em relação aos vários projectos, e ajuda a saber o ponto da situação em cada um deles...

[I] Exactamente.

[R] Portanto, é uma coisa pessoal. Não foi lida por mais ninguém, não houve ninguém que ajudasse a inserir a informação?

[I] Não. Fui eu que às medida das minhas necessidades fui percebendo mais ou menos aquilo que...

[R] Nunca foi enviado por correio electrónico para ninguém, nunca...

[I] Não. Já o mostrei a algumas pessoas, apenas como ajuda ou mesmo quando necessito de falar de um determinado projecto com determinadas pessoas posso levar uma cópia ou parte desse ficheiro só para me ajudar também a...

[R] Na construção do ficheiro há assim alguma dificuldade, alguma tarefa que venha há memória que tenha sido particularmente difícil? Talvez recolher os dados iniciais para lá colocar? Não sei, alguma dificuldade, alguma coisa que se saliente?

[I] Acho que houve várias! Primeiro porque é uma coisa um bocado complicada porque nem todos os meus projectos têm a mesma ordem e a mesma filosofia, o mesmo âmbito. Tive que



tentar arranjar uma coisa que fosse mais ou menos geral, que se aplicasse a todos, e depois como não funcionou tive que, no fundo, dentro de cada coisinha, a cada projecto, dizer se se aplicava ou se não se aplicava. Tive que partir completamente aquilo senão não conseguia...

[R] Portanto já houve várias versões do documento? Houve uma até que não funcionou muito bem e...

[I] Sim. Eu acho que aquilo é muito dinâmico, porque sempre que eu encontro alguma coisa que eu ache que deva lá estar para determinado projecto o que eu basicamente faço é ponho-a lá, digo que não é aplicável a todos os outros anteriores e depois a partir daí começo a aplicar.

[R] Provavelmente é necessário consultar outros documentos, às vezes, para depois por lá a informação?

[I] Sim. Ele é só mesmo para eu ter um panorama geral. Eu para por lá a informação tenho sempre que ir buscar a outro lado qualquer. Só me serve de painel, digamos assim.

[R] Portanto, é um documento que tem sido mexido ao longo do tempo. Provavelmente até foi mexido bastante recentemente.

[I] Sim. Ainda hoje estive a mexer nele.

[R] No local de trabalho?

[I] Sim.

[R] Que é provavelmente onde ele está armazenado, num computador algures?

[I] Sim.

[R] Existe algum backup?

[I] Eu faço backup do disco de dois em dois meses...

[R] Então, só assim duas perguntas mais genéricas sobre este documento, que é... Sobre a vida pessoal em concreto. A vida pessoal no sentido de uma coisa que pudesse estar marcada na agenda, um aniversário, o dia em que aconteceu um percalço qualquer... 'Trabalhei neste documento no dia em cheguei atrasado porque tive um furo'. Há assim alguma coisa que venha à memória ou...

[I] Relacionado com este documento acho que não... Como é um documento de trabalho normalmente não costumo por lá datas para além das datas relacionadas com os projectos não...

[R] Não precisa ser informação que foi mesmo lá colocada. Informação que venha à memória, assim mais informalmente, que nós nos lembremos. Às vezes há aquela memória associativa... O documento em que trabalhei naquele dia...

[I] Não. Não me estou a lembrar assim de qualquer coisa...

[R] E do mundo assim mais em geral, uma noticia na televisão, um amigo que tenha trocado de carro. Uma coisa assim mais...

[I] Não. Acho que não está associado com nada disso.

[R] Então quanto a este documento penso que estamos conversados. Agora repetimos este exercício, contar a história do documento, mas desta vez vamos pensar num documento mais antigo. Uma coisa que tenha sido já criada e mexida pela última vez há um ano, dois anos. Uma coisa mais antiga. Vamos primeiro identificar um documento e depois descrevê-lo.

[I] ... Ok... Podemos falar por exemplo... Eu tenho que fazer todos os anos um orçamento, uma previsão de orçamento dos meus projectos. Por exemplo, ainda hoje consultei um orçamento de 2001. Basicamente o que eu tenho que fazer é tentar achar... Fazer um orçamento anual de todos os meus projectos e quanto eu acho que... Para depois aquilo ser aprovado. E depois aí...

[R] Portanto, esse orçamento de 2001 estava guardado aonde? Foi fácil de encontrar?

[I] Foi relativamente fácil de encontrar porque eu também não trabalho há assim tanto tempo e é sempre... A rotina acaba por ser um bocado a mesma mas foi relativamente fácil, porque eu tenho aquilo numa pasta fácil de identificar.

[R] E esse documento em concreto, na altura em que foi criado, suponho que em 2001, foi criado por várias pessoas ou foi exclusivamente criado só por uma pessoa?

[I] Não. Foi criado por mim mas sou capaz de ter estado a fazê-lo com a ajuda de outras pessoas.

[R] Portanto nessa altura terá havido trocas por email...

[I] Sim. Ou trocas mesmo de informação, ou estar a fazer o documento a falar mesmo com outra pessoa, isso com certeza que aconteceu.

[R] E ainda é possível lembrar-se de qual foi essa pessoa? Não preciso de saber eu mas se fosse preciso...

[I] sim, sim, sim

[R] Se fosse preciso 'Precisava de falar com que me ajudou...'

[I] Sim, sim

[R] Pronto. Suponho mais uma vez que também haja outros documentos que seja necessário consultar para criar esse ou é uma coisa que...

[I] Não. eu preciso de consultar outra documentação para o elaborar. Preciso de consultar outra documentação para dar uma ideia. Para elaborar um orçamento preciso de saber os vários projectos que eu estou a pensar que vão existir, que eu estou a pensar que vão existir, e aí vou consultar, por exemplo, obras semelhantes que eu já tenha feito, coisas do género, coisas parecidas, outras coisas. Portanto preciso de consultar outra informação.

[R] Portanto... Também foi feito no local de trabalho, como é evidente, e foi criado em 2001, mais concretamente quando, tem ideia?

[I] Foi criado em Fevereiro de 2001. Isto é fácil de dizer porque os orçamentos é até Março e então...

[R] E é um ficheiro Word, já agora?

[I] É um Excel.

[R] E mais uma vez houve assim alguma tarefa, algum percalço? Alguma coisa assim que tenha sido ou particularmente difícil ou particularmente fácil, na construção desse documento?

[I] Pronto. Eu lembro-me que tive... normalmente eu tenho sempre que refazer orçamentos várias vezes até acertar na...

[R] Quer dizer, houve várias versões...

[I] Sim. Houve várias versões até à versão final.

[R] E as versões antigas foram mantidas? Às vezes para cenários diferentes? Ou só a versão final?

[I] Costumo... Como aquilo é em Excel faço várias folhas, várias sheets diferentes...

[R] Pronto, mais uma vez, quanto à vida pessoal ou quanto ao mundo em geral.... 'Foi no dia em que mudei de gabinete...' Há assim alguma coisa que...

[I] Não faço a mínima ideia (risos) Não faço a mínima ideia.

[R] Então, não sei se há mais alguma coisa sobre este documento que seja interessante referir?

[I] (hesita) Não. Ele só tem uma particularidade. É que, também um bocado devido à estrutura dos orçamentos é que tenho obras para várias instalações da [Empresa]. Então tenho obras para várias instalações da [Empresa] e quando faço o orçamento divido-o pelas várias instalações. É só essa particularidade. Não me lembro assim de mais nada.

[R] Então estamos quase no fim. Só falta voltarmos a fazer o exercício mas desta vez provavelmente é mais rápido porque é descrever um documento feito por outra pessoa. Um artigo que tenha sido necessário ler, um caderno de encargos que alguém tenha produzido e que tenha tido que ler... Pronto, um documento, de preferência um documento electrónico mas que tenha sido criado por outra pessoa, encontrado na Internet, uma coisa assim...

[I] Ok. Pode ser... Nós fazemos reuniões mensais de obras de ambiente, qualidade e segurança, por exemplo, e todas as semanas anteriores às reuniões é-me enviado, pelo responsável dessa área, um documento com todas as várias acções do gabinete de qualidade e segurança que são realizadas, portanto, para nos... que estão identificadas. E nós temos que analisar aquilo e levar as nossas conclusões ou as nossas alterações para a reunião.

[R] Portanto, o último ficheiro desses que foi recebido está guardado aonde?

[I] Está numa pasta só dedicada a essa área.

[R] Muito bem. É um documento que é só recebido e as alterações são só feitas na reunião ou o documento é mesmo trabalhado e trocado entre várias pessoas...

[I] Não é só para mim, é para uma série de pessoas, mas o que acontece normalmente é que as alterações são transcritas das duas uma: ou no documento se a reunião não se pode realizar e depois são enviadas para a pessoa, ou então são, se a reunião se realizou, são apontadas na reunião pelo responsável e ele depois faz as alterações e reenvia a versão final.

[R] É um documento em Word?

[I] É um documento em Word.

[R] Muito bem. Há outros documentos que estejam normalmente associados? Por exemplo, para sugerir as tais alterações é preciso ir consultar algum documento... alguma coisa mais...

[I] Não mesmo esse próprio documento eu tenho que ir consultar e verificar as minhas obras todas e ir ver quais são aquelas que se aplicam, tenho...

[R] Provavelmente até consultar o primeiro documento de que falámos hoje...

[I] Exactamente, exactamente.

[R] Muito bem. Qual foi a última vez que foi preciso lidar com um documento desses? Quando é que foi recebido o último?

[I] Há três semanas.

[R] Chegámos à parte daquelas perguntas finais. Houve algum percalço, alguma... Por exemplo, a última vez que foi preciso verificar se estava tudo bem conforme o documento houve alguma dificuldade em especial ou estava logo tudo e por isso se salientou por isso, ou houve assim alguma tarefa que tenha sido mais difícil... Conferir dezenas de números à mão, assim aquelas tarefas penosas que às vezes acontecem?

[I] Não, penso que não. As coisas também já estão um bocado estruturadas para facilitar a leitura e a... Talvez os exemplos que eu tenha dado não sejam os melhores porque não entram muito em conflito...

[R] Todos os exemplos são bons...

[I] ... mas para mim não houve assim nenhuma particularidade que...

[R] E da vida pessoal, do mundo em geral, assim alguma coisa que se saliente?

[I] Não. Também não tenho...

[R] Então, mais uma vez, a menos que haja alguma coisa que considere interessante?

[I] Não. Se calhar só dizer uma curiosidade. Não... Pronto, às vezes posso não me lembrar de imediato onde é que está um determinado documento mas consigo lá chegar porque eu tenho-os ordenados por instalações e então é muito simples.

[R] Está bem organizado por pastas e...

[I] É. É relativamente muito simples eu chegar lá.

[R] Então penso que estamos conversados. Muito obrigado!

## Interview 6

[Researcher] Aquilo que eu começava por pedir era para pensar num documento em concreto, mesmo um ficheiro concreto que tenha sido produzido nos últimos dias, até mesmo hoje. No máximo uma semana. Algo realmente recente. Começar por identificar um documento concreto

e depois dizer... no fundo contar a história desse documento. Tudo o que vem à cabeça sobre o documento. Para que é que ele serve, quando é que foi criado. Pronto, tudo o que assim de repente venha à memória.

[Interviewed] Um documento em que estive a trabalhar hoje de manhã. Foi feito no Excel, fazer um mapa para enviar para a secretaria de estado, relativo a Lisboa e foi criado na semana passada...

[R] Foi provavelmente criado no local de trabalho, não é?

[I] Sim, sim.

[R] E houve alguém que tenha ajudado na sua construção? Algum co-autor, vá lá?

[I] Só na estrutura, mas não trabalhou propriamente no computador. Só no layout e nos elementos que era necessário recolher, e deixei na rede... na pasta partilhada para o meus colegas lerem porque é informação que também lhes diz respeito.

[R] Portanto, eventualmente houve ou está neste momento a haver outras pessoas que estão a ler o documento...

[I] Exacto. Sim, sim.

[R] ... e que eventualmente poderão dar alterações.

[I] Alterações, sim.

[R] Não chegou a ser trocado por correio electrónico...

[I] Exacto. Pus na pasta. Num de vários discos... Aquilo é uma confusão.

[R] Já que falámos disso, onde é que... É fácil saber onde é que ele está, em concreto?

[I] Neste momento é porque trabalhei hoje de manhã, mas às vezes como tenho quatro pastas duplicadas na minha pasta pessoal e na pasta partilhada às vezes não sei qual é a última versão com que estou a trabalhar...

[R] Portanto costuma às vezes haver várias versões e é difícil gerir isso porque podem estar em locais diferentes. E quando o documento estava a ser construído, houve algum percalço ou alguma tarefa de especial dificuldade? Ou então de especial facilidade, que até parecia que ia dar mais trabalho mas até correu mais facilmente do que se estava à espera. Há assim alguma coisa...

[I] Houve, porque os dados foram exportados de uma base de dados e como tivemos que trabalhar os dados para irem em formato Excel... Este ficheiro era relativo a Lisboa mas havia 18 ficheiros para 18 distritos diferentes. Os outros 17 distribuí. Este trabalhei eu. Como os dados foram tirados de uma base de dados tiveram que ser filtrados porque a base de dados não vinha na nossa...

[R] ... estrutura.

[I] Exactamente. E havia informação que era... como a base de dados é relacional havia informação que saia repetida nas várias linhas e tivemos que estar a apagar a informação que estava repetida para não aparecer para o utilizador.

[R] Portanto foi mesmo necessário ir consultar outros documentos, neste caso sob a forma de bases de dados, ir consultar outras fontes de dados para construir.

[I] Sim, sim.

[R] E claro está neste momento ainda se consegue lembrar exactamente de quais foram. Se fosse preciso encontrar outra vez, ainda eram fáceis de identificar...

[I] Sim, sim.

[R] E já agora mais da vida pessoal, no sentido de coisas que até podem estar marcadas na agenda ou, por exemplo, um aniversário, uma coisa qualquer. Há alguma coisa que tenha acontecido que esteja associada a este documento? Às vezes há aquelas coisas 'este documento foi feito no dia em que alguém fez anos. Foi o documento que fiz no dia em que cheguei atrasado porque tive um furo!' Há assim alguma coisa que, algum incidente fora do computador propriamente dito?

[I] Não.

[R] E do mundo em geral, mais uma vez. Uma notícia qualquer que tenha visto na TV ou num jornal. Alguma coisa?

[I] Também não me lembro. Eu fixo mais o destinatário do ficheiro e o assunto.

[R] Então a menos que queira dizer mais alguma coisa sobre este ficheiro, podemos dar este como encerrado. Voltávamos agora a fazer este exercício, portanto, descrever um ficheiro, mas desta vez vamos pensar num documento que tenha sido produzido há já mais tempo. Tipo coisa de um ano, dois anos. Não um documento que tenha sido feito nesta altura. Às vezes há aqueles documentos que têm que ser constantemente actualizados. Mesmo um documento que fez há dois anos atrás e ficou lá arrumado. Mas agora imaginemos que por um acaso precisava de o encontrar outra vez. Portanto vamos primeiro pensar num documento e depois vamos ver do que é que nos lembramos sobre ele.

[I] Há dois anos...

[R] Um, dois anos. É preciso é que seja antigo. A idade em particular não é muito importante.

(...)

[I] Tem que ser de trabalho?

[R] Não. Pode ser pessoal. Uma carta que tenha sido escrita...

[I] As cartas de quando nós mudámos de casa!

[R] Então, quando é que isso foi.

[I] Em Fevereiro do Ano passado.

[R] Se eu agora perguntar se houve algum acontecimento na vida pessoal que esteja relacionado já está respondido não é?

[I] Exactamente. (risos)

[R] E foram feitas aonde? No local de trabalho, já em casa?

[I] Se calhar algumas no local de trabalho e enviadas para casa. Tenho tudo em casa.

[R] Neste momento está tudo no computador de casa?

[I] Sim.

[R] Existe alguma cópia, por exemplo, em CD ou numa disquete, ou...

[I] Existe porque eu tive que reinstalar o computador e tive que fazer cópias de tudo, para salvar guardar.

[R] E foram feitas só por uma pessoa?

[I] Sim. Fui eu.

[R] Alguém leu? Para eventualmente preencher alguma informação que faltasse?

[I] Não, não.

[R] E quando é que foram enviadas? Agora podemos pensar numa em concreto...

[I] Por exemplo TVCabo.

[R] Foi preciso consultar alguma documentação para...

[I] Sim. Número de cliente...

[R] E isso estava arquivado em papel ou em formato electrónico?

[I] Em papel...

[R] E foram escritas aonde. Qual é o tipo do documento? Em Word?

[I] Em Word. Aliás, neste caso nem sei se não fiz nada tipo Mail Merge com uma base de dados em Excel e com os números todos lá...

[R] Provavelmente não chegou a haver várias versões disso. Foi uma carta que...

[I] Foi uma.

[R] E chegou a ser trocada por correio electrónico para alguém? Também não...

[I] Não. Foi por carta.

[R] E do mundo em geral, há assim alguma coisa que tenha acontecido? Alguma notícia? Às vezes, sei lá, qualquer coisa que aconteceu ao vizinho da frente?

[I] Não me lembro.

[R] Então, mais uma vez, a menos que haja alguma coisa que venha à memória sobre este documento... Senão podemos passar à frente. Só falta um que é o mais fácil e o mais rápido também que é mais uma vez contar a história de um documento mas desta vez um documento que tenha sido feito por outra pessoa. Uma página na Internet que nós tenhamos visto e tenhamos

guardado. Um artigo científico ou umas folhas de uma cadeira, uma coisa qualquer. Uma circular interna. Qualquer coisa que tenha sido feita por outra pessoa e que nós tenhamos tido que guardar por alguma razão. Mais uma vez, pensar numa em concreto e depois descrever.

[I] Um manual de análise multi-critério feito por uma entidade governamental do Reino Unido.

[R] E qual era a tal entidade?

[I] O ministério dos transportes.

[R] E para que é que esse documento serviu?

[I] Faz parte da bibliografia que estou a guardar para a tese de mestrado.

[R] Muito bem. E quando é que foi encontrado?

[I] A semana passada.

[R] E estava aonde?

[I] Estava numa página da Internet de grupos relacionados com a análise multi-critério.

[R] Se fosse preciso ir lá buscar outra vez ainda era possível ir dar com aquilo ou...

[I] Eu guardei a referência no próprio ficheiro, porque ficou a... Não. Deste não guardei porque não era em Word, era em pdf... Mas fazia a pesquisa no Google. Se calhar não aparecia, mas pronto...

[R] E foi enviado a outras pessoas? Às vezes um colega que podia ter interesse e a quem foi enviado? Ou não?

[I] Não.

[R] Está guardado aonde?

[I] Está guardado no computador em casa e naquelas.. num Cell-Disk. Aqueles 'coisinhos' que...

[R] E existe algum outro documento com o qual ele se relacione? Que tenha como referência o outro?

[I] Não. Encontrei por acaso nessa página da Internet. Mas está junto com outros que se relacionam sobre o mesmo assunto.

[R] E existe mesmo um autor, uma pessoa que tenha feito...

[I] Existe.

[R] E qual é o nome?

[I] Não me lembro! (risos) Eram quatro ou cinco autores mas não me lembro.

[R] Não há um nome mais sonante?

[I] Não me lembro assim... Não.



[R] E vai ser usado como parte da bibliografia do mestrado. Está a ser reunida e depois há de ser produzido um documento que...

[I] Sim, sim.

[R] Mais uma vez, da vida pessoal, há assim alguma coisa que... É para o mestrado, evidentemente, mas há assim mais algum percalço que... A Internet não funcionou e foi preciso telefonar para a TVCabo?

[I] Isso por acaso aconteceu no dia a seguir... Nesse mês tive sorte (risos)

[R] Do mundo em geral também...

[I] Se calhar o início da cimeira das Lages. Agora como é recente lembro-me...

[R] Ótimo. Pronto. Não sei se há assim mais alguma coisa que ache importante dizer? Senão terminámos!

[I] Pronto.

[R] Obrigado.

## Interview 7

[Researcher] A primeira coisa que eu pedia era para pensares num documento recente. Pode ter sido feito hoje, ou recentemente. Uns dias uma semana, duas semanas, recente mesmo. No máximo duas semanas. E depois de ter identificado um ficheiro concreto, dizer tudo o que te lembras sobre ele. Tudo o que te vem à memória sobre ele.

[Interviewed] Tenho um documento que fiz faz hoje oito dias. E a história dele...

[R] Porque é que o fizeste?

[I] Lembro-me dele porque foi feito para o rancho. Lembro-me dele onde o escrevi. Lembro-me dele do documento em si...

[R] Qual era o tema, então?

[I] Era uma praxe. Lembro-me que o fiz no Domingo. É obvio que daqui a uns tempos já não me vou lembrar do Domingo. E lembro-me que me ocupou uma parte da tarde.

[R] E fizeste-o aonde? Em casa?

[I] Em casa.

[R] E quanto ao conteúdo do documento, a estrutura. Por exemplo, foi preciso por uma imagem que ocupou muito espaço? Está formatado em duas colunas? Tem alguns gráficos coloridos, alguma coisa que se destaque?

[I] Não. É só texto mesmo.

[R] E não houve nenhum percalço a fazer o documento? Um problema a formatar o documento? Copiaste para a disquete e depois tiveste que ir a casa buscar outra vez porque a disquete se estragou?

[I] Não...

[R] E às vezes, aquele género de coisa... Entrou lá alguém enquanto estavas a fazer o documento? Demorou mais um bocado porque a meio tiveste fome e tiveste que ir ao supermercado?

[I] Ah! Parei para... Chegou lá uma pessoa a casa para ajudar a escrever o documento. Depois parei, pus essa pessoa a fazer o documento e fui preparar o jantar. E não parei assim para mais nada. Depois tive que o gravar. Gravei para uma disquete e deixei no disco. Fui imprimi-lo a outro lado no outro dia a seguir. E mais nada... Basicamente foi isso.

[R] E foi preciso consultar outros documentos para fazer esse? Por exemplo, usar pedaços de outros que já estivessem feitos? Ou depois esse já foi usado para fazer outros.

[I] Não foram porque o único que eu teria usado, não sei dele. Foi feito há uma dúzia de meses e não sei onde é que está, nem...

[R] E foi pelo que eu percebi, feito no Word?

[I] Feito no Word, e tive que escrever tudo outra vez porque não sabia do outro.

[R] E chegaste a mandar, por exemplo, por correio electrónico a mais alguém depois? Para mais alguém ler ou mais alguém dar...

[I] Foi para mim só.

[R] Ele chegou a ter várias versões? Às vezes há aqueles documentos que uma pessoa faz uma versão com mais informação e depois...

[I] Não... Talvez outro documento fosse mais...

[R] Não! Não há problema. Falta só mais uma pergunta mais genérica, vá lá, mas que às vezes são coisas importantes. Da vida pessoal, no sentido de uma coisa que até podia estar na agenda: o aniversário de alguém, um exame, uma coisa qualquer... Houve alguma coisa assim relacionada com o documento?

[I] Teve que ser feito mais à pressa porque teve que ser usado no dia a seguir e teve que ser impresso durante o dia.

[R] E do mundo em geral? Uma notícia na televisão, foi o dia em que o vizinho partiu uma perna. Outra coisa qualquer no geral.

[I] Nesse dia não houve assim nada que tivesse marcado o dia. Foi o dia em que a [Mãe] se veio embora do [Local de Estudo] para [Casa]. O documento só foi feito a seguir.

[R] Pronto. Não sei sobre esse se te lembras de mais alguma coisa? Senão, quanto a este, estamos conversados.

[I] Não. Neste momento não me lembro de mais nada.

[R] Então agora voltávamos a fazer esta história, descrever um documento, mas desta vez um documento que já tenha sido feito há mais tempo. Tipo dois anos, três anos. Uma coisa mais antiga. Começar por identificar um documento em concreto, mais uma vez, e depois dizer a história dele.

[I] (pausa) Lembro-me de um que fiz há três anos. Práí três anos.

[R] E o que era?

[I] Era uma notícia. Uma notícia, um artigo para o jornal... para o Fórum Estudante.

[R] E?

[I] O que lembro dele foi um documento que eu fiz de madrugada, quando cheguei a casa da discoteca. E vinha... estava com o [Pessoa 1]. Fomos os dois escrever aquilo. Fui eu que escrevi. E tinha que estar pronto. Tinha que ser entregue nesse dia até à hora do almoço, razão pela qual o fizemos ainda de manhã.

[R] Sobre o que é que tratava?

[I] Tratava sobre os dirigentes associativos em geral, e os principais 'mamões' em particular. Foi o artigo que eu mais gostei de escrever até hoje, e por acaso é-me impossível encontrá-lo.

[R] Portanto não fazes ideia de onde é que ele está guardado agora no computador?

[I] Não.

[R] E teve várias versões, também? Entusiasmaram-se muito e depois tiveram que cortar coisas?

[I] Houve alturas... Não. Pelo contrário. Parávamos e depois alterávamos umas coisas... umas palavras, umas frases...

[R] Mas não houve assim mesmo... que se possa dizer que foram versões separadas e...

[I] Não.

[R] Ok. E outros documentos, consultaram para fazer esse?

[I] Não.

[R] E foi enviado por email?

[I] Foi, foi enviado por email.

[R] Não sabes a quem é que foi mandado?

[I] Não. Eu gravei-o e depois foi enviado por email. Gravei para uma disquete e foi enviado de outro lado porque eu não tinha net.

[R] E houve alguma coisa que, quando estavam a fazer o documento, parecia que ia ser mais fácil e afinal deu montes de trabalho? Ou então ao contrário. Parecia que ia dar montes de trabalho mas depois até se fez bem... Há assim algum percalço...

[I] Por um lado correu bem porque correu sempre bem. Aquilo foi fluído e escreveu-se num

instante. Por outro lado, o mais difícil era permanecer acordado, obviamente. (risos)

[R] E mais uma vez em termos da estrutura do documento em si? Estava formatado a duas colunas? Tinha um gráfico colorido?

[I] Não me lembro como é... Não. Acho que nós não fizemos porque depois no jornal é que formatavam.

[R] E da vida pessoal, ou do mundo em geral, há assim alguma coisa... Foi na altura dos exames? Foi noutra altura qualquer?

[I] Foi na altura em que o [Pessoa 1] saiu da [Organização Associativa], e fomos sair ao [Bar] com mais pessoal, e foi na altura... foi passado pouco tempo do [Pessoa 1] ter dito que o Durão Barroso era um energúmeno em geral mas que de educação não percebia mesmo nada. Razão pela qual ele saiu da [Organização Associativa] (risos). Foi a seguir a essa entrevista, mais ou menos. Práí uma semana ou duas.

[R] Portanto isso foi uma entrevista que foi dada...

[I] Foi.

[R] Então mais uma vez não sei se te lembras de mais alguma coisa sobre este documento?

[I] Há uma frase que me deu particularmente gozo que começava com qualquer coisa do género 'E a esses senhores que...' só que não me lembro depois do resto.

[R] Pronto. Então também estamos conversados quanto a este. Só falta um que provavelmente é mais fácil de descrever porque nos lembramos de menos coisas. É um documento desta vez que tenha sido feito por outra pessoa. Um artigo científico, uma coisa qualquer. Uma página na Web. Uma coisa que nós vimos no computador. Que nos mandaram por email ou que nós vimos na Internet ou que achámos de uma maneira qualquer. E até achámos aquilo interessante e guardámos no disco, ou referimos aquele documento a alguém...

[I] Pode ser qualquer documento? Pode ser um ficheiro vídeo?

[R] Sim, pode ser qualquer coisa.

[I] Lembro-me de um ficheiro dos Marretas a cantar 'Mana Mana...'

[R] E quando é que ele...

[I] E vi esse documento práí há um mês e meio, um mês...

[R] E viste-o aonde?

[I] No meu computador. Aliás, vi-o noutra computador, gravaram-mo e depois vi no meu computador.

[R] Portanto agora está guardado no disco do teu computador?

[I] Está.

[R] E não sabes quem é que o fez? O autor ou...

[I] Não faço a mínima ideia. Sei que foi um documento que foi o [Pessoa 2] que mo deu.

[R] E foi enviado depois a mais pessoas? Achaste piada e depois deste-o a mais pessoas?

[I] Não. Mostrei a mais pessoas mas não dei a mais pessoas.

[R] E houve algum problema a ver o vídeo? Alguma tarefa especial que tiveste que fazer? Tiveste que configurar qualquer coisa? Tiveste que ir buscar à Net qualquer coisa?

[I] Não. Correu tudo sem problemas... Tive que ligar as colunas, estavam desligadas, para o conseguir ouvir. Deu um trabalho considerável.

[R] E outros documentos, outras coisas que associes a esse? Que possas também ter guardadas no computador e que...

[I] Tenho outras coisas que ele me gravou. Ficheiros de ensaios de carros, e outras cenas cómicas como a cena daquele gajo, do entrevistador da NTV que se chateou com o teleponto. Coisas do género.

[R] De um modo geral, outros vídeos.

[I] Sim.

[R] E em termos da vida pessoal ou do mundo em geral, há assim mais alguma coisa? Viste isso que na casa de alguém houve uma festa?

[I] Vi na casa do [Pessoa 2] à noite estava lá eu, ele e o [Pessoa 3], salvo erro.

[R] Ok.

[I] Um dia normal, depois de um ensaio do rancho.

[R] Não foi no dia em que começo a guerra, nada assim...

[I] Não.

[R] Ok. Não sei mais uma vez se há mais alguma coisa que...

[I] Até pode ter sido no dia em que houve alguma notícia, mas não me lembro precisamente do dia...

[R] Ok. Então já está, estamos conversados. Muito obrigado.

## Interview 8

[Researcher] O que eu lhe pedia para fazer era pensar num documento concreto que tenha criado, trabalhado nos últimos dias, na última semana...

[Interviewed] Aliás, posso-lhe dar um exemplo concreto. Ontem um colega meu pediu-me um assunto, ou trabalhar sobre um assunto, do qual eu me lembrava que tinha já feito um documento para outro projecto que eventualmente poderia servir de base. Na altura em que ele me telefonou

eu estava a fazer outra coisa. Obviamente eu não encontrei o documento enquanto estava ao telefone porque estava a fazer outra coisa em simultâneo e portanto não foi fácil, mas logo que desliguei o telefone encontrei rapidamente o documento porque obviamente que depende da forma como nós... Parece-me que há dois parâmetros que são essenciais. Um é mais ou menos nós lembrarmos-nos do momento temporal em que o documento foi realizado, foi considerado como definitivo. Ou seja, situá-lo no tempo. Porque o mail hoje em dia é uma ferramenta essencial e nós há documentos que circulam por mail. E portanto se eu me lembrar mais ou menos qual foi o momento ou o mês ou enfim, o intervalo cronológico em que eu realizei o documento ao fim de obviamente que é importante porque eu vou ao mail, vou aos files, vou às datas, vou a vários sítios e se não me lembrar de nada mais de um documento pelo menos vou por aí e às vezes eu chego lá. Até porque alguns documentos são mesmo mails. Ou seja, são o próprio conteúdo escrito do mail que me importa e não o attachment que vai lá dentro. Portanto, basicamente, esse é um parâmetro. O outro parâmetro...

[R] Neste caso em particular, focando-se nesse documento em que falou. Neste momento lembra-se de qual é a data ou a altura em que ele foi produzido? Mais informações sobre ele?

[I] Lembro-me que foi na altura em que eu fiz a missão. Portanto está relacionado com um determinado elemento temporal da missão que eu fiz em Cabo Verde no ano passado portanto eu sei o documento teve que ter sido feito algures ou em Julho ou em Agosto porque era resultado desse processo e esse processo ocorreu na minha vida... Tive que ir a Cabo Verde, etc. e eu sei exactamente isso. Estava relacionado inclusive com aspectos da minha vida pessoal. Tive que fazer uma deslocação fora do país. Mas nem sempre é o caso. Neste caso foi fácil porque...

[R] E foi produzido cá ou em Cabo Verde?

[I] Foi produzido... uma primeira versão foi feita cá e depois foi finalizado em Cabo Verde. Fiz parte do relatório da missão. Eu sabia algumas informações de contexto e sabia basicamente que estava associado a um determinado acontecimento que se realizou num determinado momento do tempo. Se for isso, ou seja, se eu souber qual é o contexto em que o documento foi produzido e qual é o acontecimento temporal ou de conteúdo em que ele está inserido é relativamente fácil. É muito fácil porque com o tempo as pastas, o folders....

[R] E eventualmente a quem foi enviado? Também se consegue lembrar disso?

[I] Sim. Também se consegue perceber sobretudo noutros documentos que são menos fundamentais como este que era um anexo do relatório da missão tem que se incidir exactamente sobre a pessoa a quem foi enviado. Ou seja, é preciso a gente se lembrar com quem é que falou sobre o documento, ou com quem é que contactou com o documento. Isso também é um factor extremamente importante.

[R] E houve alguém na altura que o tivesse também lido, ou ajudado a produzir o documento?

[I] Sim. Ele foi enviado em termos de... para muita gente. Foi enviado como relatório de uma missão. Foi enviado para o chefe de projecto, para o [Local de Trabalho], foi para Cabo Verde. Foi para muitos locais, portanto.

[R] Mas já depois de terminado. Durante a altura em que estava a ser produzido não foi...

[I] De qualquer forma, como eu queria mandar a esse colega que me falou não só esse documento

mas uma versão prévia, foi mais difícil encontrar a versão prévia do que o próprio documento, porque o próprio documento estava enquadrado, digamos, num contexto bem definido, bem determinado. Enfim, arrumado no computador num sítio específico para o efeito.

[R] No computador aqui do emprego?

[I] Sim. Enquanto que o documento prévio... eu normalmente não faço, a não ser em caso de excepção, circulação de documentos entre os que eu tenho aqui e que tenho em casa, a não ser que esteja a ser trabalhado. E então ele vai ou por mail ou por uma disquete, consoante o tamanho e a vontade que eu tenho de usar o mail, mas vai mas normalmente eu não guardo em casa nenhum documento profissional. Posso lá tê-lo temporariamente se estou a trabalhar nele, mas não guardo, até porque os computadores lá de casa são mais para a minha filha e o meu marido. A minha informação está toda arrumada numa coisa chamada [nome da empresa] e só eu é que mexo naquele folder e não é difícil encontrar documentos pessoais ou profissionais em casa. E no portátil é um pouco a mesma coisa. Enfim, é uma ferramenta de transporte. Aqui é que é mais difícil porque há muitos folders, há muitas áreas, e pode estar distribuído...

[R] Portanto, neste caso em concreto do documento de que estávamos a falar acabou por ser encontrado no mail a versão final e a versão preliminar?

[I] Não, porque eu normalmente os mails eu não os reservo no arquivo de mail, porque quando eles são... Obviamente os mails de conversa normal, sem importância, ficam lá. Mas se um determinado mail eu classifico como relativamente importante ou reutilizável, eu copio o mail para um folder do assunto. Ou seja, eu guardo os mails pelos folders dos documentos. E guardo também... divido o folder em duas partes. Uma parte que é preparatória que corresponde à fase de preparação do documento e depois um folder a que chamo sempre 'mail' em que está o próprio mail de envio ou recepção e o documento final que foi enviado nele. Portanto eu sei onde é que está a versão final. Depois nesse mesmo folder eu guardo a fase preparatória ou a fase de elaboração de qualquer documento. Eu já faço essa preparação em todos os folders que tenho e também já guardo os mails importantes juntamente com os mails em que eles vão. Porque senão depois perde-se o contexto e isso às vezes é importante para perceber o próprio documento, e a forma em que foi enviado e a quem foi enviado. Como o mail tem informação de dada e destinatário, etc., é realmente uma forma de sintetizar um contexto importante para perceber e para encontrar o próprio documento.

[R] E as pastas estão organizadas por tema? Por...

[I] Por acaso é curioso vir... Eu lembrei-me quando fiz hoje de manhã. É assim. Eu normalmente tenho aquilo a que eu chamo folders genéricos, e tenho folders especializados. Se um determinado projecto ou assunto é relevante eu crio um folder específico para esse assunto. Senão ele vai para um folder genérico que eu tenho designado por SGI que é a minha unidade orgânica que tem sempre coisa... o que não é relevante para ter um carácter isolado. Mas às vezes, e foi o caso hoje de manhã, eu isolo um bloco num outro tema. Ou seja, eu crio um outro tema. E exactamente tem a ver com este documento e o assunto de que estamos a falar, que é o seguinte. Nós temos aqui um projecto internacional chamado Metaware no qual o [Pessoa 1] é um dos técnicos participantes. Este projecto já é antigo e até agora os nossos projectos de cooperação internacional resumiam-se ao Metaware. O ano passado, com a ida a Cabo Verde, acabei por ter outro projecto de cooperação internacional onde estava o dito documento de que eu andava ontem à procura. Entretanto, no dia 19, eu recebi mais um convite para ir fazer uma formação

a Moçambique. Portanto é mais um... Ou seja, neste momento começa a ter algum carácter de volume aquilo a que nós chamamos a Cooperação Internacional no meu serviço, no que me respeita, e portanto eu achei que era relevante criar hoje de manhã um folder especial para a Cooperação Internacional que preencheu para mim, na minha sensibilidade, digamos, um ponto em que, neste momento, é melhor realmente isolar isto como um assunto e criei hoje de manhã um folder Cooperação Internacional onde já estão três sub-folders: Metaware, Cabo Verde e Moçambique, que eventualmente poderá crescer. Portanto basicamente é esta a forma como eu giro. É uma forma dinâmica de alguma maneira, porque tem a ver com a situação actual. Por exemplo, quando há aspectos... também faço o contrário. Eu há quatro ou cinco anos mudei completamente de funções no [local de trabalho]. No entanto eu tinha documentos do contexto anterior que de alguma maneira eu não queria perder, porque a gente nunca sabe quando é que faz falta. Nem que seja 'zipados'. Enfim, algum arquivo histórico, chamemos-lhe assim. E nesse caso, aquilo que eu tinha em termos de composição foi tudo agrupado num único folder a quem eu chamei 'Exacto Histórico', que era o nome da unidade orgânica em que estava, e acabei por fazer exactamente o contrário. Ou seja, como não estavam activos aqueles assuntos, eu juntei os assuntos com um título mais significativo no momento para mim e arrumei-os numa área que nem sequer está aqui neste disco. Está na rede, na minha área de rede pessoal, que nós também temos, zipado, perfeitamente arrumado. De qualquer maneira eu sei que se precisar de alguma coisa que tenha feito nessa altura ele estará lá. Portanto, é dinâmico. É completamente dinâmica esta arrumação dos documentos. E é dinâmica exactamente porque a evolução obriga a que a gente ora trate disto, ora trate daquilo...

[R] No caso deste documento de Cabo Verde, houve... É um documento em Word?

[I] É em Word. Tem diagramas que estão em Visio, por exemplo...

[R] Houve alguma dificuldade especial na produção do documento? Alguma tarefa mais morosa?

[I] Foi chato! Porque foi um documento que basicamente... Eu posso explicar um pouco o contexto do documento que é o seguinte. A missão de Cabo Verde faz parte de um projecto muito alargado que vai ter quatro fases na qual estamos na segunda, e o documento correspondia ao fim da primeira fase. Esta primeira fase tinha várias actividades e terminou com a última dessas actividades que foi a demonstração que eu fui lá fazer. Eu comecei o documento em Lisboa antes da missão para ele depois ser incluído. Ele era simultaneamente um documento preparatório da missão. Um documento que foi o relatório da missão e também é o documento que iniciou a segunda fase. Era portanto um documento de transição entre as duas fases para manter a coerência e a continuidade do projecto. E portanto, de alguma maneira ele foi daqui numa versão prévia. Por isso o que me custou mais a encontrar não foi o documento mas foi a versão prévia. Eu comecei aqui numa versão prévia, que entretanto, obviamente, foi alterada, mas que tinha algumas coisas que eu queria mandar a este meu colega agora. Portanto tive que andar 'à pesca' entre a versão final e as versões anteriores para perceber exactamente o que é que iria mandar. Depois disso, num contexto completamente diferente mas que agora eventualmente se vai relacionar com o assunto que iniciamos, eu tinha diagramas e outro documento que estava noutra pasta mas eu sabia qual era a pasta e o contexto em que ele foi realizado, para reunir os três documentos no mail que enviei ainda ontem ao fim do dia para esse meu colega. Para esse meu colega. Basicamente há um documento novo que nós temos que produzir, para esse novo projecto, mas vamos aproveitar coisas que fizemos noutras áreas e noutros projectos. Portanto foi necessário reunir alguma documentação de vários tipos mas relacionada por um tema, de forma a preparar uma ou-



tra acção, uma outra tarefa, que será feita posteriormente, e que já tem o tal projecto. Portanto, o documento acaba por estar neste momento repetido, mas tem que ser. Por acaso podia fazer shortcuts, etc, mas às vezes eu perco-me um bocado na tecnologia. Acabei por repeti-lo na... Ele está num Zip de Cabo Verde mas neste momento já está fora do zip para ser trabalhado para uma nova versão.

[R] E sobre este documento, só para terminarmos sobre ele... Do mundo em geral há algum acontecimento, alguma coisa que tenha dado nas notícias, ou que tenha acontecido com o vizinho do lado, na altura em que o documento foi produzido. Há alguma coisa que venha à memória?

[I] Particularmente neste caso não, mas há muitos documentos em que isso acontece. Não sou eu que me lembro do documento mas eu peço ajuda a alguém que fez qualquer coisa no momento mais relevante do que eu fiz e que eventualmente me possa recordar qual foi o timing. Aliás, recorro...

[R] Portanto, pessoas que na altura ajudaram a produzir o documento, ou que...

[I] sim, ou até às vezes estão mais relacionadas com ele. Isto é sobretudo mais crítico quando nas minhas funções eu tenho que fazer um documento quando ando há problemas, e há sempre problemas, como sabe. Normalmente eu não giro os projectos por si. Tenho um coordenador de núcleo que conhece os projectos e tenho os chefes de projecto que cada qual gere o seu. Mas de qualquer maneira, como gestora desta área eu intervenho quando há problemas. Ou seja, quando é preciso defendermo-nos de algum ataque, e nesse caso eu preciso de ir buscar uma cronologia. Nesse caso eu preciso de ir buscar um conjunto de documentos dos quais eu às vezes tenho conhecimento, porque esses documentos como são formais, digamos assim, são mandados com o meu conhecimento mas que eu não exploro. Não posso saber os pormenores todos do que toda a gente faz. Portanto maior parte das vezes eu nem sequer olho para os pormenores dos documentos. Vejo o contexto, vejo a forma, umas vezes olho outras não olho mas não me lembro. Nem sequer faço qualquer esforço para me lembrar. Na altura em que quero, por exemplo, fazer um documento a contradizer um conjunto de acusações ou de coisas que dizem e eu quero explicar e fazer uma cronologia eu preciso de ajuda, obviamente, da pessoa que os mandou, ou das pessoas que os mandaram, para conseguir reunir essa informação. Não preciso dos mails porque eles já mos enviaram, mas preciso de saber quando é que foi mandado, porquê, mais ou menos para localizar no tempo. Portanto recorro à ajuda deles. Nem podia ser de outra maneira senão eu dava em doida.

[R] Muito bem. Então, quando a este documento, acho que podemos estar descansados. E agora gostava que pensasse num documento que tenha sido criado... Esse acabou por ser criado já mais há algum tempo. Mas uma coisa que tenha sido criada ontem, ou nos últimos dias. Para perceber até que ponto é que a informação estará mais presente ainda.

[I] Eu ontem fiz uns quatro documentos diferentes. Fiz um documento que foi o resultado de uma análise de... portanto... obrigou-me a fazer um conjunto de análises sobre uma base de dados. O documento foi-me pedido pelo telefone. Este é um dos que eu fiz ontem. Aliás ontem fiz dois ou três bastante relevantes...

[R] Basta um...

[I] Vamos continuar com este. Basicamente, o ano passado, eu já não me lembrava bem da data

e pedi-a à minha colega que mo pediu por telefone, se ela se lembrava qual era a data. Cá está o recurso. Portanto, eu o ano passado já tinha feito uma análise semelhante sobre o plano de actividades do ano de 2002 e ontem uma das minhas colegas telefonou-me a dizer 'Eu queria um documento com a mesma... formato, um ficheiro Excel, como o que me mandaste para o ano de 2002, mas agora para 2003'. E portanto eu perguntei-lhe: 'lembras-te mais ou menos qual foi a data em que ele foi mandado?' E ela disse: 'foi mais ou menos em Novembro'. Isso era um factor. Outro factor é que eu no folder, onde ele estava guardado, para além de ter, onde estava guardado todo o processo de preparação, estava exactamente encaixado num folder de assuntos que tinham a ver com aquela acção, que se chama SGI, que é o nome do projecto, da base de dados, 'SGI.' e aqui eu costumo por o nome das pessoas, [Pessoa 2] e [Pessoa 3]. Porquê? Porque como foi uma delas, eu se perguntar pelo nome da pessoa é fácil ir à procura. Eu pura e simplesmente fui ao Find ficheiros e folders e procurei pelos nomes e rapidamente o sistema deu-me o ficheiro. Depois confirmei com as datas, etc. Mais ou menos, um destes elementos, data, nome ou assunto serve-me, mas depois os outros elementos são importantes para efectivamente confirmar que é aquele documento, aquela acção ou aquele assunto que eu quero. Portanto rapidamente encontrei a base. Fiz, digamos, uma réplica para 2003. Pedi os outros documentos de que preciso para cruzamento. Basicamente eu preciso de informação do plano de actividades para ir à procura dos detalhes das actividades dentro desta base de dados. Portanto pedi o documento e elaborei a análise numa base Access que faz o attach automaticamente à base de dados do Oracle em que temos os dados. Elaborei o documento. Fiz ajustes à base de dados Access por causa do ano, etc. E pronto. Construí o documento e ao fim do dia enviei o documento às pessoas que mo pediram sem qualquer problema. E claro que ficou guardado abaixo exactamente a mesma coisa só que 2003 em vez de 2002. Portanto às vezes os nomes dos folders são um bocadinho óbvios, não é? Antigamente era um grande problema porque só tínhamos sete posições. Agora já não temos esse problema o que é muito melhor em termos de...

[R] Permite-nos registar informação que de outra forma perdíamos.

[I] É mais fácil, neste momento encontrar documentos, com o aumento do tamanho dos nomes.

[R] Houve alguém que tenha escrito ou produzido alguma parte desse documento de ontem?

[I] Não. Fui só eu. Mas há outro documento que eu fiz ontem também, e este é interessante, que não foi só feito por mim e, aliás, deu origem a um fluxo de mails entre mim e o meu chefe. Vários mails, uns três ou quatro, aliás posso-lhe mostrar, que basicamente têm a ver com o seguinte. Nós estamos numa fase de preparação ainda dos trabalhos de 2003, apesar de já terem passado dois meses. Estas coisas nunca são feitas a tempo e horas... Portanto, era preciso ver os objectivos do serviço, um conjunto de acções, um conjunto de tarefas. E entretanto eu tinha estado a falar com o meu chefe anteontem, e ele tinha-me dito que... Portanto, havia aqui alguns problemas de decisão, se havia uma reestruturação... A configuração que tínhamos em termos funcionais não é a mais adequada ao que estamos a fazer neste momento. Portanto, eu tinha feito já anteriormente uma série de propostas de vários formatos ao meu chefe, para alterar o conteúdo funcional, formalmente ou informalmente, ou seja a nível da estrutura orgânica ou sem ser mas que na prática tivesse efeitos. E portanto eu tinha falado com ele e ele tinha-me dito uma coisa muito curiosa na segunda-feira. Na quarta. Ele é director e vogal da direcção. Simultaneamente as duas coisas. E então ele dizia-me que eu tenho um conflito entre o Vogal da Direcção e o Director do departamento de Informática. Depois estivemos um bocadinho a discutir aquilo e ele dizia: 'eu não sei ainda se hei de mudar a estrutura orgânica se não hei de mudar e o director da informática quer

mudar porque tem uma coisa chamada [projecto] que já acabou, portanto convém arrumar aquilo, mas por outro lado não é o momento oportuno no [empresa] para fazer uma reestruturação, vogal da direcção, e como temos processos de auditoria fazer durante o processo de auditoria uma alteração estrutural não é bem muito... Não é um momento oportuno. Depois estivemos um pouco a discutir este assunto, ele foi-se embora, e ontem manda-me um mail. Entretanto, eu já tinha feito, antes mesmo desta conversa, tinha feito um documento de alternativas para ele, que tinha a reconfiguração do serviço em termos funcionais e em termos de arrumação de pessoas e tinha dado em mão, nem sequer tinha dado por mail, e então a primeira coisa que ele fez ontem de manhã foi manda um pedido 'reenvie-me o mail que me mandou sobre o seu serviço'. Depois eu enviei-lhe o documento e disse 'olhe, eu não mandei por mail, dei-lho em mão, mas aqui vai'. Depois ele mandou-me outro mail, e depois havia ainda umas dúvidas sobre rearrumação de pessoas e não sei o quê. Mandou-me outro mail, mandei-lhe esse mail e ele a seguir disse: 'oiça: tem que fazer as alterações de estrutura'. Portanto, resolveu o problema de conflito. Vamos fazer pequenas alterações de estrutura mas pede-me que pegue na actual estrutura orgânica e faça as alterações que forem necessárias para o seu serviço e mande-me a papinha toda feita. À moda do vogal da direcção. De maneira que eu tive que ir buscar o documento que eu lhe tinha mandado. Que lhe tinha mandado ontem de manhã por mail e tinha guardado. E sabia que não tinha mandado por mail porque não tinha no radical a palavra 'mail'. Há aqui umas ajudas de memória que correspondem à forma como eu arrumo os documentos de forma a saber se foi ou não mandado. Portanto... mandei-lhe esse documento de manhã. Tinha esse documento que já tinha mandado. Depois fui buscar a estrutura orgânica que estava na Intranet. Por acaso estava em pdf e por isso tive que fazer aquela cópia muito simpática que os pdfs têm para a trazer para outro documento Word. E depois como ele não me tinha dado instruções sobre alguns aspectos mais particulares eu acabei por fazer três versões. No mesmo documento acabei por três possibilidades, três alternativas que eu tinha feito no documento que lhe tinha entregado em mão. E mandei-lhe um mail a dizer: 'a verde vai uma versão que corresponde à alternativa A, etc.' Se me mandar a decisão final eu escolho e dou-lhe mesmo a papa toda... E disse-me 'eu decidi isto assim. Mande-me o documento final!' E foi isso que eu fiz. Voltei ao documento, retirei aquilo que eram as alternativas, foquei-me na alternativa que ele escolheu, e mandei-lhe o documento final. Este foi feito em duas missões. E tudo isso circulou por mail ontem de manhã e foi acabado conjuntamente com o outro que foi uma análise e conjuntamente com o outro de Cabo Verde e conjuntamente com outro que fiz ontem, que foi um dia de fazer documentos!

[R] Então quanto a esse penso que também estamos conversados. Para terminar há um terceiro tipo de documento com que se pode ter alguma interacção diferentes, que são documentos que não fomos nós que fizemos. Um artigo, científico ou não, que tenha sido encontrado, ou que se tenha recebido, e que por algum motivo tem interesse, e fica guardado.

[I] Sim, tenho alguns desses também.

[R] Então mais uma vez, pensando num qualquer em concreto, o que é que vem à memória sobre esse documento?

[I] Normalmente, quando eu guardo um documento, é porque ele é importante num contexto. E portanto mais uma vez ou tenho um folder específico ou um sub-folder dentro de uma área alargada... Eu não tenho muitas áreas alargadas. Eu não tenho documentos em nenhuma área das que o Office cria automaticamente. Eu não uso o que o Office faz nesses termos. Eu crio a minha própria estrutura de folders. E basicamente a minha estrutura de folders é constituída por uma

área genérica que é o saco que tem um título que é o título do serviço, e que pode ter sub-temas, que é uma área genérica. Quando os documentos assumem um carácter mais dinâmico isolo-os. E depois tenho uma área pessoal, onde tenho coisas pessoais. Os mails do meu marido, algumas questões mesmo pessoais, a título pessoal, currículo, uma série de questões mesmo pessoais. Portanto é o meu mail pessoal que se chama 'Pessoal'. Portanto, ou é o SGI ou é um assunto que está escarrapachado no título ou é pessoal. E não tenho mais hipóteses de procura. E como tenho esta forma de caracterizar, está ali. O resto é dentro daquele folder ir pesquisar. Às vezes fico indecisa se está no geral se está no particular por causa do tema, mas nem sempre. Repare, quando eu olho para um documento na intranet ou que alguém me envia, etc. das duas uma: ou esse documento é para eu dar sequência, eu vou ter que fazer qualquer coisa com esse documento para dar sequência. E isto, por exemplo, no caso da Internet, nós temos que fazer pesquisas... Não é neste caso. Alguém me manda um documento para eu dar parecer. Ainda há dias tenho aí vários desse tipo. Obviamente que esse documento para eu dar parecer fica numa pasta correspondente aquele assunto. Fica lá o documento original, no mail original, e fica depois a zona de preparação, ou de sequência. Tenho outro folder onde faço o que for necessário, mas fica, digamos, no mesmo folder principal. Portanto, faço isso. Outras vezes, e esta é uma situação curiosa, eu vou buscar um documento, ou é-me dado um documento no qual eu não tenho nenhuma acção imediata. Por exemplo: nós estamos neste momento a pensar evoluir para WebServices, para XML, etc. e quando eu tenho tempo o que nem sempre é muito real, faço pesquisas na intranet, para ver o que há, as ferramentas, os contextos. E obviamente às vezes encontro coisas interessantes e digo: 'olha... Isto se calhar tem interesse', e tiro o endereço ou tiro mesmo o documento. Portanto, se forma mesmo um site que é geral e tem interesse, tudo bem. Se for um paper, algum ficheiro específico que me interesse, eu trago o próprio ficheiro, porque depois nunca sei se ele lá está da segunda vez que lá vou. Portanto, normalmente o que eu faço é engraçado. O que eu faço é arrumo por temas. XML, WebServices, lá está os temas. Sob um título mais genérico chamado 'Pesquisa de Ferramentas', ou 'Produtos Interessantes', ou qualquer coisa assim, e depois crio um documento Word onde eu ponho os vários sites, se for um site. No site onde eu fui buscar o documento eu digo 'ver documento xpto que eu trouxe'. Portanto, basicamente o que eu faço é organizar a própria pesquisa. Portanto eu não trago a pesquisa que foi feita na net, seja feita no Google ou não sei o quê. O que trago é um documento Word que tem os links para os sites e no caso de ter trazido um ficheiro assinalo que tenho o ficheiro de nome Xpto que é interessante, que veio daquele site. Portanto, no fundo, isto ajuda-me no caso de eu mais tarde vir a reutilizar o próprio ficheiro ou aquela informação ou querer fazer uma nova pesquisa. Às vezes eu ponho comentários sobre o que é que me pareceu o site. Posso mostrar também. Tenho documentos desse tipo. Basicamente a pesquisa é guardada em termos de aquilo que no momento eu tive interesse dentro do tema que eu estava a pesquisar ou o assunto que eu estava a pesquisar.

[R] E é fácil nesses documentos, como não foi a própria pessoa que os fez, é fácil lembrar-se à posteriori do assunto concreto de um determinado documento? Por exemplo, um artigo científico, um paper. É fácil depois lembrar-se do tema do documento ou eventualmente do autor?

[I] Eu não vou muito por autor porque os autores têm nomes muito esquisitos, às vezes e eu perco-me... Tirando o Bo Sundgreen (?) que é um autor consagrado em matéria estatística, há muito poucos autores que eu me lembro de cor. O que eu me lembro normalmente é do assunto, e sobretudo da razão de eu fazer esse documento. Portanto, lembro-me do assunto do documento e lembro-me do contexto em que ele foi encontrado que era o que eu estava a fazer na altura. Ou qual era a minha necessidade na altura, ou porque é que eu fui fazer aquela pesquisa.

[R] Também sabem mais ou menos em que altura é que foi? Em termos temporais?

[I] Isso às vezes é mais difícil. É mais difícil e aí nesses casos eu normalmente não guardo as datas.

[R] E costuma trocar documentos com outras pessoas? Eventualmente outros interessados?

[I] Sim, sim. E nesse caso guardo o mail a dizer que mandei o documento. Por exemplo, encontrei um documento interessante na altura em que andava a fazer planeamento estratégico do Gartner Group sobre planeamento estratégico num dos estados americanos. Está muito bem feito e muito bem realizado. Encontrei-o e como eu sabia que o chefe estava interessado enviei-lho. Nem sequer fiz nada. Gravei o documento e mandei-lhe um mail com o attach a dizer 'olhe, se tiver tempo leia que é interessante.'

[R] E em que formato é que estava esse documento?

[I] Em pdf.

[R] E neste momento se fosse preciso ir lá buscá-lo de repente?

[I] Está na estratégia. Vou buscá-lo imediatamente.

[R] Está relacionado eventualmente com outros documentos? Foi usado depois como bibliografia para produzir outro documento? Ou foi buscá-lo porque aparecia na bibliografia de outra coisa e parecia interessante? Ou foi por acaso...

[I] fui buscá-lo posteriormente para o ler mais em detalhe para fazer uma coisa simples que era o... a nossa... os objectivos estratégicos do serviço e portanto aquilo no aspecto técnico e de conteúdo era interessante mas também me interessava em particular neste serviço. Quando defini os objectivos estratégicos deste serviço eu referi o documento, independentemente do que o meu chefe possa ter feito com ele.

[R] E quando é que foi descoberto esse documento?

[I] Em Novembro, Dezembro... foi no final do ano.

[R] Não sei se quer dizer mais alguma coisa, senão terminámos.

[I] Não.

[R] Obrigado

## Interview 9

[Researcher] Para tentar saber quais são os tais elementos a primeira coisa que lhe pedia era para pensar num documento em particular para focar a história, mas um documento que tenha sido criado recentemente. Hoje, ontem, nos últimos dias. Pensar num documento concreto, eventualmente pensar mesmo num ficheiro concreto que está guardado algures. E depois o que eu peço é que diga tudo aquilo que vem à memória sobre esse documento.

[Interviewed] Uma folha de cálculo de um orçamento doméstico.

[R] Por exemplo. Quando é que foi...

[I] No princípio do ano. Em Janeiro. O que mais é que sei sobre o documento... Sei que é uma folha de cálculo, sei onde está guardado. Está numa pasta que tem uma numeração que é manual. Está nos Meus Documentos. E neste momento assim é que me lembro...

[R] E foi produzido aqui? Em casa?

[I] Foi produzido em casa. Já o actualizei aqui e já o mandei outra vez para casa, etc....

[R] Por correio electrónico?

[I] Exacto. Por correio electrónico.

[R] E foi mandado para mais alguma pessoa? Houve mais alguém que lesse, ou foi estritamente pessoal?

[I] Não. Estritamente pessoal.

[R] Portanto também não houve mais ninguém a trabalhar nele, a ajudar a arranjar informação? Não houve mais nenhum autor...

[I] Não.

[R] Várias versões? Provavelmente existem? Ele foi alterado...

[I] Várias versões... não. Não tenho guardado as versões.

[R] E foi preciso ir buscar informação a outros documentos para colocar lá, ou foi só assim uma coisa...

[I] Foi. Foi preciso ir buscar informação a outros documentos, que não em suporte digital. Quer dizer... alguns...

[R] Se fosse preciso refazer o documento ainda é possível lembrar-se das fontes usadas?

[I] Sim.

[R] E qual é no fundo a razão de ser do documento? Para que é que ele é usado? Ou para que é que foi usado?

[I] Para controlo orçamental.

[R] Há assim algum acontecimento, alguma tarefa que tenha sido particularmente difícil na construção do documento? Há alguma coisa que lhe venha à memória? Fazer muitas contas à mão ou...

[I] Recolher a informação. Recolher alguma informação.

[R] E quanto à estrutura do documento em si, há assim alguma particularidade? Muito grande, muito pequeno, com algumas imagens, algum...

[I] Não. Tem prái umas quarenta linhas ou assim.

[R] E aspectos da vida pessoal... Há alguma coisa que consiga associar ao documento? O docu-

mento que eu fiz no dia em que tive um furo, ou... Há assim algum aniversário...

[I] Não. Só por ser no início do ano.

[R] Do mundo em geral? Do mundo em geral, alguma notícia que tenha sido...

[I] Não. Não está associado...

[R] Não sei se há mais alguma coisa que...

[I] Não. Sobre esse documento não. Por acaso não tenho assim mnemónicas para os documentos. Aparentemente não.

[R] Então tentávamos repetir este exercício mas tentávamos pensar num documento mais antigo. Uma coisa que já tenha sido feita e arrumada há um ano, dois anos... Mais uma vez pensar num documento concreto...

[I] Isso tenho que ir ver... (risos)

[R] Se preferir... No fundo, uma vez escolhido... Não importa em concreto qual é o documento. É só para ajudar a focar a história. Portanto pode ser um qualquer. Ou de trabalho, ou pessoal...

[I] Tenho aqui alguns que já não sei bem o que é que... Uma assinatura de uma revista!

[R] Serve. Quando é que foi feita, essa assinatura?

[I] Acho que em 99.

[R] E qual era a revista, já agora?

[I] Interface.

[R] Portanto, essa carta... foi feita aqui? Em casa?

[I] Foi feito aqui?

[R] E depois foi enviado por correio normal? Correio electrónico?

[I] Correio normal.

[R] E qual o tipo de documento?

[I] Uma carta.

[R] Portanto, em Word, suponho.

[I] Em Word.

[R] Mais uma vez, houve alguma tarefa que tenha tido mais dificuldade...

[I] Não. Correu tudo bem...

[R] Mas o tema sabemos, que é a assinatura...

[I] Sim, exacto. Acho que tinha que fornecer o cartão VISA.

[R] Ok. Foi preciso consultar mais algum documento? Era uma renovação e foi preciso consultar o número de assinante anterior, ou foi assim uma coisa...

[I] Não. Começou aí mesmo.

[R] Suponho que não houve ninguém a ajudar a produzi-lo, dada a natureza do documento...

[I] Não.

[R] E nessa altura, da vida pessoal houve alguma coisa... Começou a assinar porque foi um novo interesse, nessa altura?

[I] Foi um novo interesse porque era uma coisa... Era uma revista sobre as tecnologias de informação na administração pública. Portanto foi por essa razão... Por acaso não me estou a lembrar de nenhum documento... Estou a olhar assim para aqui...

[R] Quanto a esse documento não sei se há mais alguma coisa que venha à memória?

[I] Não... (pausa)

[R] Então se calhar passávamos ao terceiro e último documento que é mais ou menos... Identificá-lo e depois descrever, mas desta vez um documento que tenha sido produzido por outra pessoa qualquer. Por exemplo um artigo, científico ou não que tenha sido de interesse e que ficou guardado aí algures? E depois descrevê-lo...

[I] Um documento escrito...

[R] Escrito por outra pessoa. Que tenha sido encontrado na Internet, ou enviado por alguém.

[I] Eu tenho é um documento escrito por uma equipa, da qual eu participei...

[R] Se fosse um documento no qual não tivesse tido participação nenhuma...

[I] (pausa, procura no computador) Portanto existe um documento que eu tenho para aqui não sei aonde... Eu sei onde é que o posso arranjar rapidamente, mas...

[R] Não é preciso! É mais só para nos lembrarmos dele.

[I] Portanto é um documento que é sobre um (ininteligível) técnico, sobre portais cooperativos.

[R] E implicou o quê, em termos de tarefas, produzir esse documento?

[I] Investigação.

[R] Procurar na Web quais eram os existentes?

[I] ... e disponíveis, etc.

[R] Houve alguma coisa que tenha sido particularmente difícil? Ou particularmente fácil, que até parecia que ia dar mais trabalho mas até nem deu...

[I] A pesquisa. A pesquisa, quer dizer... Sim. Era tanta coisa que não sabia o que é que havia de escolher.



[R] E qual era o tipo de documento? Era em Word, Excel? Uma mistura...

[I] Em Word. Word com imagens, etc.

[R] E neste momento está guardado aonde?

[I] Na intranet. Portanto... No portal interno.

[R] Portanto, já não está na máquina pessoal mas está acessível...

[I] Exacto. Eu já precisei dele. Já lhe perdi o rasto na minha máquina mas sei onde está.

[R] Mais uma vez, na altura em que este documento estava a ser utilizado, há assim alguma coisa na vida pessoal que foi... Um aniversário, uma coisa qualquer. Um percalço qualquer? Aconteceu assim alguma coisa interessante nessa altura?

[I] Não. Não me recordo assim de nada.

[R] Mais uma vez, do mundo em geral? Foi na altura das eleições? Assim alguma coisa marcante?

[I] Não. Por acaso foi finalizado antes de ir para férias. Não há assim mais...

[R] Então não sei se há mais alguma coisa que venha à memória sobre este ou os outros documentos de que falámos?

[I] Não.

[R] Senão estamos conversados. Obrigado.

## Interview 10

[Researcher] A ideia da entrevista vai ser começar por pensar num documento que tenha sido feito no computador agora recentemente, nos últimos dias, na última semana. Uma coisa mesmo recente. E depois de pensares numa coisa em concreto dizeres tudo aquilo de que te lembras sobre o documento. Tudo. Contar a história do documento. Começa por identificar um concreto e depois diz-me...

[Interviewed] Fazer o historial, a razão que levou...

[R] O que quiseres. tudo aquilo de que te lembres relacionado com o documento.

[I] Posso referir concretamente aquele do serviço mais relativo a uma certa disciplina que dou.

[R] Pronto. Pensa num ficheiro concreto...

[I] Exactamente.

[R] Então diz-me lá do que é que te lembras sobre o documento.

[I] O que me lembro sobre o documento. Por exemplo, reportando-me agora à pasta sobre o (ininteligível). Temos lá o programa...

[R] Vamos falar só de uma coisa. Só do programa, por exemplo.

[I] Só sobre o programa? Referindo as componentes todas sobre...

[R] Tudo o que te lembres. Quando é que foi feito, porque é que foi feito, onde é que está, o que é que lá tem dentro.

[I] O programa foi feito em Setembro, não é. O início do ano lectivo. Contém uma introdução, os objectivos, os conteúdos, as formas de avaliação, a calendarização da disciplina e depois o cronograma das actividades que se desenrolam para o primeiro semestre e para o segundo semestre dessa mesma disciplina. Basicamente, e de forma muito geral, contém o programa. Depois há outros ficheiros relativos à mesma pasta...

[R] Mas vamos falar só desse...

[I] Só desse...

[R] Só desse. Foi feito em Setembro e foi feito aonde? Foi feito em casa?

[I] Não. No computador da escola. Só no computador da escola.

[R] E onde é que está guardado agora, se fizesse falta encontrá-lo?

[I] Está lá numa pasta que diz 'Primeiro Ciclo'.

[R] E houve alguém que tenha ajudado a escrever? Houve co-autores, vá lá? Alguém que tenha lido para corrigir?

[I] Não, não. As correcções só eu mais faço aos outros. Mas normalmente, eu este ano trabalho muito a par com o [Pessoa 1] porque ambos temos as duas turmas do terceiro ano dessa prática pedagógica e... que é o programa de prática pedagógica a que me estou a referir, e então cada um tem uma turma. De modo que temos decidido tudo, conciliado tudo, temos trabalhado em equipa e tem resultado muito. Até na redacção no computador de todos os documentos que se referem... Precisamos de avançar ou reformular relativamente à preparação das aulas, ou para a organização da disciplina porque há essa e há outra que está directamente associada que é o 'seminário de supervisão I', que funcionam a par e passo, digamos, e para qualquer das duas disciplinas, nós dois que somos os docentes, um para o turno A e outro para o turno B trabalhamos muito em equipa. De modo que a pasta até... qualquer um de nós tem acesso.

[R] Escreveram os dois e decidiram os dois o que guardaram e isso tudo.

[I] Exactamente. Tanto posso lá ir eu, porque sei a password dele destinada só a esse tipo de trabalhos, como pode lá ir ele.

[R] Então e chegaram a trocar esse ficheiro, esse documento do programa por email? Ou mandaste por email a mais alguém?

[I] Não, porque somos os únicos docentes dessa disciplina.

[R] E foi dado aos alunos, foi impresso?

[I] Impresso, impresso.

[R] Houve assim alguma dificuldade especial na produção do documento? Algum percalço, Alguma...

[I] Em termos técnicos? Não.

[R] Técnicos ou que tenha sido difícil reunir a informação. Ou uma tabela que...

[I] Não. Normalmente há aquelas pequeninas coisas que com o tempo vão surgindo. Em termos de forma, mais um bold aqui, mais um sublinhado ali mas...

[R] Mas nada assim de concreto...

[I] Não. Até porque quando vamos para o computador já temos muito bem delineado... Não em rascunho, já não fazemos rascunhos, mas já temos tudo bem conversado previamente e já sabemos o que é que vamos fazer no computador.

[R] E foi preciso consultar outros documentos para fazer esse?

[I] Para nós já não há necessidade disso. Porque já é uma disciplina que já tem havido cadeiras semelhantes, não é? E é só fazer os ajustamentos finais e as adequações relativas ao ano lectivo.

[R] E foi um documento em Word, Excel? Tinha tabelas. Tinha tabelas?

[I] Não. Tabelas claro que não tem.

[R] E houve várias versões do documento ou foi só uma que foi escrita?

[I] Houve. A calendarização, que é a segunda página, aliás, do programa. A calendarização relativa ao segundo semestre tivemos que a reformular em virtude de... Isto é relativo à prática do primeiro ciclo. Tinha-me escapado um pequeno pormenor que é a semana a seguir à Páscoa as escolas do primeiro ciclo não têm aulas. E depois tivemos que alterar em termos de datas o resto dali para a frente.

[R] Ok. Já falaste aí das datas, das férias e isso... Há mais alguma coisa do mundo em geral que associes ao documento? Aconteceu alguma coisa que...

[I] Não.

[R] Ou da vida pessoal. Algum aniversário...

[I] Não. Quando fomos fazer o cronograma aliás, vimos tudo bem. Porque há aí outro factor que tem a ver com o que te disse há pouco, que nós tivemos que ter sempre em presença. Que é ir fazendo para cada uma das semanas a articulação entre os dois dias em que vão há prática pedagógica que é a segunda e terça feira, no programa da prática pedagógica, e depois no outro programa para o seminário articular que é à quinta, que decorre do que aconteceu na prática. Portanto essa articulação dos programas teve uma calendarização com a articulação sempre coerente de um para o outro. E não houve falhas porque previamente tínhamos visto tudo. Escapou-nos foi de facto, porque desconhecíamos esse pormenor, que depois das férias da Páscoa não haver aulas ou que alterou os tais dois dias, não é?

[R] Então a menos que haja mais alguma coisa sobre esse documento do programa de que te lembres?

[I] Não, não.

[R] Podemos passar adiante. E agora é fazer a mesma coisa mas desta vez que pensasses num documento, numa coisa que fizeste no computador, mas já há muito tempo, tipo dois anos, três anos. Uma coisa qualquer.

[I] Bom, isso agora está mais complicado, há tanto tempo. (risos)

[R] Um trabalho...

[I] Dois anos, três anos... (pausa). Pois. Sei lá!

[R] Uma coisa relacionada com o Mestrado, talvez...

[I] Sim. A tese, pode ser a tese. Já fio há mais tempo. Há cinco anos, quase. Mas pode ser a tese, sim.

[R] O que é que me podes dizer sobre isso?

[I] Ainda a tenho... Tenho... no disco já não tenho. Essa foi feita em casa, não é... Mas tenho em disquetes, duas versões, por uma razão de segurança. Aí já implicou outro tipo de trabalho. Já tinha grelhas, gráficos...

[R] E acabou por ser um documento em quê?

[I] Em Word.

[R] Em Word, ok. E o que é que assim em particular deu muito trabalho ao preparar... Ou então que até parecia difícil e depois foi fácil.

[I] A questão que, de facto, não estou muito habituada a trabalhar com os gráficos.

[R] Ok. Os gráficos foram feitos mesmo em Word?

[I] Não. Em Excel.

[R] E o que é foi preciso consultar para fazer a tese?

[I] Muita documentação! Literatura...

[R] E há assim alguma referência daquelas obrigatórias de que te lembres?

[I] Agora estás a perguntar por referência mesmo para a utilização no computador?

[R] Qualquer coisa. No fundo, outro documento de que possas dizer: 'olha, fiz a tese na altura em que li o artigo tal.' ou 'para fazer a tese tive que consultar o livro tal'.

[I] Mas para fazer em termos de computador. Não em termos de fundamentação teórica para a tese, claro?

[R] Tudo. Qualquer coisa associada...

[I] Sobretudo a revisão de bibliografia que me deu muito trabalho, como em todas as teses, e principalmente por a grande parte das obras consultadas estar em inglês, que não é uma língua

que domine muito.

[R] Mas há assim algum artigo, algum livro de que te lembres...

[I] Supervisão. Era o tema, não é... Houve muitas obras de referência. Apesar de ser um tema que ainda é muito desconhecido entre os professores, mas já vai havendo alguma literatura, sobretudo anglo-saxónica. Também arranjei uma obra do Canadá. Mas em Portugal já vai havendo algumas referências porque... Há uma universidade no país que tem sido pioneira no estudo da supervisão e nas investigações relativas a essa temática, que é a Universidade de Aveiro. E há um pequenino livro que foi o primeiro contacto que tive, que é chamado 'Supervisão', da professora Isabel Alarcão e do professor José Tavares, da Universidade de Aveiro e isso seria, digamos, o impulso. Agora já o acho muito básico, claro.

[R] E na Internet, já fizeste alguma pesquisa? Alguma coisa?

[I] Na Internet nessa altura ainda não. Consultei muito com a ajuda da [Pessoa 2] que é lá a nossa bibliotecária foi... como é que se chama aquele programa de consulta bibliográfica...

[R] Pesquisa bibliográfica. Quando é que foi feito?

[I] O documento todo? 97, 98. Entreguei em Maio de 98.

[R] E houve quem lesse, e revisse, e...

[I] Sim, o [Pessoa 3] (indicado com um gesto)

[R] Ok. Chegou a ser trocado por correio electrónico com alguém? Ou enviado por algum...

[I] Não. Não tive esse tipo de...

[R] E coisas assim da vida pessoal, ou do mundo em geral. O que é que se andava a passar nessa altura. Há assim alguma coisa que venha à memória?

[I] Que me venha à memória?

[R] Quando estavas a fazer. Por exemplo, houve um dia em que até tinhas pensado trabalhar no documento mas depois atrasaste-te porque tiveste um furo e já não deu tempo, alguma coisa assim.

[I] Não, não me lembro. Particularmente não tenho referência.

[R] Ok.

[I] Quer dizer... Houve aí uma circunstância, não minha propriamente, mas que acabou por ter alguma influência no trabalho final. Na recta final. Que foi a doença do orientador. Mas eu tinha a tese toda escrita e a mim propriamente não me causou grande perturbação.

[R] Então já estamos despachados quanto a este. Agora só falta um terceiro e mais rápido, normalmente, que é pensares num documento, um artigo científico, um paper, uma coisa qualquer, mas que não tenhas sido tu a escrever. Que tenhas encontrado, ou que alguém te tenha enviado. Que tenha interesse. Um documento no computador bem entendido. Alguma coisa que descobriste na Internet...

[I] Pode ser através da Internet?

[R] Exacto. E depois guardaste porque achaste importante ou interessante. E é repetir...

[I] Guardei no próprio computador?

[R] Exacto. Um teste, uma carta pessoal...

[I] Não, a questão é às vezes imprimir quando há um site que é o Educare, que é da Porto-Editora, 'educare.pt', que às vezes consulto e há textos que depois imprimo.

[R] E então pensa num texto em particular que lá tenhas encontrado...

[I] Sim. Sobre avaliação, por exemplo.

[R] Conta-me lá o que é que...

[I] Já foi o ano passado. Foi quando foi alterado o regime de avaliação do ensino básico. Esse site normalmente está muito actualizado e é, em termos de informação para os professores, penso que tem coisas interessantes. E...

[R] E sobre o que é que tratava em concreto?

[I] Da avaliação das alterações ao regime de avaliação do ensino básico. As modalidades, as formas, etc.

[R] Quem é que escreveu?

[I] Mesmo legislativa. Portanto, foi o regulamento aprovado ministerialmente.

[R] E foi um documento que era uma página, foi um documento que viste no Word?

[I] Foi um documento que nesse Educare tinha...

[R] Viste no Internet Explorer?

[I] Sim.

[R] E houve mais alguns documentos... Pronto, outros documentos para o qual usaste esse como referência? Que consultaste na mesma altura e que estejam relacionados?

[I] Tenho uma brochura, por exemplo, que comprei o ano passado em Madrid que tem uma série de referências, de sites de todos os domínios e também sobre educação, e já imprimi também um sobre inteligências múltiplas, outro sobre questões didácticas, também, sobre a escola do primeiro ciclo. Em espanhol mas está interessante. Achei que estava muito interessante. Foi quase que assim à aventura. Deixa-me ver o que é que isto tem. Mas normalmente não guardo. O que acho interessante imprimo. Prefiro ter os papeis na mão.

[R] Pronto. Mais uma vez da vida pessoal ou do mundo em geral lembras-te de alguma coisa? Alguma notícia que tenha dado na televisão ou...

[I] Da vida pessoal, neste momento em concreto, consulto muito a Internet.

[R] Mas na altura em que encontraste este documento de que estamos a falar da revisão da

avaliação.

[I] Que eu faça assim alguma associação, não.

[R] Encontrei-o naquele dia...

[I] Não não. Não tenho essas referências

[R] Foi feito ao fim do dia e estava com pressa?

[I] Não, não tenho esse tipo de referências. Sei que decorreu quando vim de Madrid. E depois estive a ler até para outras colegas, essa brochura. E cheguei a emprestar até. Mas referências pessoais dos dias da consulta propriamente.

## Interview 11

[Researcher] O que lhe pedia para fazer era começar por pensar num documento concreto para ajudar a focar a entrevista. Pensar num documento concreto que tenha sido produzido recentemente. Hoje, até. Ontem. No prazo de uma semana. Uma coisa mesmo recente. E uma vez identificado esse documento concreto, no fundo dizer tudo o que vem à memória sobre ele. Eventualmente, para que é que foi feito, o tema. Tudo o que o documento faça recordar.

[Interviewed] Aliás, eu estou a tratar agora de um documento que são os utilizadores de todo o [Local de Trabalho] e estou a fazer um rastreio deles todos e saber... Esse documento com todos os utilizadores de todas as máquinas dá-nos a possibilidade de nós sabermos o que é que cada um utiliza, que software é que cada um utiliza, que hardware é que cada um utiliza.

[R] E está a ser produzido agora, não é?

[I] Exactamente.

[R] E em que formato?

[I] Em Excel.

[R] Em Excel... E que tarefas é que é necessário levar a cabo para...

[I] Nós temos um software que recolhe a informação toda e depois eu faço uma filtragem desse software que essa informação não nos interessa particularmente, traduzo para Excel para fazer os filtros todos e os códigos todos.

[R] E está a ser produzido aqui no local de trabalho, como é evidente...

[I] Exactamente, exactamente.

[R] E está a ser feito com a ajuda de mais alguém? Há algum co-autor?

[I] Não, não. Sou só eu.

[R] Portanto, nesta altura do campeonato ainda não houve ninguém que tivesse lido, corrigido

alguma informação.

[I] Não, ainda não. Eu estou a preparar realmente o documento para depois de ele estar mais ou menos alinhavado fazer uma apresentação dele.

[R] Portanto eventualmente será apresentado a toda a gente. Foi trocado já por correio electrónico com alguém.

[I] Não, ainda não foi. Ainda não foi.

[R] E em termos de versões já houve várias versões? Por exemplo uma primeira aproximação que depois foi refinada ou está a ser uma actualização contínua...

[I] Não. Está a ser uma actualização... é uma actualização contínua.

[R] Estará armazenada aqui nos computadores, não é?

[I] Exactamente.

[R] E em relação à consulta de outros documentos de referência para esclarecer alguns pontos?

[I] Sim, exacto. Nós temos um histórico da circulação de todo o equipamento informático e esse histórico... é cruzada a informação desse histórico é cruzado com a informação que eu estou a retirar agora.

[R] E em termos de algum acontecimento da vida pessoal, ou do mundo em geral. Há alguma coisa que se lembre? Por exemplo: 'o documento está mais atrasado do que devia estar porque tive um furo antes de chegar cá'. Alguma coisa que venha à memória?

[I] Não... não tenho assim nada que...

[R] Nem do mundo em geral? Por exemplo, casou-se a filha do vizinho...

[I] Não. Nesta altura eu ligo os acontecimentos aqui às operações estatísticas, às grandes operações estatísticas que nós temos. E normalmente até às vezes dentro dessa busco eu recorro às operações estatísticas. 'Isto foi na altura daquela operação estatística. Está aqui!. Foi na altura daquele inquérito. Está ali!' Portanto aqui funcionamos mais à base das grandes operações que nós temos e como eu estou ligado normalmente a isso consigo fixar.

[R] Neste caso concreto agora não está a ver assim nada que...

[I] Não.

[R] Pronto. Não sei sobre este ponto se há mais alguma coisa que venha à memória?

[I] Não. Normalmente aqui, além de ter a gestão toda do parque informático tenho a gestão do orçamento do equipamento. E só essas duas coisas consigo ter isso mais ou menos organizado de maneira que sei onde é que estão as coisas. Não vou muito além disso. Depois, toda aquela informação ao longo dos anos ela vai ficando e como digo são as tais informações estatísticas que me dizem... Foi no ano '2000 e tal', vou a esse ano, vou à procura, e pronto. O outro foi no ano de '1900 e troca o passo', vou a esse ano e procuro. É mais fácil do que estar ligado a outro acontecimento externo. É muito difícil. A menos que seja uma coisa muito grave. 'Olha, foi naquela altura que...' Não.



[R] Mas aproveitando que falou de outros documentos mais antigos ligados a outra... aos grandes projectos. Agora repetíamos este exercício, descrever um documento, mas agora pensando num documento mais antigo. Coisa de há dois anos, por exemplo. Mais uma vez pensar num documento concreto e depois descrevê-lo.

[I] Por exemplo, há dois anos a esta parte... Nós estamos em 2003... há dois anos, nós tivemos os censos. Essa operação censitária envolveu compra de equipamento em que além da compra houve a distribuição que não passou por aqui, foi directamente para os sítios onde... para as câmaras municipais. Teve que haver um controlo... aí já houve uma comunicação entre mim... entre as direcções regionais. Portanto as direcções regionais tinham que me enviar informação. Eu tinha que cruzar informação com eles. E depois de cruzar informação com eles tinha que cruzar informação com o nosso departamento financeiro. E tudo o que decorria dessa operação, dessa compra de equipamento, tudo passou por mim, tudo passou pelo financeiro e tudo passou pelas regionais.

[R] Mas há assim algum documento concreto que tenha sido produzido na altura? O que é que me pode dizer sobre ele?

[I] Nessa altura foram feitos vários quadros para as direcções regionais usarem, a nível de avarias, a nível de comunicações das avarias... Mesmo a nível de trabalho interno aqui dentro, quem fez a programação. Não foi o meu caso mas eu tinha esses documentos também comigo.

[R] Portanto foi trocado por email com vários...

[I] Exactamente. Isto é tudo email, não é?

[R] E qual era o tipo desses quadros? Era um documento em Excel que foi preenchido?

[I] Eram documentos em Word e documentos em Excel.

[R] Mais alguém ajudou a produzi-los? Não digo o preenchimento, como é evidente. A produzir o template?

[I] Sim, na parte... Foi aqui a parte do desenvolvimento. A parte que estava mais ligada ao desenvolvimento foram eles que fizeram a base e depois eu produzi o resto do documento, do documento completo. Também houve uma terceira parte envolvida que foi a parte do middleware que depois mais para a frente fez... eles utilizavam também portáteis e fez para... através dos portáteis, fez uma aplicação em Access, se não estou em erro, para reportar as avarias. Foi mais rápido através de Access reportarem as avarias.

[R] Houve mais uma vez, várias versões desse documento que foram produzidas...

[I] Sim, sim, sim, sim, sim.

[R] E foram todas mantidas ou acabou por ficar só a última?

[I] Não. Acabou por ficar só a última.

[R] E em termos do conteúdo propriamente dito, há alguma coisa que se destaque na estrutura? Por exemplo, tinham uma tabela particularmente complicada? Usavam cores garridas? Assim alguma coisa que venha à memória?

[I] Não. Era uma tabela bastante simples. A única coisa que tinha assim mais garrida era a distribuição dos equipamentos. Para separar aquilo por regiões era mais visível pelas cores para ver qual era a região a que pertencia do que estar a preto e branco.

[R] Ok. Foi produzido cá, como é evidente...

[I] Exactamente.

[R] Foi preciso consultar algum material já existente? Por exemplo material dos censos anteriores?

[I] Não, não, não.

[R] Então, quanto a este documento, mais uma vez é aquela pergunta da vida pessoal, ou do mundo em geral. Está associado aos censos, como é evidente. Mas houve assim alguma coisa de carácter mais pessoal. Alguma coisa que até pudesse estar na agenda. O aniversário de alguém, alguma coisa assim...

[I] Não, não houve... Bem, também não dava para pensar em nada. Não dava para pensar em nada porque o trabalho era tanto que não dava para descansar.

[R] Foi um altura que obrigou a ficar às vezes até mais tarde...

[I] Exactamente.

[R] Pronto. E agora continua a estar armazenado algures no disco? Se fosse preciso encontrar..

[I] Exactamente. Aliás, eu só ao fim de quase sete ou oito anos é que acabo por apagar os documentos porque... Aqueles que me facilitam fazer uma evolução ao longo dos anos, esses deixo-os lá estar, porque continuo a fazer uma evolução. Aqueles que, pronto, são uma coisa que já não serve para nada. Serviam naquela altura mas já não serve, vão-se embora, senão qualquer dia não há espaço que aguento.

[R] Muito bem. Não sei sobre este documento em concreto de que falámos se há mais alguma coisa assim que venha à memória?

[I] Não.

[R] Então passamos ao terceiro e último documento de que vamos falar aqui, que vai ser diferente dos outros porque agora queria que falasse de um documento que não foi produzido por si. Portanto, um paper, um artigo num jornal. Mas de preferência uma coisa em formato electrónico! Um daqueles jornais que há agora na Internet, uma página. Qualquer coisa que tenha sido produzida por outra pessoa qualquer, eventualmente até for a do [Local de Trabalho], mas que por algum motivo tenha sido de interesse e se tenha guardado. Por exemplo, um paper será um caso disso. Ou um manual de referência sobre um produto qualquer...

[I] Normalmente o que nós fazemos ali em baixo. Certa informação sobre certo software através da net, que nos convém guardar e ficar com essa informação porque mesmo a própria compra de software obriga a que nós tenhamos mais ou menos dentro da... tenhamos essa informação mais ou menos actualizada. E é mais essa informação, mais um histórico que nós precisamos, tiramos da Internet.

[R] Tente pensar num documento concreto, uma coisa em concreto que tenha sido preciso tirar. O

que é que me pode dizer sobre um documento concreto?

[I] Eles são tantos que...

[R] Um qualquer. Importa mais pensar num para centrar a história do que o documento em si.

[I] Nós temos contratos que são feitos através da net e a informação também vem toda através da net. É tudo feito através da Internet. E...

[R] Onde é que foi feito o último, por exemplo.

[I] O último vai acabar em Julho portanto... Em Julho do ano passado.

[R] E qual é a empresa... quem é fez esse contrato. Quem é que fez do lado de lá...

[I] Era a DataResults. Agora quem trabalhava com esse antivírus, a maior parte das pessoas saíram de lá, portanto não sei se a DataResults existe ou não existe, sinceramente.

[R] E há mais alguma, algum outro produto que esteja associado a esse?

[I] Associado a este não, mas nós trabalhamos muito com os produtos da Macromedia, com os produtos da Adobe... (...) Assim de repente não me lembro de mais nada que tenhamos... Consultamos muito também por causa do hardware. Por causa das impressoras e dos servidores.

[R] Finalmente, a documentação que é encontrada resulta numa página HTML, ou às vezes é mesmo um pdf...

[I] Normalmente são pdfs. Normalmente são pdfs de onde nós tiramos a informação...

[R] E depois ficam guardados aqui ou são só consultados?

[I] Não, não. Ficam guardados. Ficam guardados.

[R] Lembra-se assim do nome de algum em especial? Ou nem por isso.

[I] É difícil. Ser uma extensão pdf não me diz nada. É 'qualquer coisa pdf' mas o nome em si não me diz nada, portanto não é...

[R] É mais fácil lembrar-se, talvez, pelo tema?

[I] É exactamente, exactamente. Porque é mais fácil uma pessoa ir pelo tema. Se é da HP ou da Compaq está num sítio. Se é daquele software está naquele sítio dividido por n,n,n,n,n,n e sei mais ou menos...

[R] Ok. Portanto acaba por ser o tema, a empresa...

[I] Exactamente. A empresa do produto.

[R] A altura em que o documento foi encontrado? 'Foi aquele manual que tirei há dois meses'? Ou é mais mesmo só o tema?

[I] Não. Às vezes acontece isso. Foi o documento que eu tirei o ano passado. Deixa-me cá ver o que é que eu o que é que eu hei de este ano... Foi o ano passado, vou à procura do ano passado o que é que eu tirei de lá e que informação é que eles me puderam dar nessa altura. É mais a

referência ao espaço...

[R] Muito bem. Aquela pergunta que eu faço sempre. Da vida pessoal ou do mundo em geral há assim... Se pensarmos num documento concreto. Pode ser o tal contrato do antivírus de que estávamos a falar há pouco. Costuma ficar associada alguma informação? Ou mais uma vez são os grandes projectos...

[I] Até é capaz de ficar, só que quer dizer... São coisas... Só se for uma coisa muito grave é que uma pessoa se vai lembrar. Senão, se não é, passa ao lado e não...

[R] Na altura fica mas...

[I] Fica mas depois desvanece-se completamente. Acho que isso desvanece-se completamente. Talvez quem trabalhe pouco consiga juntar. Agora quem faz muita coisa, um grande volume de informação, é complicado. É complicado porque... Principalmente para fazer gestão de orçamentos é complicado. Uma pessoa tem que ir aos anos anteriores. Tem que fazer contas e contas e contas e agora quando foi a conversão para o Euro converter..

[R] Pronto. Então não sei se há mais alguma coisa assim dentro desta conversa que temos estado a ter ache interessante dizer, senão estamos conversados.

[I] Não. Não tenho assim mais nada específico...

[R] Então penso que terminámos. Obrigado.

## Interview 12

[Researcher] O que eu vou começar por perguntar é pensares num documento que tenhas escrito ontem, ou a semana passada no máximo. Um documento recente. E dizes tudo o que vem à memória sobre o dito documento. Quando é que foi feito, porque é que foi. Começar por pensar num concreto, uma coisa concreta, e depois descrevê-lo.

[Interviewed] Ora um documento que eu tenha feito na semana passada, em concreto...

[R] Na preparação das aulas, talvez?

[I] Portanto, lembro-me, em termos de preparação das aulas de ter sistematizado um conjunto de informação em transparências para depois servirem de suporte à aula. Comecei por pensar no tema que ia abordar...

[R] Que era?

[I] E que era sobre a investigação educativa, ou sobre a investigação em educação, e mais exactamente sobre os procedimentos a adoptar no caso de uma investigação de natureza qualitativa. Portanto, um pouco caracterizando a investigação segundo o paradigma qualitativo, e dentro desta especificamente no campo da abordagem biográfica. Previamente consultei algumas obras de referência, até aquelas que já a bibliografia geral do programa da disciplina previa para organizar de forma mais sistematizada e englobante a problemática da abordagem biográfica, ou

do método biográfico. Seguidamente comecei a redigir de acordo com os diferentes aspectos que considerava essenciais os diversos... aquilo que iriam ser as diversas transparências. Uma vez redigido o documento imprimi-o logo directamente nas folhas de acetato...

[R] E foi feito, portanto, em Powerpoint?

[I] Não, estes foram feitos em Word.

[R] E foi feito aqui ou na escola?

[I] Foi feito em casa. Aliás, como todas as aulas, são todas preparadas em casa, dadas as condições do computador da escola, onde num gabinete em que há cinco docentes há um computador para ser utilizado por todos. Além de que as tarefas a desenvolver na escola dão pouco espaço de tempo para preparar as aulas, o que tem que ser feito em casa.

[R] E mais alguém... Depois enviaste por correio electrónico ou deste alguém para ler também ou para usar também? Ou deste aos alunos?

[I] Normalmente forneço sempre aos alunos fotocópias das transparências com duas finalidades. Por um lado para que eles nas aulas possam estar com mais atenção e até intervir questionando sobre aspectos eventualmente menos perceptíveis ou dúvidas, e por outro lado para servir como suporte depois ao trabalho individual que eles fazem.

[R] E houve alguma tarefa especialmente complicada na construção do documento? Foi alguma coisa que teu bastante trabalho a encontrar, ou foi tudo mais ou menos linear.. Uma tabela que querias lá meter e aquilo se recusava a meter do tamanho que tu querias, alguma imagem...

[I] Em concreto para este último documento não. Já me tem acontecido isso em relação a outros documentos. Há determinados procedimentos informáticos que eu não domino e às vezes o computador não responde aquilo que eu pretendo, provavelmente porque eu não lhe sei dar as indicações ou os comandos necessários. Às vezes há... designadamente quando se processa texto acontecem situações que é difícil resolver, como por exemplo aparecerem espaços entre parágrafos, e que não se conseguem anular. Normalmente funciona por tentativa e erro. Vou tentando vários procedimentos até que dá... Outras vezes não consigo resolver o problema, arranjo um subterfúgio qualquer para ultrapassar a situação.

[R] E está guardado aqui no computador? Guardaste nalguma disquete também, ou...

[I] Normalmente tenho sempre no disco e guardo em disquete para se houver um problema qualquer ter sempre um suporte suplementar.

[R] Ok. E lembras-te alguma coisa da vida pessoal ou do mundo em geral, quando estavas a fazer estas transparências? Tipo foi um dia em que até tinhas pensado trabalhar naquilo mas chegaste atrasado porque encontraste alguém e estiveste a falar, ou foi preciso ir ao supermercado a correr porque faltava qualquer coisa, ou...

[I] Não, não me lembro de nada em particular em relação a... quer dizer, quando elaborei este documento e porque se tratava de uma situação específica de preparação de aulas, há sempre uma preocupação subjacente que condiciona a preparação do documento... Por um lado é sem, ou melhor, evitando que haja erros ou simplificação que possa por em causa a natureza e a qualidade dos conceitos a abordar, mas pensando sempre numa situação dos alunos, conhecidas as dificul-

dades que só alunos às vezes têm, quer ao nível da compreensão de uma linguagem mais técnica ou científica, quer ao nível da complexidade dos conceitos. Isso de facto constitui um condicionamento que não foi específico deste documento porque são condições que normalmente quando elaboro documentos para as aulas tenho sempre em atenção mas que de algum modo constituem um factor que determina o modo de elaborar o próprio documento, e que têm a ver com a própria forma de apresentar os conteúdos do documento.

[R] Ok. Quanto a este documento acho que estamos conversados. Agora queria que fizesses a mesma coisa, dizer tudo o que te vem à memória sobre o documento, mas desta vez uma coisa feita há dois anos, três anos. Uma coisa mais antiga. Mais uma vez começar por identificar um documento concreto, e depois descrevê-lo.

[I] Um documento marcante e que já tem alguns anos, e que é um documento de maior dimensão quer em termos de número de páginas quer em termos de tempo necessário, posso referir a elaboração da tese de doutoramento. Evidentemente que aí houve vários tipos de dificuldades na medida em que o documento é... a tese é constituída não só pelo corpo do trabalho propriamente dito mas também pelos anexos, que implicou o recurso ao uso de, por exemplo, de... para fazer os quadros, do Excel. E a elaboração de quadros, tabelas, e de gráficos. E de facto havia um certo conjunto de procedimentos de informática que eu não dominava e que, em primeiro lugar, tive que perguntar a quem soubesse que me desse algumas indicações. E depois fui procedendo por tentativa e erro. Numa situação de naufrágio, de não saber ultrapassar uma dificuldade intransponível, pedia apoio para lhe dar realização. Uma outra dificuldade que encontrei na elaboração do documento é que, tratando-se de um documento muito longo, tive que o organizar em diversos ficheiros. O que teria sido eventualmente ultrapassável se em vez de ter utilizado apenas o disco, o computador como tinha pouca capacidade de disco e portanto no suporte de reserva, as disquetes, e que eram sempre a dobrar para garantir que se falhasse o disco e falhasse uma disquete houvesse sempre uma segunda disquete e portanto um terceiro suporte. Se eventualmente nessa altura tivesse recorrido a um computador mais potente e tivesse capacidade de utilizar CD teria sido mais fácil, mas... Portanto, essa foi uma das dificuldades. Outra dificuldade no processamento de texto foi localizar... Por exemplo, os gráficos, no contexto geral do texto, portanto, onde eles deviam de facto... porque estavam, foram feitos em ficheiros separados e na transposição dos gráficos para o ficheiro geral do texto muitas vezes havia problemas, designadamente quando os gráficos desapareciam todos e em sua substituição aparecia um quadradinho no centro da página, que era uma carga de trabalhos. Assim, em termos globais, quer dizer...

[R] E alguém leu o documento mais, para além de...

[I] Evidentemente, tratando-se de um documento desta natureza, não só por uma questão de lógica organizacional, que é da responsabilidade do orientador, e que leu o documento, mas em termos sobretudo de gralhas, de processamento, trocas de letras, palavras às vezes falta de espaços entre as palavras, etc., pedi que me fossem lendo o documento, para além das próprias correcções que eu fazia, para me ajudar a superar essas dificuldades e até em termos de estética, de apresentação do documento em si, da localização, sei lá, o tipo de letra não mas o tamanho das letras, a articulação entre os quadros, as tabelas e o próprio texto, em termos de apresentação, também pedi aconselhamentos sobre esses aspectos.

[R] E agora continua guardado aqui no disco e em disquetes ou está só em disquetes?

[I] Continua a estar no disco e em disquetes.

[R] E foi feito cá em casa?

[I] Foi feito tudo em casa.

[R] E em que altura é que isso foi?

[I] Portanto... Terminou em Dezembro de 2000.

[R] Ok. E foi alguma vez enviada alguma parte por correio electrónico para alguém? Foi trocado com alguém?

[I] Não. Foi sempre entregue em suporte de papel.

[R] Mais uma vez, assim da vida pessoal ou do mundo em geral, há alguma coisa que esteja ligada...

[I] Pois... Nesse período, que foi um período relativamente longo eu iniciei um processo... Era presidente do Conselho Directivo, terminei o mandato, depois disso... Só depois de terminar o mandato é que pude investir verdadeiramente... Tive licença, licença para... tive nesse espaço um período de licença sabática de seis meses e depois tive a licença para a dispensa do Prodep de três anos e durante esses três anos foram anos de trabalho contínuo de manhã à noite em que, em termos pessoais não aconteceu assim nada de especial... O que aconteceu de especial foi que trabalhando os dias continuamente, muitas vezes já estava de tal maneira saturado e cansado que tinha que ir ao supermercado quanto mais não fosse comprar pão mole para aliviar um bocado.

[R] Ok. Quanto a este também já estamos conversados. Só falta um, que normalmente até é mais rápido, que é descrever um documento que não tenhas sido tu a fazer. Um paper qualquer que tenhas recebido no computador e que tenhas agora guardado no computador. Ou uma carta que alguém te enviou, um documento que recebeste por email, por exemplo, ou que viste na Internet, e que achaste interessante e guardaste no computador.

[I] Posso, por exemplo... A última coisa que recebi em termos de... via correio electrónico é um projecto, parte de um projecto de trabalho de uma aluna, uma formanda, que estou a orientar em termos de Mestrado, e que me manda um conjunto de linhas orientadoras que pensa seguir na dissertação de mestrado que está a começar, juntamente com um questionário, ou melhor, uma proposta de questionário, para aplicar no âmbito do estudo que vai fazer.

[R] E qual é o tema desse estudo?

[I] O estudo que vai fazer é a caracterização... O mestrado é na área da educação de adultos e da formação comunitária e o estudo que ela vai fazer é uma caracterização de uma comunidade rural no plano... sobretudo sobre a preservação das suas tradições e na necessidade da formação para que essas tradições, de facto, possam ser preservadas e transmitidas.

[R] E quando é que recebeste esse documento?

[I] Recebi na semana passada na terça-feira. Portanto há quatro dias.

[R] E em que formato é que era o documento?

[I] Era em Word mas em..., portanto, no sistema Zip.

[R] E houve alguma dificuldade, algum problema, para conseguires ver o ficheiro, ou foi...

[I] Não houve problema. Fiz download do ficheiro e como o computador tem o programa Zip...

[R] Ok. E está guardado aonde, agora?

[I] Está guardado num ficheiro próprio que se chama tese e depois o nome da formanda.

[R] Relacionas isso com outros documentos? Há alguma coisa que esteja relacionada com esse que ela te mandou?

[I] Pois, este documento está relacionado com um outro que ela já me apresentou e eu já analisei que foi o projecto de tese.

[R] Pronto... E viste o documento em casa, ou lá na escola?

[I] Vi em casa. E está arquivado em casa.

[R] E mais uma vez, da vida pessoal, ou do mundo em geral, há alguma coisa que te venha à memória quando pensas nesse documento?

[I] Pois... esse documento, em termos pessoais, quer dizer... Foi a primeira experiência em receber um documento de forma concentrada, digamos assim. Foi o primeiro que me foi enviado. Pronto, quer dizer. E da minha vida pessoal o que é que eu posso associar? Dada a natureza do mestrado em que a tese vai ser elaborada, do meu ponto de vista pessoal as questões que no documento são levantadas lembram-me do tempo em que eu próprio trabalhei em alfabetização, após o 25 de Abril. Fiz algum trabalho enquanto à alfabetização e de facto o documento recorda-me isso. Há alguns aspectos que têm a ver com as culturas locais, designadamente com a manutenção, a preservação das características das culturas locais. O documento lembra-me isso. Por outro lado também me faz lembrar uma outra coisa. É que eu estou neste momento a leccionar uma disciplina de diferenciação pedagógica e de facto a diferenciação pedagógica que é uma abordagem científico-pedagógica que visa dar resposta à diversidade dos alunos que acontecem, que constituem uma turma, um grupo, não apenas as diferenças individuais mas as diferenças culturais, étnicas, sócio-culturais, etc. Portanto também está de algum modo relacionado com isso.

[R] Ok. Não houve assim nenhum percalço... Por exemplo, estavas a tentar tirar o ficheiro e a Internet foi abaixo. Ou acabou-se a luz. Não assim nada...

[I] Durante esse não. Mas não é pouco frequente acontecer isso. Não sei se é do Windows se o que é. De vez em quando o rato fica paralisado e o computador vai abaixo. Isso acontece com alguma frequência. Mas durante este documento não.

[R] Pronto. Então, já está. Terminámos.

## Interview 13

[Researcher] Para começarmos eu começava por lhe pedir para pensar num documento concreto, numa coisa concreta que tenha sido criado nos últimos dias, na última semana, no máximo. Por-



tanto, uma coisa recente.

**[Interviewed]** Na última semana?

**[R]** Sim, uma coisa recente. Não importa propriamente o dia concreto mas é só importante que seja um documento recente. E depois uma vez identificado é simplesmente dizer tudo aquilo que vem à memória sobre ele. Qual é o tema, quando é que foi feito, algum percalço na sua criação?

**[I]** Um documento que tenha criado... Pode ser mais difícil. Estou-me a lembrar dos que escrevi hoje...

**[R]** Os de hoje também servem. Dentro do prazo de até uma semana. Portanto os de hoje servem perfeitamente.

**[I]** Ai é? É o que me veio primeiro à cabeça.

**[R]** Claro.

**[I]** O que eu fiz ontem. O que é que eu lhe digo agora?

**[R]** Qual era o tema dele, para que é que foi feito...

**[I]** Está bem. Era uma apresentação em Powerpoint. Foi feito para uma formação, para um seminário. Fiz em casa no portátil. Criei uma pasta própria pendurada numa directoria que eu tenho quando estou a trabalhar em casa.

**[R]** Neste momento ainda está no portátil ou já foi copiado...

**[I]** Já foi copiado... Não, não foi copiado! Copiei para Cd só. Foi copiado para CD. Esqueci-me de o copiar para o rígido do trabalho.

**[R]** E houve algum co-autor ou alguém que tenha lido e dado comentários para rever ou...

**[I]** Não, não.

**[R]** E qual era o tema, já agora?

**[I]** O tema é cartografia temática.

**[R]** E foi preciso consultar alguns outros documentos para fazer isso?

**[I]** Foi, foi. Isso é que foi mais complicado. Porque curiosamente neste caso o que eu fiz foi fazer um copy de todos os documentos de que ia precisar para essa pasta, por causa de que estava no portátil. E então foi fácil encontrá-los.

**[R]** E na construção do documento houve alguma tarefa que tenha sido muito difícil, tipo reunir dados de muitos sítios e construir uma tabela complexa. Coisas desse género. Ou então particularmente fácil.

**[I]** Não. O complicado foi que eu me lembrava de uma vez ter gravado uma imagem, e andei à procura dela e não a consegui encontrar (riso).

**[R]** E depreendo que acabou por ter imagens, já agora?

[I] Não... Ah! A apresentação tem.

[R] Não só essa mas outras.

[I] Não, acabei por criar uma imagem nova. Isso é outra coisa. Eu outras vezes demoro mais tempo a encontrar o que quero do que a fazer de novo. Por isso faço de novo.

[R] Exactamente. Em relação a algum elemento da vida pessoal, no sentido de alguma coisa que pudesse estar guardada mesmo na agenda, marcada mesmo na agenda. Tipo coincidiu com o aniversário de alguém.

[I] Não. O que eu faço às vezes para recuperar documentos que eu escrevi atrás é ver, em vez de me lembrar do sítio em que os guardei, olho para a agenda, vejo quando é que os fiz, e faço um search pela data, para encontrar.

[R] Aspectos em concreto do mundo em geral, do tipo: o dia em que começou a guerra?

[I] Sim. Nalgumas coisas sim. Da minha vida pessoal sim, definitivamente. Fiz o documento na véspera do aniversário, ou antes do 25 de Abril, ou... Pronto, qualquer coisa assim.

[R] Agora era repetir o mesmo exercício, descrever o documento, mas desta vez um documento consideravelmente mais antigo. Dois anos, três anos. Uma coisa mesmo mais antiga. Começar por identificar um, mais uma vez.

[I] Está bem. Dois anos... Pode ser... dois anos... Quer dizer. Eu lembro-me de coisas de há dois anos mas estão muito bem guardadas e lembro-me de onde é que as encontro, e acho que não é isso... Quer isso? Quer esse exemplo?

[R] Se não vierem outras à memória... Porque apesar de saber onde se encontram interessa-me também mais aquela informação adicional de onde é que foi, porque é que foi, talvez...

[I] Sim. Então pode ser há dois anos. Mais de dois anos, não é. Sim. Um artigo que tenha escrito. Um artigo.

[R] Sim. E qual era o título, qual era o tema?

[I] O tema era detecção remota, imagens de satélite. Era sobre... Aí como eu recupero é muito simplesmente... tenho a directoria organizada por conferências ou por revistas e encontro facilmente.

[R] E este em particular para que conferência é que foi?

[I] Foi para uma em Barcelona.

[R] O nome exactamente já não...

[I] O nome da conferência? Sim... O nome... era... eu tenho... É um nome inventado, que é Aerospace 2000. Não era o nome da conferência.

[R] Mas serve de referência?

[I] Sim.

[R] Foi feito há cerca de dois anos, não é?

[I] Sim. Dois anos sim. Mais de dois anos. Há mais de dois anos.

[R] E houve algum co-autor?

[I] Houve. Esse tinha co-autores.

[R] E ainda é possível, se agora fosse preciso falar com eles, ainda é possível contactá-los, saber quem eles são...

[I] Sim. Sim, era era, claro. Mas acho que não iria por essa via. Tentar recuperar o documento por aí? Acho que não ia por aí, não.

[R] E foi feito aonde? Foi feito aqui, em casa?

[I] Foi feito parte em casa e parte no trabalho.

[R] Portanto houve trocas por email, eventualmente, mesmo com os co-autores?

[I] Houve, sim, houve.

[R] Depois provavelmente foi submetido pela Internet para a conferência?

[I] Sim. Sim.

[R] E era um documento de que tipo? Word?

[I] Word.

[R] E tinha alguma coisa na sua estrutura, algumas tabelas complicadas?

[I] Tinha tabelas e figuras. Tinha figuras.

[R] E alguma delas, em relação àquela história das tarefas difíceis ou peculiares. Houve assim alguma tarefa que...

[I] Não. Eu aí tenho um problema que é precisamente andar com as análises, que eu às vezes faço várias versões do documento e depois não consigo encontrar a versão certa.

[R] Portanto presumo que por vezes apesar de até ser involuntário essas versões acabam por persistir algures no disco?

[I] Há, sim! Persistem sempre. Eu nunca apago nada. É muito raro...

[R] Eu também sou assim (risos). Outros documentos como é evidente fizeram falta como referências na bibliografia, mas há assim alguma referência que tenha sido usada que venha agora à memória?

[I] Referência que tenha sido usada que venha agora à memória... Sim, um artigo de um orientador meu, sim.

[R] Muito bem. E está neste momento armazenado aonde? Está no computador de casa?

[I] Está no trabalho.

[R] Mais uma vez, só para terminarmos sobre este documento. Sobre a vida pessoal, ou o mundo em geral, há assim algum landmark que...

[I] Não. Às vezes, quando eu me quero lembrar de um determinado documento, posso não me lembrar se foi escrito em 2000 ou em 99. Sim, isso tenho sempre marcos pessoais para encontrar. E marcos pessoais pode ser ter deixado de fumar. Entretanto agora já fumo outra vez... Ou ter passado umas férias não sei quanto e lembro-me que foi quando escrevi uma parte de um relatório quando estava na casa de uns amigos, ou sim...

[R] Neste caso em concreto deste paper de que estávamos a falar, há assim alguma coisa que venha à memória?

[I] Não.

[R] Então, não sei sobre este documento há alguma coisa que ache interessante?

[I] Eu mencionar?

[R] Sim.

[I] Não.

[R] Só falta um terceiro documento, que normalmente é o mais rápido de todos, porque desta vez é um documento que não tenha sido feito por nós. Portanto, um artigo que tenha sido lido por outra coisa qualquer...

[I] Tenho que me lembrar, não é?

[R] Tem que se lembrar de um. Desta vez não há limitação em termos temporais, pode ser um qualquer.

[I] Ok. Já me lembrei.

[R] E descrevê-lo...

[I] Descrevê-lo...

[R] Sobre o que é que tratava...

[I] Pronto. Era um documento da comissão europeia para política... Definição da política... É um livro verde da política para o espaço da UE. É um pdf. Estava criado... Sabia que o podia encontrar em dois sítios. Ou nas mensagens de email que eu nunca apago, também, ou na directoria onde... Eu hoje andei à procura dele e foi fácil de encontrá-lo porque criei uma pasta... Aquilo tem a ver com discussões de coisas do espaço e eu criei uma pasta para questões políticas do espaço. Mais coisas... Uma coisa que eu me lembrei foi... lembrei-me de que tinha criado uma directoria há pouco tempo para meter este tipo de coisas. Por isso se calhar facilmente encontrei-o.

[R] Mas quando é que ele foi encontrado pela primeira vez? Ou foi recebido pela primeira vez?

[I] Há pouco tempo. Há duas semanas.

[R] E quem foi que o enviou? Foi mesmo...

[I] Fui buscá-lo. Descarreguei-o. Fiz download da net.

[R] Existe algum outro documento que apareça associado de memória? Por exemplo, foi buscar esse para servir de referência noutro lado, ou porque fez falta para produzir algum outro documento?

[I] Não. Fui buscá-lo agora porque um colega meu me pediu para enviar por email. Porque a pessoa não conseguia ir buscar...

[R] Portanto acabou por ser enviado a outra pessoa...

[I] A outra pessoa, sim.

[R] Não houve nenhum percalço em o recuperar? Às vezes aqueles dias em que a rede está lenta e vai abaixo ou que...

[I] Não, não.

[R] E só falta fazer aquela pergunta mais genérica. Em termos da vida pessoal ou do mundo em geral há algum evento que...

[I] Neste caso?

[R] Neste caso. Que esteja relacionado com...

[I] Sim. A minha última ida a Bruxelas, porque teve a ver com isso.

[R] Muito bem. Só mais uma coisa que ia ficando esquecida. Em termos do conteúdo do documento, há alguma coisa que se saliente, como por exemplo um gráfico particularmente informativo, algum...

[I] Não. Há o formato pdf que eu detesto.

[R] Portanto estamos. A menos que haja alguma coisa a dizer? Só ultrapassámos em dois minutos, apesar do percalço. Obrigado.

## Interview 14

[**Researcher**] Para começar, o que eu pedia era para pensar num documento concreto que tenha sido feito ainda hoje ou nos últimos dias, no máximo no prazo de uma semana, e depois de identificar o documento concreto, descrever tudo o que vem à memória sobre esse documento. Quando é que foi feito, para que era...

[**Interviewed**] Ora bem. Assim o último foi à coisa de uma semana. Foi um documento de um trabalho de estudos de mercado. Portanto, há aqueles documentos pessoais todos os dias, mas o mais importante que eu fiz ultimamente foi há uma semana. Era um trabalho sobre... da cadeira de estudos de mercado, da parte exploratória. Foi um trabalho que foi feito em grupo, portanto foi difícil compactá-lo todo num só porque havia coisas em vários computadores. Nomeadamente também no da minha colega de grupo. Eu chamei-lhe, tentei-lhe chamar 'trabem1' para saber que

era o primeiro trabalho de estudos de mercado...

[R] Portanto houve co-autores. Para além das pessoas que ajudaram a escrevê-lo houve mais alguém que tenha lido para corrigir algumas coisas?

[I] Há, sim. O meu namorado, geralmente, por acaso lê os trabalhos para ajudar algumas coisinhas de português, alguns erros que nos passam. Às vezes colegas da escola, peço aos meus colegas de português da escola.

[R] Neste caso concreto o documento foi enviado por correio electrónico, talvez?

[I] Neste caso concreto não, porque os professores exigem a maioria deles, que os trabalhos sejam entregues em suporte de papel, uma vez que poderão... não querem estar a imprimir. Às vezes são muitas folhas. Mas já tenho trabalhos que entrego por email.

[R] Mesmo entre outras pessoas, entre os vários colegas do grupo, quando o estavam a fazer, houve trocas?

[I] Sim, sim sim. Houve trocas por email. Aliás, a maioria dos trabalhos de grupo, acho que isto ajuda, pelo menos com a pessoa com que eu trabalho, uso muito a troca do que é que já fiz por email para irmos confrontando as ideias e ver em que ponto é que temos o trabalho.

[R] E era um documento Word?

[I] Era Word.

[R] Tinha algumas tabelas, alguns gráficos, alguma coisa assim que...

[I] Coisas muito simples. Tinha uma tabela que nem fizemos em Excel, foi apenas necessário fazer em Word e gráficos não, não tinha.

[R] E foi preciso consultar outros documentos para fazer esse?

[I] Propriamente não. Tínhamos um guião de um trabalho de grupo, tínhamos os documentos estudados pelas aulas, mas não foi preciso assim nenhuma bibliografia específica.

[R] Exacto. E foi feito aonde? Foi feito em casa e depois juntaram? Foi feito na [Universidade]?

[I] Então é assim: foi feita uma parte na [Universidade], foi feita outra parte na minha casa, em [Localidade], outra parte foi feita em Lisboa, cá em casa. Depois outra parte foi feita pela minha colega. Depois juntámo-nos um dia num fim-de-semana e compilávamos tudo.

[R] E agora onde é que está a versão final guardada? Está em disquetes, está no computador?

[I] Está em vários sítios. Portanto, eu uma coisa que faço, uma vez que os computadores não são de fiar e não sabemos se eles um dia endoidecem e perdemos os documentos, tal como fiz no primeiro semestre das cadeiras, vou guardando numa pasta do computador chamada [Universidade], e guardo... dentro do [Universidade] há o primeiro e segundo semestre, e dentro do semestre tenho as pastas das várias cadeiras, uma vez que já estava escaldada com essas coisas. Portanto está na pasta de Estudos de Mercado, e ainda tenho uma cópia em [Localidade]. Vou transportando as coisas que faço em [Localidade] e tenho feito as pastas actualizadas. A ideia é, no fim de ter material suficiente para gravar um CD, gravar então o CD, aliás dois. Eu nessas

coisas costume gravar dois (risos).

[R] E qual era o tema do...

[I] Festas de crianças. Era sobre festas de crianças. O que é que as pessoas gostavam, qual é que era o sonho das pessoas para realizar uma festa para as suas crianças...

[R] Quando estava a ser produzido o documento houve alguma tarefa particularmente penosa? Às vezes há aqueles percalços e é preciso, sei lá, rescrever parte do documento, ou fazer umas contas à mão que foi muito chato...

[I] Por acaso não. Mas já tive casos desses. Perdi completamente os documentos. Não sei se quer que refira?

[R] Não, para já não. É melhor cingirmo-nos a este documento. Ainda sobre este documento, e para terminarmos sobre ele, sobre a vida pessoal, no sentido de uma coisa que pudesse estar marcada na agenda, há alguma coisa que vem à memória? Por exemplo, foi no dia do aniversário de alguém, ou foi no dia em que faltou a gasolina e por isso levei mais tempo a chegar ao trabalho, uma coisa assim. Alguma coisa do mundo em geral, talvez. Alguma notícia, alguma...

[I] Não porque foi num Domingo, portanto... Foi nas vésperas da Guerra do Golfo, do Golfo não, desculpe, do Iraque, mas não houve assim nenhuma data especial. Não, não há nada assim...

[R] Então quanto a este documento estamos conversados. Agora fazíamos o mesmo exercício mas para um documento que tenha sido produzido há mais tempo. Um ano, dois anos, uma coisa mais antiga.

[I] Foi um terrível!

[R] Então, vamos a ele!

[I] Foi um trabalho que era... Eu fiz estágio, completei o estágio há coisa de mais ou menos dois anos, e tinha que fazer uma monografia científica. Uma primeira parte da monografia era sobre a matemática na antiguidade, mas a segunda parte é que foi mesmo complicada, porque a segunda parte era pegar numa matéria qualquer de matemática que fosse leccionada, nos fizemos sobre trigonometria, pegar como ela foi dada no 9º, 10º, 11º, 12º e fazer alguma ponte com a faculdade. Portanto, nessa altura eu tinha um hábito que hoje considero terrível que era escrever tudo em papel e depois então fazer uma coisa que era passar tudo a computador. O que é que me aconteceu? Eu já tinha cerca de 20 páginas mais ou menos de trabalho completo. Era um trabalho muito chato porque era um trabalho em que aquilo estava dividido em tabelas. Portanto, em que uma parte eu dava a parte lúdica, como é que se poderia introduzir a matéria, e na segunda coluna estavam realmente as definições matemáticas correctas. Tinha muitos desenhos, feitos no Word. Tinha muitos gráficos. Tinha muitos rococos. Isto é em itálico, isto é em negrito. Portanto, foi um trabalho que eu quando o perdi fiquei mesmo muito em pânico. Perdi porque não o consegui salvar, não tinha cópias de segurança, ou o que estava estava muito, muito desactualizado, não me servia de nada. Foi terrível. Foi mesmo terrível.

[R] Mas depois acabou por ter que ser feito outra vez e agora está guardado...

[I] Tudo de novo, outra vez. Agora está. Aliás, acho que isto é importante frisar. Eu a partir daí comecei a ter uma técnica quando faço os trabalhos. Vamos imaginar, esse trabalho era uma

monografia, e então se estivesse a fazer ontem, eu gravei como monografia 24-3. Hoje, o que eu escreveria seria 25-3. Portanto o que eu perco é o trabalho de um dia. Às vezes quando são mesmo coisas muito importantes, com gráficos, com tabelas, coisas muito chatas e morosas, até chego ao cúmulo de gravar 25-3a, de a.m. e 25-3pm. Faço isso.

[R] E então...

[I] ... Onde ele está guardado agora.

[R] Está guardado num CD, não é?

[I] Está guardado, mas posso falar, que foi outra bronca. Eu acabei o estágio, tinha muitos materiais e fui toda contente gravar um CD. Gravei o CD mas aquilo correu muito mal. Eu pensei... abri o CD novinho gravado, vi lá as pastas todas, tive o azar de abrir uma ou duas pastas em que na realidade estavam lá os documentos, mas passados uns meses já não estavam lá alguns dos documentos que tinha guardado em algumas pastas. Portanto, o que é que me tinha acontecido? Eu tinha tido problemas no computador, decidi formatá-lo, e tinha tudo perdido. Tive que contactar uma empresa para me recuperar os dados, o que ficou caríssimo. Mas era o trabalho de um ano. Valeu a pena.

[R] Houve alguém mais que tivesse ajudado a produzir esse documento, ou foi só uma pessoa.

[I] Era também um trabalho de grupo.

[R] Já nessa altura houve trocas de email, também?

[I] Nessa altura... nesse trabalho em concreto não porque estávamos as duas a fazer estágio e viamo-nos todos os dias.

[R] Portanto ele produzido no local de trabalho, vá, na escola?

[I] Não. Foi, novamente, em minha casa, em casa dela e...

[R] ...juntavam-se ora na casa de uma ora na casa de outra...

[I] Exactamente. E compilávamos as coisas.

[R] E foi para quê, o trabalho?

[I] O trabalho foi uma monografia científica de quando eu estava a fazer estágio.

[R] Portanto, era um daqueles documentos que faz o resumo do estágio?

[I] Não. Aquilo era assim. Nós tínhamos três partes distintas. Tínhamos a orientadora pedagógica, que fizemos uma monografia para ela. Tínhamos a orientadora científica que a monografia era um trabalho de quatro que era sobre a história da matemática. Era este trabalho que eu estava a falar, que era um grupo de duas pessoas, e era um trabalho individual. O dossier de estágio já era com uma orientadora da escola. Eram várias coisas distintas.

[R] E foi preciso consultar outros documentos para produzir esse?

[I] Sim, sim. Muitos.



[R] E há assim algum que venha assim à memória ainda agora?

[I] Sim. Há um site muito bom sobre a trigonometria. Eu tenho a morada em casa, não sei qual é, mas é um site muito bom. Tinha figuras animadas. Obviamente não podemos por no trabalho, mas referimos e explicámos o que é que se podia fazer para fazer aulas diferentes. Esse documento foi... E fora os livros do secundário...

[R] Em termos do mundo em geral. Mais uma vez, há assim alguma coisa que tenha acontecido nessa altura que esteja presente?

[I] Eu lembro-me que foi... A entrega do trabalho final foi no dia 31 de Maio. Portanto, este não foi bem entregue no dia 31 de Maio mas é como se tivesse que entregar todos no dia 31 de Maio. Lembro-me que nessa altura foi uma altura muito complicada da minha vida porque estava muito cansada. O estágio não me tinha corrido às 1000 maravilhas e até me recordo de ter estado no dia 1 de Maio a acabar o trabalho. Portanto eu devo ter entregue no dia 2, dia 3, e o trabalho individual e todos os três juntos no dia 31.

[R] Não sei se há mais alguma coisa que pareça interessante sobre este documento?

[I] Sobre este? Foi o facto de desconhecer, se calhar, algumas coisas, que na altura, pronto, adoptei. Quer dizer. Depois adoptei aquela história do 25-3, etc. Na altura nem sei se havia os USB disk, que são muito melhores que as disquetes. Permitem... porque é assim. Aquele trabalho, aquelas 20 páginas nem em três disquetes cabiam. Portanto era um trabalho pesadíssimo. Portanto era um trabalho difícil de fazer cópia de segurança sem ser no próprio computador. E lembro-me que nessa altura fiquei em pânico, e desesperada e... Foi mesmo...

[R] Então, já despachámos este. Falta só um terceiro documento. Normalmente até é o mais rápido de descrever, porque é um documento que tenha sido feito por outra pessoa. Feito por outra pessoa nossa conhecida ou por exemplo um artigo científico que tenha... Qualquer coisa que nos tenha chegado às mãos. Que alguém tenha mandado, ou que nós tenhamos encontrado na Internet, por exemplo. E que tenha tido interesse suficiente para ser guardado algures no disco. Portanto mais uma vez, identificar um documento...

[I] Ok. Então identifico... É assim. A minha irmã era enfermeira, e fez... Na altura era uma licenciatura de três anos, teve que fazer um chamado complemento que foi o ano passado. Portanto, eu ajudei-lhe a fazer bastantes trabalhos, a formatar, etc. Portanto, o que é que acontecia. Ela na altura ainda não tinha computador e guardei muitas coisas dela em casa. Além de guardar muitas coisas dela tinha documentos que recebia dela, entre outras, em disquetes. Documentos partes de trabalho que as colegas davam. Estes não foram muito problemáticos. Abri uma pasta dela. tal como fiz para a [Universidade] abri uma pasta chamada [Nome da Irmã], uma pasta das cadeiras que ela me dizia, o nome das cadeiras, e tinha tudo assim organizado. O problema com este trabalho, uma vez que eu já estava escaldada tive umas certas precauções, não é, mas o único problema que tive foi um documento que perdi. Ainda hoje estou para saber como é que o perdi. Ou se foi a colega que não gravou na disquete, realmente, e não havia. Porque eu lembro-me de abrir o documento, mas o documento estava estranho. Como é que eu hei de explicar... Estava com uns símbolos muito estranhos, e depois dizia-me, quando eu via o simbolzinho, não tinha o símbolo do Word. Tinha o símbolo de um fragmento. Julgo eu que era um fragmento. Eu não sabia que símbolo era aquele. Aliás, tive que pedir para a minha irmã pedir à colega para gravar novamente, foi novamente com esse problema, até que decidi a tal história de ela me mandar por

mail. E aí não sei. Veio o documento direitinho. Não sei qual era o problema. Se era na disquete se era... Não faço ideia.

[R] E isso foi mais ou menos há um ano?

[I] Não foi... ela acabou agora em Março... Deve ter sido na altura de Junho, mais ou menos.

[R] E sobre o que é que tratava esse documento, no final dessa história toda?

[I] Esse documento era para uma cadeira chamada acho que era metodologias da investigação, não tenho a certeza, e era sobre... o objectivo do trabalho era os primeiros socorros nas escolas. Ver quais é que eram as lacunas que haviam na formação dos professores primários, principalmente. Tentaram arranjar uma caixa de primeiros socorros com tudo o que seria necessário. Uma listagem para a pessoa verificar de dois em dois meses se tinham tudo, se faltava alguma coisa. O que é que se devia fazer se por exemplo um miúdo fosse picado por uma abelha, o que é que eu faço. Era um trabalho assim desse tipo. Recolhi fontes eu própria. Já sei mais ou menos a estrutura da entrevista (risos). Recolhi fontes eu própria uma vez que era uma coisa para professoras primárias e até podia ajudar os miúdos fazerem trabalhos, tinha que ser um trabalho relativo, com imagens. E lembro-me de ir buscar imagens ao site do Sapo. Há lá uma coisinha que se chama Primeiros Socorros. Tem umas imagens muito engraçadas. Tem um manual de primeiros socorros também. Tem umas imagens muito giras de picadas. Por exemplo, na fractura de uma perna como é que se deve fazer a tala. Isso lembro-me de ir buscar.

[R] E depois da... do trabalho feito, acabou por ser um documento Word. Uma vez inseridas as imagens...

[I] Sim, sim.

[R] Foi feito em casa, suponho?

[I] Foi, foi. Mas mais uma vez é a mesma história. Era um trabalho de grupo, mas o trabalho final foi feito... Nessa altura foi iniciado no meu computador. Entretanto a minha irmã comprou um computador e finalizou o trabalho lá. Não tenho a certeza de que ele esteja seguro, porque ela faz desaparecer pastas miraculosamente. Portanto não tenho a certeza se ele ainda existe.

[R] Houve várias versões do trabalho que tenham sido produzidas explicitamente, de propósito, ou acabou por ser só a versão final que ficou guardada...

[I] Não. Porque é assim. O trabalho era a versão final. Nos anexos constava o manual para entregar nas escolas. Elas entregaram completamente o manual, a tal ficha que eu referi de verificação de dois em dois meses, e constava também de umas reflexões que elas foram fazendo ao longo do trabalho, que foi aproveitado também para outra cadeira delas. Portanto o trabalho está um bocadinho partido. Está a versão final que eu tive o cuidado de por os anexos e o tal guião, o manual, já não sei como é que lhe chamaram, mas há várias partes do trabalho. Estão é todos na mesma pastinha a dizer trabalho de...

[R] Pronto. Então, para terminar, da vida pessoa, dessas coisas, há assim alguma...

[I] Ainda me lembro. Lembro-me que foi uma altura que estava com muitos, muitos testes para corrigir. Eu tenho praticamente a certeza que deve ter rondado aquela semana antes da... do Natal. Deve ter sido mais ou menos entre o 8 de Dezembro e o 15 de Dezembro. Tenho a ideia

de que foi isso, porque eles eram para entregar num dia, na ante-véspera do Natal, portanto foi nessa altura.

[R] Mais uma vez, a menos que haja mais alguma coisa que pareça interessante?

[I] Eu, por segurança, não sei se ela vai precisar daquele trabalho, confesso, surripiei-lhe o trabalho do computador e tenho no meu computador. Porque ela, como eu referi ela desaparece com pastas não sei como (risos)

[R] Então já está. Obrigado!

## Interview 15

[Researcher] O que te peço que faças é que penses num documento concreto, numa coisa qualquer que tenhas feito hoje ou nos últimos dias, há uma semana no máximo, e depois de ter identificado esse documento concreto, descreveres tudo o que te lembras sobre o documento.

[Interviewed] Tenho que me lembrar de um documento concreto. Sim... Acho que já me lembrei.

[R] Então, do que é que te lembras sobre ele?

[I] Do que é que me lembro. Foi prái há uma semana. Era para uma apresentação de uma reunião semanal. E é um documento em Powerpoint e tinha uma parte duns gráficos que foram feitos noutra programa e que depois coloquei...

[R] Qual era o tema do documento?

[I] Eram os resultados do trabalho semanal.

[R] Foste só tu que fizeste ou houve algum co-autor, alguém leu para tirar umas gralhas...

[I] Por acaso nesse fui eu que fiz mas até é um trabalho que estou a fazer com outra pessoa. Mas fui eu que fiz a apresentação. E depois mostrei-o antes de fazermos a apresentação.

[R] E fizeste aonde? Em casa, aqui?

[I] Fiz aqui.

[R] E agora onde é que está o ficheiro, se fosse preciso...

[I] O ficheiro está no portátil.

[R] Pronto. Chegou a ter várias versões ou foi só... Às vezes uma versão para ti, uma versão com menos informação para apresentar...

[I] Não. Foi uma versão pronta do documento.

[R] E precisaste de consultar outros documentos para fazer esse? Ou já usaste esse depois para fazer outros?

[I] Não. Precisei já de outros anteriores para fazer esse.

[R] E sabes quais são ainda? Se fosse preciso refazer tudo conseguida?

[I] Sim. Também eram os resultados da semana anterior.

[R] E quando estiveste a fazer o documento houve alguma complicação especial, alguma tarefa que tenha dado assim particularmente mais trabalho? Ou correu tudo...

[I] Correu normal.

[R] Nada que se destaque nem pela dificuldade nem pela facilidade? Nenhuma tarefa que...

[I] Por acaso tinha lá uns gráficos. Eu estava atentar fazer coisas novas. Estava a tentar por uns ao lado dos outros só que não deu.

[R] Então e em termos de trocas de email. Mandaste o documento a alguém por mail ou não? Trocaste com uma disquete?

[I] Não. Depois passei para a rede para a outra pessoa que está a fazer... a trabalhar nisso, ter acesso.

[R] E da vida pessoal ou do mundo em geral há assim alguma coisa que esteja associada ao documento? Por exemplo demorou mais tempo a fazer porque tiveste que ir a qualquer lado. Ou até ia começar a ser feito antes mas foi um dia que tiveste que ir ao supermercado de emergência e...

[I] Não. Foi normal.

[R] Não há assim também do mundo em geral ou das notícias da televisão, algo... Foi o dia em que a vizinha da frente escorregou e partiu uma perna...

[I] Não associo a nada.

[R] Não sei sobre esse documento se te lembras de mais alguma coisa?

[I] Assim qualquer coisa... Estou a visualizar-lo, neste momento.

[R] Em termos do conteúdo já sei que há as tais imagens. Mas não há assim mais nada que se destaque?

[I] Não. Se calhar não foi muito bom mas foi do que me lembrei logo, desse documento...

[R] Então agora vais-te tentar lembrar doutro, mas de um documento antigo. Uma coisa com dois anos, três anos. E vamos repetir o exercício, de descrever o documento.

[I] Sim... (pausa) Isso começa a ser mais difícil... (pausa)

[R] Um relatório de uma cadeira, uma carta, um...

[I] Sim, sim... (longa pausa). Sim. Já me lembro.

[R] Então, do que é que te lembraste?

[I] Só pode se até um ano?

[R] Não. De um ano, dois anos, três anos...

[I] Antigo. Já me lembrei de um trabalho escrito que foi feito para uma cadeira.

[R] E foi feito quando, qual era a cadeira?

[I] Era uma cadeira de Engenharia de Sólidos e de Fluidos.

[R] E foi quando?

[I] Foi no primeiro semestre do quinto ano... Sim, no primeiro semestre do quinto ano.

[R] Onde é que o fizeste?

[I] Fiz em casa e aqui na [Universidade]

[R] Portanto andavas com ele numa disquete, ou coisa que o valha?

[I] Sim. Provavelmente numa Zip, porque era grande.

[R] E onde é que agora ele está guardado?

[I] Agora ele está guardado... não sei... Provavelmente está no computador lá de casa ou na Zip. Eu tenho uma Zip que foi de backup do [Computador na Universidade], mas se me perguntares exactamente... Está num desses sítios.

[R] Era trabalho de grupo? Houve alguém que te tenha ajudado a fazer?

[I] Era individual.

[R] Mas houve alguém que leu para dar algumas ideias ou assim, ou...

[I] Não.

[R] E foi trocado por email com alguém?

[I] Também não.

[R] E era de que tipo, o documento?

[I] Era um documento Word.

[R] E lembras-te de alguma coisa no seu conteúdo que se destaque? Algumas cores, algum gráfico vistoso ou difícil de calcular?

[I] Não. Era um trabalho sobre um tema e tinha umas figuras...

[R] Qual é o tema, já agora?

[I] Era sobre encapsulação de moléculas de ADN...

(telefonema curto...)

[R] E não houve nenhum co-autor nem nada. Foi individual.

[I] Não. Foi individual.

[R] Ok. E outros documentos foi preciso consultar? Ou consultar para fazer ou então outros de que te lembres que leste na mesma altura ou...

[I] Aquilo era com base em artigos. Fiz pesquisa de Internet.

[R] Há assim alguma referência que te venha à memória? Sabes que foram artigos, mas...

[I] Sei que vi alguns livros. O nome dos livros não me lembro...

[R] E mais uma vez, quando estavas a fazer o documento houve assim alguma coisa particularmente difícil ou particularmente fácil? Alguma tarefa que se destaque? Foi preciso fazer umas experiências no laboratório ou...

[I] Não, não. Era só teórico.

[R] Ok. Da vida pessoal e do mundo em geral?

[I] Se alguma coisa me afectou enquanto estava a fazer o documento?

[R] Sim. Se te lembras nessa altura. De que te lembres nessa altura.

[I] Nessa altura estava a fazer um projecto. Portanto... Já não me lembro exactamente em quê mas de certeza que afectou a disponibilidade ou os tempos que tinha reservados para uma coisa e para outra.

[R] E do mundo em geral, lembras-te de alguma coisa que tenha acontecido, alguma notícia? Algum percalço? Viste algum acidente de automóvel nesse dia?

[I] Não. Que me lembre não (risos).

[R] Ok. Várias versões do documento. Teve, não te lembras?

[I] Não sei exactamente versões, mas foi sendo construído. Sei que imprimir prai umas duas vezes para ler e corrigir e meter outras coisas pelo meio.

[R] Quanto a este também estamos conversados, a menos que te lembres de mais alguma coisa.

[I] Não.

[R] Então falta o terceiro documento, que é um bocadinho mais rápido, normalmente. É um documento que tenha sido feito por outra pessoa. Um paper que tenhas lido, uma carta que tenhas recebido. Em formato electrónico, evidentemente.

[I] Com idade... Pode ter sido ontem como pode ter sido...

[R] Sim. De preferência uma coisa assim mais de um mês, dois meses...

[I] (pausa) Um paper...

[R] Um paper ou outra coisa qualquer...

[I] Um documento que me tenham dado...

[R] Ou que tenhas encontrado na net.

[I] (pausa) Como li tantos...

[R] Tens é que pensar num em concreto.

[I] (pausa)

[R] Não precisas de saber tudo sobre ele. Escolhe um e melhor ou pior...

[I] Sim. Já.

[R] Então força. Era sobre o quê, quando é que o viste?

[I] Era um... Por acaso era um relatório de uma pessoa com um trabalho que tem a ver com o meu e já o vi pela primeira vez há mais de um ano, e depois, de vez em quando... Acho que já li mais uma vez ou duas.

[R] E onde é que ele está guardado?

[I] Está num dossier. Tenho é que descobrir qual.

[R] Mas existe em formato electrónico?

[I] Eu não tenho em formato electrónico. Tem que ser uma coisa em formato electrónico?

[R] Ajudava...

[I] Pronto... (pausa) Sim. Pronto. Já sei!

[R] Força, então!

[I] É um paper.

[R] Quando é que o viste? Quando é que o arranjaste?

[I] Há seis meses, talvez. Foi-me mandado por email e tenho uma cópia em papel.

[R] Sabes quem foi que te mandou? Se precisasses de pedir outra vez, por exemplo?

[I] Sei.

[R] E trata sobre?

[I] Trata sobre uma reacção que eu preciso de fazer. Portanto, é aquela que neste momento estou a fazer.

[R] E onde é que ele está armazenado agora? Sem ser a versão em papel, evidentemente?

[I] Está neste computador.

[R] Sabes o nome do autor?

[I] Sei.

[R] Qual é?

[I] Martin Stitsman (?)

[R] Sabes mesmo (risos).

[I] E teve outros co-autores. Esses é que já não me lembro.

[R] Está relacionado com outros documentos? Há outros documentos associados a esse?

[I] Sim. Outros artigos, outros...

[R] Coisas que tu tenhas feito, eventualmente, e que tenhas usado esse como referência?

[I] Tenho trabalhado naquilo.

[R] Portanto, da tua vida pessoal, está relacionado com o teu trabalho?

[I] Com o meu trabalho, sim.

[R] Mas assim outro tipo... Encontraste-o no dia do aniversário de alguém...

[I] Não. Foi-me enviado na altura que se sabia que eu ia fazer isto e...

[R] Lembras-te do conteúdo dele? Assim algum gráfico, algum padrão, algumas cores assinaláveis?

[I] Cores não porque era uma versão a preto-e-branco. Eles não... tinha umas figuras, uns gráficos. Também tinha.

[R] E do mundo em geral?

[I] Não me lembro de nada...

[R] Foi no dia em que começou a guerra...

[I] Sobre o dia em que me mandaram?

[R] Ou o dia em que o leste, eventualmente, ou ias formatar a máquina e quase que te esquecias dele?

[I] Não (risos).

[R] E já agora, leste-o aonde?

[I] Li-o aqui na [Universidade]. Pela primeira vez. Depois acho que li mais uma duas vezes...

[R] Então, já agora, imprimiste-o, não foi? Não houve problema com a impressão?

[I] Não. Não houve problemas.

[R] Então a menos que haja mais alguma coisa que aches interessante?

[I] Não.

[R] Obrigado.



## Interview 16

[**Researcher**] O que eu começava por pedir era para pensar num documento concreto que tenha produzido hoje ou nos últimos dias. No máximo uma semana. E uma vez identificado esse documento concreto, que provavelmente será um ficheiro só no computador, contar a história do documento. Dizer tudo o que vem à cabeça sobre ele. Porque é que foi feito, qual era o tema...

[**Interviewed**] Aquilo que eu me lembro que tenha produzido assim um documento começado e acabado numa semana tem a ver com uma exposição que eu fiz ao provedor de telecomunicações. E isto prende-se com um contencioso que eu tenho com a PT comunicações, que embora isto pareça de alguma forma irrisório quando se fala no valor envolvido, eu penso que as grandes empresas têm que ter tanta honestidade como os indivíduos mais paupérrimos que existem na nossa sociedade, e penso ainda mais que as empresas grandes devem dar até o exemplo, quando ainda por cima são empresas com uma carga institucional muito forte. Em Novembro resolvi experimentar ADSL e nesse sentido aproveitei uma campanha que estava a ser feita em que o custo de adesão era zero e em letras pequeninas estava lá escrito que teria que se pagar 50 euros de activação. De qualquer modo a campanha garantia que eu podia fazer a experiência durante um mês sem qualquer compromisso da minha parte. Ao fim de pouco mais de uma semana, uma semana e meia, concluí que não me interessava aquela tecnologia e resolvi contactar exactamente a PT comunicações, que era o que estava nas cartas que acompanhavam o equipamento, manifestando a vontade de desistir do serviço e solicitando informações sobre como devolver o kit. Tudo isso funcionou bem. De acordo com as instruções que eu recebi enviei para a morada que eles me indicaram. De acordo com as instruções que eles me deram. Passado um mês e meio, ainda não havia qualquer reembolso dos tais 50 euros. Fiz meia dúzia de contactos no sentido de saber em situação estava a minha reclamação. Como nada disso surtiu efeito chegou à altura de recorrer a um dos vários provedores que nós temos, especializados com certeza, neste caso, em comunicações. Portanto o documento que eu elaborei foi uma exposição em que eu contava toda esta história com datas, com a minha argumentação, e com os anexos, fotocópias dos documentos que recebi. Sei lá, os talões de registo, essas coisas, e enviei então no início da semana para o provedor de telecomunicações. Foi o documento assim que eu me lembre de ter começado e acabado. Foi o único porque depois nós temos aqui no âmbito do meu trabalho. No âmbito do trabalho trabalhamos com muitos ficheiros mas normalmente são sistemas que...

[**R**] Que vão evoluindo mais ao longo do tempo?

[**I**] Não é bem isso que eu quero dizer. Eles foram construídos num determinado momento e neste momento estão em fase de exploração. Periodicamente nós produzimos dados, normalmente com uma referência anual. Não são coisas que eu construísse na semana passada. Demoram meses.

[**R**] E quanto a esse documento, sobre a PT Comunicações, houve mais alguém que o tenha ajudado a escrever, ou que eventualmente tenha lido?

[**I**] Não.

[**R**] E foi feito aqui ou em casa? Ou nos dois sítios, talvez?

[**I**] Foi feito aqui, feito aqui.

[R] Foi enviado por email a alguém ou...

[I] Não... Uma das coisas também foi feita via mail mas eu digo-lhe: mesmo ao nível desta instituição onde eu trabalho os emails institucionais geralmente não funcionam. Eu utilizo vários endereços de email para outro tipo de problemas. Por razões fiscais por exemplo, uso os endereços institucionais do Ministério das Finanças, em geral não recebo resposta. E aqui na [Empresa] também tenho feito isso e não recebo resposta. O que me dá a... Parece querer dizer que eles existem lá por existir, mas normalmente como eles não estão afectos a uma pessoa de vez em quando é capaz de passar por lá alguém para ver se tem emails mas não tem os resultados de imediatismo que pelo menos eu esperava. Eu tenho inclusivamente um caso no ano passado com um programa no âmbito da divulgação das ciências por todas as camadas da população... Acho que é 'Ciência e Vida', uma coisa assim. Havia várias manifestações programadas. Usei o email no sentido de saber informações sobre um determinado evento que iria acontecer em Ferreira do Zêzere, e o certo é que quando eu precisava da informação mais interessante que era saber se se confirmava que ele ia acontecer, eles não me responderam. Se fosse o endereço 'fulanotal@qualquercoisa' se calhar ele respondia. Como era institucional não responderam.

[R] E foi um documento escrito no Word?

[I] Foi. foi, foi. Eu tudo o que são textos utilizo o Word. Normalmente costumo distinguir bem o que devo utilizar. Em geral costumo analisar e escolher o produto que se adequa mais à necessidade que eu quero resolver. Se é um texto, o Word. Se é um cálculo o Excel, se for manipulação de tabelas um SGBD qualquer. Costuma ser o Access que é o que temos aqui mais disponível.

[R] Houve alguma tarefa que se destacasse na produção do documento? Por exemplo uma coisa que deu especial trabalho, ou que até parecia que ia dar mais trabalho mas depois na realidade até foi mais fácil do que se pensava...

[I] Sei lá... quando nós estamos a escrever um texto e vamos organizá-lo por pontos costumo utilizar a facilidade que o Word tem de ir criando os pontos automaticamente.

[R] Exacto. Mas neste caso em concreto não houve assim nada que...

[I] Não, utilizei, utilizei. Quando estava a especificar tudo o que aconteceu desde o momento em que eu resolvi fazer a experiência até ao momento em que eu resolvi que teria que recorrer à influência desse provedor, pus 'ponto 1 tal, ponto 2 tal...' E aí usei essa facilidade que o Word tem.

[R] E neste momento onde é que ficou guardado o documento? Se fosse preciso imprimir outra vez, por exemplo?

[I] Isso então já temos que recorrer aqui a outra secção porque a máquina usada... eu sou um inovador, pelo menos classifico-me como tal... Portanto, está aqui dentro (mostra uma flash-drive USB).

[R] É um daqueles... USB.

[I] Um disco, uma memória que funciona como disco.

(...)

[R] E quando... Chegaram a existir várias versões do documento? Uma versão preliminar que depois foi...

[I] Vamos lá ver... Isto é um bocado como antigamente se fazia o rascunho, se dava à dactilógrafa e ela fazia-nos a prova e às vezes ainda dizíamos 'olha, agora vais ter que escrever tudo outra vez mas este parágrafo aqui não está bem, não fica aqui bem desta forma como está, substitui por este.' Claro que com as facilidades de hoje é muito mais fácil. Houve várias revisões. Aliás o documento não demorou um dia a fazer. Como eu fui fazendo aqui, às vezes eu escrevia um bocadinho, e parava. Tudo aquilo que era factual eu tinha uma listagem. Depois a argumentação, digamos, foi sendo feita e foi sendo melhorada. Mas não há v1, v2, v3.

[R] Então só para terminar sobre este documento... Sobre a vida pessoal no sentido de uma coisa que até pudesse estar marcada na agenda... Por exemplo, será que foi feito no dia do aniversário de alguém conhecido? Alguma coisa desse género?

[I] Não, não, não.

[R] Ou do mundo em geral, talvez? O dia em que houve uma notícia qualquer importante na televisão?

[I] Não, não, não. Isto calhou a ser naquela altura... não foi porque tivesse feito um mês sobre a reclamação que eu fiz escrita e que enviei registada. Nada disso. Aliás, faltavam ainda uns dias para um mês. Não houve assim nada que despoletasse a acção e que me fizesse decidir por fazer. Um dia lembrei-me, fui aqui à lista telefónica, consultei a lista telefónica da PT. Vi que havia um provedor. Digo assim 'pronto, deixa lá, pelo menos chatear-vos hei de vos continuar a chatear' e resolvi fazer a exposição e enviar. Não há portanto nenhum evento que tenha assim...

[R] Ok. Portanto, quanto a este documento, terminámos. Agora repetíamos o exercício, descrever um documento, mas desta vez tentando lembrar um documento que tenha sido feito há já mais tempo. Dois anos, uma coisa nessa ordem de grandeza.

[I] Dois anos. Isso obriga-me então...

[R] Pode ser mais recente, se for mais fácil lembrar...

[I] Bom, pode ser este sistema, onde eu participei. Pode ser... Andará por aí dois, três anos. Claro que não é um documento em Word, é um sistema de ficheiros em Excel, não sei se quer que eu vá por aí?

[R] Sim, é possível. Então sobre o que é que tratava...

[I] Olhe, a necessidade surgiu aqui no ambiente de trabalho. Era necessário agregar uma série de dados produzidos por vários colegas que no fim tinha que sintetizar-se num único. E havia... havia uma experiência do passado que passava quase por soluções ad-hoc. Cada ano era uma solução diferente, e eu resolvi propor que se introduzisse alguma ordem no caos, digamos assim, e propus que fossem definidas algumas regras, fossem definidas algumas estruturas que se mantêm fixas de um momento para o outro. Embora isto não seja mesmo, digamos, a forma mais simples, mais segura e mais uma série de características que eu tenho a certeza que este sistema não é adequado, mas tinha que haver um equilíbrio entre os gostos, as preferências das pessoas, por um lado, e por outro esta necessidade de agregação no sentido de sintetizar o documento. Conjugando estas duas coisas eu acabei então por propor que se construísse então um sistema

baseado em ficheiros de Excel, ficheiros de Excel que se interligavam através de links célula a célula. Que havia mesmo alguns procedimentos de verificação de coerência, a que eu costumo chamar bandeirinhas, que nos mostravam se alguma coisa estava mal. Pronto, e isso foi aceite, foi desenvolvido, e neste momento é o que está a ser usado. Foi melhorado, claro que foi melhorado. Nunca se conseguem produtos finais perfeitos. Se me disser que tabelas de dados era melhor, eu concordo, mas há pessoas aí que não dispensam a facilidade do Excel.

[R] Portanto, pelo que percebi, desta vez houve co-autores, ou pelo menos houve opiniões dadas por outras pessoas sobre...

[I] Houve até porque é assim. Houve porque quando existem conjuntos de pessoas, grupos, que têm que atingir um objectivo que é comum, claro que quem está com uma tarefa de organizar a informação em termos informáticos terá que consultá-los. não pode ser de outra forma. Senão, já se está a ver, uma pessoa que não colabora, coopere e... Eu até costumo dizer que confrontação e colaboração começam ambos por 'c'.

[R] E neste momento ainda é possível, se fosse preciso interpelar outra vez essas pessoas, ainda é possível saber exactamente quem foram...

[I] Sim, sim sim. Aliás, há um pequeno relatório em que é descrito o que se fez, o que se pretendia fazer. Isto é um processo sempre em melhorias sucessivas. Neste momento eu estou a propor outra solução. Mais ao nível das regras.

[R] Esse documentos em Excel foram trocados depois com essas pessoas, ou enviados por email? Ou foram colocados na rede interna, talvez?

[I] Isto é assim... Isto é como se nós tivéssemos aqui vários post-its, e cada um é dono de um. Mas este post-it, à partida, está dividido segundo uma determinada quadrícula, uma estrutura. De tal forma que depois eu consigo através dele, outro exactamente igual, agregar todas as que estão a ser produzidas pelos meus colegas. Em termos de records isso já é outra organização que já não me diz respeito, é mais geral. Desse ponto para a frente diz-me respeito, também a forma como eu organizei a árvore de directorias, tem a ver com esta particularidade, que é cada um... eles começavam com estes post-its e depois há um, que está numa directoria diferente, obviamente, que agrega essa coisa toda.

[R] Portanto o local em que as coisas estão armazenadas é fulcral para...

[I] Sim, sim, sim, sim. Onde estão, como estão...

[R] Houve, mais uma vez, alguma tarefa que tenha sido particularmente penosa na produção dessas folhas de cálculo? Ou então particularmente fácil? Há assim alguma coisa que se destaque?

[I] Se aqui a reacção de alguns colegas também é interessante, houve alguma reacção contra. Foi preciso mostrar resultados para perceberem que valia a pena. E é curioso, normalmente a níveis de pessoal de chefia... Não sei porquê, mas nesta [Empresa], pessoal de chefia não mostra inovação, dinamismo, nem coisa nenhuma. 'Laissez Faire, Laissez Passer'. O que eles conhecem e andar. Tenho a impressão que é a influência forte da postura de funcionário público. Muito enraizado...

[R] Então, mais uma vez para terminar sobre este documento, sobre a vida pessoal houve alguma coisa na altura que...

[I] Eu penso aí que talvez a minha formação na área da informática... Eu na área da informática, para além das cadeiras que se têm habitualmente num curso do género que eu tirei... Eu costumo rir-me quando falo nisto, mas deram-me o Pascal, a linguagem C, o Assembly, mas claro que só os rudimentos, e eu perguntei muitas vezes 'Para que é que eu quero isto? Isto não me serve para nada'. Mas claro, depois também tive outras coisas que me ajudaram. Tive arquitectura de sistemas de informação, que foi importante, por exemplo, neste caso. E depois sou um autodidacta sobre estas coisas. Interesse-me sobre estas coisas e vou apanhando um bocadinho aqui, um bocadinho lá. Nós hoje em dia temos uma fonte de informação que é a Internet. Quase qualquer problema é resolvido por lá. Eu há pouco tempo, e no âmbito do aperfeiçoamento deste sistema que como já lhe disse funciona em Excel, há um pequeno senão nos sistemas desenvolvidos a partir de várias folhas de Excel, que é o facto de quando produz alterações numa folha de Excel que, digamos, está a jusante, e quero que essa alteração se repercuta a montante. Ou seja, se quiser, naquelas que são as folhas mais elementares e depois quando ela sintetiza no fim, tenho que abrir todas e fechá-las. E eu só para dizer que isto recolher informações e ir aprendendo com o tempo, surte o seu efeito e hoje a Internet consegue suprir muitas destas pequenas lacunas que eu tenho. Eu sabia como resolver, desde que eu tivesse à minha disposição um pequeno conjunto de procedimentos que me permitissem abrir e fechar, conseguia fazer isso. Eu fazia isso em Clipper, por exemplo. Nessa linguagem... o dBase também permitia programar. O Clipper é muito semelhante. Até acho que era da mesma software house que o Clipper. E então eu isso sabia fazer. Agora no Excel dentro do Windows, no Office, não sabia. Mas quando resolvi que era talvez a forma de eu garantir que quando fosse produzir uma qualquer alteração, ela se repercutisse em todo o sistema, era abrir todos e fechar todos, e quando senti a necessidade, eu fui à Internet resolver o problema. Outra coisa que eu também gostava que ficasse registado algures, numa folha, numa sheet, era a data em que foi feita essa... foram corridos esses conjuntos de procedimentos. Fui lá e também soube. Portanto, tenho aprendido assim. É aí que normalmente eu costumo aprender. Nesta fase, porque já houve uma fase em que eu recorria às pessoas da informática, com quem me dava bastante bem... Hoje não me dou mal, mas não estão tão disponíveis, porque estão longe. A distância também é uma barreira...

[R] só para terminar com este documento, faltam duas coisas muito breves. A primeira é confirmar o local em que ele foi feito, que deve ter sido aqui?

[I] Completamente.

[R] E a segunda é sobre o mundo em geral, se houve alguma notícia que... Foi feito no mês em que houve uma cimeira algures. Alguma coisa que... No dia em que o vizinho da frente teve um acidente.

[I] Não, não. Absolutamente não.

[R] Então quanto a este documento também terminámos. E quanto à entrevista estamos quase, porque só falta um documento, um terceiro documento, mas que normalmente é mais rápido de descrever, porque é um documento que tenha sido produzido por outra pessoa. Por exemplo um artigo científico, um artigo daqueles jornais online, uma página Web. Qualquer coisa que nos tenha vindo parar às mãos, e que tenhamos achado interessante, e que por algum motivo tenhamos decidido guardar, ou manter uma referência.

[I] Isso é, talvez mais complicado. Eu acho que foi o Virgílio Ferreira que um dia escreveu um texto em que dizia que um dia devíamos ter uma grande dificuldade do exagero de informação

que temos à nossa disposição escolher aquilo que é necessário. Isto é assim. Eu recebo tanta coisa aqui durante o dia, que agora lembrar-me de qualquer coisa que eu tenha guardado... Eu estou-me neste momento a recordar. Não sei é se será bom. Foi um Powerpoint em que as imagens que vinham nesse Powerpoint, nessa apresentação, eram imagens, algumas nem sei se seriam reais se seriam ficções, que versavam aspectos interessantes da natureza. A forma, as cores, pronto. Era assim este tipo de coisas. O deserto pode parecer uma coisa completamente desinteressante mas se for visto de um determinado prisma em que as sobras que são formadas pelas dunas... Sei lá. A imagem fica com um interesse, uma beleza que eu não imaginava. Antes de ver aquelas imagens não imaginava. Porque às vezes nós até nos podemos aqui cruzar com algo que parece sem qualquer interesse. No entanto os indivíduos que se dedicam a fixar imagens fixam aquela imagem porque nós passamos tantas vezes e nunca achámos qualquer interesse, e fixam a imagem e ali na imagem vemos realmente que é interessante. Eu imagino um mar de areia. No entanto aquela imagem achei muito interessante. E guardei isso aqui no meu caixote do lixo, ente aspas.

[R] E quando é que foi recebido, esse documento?

[I] Terá sido aí pelo início do mês de Janeiro. Portanto, aquilo, depois de passar os postais de Natal e Boas Festas que eu não guardei nenhum, penso que aquilo que guardei foi esse. Aliás, eu a organização dos emails, já que isso veio através de email, eu os emails tenho organizados por trimestre. Aliás, esse, não fazendo parte do trabalho, é o que eu me estou neste momento a lembrar que guardei.

[R] Não sabe o autor mesmo do documento?

[I] Aquilo é completamente anónimo. Não há qualquer referência a autores. Talvez em conseguisse através das características do documento. Aquilo que nós nunca preenchemos. Aquilo que é comum a qualquer aplicação, principalmente no Office, que é exactamente as propriedades. Dá para por lá o um sumário, etc, mas as pessoas não põem. Nunca pensam que isso é necessário...

[R] Quem enviou, por outro lado, é mais fácil lembrar-se, ou...

[I] Sim, sim. São pessoas conhecidas. Neste caso concreto até foi um familiar.

[R] E já, depois, reenviou a outras pessoas?

[I] Ah! Claro! Logo!

[R] Não houve nenhum percalço a tentar ver o documento? Por exemplo a extraí-lo do mail, ou...

[I] Não, não. Que me recorde não.

[R] E outros documentos? Há assim outros documentos que venham à memória quando pensa nesse? Talvez um documento semelhante, ou outro documento que tenha recebido na mesma altura?

[I] Lembro-me de qualquer coisa que recebi da BP, mas não me... eu sou cliente da empresa. Julgo que recebi alguma coisa da BP.

[R] Então só para terminar, mais uma vez aquelas perguntas um bocado mais genéricas, se quisermos, sobre a vida pessoal ou o mundo em geral. Se há alguma coisa que...

[I] Neste caso da BP o que eu me lembrei é que eu sou cliente da BP, e quando vi lá a referência suscitou-me algum interesse. Mas esse não o guardei! Aquele que eu guardei não está relacionado com absolutamente nada. Apenas o gosto que eu tenho por fotografias da natureza. Que versam o tema da natureza.

[R] Pronto. Não sei se há mais alguma coisa na linha do que temos vindo a falar que ache interessante referir? Senão...

[I] Não me recordo assim de nada. Eu pensava que me viesse perguntar coisas mais concretas. Por exemplo, eu quando atribuo um nome, normalmente esse nome eu tenho a preocupação que ele seja curto, mas que diga alguma coisa. Por exemplo, naquele caso concreto de que falámos, da PT, do provedor. Eu pus 'PTCom'. Se precisar de encontrá-lo, é um instantinho. Porque, por exemplo, aqui, nós trabalhamos com contas. Todos os ficheiros que são iniciados por 'conta' nós sabemos exactamente do que é que se trata. Se for 'ramos', por exemplo, se começarem por 'ramos', também.

[R] Existe uma nomenclatura, vá lá, bem definida...

[I] Sim, nós trabalhamos com muitas nomenclaturas. Mas depois há estas nomenclaturas que são criadas por nós próprios. Claro, há colegas que seguem outros critérios. Como eu lhe disse há bocado, o arquivo dos emails é feito por trimestres. Eu todos os trimestres inicio uma nova pasta, transfiro o reencaminhamento da pasta do trimestre que terminou para a pasta que eu criei nova do novo trimestre, e pronto. A partir dali sei que aquela pasta tem aquele correio. A forma como dei o nome também é fácil de ir ver o que lá está, portanto, 'email2003primeirotri, segundotri' e por aí fora. Talvez a esse nível talvez o facto de eu ter começado logo a brincar com estas coisas logo com o Spectrum, e depois a seguir quando começaram a aparecer os primeiros PCs, havia uma certa disciplina que nos era imposta pelas regras do DOS. Por exemplo, não podíamos ultrapassar os oito caracteres, excluindo a extensão. Isso de alguma forma obrigava a que existisse um esforço no sentido da atribuição de um nome que tivesse alguma coisa a ver com o que estava no ficheiro. Talvez por isso, ainda hoje, tenho essa preocupação. Hoje não com o limite dos oito, felizmente, mas de alguma forma tento incluir no nome o máximo de informação que me seja possível. Através de sigla, através de iniciais. Muitas vezes até escrevendo o nome.

## Interview 17

[Researcher] Portanto, um documento concreto que tenha sido criado hoje, no máximo numa semana. E depois de ter pensado num documento concreto contar a história dele. Dizer o que é que vem à memória sobre ele.

[Interviewed] Ainda criei hoje um. Por isso...

[R] Então?

[I] Basicamente o documento é uma folha Excel. Uma folha que nos permite... o objectivo é criar uma tabela. Há uma norma interna aqui na [Empresa]. Tu crias esse documento Excel. Incluí basicamente os campos que a tabela vai ter, índice, etc. Esse documento é enviado à administração de bases de dados, e através disso é gerado um conjunto de documentos, de ficheiros, que são os

scripts de criação da tabela e dos índices. Basicamente...

[R] Esse foi criado hoje, pelo que eu percebi?

[I] Exactamente.

[R] E aqui no local de trabalho, como é evidente.

[I] Exactamente.

[R] Houve alguém que tenha ajudado a fazer? Às vezes há documentos muito grandes em que uma pessoa trabalha numa parte e outra pessoa trabalha noutra, ou alguém que tenha revisto para acrescentar alguma informação que faltava?

[I] Aquilo acaba por ser uma macro em que apenas eu tenho que preencher alguns itens. Fí-lo sozinha.

[R] Ok. Portanto tem uma estrutura bem definida?

[I] Exacto. A estrutura do documento já está definida.

[R] Ok. E é sobre... é sobre o quê? Não é propriamente o tema porque era preencher a tal estrutura, mas suponho que isso seja para ajudar na gestão da base de dados e neste caso em concreto quais eram os tipos de dados que lá se iam meter? É fácil lembrar? Foi as tabelas que eu criei para os dados de...

[I] Para os dados que vêm do IRC, o imposto sobre o rendimento colectivo.

[R] Ok. E foi preciso consultar alguns outros documentos? Por exemplo, outras grelhas que já tenham sido feitas, para fazer esta? 'Posso aproveitar isto aqui', 'posso...'

[I] Exactamente, foi. Prá uns dois ou três documentos, ou tabelas parecidas, tabelas que tinham campos parecidos...

[R] Foi só alterar um bocadinho... Muito bem. Portanto no preenchimento do documento, ao fazer o documento, não houve nenhuma tarefa que tenha sido particularmente trabalhosa? Por exemplo, algo que até parecia que ia ser fácil mas que depois deu trabalho a encontrar? Coisas que foi preciso verificar? Foi tudo linear?

[I] Sim, porque neste caso como são documentos que fazes normalmente, já está sistematizado.

[R] E está armazenado aonde? Aqui no computador?

[I] Está na rede.

[R] E foi também enviado por correio electrónico, para a gestão da base de dados?

[I] Foi enviado... eles têm uma página na Web e que nós entregamos através dessa página.

[R] E em termos de algum acontecimento da vida pessoal? No sentido de uma coisa que até podia estar marcada na agenda? Por exemplo, hoje foi o aniversário de alguém? Houve algum percalço? Hoje ao vir para cá havia muito trânsito? Alguma coisa sobre o dia de hoje, ou sobre o mundo em geral? Alguma notícia que tenha passado hoje na televisão ou na rádio que seja marcante?



[I] Que eu me lembre...

[R] De que uma pessoa depois se lembre: 'fiz o documento no dia em que houve o engarrafamento'

[I] Hoje por acaso não associaria o documento a nenhum...

[R] A nenhum... nada de fora. Pronto. Quanto a este documento penso que estamos conversados. Foi rápido. Agora íamos repetir este exercício, dizer tudo o que vem à memória sobre um documento, mas íamos pensar num documento que já tenha sido feito há muitos anos, dois três anos. Mais uma vez pensar numa coisa concreta e depois descrever, para ver o que é que sobreviveu este tempo todo na memória.

[I] Está bem... Há dois, três anos...

[R] Pode ser qualquer coisa fora daqui. Durante o mestrado...

[I] Sim era o que me estava a tentar lembrar.

[R] Ok. Então o que é que acabou por ser lembrado.

[I] Basicamente foi mesmo sobre a minha tese. Há uns tempos pediram-me a minha tese e fui imprimi-la e já não me lembrava muito bem onde é que tinha posto isso lá em casa. Mas a estrutura em minha casa foi relativamente fácil de encontrar.

[R] Portanto estava armazenado em casa, pelo que eu percebi?

[I] Sim, em casa.

[R] O autor é evidente quem foi. Mas teve provavelmente co-autores, no sentido de revisores, pessoas que leram em todo ou em parte, que sugeriram alterações?

[I] Sim. Sim, sim, sim. A minha orientadora. Dei a mais duas ou três pessoas para ler.

[R] Ok. E foi trocado por email com essas pessoas?

[I] Sim.

[R] E foi quando, já agora?

[I] Isto foi cerca... foi há cerca de um ano, porque eu entreguei a tese em Abril, por isso...

[R] Ok. E foi feita em casa, aqui? Partes num sítio partes noutra?

[I] Foi feita parte em casa, às vezes partes na faculdade. E foi nestes dois sítios. Porque na faculdade, basicamente, nós tínhamos uma pasta nossa lá, e às vezes ficava lá algum tempo, porque também tinha que fazer umas coisas. Gravava aí e levava para casa.

[R] E foi escrito em quê? Em Word?

[I] Word.

[R] Mas em termos de estrutura, depois tinha alguma... às vezes há aquelas coisas... Lembramos de um documento porque era aquele que tinha aquele gráfico colorido, ou que tinha aquela

tabela com dados importantes. Há assim alguma coisa em termos de estrutura que venha à memória?

[I] O que tinha era uma estrutura de capítulos e então sei do documento não por gráficos mas por estruturação de capítulos. Em que parte da tese estava o quê.

[R] E qual era o tema da tese?

[I] Era aplicação do método de CART, que é Classification and Regression Trees à estimação de uma variável. Acho que não vale a pena entrar em mais detalhes...

[R] Não. Já deu para ver que está presente na memória.

[I] Sim, a tese acho que não te esqueces tão cedo.

[R] Em termos de alguma tarefa que tenha sido particularmente complicada na produção do documento? Há assim alguma coisa que venha à memória? Foi preciso recolher alguns dados o que deu muito trabalho? Ou então o inverso. Alguma coisa que até parecia que ia dar muito trabalho e depois até foi mais fácil do que se estava à espera. Há assim alguma...

[I] E documento que está mais presente é mesmo as revisões a nível do documento que tiveram que ser feitas.

[R] Portanto houve várias versões do documento que foram sendo trabalhadas... E essas versões acabaram por ser guardadas ou foi só a última?

[I] eu gravava cada versão. 'Esta é a versão deste mês', gravava. 'Esta é a versão daquele mês', gravava.

[R] E ainda existem?

[I] Existem. Sim, essas coisas...

[R] E outros documentos? É evidente que foi necessário consultar outros documentos para fazer a tese, mas há assim algum... Tipo, se fosse preciso sugerir agora um documento qualquer há algum que venha à memória? Lembro-me DAQUELE artigo ou DAQUELE livro?

[I] Sim, há uns papers de que me lembro. Foram aqueles em que me baseei mais. E se decidir ir ver o documento, sim. Há uns quantos que tenho gravados.

[R] Pronto. Em termos da vida pessoal, há o mestrado, que é um tema importante, é evidente. Mas há mais alguma coisa que venha à memória? Como aqueles exemplos que eu dei. 'Um dia era suposto trabalhar naquilo para depois dar ao orientador mas o carro teve um furo e já não deu tanto tempo.' Houve assim algum percalço, alguma coisa? Ou uma notícia na televisão? Foi na altura em que aconteceu qualquer coisa no país, ou no mundo?

[I] Oh Daniel. Desculpa mas eu não associo também...

[R] Não associares também é informação útil. Eu estou a tentar descobrir exactamente ao que é que as pessoas associam, portanto... Não associar também é bom. Pronto, não sei se há mais alguma coisa que te pareça interessante dentro desta conversa dizer sobre o documento?

[I] Assim à partida...

[R] Passamos adiante, então. Só falta um terceiro tipo de documento, para mais uma vez contar a história sobre ele, mas desta vez é um documento que tenha sido feito por outra pessoa. Por exemplo, um artigo científico, qualquer coisa que nós tenhamos arranjado na Internet, ou que nos tenham mandado por email, que de alguma maneira chegou às nossas mãos, e que por ser interessante, ou por algum outro motivo, nós guardámos em qualquer lado.

[I] Não pode ser um documento de trabalho que te enviam especificamente para ti?

[R] Pode ser. Deste que tenha sido feito por outra pessoa e nos tenha chegado às mãos...

[I] Sim... É que como eu recebi há pouco tempo. Pode ser?

[R] Pode. Há pouco tempo quando?

[I] Ontem. É muito...

[R] E o que é que...

[I] Basicamente era um documento de trabalho, um documento a expor as necessidades deles, os pedidos deles, e eu tenho que transpor isso para o projecto me que estou a trabalhar.

[R] Ok. Portanto, quais eram... Quem foi o autor? Quem foi que o fez?

[I] É um colega de trabalho.

[R] Está bem identificado? Se fosse preciso pedir outra vez o documento sabemos a quem é que se pode...

[I] Eu como até enviei por email e eu guardo sempre os emails, quando guardar na pasta guardo também na área do projecto.

[R] E foi só esse autor ou houve co-autores, que se saibam?

[I] Provavelmente houve co-autores, mas a pessoa que me enviou o documento não fez menção disso.

[R] E é de que tipo?

[I] São dois documentos Word e depois vem informação no próprio mail.

[R] Em termos do conteúdo, da estrutura dele, há alguma coisa que se saliente? Tem um gráfico ou uma tabela grande?

[I] Basicamente são dois documentos e há um deles que tem uma tabela enorme Word com linhas das especificações, que a nível de trabalho aquilo é um bocado assustador de fazer, por isso...

[R] Por isso uma pessoa lembra-se... Muito bem. E outros documentos que possam estar relacionados com esse?

[I] Sim, há vários. Mas queres saber se me lembro de onde é que eles estão?

[R] Não. Se é possível lembrar mesmo de outros documentos. Tipo, foi este documento que mandaram no seguimento daquele outro que me tinham mandado. Coisas assim do género.

[I] Ah, sim, sim. Porque isto é um projecto em várias fases, este documento foi associado a esse projecto que já teve outras fases anteriores. Lembro-me de todos os documentos anteriores.

[R] Ok. Já agora, antes de terminar, houve alguma tarefa... por exemplo, foi difícil tirar o ficheiro do mail? Houve algum problema com o attachment? Houve algum detalhe mais...

[I] Foi fácil.

[R] Pronto. Então agora sim, em termos da vida pessoal ou do mundo em geral lembras-te de alguma coisa que...

[I] Eu acho que divido muito o campo pessoal e o trabalho. Então não... É assim, se tivesse que recordar aquele documento nunca iria recordar nenhuma... algo que aconteceu nesse dia.

[R] Então mais uma vez, a menos que haja alguma coisa que pareça ainda importante dizer?

[I] Acho que não.

## Interview 18

[**Researcher**] O que eu começava por pedir era para identificar um documento concreto que tenha sido feito recentemente. Hoje, ontem. No máximo uma semana, vá lá. E depois de identificado o documento concreto, o ficheiro concreto, dizer-me, no fundo, contar a história desse documento. Dizer tudo o que vem à memória sobre ele. Porque é que foi feito, sobre o que é que tratava?

[**Interviewed**] Pode ser um documento completamente simples ou...

[R] Sim. Pode ser uma coisa simples, desde que seja uma coisa concreta, não é preciso ser nada rebuscado...

[I] Lembro-me que fiz umas propostas internamente. Já não me recordo foi de quê. Tenho que ir ver.

[R] Não saber o que é também é informação.

[I] Foi uma proposta de aquisição de material. Já não me lembro do que é que foi.

[R] Muito bem. E quando é que foi feita?

[I] Não faço a menor ideia. Talvez na segunda feira.

[R] Foi feita aqui, como é evidente?

[I] Foi.

[R] E houve algum co-autor, vá lá. Um revisor, por exemplo. Foi feita uma primeira versão e depois foi lida por outras pessoas para...

[I] Não, não não.

[R] Portanto, não foi trocado por email com ninguém, ou acabou depois por ser enviado...

[I] Foi enviado por email para a direcção e depois para o departamento financeiro e administrativo.

[R] E qual era o tipo de documento?

[I] Aquilo é um memorando interno de pedido... de pedido não, de proposta de aquisição.

[R] E foi feito no Excel, foi feito no Word, ou umas partes...

[I] Umas partes em Excel, outras no Word, e mais uns gráficos de uma aplicação.

[R] Chegou a ter várias versões ou foi só...

[I] Foi directo.

[R] E neste momento está armazenado aonde? Se fizesse falta ir lá...

[I] Está no meu portátil.

[R] Houve alguma tarefa que tenha sido ou particularmente difícil ou particularmente fácil, que até parecia que ia ser pior na produção do documento. Foi preciso, sei lá, consultar muitas coisas, foi preciso fazer muitas contas? Há assim alguma coisa que...

[I] Não. Foi um documento simples. Foi só ir sacar uns gráficos, fazer uns copy-pastes e construir o texto.

[R] E, portanto, depreendo por aquilo que acabou de dizer que foi preciso consultar outros documentos que já existiam?

[I] É uma aplicação que gera gráficos. Foi aquela que acabámos de ver ali há bocado. Uma aplicação de gestão de largura de banda.

[R] Ok. E em termos da vida pessoal... Vida pessoal no sentido de coisas que até podiam estar marcadas na agenda. Por exemplo, 'fiz o documento no dia em que era o aniversário de alguém'. Ou 'foi no dia em que até me atrasei a fazer um documento porque tinha havido um acidente na 2a. circular e demorei muito tempo a cá chegar.'

[I] Alguma coisa que me marcou nesse dia?

[R] Exacto. Ou então mesmo do mundo em geral. 'Vi uma notícia que me chamou a atenção...'

[I] Não faço a menor ideia. Não me recordo de nada. Esta semana é especialmente difícil para mim.

[R] Ok. Não sei se há assim mais alguma coisa que possa parecer interessante?

[I] Não. Pronto. Agora já me estou a recordar. Aquilo deve ter sido... Aquilo foi numa segunda feira ao final do dia. Agora recordo-me por causa dos gráficos. Porque apanhei um Domingo e vi no gráfico que não havia cá ninguém.

[R] Então quanto a este documento estamos conversados. Passávamos a um segundo documento em que íamos fazer a mesma coisa. No fundo contar a história dele, o que é que vem à memória, mas desta vez um documento que já tenha sido feito há dois, três anos. Uma coisa antiga.

[I] Dois anos, então. É o mais antigo. Sim... estou-me a recordar por exemplos de um relatório também na área das comunicações.

[R] E foi feito para quê? Serviu como base para comprar novo equipamento, também? Foi só um relatório que...

[I] Foi um relatório de análise à infra-estrutura de comunicações da [Empresa] e depois veio para alocar a um projecto.

[R] Portanto também foi feito aqui, no local de trabalho?

[I] Exactamente.

[R] Há cerca de dois anos?

[I] (assentimento)

[R] E co-autores, teve, mais uma vez? Ou pessoas que tenham lido para dar algumas ideias? Ou foi feito só por uma pessoa?

[I] (pausa) Houve contributos. Mas pessoas que tenham ajudado à construção do documento propriamente dito não. Houve pessoas com que se falou para...

[R] E trocas de emails, chegou a haver, sobre esse documento, ou sobre a informação que...

[I] Sim.

[R] Se fosse preciso agora ir chatear, entre aspas, as pessoas que ajudaram nessa altura, era fácil saber quem elas foram?

[I] Nem por isso...

[R] E foi feito também no Word, Excel?

[I] Word, Excel, Visio.

[R] Ok. Há assim alguma figura, algum diagrama que ainda esteja presente na memória porque deu muito trabalho a fazer?

[I] Sim, há um sobre as tipologias de rede...

[R] E neste momento está armazenado aonde?

[I] Está no meu portátil.

[R] Também foi preciso consultar outros documentos ou esse próprio documento foi depois usado para fazer outros?

[I] Esse serviu de base para outros.

[R] Pronto, chegámos àquelas perguntas genéricas. Em termos da vida pessoal, ou do mundo em geral, há assim alguma coisa que agora seja fácil lembrar?

[I] No momento em que eu estava a fazer o documento não me recordo. Mas na altura recordo-

me que tinha acabado de entrar para a [Empresa]. Estava a ter o meu primeiro contacto com a [Empresa].

[R] Já agora só para terminar. Houve várias versões do documento também ou foi só uma?

[I] Não, houve várias versões do documento.

[R] E acabaram por ser mantidas? Neste momento ainda existem, ou ficou só a última?

[I] Não, foi ficando sempre a última.

[R] Pronto. Agora, sim, terminámos sobre ele, a menos, mais uma vez, que exista alguma coisa assim que possa...

[I] Não, não me recordo assim de mais nada.

[R] Pronto. Então chegámos ao terceiro documento, que é ligeiramente diferente dos outros porque é um documento que não tenha sido feito por si. Portanto, um artigo que tenha sido necessário ler, ou alguma coisa que alguém enviou por correio electrónico ou que tenha encontrado na Internet e achou interessante e acabou por ficar guardado algures? E uma vez identificado um documento em concreto, descrevê-lo.

[I] Sei lá, há tantos documentos... Vêm-me tantos à memória que não sei... Quer o último ou quer um qualquer?

[R] Um qualquer. Aquele que veio primeiro á memória, por exemplo.

[I] Pode ser um documento sobre a utilização do material de redes (?).

[R] Serve perfeitamente. Bem, o tema já percebemos. E foi feito porquê? Havia alguns problemas e por isso foi feito o documento?

[I] É um documento da Gartner, é um estudo da Gartner, e aquilo tem mais um carácter de evangelização do que...

[R] Ok. Nomes de pessoas concretas que o tenham produzido?

[I] Que o tenham produzido não sei. Foi pela Gartner mas não me recordo agora de nenhuma pessoa em especial.

[R] E quando é que foi encontrado esse documento? Ou quando é que foi lido?

[I] Sensivelmente há um mês.

[R] Ok. E aqui? Foi lido aqui?

[I] Aqui.

[R] E está associado também com outros documentos? Esse foi encontrado porque aparecia como referência noutro? Ou depois foi usado como referência?

[I] não. Foi usado dentro de outro... dentro de uma outra área. Eu estava a consultar um dossier que eram os 'perdicts' de 2003. Eram as previsões deles para 2003. Encontrei isso lá dentro.

[R] Portanto estava no site deles, provavelmente?

[I] Sim. Aquilo é um site reservado mas estava no site deles.

[R] E em termos do conteúdo, da estrutura. Há assim alguma coisa que se destaque? Aquilo tem algumas figuras grandes numa página, assim alguma coisa?

[I] Não. Tem aquele layout standard da Gartner. Ou seja tem um... esqueço-me agora o nome... um... um summary?

[R] Um abstract...

[I] Um abstract! É daqueles documentos... layouts científicos.

[R] Ok. E foi depois enviado a alguém? Por exemplo: 'olha, isto pode-te parecer interessante', e seguiu por mail?

[I] Foi. Foi para muita gente. Seguiu para o [Director] e para os meus colegas daqui. E para algumas pessoas com quem estava no projecto na altura e tem vindo a ser utilizado várias vezes.

[R] E é um pdf, é um doc?

[I] No formato que eu estou a enviar é pdf.

[R] E, por exemplo, quando foi preciso tirá-lo da Web, houve algum problema? Às vezes a ligação está má e vai abaixo? Foi preciso tentar outra vez? Houve assim algum percalço que...

[I] Não.

[R] Pronto. Então chegámos assim outra vez àquelas perguntas mais genéricas do mundo em geral, ou da vida pessoal em particular. Foi no dia de um aniversário? No dia em que o vizinho fez muito barulho de manhã? Esse género de coisas...

[I] (pausa)

[R] Quando estava a ler o documento entrou alguém de repente no gabinete porque tinha acontecido qualquer coisa?

[I] Não me recordo...

[R] Pronto. Então, a menos que haja mais alguma coisa que pareça interessante, terminámos.

[I] Não.

[R] Obrigado.

## Interview 19

[Researcher] O que eu pedia era para começar por pensar num documento concreto, um ficheiro concreto. uma coisa que tenha sido produzida hoje ou nos últimos dias no máximo numa semana, e depois de ter pensado nesse documento concreto, contar a história dele. Dizer tudo o que vem



à memória sobre ele. Porque é que foi feito, quando é que foi, a quem é que foi enviado. Portanto, tudo o que vem à memória sobre esse documento. Portanto, vamos começar por pensar numa coisa concreta que tenha sido feita...

[Interviewed] Por acaso estas semanas o que tenho andado a fazer é relatórios de queries. Não sei se...

[R] Desde que seja um documento, que esteja aí guardado algures...

[I] Sim.

[R] Pronto. Então, pensando num em concreto, o que é que...

[I] Portanto, o que é que eu estive a fazer. Estive a fazer um relatório com uma análise que fiz a uma query que tinha problemas de performance.

[R] E quando é que isso aconteceu? Quando é que foi feito?

[I] A semana passada.

[R] Foi feito aqui, obviamente...

[I] Sim.

[R] E houve alguém que tenha ajudado a fazer, que tenha lido para dar algumas ideias, ou nem por isso?

[I] Penso que foi lido ali pelo [Pessoa 1].

[R] Ok. E foi... Foi trocado por email?

[I] Não. Ele veio aqui ao meu PC e esteve a ver.

[R] E, portanto, não acabou por ser enviado por email a ninguém ou depois de estar pronto...

[I] Não. Depois foi enviado.

[R] Muito bem. E foi feito aonde? É um Word, Excel? Tem misturas?

[I] O resultado foi Word, mas... É assim, eu gerei vários ficheiros, porque aquilo foi a análise de uma query. Eu geri vários ficheiros e depois fui buscar, produzir o documento final com base em bocados dos ficheiros.

[R] Ok. Nessa produção houve alguma tarefa particularmente difícil? Ou então particularmente fácil, que até parecia que ia dar mais trabalho e depois não deu.

[I] Mas em relação ao quê, ao... Quer dizer, não é ao relatório em si, não é... O relatório deu muito trabalho, mas foi a análise que tive que fazer da query, não a construção do relatório.

[R] A construção no sentido mais lato, exactamente.

[I] A análise deu muito trabalho, não é? Porque foi a análise que eu tive que fazer de uma query. Eu peguei numa query que nem sequer era minha, ter que perceber o que é que aquilo minimamente estava a fazer, e depois gerar com a ajuda de outro software, gerar queries que produzissem

os mesmos resultados em melhores tempos.

[R] Foi armazenado aonde, no fim? Agora se fosse preciso ir lá buscá-lo, onde é que ele está?

[I] Agora tinha duas possibilidades. Ou ia ao mail que enviei às pessoas, e a forma que ele ia procurar o mail, para além do subject, ia pensar a que projecto é que pertencia a query, provavelmente, e porque o nome do ficheiro, honestamente, já não me lembro que é que é. Vou à procura do projecto em que se passou a questão, e fazer um find dos mails que enviei... Nesse caso eu sabia as pessoas a que enviei e facilmente ia. Mas ao disco, normalmente eu tenho uma pasta onde ponho os documentos todos do trabalho. E aí eu ia à procura e algum havia de ter um nome parecido ou então fazer um find com o... Eu uso muitas vezes, quando não me recordo do nome do ficheiro, o find no NT, não é? Procurar dentro do ficheiro. Meto algumas palavras chave e...

[R] Portanto mesmo do conteúdo do ficheiro, e não do nome.

[I] Exacto.

[R] Já que falámos do conteúdo do ficheiro, há alguma coisa na estrutura do ficheiro que venha à cabeça? Por exemplo, acabou por ter um gráfico muito colorido, ou tem uma tabela enorme que deu trabalho a formatar? Há assim algum...

[I] Não, não tem nenhum gráfico colorido. O que eu me recordo do ficheiro é que ele tem traces de queries. O que é que isto significa? Eu tenho traces de queries, que me dizem que primeiro foi a esta tabela, depois a esta, esta, esta. Ele tem uma indentação da forma na qual foi executado. Não tem gráficos nenhuns. É texto mas tem... são coisas típicas do nosso trabalho.

[R] Exacto. Que segue um padrão bem definido de...

[I] Exacto.

[R] Para terminar sobre este documento. Sobre a vida pessoal, no sentido de alguma coisa que pudesse estar marcada na agenda. Por exemplo, foi feito no dia do aniversário de alguém, foi feito, por exemplo era para ter sido mais cedo mas havia muito trânsito e foi difícil chegar a horas nesse dia. Há assim algum...

[I] Não... Neste... Se queres que te diga não me recordo assim...

[R] Do mundo em geral, talvez? Alguma notícia que tenha dado nesse dia...

[I] Não, nesse dia por acaso com esse documento houve um percalço. O mail da [Empresa] não estava a funcionar, o servidor de mail não estava a funcionar, e eu vi-me aflita para mandar o mail. E eu queria mandar o mail naquela hora exacta e... Portanto não foi coisas do dia extra-trabalho. Foi aqui. O nosso servidor de mail não funcionava e eu mandei-o através do nosso email externo. Por acaso esse documento específico foi assim.

[R] Ok. Então quanto a este documento, a menos que haja mais alguma coisa que venha à memória, estamos conversados. Passamos a um segundo documento em que vamos fazer exactamente a mesma coisa, dizer tudo o que vem à memória sobre ele, mas desta vez é um documento antigo. Uma coisa que tenha sido feita há dois anos, três anos.

[I] Podemos falar do caso do meu Currículo Vitae que eu não faço ideia de onde é que o meti (risos).

[R] Por exemplo. Portanto presumo que foi escrito já há algum tempo, não é?

[I] Pois já foi... Já deve ter sido prái há uns quatro anos, a última vez que o refiz...

[R] E foi aonde? Onde é que foi feito nessa altura?

[I] No computador de casa. Nessa altura foi no computador pessoal de casa.

[R] E eu ia perguntar onde é que ele está armazenado agora mas como já percebi... (risos)

[I] É assim. Há de estar numa Zip drive... numa Zip em qualquer lado. E não sei qual delas é que é.

[R] Teve várias versões?

[I] Teve.

[R] Às vezes acontece encontrar a versão errada...

[I] Exacto, já me aconteceu. Já encontrei uma das versões, aquela de quando uma pessoa sai da faculdade, e não consigo encontrar a última.

[R] Houve alguém que tenha ajudado a fazer o documento?

[I] Ajudado... sim. Com ideias, sim.

[R] E foi trocado por email com essa pessoa nessa altura ou nem por isso?

[I] Não me recordo.

[R] E está escrito em Word, suponho.

[I] Sim.

[R] Em termos de estrutura há assim alguma coisa que se destaque?

[I] É a estrutura de um currículo normal.

[R] Não tem nenhuma tabela que tenha sido formatada, nenhuma imagem?

[I] Não, não.

[R] Foi preciso consultar outros documentos para fazer esse? Ou já foi usado como base de outros documentos?

[I] Foi usado como base de outros, não é?

[R] Houve, mais alguma vez, algum percalço na sua construção? Por exemplo foi preciso imprimir e a impressora encravou?

[I] Não me recordo...

[R] E da vida pessoal ou do mundo em geral? 'A última vez que lhe mexi foi naquele dia em que tive ir a correr a...'

[I] Não, não não. Não me recordo.

[R] Pronto. Mais uma vez, não sei se há assim mais alguma coisa que venha à memória?

[I] Não. Desse aí não, infelizmente.

[R] Ok. Podemos passar adiante, então. Chegámos ao terceiro e último documento que é um documento que vai ser diferente porque desta vez vai ser um documento que não foi o próprio a fazer. Portanto, um artigo qualquer que tenha sido lido, uma página na Web, um artigo daqueles publicados online, por exemplo. Qualquer coisa que tenha sido interessante para ler e depois para guardar, mas que tenha sido outra pessoa a fazer.

[I] Sim... Pode ser coisas... páginas de Web?

[R] Pode.

[I] Porque aqui há umas semanas recebi informação de um software que eu achei bastante interessante. Veio por mail. Consultei a informação que vinha no mail e depois fui consultar a página. Se me perguntares agora qual é o nome do software não sei, não me recordo. Teria que ir à procura do mail que ele me enviou.

[R] O sítio onde está a página propriamente dita também...

[I] Teria que ir a partir do mail, clicar lá para ir onde ele me enviar.

[R] E quem é que tinha feito a página? No fundo suponho, quem é que será a empresa desse software? Ainda te é possível lembrares disso?

[I] Eu não me recordo do nome daquilo... Mas tenho ideia de um gráfico que aquilo tinha. Tinha um gráfico a explicar como é que funcionava o software tudo. Portanto tenho a ideia de como é que o software funciona, da arquitectura daquilo. Agora neste momento para me lembrar do nome da empresa não me estou a conseguir lembrar. Lembro-me das pessoas que estavam envolvidas. Foi o [Pessoa 2] que me enviou o mail e aquele mail ia para mais pessoas para debater o assunto. Consigo-me lembrar disso mais ou menos. Iria conseguir chegar lá ao software. Mas não me recordo do nome.

[R] E era um software para quê?

[I] Para DataWarehousing.

[R] Muito bem. Isto está associado a outros documentos? Por exemplo, foi um de vários que foi consultado? Com várias possibilidades? Foi usado como referência para produzir um relatório, com vários produtos?

[I] É assim... Aquele documento, a [Pessoa 2] enviou-me para eu dizer o que é que achava sobre aquele software. Aquilo que eu fiz, fiz uma análise do que vi, mandei-lhe um mail de volta a dizer o que é que achava. O software era bastante interessante, e dois ou três pontos de vista. Não fiz... fiz um documento no mail.

[R] Exacto. Está escrito no corpo do mail propriamente dito. Chegou a ser guardada alguma coisa tirada desse site, ou foi só consultado...

[I] Do próprio site suponho que quando fiz o relatório de volta capturei uma das imagens que estava lá que é com a arquitectura.

[R] Pronto. Chegámos aquelas perguntas mais genéricas da vida pessoal, do mundo em geral. Foi na altura em que um vizinho fez muito barulho de manhã?

[I] Não me recordo. Eu percebo isso mas este especificamente não me recordo assim de nada. Escolhi uns documentos muito maus (risos)

[R] Não. São documentos típicos, portanto... Não sei se há mais alguma coisa que pareça importante dizer senão estou contente. Estamos conversados.

## Interview 20

[Researcher] Vou começar por pedir para pensares num documento concreto, uma coisa concreta que tenha sido feita hoje ou no máximo numa semana. Um ficheiro qualquer recente. E depois de ter sido identificado esse ficheiro, descrever o que vem à memória sobre ele. Porque é que foi feito, tudo o que vem à memória sobre o ficheiro.

[Interviewed] Ok.

[R] Então?

[I] Mas pronto, é assim... É um bocadinho complicado porque eu faço ficheiros que são muito importantes. Sempre dou nomes que são muito significativos. Por exemplo, se trabalho num artigo sobre a teoria do caos, controlo do caos, logo crio uma coisa e tudo está lá dentro. E está tudo mesmo em termos de gráficos que saem, diagramas, tudo está assim com nomes muito sugestivos. Porque eu em princípio assim, com as datas sou péssima. Péssima. Então aqueles que não têm importância... Ok. Já está. Não pensamos em coisas muito importantes, não é? Ou é melhor pensar em...

[R] Um documento qualquer que venha à cabeça. Não é preciso pensar em nada específico. O importante nesta fase é que seja um documento concreto que tenha sido feito nos últimos dias. Uma coisa recente.

[I] Ok. Já está.

[R] E então, qual foi o que veio à memória, então?

[I] Então foi uma carta que fiz para a minha naturalização.

[R] Ok. E quando é que isso foi?

[I] Isso foi a semana passada... Acho que foi terça-feira. Lembro-me que foi terça-feira porque foi no dia em que o [Filho] fez uma crise muito feia. E pronto, exactamente por isso é que fiz aquela. Para não adormecer, sabes?

[R] E foi feito em casa?

[I] E casa. Foi feito em casa.

[R] Onde é que está guardado, agora? No computador de casa?

[I] Sim, no computador de casa.

[R] E alguém ajudou a... Leu? Ajudou a fazer? Fez parte?

[I] Não. Odeio quando alguém está nas minhas costas. Não gosto...

[R] E foi trocado por email eventualmente com alguém? Ou foi enviado por email no fim?

[I] Não. Depois foi imprimir e foi mesmo...

[R] Ok. Foi escrito no Word, suponho?

[I] Não, eu escrevo... Eu não escrevo no Word porque eu não gosto muito... Eu escrevo em LaTeX.

(...)

[R] Foi preciso consultar outros documentos para fazer esse?

[I] Não que eu já tinha assim... Já conheço a fórmula muito bem.

[R] E houve assim alguma tarefa na construção do documento que foi mais difícil do que o que se estava à espera? Ou então mais fácil do que se estava à espera? Ou correu tudo bem?

[I] Sim, foi tudo normal.

[R] E do mundo em geral há assim alguma coisa que venha à memória? Foi o dia em que começou a guerra. Foi o dia em que...

[I] Sim, foi o dia em que o [Filho] fez a crise.

[R] Sim, exacto, mas para além disso? Uma coisa nas notícias? Do mundo em geral?

[I] Não. Não porque aquela coisa perturba tanto que já não me lembro de mais nada. Fico só com essa coisa má na cabeça.

[R] Ok. Em termos do conteúdo, da estrutura. Há assim alguma coisa que se saliente? Por exemplo, às vezes acontece ter uma tabela muito grande que deu trabalho a formatar, ter uma imagem colorida, ter um gráfico. Há assim algum...

[I] Não... O pá, por acaso pensei numa coisa muito simples. Não... Isso foi sem problemas.

[R] Quanto a este documento estamos conversados. Agora fazíamos a mesma coisa, descrever um documento, mas desta vez um documento que tenha sido feito há muito tempo. Dois anos, três anos. Uma coisa já antiga.

[I] Ok.

[R] Então qual é o documento?

[I] Um artigo... Vamos dizer... Não é bem um artigo. Foi a apresentação de uma conferência. Foi a primeira conferência aqui em Portugal.

[R] E sobre o que é que era?

[I] Análise de dados.

[R] E houve algum co-autor?

[I] Sim.

[R] E se fosse preciso ir chateá-los agora ainda te lembrás de quem eles foram?

[I] Claro.

[R] Foi feito aonde?

[I] Principalmente foi no computador de casa e no computador da casa do co-autor.

[R] E chegou a haver trocas...

[I] Por mail. Exactamente, tudo isso.

[R] E foi escrito... Já nessa altura foi em LaTeX ou foi em Word?

[I] Naquele caso foi no Word porque foi especificado ser escrito em Word. Foi horrível.

[R] E outros documentos que tenha sido preciso consultar para fazer esse? Há assim algum concreto que venha à memória?

[I] Sim. Foi um livro horrível que foi necessário mesmo consultar, porque o homem era o presidente mesmo da conferência... E tivemos grandes dificuldades em ler o livro (risos).

[R] E quando é que foi a conferência?

[I] A conferência foi em Fevereiro de 2001.

[R] E agora onde é que está guardado...

[I] O documento? Tenho uma pasta que se chama arquivo.

[R] Teve várias versões ou foi só sendo trabalhado...

[I] Não, tem várias versões. Muitas são apagadas, outras sobreviveram. Outras são coisas de que me esqueci, ou gráficos. Às vezes abro e 'o que é isso?'. Um gráfico, vai-se embora.

[R] Já que falámos de gráficos, em termos de conteúdo, de estrutura, há assim algum...

[I] Há gráficos, há quadros, há tabelas, há tudo.

[R] Mas não há assim nenhum que tenha sido muito grande ou muito trabalhoso?

[I] Foi, foi. Porque foi mesmo análise de dados. Então são aqueles gráficos que tenho que interpretar mesmo... E depois não saíram muito bem e no quadrante quatro saiu no quadrante três. E fizemos umas transições pequeninas, mas afinal saiu porque percebemos que nos esquecemos de ponderar não sei o quê e pronto. Aquilo foi...

[R] Ok. Da vida pessoal ou do mundo em geral? Naquela altura estava a acontecer ou aconteceu

alguma coisa? Às vezes aquelas coisas do tipo: no dia em que se ia entregar aquilo a impressora encravou. Ou foi no dia em que havia muito trânsito e quase que chegava atrasada à entrega? O dia perto de um aniversário.

[I] Não. Nada disso. Mas lembro-me muito bem porque naquele dia... bem, não é bem no dia em que entreguei, mas no dia em que ia à conferência finalmente o documento foi naquele dia quando a [Chefe] me disse que fico aqui. Fico aqui a trabalhar.

[R] Ok. Pronto. Quanto a este também estamos conversados, a menos que haja assim mais alguma coisa que venha à memória e que seja interessante falar?

[I] Não.

[R] Então podemos passar para o último documento, que normalmente é o mais fácil de descrever. Um documento que tenha sido recebido ou encontrado, que tenha sido feito por outra pessoa. Por exemplo, um artigo que tenha feito falta ler, uma coisa que nós tenhamos arranjado com interesse suficiente para se guardar. Mais uma vez pensar num documento concreto e depois...

[I] E também importa o tempo? Um mais antigo?

[R] Não, não importa.

[I] (pausa) Sim...

[R] Então, qual foi o documento?

[I] Isso foi... Não é bem um artigo, um semi-artigo que apareceu há uns anos atrás e tivemos que arranjar. Sobre... é mesmo sobre o controlo de caos na bolsa de Nova Iorque. Estava interessada nele porque um dos autores é uma pessoa genial, um grande cérebro, e que estou a fazer tudo o possível para conhecer pessoalmente. Outra coisa... Está no computador de casa. Foi enviado por mail. Alguém encontrou para mim, porque eu não conseguia encontrar...

[R] Está em PDF, em Word?

[I] Está em PDF.

[R] E teve vários autores ou só esse?

[I] Não, só um autor. Não é bem um artigo... é mais ou menos uma receita para fazer controlo que caos.

[R] Há assim outro documento que venha à memória? Esse foi usado como referência para qualquer coisa, ou estava referido...

[I] Sim. Por acaso estava referido noutra artigo. Foi muito difícil para mim encontrá-lo porque praticamente não está publicado. Funciona, vamos dizer que é assim: a bolsa de Nova Iorque já há três anos que contratou um grande grupo de físicos e matemáticos para trabalhar em teoria do caos. Eles fazem previsões das cotações da bolsa e aplicam a teoria do caos para resolver... É uma previsão a curto prazo mas dá melhores resultados do que as técnicas clássicas. Então o senhor faz com essa receita de controlo do caos, naquele curto prazo de controlo, ele prolonga o prazo. Transforma um bocadinho o curto prazo num médio prazo. Não é mesmo um médio prazo. Tem grandes aplicações. Em princípio eu estava a pensar fazer uma coisa parecida, só que é muito



difícil. Ainda estou assim um bocadinho super perdida.

[R] Foi lido aonde, em casa?

[I] Sim.

[R] E pronto. Já percebi que deu algum trabalho porque não se conseguia encontrar, não é?

[I] Exactamente. E também para perceber porque é muita matemática. Computação. São muitas coisas que... Acho que são precisos seis meses para perceber (risos).

[R] E da vida pessoal, mais uma vez, ou do mundo em geral...

[I] O dia em que chegou foi exactamente o dia de Natal.

[R] Esse é difícil esquecer...

[I] Pois é. Foi tipo prenda de Natal. Aquilo ficou com esse título. Está lá no arquivo do caos.

[R] Chama-se mesmo 'prenda de Natal'.

[I] Exactamente. Prenda de Natal.

[R] Muito bem. Em termos do conteúdo, mais uma vez, há alguma coisa... Um gráfico numa página que foi difícil de esquecer? Muitas fórmulas? Há assim alguma coisa?

[I] Sim, sim. Muita matemática. Tem muita matemática.

[R] Ok. Pronto. Não sei se há mais alguma coisa que venha à memória sobre ele?

[I] Sobre este último? Fiquei super assustada com a parte de programação.

[R] Ok.

[I] É tudo?

[R] É tudo. Obrigado.

## Interview 21

[Researcher] Vamos começar. Para descobrir quais são os tais elementos o que eu lhe pedia era para pensar num documento em concreto, que eventualmente será um só ficheiro no computador. Um documento que tenha sido trabalhado, criado hoje, nos últimos dias. Uma coisa recente. Uma semana, quanto muito. E depois de identificado um documento concreto, para ajudar a focar as histórias que vamos aqui contar, era justamente dizer... no fundo era descrever esse documento. Dizer tudo o que vem à memória sobre o documento, não só em termos de interacção com o computador mas eventualmente outras informações do mundo em geral que possam estar associadas. Quando o documento estava a ser criado houve um determinado percalço. Portanto, tudo o que venha assim à memória relacionado com o documento.

[Interviewed] Vou falar em termos de empresa. Depois posso fazer uma alusão à minha localização

peçoal. Em termos de empresa o que acontece é que está estruturado por departamentos, e esse é o primeiro factor da documentação. O departamento de database, de financeira, logística, o SAC, apoio ao cliente. Existe cobranças e por aí além. Marketing. Esse é o primeiro factor na utilização do documento. Associado a um departamento concreto. Em termos pessoais eu procuro um assunto, ou seja, que está relacionado com o documento...

[R] Se calhar o melhor é mesmo pensar num documento concreto para ajudar a...

[I] Penso num documento concreto e vou buscar essencialmente um documento que fizemos inicialmente que tem as competências de cada um e de cada departamento. As nossas funções estão definidas e logo se o documento tem a ver com as cartas de renovação vou pensar: quem é que emite as cartas de renovação? É o departamento de logística que faz tudo o que for fulfillment das acções: inscrição, essas coisas do género. Daí vou buscar a uma pasta de logística, (...) está em 2001, logo é lá que eu vou encontrar. E é mais ou menos pelo nome do documento ou pelo nome da tarefa a que está o documento associado. Depois há de haver documentos parecidos mas essencialmente há de estar dentro de um mapa. Nós controlamos através de mapas e há de estar... vou começar por procurar o mapa. Dentro desse mapas, dentro das renovações, dentro das assinaturas, há de estar o...

[R] Como a hierarquia até está bem estruturada, até costuma ser fácil de encontrar...

[I] Exactamente. Talvez o ponto de partida seja precisamente esse. Ter uma hierarquia e uma definição de funções bem estruturada que permite logo enquadrar e depois ir mais facilmente encontrar os documentos.

[R] Exacto. Mas será que vem assim à memória mais alguma informação... Pronto, se pensarmos num ficheiro concreto, para além da hierarquia, com que já vi que é fácil descobri-lo. Mas...

[I] Há outra questão que é... Cada pessoa só consegue, e eu consigo, mesmo dentro das actividades da empresa, documentos quer pessoais quer do âmbito da minha actividade, eu consigo lembrar-me perfeitamente de onde é que eles estão porque só me cinjo àquela informação e documentos. É a minha área. Por exemplo, se eu quiser procurar alguma coisa de que vou precisar para trabalhar, eu vou ter com uma pessoa dessa área e ela sabe dizer onde é que está. Se eu tentar fixar todos os documentos de todas as áreas eu não consigo.

[R] Só aqueles que foram feitos...

[I] ... dentro da minha estrutura da minha área. Se eu precisar de uma coisa do database eu não sei localizá-la. Mesmo que já tenha usado eu não sei localizá-la. Portanto vou ter com alguém do database e ele vai-me dizer se está nas assinaturas do ano 2000, 2003. Se está na pasta X ou na pasta Y. É o...

[R] E acontece, por exemplo, haver casos de documentos que foram trabalhados no local de trabalho e depois também em casa, ou...

[I] É muito frequente também a nível das chefias trabalharmos em casa e depois colocarmos na rede, na rede da empresa. E depois o que fazemos por norma é cada vez que colocamos um documento novo na rede comunicar à estrutura directamente envolvida a localização dessa pasta, para poder aceder e verificar.

[R] E em relação a trocas de documentos por correio electrónico enquanto está a ser produzido,

para outra pessoa ler. Isso tem lugar ou normalmente é produzido só por uma pessoa? Como é que isso se costuma processar?

[I] Existem mapas e reports que... Bem, em termos de produção, da criação do documento, normalmente só participa uma pessoa. Eu faço para os outros acederem e consultarem e geralmente colocarem informação de controlo, ou simplesmente validarem uma informação qualquer. O que acontece é que realmente são várias as pessoas que interagem com esse documento. E o que acaba por acontecer é que como trabalhamos por ciclos, por fases, ele acaba por se instalar na nossa memória sem ter que recorrer a grandes soluções de localização para conseguir encontrar o documento. Eu, de cabeça, devo ter centenas de pastas porque sou obrigado a verificar a informação e...

[R] Mas mais uma vez se fosse obrigado a pensar num documento concreto e dizer 'neste documento também trabalhou determinada pessoa'. É possível dizer uma pessoa em concreto? Imaginemos que por alguma razão era necessário entrar em contacto com todas as pessoas que ajudaram no documento X. É fácil lembrar quais foram essas pessoas, ou...

[I] A ver se consigo um exemplo. A nível da logística, eu tenho um mapa de controlo das expedições das publicações, para que um dia haja uma hierarquia qualquer que me venha pedir em que dia é que foram para os correios a revista X e qual foi a quantidade expedida, quantas foram as ofertas, qual foi o brinde de capa, se houve atraso, se não houve e tenho um documento em que todos os elementos da área logística diariamente acedem e colocam informação... Neste momento é fácil lembrar quais foram as pessoas que...

[R] Fica registado que colocou essa informação?

[I] Para já não fica registado, até porque os meus conhecimentos de informática não o permitem. Mas à partida...

[R] Cada departamento tem um responsável e portanto terá sido esse...

[I] Exactamente. Neste caso, neste mapa, há três pessoas que interagem com ele diariamente, e só estas três pessoas é que vão. Quando é preciso saber alguma coisa deste mapa, alguma informação, o que acontece é falamos connosco e dizemos: 'olha, gostava de saber em que dia é que a revista X foi para o correio'. Se eu tiver tempo, eu digo. Eu abro o mapa e vejo que foi no dia X. Se não tiver tempo 'está no plano nacional de assinaturas, mapas, etc.'. Digo-lhe o nome do documento e ele vai lá buscá-lo e verifica. Norma geral não fazem alterações porque só nós é que fazemos... É uma ferramenta de trabalho nossa. Eles só consultam.

[R] E quanto ao tipo de documentos é fácil lembrar? Se é um documento Word, Excel?

[I] É. Começa pela denominação que damos a cada documento, e o cuidado que temos em colocá-lo num âmbito em que seja fácil distingui-lo e localizá-lo. Tudo o que tenha a ver com mapas, quer seja renovação, captação, expedição, fica numa pasta mapas, na logística, porque é a logística que faz a expedição e a captação e contabiliza os portes... Logo, é assim que se vai.

[R] Portanto, de acordo com a pasta e a designação dos documentos que lá estão sabe-se logo quem é o responsável por ele, qual é o tipo de documento e qual é a informação que nele estará contida. E assim mais genericamente, em termos de acontecimentos da vida pessoal, no sentido de algo que poderia estar registado até na agenda, ou do mundo mais em geral, há assim alguma

coisa que às vezes fique associada a algum documento? Algum percalço na construção do documento. Algo que quando voltamos a abrir o documento nos lembremos: 'olha, este foi aquele documento que fiz no dia em que tive um furo'.

[I] Uma coisa temporal e associada a outros factores que não propriamente... eu na... pelo menos na minha óptica não vejo, não consigo associar nada nesse sentido. Posso associar o documento à data em que se faz, em que se procede a alterações. Isso acontece realmente associado aos fechos contabilísticos ou ao... ao fecho do ano... o fecho civil... não está necessariamente associado à facturação. Cada vez que transitamos de ano... O Ano Económico! Assim é que é. O ano económico é importante em relação à identificação de certos documentos.

[R] E os documentos costumam ter várias versões que sejam mantidas explicitamente, ou mantêm-se só a última versão?

[I] Eu, em termos de organização pessoal, quando abro um documento ele fica com a designação inicial e depois vou criando pastas com actualizações mas o documento anterior fica sempre, até porque fica um histórico, um background para onde posso recorrer e ir buscar informação que é fundamental. Mas o que faço é renomeá-lo com a data em faço as alterações e...

[R] Portanto a data passa a fazer parte do nome para identificar qual foi a versão.

[I] Exactamente. E fico a saber que o documento em que estive a trabalhar é o que tem a data mais recente. Isto ajuda porque tenho que controlar datas de facturação. Essencialmente está associado a tempo de efectuar tarefas. Vou buscar o documento mais recente...

[R] Portanto os documentos acabam por estar associados a tarefas que é necessário realizar dentro da empresa num dado momento.

[I] (assentimento)

[R] E quanto a documentos que não foram criados por si, criados até fora da empresa. Uma coisa que tenha sido recebida de fora e que tenha sido suficientemente importante para guardar...

[I] Na minha Inbox caem constantemente newsletters de pessoas a meu serviço. E logo à partida tenho que fazer uma filtragem. No caso do material de economato... é frequente haver tipo Office1Store, tipo Stapples Office center. Eles fazem newsletter e o que é que acontece. Eles caem no meu... economato e o que é que eu penso: eu não vou comprar economato, o database não compra economato. Quem compra é o secretariado. Faço logo um forward para o secretariado. Dessa forma restrinjo logo ao departamento a quem diz respeito essa informação e não fico a pensar mais naquilo.

[R] Nunca aconteceu um dia mais tarde fazer falta alguma coisa e 'por acaso de relance vi uma coisa na newsletter que enviei para o economato e que até era uma coisa parecida...'

[I] Exacto. Aí o que acontece? Para já, não apago. Tudo o que eu recebo vai para uma pasta de mails lidos. E depois tudo o que eu envio vou aos sent itens e coloco na pasta da pessoa para quem enviei. Assim eu pensei: a [Pessoa 1], que é colega minha, é responsável pelo secretariado é a pessoa da direcção. E é ela que faz as compras de economato, que é o exemplo que estamos a dar. Eu vou à pasta da [Pessoa 1], procuro as Newsletters e há de lá aparecer.

[R] Portanto, consegue associar mais ou menos à pessoa para quem foi enviado porque aquilo

está mais ou menos organizado. E consegue lembrar-se mais ou menos da altura em que foi?

[I] É! Já me aconteceu muitas vezes fazer a pesquisa por data. Ter uma ideia que foi no mês X, entre o dia tal e tal e procurar o assunto nesses dias. Também é possível fazê-lo em termos de Outlook. É uma questão de quando a informação 'cai' distribuí-la logo, arrumar, e associar logo ao departamento a que poderá ter interesse essa informação, ou que vai gerir essa informação.

[R] Acaba no fundo por ser possível lembrar do documento pelo destino que lhe foi dado, pelo objectivo que ele tinha.

[I] É. O subject e a dimensão do documento são fundamentais para a detecção e para a localização. Estou-me a lembrar ainda hoje eu recebi um documento que tinha a ver com a criatividade. Nós temos a criatividade em outsourcing. Temos elementos do interior mas temos uma empresa externa que... Eu não sou criativo. Eu não sei fazer criatividade. Eu faço a parte final dos projectos, o fulfillment. E o que é que eu fiz logo: o reencaminhamento para a secção comercial e de marketing. São as pessoas que vão aproveitar aquela informação, ou para fazer contactos ou para pedir uma informação, ou para vender um serviço. E quem é que faz isso? É a secção de marketing que é chefiada pela pessoa X. E a seguir vou aos Sent Itens, pego na informação e coloco na pasta dessa pessoa. E assim se algum dia for preciso... 'mandaste-me um mail mas eu perdi!' vou à pasta e pronto.

[R] Não sei se há assim mais alguma coisa que ache interessante dizer?

[I] Eu não sou muito organizado, não sou das pessoas mais organizadas. Por uma questão de tempo. Eu não consigo organizar-me muito bem. Mas para mim é fundamental a questão da organização da informação. Associá-la a uma área de trabalho e com base nisso localizá-la. E eu posso localizá-la pela data e essencialmente pela nomeação que eu dei ao ficheiro. É fundamental a nomeação. E sobretudo o âmbito. A família de assuntos em que ele se irá enquadrar é fundamental. E é com isso que eu trabalho normalmente. Eu não tenho cabeça para fixar tudo mas quando cai um assunto qualquer da área de venda de, por exemplo, listas de mail ou uma base de dados dum assunto qualquer eu encaminho logo, sei logo que é database...

[R] Penso que está... Obrigado.

## Interview 22

[Researcher] Então, a primeira coisa que eu pedia era para pensar num documento concreto, uma coisa concreta que tenha sido feita hoje ou nos últimos dias. Recente. No máximo uma semana, e depois de ter pensado mesmo num documento concreto, num ficheiro mesmo concreto, dizer tudo o que vem... No fundo contar a tal historiazinha. Dizer tudo o que vem à memória sobre o documento. Do que é que ele trata? Para que é que foi feito? Tudo isso.

[Interviewed] Há um projecto em que eu estou envolvido. É um projecto internacional chamado MetaWare, de tratamento de informação estatística. Existem uma série de documentos produzidos alguns por mim, outros por outros elementos da equipa. É uma equipa internacional.

[R] Vamos pensar num...

[I] Qualquer um é um exemplo prático da mesma coisa. Existe uma referência por mails, que era uma estrutura no Outlook em particular que está dividido em actualidade, se é mais recente, se é mais antigo. Se é antes ou depois de determinados marcos. Existe uma referência documental, dentro de uma estrutura de folders, e existe uma descrição sucinta noutros documentos. Ou seja, temos três entradas para a mesma coisa. Qualquer uma delas pode... quando eu preciso de encontrar um documento, não é sempre o caso que me lembre duma dessas três possibilidades, em vez das outras. Hoje sou capaz de relacionar um documento com um mail que recebi depois de um determinado marco, vou buscar a informação de uma forma, e amanhã sou capaz de me lembrar do mesmo documento por um tema qualquer que estava lá dentro e sou capaz de e sou capaz de começar por outro lado.

[R] Se pensarmos num concreto é melhor porque assim podemos ver exactamente o que é que, nesse caso em concreto foi usado.

[I] Por exemplo, um documento produzido por mim. Existe referência nas minhas designações, que falam sobre o conteúdo do documento, a altura em que o fiz, o sítio em que ele está e o tema em que se insere. Existe referências nos mails que enviei a avisar todos os outros, que estão depois de um determinado marco e tem um determinado nível de actualidade, isto dentro do Outlook. Existe o ficheiro propriamente dito que está dentro de directórios que são os meus, referenciados a cada altura.

[R] Portanto, já vi que o tempo é uma coisa fulcral para saber o que é que o documento tem. Então e, por exemplo, se tentarmos pensar num documento concreto, co-autores? Por exemplo, é fácil lembrar quem é que ajudou, ou se ninguém ajudou, mas quem é que ajudou a fazer um determinado documento?

[I] Isso aí é mais difícil. É mais difícil porque a referência aos autores neste caso aparece em trabalho conjunto com que alguém concordou e essa concordância existe noutro documento algures. Essa parte aí já é mais difícil.

[R] Mas assim de memória é possível lembrar-se de mais alguém?

[I] De memória é possível, porque não existem muitas combinações. Se não fosse, já teria tomado as devidas providências para que fosse mais fácil.

[R] Portanto daquilo que eu percebi, normalmente um documento acaba por estar relacionado com outros, não é? Há sítios em que ele está descrito...

[I] Existe um plano de trabalho e existem divisões do trabalho e deliverables a promover, e dentro dos deliverables as tarefas de cada um. E esse é o trabalho interessante, como eu reparei.

[R] E coisas da vida pessoal, ou... Da vida pessoal no sentido daquelas coisas que até poderiam estar marcadas na agenda. Calhar ao pé do aniversário de alguém, ou ter sido no dia em que, sei lá, ao vir para cá havia um engarrafamento muito grande e por isso o documento foi feito mais à pressa porque não houve tanto tempo. Assim qualquer percalço...

[I] Existem três coisas. Existe um documento que abro como startup, que é o BDI (?) onde coloco tudo e mais alguma coisa de que me lembro. E de quando em vez vou fazer uma limpeza e uma reclassificação. Existem aqui umas coisas que são próprias. Que estão em folders particulares. Que são próprias a dois níveis: ou são pessoais, ou são estudos. Não são directamente relaciona-

dos com trabalho. Nesta altura podem vir a ser mais tarde. E existem novamente os Outlooks e as agendas. Portanto, coisas pessoais não me lembro assim de nenhuma.

[R] Então e mundo em geral? Às vezes lembrar: aquele documento fiz no dia em que começou a guerra. Ou no dia em que vi nas notícias que houve um acidente muito grande na 2a. circular. Assim uma coisa... Normalmente essa informação vem assim à memória ou...

[I] Não, não.

[R] Ou às vezes algum percalço mesmo tipo... Estava eu a fazer o documento e entrou alguém no gabinete a pedir qualquer coisa...

[I] Não porque isso é muito usual que seja interrompido durante a execução de um documento.

[R] Portanto não se verifica nada em particular porque acontece com todos, praticamente... Pronto. Então quanto a este estamos conversados. Agora pedia para fazer uma coisa semelhante, pensar num documento e depois dizer o que é que costuma vir à memória sobre ele, mas desta vez pensando num documento que tenha sido feito já há dois anos, três anos. Uma coisa já consideravelmente antiga. Mais uma vez, é mais fácil, talvez, pensar num documento concreto, num ficheiro concreto para descrever. Como este é mais antigo...

[I] Aqueles documentos que considero que são importantes acabo por tratá-los como tratava quando era estudante: fazia resumos da matéria e resumos dos resumos. E isto acabava por estudar os resumos dos resumos, porque o resto era suposto estar sabido. Nos documento importantes ele acaba por cair nos resumos dos resumos e...

[R] Ok.

[I] E têm uma certa referenciação. Naqueles que são menos importantes, é o caos. Estão mais ou menos no arquivo morto e não se previu a reutilização dos ditos.

[R] Exactamente. Mas se calhar, às vezes, sobre um documento em concreto, apesar de nós já não sabermos exactamente onde é que ele está, lembramo-nos: 'olha, foi o documento que fiz sobre não sei o quê', ou...

[I] Em geral, cai na pesquisa textual, no trabalho de detective de procurar palavras.

[R] Portanto, coisas que estejam no conteúdo do documento. E em relação à data em que foi feito? Foi o documento que fiz há 6 meses, há um ano... Foi o documento que fiz...

[I] Aprendi a não confiar nisso, porque a referenciação de data nos computadores é um bocadinho pobre. Quando eu mando de um lado para o outro às vezes a referenciação que fica no documento não tem nada a ver em termos de data com a execução do dito... ou bem que a data está expressa com palavras dentro do próprio documento, e aí é usável, ou se e é uma referenciação à data do dito na... referenciada à máquina ou ao sistema operativo, não é de confiar.

[R] Ok e o lugar em que ele foi feito. Se foi feito dentro do escritório, em casa, noutra sítio. Durante uma reunião... Isso é normalmente algo que é fácil lembrar ou nem por isso?

[I] Nem por isso, porque é muito... Eu tenho o hábito de levar o trabalho para casa e trazer o trabalho de casa. Um documento, há documentos que são feitos em casa, há documentos que são feitos no trabalho, e há documentos que são feitos nos dois lados. Passado bastante tempo não há

noção de onde é que ele foi feito, foi feito aonde.

[R] Ok. E o tema do documento? Isso talvez seja mais fácil de recordar. Fiz o documento sobre ESTE tema. E para além do tema, a razão pelo qual foi feito. Será que também vem à memória? A razão no sentido de 'fiz porque precisava de apresentar numa reunião' ou 'porque precisava de transmitir a outras pessoas'. A razão de ser...

[I] É um dos elementos da associação...

[R] E co-autores ou revisores? Há alguém que tenha lido para completar informação que faltasse, por exemplo, ou um daqueles documentos em que metade faz uma pessoa e outra metade faz outra. Isso também é possível lembrar isso?

[I] Sim. Também é possível lembrar isso.

[R] Ok. E, às vezes, os documentos dentro do seu conteúdo têm algum elemento que se destaca. Por exemplo, foi aquele documento que tinha aquele gráfico colorido. Ou foi o documento que deu trabalho porque tive que formatar em duas colunas e havia uma figura que não cabia. Há assim alguma coisa desse género de informação?

[I] Eu julgo que sobretudo quando há imagens de visão de helicóptero. Porque essa imagem acaba por funcionar como um mapa das estradas dentro do próprio documento. E muitas das vezes quando eu estou à procura de determinado documento ou de determinado conteúdo dentro do documento, uma das primeiras ideias que surgem na cabeça é onde é que está o mapa das estradas disto. Se é uma determinada imagem nós sabemos que aquele documento. Pode até nem estar relacionado com uma imagem do próprio documento. Essa que se vai buscar para fazer uma aproximação sucessiva.

[R] Então só falta mais uma vez aquela pergunta de coisas mais fora do documento em si. Da vida pessoal ou do mundo há alguma coisa que associe...

[I] Algumas colectâneas de URLs e a explicação sobre os ditos. Acaba por cair mais... uns mais que outros, mas acabam por cair dentro das colecções.

[R] Ok. Então, se calhar, podíamos passar a diante e só falta um terceiro tipo de documento, e estes normalmente são mais fáceis de descrever, porque desta vez é um documento que tenha sido feito por outra pessoa. Um artigo, uma carta que tenha sido recebida, uma coisa assim. Como não fomos nós que fizemos provavelmente lembramo-nos de coisas diferentes. Pensando num documento qualquer que tenha sido recebido. Ou encontrado na Internet, ou recebido por email, o que é que normalmente é mais fácil de lembrar sobre um documento desse género?

[I] A forma de o ir buscar. Se vier por email é fácil, foi a pessoa que o enviou. A primeira pesquisa é por pessoa. Só quando o documento tem uma abrangência mais lata é que ele sai da referência do email e passa para outro sítio e começa a ter outras referências.

[R] E em termos da altura em que ele foi enviado ou descoberto? 'Foi um documento que lembro que alguém me mandou há seis meses'. Isso normalmente ajuda?

[I] Acaba por não ajudar muito. Embora nós tenhamos a noção de que foi enviado num passado mais ou menos longínquo...



[R] ...exactamente quando já se perdeu. Foi há muito tempo, mas exactamente quando...

[I] Muito embora exista a ideia de que foi há três meses atrás a experiência, pelo menos a minha, diz que maior parte das vezes engano-me.

[R] Mas há outro género de informação como por exemplo do que é que o documento tratava, ou por exemplo, foi um documento que eu li ou que alguém me mandou porque era sobre um determinado projecto, e era interessante para mim por uma determinada razão. Isso é mais fácil lembrar? O tema e a razão de ser do documento em si?

[I] Muitas vezes o que acontece é que documentos antigos, onde a informação, as nossas referências não estão bem organizadas, nós acabamos por aceder a elas em tempos diferentes. Sou capaz de dizer 'isto foi há três meses atrás' e fazer uma primeira pesquisa e não encontrar nada. Depois vou fazer outra coisa qualquer e procuro o tema. E vou associando, construindo associações de ideias até chegar a um ponto, a uma ponta da meada e começar a tirar os nós todos para trás.

[R] E o que é que será que está nessa meada? Já falámos do tempo, do tema, das pessoas que o enviaram, portanto. O nome do autor? Há aqueles artigos que são feitos por determinadas pessoas. É uma coisa que é fácil de lembrar, quem fez o artigo?

[I] Há alguma facilidade nisso.

[R] E o local em que ele está guardado ou estava guardado quando foi encontrado? Por exemplo, se foi tirado de um site saber em que site é que estava. Imaginemos que se apagava o ficheiro por engano e tinha que se ir lá buscar outra vez. É fácil encontrar o sítio onde ele estava?

[I] Não é muito fácil, e acabamos por cair em motores de pesquisa. E o facto de não ser fácil é comprovado exactamente por isso, pela necessidade de ir através de motores de pesquisa.

[R] Exactamente. E em relação ao tipo dos documentos? Se era um documento Word, Excel. Isso é fácil lembrar ou é uma coisa que já não fica muito...

[I] Não, não. Porque colocamos Words dentro de 'Exceis' e 'Exceis' dentro de Words. E outras combinações.

[R] Falta só saber se em termos de outros documentos, por exemplo, às vezes nós vamos buscar qualquer coisa porque depois vai servir de referência para alguma coisa que nós estamos a fazer, ou assim. É fácil associar esse género de informação? 'Aquele documento que eu fui buscar porque precisava não sei aonde'. Costuma ser fácil lembrar assim disso?

[I] É mais um daqueles condimentos da grande sopa de recordações que não são imediatas.

[R] Mas se for preciso lá chegar?

[I] É mais um condimento.

[R] Ok... Então mesmo para terminar, é aquela pergunta que eu faço sempre no fim de cada documento que é a da vida pessoal ou do mundo em geral em relação ao documento. Portanto, dos que nós estamos a falar agora, feitos por terceiros, é fácil lembrar-se de alguma coisa ou é mais difícil do que em relação àqueles que nós fizemos?

[I] Existem documentos que são especiais. Esses são tratados de forma diferente e são fáceis de ir lá buscar. E há outros... Há aqueles que são fáceis e aqueles que são quase impossíveis. É dependente da importância que lhe demos na altura de classificar.

[R] Portanto acaba por estar muito dependente do nosso esforço na altura da classificação. Ok. Penso que é tudo. Não sei se há mais alguma coisa que possa parecer interessante? Senão estou satisfeito...

[I] Muito bem.

[R] Obrigado.