Telling Stories with Dialogue Boxes to Retrieve Documents

Daniel Gonçalves, Joaquim Jorge

Computer Science Department, Instituto Superior Técnico, Av. Rovisco Pais
1049-001 Lisboa Portugal
djvg@gia.ist.utl.pt, jorgej@acm.org

Abstract. Nowadays, it is common for users to handle large numbers of documents. Organizing and retrieving those documents is extremely difficult using the tools commonly provided for those tasks. The use of document-describing narratives constitutes an alternate, easier way of allowing the users to do so. Narratives can help them remember important information about documents and are a natural way to convey that information to computers. In order to develop this approach, it is necessary to understand what shape do document-describing stories have. To this end we interviewed 20 users and collected 60 stories about documents. Analyzing these stories yielded a thorough characterization of their contents and structure and to extract guidelines on how to deal with them. We then validated those results by creating and evaluating two low-fidelity prototypes for possible story-capturing interfaces. We found that stories told to computers can be very similar to those told to humans, if the interface is properly designed. These results seem to suggest that structured text entry is a promising design for this interface.

1 Introduction

Computers are part of most people's everyday life. It is now common for typical users to have access to several different machines, both at home and in the workplace. Furthermore, many common tasks are now performed with the help of computers, from purchasing goods on the Internet to turning in tax forms. All this has caused most users become swamped in documents.

The usual ways to organize and retrieve documents, usually based on hierarchies, are becoming harder to use as these trends progress. Such hierarchical approaches require users to explicitly classify their documents. However, this is a problem-laden task. Two of the most common difficulties appear when a document seems not to fit any of the existing categories (and might not justify the creation of a new one by itself), or when more than one category seems to apply. The users are aware that their ability to later retrieve the document is strongly dependant on this classification, causing undue cognitive loads. In fact, once stored into the hierarchy, the documents become invisible until found again, the only clue to their whereabouts being the aforementioned classification. Thomas Malone [4] found that many users try to avoid classifying their documents, just storing them in undifferentiated collections ('*piles*') and

resorting to additional clues, such as their location, to find them. More recently, while looking at the usage of email tools [10], Whittaker et al witnessed that some people use those tools not just to read and send email, but overloading them with other functions for which those tools were not designed for, such as managing their agenda or storing their documents. This occurs because email messages are not hierarchically classified. Instead, they have associated to themselves all sorts of specific context-dependant information, such as the sender, date, or subject. Using this information instead of an explicit classification helps the users to find the messages or documents stored therein.

Several approaches have appeared that try to make use of additional information for document organization and retrieving. For some, time is the most relevant factor. It is the case of Lifestreams [2], in which all documents are organized on a temporal stream that can be navigated or filtered. In the Timescape system [9], the desktop displays collections of objects bound to a certain time period, and can be moved to past or future states. These approaches have the disadvantage that, by giving time a special role, they can disregard other potentially useful information.

More general are the approaches based on Gifford's Semantic File Systems [3]. In them, there are no directories or folders per se. Instead, the user faces virtual directories whose contents are the result of queries for specific values of certain properties of the files. More recent approaches that embody this idea are Dourish et al's Placeless Documents [2] and Ricardo Baeza-Yates et al's PACO [1]. These approaches, however, shift the burden from classifying the documents to the need for the user to remember the names and possible values of an arbitrary set of properties.

We argue that narratives can provide a better way to retrieve documents taking advantage of autobiographic information. In fact, humans are natural-born storytellers. We start our life listening to stories told by our parents or grand-parents, and tell them to others on a daily basis. On human-computer interaction research, it is common for storytelling to be investigated as a way for the computer to interact and convey information to a human user. Here, we explore the opposite notion, that using stories to convey information to the computer is also a useful form of interaction. In fact, stories help us to remember related pieces of information. Thus, rather than the effort required to remember unrelated properties, stories will allow the users to naturally and easily remember a wealth of information about their desired documents that can then be used by the computer to find them.

To better understand how interfaces for document retrieval using narratives can be built, we performed a study in which twenty users were asked to tell document-describing stories. Sixty such stories were collected and analyzed, allowing us to understand what story elements are more common and what structure is to be expected in such a story. From this data we were able to extract several guidelines on how narrative-based interfaces for document retrieval can be built. However, those stories were told to a human interviewer. Would stories told in a structured environment, like a computer interface, share the properties we had just found? To answer this question we used the guidelines to create two low-fidelity prototypes of possible story-based interfaces and, again, asked twenty users to tell their stories, but this time using the prototypes. Both resorted to dialogue boxes for the introduction of the several story elements, and one represented the story as a set of graphically separate elements,

while the other was based on structured text entry. For several reasons we chose not to consider an interface allowing unrestrained natural language interaction with a software agent. First, it would be hard to simulate such an interface on a low-fidelity-prototype without falling back to telling the stories to a human. Second, given the wide range of sentences the users can utter, such an interface would have a hard time establishing meaningful dialogues, given the current state of speech recognition. Third, the document-retrieval task should be performed as efficiently and robustly as possible within the current technological constraints. Finally, we want to take advantage of the users' familiarity with existing solutions to the retrieval problem.

We found that stories told to computers can be as rich as those told to humans and share a similar structure. However, for this to happen, the interface must be designed with care.

In the following section, we describe the first part of the study, in which the stories' contents and structure were analyzed. Next, we'll show two possible approaches for capturing document-describing stories and their evaluation. Finally, we'll discuss overall results and possible future work.

2 Analyzing Stories

This part of the study aims at getting answers to two important research questions: 1) what is the expected content of document-describing stories, and 2) what structure is typical of such stories. In order to answer those questions, we conducted a set of interviews in which 20 users were asked to tell stories about their documents Given that we did not know what to expect, we chose a semi-structured interview method.

2.1 Procedure

The subjects could tell their stories in any form they chose, but the interviewer had a set of questions prepared to keep them on track if they started digressing. The interviews were recorded with subject consent. Participants were asked to tell three stories, describing three different document types: Recent Documents, created by the user up to two weeks ago; Old Documents, created by the user over six months ago; and Other Documents, which the users had read but not created. Special care was taken to find a diverse user sample. They were balanced in gender, with professions ranging from computer programmer to social sciences professor. Ages varied from 24 to 56. We also accounted for a wide range of computer expertise from casual users to nerds.

2.2 Analyzing the Interviews

All interviews were transcripted and submitted to a formal Contents Analysis [7,11]. We coded for the elements in Table 1, the only ones found during the analysis phase.

The coding was made by hand rather than automatically, which would require a dictionary of words belonging to the several elements, forcing the researcher to antici-

pate all possible relevant words. This is an impossible task on our study given the open nature of the users' responses. We coded for frequency rather than for occurrence to be able to understand not only what elements occur more frequently, but also if they are repeatedly in a story. Finally, we took note of whether a particular element was *spontaneous* (occurring normally in the course of a story) or *induced* (remembered by the users after some intervention from the interviewer).

Table 1. Story Elements

Time	Place	Co-Authors	Purpose
Author	Subject	Other Docs.	Personal Life
World Events	Doc.Exchanges	Doc. Type	Tasks
Storage	Versions	Contents	Events
Name			

We also conducted a Relational Analysis. A transition between two elements was considered to have occurred when they immediately followed each other in a story. No transition was recorded when the destination element was induced, since in that case no real connection between the elements existed on the user's mind. This analysis allowed us to estimate how the stories are structured.

2.3 Results

The stories usually took five minutes to be told. Their transcripts averaged two to three plain text pages, although some users told longer stories. A typical story might start like this translated excerpt from a real interview:

Interviewer: So, now that you have thought of a document, please tell me its story…
Interviewee: It's a paper I had sent to my supervisor. We had sent it to a conference some time ago. It was rejected… meanwhile I had placed the document on my UNIX account…

The data collected from the Contents and Relational Analysis was submitted to statistical tests in which the values for the different document kinds and story properties were compared. All results are statistically significant with 95% confidence. A detailed technical report describing this part of the study and its results can be found in [5].

We found stories to be 15.85 elements long, on average (st.dev.=5.97). A significant difference was found between documents created by the user and Other Documents (17.7 and 12.15, respectively). Story length seems to be independent of age, but women tell longer stories than men (16.81 vs. 14.67 elements).

Since a transition between two elements was considered only when the second one wasn't induced, comparing the number of transitions with the total elements number gives us an estimate of how spontaneous the story was. We found that 47% of stories were spontaneous, regardless of document type.

Each continuous sequence of spontaneous story elements was called a story train. The entire story is composed of several of those trains, separated by some induced element. Over 75% of the elements in a story are contained in just three different story trains: the first two, containing 50% of the story, and the last (another 25%), where a final burst of information exhausts the available data. Stories have 2.87 trains, on average, regardless of document type.

Table 2. Element frequencies and avg. percentage of induced occurrences in stories

	Element	Recent	Old	Other	Overall	Element	Recent	Old	Other	Overall
Frequency	**Time**	38	34	28	**100**	**Exch.**	25	18	23	**66**
Induced per story (avg%)		35.0	54.2	59.6	**49.6**		47.7	56.3	34.2	**46.0**
Frequency	**Storage**	31	27	25	**83**	**Place**	27	27	6	**60**
Induced per story (avg%)		47.9	43.3	40.0	**43.8**		65.4	60.0	22.5	**49.3**
Frequency	**Purpose**	34	32	14	**80**	**Personal**	12	13	11	**36**
Induced per story (avg%)		16.7	23.3	11.3	**17.1**		17.5	12.5	21.7	**17.2**
Frequency	**Tasks**	31	24	22	**77**	**Version**	15	16	0	**31**
Induced per story (avg%)		45.8	45.0	34.2	**41.7**		46.3	35.0	0.0	**27.1**
Frequency	**Content**	26	29	21	**76**	**Author**	4	4	15	**23**
Induced per story (avg%)		23.8	37.5	5.8	**22.4**		5.0	15.0	22.5	**14.2**
Frequency	**Other Doc.**	24	29	21	**74**	**Name**	8	5	5	**18**
Induced per story (avg%)		40.0	52.5	57.5	**50.0**		0.0	10.0	1.7	**3.9**
Frequency	**Subject**	28	17	25	**70**	**World**	8	5	2	**15**
Induced per story (avg%)		35.0	40.0	21.7	**32.2**		12.5	5.0	5.0	**7.5**
Frequency	**Co-Author**	30	33	5	**68**	**Events**	2	5	0	**7**
Induced per story (avg%)		55.2	66.3	10.0	**43.8**		2.5	7.5	0.0	**3.3**

2.3.1 Story Elements

The most common overall story elements were **Time**, **Place**, **Co-Author**, **Purpose**, **Subject**, **Other Documents**, **Exchanges**, **Type**, **Tasks**, **Storage** and **Content**. (Table 2). Some were mentioned more than once in the stories (100 occurrences of Time in 60 stories, for instance), whenever the user felt the need to clarify some information. Elements such as **Authors**, **Personal Events**, **World Events**, **Versions**, **Events**, and **Names** were the least often mentioned.

Among document types, the larger differences were found comparing documents created by the users and Other Documents, namely regarding **Co-Authors** (usually only the main author of Other Documents is remembered), **Author** (taken for granted when describing own documents), and **Version** (hard to know for documents of other authors). **Place** and **Purpose** were also different appearing to be easier to remember for documents created by the users, with which they closely interacted.

Less often induced were **Purpose**, **Author**, **Personal Events**, **World Events**, **Events** and **Name** (Table 2). With the exception of Purpose, these are the least frequent elements. Thus, they are rarely and spontaneously mentioned, suggesting that no amount of persuasion can make the users remember them. Purpose is a frequent element, showing it to be important and easy to remember. The more often induced ele-

ments are **Time**, **Place**, **Co-Author**, **Other Documents**, **Exchanges**, **Tasks** and **Storage**, all of which are very frequent, suggesting they are important but require external stimuli to be mentioned. Few noteworthy differences in the percentages of induced elements were found when comparing the several document types.

2.3.2 Element Transitions

We verified that some transitions do occur more frequently than others, suggesting an underlying structure. Only 36.7% of all possible transitions occurred more than once, the most common being **Time-Purpose**, **Tasks-Content**, **Subject-Time**, **Type-Purpose**, and **Storage-Type**, and also the reflexive transitions involving **Content, Place, Time**, and **Storage**. Normalized transition frequency values were calculated, accounting for the relative frequency of the involved elements. This was done to look for biases when infrequent elements appearing together, causing the transition not to be significant. No such bias was detected. With few exceptions, most transitions have low probabilities of occurring, enough to have some expectations but no certainties.

2.4 Discussion

There was little dependency of personal factors such as gender or age to the way in which the stories were told. *No user customization will be necessary in relation to what to expect from a story*. The interface might still need to adapt to specific needs of the users, but not regarding the stories themselves. We did find some differences among stories describing different document kinds. *It is important to determine early in the narrative what kind of document is being described, in order to correctly form expectations about what can be found ahead in the story*.

Several elements appearing in stories are induced, the result of an intervention by the researcher. Since those generally occurred as the result of the storyteller being at a loss for what to say next, it becomes evident that the listener plays an important role, by encouraging the storyteller to continue and remembering more information. *It is important to establish dialogues with users in order to obtain all information they can actually remember*. To know what the storyteller might still add to the story, the data about the frequency of the story elements and whether they were induced or not can be used. That data and the expected story structures can also be used to build expectations about the stories, help disambiguate them, and know how better to encourage the storyteller go on. Particularly relevant in this regard is the fact that events occurring during the interactions with the documents mentioned. *They appear not to be relevant in the stories.* Also important is *a model of the users and their world*, to understand some context-dependent information the usually for granted when telling the stories.

Stories are inherently ambiguous and inaccurate. The precision of the data can vary a lot. For instance, Time information ranged from specific dates and times (“last Friday at 14:00”) to vague references such as “around summer last year”. Also, the remembered information is often incomplete. *Some level of ambiguity must be tolerated by narrative-based interfaces.* A particular aspect of this inaccuracy occurs when the users often remember the overall look of a document rather than specific images of phrases therein. *Some technique that identifies the overall structure or visual appearance of a document and can use that information to differentiate among*

ance of a document and can use that information to differentiate among several documents would be useful.

Often, the users describe documents related to the ones they want to find, in the form of short recursive stories. *Those stories must be captured and analyzed, taking care to prevent the users from digressing from the main story.*

3 Telling Stories in Structured Environments

At this point, we have a thorough characterization of document-describing stories and a set of design guidelines for interfaces that make use of them. However, it is necessary to validate those findings, and verify if stories told to a computer are similar to those told to a human listener. Hence, we developed two low-fidelity prototypes that embody the guidelines above and collected new stories using those prototypes. Those stories were then compared with the ones previously collected.

3.1 The Prototypes

Prototype A (Figure 1) is based on the sequential entry of story elements using dialogues. The screen is divided into three different areas. In Area 1 the several dialogues, one for each possible story element, appear. The order in which they do so reflects the structure found in the previous part of the study. A drop-down list can be used to select a different story element. In Area 2 each story element is displayed as a small box that can be edited, dragged, and deleted. That area is divided into three sections representing, from left to right, the past, present and future. It is possible to drag a story element to those areas establishing a temporal restriction.

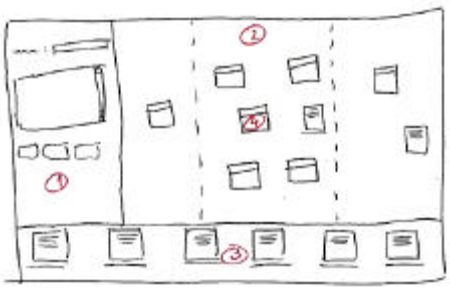

Figure 1. Prototype A

Figure 2. Prototype B

In Area 3 the interface present a list of candidate documents that, from the story told so far, could be the one sought by the user. Those documents can be dragged to the main story area if they seem somehow related with the target-document (written by the same author, at the same time, etc.).

Unlike Prototype A, Prototype B (**Figure 2**) represents the story in a textual manner. Instead of just presenting the dialogues to the users, it displays, on Area 2, incomplete natural-language sentences. The blanks in those sentences can be filled with the help of dialogues similar to those used in Prototype A (Area 1). For instance, the computer could present the user with the following sentence:

This is a document I read **time**

At the same time, the dialogue where information about Time can be entered will be visible. If the user specifies July 2003 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 **last July**. Its author is **author.**

This will go on until the target-document is found. In Area 3, promising documents are shown to the user, as in Prototype A. Finally, although the order in which the several story elements are asked to the users reflects the expected story structures, they can control it with the help of the buttons labeled 4 to 6. The first one, "I don't remember", can be pressed when the user can't remember the element being asked at the time. The interface will just ask the next one. The second, "It didn't happen", is used to indicate that something being asked didn't take place (a document didn't have co-authors, for instance). The third, "I want another", shows the users a list of all possible story elements allowing them to select the element to be mentioned next.

Both these interfaces follow the guidelines found above, albeit in different manners. Time is given an important role, especially in Prototype A. Dialogues with the user are inherent to both prototypes, as is the easy access to other documents and ways to relate them to the target document. The dialogues themselves are built in a way that takes into account possible ambiguities and uncertainties. More details can be found in the technical report that describes this study [6].

3.2 Procedure

A Wizard-of-Oz methodology was used when evaluating the prototypes. Two researchers were present in all interviews. The Wizard was responsible for simulating the prototype's reactions to the users' actions. He possessed a list of sentences to be used in Prototype B, and knew the order in which the several elements were to be asked to the users. The Observer took note of the users' reactions, comments, and all other relevant information, including the order in which the story elements were told, their values, what features of the interfaces were used, and at what time.

The users were asked to tell stories about the same document kinds considered previously (Recent, Old, and Other). Afterwards, they were asked to fill in a small questionnaire, allowing us to evaluate their subjective satisfaction with the prototype.

3.3 Prototype Evaluation Results

Again, the user sample was fairly diversified, to prevent biasing the results. Ten users evaluated each of the prototypes, allowing us to collect thirty stories for each. All results are statistically significant with 95% confidence, unless otherwise stated.

Regarding story length, we found that they were, in fact, longer than the ones previously analyzed! For Prototype A they were, on average, 16% longer, regardless of document type, and those told using Prototype B were 20% longer for documents created by the users, and 35% longer for Other Documents! We attribute this increase in story length to the fact that the dialogues presented to the users help to them to remember more information, as does the story itself, visible at all times.

The elements were presented to the users in an order that directly reflected the underlying structure found in the previous study. Hence, changes in that order are a good estimate of how structure differs from stories told to humans to those told using the prototypes. We found that although 50% of the users of Prototype A chose new elements 1.23 times per story, on 43% of stories, only 10% of the users of Prototype B did so, for 3% of the stories (0.07 times per story). This is a surprising result, considering that the order in which the elements were presented to the users was the same for both prototypes. We see that the form of interaction favored by Prototype B is better suited for storytelling, better mimicking the experience of telling stories to humans. We can, thus, conclude that if the interface is build correctly, the element order found in the previous study remains valid.

We found some differences in the frequencies in which story elements were mentioned. With few exceptions (such as Tasks, found less often when using the prototypes), this is consistent with the longer stories.

Also important is the relative order of the different story elements. We divided them into two categories, for each document type: frequent and rare. In fact, it was noticeable in the occurrences of story elements in the previous study that some were mentioned very frequently, and some were rarely referred to. There is a gap of at least 20% between the two groups. We did this division for the stories told to the prototypes and compared the groups in which the story elements were placed in both story sets. With few exceptions, the relative importance of the elements remains the same. **Name** seems to become somewhat more important in Recent and Old Documents. **Personal Life** information becomes even more infrequent. For Other Documents, the differences are larger, with **Place** and **Version** given more importance.

3.3.1 Comparing the Prototypes

With the help of a questionnaire, we were able to compare the prototypes in terms of the quality of the subjective user experience they produce. The users were asked to rate a set of sentences with a value from “1-Strongly Disagree” to “4-Strongly Agree”. From those ratings, we were able to infer their opinion about the prototypes. Given the low (ten) number of users of each prototype, the usual statistical significance tests could not be used. However, some values can be directly compared to provide insights into the users’ subjective feelings.

Both prototypes were considered equally **satisfying**, **flexible**, and **complete**, in terms of the information they allow the users to mention (with averages of 3.35, 3.2,

and 3.05, and standard deviations of 0.58, 0.61 and 0.82, respectively). This is unsurprising, since similar dialogues were used to collect information in both prototypes.

Although both prototypes were considered **simple** and easy to use, Prototype B was considered the simpler (3.7 vs. 3.3 with a std. dev. of 0.48 in both cases). It would seem that the textual representation of narratives is felt as more natural by the users. In terms of **novelty**, Prototype B was also the winner (3.9 vs. 3.6 with std. devs. of 0.52 and 0.32, respectively). The approach followed in Prototype A was perceived as more similar to existing solutions. Finally, with respect to **understandability**, Prototype B is, again, superior to Prototype A (3.7 vs. 3.3 with std. devs. of 0.82 and 0.48). This leads us to conclude that presenting the story using natural language sentences is less demanding, cognitively-wise, than graphically separating the story into its elements.

On Table 3 we find a summary of the usage of the main features of both interfaces. As we had already mentioned above, half the users of Prototype A chose different elements than those suggested by the interface 1.27 times per story, and for Prototype B that value is of only 0.07. This reflects the fact that Prototype A is more confusing, making it hard for users to see how certain elements fit in the stories.

The values are much more similar regarding the number of times the users just let the prototypes choose a new element for them. The decisive factor here seems to be only the users' memories, and not the prototype in use. Virtually never in both prototypes was the ability to correct an already introduced story element used. It seems that users have great confidence in what they remember. In a real system it is conceivable that, after failing to find a document, this feature could be used.

Dragging documents from the suggestion list to the main area of the interface was seldom done in Prototype A (0.4 times per story for 23% of stories). This could be due to the fact that no real documents were used. In a real usage situation, those documents could elicit better responses from the users in this aspect. Even less used was the ability to move elements to the past and future areas of the interface. In fact, this was considered to be one of the most confusing aspects of Prototype A.

Table 3. Average usage frequency of interface features (per story)

		Rec.	Old	Oth.	Over-all	Stdv	%Stories	%Us-ers
PA	**Chose**	1.40	1.30	1.10	1.27	0.15	43.33	50.00
	Moved On	2.44	2.80	5.20	3.48	1.50	76.67	80.00
	Correct Element	0.00	0.10	0.00	0.03	0.06	3.33	10.00
	Drag Document	0.60	0.40	0.20	0.40	0.20	23.33	50.00
	Drag Past/Fut.	0.27	0.18	0.18	0.21	0.05	23.33	40.00
PB	**Chose**	0.00	0.00	0.20	0.07	0.12	3.33	10.00
	Moved On	2.70	3.50	3.44	3.21	0.45	93.33	100.00
	Correct Element	0.00	0.00	0.00	0.00	0.00	0.00	0.00

3.3.2 User Comments

From the users' comments and reactions, it was possible to detect important limitations of the prototypes. First of all, although there was a dialogue for entering infor-

mation about Other Documents, the users felt it was not enough. Some true support for recursive stories is required. It was also mentioned that the Personal Life Events, World Events, and Events story elements are confusing and even useless. It was hard to tell them apart, and they were often confused with other story elements. They were explicitly described as “superfluous”, “unnecessary”, and “useless” by some users. Some users did like them, though, as the frequency of their occurrence shows. It would seem that their usefulness is highly dependant of the user.

Many users of Prototype A complained about the order in which the story elements were suggested to them. Not even one user of Prototype B did so! We can conclude that Prototype A is not mimicking the storytelling process adequately.

Some users felt that some of the dialogues became redundant, in some situation. Overall, several suggestions were made regarding the improvement of the dialogues, a full account of which can be found in the study’s technical report [6]. Notably, there were complains about the lack of better support for recursive stories about related documents, and some confusion between Purpose and Subject.

3.3.3 Discussion

The most important conclusion that this study allows us to achieve is that, indeed, *stories can be told in structured environments just like they are told to humans*. The differences are, if anything, advantageous. For instance, stories told to the prototypes were larger than those told to humans, conveying more information. Story structure remains largely unchanged, and the elements maintain their relative importance.

For this to happen, it is important to maintain the illusion of storytelling with an adequately conceived interface. Prototype B, based on text-represented stories, was clearly better suited for that task. Not only were the stories told using it longer but also the users hardly felt the need to choose different story elements just “going with the flow”, as if telling a real story. This means that the elements were easier to recall in that prototype, reinforcing the conclusion that it is simpler and easier to understand.

5 Conclusions

We verified that, indeed, narratives can be a natural, efficient way for users to remember important information about their documents, and to convey it to a computer. This makes stories an important alternative to traditional document retrieving approaches which are quickly becoming ineffective given the ever growing document numbers the users must face nowadays.

After collecting sixty document-describing stories, we were able to devise a thorough characterization of those narratives, identifying not only their overall structure, but also what elements they contain, and which were more important (Time, Purpose, etc.). From this we extracted a set of user interface design guidelines.

When trying to retrieve a document, the stories will be told to a computer, and not to a human. Hence, we performed another study in which the previous findings were validated using two low-fidelity prototypes. The stories collected using them were, indeed, similar to those told to humans. We soon verified that it is important to main-

tain the illusion of telling a story in the users. If that illusion shatters, the stories immediately suffer in terms of quality. This explains the differences found when comparing both prototypes. Prototype A, where the different story elements were separately represented, was clearly unable to maintain that illusion, resulting in a worse user experience. Prototype B, based on structured text entry and presenting stories in a textual manner, better mimicked the experience of telling stories to human listeners. In short, it is important for the interface to be correctly designed and tested with the users' help.

The next step in our research will be to build a functional prototype based on Prototype B and improved with the help of the suggestions made by the users. This will allow us to further validate or findings, and to answer some important research questions that have yet to been addressed, regarding narrative-based document retrieval. The most important unresolved issues concern the accuracy and discriminative power of stories, so far unexplored since no real documents have been considered in our studies, allowing us to validate the stories in a real situation.

References

1. Baeza-Yates, R., Jones, T. and Rawlins, G. A New Data Model: Persistent Attribute-Centric Objects, Technical Report, University of Chile, 1996
2. Dourish, P. *et al*. Extending Document Management Systems with User-Specific Active Properties. *ACM Transactions on Information Syst.*,18(2), pp 140-170,ACM Press 2000.
3. Freeman, E. and Gelernter, D. Lifestreams: A Storage Model for Personal Data, *ACM SIGMOD Record*,25(1), pp 80-86, ACM Press 1996
4. Gifford, D., Jouvelot, P., Sheldon, M. and O'Toole, J. Semantic File Systems. *13th ACM Symposium on Principles of Programming Languages*, October 1991.
5. Gonçalves, D. Telling Stories About Documents, Technical Report, Instituto Superior Técnico, 2003 (http://www.gia.ist.utl.pt/~djvg/phd/files/telling_stories.zip)
6. Gonçalves, D. 'Telling Stories to Computers'. Technical Report, Instituto Superior Técnico, December 2003. http://narrative.shorturl.com/files/telling_stories_to_computers.zip.
7. Huberman, M. & Miles, M. Analyse des données qualitatives. Recuil de nouvelles méthods.. Bruxelles, De Boeck 1991.
8. Malone, T. How do People Organize their Desks? Implications for the Design of Office Information Systems, *ACM Transactions on Office Information Systems*, 1(1), pp 99-112, ACM Press 1983.
9. Rekimoto, J. 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, 1999.
10. Whittaker, S., Sidner, C. Email overload exploring personal information management of email. In *Conference proceedings on Human factors in computing systems*, pages 276-283, ACM Press, 1996.
11. Yin. R. Case Study. Design and Methods. London, Sage Publications 1989.