Surpassing Farley Files: Opportunities and Challenges on Obtaining Personally Relevant Information

João Guerreiro, Tiago Guerreiro, Daniel Gonçalves

Technical University of Lisbon / INESC-ID
R. Alves Redol, 9, 1000-029, Lisbon, Portugal

+351 21 4233565

joao.p.guerreiro@ist.utl.pt, tjvg@vimmi.inesc-id.pt, daniel.goncalves@inesc-id.pt

ABSTRACT

The proliferation of personal devices and their constant awareness of our interactions have generated an enormous amount of data that can be useful to help the user obtaining relevant information when needed. Our approach uses the personal information on users' devices, together with public online sources, to provide relevant information from the user point of view. The information from the users' devices, due to its personal and credible character, works as a filter to the retrieved from other less trustable and structured sources. A preliminary evaluation, suggested that we can provide the user with inter-connected relevant information from heterogeneous sources. However, we found some limitations that led us to our current research challenges.

Categories and Subject Descriptors

H.5.m [**Information interfaces and presentation (e.g., HCI)**]: Miscellaneous.

General Terms

Design, Experimentation, Human Factors.

Keywords

Personal and Public Information, Mobile, Information Filtering.

1. INTRODUCTION

Not so long ago, our personal information (PI), whether documents, agendas or letters, was restricted to limited, private environments. Times have changed and the Personal Information Management (PIM) research field has emerged to face the challenges of an immense universe of information. Never have we had so much personally relevant information almost instantaneously at our disposal. In particular, mobile devices are widely used and have become essential tools for most of us. They contain information about the users, their habits and daily interactions, as no other person or device. There are some projects that try to make use of PI to help the user, however they are restricted to a limited set of data (ex: [1-3, 5, 6]). There is no way to get simple, coherent answers to questions about people, places, subjects or events. The information in public online sources (social networks, blogs, search engines, etc) tends to be either generic or tailored to a particular audience, and much of it might just not be relevant to a particular user. A way to garner all available information combining it with the users' own PI to produce personally-relevant results is lacking.

James Farley, Franklin Roosevelt's campaign manager, kept a file of every person the President met, so that in future occasions he could provide him useful information about past meetings or even more intimate details. Our approach has the same essence. According to preliminary results, it can help the user by relating, summarizing and providing context- and personally relevant information when required [4]. It does so by using data from the user's devices enriched with that from public sources. The credibility and individual character of PI can help filter the information from online sources, resolving their inherent ambiguity and presenting results with meaning and interest to the user. All this information is interrelated in a structured and meaningful knowledge base. While laying the ground work for applications in several contexts and domains, we believe that the developed framework is likely to have great benefits in a mobile context, one where, nowadays, the user is highly PI-deprived and where that information is likely to be required on time.

2. POSSIBLE SCENARIO AND SOURCES

To avoid privacy issues and dependency on what others may provide, we consider only the PI existent in the user's devices and public sources of information. Below is a possible usage scenario:

*"At a coffee break I find someone that seems familiar talking to a friend of mine. Using his **Bluetooth ID**, I ask the system about our past interactions, finding that he is Tom Parker, that we were together 2 years ago in Rome and that we exchanged two **email messages** and a **document**. The subject is also shown, as are the **people** I have forwarded it to. It also shows, resorting to my mobile device **agenda**, that I had a meeting with him and other three colleagues about PIM at that time. I also found a **SMS** postponing it, so I remember he was not very rigorous with schedules. I ask for more information. I drink my coffee and see that according to his **personal webpage** and **blog** he is currently working on the same subjects I am, his wife is Marcia Parker and he one kid, John. His **social network** profiles also show that he is a soccer fanatic, he is from Dallas and loves Jazz. Ok, now that I know all that useful information, I can approach Tom Parker."*

3. SURPASSING FARLEY FILES

We implemented a framework that uses the PI on the users' devices, together with public online sources, to provide useful and relevant information from the user perspective, in an iterative information seeking process. The information retrieved from the users' personal devices, due to its personal and trustable character, works as a filter to the information retrieved from other less trustable and structured sources. We defined a single structure to inter-relate heterogeneous information as a coherent whole.

3.1 Personal Filtering of Public Sources

We are dealing with two different types of information: PI existent in our devices; and public information mainly from online sources. Although the users' devices provide reliable and relevant information from the users' point of view, with a proper meaning to them, public sources are generally much more ambiguous. Also, there can be a lot of information about the same person, how can the user get information about the subjects that really matter to him/her? The PI retrieved from the user's devices, is the perfect candidate to help filtering the ambiguous information that public sources provide. Thus, it is possible to identify which data is related to our search and at the same time collect relevant results from the user point of view. The information that matters to a certain user may be irrelevant to others and vice-versa.

4. PRELIMINARY RESULTS

We developed an example application with the goal of obtaining information about a person. This application resorts to four different types of information sources: PI on the devices (e-mails); social networks (facebook); blogs (Blogspot) and Wikipedia. We limited the sources in this proof-of-concept application to these as they represent different types, enough to prove our assumptions and keep the system evaluation simple and quicker.

Preliminary evaluation with users showed that it is possible to provide the user with inter-connected relevant information using personal and public heterogeneous sources. The user had to search for persons with three degrees of relationship (close friend, "known" and famous). Results showed that our framework could retrieve relevant information that the user was expecting before the search (information that characterizes the person), together with other the user was not expecting but also found relevant. Results also suggest that PI can be used to guide searches and filter information from more ambiguous sources (public).

5. NEXT STEPS AND IMPROVEMENTS

Our preliminary results indicate that our approach is feasible. However, some limitations need to be addressed:

Personal Filtering of Public Sources. In the evaluation some users had friends in common, so we used that to compare their searches and check if their PI influenced/filtered the remaining results. In some cases, we observed that only users with PI related to the searched person could find and present results from his/her blog. It was possible only because the PI was used as context to further searches. However, we were expecting more information about this feature, which we think can be better tested when we extend our number of information sources. For instance, if we introduce search engines, when a user searches for a person there will be thousands of results relative to several different persons with that name. We are studying clustering the different results, so we can separate the web pages that talk about different persons. The PI is crucial to identify the right cluster.

Dealing with Unstructured Information. In our first prototype we made a very simple algorithm to deal with the unstructured information, since it was not our main focus. Now we are trying to improve it, studying how Ontologies or Natural Language Processing can assist us in this regard. Also, we previously focused only on web pages that had known structures, which limits our possible information sources. We will try to recognize patterns to identify which information might be relevant or not in some webpage and even on our PI. Email signatures are an example of patterns that could add value to our framework.

Extending Searches Context. Our framework was built with the intent of organizing information no matter what is its nature. Although the example application deals with person searches, in future work, other applications will be explored focusing on other types of information (events, locations,..).

Mobile Context. Mobile devices fill great part of the scenarios where our framework can be used. It is most likely that users want immediate answers when they need information and it may not be useful when they reach their personal computers. We will study how the presentation of the information can be influenced by the usage of mobile devices and how it can be of use providing personal information (SMS, phone calls, agenda, Bluetooth, etc).

User Studies. We will analyze how users employ different sources when seeking personally relevant information, to find the most promising ones, establish confidence values based on credibility and find usage patterns for each one of them. It may allow us to adjust the credibility value based on the type of search.

6. CONCLUSIONS

Trying to get information about someone or something in a mobile context is a recurrent task for many users. Our approach gathers PI from the user's devices, and uses it as a filter to the information available in public sources like search engines and social networks. After an iterative process of searching and improving the information retrieved, from the user point of view, it is able to present contextualized personally relevant structured information. We evaluated an example application, which suggested that our approach can provide good results. However, we identified some flaws that we are trying to suppress.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Ankolekar, A., et al.. 2009. Friendlee: A mobile application for your social life. In Mobile HCI 09, ACM.

[2] Beach, A., et al. 2008. WhozThat? Evolving an ecosystem for context-aware mobile social networks. Network, IEEE, 22(4), 50-55.

[3] Falke, E. 2008. The associative pda 2.0. In CHI '08 EA on Human factors in computing systems, ACM, 3807–3812.

[4] Guerreiro, J., et al. 2010. Relating Personally Relevant Social Information from Heterogeneous Sources. IUI Workshop on UDISW, ACM Press.

[5] Lamming, M. et al., 2004. "Forget-me-not: Intimate Computing in Support of Human Memory. Proc. of Next Generation Human Interfaces.

[6] Rhodes, B. J. and Starner, T. E. 1996. Remembrance agent: A continuously running automated information retrieval system. In Proc. of PAAM '96.