In Search of Personal Information: Narrative-Based Interfaces

Daniel Gonçalves, Joaquim A. Jorge
Instituto Superior Técnico
Universidade Técnica de Lisboa
Av. Rovisco Pais, 49
1050-001 Lisboa, Portugal
Tel: +351-213100289
{daniel.goncalves,jaj}@inesc-id.pt

ABSTRACT
Most computer users find organizing large amounts of personal information problematic. Often, hierarchies are the sole means to do it. However, users can remember a broader range of autobiographic contextual data about their personal items. Unfortunately, it can seldom be used to manage and retrieve them. Even when this is possible, it is often done by asking users to fill in values for arbitrary properties in dialog boxes or wizards.

We propose that narrative-based interfaces can be a natural and effective way to help users recall relevant autobiographic data about their personal items and convey it to the computer. Using Quill, a narrative-based personal document retrieval interface, as a case-study, we show how such an interface can be designed. We demonstrate the approach's validity based on a set of user studies, discussing how the problems raised by the evaluation of such an interface were overcome.

Author Keywords
Narrative-based Interfaces, Document Retrieval, Personal Information Management

ACM Classification Keywords
H.5.2 User Interfaces: Interaction Styles, H.5.2 User Interfaces: User Centered Design, H.3.3 Information Search and Retrieval: Query formulation

INTRODUCTION
One of the problems that computers users nowadays face is the difficulty in managing their personal information. This information is spread throughout different applications with little or no relation between themselves. This Fragmentation Problem can lead to severe difficulties when trying to organize and retrieve relevant personal information, such as documents, emails, or bookmarks [2]. While those items are undoubtedly connected to each other, this connection exists in the users' minds only. For instance, it is impossible, from a document in the filesystem, to find the email message it came attached to.

The same organizing principle is recurrently used for most personal information: hierarchies. Indeed, they are the most commonly found means provided by applications to help users organize their data. Each element must be classified into a specific category in the hierarchy. This is a task many users find daunting. Often, none of the existing categories seems to apply. Sometimes, several appear to be relevant. In the end, only one can be chosen. This leads to high cognitive loads in users, distracting them from their main tasks. Even worse, they know that a bad choice at classification time can result in a difficult (if not impossible) retrieval later on. This occurs because the item's location in the hierarchy is the main (if not the sole) hint to its whereabouts. Hence, its classification must be done with care, under the penalty of never finding it again!

So troublesome is the classification task that it was identified as a problem by one of the first studies about the users' organization of their documents. In 1983, Malone found that many users avoid classifying their documents altogether, relying instead on other clues (their physical location, for instance) to find them [18]. Nowadays, with the large quantities of information at the disposal of computer users, the problem has become more serious. It is often the case when it seems that "stuff goes into the computer and doesn't come out" [3].

A way alleviate these problems is to allow users to retrieve their personal information based not only on arbitrary classifications in a hierarchy, but reporting to a wider range of contextual autobiographic information. The users have interacted with their personal items (documents, emails,

IUI'08, January 13-16, 2008, Maspalomas, Gran Canaria, Spain.

etc.) in the past, in specific contexts and situations, information about which, being more meaningful to users, is also easier to recall. Also, if users are able to refer to a wider range of relevant information when describing their items, the retrieval will not be as prone to mistakes as when relying on a single location clue.

All this was confirmed by different studies in recent years. An analysis of the users' email inboxes [27] showed that many relied on their email tools to find documents or contacts, even when that is not those tools' purpose, and no special support for those tasks is given. Storing relevant information in email messages is seen as advantageous since those messages are associated to a wide range of meta-data that can be used to successfully retrieve them: their subject, the time when they were sent / received, their sender and recipient(s), etc.

All data necessary to help users manage their personal items is present, in different ways, in today's applications and operating systems. What is required is a way to consistently gather it and use it to manage those items, and a way to help users convey it to the computer in meaningful ways. This is, inherently, a user interface problem. A properly designed interface should facilitate the recall of relevant autobiographic information. Having to remember and fill in values into arbitrary property fields can be as daunting as classifying documents into hierarchies.

We propose that narrative-based interfaces can provide a way to help users remember and convey relevant autobiographic information to the computer. Humans are natural-born storytellers. Stories have been told since the dawn of mankind, across all ages and cultures. We tell them from birth to our deathbeds. In fact, anthropology studies show storytelling to be a universal human activity [4]. Thus, an interface that allows users to tell their stories about their personal information items (documents, emails, etc.) would be a natural and efficient way to help users mention relevant information about them.

In this paper, we'll discuss the design of narrative-based interfaces for personal information retrieval, with the help of a case study: the creation of Quill, a narrative-based interface that allows users to find their personal documents by telling stories about them. We will describe its design process, showing how it can be applied to other domains, discussing the issues and challenges it raises, and showing how they can be overcome.

In the next section, we will describe relevant related work in the area. Next, we will show how Quill was created, following which the interface and underlying system will be succinctly described. Then, we will analyze the issues raised by the evaluation of narrative-based interfaces and show, based on Quill's evaluation, our approach's validity. We will then conclude, pointing to relevant future work in the area.

RELATED WORK

Recognizing the limitations of relying solely on hierarchic locations when retrieving personal items, several works have tried to help users employ a wider range of information.

In Lifestreams [7], all files are presented to the user sequentially, organized in temporally sorted streams, theoretically starting at the user's birth. It is possible to filter the files in the streams to create substreams, more relevant to certain tasks. Also giving time a central role, Rekimoto's Timescape [21] presents the idea of a temporal desktop, in which its contents appear only in certain time spans. Older, unused, items fade with time until they disappear, and it is possible to visit past and future versions of the desktop.

While interesting, temporal-based solutions like the ones we just described do not fully achieve the goal of allowing users to report to the wider range of autobiographic information that surrounds them. By giving time a central role, other information items are underused or ignored. More encompassing are property-based systems. Seminal in this approach is Gifford's Semantic File System [9]. There, virtual folders based on the values of predefined properties can be used to find items in the filesystem. More recently, and using the same principle, more sophisticated systems have been created. Paul Dourish's Placeless Documents [5] and Ricardo Baeza-Yates' PACO [1] both use search criteria to create document collections that are automatically updated when new document match them. Such virtual folders have become common, appearing in mainstream systems such as the recent Microsoft Vista OS. However, remembering values of arbitrary properties to enter in text fields is, in itself, problematic.

The previous solutions are focused on data that can be gleaned directly from the files. Others try to make explicit use of other related information sources. It is the case of Haystack [17], where a semantic interface is created based on the types of elements displayed. All of a users' personal information can be displayed and browsed, interrelated with each other. The actions allowed for each item depend on their semantics. The Stuff-I've-Seen system [6] integrates information gathered from the documents and the different applications executed by the users (email, etc.) providing a richer context. MyLifeBits [8] goes further and allows the use of information gathered by other devices, with the ultimate goal of collecting all information pertaining a user.

A more unstructured approach is tagging, popular in web-based applications. Tagging allows users to easily annotate their files. More than one tag can be used per item, partially solving the problem of choosing a single category for each. However, the lack of structure can lead to inconsistent tags and low tag reuse, compromising the retrieval task [10].

All these solutions share the same inherent limitation: the

interfaces are not designed in a way that helps users recall relevant information. Most rely on browsing or keyword search to help users find their files (as is also the case of most desktop search solutions, such as Google Desktop). Even when more contextual data is available, usually only after a keyword search has been performed can it be used to sort or filter the results. A way to allow users to naturally and efficiently use that data as the primary means to describe personal files is required.

DESIGNING A NARRATIVE-BASED INTERFACE

One of the most common types of personal items users must manage in their computers are documents. Organizing them poses a particularly hard challenge, as they are the most heterogeneous type of personal information that users must manage. Managing varied document types is not easy, as proved by the myriad of special-purpose management and retrieval applications that exist for different document types (texts, photos, music, etc.). We decided to design a narrative-based interface for personal document retrieval, regardless of document type.

Understanding Stories

The first step that has to be given when designing a narrative-based interface (and one of the most important) is to gain an intimate knowledge of narratives for the chosen application domain. If the interface is to be able to allow users to tell their stories and to understand them, it should be designed based on what can be expected of those stories. In our case, it was important to know what narratives about documents are like.

Not only the contents of stories are important. Indeed, one of the most useful characteristics of stories is that the different information elements in them do not appear independently from each other. Rather, they are interconnected as a coherent whole. Thus, the stories' structures must also be understood. This will allow the interface to help the users to maintain the illusion of storytelling and prevent the disruption narrative's flow.

The best way to gather information about stories is to ask users to tell them and then to analyze those stories. Since it is desirable that stories are as complete and relevant as possible, some way of helping the story move forward when users seem at a loss about what to mention next should be considered. However, this should be done as little as possible and in a natural and unobtrusive way, as not to disrupt the story's flow. For this reason, we opted for semi-structured interviews. The users were allowed to tell their stories with no restrictions. Whenever they seemed stuck the interviewer had a set of questions that could be asked in order to help the users to continue. While a fully unstructured interview might have best served the purpose of collecting “untainted” stories, the amount of information in each story, and its quality, might have suffered, rendering the analysis harder and the results weaker.

A wide range of stories should be considered, to get a general idea of what to expect. In our domain, documents, we needed to take into account domain-specific features that could be determinant for the stories. It is the case of document age. Older documents, handled long ago, could result in different stories than more recent ones. Also, the document's author (the user or someone else), might be of influence. To account for possible differences in stories due to those factors, each user was asked to tell three different stories: about a Recent document, created by the user up to two weeks ago; about an Old document, created by the user at least 6 months ago; and describing a document with Other authors.

Analyzing the Interviews

We interviewed 20 users and collected 60 stories [12]. Those stories were transcribed, omitting all references that could help identify the users. Ensuring the total privacy of this study was paramount in ensuring the users' cooperation. The transcripts were subjected to a contents analysis [16]. Each was coded by hand, and the different phrases in the stories classified as belonging to one of several possible story element categories. Those categories had been determined beforehand with the help of a set of test interviews (although they could have changed if deemed necessary during the analysis). Those elements were: **Time, Place, Co-Author, Purpose, Subject, Other Documents, Exchanges, Type, Tasks, Storage, Contents, Authors, Personal Events, World Events, Versions, Events** (involving the user when handling the document)**,** and **Names**. Although there are tools to perform this coding automatically, they cannot be used when analyzing stories, as they rely on predetermined words or phrases as belonging to each category. Narratives about personal items are so unconstrained in their scopes that finding an appropriate set of such words would not be possible.

All the quantitative data was subjected to statistical tests (with 95% confidence) to extract overall trends for document-describing stories [12]. We found that stories are, on average, 16 elements long, being longer for documents written by the users than for those written by others. Also, there are no relevant differences regarding user gender or age. The relative frequencies and importance of the different story elements were also found. For instance, information about when a document was created, where this took place and its purpose are very frequent, occurring in nearly all stories. The elements themselves were studied, allowing us to gather qualitative data about stories. For instance: when users mention the time when they read a document, how do they describe it?

To understand the stories' structures, we performed a relational analysis, in which the transitions between story elements were recorded, and their probabilities noted. In our analysis, we considered two different kinds of story elements: spontaneous and induced. Spontaneous elements

are those mentioned normally in the course of a story. Induced elements are those that were mentioned only after the researcher had made some question. As those elements might not have appeared if not for our intervention, special care was taken when analyzing stories containing them. The most important consequence of this is that no transition between two story elements was considered to exist when the second was induced, as the connection probably does not exist in the user's mind.

Archetypical Story Structures

One important result from the analysis of stories are archetypical story structures. Those structures can then be used in the interface to know what to expect next in a story, making its understanding easier. Also, whenever the user seems at a loss as what to mention next, the interface might want to coax the users to continue, by asking them about the element that would most likely be mentioned next, so that the storytelling can proceed as naturally as possible.

To infer archetypical structures, we used the transition probabilities obtained from the stories and the relative frequencies with which the different story elements appear to train Hidden Markov Models. Those models were then able to generate the most likely element order for a story.

Interface Design Guidelines

All the statistical data we collected is, in itself important. However, it can be difficult to apply it directly to the design of an interface. We abstracted all quantitative and qualitative results into 9 guidelines for the design of narrative-based interfaces that allow the description of personal items:

1. Personal factors like age or gender are not relevant to stories. Some user customization should be considered, but overall stories remain the same.
2. There is no need to adapt the interface to expect different stories for items that were handled a long time ago, since all stories are similar, regardless of when the interaction took place.
3. Maintaining a dialog with the users when they their stories is important, preventing them from digressing and to elicit all the information they might remember.
4. Much of the information in stories is context-dependent.
5. Some level of ambiguity and inaccuracy is to be expected in stories, suggesting that information in stories should not be seen as an absolute truth.
6. World and user models are required to understand stories. This includes personal information about the users' typical activities, family, etc.
7. While some user variability exists, there are overall trends in the stories' structures that can be used to guide the storytelling and better understand them.
8. Events occurring in the world and, especially, around the user while the item is being handled are very rarely mentioned by users, even when prompted to do so.
9. It is common for recursive stories (stories about related items) to occur. They should be handled taking care not to confuse information about different items.

Shaping the Interface

The study we just described gave us a thorough understanding of document describing stories, and yielded a set of guidelines for the design of an interface that is able to collect and understand those stories. However, the actual shape of that interface remained to be found. In order to do so, a new study was performed.

We used the design guidelines to create two low-fidelity prototypes of possible interfaces (Figure 1). Prototype A was based on the direct manipulation of story elements using a point-and-click interface. Prototype B, on the other hand, represented the stories textually [11]. Different sentences corresponding to the different elements would be displayed in turn, with blanks to be filled in with the appropriate information. We tried to limit the number of variables in play by using the same dialogs to enter information in both prototypes. The different possible story elements are suggested to the users in turn, in the most likely order found on the previous study. The users could then enter the corresponding information, choose another story element to mention, state that something didn't take place, or that they don't remember anything relevant.

We did not create a prototype based on the most direct approach possible: allowing the users to write the entire story as free form text. Indeed, while this might appear a good idea at first, it should be avoided when creating narrative-based interfaces. It directly conflicts with several design guidelines. One of its major problems is that it precludes the possibility of guiding or helping the users to tell their stories and, as we have seen (Guideline 3), some level of dialog between the storyteller and the listener is necessary. A more pragmatic issue is that, given the unrestricted nature of stories, in which almost everything can be mentioned, and the current state-of-the-art of natural language processing research, it would be hard to understand what is being said. Finally, asking users to write the entire story would be too time-consuming and tiresome. In the interviews, they explicitly made clear they would not be willing to spend such effort to perform tasks that could be performed in some other, albeit less perfect, way.

Twenty users were interviewed, ten of which told their stories using one of the prototypes. Three stories were told by each user. The stories collected using both prototypes were compared with those told to humans, looking for statistically significant differences.

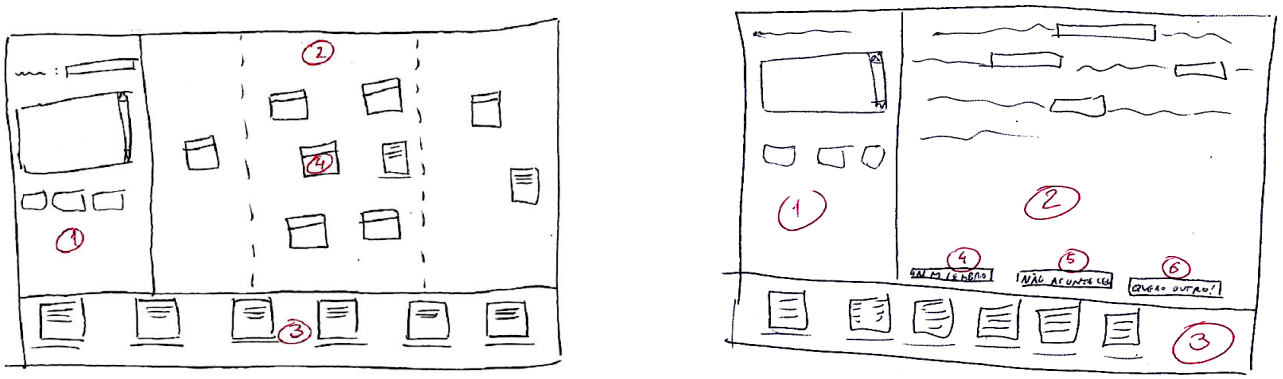

Figure 1 – Low-Fidelity Prototypes A and B

Both in terms of contents and structure, the stories told using Prototype B were much more similar to those told to humans than those collected with the help of Prototype A [11]. The same elements appear with the same relative importance, and only one user once deviated from the order found to be most likely in the previous study (unlike half the users of Prototype A, for 43% of stories). Also, the storytelling process was much easier and straightforward using Prototype A.

These results, led us to choose Prototype B, where the story was displayed textually, as a good interface for users to enter their stories. It, as a whole, better replicates the experience of telling stories to humans.

The Fill-In-The-Blanks Approach

The structured storytelling approach used in Prototype B has the advantage of representing the story as text, while avoiding the problem of asking users to write the entire story themselves. Only the relevant parts of the story need to be entered. Also, the entire story is visible at all times in an easy to understand format, allowing users to scan it and helping them remember additional information. As an example, the interface might start by suggesting the user can mention when a document was created:

> The document was read around Time.

The user then fills the blank with appropriate information. The sentence changes to reflect this, and the next story element is suggested to the user:

> The document was read around ***January 2006***. I read it because I Purpose.

This will continue until the entire story has been told.

THE QUILL SYSTEM

Quill's interface can be seen in Figure 2. It directly mimics the design of Low-Fidelity Prototype B, described in the previous section [14]. 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, 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 sentence when entering 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, to prevent distracting and confusing the users.

The information is entered with the help of specialized dialogs, one for each element. These are placed to the left of the story area (the one for the Appearance element is shown). The different elements are suggested to the users in the order inferred from stories told to human interviewers. 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. The fail-soft principle was used when implementing Quill: even if the system is unable to understand what was entered by the user in a given dialog, it will carry on, no to distract the users whenever, and preventing time-consuming trial and error iterations.

Visible at all times at the bottom of the element entry dialogs, 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 dialog. 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): not knowing something is different from knowing something not to have happened. Clicking on the dialog's title causes a list to pop up, from which the users can choose any of the elements to mention next.

The system is continuously looking for the target document, based on the story told that far. The most likely candidates

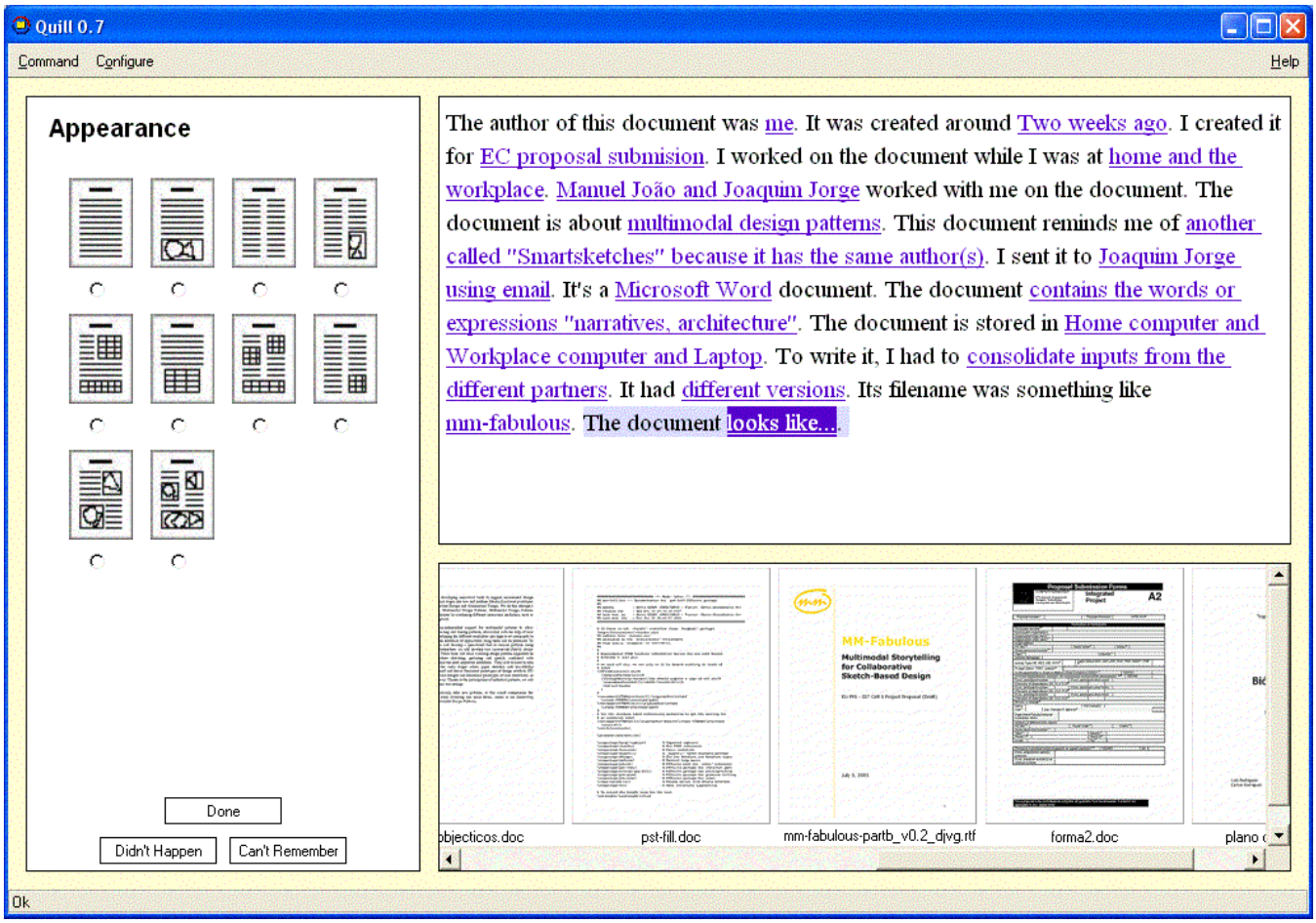

Figure 2 – The Quill Interface

are displayed at the bottom of the interface window. For each, a thumbnail is generated, whenever possible. This allows users to easily identify the document they are looking for without getting distracted from the story, and capitalizes on the fact that one of the things the users remember well about documents are their visual aspects.

While the interface in Figure 3 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.

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

The KB uses RDF and RDF Schema as knowledge representation formalisms. [26]. A special-purpose library, Scroll, was created to build an abstraction layer on top of RDF, allowing more complex constructs than RDF triples to be handled with ease. It allows path- and node-based inference, as well as the creation and evaluation of first-order-like inference rules.

All the information is fed into the KB by an automatic monitoring system that gathers it from different sources. Currently, it is able to gather (and keep updated)

information about the users' documents, emails sent and received (and the attachments therein), web pages visited, applications ran, the users' datebook or agenda, and printed versions of documents (with the help of RFID technology) The separation between the components ensures extensibility of the system, 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 of 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. Whenever the user enters a new element into the story, a new set of inference rules is created. The Document Searcher sub-module of Quill evaluates each of those inference rules in the KB. It is expected for one of those rules to identify, as the possible bindings of a determined variable, a set of documents that match them. Each of those documents is assigned a score. The sum of the partial scores, derived from the inference rules resulting from each story element provides an overall ranking score of the document. Those with higher scores are shown to the user.

EVALUATING QUILL

Based on the results of the user studies already described, we created Quill, a narrative-based interface for personal document retrieval. However, the effectiveness of that interface in particular, and the entire approach in general, remained to be evaluated. Depending on the domain and the types of personal items described in stories, it might be hard or impossible to create an interface that allows users to tell their stories.

With the help of Quill, we performed a set of user studies in which we were able to prove that narrative-based interfaces can be used help users describe personal documents. It was important to measure the inherent quality of the approach, rather than software limitations of the prototype. A more polished system would undoubtedly provide better results, but it is important to evaluate the interface early in its development to be able to easily correct any problems. To validate Quill's interface, it is necessary to affirmatively answer four different questions:

Are stories similar to those told to humans?

> When users told stories in the previous studies there was a human researcher present. This could have unwillingly influenced the results. It is important to see if stories told using Quill share the properties of those told to human, showing the interface does not hinder the storytelling process.

Are stories accurate and trustworthy?

> It might happen that stories contain lots of incorrect or inaccurate information. The users' memories are not perfect, and mistakes might occur. It is important to understand to what extent this can happen and whether it might compromise the retrieval process.

What is the discriminative power of stories?

> A document's description in a story might omit some distinctive feature, thus preventing that document from being identified among several possible candidates. If a story is only able to discriminate between hundreds of personal documents, it won't be useful to help users find a specific document.

What retrieval rate can be achieved using Quill?

> Even if accurate and discriminative stories can be told, it might still not be possible to successfully retrieve specific documents. The knowledge base might not contain enough information to facilitate it, some documents may have been incorrectly indexed, or some other practical aspect might hinder the retrieval process. It is of capital importance to estimate the actual retrieval success rate.

The first two questions apply to any narrative-based interface, regardless of their domain. The latter two are directly related to Quill's application domain, and can provide an idea of whether a narrative-based interface is suited for that domain or not.

We performed two user studies. The first allowed us to obtain results regarding the first two questions, while the second's results were of help in answering the other two. All results were verified to be statistically significant with 95% confidence.

When evaluating narrative-based interfaces, we found it to be important to keep some issues in mind. Firstly, to get the users' cooperation, it was important to ensure their privacy, in clear and believable ways. A lot of sensitive data resides on their computers, and even friends and family can refuse to cooperate if privacy is not taken into account. Also, it was important for users that the prototype would not leave behind any leftover traces on their computers.

All user studies have to be performed at the users' own computers, in their homes or workplaces. As we needed users to describe personal documents (although this is true for any electronic items), it is important that they have access to those documents. For the same reason, prearranged document sets cannot be used. The researcher's documents are not adequate. It would be possible to build a set of the users' documents only with their help, which would bring those documents to the users' attention and invalidate the study by reminding them of information about those documents when handling them.

Story Quality and Trustworthiness

As we were concerned with the quality of the narrative-based interface, and not the underlying system, it was not necessary for it to be fully working. This allowed this study to be performed in the very early stages of development.

We installed Quill on the users' machines, and allowed it to perform only a partial indexing of the users' documents (the process still hadn't been optimized). This was not a problem since it did not influence the way the interface works.

Ten users were interviewed, and 30 stories collected [13]. To evaluate if stories told using Quill can be similar to those told to humans (in short, it the storytelling illusion can be maintained by Quill), we compared those stories to those collected in the interviews in the first study described in this paper. This comparison was similar to that made for stories collected using the low-fidelity prototypes. We found that, both in terms of contents and structure, there are no major statistically significant differences. This leads us to conclude that users can tell their stories to Quill as if to a human listener. The interface does not unduly intrude in the storytelling process.

To verify to what extent are stories told using Quill truthful and accurate when describing documents, we compared the information in each story with actual data from the document described in the story and the users' system.

As the users were free to describe whatever documents they wished in their stories, we didn't know beforehand the "correct" information that might appear in the stories. Hence, we had to carefully assess the accuracy of each story element. We considered three accuracy levels. *Inaccurate* story elements were those that were undoubtedly incorrect. *Accurate Verified* were those we were able to verify from hard data in the users' computers (a file name or a keyword's occurrence in a document, for instance). Finally, *Accurate Unverified* were those that we were unable to confirm from hard data but were nevertheless convinced to be true. To attribute this classification to an element, the interviewer needed to be fully convinced, beyond any reasonable doubt, by questioning users, opening documents and email messages, etc.

We found that on average 91% of the information in a story is accurate (81%, considering only Accurate Verified elements). This corresponds to 1 to 3 inaccurate elements per story. It reinforces the Design Guideline 5, showing stories not to be 100% correct, but also indicates that the inaccuracy level is low, and can be easily handled by a properly designed system.

Discriminative Power of Stories and Retrieval Rate

It was necessary to verify if storytelling was versatile enough to help users retrieve their documents. We did not use traditional Information Retrieval metrics such as Precision and Recall because we are dealing with personal documents. As discussed at the beginning of this section, the study of narrative-based interfaces prevents us from using predetermined test sets, as it only makes sense for stories to be told about personal items that have been handled sometime in the past and about which meaningful stories can be told.

To measure Precision (how many of the documents are relevant) makes little sense. In the general case of web search, for instance, users are looking for different possible matches about a given subject. As such it is interesting to know how many suggestions were actually interesting. When looking for a personal document, there are no several potentially good results: the user is looking for a specific document, handled in the past. Likewise, to measure Recall (how many of the existing relevant documents were returned) would be very hard. We would need to know beforehand how many documents in a user's hard drive are relevant to a search, requiring require the researchers to know all documents in a user's computer to identify which would be relevant. As constructing test sets is impossible, this cannot be done (even if it was somehow practical).

For all the aforementioned reasons, to estimate the effectiveness of our narrative-based personal document retrieval interface, we chose to measure Quill's discriminative power and retrieval. To that end, we performed a second study, with a more refined version of the prototype, in which 21 users were interviewed and 63 stories collected and analyzed [15]. This time, not only the storytelling process was important, but also the actual results it could lead to. Hence, it was important to let Quill index the entirety of the users' documents and, whenever possible, other sources of autobiographic information such as their emails and agendas. This proved to be one large obstacle to the study's success, as the entire indexing process could take several hours. Whenever it seemed this might happen, the system was allowed to perform the indexing overnight and the rest of the interview took place the following day.

Using the KB Explorer, a special-purpose tool that allows us to manually inspect the index's contents, we were able to identify the documents that matched all information in each story. We found that, on average, *stories have a discriminative power of 2.51*. Furthermore, 55% of stories identify a single document, and only 4.7% describe five or more. Five is a number of documents whose thumbnails can be easily displayed at once in most screen resolutions, allowing users to scan them at a glance, without disrupting the storytelling process.

As for Quill's overall retrieval rate, we measured retrieval success for the different stories. Overall, 87.9% of all documents sought using Quill were retrieved. What is more, 95.2% of all text documents (office documents, plain text files, etc.) were found. The value drops to 68.8% for non-text documents (photos, music, videos, etc.). Considering

that no tags or other manually entered meta-data was present, that Quill is a general-purpose tool, and that it is only a prototype, this is a good result.

Of all documents, 8.6% were not found due to some mistake in the stories that Quill was unable to cope with, and 3.5% (two documents) were not found even when all information in the story was correct. A closer look reveals the reason for those failures. In one case, Quill had been unable to index the users' email, where the document resided as an attachment. In the other case, data that would have been collected had Quill been running for longer would have allowed it to find the document. In short, the failures were due not to any intrinsic fault with our approach, but by practical aspects of the study.

It is also interesting to note that 7.84% of documents found by Quill would not have been found by traditional, keyword-based, search tools. They were found with the help of other kinds of autobiographic information contained in the stories, reinforcing their importance over that of mere keywords.

DISCUSSION

As shown in the different user studies we just described, narratives can be successfully used as a way to help users recall important autobiographic information about their documents and convey that information to the computer.

Providing that the interface is properly built, stories told to a computer can be similar to those told to human listeners. Quill's evaluation showed this for to be true for document-describing stories. However, there is no reason for narrative-based interfaces not to be used in other domains. The main limitation is that there must be a story to tell. As such, domains in which users have only a passing knowledge of what they are looking for, or don't know exactly what is to be found (as is the general case of web search), are not amenable to this solution. Narrative-based interfaces would be applicable to domains such as the retrieval of certain types of documents (using stories specific to them), annotating personal items, managing personal objects in the real world (with the help of RFID technology, for instance), or even as a collaborative tool to manage project files.

Once an appropriate domain is identified, it will be useful to follow the steps described in this paper leading to Quill's creation. The several story elements and expected structure of stories must be identified, and the lessons learned then used to design the interface. There are some issues that, regardless of the application domain, should considered:

Establish dialogs with users: it is common for the users to digress when telling stories. This should be prevented by limiting what can be entered to reasonable information about the problem domain, taking care not to limit flexibility. Dialogs will also help users move forward.

Avoid the temptation of using a free-form natural-language based interface: it is very hard if not impossible to foresee the contents of stories, making it very difficult to make sense out of them. We've shown that structured storytelling, using the fill-in-the-blanks approach, can help users tell their stories as text, without the inherent disadvantages of having to write the entire narrative.

User knowledge about the world: in stories some information will undoubtedly assume the understanding of some common-sense world knowledge. A properly constructed knowledge base will help, by including knowledge about the user and problem domain.

Mine the users' environment for information: no one will annotate or describe all objects in the search domain in a consistent or effective way. A monitoring system that takes note of all relevant information should be used.

Evaluate what needs to be evaluated: it easy to try to perform user studies with predefined test sets, so that the resulting narratives can be matched against expected results. This is to be avoided, as it will taint the results.

Evidently, domain-specific characteristics might prompt particular adaptations. Ultimately, it should always be ensured that illusion of telling a story is not dispelled.

CONCLUSIONS

It is hard for users to manage their personal items, such as documents, email messages, or bookmarks, using today's tools. Hierarchies are the main approach to their organization, but are growing increasingly ineffective with the growing numbers of items that must e dealt with by users. Recent solutions, such as desktop search systems, try to help users by allowing them to efficiently search their hard-drives, instead of browsing the hierarchy. However, those systems are often based on keyword search, and do not take advantage of a wide range of autobiographic information that users remember about their items, stemming from past interactions with them.

Conveying contextual autobiographic information to the computer can be problematic. The interface should be designed in a way that allows that information to be used naturally and efficiently. We've shown that narratives can be the basis for such an interface.

With the help of a prototype system for the domain of personal document retrieval, Quill, we were able to evaluate the interface with user tests, carefully designed given the personal nature of narratives. The evaluation's results show that narrative-based interfaces allow users to naturally and efficiently tell their stories, that are trustworthy, accurate and discriminative, when describing documents. Furthermore, with the help of knowledge about the world and the user, the stories can be understood and used by the computer, as shown by the document retrieval rates that can be achieved using Quill. The same approach can be used for

other problems in Personal Information Management.

In the future, we plan to see how other interaction modalities (speech recognition, for instance), can improve the storytelling process. Also, we'll try to extend the approach to help users find objects in the real world. For them, as for electronic documents, stories can be (and often are) told. Another aspect that we plan to deal with in the future is the ability to allow users to state the certainty degree with which they know some information in a story.

ACKNOWLEDGMENTS

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

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. Bergman, O., Beyth-Marom, R. and Nachmias, R. The project fragmentation problem in personal information management. *In Proceedings CHI 2006*, pp. 271-274. Montreal, April 2006. ACM Press
3. Boardman, R. and Sasse, M.. "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*, pp. 583–590. ACM Press, New York, NY, USA.
4. Brown, D. E.. Human Universals. New York: McGraw-Hill 1991.
5. Dourish, P. et al. Extending Document Management Systems with User-Specific Active Properties. *ACM Transactions on Information Systems, 18(2)*, pp 140-170, ACM Press 2000.
6. Dumais, S. T. et al. Stuff I´ve Seen: A system for personal information retrieval and re-use. In *Proceedings of SIGIR 2003*.
7. Freeman, E. and Gelernter, D. Lifestreams: A Storage Model for Personal Data, *ACM SIGMOD Record,25(1), pp 80-86*, ACM Press 1996.
8. Gemmell, J., Bell, G. and Lueder, R.. Mylifebits: a personal database for everything. *Communications of the ACM*, 49(1):88–95, 2006.
9. Gifford, D., Jouvelot, P., Sheldon, M. and O'Toole, J. Semantic File Systems. *13th ACM Symposium on Principles of Programming Languages*, October 1991.
10. Golder, S. and Huberman, B. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2006.
11. Gonçalves, D. and Jorge, J. Telling Stories to Computers. *In Proceedings CHI2004*, ACM Press, 27-29 April 2004, Vienna, Austria.
12. Gonçalves, D. and Jorge, J., "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.
13. Gonçalves, D. and Jorge, J. Evaluating Stories in Narrative-Based Interfaces. In *Proceedings IUI 2006*. January 2006, Sydney, Australia. ACM Press
14. Gonçalves, D. and Jorge, J., Quill: A Narrative-Based Interface for Personal Document Retrieval. In Proceedings ALT.CHI2006. April 2006, Montreal, Canada. ACM Press.
15. Gonçalves, D. Evaluating the Discriminative Power and Retrieval Rate of Document-Describing Stories. *Technical Report*, *IST/UTL*. January 2007, (at http://immi.inesc.pt/~djvg/phd/files/final_evaluation.pdf)
16. Huberman, M. and Miles, M. Analyse des données qualitatives. Recueil de nouvelles méthodes. Bruxelles, De Boeck, 1991.
17. Karger, D. et al. Haystack: A platform for creating, organizing and visualizing information using RDF. In Proceedings *Semantic Web Workshop, The Eleventh World Wide Web Conference 2002 (WWW2002)*. 2002.
18. 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.
19. NLTK natural language toolkit, available online at http://nltk.sourceforge.net/
20. Porter, M. F. 1980. An algorithm for suffix stripping. *Program* 14, 130-137.
21. 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.
22. RDFLib homepage: http://rdflib.net/
23. Russel, S. and Norvig, P. Artificial Intelligence: A Modern Approach, 2nd. Ed. Prentice Hall, USA. 2003
24. Salton, G. Automatic text processing, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 1988.
25. Shamir, A.. A View on Views. In *Proceedings SmartGraphics04*, Banff Center, Canada, May 2004.
26. W3C Semantic Web, http://www.w3.org/2001/sw/
27. 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