VisMe: An Integrated Visualization of Heterogeneous Personal Information

Paulo Gomes, Sandra Gama, Daniel Gonçalves
Instituto Superior Técnico
Av. Rovisco Pais, 49
1050-001 Lisboa, Portugal
+351-213100289
{paulo.gomes, sandra.gama}@ist.utl.pt, daniel.goncalves@inesc-id.pt

ABSTRACT
Our personal information is dispersed throughout a multitude of devices and applications, making an overall integrated view difficult to achieve. This not only hinders retrieval and management tasks, but also introspection and the discovery of personally meaningful patterns. We present VisMe, a dynamic interactive personal information visualization tool that uniformly displays personal information from the heterogeneous sources that make up our digital identities. VisMe presents the information in interconnected timelines and lets the users progressively explore it while maintaining the overall context to help make sense of the results. We describe our solution and discuss the design decisions underlying its development. Usability tests, in which users performed typical tasks and were encouraged to find personally meaningful trends and patterns, confirmed the validity of this solution as a personal information visualization tool.

Author Keywords
Information Visualization, Personal Information Management, User-Centered Design.

ACM Classification Keywords
H.5.2 User Interfaces – Graphical User Interfaces; H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Given the pervasive nature of computers in our lives, it is likely that there is much to be learned about ourselves in the immense amounts of information that we continuously accrue. That analysis, however, is not facilitated by current tools. Our information is scattered around distinct document types, which in turn are commonly handled by different and unrelated applications, making it very hard to get a global view of our digital lives.

While the information is dispersed and heterogeneous in nature, it is often the case that it would be possible to interconnect different pieces of information. For instance, there is a conceptual link between a file in the filesystem and the email message it came attached to. Unfortunately, there is little or no support for this kind of association in current tools and operating systems. Thus, the different pieces of information that comprise our Personal Information Spaces remain unconnected.

All of these problems place on the shoulders of users the burden of understanding and managing their information in an integrated way. It should be possible to facilitate such tasks by providing users with a way to navigate their personal information spaces in a straightforward, unified, way. We propose that using information visualization techniques, where heterogeneous information is displayed in a homogeneous way, can help alleviate this problem. This will give the users the basis for gaining insights and perceiving patterns in their personal information (and, ultimately, their lives) that would otherwise remain hidden.

Our goal is to visualize the content of the personal information spaces that constitute a person’s digital presence; including text documents of all sorts, instant messaging conversations, and emails, in a way that allows users to identify personally relevant trends and patterns. This visualization should integrate different kinds of personal information and display them uniformly, to allow users to inspect their personal information in a unified way. It should be interactive, to facilitate exploration and, throughout this exploration, it should preserve the surrounding context, allowing users to interconnect their discoveries and make sense of all that information.

Additionally, it would be interesting if such a system could provide some means for finding and retrieving the documents themselves, supplementing existing browsing and searching tools, making use of the contextual exploration to facilitate those tasks. The ability to retrieve and inspect particular pieces of information will also be important as a way for the users to gain a deeper understanding of their personal information.

We studied the best ways to visualize personal information and how the interaction with that information could take place. Our approach led to the creation of a prototype system, VisMe, which presents information (the content of the documents in the form of keywords; the contacts responsible for creating, sending, and receiving those documents; and document names or titles) in timelines. Starting from the broad view of all the available
information, users can progressively narrow their search by opening new timelines from any element in the visualization.

This solution was continuously developed and tested over a year. This included two rounds of usability tests that confirmed the solution facilitates the discovery of personally relevant trends and patterns, as well as the retrieval of documents taking advantage of the visibility of their surrounding context.

In the following section, we will describe some works that have tried to address this problem in the past, in different ways. Next, we will describe our approach and the design rationale behind it. The different user studies where our ideas were validated with the help of a prototype system are described next, after which we conclude by presenting possible avenues for future work.

RELATED WORK
Once we acknowledge the vast stores of personal information kept by the average computer user, accessing and analyzing that information in search of meaningful insights becomes a logical pursuit. Indeed, several personal information visualization tools have been developed over the years, often focusing on particular sources or aiming at different specific goals, from a practical standpoint of document retrieval to a purely recreational exploration of one’s past. Considering the scope of this project, we are particularly interested in visualizations of document collections, rather than of individual documents, from the point of view of a single user when communication and authorship properties are involved.

Email visualizations abound, often attempting to aid personal information management rather than providing meaningful overviews of the email collections. TheMail [8] distinguishes itself from the remaining solutions with its temporal display of the content of an entire collection of emails that is both aesthetically pleasing and easy to read and comprehend, providing the ability to identify trends and patterns in email content.

When it comes to visualizing instant digital conversations, visualization research is tilted towards conversations between multiple contacts. Still, CrystalChat [7] displays an egocentric conversation space in a 3D structure, though its content representation, besides the textual display of the actual messages on demand, is limited to a peripheral mood indicator.

In terms of visualizing text document collections, ThemeRiver [4] displays a visual overview of the thematic content of a collection, allowing users to discern trends in its contents, although it is more limited in terms of searching and browsing than other document visualizations.

Finally, there are some systems which attempt to display and facilitate access to heterogeneous document collections. The Stuff I’ve Seen [2] system indexes information from multiple sources (emails, web pages, documents, media files, etc.) but it does not, by itself, provide a view of the content of such a collection beyond plain lists of results. Milestones in time [5] builds on top of that indexing system and goes beyond a familiar list display by coupling it with a landmarked timeline to provide a simple and appealing interface for multimedia history search and browsing. Feldspar [1] allows users to find personal information in their computers by interactively and incrementally specifying chains of associative queries. Its strength lies in the ability to efficiently find documents based on a combination of known facets. However, although it presents an aesthetically pleasing and direct graphical representation of the queries, it does not provide a visualization of the content of the entire information space under analysis. FacetMap [6] also ties the visual representation, facets as bubbles, with the underlying searching mechanism in a simple and effective way.

VisMe goes beyond these solutions by offering a unified visualization of the content of several personal information sources with which a person commonly interacts. It goes beyond lists of results by providing a graphically rich integrated interactive view allowing personally relevant patterns to be identified by users.

PROPOSED SOLUTION
In the context of this project, personal information can be understood to be contained in the documents one usually accesses and interacts with in regular computer use, in the wider sense of text documents in several different formats, websites, instant messaging logs, and emails. Attempting to extract and visualize information from all the aforementioned sources begets the issue of heterogeneity. At the same time, all sources contain common related properties that can be used to integrate them in a single, all encompassing visualization. Our solution for the interactive visualization of personal information tries to solve the problems and take advantage of the opportunities presented by the targeted sources.

Architecture
The information is indexed using an external indexing application, Scribe, which is capable of indexing and interconnecting emails, documents, instant messaging logs, web pages, etc. On top of this indexed data, a layer was developed to facilitate integration and to provide efficient access.

The interactive visualization interface was implemented in Python using Pyglet (OpenGL).
Visualization
The development of an information visualization tool must necessarily take into consideration the specificities of the information it is meant to visualize.

Facets
Time has been shown to be extremely important in terms of how people think about their personal information [3]. Additionally, all document types under consideration for this visualization have some sort of temporal presence. As such, time is a straightforward choice.

From the beginning, we identified the content of the textual documents as something we had strong interest in visualizing. As such, and again considering the content analysis of the document types we are working with, we chose to represent the most significant keywords extracted from each document according to their tf-idf weight.

It is also noticeable that many of these documents exist within a context of interpersonal communication; we decided it would be equally relevant to display contacts, the different people who create, send, or receive the documents that make up a personal collection.

Finally, we need to identify the documents themselves, so the name or title of each document is also integrated into the visualization.

Most of these facets can be reliably extracted from all documents, with a few exceptions: some documents may simply have no defined creator (though if they are attached to emails they can still be connected to the participants); others may come from the same contact, but under different aliases that the underlying indexing system is not able to connect. Still, there is generally still enough information available to create a useful visualization from these facets.

Organization
We have already acknowledged the importance of time, as well as its prevalence in the information we are contemplating. Thus, we have chosen to organize information using timelines.

Besides being applicable to all the chosen facets, timelines have been shown to be effective in presenting the content of document collections in a way that allows for the detection of personally relevant trends and patterns, as observed in the aforementioned Themail visualization.

The most significant elements (keywords, contacts, and documents) appear larger and at the bottom of the timeline (Figure 1), with less relevant elements appearing smaller and further up. Font size is kept between fixed minimum and maximum values to assure legibility. The granularity of the timeline can be controlled by clicking on the times at the bottom (to zoom in; for instance, from months to days) and bottom left (to zoom out).

Exploration
The information to be displayed in the timelines is not selected through a traditional process of search terms and filters. Rather, we allow users to observe timelines containing one of three facets (together with time) as they relate to any particular chosen element.

The visualization starts with a single element, “ego”, representing the owner of the information. From this element, users can expand timelines with keywords, contacts, and documents by clicking or dragging from the respective icons (“k”, “c”, and “d”) at the bottom of the word. The same action can then be applied to any element in any timeline (Figure 2). These timelines contain all the elements of a chosen facet that are related to the selected element. For instance, we can extract a timeline with the names of all the documents containing a particular keyword, followed by the contacts related to one of those documents, followed in turn by the contacts related to one of those contacts, etc. There is, potentially, no limit to this; users are allowed a progressive and free exploration of any element in the visualization.

View
Since we allow users to expand timelines sequentially from any element in the visualization, we considered implementing two alternative solutions: limiting the view to a single timeline at the time, with the option to go back and forth the browsing history, or giving the users control of the view and allowing them to expand interconnected timelines in an open canvas. The first option may be easier to control, but it limits the view of the surrounding context and discourages branching exploration. The second, while requiring greater skill to manage both the timeline structure as it evolves and the view to encompass it, has the benefit of allowing users to freely explore information while keeping the overall context, which they can then use to interrelate and interconnect the results of their ongoing research. Given our goals, we decided to develop the second option.
As they have control of the view, users can either click on one of the icons to open a perpendicular timeline, or drag and drop a timeline to any position and orientation, as indicated by a line that is drawn on the screen while the action is performed.

To control the view, users can use the mouse to pan (dragging the left mouse button over empty space), rotate (dragging the right mouse button), and zoom in and out (using the mouse wheel and, alternatively, the Page Up and Page Down keys) of the view to observe as much information as they want. Still, we consider the option to focus on a timeline important, so we allow that by double clicking on a time, which gradually and smoothly changes the view to encompass only the timeline in question. Double clicking on an empty space will change the view to encompass all timelines.

This automatic view control can be extended to become constant, meaning the view will automatically encompass the entire visualization structure after every action that modifies it. This option can be turned on and off at any time by clicking a small button on the top right corner of the screen.

Highlights
Since the shape of a word alone may not be enough for an optimal appreciation of the evolution of an element along the timeline, we added the possibility of highlighting any instance of an element throughout the visualization by placing the mouse over that element. This changes color from the regular black to one within a range from red to blue. The position of the mouse along the length of the element determines the exact color, sparing the need for any additional interface elements. A single left click fixes the color and a second left click undoes that action. Right clicking on a colored element copies the color, allowing its application on any element by left clicking on an uncolored element afterwards. Right clicking on an empty space will clear this selection.

Text Search
Finding particular keywords, contacts, or document titles in any significantly large collection of documents would be unfeasible without a straightforward search mechanism. Keeping with the minimalistic design of the interface, simply pressing a character on the keyboard prompts the display of a search input box on the top left corner of the screen containing the search string as it is written, as well an indication of the number of possible results. Pressing the tab key will complete the string to match the currently visible result, pressing the up and down keys will cycle through results. Clicking on the “k”, “c”, and “d” buttons, or on the right and left arrow keys, will cycle through the results.

Figure 2: New timelines can be expanded from any element.

Figure 3: Text search.
results for each of the three facets. If the current string appears anywhere on the expanded timelines, the respective elements will be highlighted. If there is an element that matches the search string but is not representative enough to be displayed on the timeline, it will appear on top the respective column. It will be significantly larger than the element before it, showing that the element does not follow the same size convention as the rest of the timeline. Also, whether or not there is an exact match in any time period, the respective time will be highlighted at the bottom of the timeline (Figure 3).

**Filters**
To give users control over how timelines are populated, beyond the selection of a single base element, they can drag any keyword, contact, and file name, into any timeline, as many times and in any combination, to filter it. For example, clicking and dragging the mouse from a keyword in one timeline to the space occupied by a second timeline will add the keyword as a filter to it. Active filters appear to the left of the timeline and a simple click will remove it (Figure 4). This has been extended to the search string that appears on the top left corner of the screen, which can also be dragged into any timeline, making it extremely easy to filter a timeline according to any existing facet that users find through textual search.

**Managing Clutter**
There is a limit to how much information can be displayed in the same screen before it becomes too cluttered. This has led us to test and implement a number of coping measures. When the mouse is not above a timeline, only the most representative elements at the bottom are shown as the rest gradually fades out. Also, timelines initially appear in a horizontal state, with the most representative elements from all time periods. As such, only relatively thin lines of important elements are seen most of the time. We also let users reposition or hide existing timelines (and, indeed, entire branches) they may consider less relevant. By moving the view and reorganizing information, the necessary relevant information can be kept in view. Besides manually controlling the positioning of the timelines, we have also implemented collision detection with gradual separation of overlapping timelines.

**Help**
At any time during the use of the application, users can press a button at the bottom left of the screen to activate the display of tooltips at the bottom of the screen. These tooltips show up when the mouse goes over any element in the visualization and contain concise instructions related to their use.

**WHY TWO USER EVALUATIONS?**
Our prototype solution was subjected to a heuristic evaluation with four usability experts, which allowed us to detect and correct several usability issues. Once they were solved, we moved on to user tests.

First and foremost, we wanted to verify if our design choices made sense and were understood by users. To do so, we needed to ask them to perform some tasks that would be typical for a tool such as ours, pertaining document retrieval and pattern discovery. However, since we’re dealing with personal information, it would be impossible to have an a priori knowledge of each user’s personal information space. Thus, it would be impossible to ask for very concrete tasks (“find document X”). More generic tasks would be possible, but this would make the comparison of results from different users hard to do, as they would be looking at different data sets from which different conclusions could be reached.

For these motives, we performed two rounds of user evaluation. In the first, we gave all the users the same information to explore. While we took care to simulate the contents of an average user, no deep insights could be produced, as there is no intimate relationship between the users and it. On the other hand, all users were equally (un)familiarized with the data set, providing us a baseline for comparison. This first test showed us that the user interface design is sound, understandable and can be used effectively by users. Then in a second moment, we conducted a study where the users’ actual personal information was indexed and used. This second study showed us how our approach, already validated in general, can actually be used with real data. It can cope with its inherent noise and scruffiness and still be used to help users in retrieval and pattern discovery tasks.

**FIRST USER EVALUATION**
In the first attempt to validate our solution we asked 20 volunteers, aged 17 to 29 (\(\bar{x} = 23.7, \sigma = 2.7\), mode = 24) and with self reported high level of experience with computers (\(\bar{x} = 3.7, \sigma = 4, \sigma = 0.5\), in a four point scale), to perform a series of tasks using the prototype over a set of 1004 text documents authored by 102 people. This data set was crafted, based on our knowledge of previous experiments, to be representative of a real, albeit small, collection of documents, with realistic trends and patterns for each tested combination of facets and enough documents and authors to make it hard to just stumble upon them without significant help from the interface. Having full control of the data set allowed us to craft more
complete and specifically targeted tasks, to validate our interface design decisions. A second evaluation, described later, was performed to validate the solution using actual personal information from the users.

Eight of the tasks consisted of finding a document based on the knowledge of one or of a combination of its facets (time, most representative keyword, and author, as well as a single task in which the actual file name was given). For instance, "Find a document written by Bob about the Internet that you read in May of 2009" or just "Find a document about guitars".

The remaining nine tasks required users to determine the most significant keyword, contact, and a combination of the two in a time period (and the other way around) as well as its evolution throughout that period (when it started and stopped appearing, and whether it evolved to become more or less important, or stayed the same). For example, "Who did you contact the most in January 2010" and "How did that contact evolve during that year".

These tasks were timed and recorded for later analysis. Users were also asked to grade the difficulty of each task with a four point scale upon completion. The time limit for each task was 150 seconds, after which users would be told to move on to the next task. The order in which tasks were performed was random, in order to prevent result biases.

Users were first given a 5 minute demonstration of the prototype and all its features followed by 5 minutes of free experimentation with the interface. They also answered a questionnaire at the end of the session to determine their satisfaction with the interface using a four point scale, with 4 being the least difficult (95% confidence interval).

Results
We recorded success, completion time, and difficulty for each task (as rated by the users), as well as the answers to the questionnaire, and later performed a statistical analysis of that data. Bellow, we present the major findings of that analysis relative to three main groups: document retrieval, pattern recognition, and satisfaction. These results are shown through averages ($\bar{x}$), medians ($\tilde{x}$), standard deviations ($\sigma$), confidence intervals (CI, 95%), and, in the difficulty measurements and in the questionnaires, modes.

**Document Retrieval**
Finding documents based on a single facet (figure 5, tasks 1.1 to 1.4) or on a combination of one facet with time (figure 5, tasks 1.5 and 1.6) were generally performed successfully (2 failures in 20 sections for task 1.5, finding documents based on a keyword and a date; no failures for the remaining tasks) in under a minute, and they were considered very easy (figure 6, tasks 1.1 to 1.6).

Two tasks (tasks 1.7 and 1.8), in which the users had to find a document based on both a keyword and an author (and, in one, also time) as opposed to doing the same based only on a keyword or author (with and without time), stand out by having the greatest number of failures (6 and 5 in 20 sessions, respectively), the highest average completion times ($\bar{x} = 84.5, \tilde{x} = 77, \sigma = 27.5$; and $\bar{x} = 88.7, \tilde{x} = 84, \sigma = 27.6$) and being considered the most difficult tasks ($\bar{x} = 2.6, \tilde{x} = 3, \sigma = 1.4, \text{mode} = 3$; and $\bar{x} = 2.6, \tilde{x} = 2.5, \sigma = 0.9, \text{mode} = 2$). To complete these two tasks, users had to either expand documents from two facets and cross check the results, or expand documents from one facet and expand the other facet out of each document one by one. Although that is not overly complicated, especially given the fact that the number of documents that matched one of the facets (or one of the facets plus time) ranged from only two to a dozen, it was not straightforward for many users. This strengthened the need for a more direct method of combining facets, which we attempted to provide with the addition of the filtering mechanism described earlier.

**Pattern Recognition**
Trend and pattern detection tasks were generally completed successfully (4 out of 9 tasks with 1 failure in 20 sections each) in under a minute (figure 5, tasks 2.1 to 2.9). They were also considered very easy (figure 6, tasks 2.1 to 2.9).

Tasks 2.1 to 2.4, besides asking users to find the most relevant keyword or contact in a year, or the year in which a
keyword or contact were the most relevant, demanded a description of the evolution of that element over the respective year, while the remaining tasks asked only for the identification of the most relevant facets in different combinations without the need to describe their evolution. These four tasks present similar times and difficulties, evidencing no statistically significant difference between them. Among the other subgroups of directly comparable tasks (2.5 and 2.6; 2.7 and 2.8) there is also no statistically relevant difference in times and difficulties. This gives us an indication that the unified representation of different facets in timelines allows users to identify the most significant elements in a time period and their evolution over time with uniform ease.

**Satisfaction**
The questionnaire shows that users were generally satisfied with the system ($\bar{x} = 3.4$, $\tilde{x} = 3$, $\sigma = 0.5$, mode 3, four point scale). For the most part, they did not find it difficult to use ($\bar{x} = 3.2$, $\tilde{x} = 3$, $\sigma = 0.7$, mode = 3), but a few did find it somewhat difficult to learn ($\bar{x} = 2.8$, $\tilde{x} = 3$, $\sigma = 0.9$, mode 3).

**Discussion**
Using the VisMe prototype, users were capable of identifying and following the relative importance of keywords and contacts over time. They were also capable searching for documents with the initial knowledge of different facets. While these results do not fully validate our solution, they show that interface is suited for the performance of the essential pattern detection and document retrieval operations we aimed to facilitate. Also, understanding what worked and what didn’t work guided the development of the solution and its prototype from this point onwards.

**SECOND USER EVALUATION**
Following the conclusion of the final VisMe prototype, we conducted a second set of usability tests. Having validated the main ideas behind the solution using an artificial collection of documents, this final evaluation is expected to not only test the added functionalities and modifications but also to more fully validate the entire solution by testing it over real user data. This will determine if the same interaction and visualization techniques that were shown to work with a controlled data set, are also effective with the added size and indexing noise of real user data. It is also crucial to evaluate the applicability of these techniques when it comes to discovering personally interesting patterns in the visualized information.

**Method**
We asked 10 volunteers, aged 17 to 53 ($\bar{x} = 25.6$, $\tilde{x} = 24$, $\sigma = 10.2$, mode = 24) and with a high level of self-evaluated experience with computers ($\bar{x} = 4.3$, $\tilde{x} = 4.5$, $\sigma = 0.8$, mode = 5, five point scale), to index their email accounts and their document folders prior to executing a series of tasks designed to test all the features of the VisMe prototype and also to validate its ability to cope with actual personal information and provide personally relevant results.

The tasks can be split into two main groups: one with three subgroups of simple pattern detection and document retrieval tasks; and the other with a single large task of free exploration. They were necessarily more generic than those of the previous study as it was impossible to know beforehand what would be found in the users’ own information.

Two pattern detection and document retrieval tasks required users to alternately identify the most important keyword and contact in a given year and how it evolved over the months, as well as to retrieve the documents related to that keyword and the ones related to that contact. The third task asked users to find the most important keyword related to the contact found in one of the previous tasks before requesting users to find a document related to both those facets. Each individual task had a time limit of 150 seconds, after which users classified its difficulty using a five point Likert scale.

The exploration task, with a 15-minute time limit, required users to explore their personal information freely in search of interesting trends and patterns. As they did this, they were instructed to vocally communicate their thought processes and their findings to the accompanying examiner.

The first two tasks were randomized, with half the testers performing the task pertaining to a contact first, and the second doing the same with a keyword. Otherwise, the tasks were performed in the order in which they were described here.

Before performing the tasks, each user was given a 5 minute demonstration of the prototype, followed by a 5 minute period of acclimatization to the software. This was done with a separate collection of documents.

After the tasks, the users answered a questionnaire with a series of statements about general and specific aspects of the VisMe interface to be rated according to a five point Likert scale.

**Results**
As in the previous evaluation, we recorded success, completion time, and difficulty for each task (as rated by the users), as well as the answers to the questionnaire, and later performed a statistical analysis of that data. Below, we present the major findings of that analysis relative to four main groups: document retrieval, pattern recognition, exploration, and satisfaction.

**Document Retrieval**
Obtaining the documents related to a keyword (task 1.3, figure 7, $\bar{x} = 7.3$, $\tilde{x} = 5.5$, $\sigma = 5.3$) and to a contact (task 2.3, figure 7, $\bar{x} = 8.5$, $\tilde{x} = 6$, $\sigma = 6.3$) are fast operations. These were also considered very easy tasks (task 1.3, figure 8, $\bar{x} = 5$, $\tilde{x} = 5$, $\sigma = 0$, mode = 5, and task 2.3, figure 8, $\bar{x} = 4.9$, $\tilde{x} = 4.9$, $\sigma = 0$, mode = 5, and task 2.3, figure 8, $\bar{x} = 4.9$, $\tilde{x} = 4.9$, $\sigma = 0$, mode = 5, and
5, $\sigma = 0.3$, mode = 5). This is true regardless of the type of facet, as the interface allows users to directly expand timelines filled with the documents from any keyword or contact in the visualization in a uniform way.

These times reflect only the act of obtaining the documents; they do not include the initial search for the desired facet as it was required in the more complete document retrieval tasks of the first evaluation.

We also tested the filtering mechanism through which users can combine several facets in the same timeline, particularly its application to document retrieval. Users were generally capable of performing this task successfully in under 20 seconds (task 3.1, figure 7, $\bar{x} = 25.1$, $\bar{\sigma} = 6.5$). Some users found it difficult and this was considered more complicated than directly retrieving documents from a single facet, but still easy (task 3.1, figure 8, $\bar{x} = 3.8$, $\bar{\sigma} = 5$, mode = 5). This means that the combination of a keyword and a contact to obtain a document is a feasible and not overly lengthy or complicated operation.

**Pattern Recognition**

The ability to perceive the relative importance of keywords and contacts, and to follow their evolution over time is fundamental to being able recognize trends and patterns.

Identifying the most relevant keyword and contact was found to be fast (task 1.1, figure 7, $\bar{x} = 16.3$, $\bar{\sigma} = 7.2$; and task 2.1, figure 7, $\bar{x} = 16.2$, $\bar{\sigma} = 7.8$) and easy (task 1.1, figure 8, $\bar{x} = 5$, $\bar{\sigma} = 0$, mode = 5; and task 2.1, figure 8, $\bar{x} = 4.7$, $\bar{\sigma} = 0.5$, mode = 5).

Describing the evolution of these elements over time takes longer (task 1.2, figure 7, $\bar{x} = 36.3$, $\bar{\sigma} = 27$, mode = 5; and task 2.2, figure 7, $\bar{x} = 27.9$, $\bar{\sigma} = 6.3$) and is slightly more complicated (task 1.2, figure 8, $\bar{x} = 4.5$, $\bar{\sigma} = 0.7$, mode = 5; and task 2.2, figure 8, $\bar{x} = 4.3$, $\bar{\sigma} = 1.3$, mode = 5). Task 1.2, concerning the description of the evolution of a keyword over time, has a larger standard deviation than its contact related counterpart. Removing one clear outlier in this task’s recorded times, however, lowers its standard deviation to a considerably closer value ($\sigma = 8.3$).

**Extended Exploration**

The extended and unrestricted exploration task is more subjective. Users were told to report any interesting findings and the level of interest for similar discoveries naturally varies from person to person. We also need to take into account that people may not have been comfortable exposing certain parts of their personal information, no matter how relevant to the study. It is, however, possible to gain valuable insights about the possibility of finding personally relevant trends and patterns with VisMe, along with a better understanding of how people use the interface, by observing the task in progress and by analyzing what information the users were able to gather.

All users were capable of identifying something interesting within their personal information, from singular events to more spread out communication patterns.

Two distinct users noted similar stories featuring the sudden appearance of a contact, followed by a strong presence for a few months, and ending with an abrupt disappearance, after which the person is never seen again.

One user expressed his surprise and frustration over the fact that most of his top keywords were coming from a few websites that sent him periodic automatic emails. He did not consider these to be spam, but he did admit he did not really care for their specific content. He also noted that it would be interesting to be able to filter out these emails, something which ideally should be possible with the current interface, were it not for the lack of exclusive filtering, a clear flaw that we will discuss in a latter section.

Six users described the visible appearance and ensuing constant presence of certain friends or colleagues. This was met with no surprise from the users, with the exception of one user who was, on one occasion, led to remember the few times when emails were sent to a particular contact, a friend with whom this person did not usually communicate online.

**Satisfaction**

Users were generally satisfied with the system ($\bar{x} = 4.5$, $\bar{\sigma} = 5$, mode = 5, five point scale); they considered it
aesthetically pleasing ($\bar{x} = 4.9, \bar{x} = 5, \sigma = 0.3, \text{mode} = 5$); and somewhat easy to use ($\bar{x} = 3.8, \bar{x} = 4, \sigma = 0.9, \text{mode} = 4$), learn ($\bar{x} = 3.9, \bar{x} = 4, \sigma = 0.7, \text{mode} = 4$), and understand ($\bar{x} = 3.6, \bar{x} = 4, \sigma = 0.8, \text{mode} = 4$).

In terms of specific functionalities, the constant automatic view control was not considered very useful ($\bar{x} = 3.7, \bar{x} = 3.5, \sigma = 0.8, \text{mode} = 3$), in contrast with the occasional automatic view control which was considered very useful ($\bar{x} = 4.7, \bar{x} = 5, \sigma = 0.5, \text{mode} = 5$). This matches our observations, as users rarely turned on the option to maintain constant automatic view control, but many made ongoing use of the second.

Manual view control was considered easy, but some users did find it complicated ($\bar{x} = 3.9, \bar{x} = 4, \sigma = 1, \text{mode} = 4$), again matching our observations. Though some users were not completely comfortable with the method of interaction, they still used it successfully without finding it obstructive.

The choice of facets was considered adequate ($\bar{x} = 4.9, \bar{x} = 5, \sigma = 0.3, \text{mode} = 5$), although users may not have been aware of the implications of the different possibilities.

**DISCUSSION**

On the whole, taking into consideration the results of both evaluations, the system behaved as expected and provided the results we worked for.

Trend and pattern recognition was shown to be facilitated by the visualization, as users could successfully detect the relative importance of a variety of contacts and topics as well as their evolution over time.

We verified that document retrieval is also well supported, as users have the ability to obtain timelines with the documents related to any element they find throughout their research. The filtering system also allowed users to combine multiple facets in the same timeline with relative ease to find the documents related to those facets. This functionality also has implications in pattern recognition (although that was not tested directly), and it was on the extended exploration task that a flaw in this functionality was made evident. Although it is easy to obtain a timeline with the elements related to an undetermined number and combination of different facets, it is currently impossible to do filter out the elements according to selected filters. The solution does not require an extensive reworking of the interface or of the underlying document access layer. To maintain the simplicity, dragging and dropping elements into a timeline will create inclusive filters, while clicking on an existing filter will toggle it from inclusive to exclusive and vice versa. Exclusive filters can be identified by a black background and white text, in the same manner that textual search results are highlighted. To remove a filter, users will have to double click over it.

Users were also capable of identifying trends and patterns of personal interest by observing the evolving relative importance of different keywords and contacts.

Overall, three things were determining factors for this approach’s success, and might be taken into consideration when developing similar solutions in the future. Firstly, the minimal, aesthetic interface, stripped of all extraneous, distracting, elements, allowed users to focus on their information. This information is complex and exists in large quantities, so the more space is occupied by it the better. Secondly, the ability to continuously explore the personal information space placed the users at ease and let them “go with the flow”, facilitating the exploration of an otherwise complex data set. Finally, integrating all information regardless of its original nature was, as we expected, of capital importance.

**CONCLUSION**

There is much to be gained from a competent visualization of all the personal information stored in our computers, from simple reminiscing about nearly forgotten moments, to efficiently locating lost documents based on their content and surrounding context. But existing solutions do not provide an integrated visualization of the content of the heterogeneous collection of documents that define that personal information. In an attempt to discover the best way to achieve such a solution, we developed a tool, VisMe, in which the content of those documents is extracted into representative keywords which, together with the connected contacts and document titles, are displayed in uniform timelines from which more timelines, pertaining to the facets one wishes to explore, can be extracted.

Usability tests validated VisMe’s approach to the problem, providing evidence for its ability to facilitate finding and retrieving documents based on the knowledge of different associated facets, detecting the most relevant keywords and contacts together with their evolution over time, and identifying personally relevant trends and patterns in the users’ own personal information. Displaying information in timelines makes their relative importance and how it changes over time clearly visible, and opening, keeping, and manipulating several timelines in the same canvas allows users to freely investigate several avenues while having access to the branching results of their explorations to make sense of all their personal information.

Still, the solution is not final. Besides augmenting the filtering mechanism by adding exclusive filters, there are some underlying technical issues which can still be worked on to improve the implementation and may require alternative solutions for visualization and interaction.

One is the direct access to the documents found through the visualization. The indexing application keeps a path of the individual files, but accessing the full email messages would require some alterations and it would leave us with an interesting choice regarding how to display those emails: either in the application itself or in an email application specified by the user. The former option would drag with it the problem of being a simple preview without the features one expects when browsing emails, unless we add
significant complexity to its interface, while the latter can deégrade the user experience if it is not handled correctly, since it could potentially require an erratic configuration and not very seamless transitions to other applications. Further research will be necessary to determine the best solution.

Also concerning the indexing application, its range is still limited, lacking proper indexation of instant messaging logs, website, and, most notably in this day and age, of social networking sites. Still, the information coming from those sources could be formatted in terms of keywords and egocentric communication networks, meaning the system is ready to accommodate it. This does not exclude potential improvements to take full advantage of all the particular features of each new source, as long as the coherence is preserved and the visualization is kept uniform.

Besides a more comprehensive indexation of personal electronic documents, we can consider the possibility that different sources of data, not necessarily personal, can also be visualized using the same basic principles of the visualization and interaction techniques developed for VisMe. As an example, we can imagine exploring the results of typical web search engines by iteratively dragging searches out of any visible word, starting with an initial search term and eventually developing into a rich, branching exploration.

ACKNOWLEDGMENTS
We would like to thank all the participants in our studies, particularly the ones involved in our last usability tests for suffering through the indexing process and allowing us to take a look at their personal information.

REFERENCES