

# Designing a Personal Information Visualization Tool

**Paulo Gomes**

Instituto Superior Técnico  
Av. Rovisco Pais, 49  
1050-001 Lisboa, Portugal  
+351-213100289  
paulo.gomes@ist.utl.pt

**Sandra Gama**

Instituto Superior Técnico  
Av. Rovisco Pais, 49  
1050-001 Lisboa, Portugal  
+351-213100289  
sandra.gama@ist.utl.pt

**Daniel Gonçalves**

Instituto Superior Técnico  
Av. Rovisco Pais, 49  
1050-001 Lisboa, Portugal  
+351-213100289  
daniel.goncalves@inesc-id.pt

## ABSTRACT

Interacting with computers has become part of our daily lives. This interaction results in large amounts of personal information, spread throughout places and applications. As a consequence, it is quite difficult to get an overall view of all our information or to find a specific item we are looking for. A meaningful visualization technique may be the solution to this problem. We present VisMe, an interactive visualization tool that allows users to explore personal information. It integrates and uniformly displays relevant concepts in interconnected timelines. Each of these items (people, subjects and documents) can be progressively explored, creating new timelines, where several avenues can be simultaneously explored in context. VisMe allows relations between concepts to be explored in a straightforward way. By providing the means to interactively find relations between different kinds of information in order to retrieve personally relevant data, VisMe gives users insight into their digital selves that current tools have a hard time providing.

## 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

An increasing number of devices, such as laptop computers and smartphones, have pervaded our daily lives, providing us with the means to generate large amounts of personal information, from the documents we write to the emails we exchange. This data can help us understand who we are,

what we do, and what we are interested in. "What was I doing in January 2005? With whom did I ever discuss classical music? John sent me a paper about a really vague subject when I was working on my thesis, where can I find it now?" Despite these being questions that an effective personal information analysis should answer, today's tools do not facilitate that task. In fact storing capacity has considerably grown over the last years, but methods and applications for managing and retrieving personal information have not suffered substantial improvements. Although hierarchical organization is widely used as an organizational system for electronic information, it has several shortcomings. It requires permanent effort from users to classify every piece of data, which is hard to achieve, and different kinds of information are managed by different applications, with little or no links among them. Personal information is, thus, scattered and difficult to find.

Several information visualization applications have been developed, often focusing on a single information source, such as email or text documents. Themail [7] stands out from other email visualizations with its simple and attractive interface and the ability to display patterns in email content. CrystalChat [6] displays an egocentric conversation space in a 3D structure, but its content representation, besides the textual display of the actual messages on demand, is limited to a peripheral mood indicator. ThemeRiver [3] displays a visual overview of the thematic content of a collection of documents, allowing users to discern trends in the content of document collections, but it is much more limited in terms of searching and browsing than other document visualizations.

Some systems combine multiple sources of information. However, they are mainly focused on searching and browsing. Still, Milestones in time [4] takes a familiar list display and couples it with a landmarked timeline to provide a simple and appealing interface for multimedia history search and browsing. FacetMap [5] goes beyond simple lists and joins a visual representation with the underlying searching mechanism in a relatively effective way using facet bubbles. Feldspar [1] allows users to interactively and incrementally construct association queries to find things about the individuals that wouldn't be found if searching items separately, by focusing more on connections between entities than on entities themselves.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

NordiCHI 2010, October 16–20, 2010, Reykjavik, Iceland.  
Copyright 2010 ACM ISBN: 978-1-60558-934-3...\$5.00.

But none of these visualizations provides a unified content overview of a heterogeneous collection. That is the void VisMe attempts to fill: an interactive visualization of personal information that displays relevant content patterns, allowing micro-data (individual documents, emails, etc.) to be retrieved in context. As such, all relevant personal information (documents, emails, instant messages, etc.) is indexed as a whole. Links lost by applications, such as between a document in the file system and the email it was attached to, are recreated. Our visualization is centered on relevant autobiographic clues and allows the exploration of personal information in an efficient, understandable way. It abstracts from data sources to present semantically relevant information instead, allowing several avenues to be explored simultaneously. By providing a synergistic visualization, VisMe allows users to efficiently navigate information and find personally relevant patterns.

### PROPOSED SOLUTION

The development of this visualization technique was carried out through iterative sketching and prototyping, keeping in mind three main goals for the intended interface: a uniform representation of heterogeneous information; a simple and uncluttered interface; and the possibility of exploring information in context, as opposed to displaying individual queries by themselves. This posed a series of challenges and conditioned the development of our solution, as explained below.

### Indexation

A solution to the problem of personal information visualization first requires a way to gather and index that information. An automatic indexing application, Scribe, which is not the focus of this paper, is used for that purpose. It is capable of indexing and interconnecting emails, documents, instant messaging logs, web pages, etc. Above this indexed data, a layer was developed to facilitate integration and to provide efficient access.

### Facets

There are a number of document properties that people consider relevant [2]. Of those, we selected four that can be indexed automatically and that fit our particular needs. Our first requirement was to represent document content in a meaningful way, so we chose to represent the most significant words extracted from each document according to their tf-idf weight. Time is very significant to users for organizing and remembering personal information, so it was also identified as a necessity from the start. We also needed to represent the documents themselves, so we use the file names or the titles. Finally, considering that much of the personal information we handle in our computers is a product of communicating with other people, we chose to represent contacts, people that send, receive, or create documents.

### Organization

We experimented with a number of different organization methods but ultimately settled on a temporal organization in the form of simple timelines, since time is a facet of great relevance to people and timelines can be applied uniformly to the other three facets we handle.



Figure 1: Timeline.

The most significant elements (keywords, contacts, or documents) appear larger and at the bottom of the timeline (Figure 1). 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 and bottom left.

### Progressive Exploration

The visualization starts with a single element, the “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). For instance, we can extract a timeline with all the documents containing a particular keyword, then the contacts related to one of those documents, and then the contacts related to one of those contacts, etc. Users are allowed a progressive and free exploration of any element in the visualization.

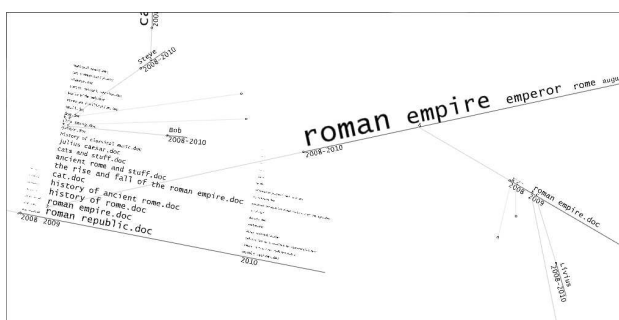


Figure 2: Timelines expanded from several different elements.

### View Control

The ability to expand timelines from any element in the visualization left us with two possibilities regarding the view: focusing on a single timeline at any time, which is simpler to control but makes the overall context difficult to retrieve; or viewing all timelines as a connected structure, at

the expense of a more complicated control mechanism. We experimented with both options and finally settled on the second. Users can use the mouse to pan, rotate, and zoom in and out 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 users to that by double clicking on a time, which gradually and smoothly changes the view to encompass only the timeline in question.

**Highlights**

The shape of a word alone is not enough for an optimal appreciation of the evolution of an element along the timeline. Therefore, we added the possibility of highlighting any instance of an element throughout the visualization. By placing the mouse over an element, it 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 click fixes the color while a second click undoes that action.

**Text search**

As we prepared the usability tests, it became apparent that finding particular keywords, contacts, or document titles in any significantly large collection of documents would be unfeasible without a more 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 (Figure 3) containing the search string as it is written, as well an indication of the number of possible results and one possible result in grey. Pressing the tab key will complete the string to match the currently visible result, pressing the up and down keys will cycle through results, left and right switches from keywords, contacts, and documents. 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 larger than the element below it, showing that it 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.

**Clutter**

There is a limit to how much information can be displayed in the same screen before it becomes too cluttered, and that is a problem that has led us to test and implement a number of measures. When the mouse is not above a timeline, only the most representative elements at the bottom are shown. 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 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 implemented collision detection with gradual separation of overlapping timelines.

**EVALUATION**

To validate our solution we asked 20 volunteers, aged 17 to 29 ( $\bar{x} = 23.7$ ,  $\tilde{x} = 23.5$ ,  $\sigma = 2.73$ ) and with self reported high level of experience with computers ( $\bar{x} = 3.7$ ,  $\tilde{x} = 4$ ,  $\sigma = 0.47$ ), 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 to be representative of a real 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.

Eight of these 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, when it ended, and whether it went up, down, 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 150s, 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 five minute demonstration of the prototype and all its features followed by five 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.

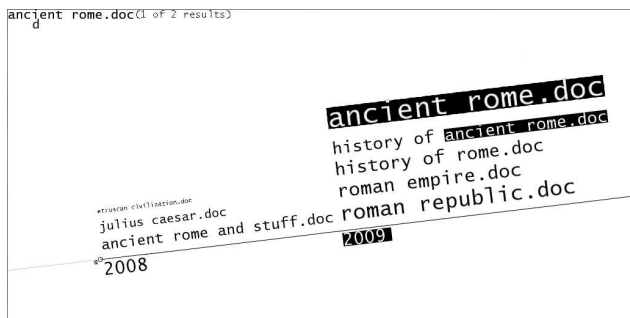


Figure 3: Text search.

## Results

The information gathered throughout the user tests was subjected to a statistical analysis, which we present through averages ( $\bar{x}$ ), standard deviations ( $\sigma$ ), and medians ( $\tilde{x}$ ).

Most document retrieval tasks were completed successfully (failures per task in 20 sessions:  $\bar{x} = 2$ ,  $\tilde{x} = 1$ ,  $\sigma = 2.4$ ). On average, successful document search tasks were completed in about 52s ( $\bar{x} = 52.2$ ,  $\tilde{x} = 45.5$ ,  $\sigma = 27.1$ ). Users also considered the tasks to be easy ( $\bar{x} = 3.4$ ,  $\tilde{x} = 4$ ,  $\sigma = 0.9$ ). This provides evidence that VisMe can be an effective retrieval tool for documents based on time, keywords, and authors.

Two tasks, 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), 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 were considered to be the most difficult tasks ( $\bar{x} = 2.6$ ,  $\tilde{x} = 3$ ,  $\sigma = 1.4$  and  $\bar{x} = 2.6$ ,  $\tilde{x} = 2.5$ ,  $\sigma = 0.9$ ). 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.

Trend and pattern detection tasks were generally completed successfully (failures per task in 20 sessions:  $\bar{x} = 0.4$ ,  $\tilde{x} = 0$ ,  $\sigma = 0.5$ ) in about 35s ( $\bar{x} = 35.5$ ,  $\tilde{x} = 30$ ,  $\sigma = 18.99$ ). No single task stood out in terms of failures, time, or difficulty, which 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.

The questionnaire shows that users were generally satisfied with the system ( $\bar{x} = 3.35$ ,  $\tilde{x} = 3$ ,  $\sigma = 0.49$ ). They did not find it difficult to use for the most part ( $\bar{x} = 3.20$ ,  $\tilde{x} = 3$ ,  $\sigma = 0.69$ ), but a few did find it somewhat difficult to learn ( $\bar{x} = 2.75$ ,  $\tilde{x} = 3$ ,  $\sigma = 0.85$ ).

## CONCLUSION

Visualizing the extensive personal information we generate throughout our daily use of computers could provide us with valuable insights and even help us find documents. But an interactive visualization tool that uniformly presents the overall contents of such a heterogeneous collection of documents does not currently exist. We presented our solution, VisMe, an interactive personal information visualization system. VisMe allows users to progressively explore their personal information by observing keywords, contacts, and documents in coherent and easily intelligible timelines, maintaining as much of the surrounding context

as they need.

Usability tests validated VisMe's ability to display the most relevant facets of both entire collection of documents and single documents, as well as its capabilities to facilitate document search and retrieval, although it exposed the lack of a mechanism to combine several facets in a single timeline for a more direct search. We have since developed a solution that allows users to simply drag any keyword, contact, and file name, into any timeline, as many times and in any combination they want, to filter it. For instance, clicking and dragging the mouse from a keyword in one timeline to a second timeline will add the keyword as a filter to it in a way that shows only the elements which are related to the chosen keyword. Active filters appear to the left of the timeline and can be removed with a single click.

It is also clear that testing the interface with an artificial collection of documents does not provide a definitive validation of our solution. Still, the goal of this evaluation was to obtain an initial validation of the main ideas behind our solution and we are preparing usability tests in which we will use the personal information of each tested user.

## REFERENCES

1. D. H. Chau, B. Myers, and A. Faulring. Feldspar: A system for finding information by association. In ACM SIGCHI PIM2008, the Third International Workshop on Personal Information Management, Florence, Italy, 2008.
2. D. Gonçalves and J. Jorge, Tell Me a Story: Issues on the Design of Document Retrieval Systems. In Proceedings DSV-IS04, Lecture Notes on Computer Science, Springer-Verlag, July 2004, Hamburg, Germany
3. S. Havre, B. Hetzler, L. Nowell, ThemeRiver: Visualizing Theme Changes over Time, Proc. IEEE Symposium on Information Visualization (2000), pp. 115-123.
4. M. Ringel, E. Cutrell, S. T. Dumais, and E. Horvitz. Milestones in time: The value of landmarks in retrieving information from personal stores. In INTERACT, 2003.
5. G. Smith, M. Czerwinski, B. Meyers, D. Robbins, G. Robertson, and D. S. Tan. Facetmap: A scalable search and browse visualization. IEEE Transactions on Visualization and Computer Graphics, 12(5):797-804, 2006.
6. A. Tat and S. Carpendale. Crystalchat: Visualizing personal chat history. Hawaii International Conference on System Sciences, 3:58c, 2006.
7. F. Viégas, S. Golder, and J. Donath. Visualizing email content: portraying relationships from conversational histories. In CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, pages 979-988, New York, NY, USA, 2006. ACM.