Visualiz'em: "Show me more about him!"

João Guerreiro, Daniel Gonçalves
Technical University of Lisbon / INESC-ID
Rua Alves Redol, 9. 1000-029, Lisboa, Portugal.
joao.p.guerreiro@ist.utl.pt, daniel.goncalves@inesc-id.pt

ABSTRACT
Our Personal Information is now scattered among several applications and personal devices in an unrelated, but still connected, information network. We have never had so much information at our disposal as we have now, giving us the opportunities to find information about other people, a very frequent need. While previous works have proposed several solutions to relate data from multiple sources that does not happen at the presentation level. When needing information about people, we still have to navigate among sources to find it. Besides troublesome, it is difficult to have an overall characterization of what each person represents to us. We argue that gathering information from all these sources and presenting it in a unified interface benefits the user by providing a quick and meaningful overview of who that person is and how he/she is related to him/her. We present *Visualiz'em*, a personal visualization tool based on three interconnected views: *Profile, Tagcloud* and *Timeline*. When compared to traditional applications *Visualiz'em* provides a faster and richer overview of whom a person is and his/her relationship with the user. Moreover, results show that our visualization tool promotes serendipitous behaviors, allowing users to easily explore data and find interaction patterns.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *graphical user interfaces, screen design, user-centered design.*

General Terms
Design, Experimentation, Human Factors.

Keywords
Person Search, Information Visualization, Presentation-Level Integration, Social Data.

1. INTRODUCTION
With the advent of the internet and its underlying services along with a proliferation of personal devices with increasingly large storing capacities, our Personal Information (PI) is now scattered in an unrelated, but still connected network. The *Fragmentation Problem* [1] stems from the fact that current architectures spread the information among non-interoperable applications. Whilst the great amount of information can be seen as a problem and a challenge at several levels, it can also be faced as an opportunity. It is at our disposal and available for us to use. Social Networks (SNs) play an important role in this growth of social data; people add acquaintances or unknown people as friends, a tendency that extends the accessible information beyond our close relationships [6]. Moreover, PI available is not restricted to these sites but distributed across many different websites and services.

The boost of social data comes along with the frequent pursue for this type of information. Search engines are regularly used to find information about people [7] and one of the most popular uses given to *Facebook* is social surveillance/investigation [5] as users want to know more information about others or just keep *up-to-date*. In a social environment, it is natural to wonder about someone, even with suspicion of previous interactions ("*I know him, but where from?*" or "*We are Facebook friends but I do not recall what he does for a living*"), which is possible to be answered resorting to all this available data. Even so, when we need information about someone, we still have to navigate among all these information repositories to find it. Besides troublesome it is difficult to have an overview of what each person represents.

Visual presentations are known to be easier to use and understand than textual descriptions for many tasks. There are several successful visualizations about social relationships [4, 8], however most of them focused on a single source or information type. We have already tried to combine information from multiple sources [3], but limited the results to profile information and the most recent interactions. Our solution tries to go beyond the state-of-the-art by integrating information from multiple sources in a unified visualization that allows users to get a quick overview of whom a person is and what she/he means to them. In addition it enables a more detailed description and provides cross-source and temporal information about their relationship. It is based on three synchronized views which aim at helping the user in different ways: a *Tagcloud* for a quick overview; a *Profile* describing the person in more detail and a *Timeline* with past interactions for a better understanding of their relationship and interaction patterns. An evaluation showed that, comparing to traditional applications, our approach provides a faster and richer insight of what each person means to the user and the opportunity to quickly navigate/explore the data to find specific information.

2. DESIGN AND RATIONALE
We frequently want to know information about someone; however those specific details are spread out on different applications. Although it is troublesome when we know where that information is, it is even harder if we want to get an overview of all this information. As "*the whole is greater than the sum of its parts*", we expect that by gathering information from several sources and integrating it in a single interface, we will be able to provide a faster and richer overview of who a person is and how they are related. We also expect users to easily find specific information as it is integrated as a coherent whole instead of separate chunks.

AVI'12, May 21–25, 2012, Capri Island, Italy.

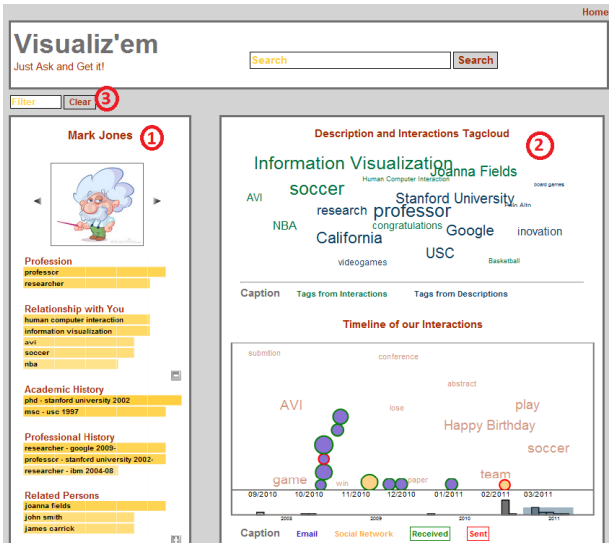

Figure 1. The results page with the different views: *Profile, Tagcloud* and *Timeline*

2.1 Visual Design

We designed an application that, with on-demand searches about a person, presents a visualization that integrates data from multiple sources (SNs, mail, search engine results, IM and *SMS*s). It has 3 synchronized views that aim to provide information related to the person's description and interactions with the user. It includes a *Profile* where the person is described, a *Timeline* with all the interactions between them, and a *Tagcloud* containing the most relevant tags. The interface is based in HTML5, *Javascript*, and *Protovis*.

2.1.1 Weighted Profile

This view intends to describe the user. The *Weighted Profile* (Figure 1-1) has a photo on the top (with the option to see more if they exist). Below are the attributes and the corresponding items. As retrieving information from the web and personal devices is not a straightforward task due to unstructured sources and name ambiguities, we added the information on how trustful and relevant each piece of information is. That value is visible by the bars' length (0-100%) and only information above 50% appears in the results. This value can be easily configured after empirical validation. To avoid endless lists of data, we opted to show just the best rated for each attribute. Yet, an option to see all results is provided (active in *Relationship with You* attribute).

2.1.2 Overview Tagcloud

The *Tagcloud* (Figure 1-2) aims at providing a quick overview of what a person represents to the user. The two colors specify whether the tag is related to the person's description or to her/his interactions with the user. The font size shows the importance of each tag and is based on the tag's frequency and/or confidence.

2.1.3 Interactions Timeline

This view shows a timeline with all interactions between the user and the person (Figure 2). It allows them to glimpse when, how and about what they interacted. If needed, users can obtain more details about each interaction. A new interaction is every sent or received event between them. We opted for this over considering communication threads due to the time those can take. As an example, a response to an email can be given 1 month later, but it does not mean they interacted during that entire time span.

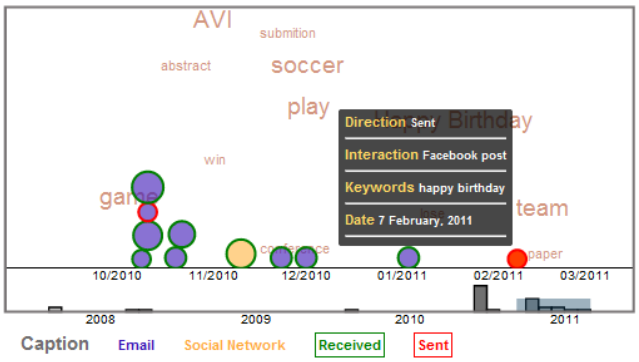

Figure 2. *Timeline* showing a preview of an interaction

It is based on the *Focus+Context* technique that allows analyzing a specific period of time without losing the global context. There is one section (at the bottom) providing the context of the interactions over time where it is possible to select the period of time to visualize (Figure 2). The bars indicate when there were interactions and their frequency. Each bar represents a relative frequency of interactions (per month). In the main chart, each interaction corresponds to a bubble and its size depends on the quantity of information exchanged in the interaction. The color specifies the means of communication and the border indicates the direction of the communication (sent or received), as depicted in the interface caption. In Figure 2 we can observe that most interactions (in the period selected) were received e-mails. The position is relevant for the x-axis, since it represents the time, but the y-axis is used to avoid bubbles to overlap. The words in the background are the most representative of the interactions in the period of time selected (position is random). This allows us to get an insight of the interactions' subjects without the need to demand more details. Yet, there may be the case where the user needs to know more about some interaction. By passing the cursor over a bubble, a brief summary is shown. It may include, for example, the date, keywords or participants. We wanted to allow the same detail given in the applications where it occurred, so we also show the whole content of the interaction by clicking on the bubble.

2.1.4 Filtering and Synchronized Views

As all views have data about one person, interacting with one view is expectable to influence the others. *Brushing and Linking* refers to the linkage of views, such that a change in one view influences the others as well. We needed it so the user, when interacting with a view, could understand how the data is shown in the remaining. For a more permanent view of the connections among views it is possible to fix a maximum of 3 highlights. This can be done by clicking on the tags/bars, or using a filter box (Figure 1-3) where the user may write what he wants to highlight.

3. EVALUATION

We carried out an evaluation aimed at ascertaining if *Visualiz'em* could provide a richer and faster overall characterization of whom a person is and if it is also suitable to quickly find specific information. We conducted a between-subjects study in which the participants were assigned to one of two different conditions wherein they had to perform the exact same tasks but using different tools: 1) *Visualiz'em*; 2) applications generally used to keep person-related information (*Usual Apps*) as we found crucial to compare our visualization with the way how people usually try to find information about someone. We resorted to: *Google*, *Gmail*, *Hotmail*, *Facebook*, *LinkedIn*, *Google+* and *SMSs*.

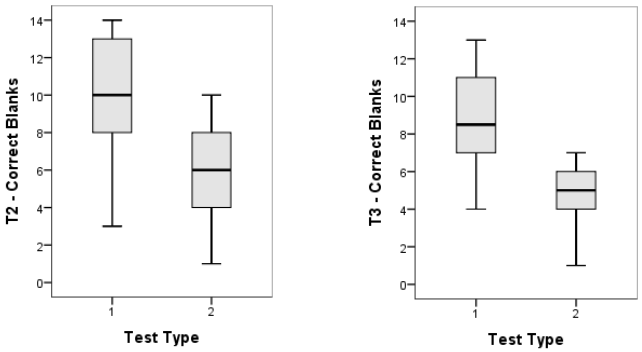

Figure 3. Box plots of the number of blanks correctly filled in tasks 2 and 3 for *Visualiz'em* (1) and *Usual Apps* (2).

3.1 Methodology

We recruited 40 volunteers from our university including students (18), researchers (9) and recently graduated (13) with ages between 17 and 28 (M=24.45), 28 males and 12 females.

Comparing approaches raises an inherent difficulty as the results can be influenced by the users' PI instead of the experimental conditions. To control this issue, we extracted information about some people resorting to single-user PI and public information available on the web. The data appearing on the visualization was extracted semi-automatically. Such process allowed us to feed the interface with real data, but control it so the visualization could not be affected by incorrect or incoherent data due to a less able *Information Retrieval* module. The common ground was established resorting to the concept of *persona*, a technique usually used in the design phase but also believed to be equally effective on early evaluation stages [2]. In our evaluation, users interpret a real person instead of one imaginary built from many real ones. We also assured that the participants did not know the person they would have to search for. This approach guarantees that the results were not affected by previous knowledge.

Participants were randomly assigned and evenly distributed over the two conditions. Those assigned to *Visualiz'em* were only allowed to use our visualization. The ones with *Usual Apps* had at their disposal the previously mentioned applications logged on (in the *persona* account) in different browser tabs. The experimental setting (independent of the experimental condition) consisted in: the **characterization questionnaire**, the **persona description** to get users into character; **tasks**; and a **satisfaction questionnaire**.

The participants performed 4 tasks in which the first was to adapt to the applications they were allowed to use *(Visualiz'em* o*r Usual Apps)*. The three remaining tasks intended to perceive if they could find/retain information in three person search scenarios. They were based i*n fill-in-the-blanks* texts to complete. The blanks included information about the interactions and profile attributes. The last 3 tasks are detailed below:

Task 2. We wanted to verify if *Visualiz'em* could provide a quick overview of whom a person is and how she/he is related to the user. To accomplish that, we asked the users to search for one person for **1 minute** and to retain the most information possible. Only after that we provided the text with the (22) blanks to complete. This reproduces the scenario where we want to know more about someone but without a focus on specific information.

Task 3. Users could read the (22) blanks text before starting the search, but they could not consult it while performing it. After the **1 minute** search, we handed them (again) the text. This reproduces the scenario where we have an idea of the information needed but are not searching for well-known specific attributes.

Task 4. The last task tried to find out the effectiveness of our visualization when the users know what they are looking for. The *fill-in-the-(24)-blanks* text was delivered before the search was performed and they could consult it while exploring the tools. We measured the time they took to finish it, with a 15 minute limit.

The quantitative data was obtained from these tasks, where we collected the blanks filled, the time to perform the fourth task and the number o*f mouseover* an*d clicks* on *Visualiz'em* views. The qualitative data was obtained through a questionnaire to determine their satisfaction and particular opinion of each view/source. They had to classify several sentences using a 5-point *Likert-Scale*, ranging from 5-totally agree to 1-totally disagree.

3.2 Results

Herein, we compare results from both conditions and the ratings given to the contribution and usefulness of *Visualiz'em* views.

3.2.1 Faster and Richer Overviews

To compare both alternatives performance in providing quick overviews about people and their relationship with the user we relied on the results of Task 2 and 3. Users had 1 minute to search for the person. In Figure 3 we can observe that *Visualiz'em* users filled more blanks correctly in both tasks (in a total of 22 blanks per task). An *Independent-Samples T-Test* showed a significant difference ($p<.005$; $\eta^2=.34$) for *Task 2* between *Visualiz'em* (M=9.95) and *Usual Apps* (M=6.05). A *Mann-Whitney U Test* (the samples did not show a normal distribution) also showed a significant difference for *Task 3* ($p < .005$; $r=.72$) between *Visualiz'em* (Median(MD)=8.5) and *Usual Apps* (MD=5). Besides performing better with *Visualiz'em*, users also classified it better for overview matters. When asked if the tools they used provided an overview of who a person is, most of *Visualiz'em* users totally agreed (75%, MD=5) or agreed (25%). *Usual Apps* ratings were lower (MD=4), but in general (90%) users were satisfied. The difference is higher when asked about the overview of the relationship and, overall, *Visualiz'em* users totally agreed (MD=5) and *Usual Apps* users agreed (MD=4). *Mann-Whitney U Tests* revealed significant differences for both cases ($p<.05$).

3.2.2 Faster Pursuing Specific Information

In task 4, users knew which specific features they needed to collect. Overall, *Visualiz'em* users filled more blanks correctly and took less time to accomplish that, as indicated by the number of blanks correctly filled per minute. Figure 4 shows that *Visualiz'em* users needed less time to find information ($M_{Visualiz'em}$ =2.84 blanks per minute; $M_{UsualApps}$ =1.35). An *Independent-*

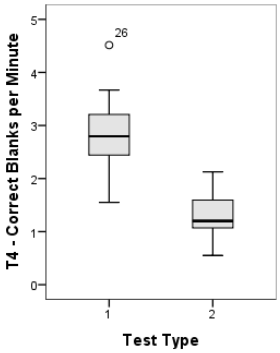

Figure 4. Box plot of the number of blanks correctly filled per minute in task 4, for *Visualiz'em* (1) and *Usual Apps* (2).

Samples T-Test showed significant differences ($p<.001$, $\eta 2=.67$). It is supported by the ratings users gave when asked if it was easy and quick to explore the results to find the desired information ($MD_{Visualiz'em}=4$; $MD_{UsualApps}=3$ for both). *Mann-Whitney U Tests* showed significant differences ($p<.05$), including when asked about the tools ability to obtain specific data (both MD=4).

3.2.3 Analyzing Visualiz'em

To understand how users interacted with the interface we resorted to the frequency of clicks/*mouseovers* in profile bars, tags and timeline bubbles. In tasks 2 and 3 most users did not click on any view, as they were focused on quickly obtaining the most information possible. Since they had more time in task 4 they interacted more, mainly with the *Timeline*. This is explained by the need to see some of the interactions in detail (MD=5; IQR=4). Likewise, it got more *mouseovers* in all tasks. To select filters, users resorted more often to the profile bars (MD=2, IQR=2) than to the tags (MD=0, IQR=1). It is supported by the *mouseover* frequencies in all tasks. Yet, a *Wilcoxon Signed Ranks Test* showed a significant difference only in task 3 ($p<.005$, $z=-3.072$).

We asked the users to classify sentences related to each view's relevance to provide a person/relationship overview, specific information and the overall relevance. Users classified the profile as the most relevant view for *Person Overview* and a *Friedman Test* and post-hoc *Wilcoxon Signed Rank Tests* revealed that the difference to the other views is statistically significant ($p<.05$ $r_{profile\text{-}tagcloud}=.46$, $r_{profile\text{-}timeline}=.37$). The same occurs respecting *Specific Information* but with a larger effect size ($r_{profile\text{-}tagcloud}=.56$, $r_{profile\text{-}timeline}=.50$). In this case the difference between the *Timeline* and *Tagcloud* is also significant ($p<.05$, $r=.37$). To provide a *Relationship Overview*, the most relevant is the *Timeline* with significant difference ($p<.05$, $r_{timeline\text{-}tagcloud}=.46$, $r_{timeline\text{-}profile}=.50$) to the other views. These results suggest that the *Tagcloud* plays a minor role in comparison with the *Timeline* and *Profile,* which is supported by the *Overall Relevance* classification. The *Profile* was classified as the most relevant, with a significant difference ($p<.05$, $r_{profile\text{-}tagcloud}=.42$, $r_{profile\text{-}timeline}=.38$).

3.3 Discussion

The final questionnaire suggests that users were overall satisfied with both alternatives, either to provide person and relationship overviews or to obtain specific information, yet with a significant advantage to *Visualiz'em*. While expecting an advantage we believed that *Usual Apps* results would be worse, but users referred *Facebook* and *LinkedIn* profiles as a good combination to obtain specific information and a person overview. As this may be true, tasks analysis shown that it was more difficult to obtain a person overview in a small period of time and to find specific information was considerably slower. Likewise, most *Usual Apps* users rated it positively for relationship overview due to *Facebook's See Friendship* feature and mail messages; still, the blanks related to the interactions showed that it was also more difficult to gather that information in *Usual Apps*. Having the interactions spread out by several sources makes it more troublesome to find specific information and lack an overview of the interactions history and its patterns (who, how, when, what,..).

Visualiz'em results suggested that it is a step ahead from our usual methods to search information about people but they also allowed us to consider the options we made. The *Timeline* revealed its capability to present cross-source interactions and provide relationship overviews; the *Profile* was high rated to find specific information and to provide a person's overview by resorting to the integration of SNs profile information; but the *Tagcloud* results did not match its purpose. We expected it would be important to provide quick overviews; and as we think it may still be true for shorter periods of time where users do not have time to explore the visualization we also believe that, at least, it is oversized in our interface and its contribution needs to be discussed.

4. CONCLUSIONS

There are many applications dealing with information about people and in most cases they are designed for different purposes and have different information therein. Even applications with similar aims may deal with different information content and searching them separately only provides part of the knowledge. As it may be difficult to integrate it within applications by deflecting their rationale, gathering that information and integrating it in a single, organized visualization tool should be considered to reduce unnecessary navigation. The same happens when taking interactions into account but with wider inherent consequences, since it is more difficult to obtain all the information and insights that are provided in an integrated solution. Future research in this area should consider that "*the whole is greater than the sum of its parts"* as sources alone are much less expressive The results have shown that an integrated visualization is able to provide richer and faster overviews of whom a person is and of her/his relationship with the user, as well as it is faster when seeking specific information. We wanted the users to perform exactly the same tasks and that included searching for the same people to use the same fill-in-the-blanks text in both conditions. This was possible by resorting to the real data of an external person. The results provided good insights, but we find very important to evaluate our approach with the users' data in the future.

5. ACKNOWLEDGMENTS

Work supported by the Portuguese Foundation for Science and Technology (FCT): João Guerreiro's scholarship, grant SFRH/ BD/66550/2009; INESC-ID multiannual funding - PIDDAC Program funds; project PAELife, reference AAL/0014/2009.

6. REFERENCES

[1] Bergman, O et al. The project fragmentation problem in personal information management. *In proc. of CHI* (2006).

[2] Campos, J., Paiva, A. A personal approach: the persona technique in a companion's design lifecycle. In *Proc. of Interact* (2011).

[3] Guerreiro, J. et al. Relating Personally Relevant Social Information from Heterogeneous Sources. *In proc. of UDISW (2010)*.

[4] Heer, J.; Boyd, D.; , "Vizster: visualizing online social networks. In *Proceedings of INFOVIS* (2005)

[5] Joinson, A. N. Looking at', " Looking up " or " Keeping up with " People ? Motives and Uses of Facebook. *Social Networks*, 5 (2008).

[6] Lugano, G. Mobile social networking in theory and practice. *First Monday 13*, 11 (2008).

[7] Spink, A., Jansen, B.J., Pedersen, J. Searching for people on web search engines. *Journal of Documentation* 60 (2004).

[8] Tat, A. and Carpendale, S.. Crystal Chat: Visualizing Personal Chat History. *In Proceedings HICSS,* IEEE (2006).