A user-centric design for the BiographyNet Linked Data interface

Wouter van den Hoven

MSc Information Sciences
VU University Amsterdam, Amsterdam, The Netherlands
wnvdhoven@gmail.com

Abstract. In order to fully exploit the benefits of modern computer technology in non-computer oriented fields of research, in this case history, researchers of the VU University Amsterdam have started the BiographyNet project. The goal of this project is to facilitate historical research through computational and semantic technology. Part of this project is the development of a demonstrator. This research proposal focuses on the design of the user interface of such a demonstrator, where multiple relations and points of view have to be visualized. Challenges occurring are finding out the historians information needs, facilitating the selection of multiple conflicting or aggregated views, giving the user insight in the data selecting and manipulating process, and visualizing context of the data. The approach will be divided in three steps, namely 1) Information gathering and Requirements Engineering, 2) Prototype developing and evaluating, and 3) Evaluating the final prototype. The final prototype will be considered successful when the majority of prescribed tasks can be executed and the System Usability Score is above average.

1 Introduction

In the past decades computer technology and the internet became an increasingly big part of our lives. Where first only experts used computers daily, now every person has to use some sort of computer in their everyday life. This is also the case with academics, using computer technology in their research.

In order to fully exploit the benefits of modern computer technology in non-computer oriented fields of research, in this case history, researchers of the VU University Amsterdam have started a project called BiographyNet[1]. Other partners in the project include the Netherlands eScience Center[2] and Huygens ING[3]. *"The BiographyNet project is an e-history project bringing together researchers from history, computational linguistics and computer science."* (Ockeloen et al., 2013). The goal of this project is to facilitate historical research through computational and semantic technology. Part of the project is to develop a demonstrator where historians should

[1] http://www.biographynet.nl/
[2] http://esciencecenter.nl/
[3] http://www.huygens.knaw.nl/

be able to browse through biographical data to discover interesting relations between people, events, places, and time periods. Historians can then use these interesting relations to start new research. The BiographyNet project uses data of the Biography Portal of the Netherlands (BP)[4] which contains 125,000 Dutch biographies of over 75,000 people. In this document we describe what challenges have to be taken into account while researching how to design the user interface of such a demonstrator and how we plan to solve these challenges by performing requirements engineering and prototyping.

2 Related work

In order to combine the data from the various sources the BiographyNet schema[5] was designed. Using this schema provenance of the data can be managed, which is important for both the end-users, namely historians, and the computer scientists involved in the project (Ockeloen et al., 2013). The schema also allows multiple sets of, possibly conflicting, meta data to be stored. The BiographyNet schema uses Linked Data, using the RDF[6] (Resource Description Framework) format. Linked Data is *"data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets."* (Bizer, 2009). The main challenges in the development of user interfaces for applications that use Linked Data according to Heath (2008) are offering the appropriate functionalities and providing guidance to users.

Berners-Lee et al. (2006) describe an RDF browser called Tabulator used on the Semantic Web. The browser focuses on new users who are not familiar with the Semantic Web, and on developers to stimulate them in using RDF, help improving RDF linking standards, and letting providers of the data see how it is linked to the Semantic Web. The browser has two modes, exploration and analysis, that show information that fits the users curiosity, an aspect we would much like to see in the BiographyNet demonstrator, as the main goal is "*to inspire historians when setting up new research projects*" (Ockeloen et al., 2013).

Lei, Uren, and Motta (2006) describe a semantic search engine, SemSearch, which makes semantic search accessible for end users by hiding its complexity. Users do not need to have any foreknowledge to create queries, but it is still possible to create complex queries in SemSearch. This makes the search engine useful for a diverse group of end users. Furthermore, the reason why specific results are displayed is explicitly explained by showing how the search terms are related to the results.

The CultureSampo[7] project uses semantic web technologies like Linked Data to create links between heterogeneous cultural objects (Mäkelä et al., 2012). The user interface offers multiple approaches to the data, such as map browsing and searching,

[4] http://www.biografischportaal.nl/en
[5] http://www.biographynet.nl/schema/
[6] http://www.w3.org/TR/rdf-schema/
[7] http://www.kulttuurisampo.fi/?lang=en

relation search between two persons, a timeline view of events, and biographies enriched with data from the semantic web.

In the Open PHACTS[8] project a discovery platform is built in order to reduce the barriers in drug discovery in industry, academia, and small businesses. The discovery platform will be "*integrating pharmacological data from a variety of information resources and providing tools and services to question this integrated data to support pharmacological research.*". The discovery platform supports the management of provenance, and offers the user the option to toggle provenance display on or off. The Open PHACTS discovery platform and the BiographyNet demonstrator have a couple of similarities in their key-elements, namely combining multiple heterogeneous sources in a Linked Data schema, managing and visualizing provenance, and the goal of supporting research.

3 Problem statement

Developing a demonstrator that allows end users who have standard computer skills to browse the complex linked data schema of BiographyNet is the main challenge of this project. The goal of the demonstrator is to facilitate new ways of data exploring in addition to current methods already used by historians. The possibilities of a demonstrator include browsing through biographies in a faster way than the current classical archive work, and thereby showing interesting relations between people, events, places, and time periods. In order to establish such a demonstrator there are multiple problems to overcome.

First, the demonstrator should allow the user to trace back different data sources for the same object, since information about one object can be composed from multiple data sources. The challenge here is to show what part of the information originates from which source. Also, different sources could provide multiple views on the same object, therefore it needs to be determined how many of these views should be visualized and in which way.

Second, the end user must have insight in the process of manipulating and selecting the data that was used to generate information, also displaying potentially relevant data that was left out. This way the end user can be reassured the information is reliable, and possible errors caused by the automated text interpretation that is used to select the data can be exposed this way.

Third, using the demonstrator should aim to give as much insight into the data and the origins of the data as usual archive work does. Therefore it is important to visualize as much of the context of the data as possible. The context can be, for example, the text surrounding the data in the original source.

Finally, the end users, namely historians, cannot be expected to possess extensive computer skills. Therefore the interface should be simple to use, while still visualizing all the necessary information. The challenge here is to display all the information

[8] http://www.openphacts.org/

needed in a clear way, so that the end user will not get overwhelmed and lost in loads of information.

4 Research question

Based upon the problem statement, we can formulate the following research question and subquestions:

- How can relations between people, events, places, and time periods from multiple points of view be visualized in order to support academic research?
    - What are the information needs for historians and how can all needs be met and displayed without losing overview?
    - How can the user select between multiple conflicting or aggregated views?
    - How can the user get insight in the process of manipulating and selecting the data used?
    - How can a form of context of the original data source be visualized?

5 Method

The approach for this research will be divided in three steps, namely 1) Information gathering and Requirements Engineering, 2) Prototype developing and evaluating, and 3) Evaluating the final prototype.

First, an extensive requirements engineering phase will take place. A questionnaire will be distributed among possible end users, and interviews will be held. Participants for the questionnaire and the interviews will be recruited from the Faculty of Arts, more specifically students and staff members of the BSc and MSc History program. In order to make sure the right questions are asked in both the questionnaire and interviews and to gather information about the end user, interviews will be held with experts in the field of research, and especially in the BiographyNet project. Those experts include the researchers involved in the BiographyNet project and partners involved in the project. Both the interviews with the experts and with the end users will be flexible interviews, to make sure topics that arise during the interview can also be treated. After the information from the questionnaires and the interviews is collected, a list of requirements will be drawn up, as well as a set of prescribed tasks the user must be able to execute.

The second step is to develop and evaluate prototypes in two cycli, using the information gathered during the requirements engineering. In the first cycle a low-fidelity prototype will be developed using screen mockups. This prototype will be evaluated by a minimum of five end users, preferably a mix of the interviewees of the requirements engineering phase and new participants not yet familiar with the project in order to get new insights, in one-to-one meetings, where the participants will be

asked to fill in the System Usability Scale (SUS) (Brooke, 1996), and make suggestions for the second, high-fidelity, prototype. Also, other remarks by the participant will be noted and taken into account.

The second, high-fidelity, prototype will be developed in the form of a live-demo. This demo will be constructed using HTML and example data. By using a demo the participant can get a better impression of the feel, layout, and behavior of the user interface during evaluation (Stone et al., 2005).

After the high-fidelity prototype is developed it will be evaluated, preferably by the same participants that evaluated the low-fidelity prototype and again new participants to get new insights. Again in one-to-one meetings the participants will be observed while using the demo, executing a selection of prescribed tasks, following the Think-Aloud protocol. Also, the participants will be asked to fill in the SUS for this prototype, and to conclude the meeting a post-session interview will be held to debrief the session. The prototype will be considered a success when the majority of the tasks can be executed, and the SUS-score is higher than 70, as this is considered to be above average (Bangor, Kortum, and Miller, 2009).

6 Planning

Week	**Activities**
6. 3-feb	Interview experts. Find participants
7. 10-feb	Interview experts. Find participants
8. 17-feb	Make questionnaires and interviews. Find participants
9. 24-feb	Interview participants
10. 3-mar	Buffer
11. 10-mar	Make list of requirements
12. 17-mar	Make low-fidelity prototypes, screen mockups
13. 24-mar	Buffer
14. 31-mar	Evaluate low-fidelity prototypes
15. 7-apr	Buffer
16. 14-apr	Adjust requirements list using information from first evaluation
17. 21-apr	Make high-fidelity prototype, HTML demo
18. 28-apr	Make high-fidelity prototype, HTML demo
19. 5-may	Evaluate high-fidelity prototype
20. 12-may	Evaluate high-fidelity prototype
21. 19-may	Buffer
22. 26-may	Make midterm presentation **(May 31: Midterm presentation)**
23. 2-jun	Write thesis
24. 9-jun	Write thesis
25. 16-jun	Write thesis
26. 23-jun	Write thesis, make final presentation

27. 30-jun	Make final presentation **(July 5: Final presentation)**

References

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), 114-123.

Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., ... & Sheets, D. (2006, November). Tabulator: Exploring and analyzing linked data on the semantic web. In *Proceedings of the 3rd International Semantic Web User Interaction Workshop* (Vol. 2006).

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *5*(3), 1-22.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*, 194.

Heath, T. (2008). How will we interact with the web of data?. *Internet Computing, IEEE*, *12*(5), 88-91.

Lei, Y., Uren, V., & Motta, E. (2006). Semsearch: A search engine for the semantic web. In *Managing Knowledge in a World of Networks* (pp. 238-245). Springer Berlin Heidelberg.

Mäkelä, E., Hyvönen, E., & Ruotsalo, T. (2012). How to deal with massively heterogeneous cultural heritage data–lessons learned in culturesampo. *Semantic Web*, *3*(1), 85-109.

Stone, D., Jarrett, C., Woodroffe, M., & Minocha, S. (2005). *User interface design and evaluation*. Morgan Kaufmann.