Can Linked Data improve the User Experience (UX)?

Rosa Gil, Josep Maria Brunetti, Juan Manuel Gimeno, Roberto García

Universitat de Lleida. Jaume II, 69. 25001 Lleida, Spain
{rgil, josepmbrunetti, jmgimeno, rgarcia}@diei.udl.cat

Abstract. The amount of data in different forms, from CSVs to Linked Data (LD), is rapidly increasing. The more sophisticate the way of publishing it, the more computers can help us dealing with it. Our sense best prepared to deal with these amounts of data is sight so the way to establish this communication among computers and us seems to be quite dependent on visualisation techniques, Interaction Patterns and Information Architecture. We have developed a first prototype of a Linked Data publishing tool enriched with concepts from the previous disciplines and performed a preliminary test with end-users to show Linked Data can improve the User Experience (UX).

Keywords: Linked Data, Human Computer Interaction, Information Architecture, User Experience, Usability.

1 Introduction

The amount of data available in the Web, from file dumps to Linked Data, is becoming so big that computerised mechanisms are required to help us deal with it. However, at last, it is our responsibility to make sense of all this data in order to discover unforeseen patterns, make decisions, etc. Moreover, the range of devices where users can consume data is widening so even more care should be taken when building user interfaces for all this data.

The potential of this huge amount of data is enormous but it is not being fully realized as end-users find a great barrier when facing it. The barrier is that most of this data is available as data dumps or SPARQL endpoints.

For data dumps, it is really complicated to realise what data does one have at hand, what it refers to and what kind of terms are used. And it requires some experience in Semantic Web tools in order to do so.

For SPARQL endpoints, the amount of work required for grasping the internalities of the data set might be reduced. However, a good knowledge of SPARQL is required in order to generate and understand a set of queries that allow realising how big the data set is, which are the main kinds of things, how are they interrelated, etc.

Consequently, computers need a powerful way to communicate with us when such amounts of data are into play. Our approach focuses in users and not in a particular browse process or search algorithm the reason is double: users change their behaviour depending on context, mood or topic when they consume data [1] and secondly, in some data sets the browse or search process changes depending on user's mental model [2].

Users process best great amounts of information using the fastest sense, sight [3]. As our sense best prepared to deal with these amounts of data is sight, the way to establish this communication among computers and us seems to be quite dependent on disciplines like Visualisation, Interaction Patterns and Information Architecture.

From the Interaction Patterns point of view, we have started from the fundamental set of tasks for data analysis proposed by Shneiderman [4]. Below, there are these tasks associated to the Interaction Patterns that we propose to apply in Linked Data scenarios. This is just a preliminary proposal based on simple Interaction Patterns and future work now concentrates on exploring richer ones:

- *Overview*: get the full picture of the data set at hand. At this stage we propose to apply the Global Navigation interaction pattern. In the context of Information Architecture (IA), it corresponds to the navigation bars users are used to see at the top or on the right of web sites.
- *Zoom & Filter*: zoom in on items of interest and filter out uninteresting items. Here the proposal is Faceted Navigation, facets in IA. Once we have zoomed by selecting the kind of things we are interested in from the navigation bar, facets for that set of things help us filtering out those we are not interested in.
- *Details on Demand*: after zooming and filtering the user arrives to the concrete resources of interest. At this point, the user can get the details for those resources, which in the case of Linked Data is to get the properties for the resources plus those properties pointing to them. This is related to the Details on Demand interaction pattern and can be implemented as a simple list of properties and values of the resource of interest or as a specific visualization tailored to the kind of resource at hand, e.g. a map for geo-located resources.

Our proposal is to elaborate these interaction patterns in the context of Linked Data. We have chosen them because they are simple and very common so users are very confortable with them. They are currently part of the "culture" about how information is structured in the Web so they have been deeply studied in Information Architecture (IA) domain [5].

The drawback of all these IA systems is that they are quite expensive to develop and maintain. Fortunately, when they are built on top of the structured data typical in the Semantic Web and Linked Data, it is possible to automate most of the development and maintenance work.

We are currently testing all these interaction patterns in a Linked Data publishing tool called Rhizomer[1]. It features navigation bars automatically generated and maintained starting from the underlying thesaurus and ontologies. A similar approach is followed for generating facets for each kind of entity in the data set. More details about Rhizomer are available from [6].

2 Evaluation

Rhizomer, though it is still at a prototype stage, has already been tested with end-users in order to evaluate its functionality and usability. The goal of the test conduced

[1] http://rhizomik.net/rhizomer/

so far was to preliminary evaluation of the Information Architecture components, if they are understood and if they improve the awareness of the structure of a particular dataset by improving user performance when looking for a specific piece of information.

We registered end-users sessions with the system and analysed user test data. For the usability test metrics we chose effectiveness (percentage of tasks completed) and efficiency (time to complete a task).

We have used a real test dataset called the Linked Movie Data Base (LinkedMDB)[2]. We chose the movies domain because it is well known for most users and quite appealing. LinkedMDB is generated from the Internet Movie Database[3] (IMDB).

Therefore, we considered interesting to compare the evaluation results with those for IMDB and thus be able to test if the same data as Linked Data can become more usable than from the original web site an improve the user experience. Consequently, we established one task to be performed with IMDB and another one with Rhizomer. The order of the tasks randomized in order to avoid memory effects.

Six participants were selected, with a unique profile characterized by good knowledge of information technology, limited knowledge about Semantic Web technologies and interest in movies.

The test facilitator proposed users the two tasks, but not necessarily in this order:

- **Task A: "**Find three films where Woody Allen is director and actor at the same time" using IMdB.
- **Task B: "**Find three films where Clint Eastwood is director and actor at the same time" using Rhizomer.

The main findings from the test were:

- Only one participant was able to complete the **Task A** without assistance.
- **100%** of participants needed in at least one occasion the guidance of the facilitator to successfully complete **Task B**.
- In **Task B**, **all** participants began the navigation from actors instead than from movies. This was the reason why users required assistance but as soon as they realized they were able to start from movies, the task was easily solved.

Efficiency based on the degree of completeness is relatively low for both tasks. 32% on average in the first task, and 54% in the second task. Only one participant approached 100%, giving an efficiency of 95% for the second task.

- 83% of participants completed the second task in less time than the first. Just one user completed the first task in less time than the second.

3 Conclusions and future work

The main conclusion of the evaluation, apart from the initial problems that make users require assistance in almost all cases, is that Rhizomer improved the user experience

[2] LinkedMDB http://www.linkedmdb.org

[3] IMDB database http://www.imdb.com

for the users interacting with the Linked Data version of IMdB. This was observed not just quantitatively in the fact that users were more efficient when completing the Rhizomer task, but also qualitatively through their comments recorded during the evaluation.

The increased efficiency was observed not just when Task B was completed after Task A but also when Task B was the first to complete. This increase in efficiency would have been even more evident if the users had not required assistance during the interaction.

Consequently, an additional conclusion is that though the proposed approach based on facilitating to the user mechanisms to attain the interaction patterns proposed by Shneiderman using Information Architecture components that are familiar for them, the automatisation of these components has made them too dependant on the structure of the underlying data.

Consequently, the user interaction is currently too constrained by how the underlying data is structured. In this test, the result was that the task was performed differently from it was expected and this confused all users. They were looking for movies where actor and director were the same but, instead of initiating their interaction from the "Movies" menu option, all users started from "Actor".

From there, as the underlying data just modelled actors per film but not the reverse, it was impossible to filter those films where the same person was the director. The easy way was to look for movies and to filter by director and actor using the corresponding facets, as the underlying data has these two properties associated to every film.

The impression is that users tend to think first about persons and consider films a secondary entity. The idea here is to exploit the possibilities of the underlying conceptual model and derive implicit properties, for instance reverse properties, in order to provide users with alternative paths. In this particular case, there will be reverse properties from actors to films. Moreover, it will be necessary then to focus on the set of films for an actor and filter it by director.

References

1. Heinström, J. Five Personality Dimensions and Their Influence on Information Behavior. Information Research. Vol 9, n.1, October 2003. (2003)
2. Young, I. Mental Models. Aligning Design Strategy with Human Behavior. Rosenfeld. (2008).
3. McCandless. D. The beauty of data visualization. TEDGlobal 2010, July 2010, Oxford, UK (2010).
4. Shneiderman, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proceedings of the 1996 IEEE Symposium on Visual Languages (VL '96), pp. 336-343 (1996).
5. Morville, P. & Rosenfeld, L. Information Architecture for the World Wide Web. O'Reilly Media. (2006)
6. García, R.; Brunetti, J.M.; López-Muzás, A.; Gimeno, J.M. & Gil, R. (2011). Publishing and Interacting with Linked Data. 1st International Conference on Web Intelligence, Mining and Semantics, WIMS'11, May 25-27, 2011, Sogndal,