

Evaluating Interaction Patterns for Linked Data

Rosa Gil, Antonio López-Muzás, Josep Maria Brunetti, Juan Manuel Gimeno, Roberto García
Universitat de Lleida
Jaume II, 69
25001 Lleida, Spain
+34 973702742
{rgil, jmbrunetti, jmgimeno, rgarcia}@diei.udl.cat, lopezmuzas@gmail.com

ABSTRACT

The amount of data in different forms, from CSVs to Linked Data, is rapidly increasing. The more sophisticated 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.

Categories and Subject Descriptors

H.5 [Information Interfaces and Presentation]

Keywords

Semantic Web, Linked Data, interaction, information architecture, visualisation.

1. INTRODUCTION

The amount of data in different forms, from CSVs to Linked Data, is rapidly increasing. The more sophisticated the way of publishing it, the more computers can help us dealing 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.

Linked Data is just one example, and the potential of this huge amount of data is enormous but it is not being fully realised 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. Users process best great amounts of information using the fastest sense, sight [1]. 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 [2]. 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 pattern¹. 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 Navigation², 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:** 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 Demand³ interaction pattern and can be implemented as a simple list of properties and values of the resource of interest or as a specific visualisation 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 comfortable 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 [3].

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

¹ <http://www.welie.com/patterns/showPattern.php?patternID=main-navigation>

² <http://www.welie.com/patterns/showPattern.php?patternID=faceted-navigation>

³ <http://www.welie.com/patterns/showPattern.php?patternID=details-on-demand>

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 Rhizomer⁴. 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 [4].

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 conducted 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)⁵. We chose the movies domain because it is well known for most users and quite appealing. LinkedMDB is generated from the Internet Movie Database⁶ (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. Consequently, we established one scenario with one task to be performed with IMDB and another one with one task for Rhizomer.

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 scenarios and tasks, but not necessarily in the same 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 first task without assistance.
- **100%** of participants needed in at least one occasion the guidance of the facilitator to successfully complete the second task.
- 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.

⁴ <http://rhizomik.net/rhizomer/>

⁵ LinkedMDB <http://www.linkedmdb.org>

⁶ IMDB database <http://www.imdb.com>

- 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

From test results and their analysis, we have elaborated these proposals for Rhizomer improvement:

- Navigation must be better contextualised. The interface should provide more mechanisms to inform the user where she is, where she can go and where she has been. For that, the proposal is to integrate some kind of breadcrumbs that summarise the navigation steps though navigation menus and facets.
- To improve how facets are presented to the user, especially when there are a lot of values. For that, the proposal is to use values indexes or graphical representations for numeric values, e.g. histograms.
- One of the main issues detected is that 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.

4. ACKNOWLEDGMENTS

The work described in this paper has been partially supported by Spanish Ministry of Science and Innovation through the Open Platform for Multichannel Content Distribution Management (OMediaDis) research project (TIN2008-06228).

5. REFERENCES

- [1] David McCandless. (2010). The beauty of data visualization. TEDGlobal 2010, July 2010, Oxford, UK
- [2] Shneiderman, B. (1996). 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.
- [3] Morville, P. & Rosenfeld, L. (2006). Information Architecture for the World Wide Web. O'Reilly Media.
- [4] 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, Norway.