

Information Architecture Automatization for the Semantic Web¹

Josep Maria Brunetti²

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

Abstract. Our proposal is to develop generic Information Architecture components to facilitate publishing and browsing semantic data in the web, improving its usability and accessibility.

Keywords: Semantic Web, Linked Data, Information Architecture, Usability, Accessibility, Navigation, Metadata.

1 Introduction

Despite the Semantic Web was proposed ten years ago [1], it hasn't been until recently when it has started to become popular. In the last years, the amount of semantic data available in the Web has increased in a spectacular way, especially thanks to initiatives like Linked Open Data (LOD). The objective of this initiative is to motivate the publication of Open Data in formats that are more easily integrable, queryable and that facilitate its reuse. The cloud of interrelated and open datasets included in the LOD cloud has rapidly evolved, from the 2 billion statements and 30 datasets one year after its creation in February 2007, to more than 25 billion statements and 200 datasets in September 2010³.

The potential of this huge amount of data is enormous but in most cases it is very difficult for users to explore and use these data. The problem is that most of this data is available as raw data dumps or SPARQL [2] semantic query services, being it complicated to realise what the data is about, which are the main kinds of things, how are they interrelated, etc. Moreover, it requires some experience in Semantic Web tools and, in any case, the results are not very usable. Therefore, the challenge is to structure all this information making it more usable and accessible.

Information Architecture (IA) is a discipline that focuses its efforts in this problem, especially in complex systems and situations with great amounts of information. IA

¹ The work described in this paper has been partially supported by Spanish Ministry of Science and Innovation through the OMediaDis research project (TIN2008-06228).

² This dissertation work is supervised by Roberto García

³ <http://linkeddata.org/>

describes the structure of a website, how its information is grouped, navigation methods and terminology used. A good IA can improve the quality of a website and users can find more easily the information they are looking for.

The proposal of this project is to draw from the experience accumulated in the Information Architecture domain [3] and adapt existing IA components to these situations with large amounts of heterogeneous semantic data. The rest of this paper is structured as follows. In Section 2 we present the state of the art and other approaches to this problem. Section 3 is about our proposed approach and in Section 4 we explain the methodology we will follow during this PhD and evaluation methods. Finally, in Section 5 we present our conclusions and future work.

2 State of Art

The best approach to make a semantic dataset more usable to a wider range of users is to use a Web data publishing tool or a Semantic Web browser, e.g. Disco [4]. These kinds of tools provide an HTML rendering for each resource in the dataset, listing all their properties and values. HTML pages are interlinked based on the connections between resources in the underlying dataset. However, this feature is only useful if the user has some a priori knowledge of the data structure. There is no way to get an overview of the kinds of resources in the dataset, e.g. person, address, etc.

Explorator [5] is another tool that makes it possible to browse a dataset combining search, facets and operations on sets of resources. However, it is also difficult for users to get a broader view on the dataset. Other tools also provide faceted views but not as a generic browser, i.e. just for a specific dataset like the case of the DBPedia Faceted Browser [6] or \facet [7].

Consequently, existing tools make very difficult for users to explore a dataset, realize what kind of resources there are, what properties they have and how they are related.

3 Proposed Approach

As the volume of information in the Semantic Web increases, interacting with information becomes a more difficult task. To interact with these amounts of data, users use different ways or strategies depending on their goals.

Starting from the fundamental set of tasks for data analysis proposed by Schneiderman [8], we have explored the most appropriate Interaction Patterns to perform these tasks in the context of the Semantic Web:

- **Overview:** the goal of this task is to get a full view of the entire collection. We propose to apply the Global Navigation interaction pattern. In the context of IA it corresponds to **navigation menus**.
- **Filter:** to select items of interest and filter out uninteresting items through exploratory search. The proposal is the Faceted Navigation pattern, which in the context of IA corresponds to **facets**.

- **History:** to keep a history of actions and support undo and replay actions. We propose to apply the Breadcrumb Navigation pattern. In the context of IA it corresponds to **breadcrumbs**.

We have chosen these IA components because they are simple and very common in websites, so users are comfortable with them. Besides them, we are exploring other IA components such as **sitemaps** and **search systems**.

The drawback of all these IA systems is that they are quite expensive to develop and maintain, even more with the large amount of heterogeneous semantic data we are dealing with. Fortunately, when these IA systems are built on top of structured data and formal languages, it is possible to automate most of this development and maintenance work.

Concretely, the Semantic Web, with the use of ontologies and semantic languages like RDF, allows the automatization of these components. The Semantic Web provides methods and tools to model the Information Architecture for a domain with much more detail and in a formal way. It is possible to define concepts, properties and relations between them. This allows the use of automated tools to process these formal descriptions, create and maintain the Information Architecture.

The main goal of this PhD is to define a methodology and develop the necessary components to assist during the definition and maintenance of the Information Architecture of websites from a formal description of the site's domain. It is also intended, in parallel, to improve the usability and accessibility of the generated Information Architecture. By using these components, the final users will be able to easily explore semantic data, which at this moment is not usable and accessible for them.

4 Methodology and evaluation

The methodology to follow in this project is based on the MPIu+a [9] development process, which integrates the disciplines of Software Engineering, Human-Computer Interaction and Accessibility. Besides the traditional Software Engineering lifecycle, MPIu+a has two other main blocks, prototyping and evaluation, which are also two main parts of this project.

We plan to develop prototypes for each of the identified Information Architecture components, test and evaluate them with users. The evaluation will be conducted with common evaluation methods such as user tests, card sorting or heuristic evaluation. The objective of the evaluation is to test and compare the generated Information Architecture with a traditional one.

Although this project has just barely started, a prototype⁴ has already been tested with end-users in order to evaluate its functionality and usability. The goal of the tests conducted so far was to do a preliminary evaluation of the menu and facet components, to check if they are understood and they improve user performance when looking for a specific piece of information. The results so far show that LinkedData can improve the user experience, providing more ways to interact with data [10].

⁴ <http://rhizomik.net/linkedmdb/>

5 Conclusions and Future Work

In this paper we have presented our vision of this doctoral thesis in the context of the Semantic Web and Information Architecture. We have started developing generic IA components that can be reused in websites with totally different domains as long as they are based on Semantic Web technologies.

The identified IA components are able to use the semantics captured by ontologies and semantic data, providing users different ways to access and interact with the data. Those components, which are automatically generated and maintained, facilitate publishing and browsing a dataset without requiring a priori knowledge of it or experience in Semantic Web tools.

The short-term future work focuses now on implementing the improvements detected during the user tests, basically those that affect the navigability between resources in the dataset. The main problem detected is that the user interaction is currently too constrained by how the underlying data is structured. The objective is to improve IA flexibility and obtain implicit properties and relations, for instance reverse properties, in order to provide users with alternative paths.

Then, we also plan to develop prototypes for the other identified IA components, test and evaluate them with users.

References

1. Tim Berners-lee, James Hendler and Ora Lassila. The semantic web. *Scientific American*, 284:34–43, 2001.
2. Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF. Recommendation, World Wide Web Consortium (2008). <http://www.w3.org/TR/rdf-sparql-query>.
3. Morville, P., Rosenfeld, L: *Information Architecture for the World Wide Web*. O'Reilly Media, Cambridge, MA (2006).
4. Bojars, U., Passant, A., Giasson, F., Breslin, J. G.: An Architecture to Discover and Query Decentralized RDF Data. *Proceedings of Workshop on Scripting for the Semantic Web, SFSW 2007. CEUR Workshop Proceedings, Vol. 248 (2007)*.
5. Araujo S., Schwabe D., Barbosa S.: Experimenting with Explorator: a Direct Manipulation Generic RDF Browser and Querying Tool. *Visual Interfaces to the Social and the Semantic Web (VISSW 2009)*, Sanibel Island, Florida (2009).
6. Hahn, R., Bizer, B., Sahnwaldt, C., Herta, C., Robinson, S., Bürgle, M., Düwiger, H., Scheel, U.: Faceted Wikipedia Search. *13th International Conference on Business Information Systems (BIS'10)*, Berlin, Germany (2010).
7. Hildebrand, M., Ossenbruggen, J., and Hardman, L.: /facet: A Browser for Heterogeneous Semantic Web Repositories. *The Semantic Web - ISWC 2006*. Springer. (2006) 272-285.
8. 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.
9. Granollers, T.: 'User centred design process model, integration of usability engineering and software engineering', *Proc. INTERACT 2003*, Zurich, Switzerland, 2003, pp. 673-675.
10. Gil, R., Brunetti, J.M., López-Muzás, A., Gimeno, J.M., García, R.: Improved Linked Data Interacting with the Web of Data through Information Architecture Components. *INTERACT 2011*, Submitted.