Rhizomer: Overview, Facets and Pivoting for Semantic Data Exploration

Roberto García
Universitat de Lleida
Jaume II, 69
25001 Lleida, Spain
roberto.garcia@udl.cat

Josep Maria Brunetti
Universitat de Lleida
Jaume II, 69
25001 Lleida, Spain
josepmbrunetti@diei.udl.cat

Rosa Gil
Universitat de Lleida
Jaume II, 69
25001 Lleida, Spain
rgil@diei.udl.cat

Juan Manuel Gimeno
Universitat de Lleida
Jaume II, 69
25001 Lleida, Spain
jmgimeno@diei.udl.cat

ABSTRACT

Rhizomer addresses the exploration of semantic data by lay users by applying the data analysis mantra of overview, zoom and filter. These interaction patterns are implemented using information architecture components users are already familiar with but that are automatically generated from data and ontologies. This approach makes it possible to obtain an overview of the dataset being explored using navigation menus, treemaps or sitemaps. From there, users can interactively explore the data using facets. Moreover, facets also feature a pivoting operation, motivated during tests with lay users, that removes the main constraint of most faceted browsers, i.e. the inability to combine filters for different faceted views to build complex queries.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.5 [**Information Interfaces and Presentation**]: Miscellaneous

General Terms

Design, Experimentation, Human Factors

Keywords

Ontology, Semantic Web, Linked Data, metadata, human-computer interaction, usability, visualisation

1. NOVELTY

The amount semantic data available in the Web is rapidly increasing, for instance as part of the Linked Open Data cloud [1]. However, from the end-user perspective, the situation continues to be that the available datasets are monolithic and opaque files, which usually can just be explored using complex semantic queries or complex user interfaces. The objective is now to make this data more usable so that non Semantic Web experts can easily grasp what kind of entities are contained in a dataset, how they are interrelated, what are the main properties and values, etc. This will increase the awareness of the semantic data currently available on the Web and also facilitate the development of new and innovative applications on top of this data.

Rhizomer contribution draws from the experience accumulated in the Information Architecture (IA) domain [2], reusing and adapting existing IA components to provide browsing, exploration and visualisation guidance to users. Such IA components are well known to Web users, as they are present in most web pages: navigation bars, facets, sitemaps, breadcrumbs, etc. This approach is implemented in Rhizomer, a tool for publishing Semantic Web datasets while facilitating user awareness of the published content. It is also being evaluated with lay users as part of a User Centred Design development process. Iterative evaluations have motivated and guided the introduction of new features, like pivoting, and validated improvements in the context of a quality in use model [3].

Evaluations with users show the usefulness of an approach based on interface components that provide an overview of the explored dataset and faceted navigation, especially when dealing with highly structured data like Semantic Web data. Moreover, due to its richness, it also shows that it is fundamental to provide a pivoting operation. Facets are sufficient when the data model is simple, comprising a main type of resources described with a set of attributes and relations that are used to generate the facets. Similarly, facets are suited when the data is explored in a fragmented way, without requiring the combination of constraints on facets for different types of resources, i.e. different faceted views.

However, if full power to explore the data is required, for example, to express complex queries such as “actors from Spain, which have acted in films directed by Woody Allen”, it is necessary to be able to pivot from the Actors’ faceted view to the Films view. The proposed approach and its

implementation in Rhizomer, are among the few Semantic Web data exploration tools offering this functionality. Moreover, as our evaluation shows, Rhizomer provides the best user experience when compared to the two other main tools featuring pivoting [4], SParallax [1] and Virtuoso Facets [2].

2. INTENDED USERS

Rhizomer is based on applying existing information architecture components in the context of semantic data. They have been chosen because they are simple and very common so users are very comfortable using them. They are part of the "culture" about how information is presented in the Web so they can be easily learned. However, though they look like the common ones, these ones should be capable of giving access to the richer semantic data they are built on top of.

The aim is to make it possible that lay users, not just Semantic Web technologies experts, can reach the Semantic Web. This does not necessarily mean that any user will directly use these interfaces; usually it is about making it easier for application developers to discover and reuse semantic data while developing innovative products that ultimately spread the benefits of semantic data through the Web. For instance, using the Personas approach [5], we can illustrate the target audience as shown in Table 1.

flickr.com/photos/electricnerve

Christina Warren is a 23 years old journalist that is currently in charge of the Films section of an online journal. She likes to write about curious facts like "who appears most in films where Woody Allen is both the director and an actor". However, these kinds of things are really difficult to find out using resources like Wikipedia or IMDb.

Michael Harper is a 30 years old freelance developer that creates and commercialises mobile applications using online application stores. He is currently developing a mobile application that supports bird watching and as a way to reduce development costs to a minimum he is trying to reuse as much as possible available data about bird species, habitats, etc.

Table 1: Personas [5] illustrating the intended users

3. DEMO

Rhizomer demos for different datasets, like DBpedia or LinkedMDB, are available online[3]. It is important to note that all these demos are generated automatically just by deploying Rhizomer on top of the corresponding dataset, served using Jena, Virtuoso or OWLIM.

[1]SParallax, `http://sparallax.deri.ie`
[2]Virtuoso Facets, `http://www.openlinksw.com/dataspace/dav/wiki/Main/VirtuosoFacetsWebService`
[3]Rhizomer demos, `http://rhizomik.net/rhizomer`

4. KEY ASPECTS

Key aspects of Rhizomer are analysed in the next subsections from the human factors, computational models and application domains points of view.

4.1 Human factors

The overall user experience when interacting with semantic data using Rhizomer has been evaluated using the Semantic Web Exploration Tools - Quality in Use Model (SWET-QUM) [3]. The model proposes a set of quality factors to take into account and how to measure them through a set of metrics, detailed in [4].

SWET-QUM has guided a comparative evaluation of three semantic data exploration tools featuring facets and pivoting: Rhizomer, SParallax and Virtuoso Facets. Moreover, it is important to note that SWET-QUM is also being used to guide the User Centred Design (UCD) development process of Rhizomer.

Overall, 19 users were involved during the two last evaluations, all of them in the lay-user profile. 3 groups were defined: 6 for the Rhizomer pre-pivoting test, 7 users for the Rhizomer post-pivoting test and 6 users to test SParallax and Virtuoso Facets. None of the users received any a priori training about the evaluated tools, to try to mimic as much as possible the situation when the user arrives at the tool by browsing the Web.

The test facilitator proposed the following tasks to the users, to be completed using Rhizomer deployed on top of a relatively big dataset, LinkedMDB[4], which includes, among other entities, 85,620 films:

- **Task 1** - "Find three films where Clint Eastwood is director and also actor".
- **Task 2** - "Who has directed more films in countries located in Oceania?".

The objective of Task 1 was to evaluate a task that did not require pivoting and could be completed using only facets. However, pivoting can facilitate completing it producing a more efficient interaction. As shown in the left part of Fig. 1, the time to complete Task 1 with Rhizomer is smaller than for Virtuoso Facets and SParallax. Consequently, it is more efficient. On the other hand, Task 2 is more complex and it can only be solved using pivoting. Rhizomer is also more efficient compared with SParallax and Virtuoso Facets, as shown in the right part of Fig. 1.

4.2 Computational models

Rhizomer starting point is the fundamental set of tasks for data analysis proposed by Shneiderman [6]. In the following, we present each task associated with the chosen interaction pattern and Information Architecture component [2] to implement the pattern:

- **Overview**: obtain a full picture of the data set at hand. At this stage we propose to apply the Global Navigation interaction pattern or the Directory Navigation pattern. In Information Architecture terms, the

[4]LinkedMDB, `http://www.linkedmdb.org`

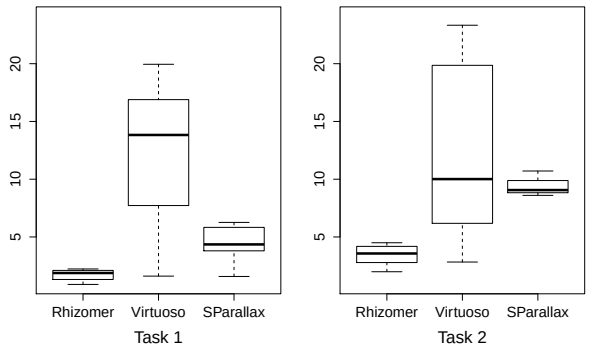

Figure 1: Comparing Rhizomer, Virtuoso Facets and SParallax efficiencies using the Task Time metric.

former can be implemented using navigation menus users are used to see at the top or on the left of web sites. The later sitemaps that can be represented using text or graphically, for instance using TreeMaps [7].

- **Zoom and Filter**: zoom in on items of interest and filter out uninteresting items. Here the proposal is to employ some form of Faceted Navigation. Once we have zoomed in by selecting from the navigation bar the kind of things we are interested in, facets are the Information Architecture components that help users to filter out those data items that are not interesting to them.
- **Details**: after zooming and filtering the user arrives at concrete resources of interest. At this point, the user can obtain details for those resources, which in the case of Semantic Web data means to retrieve properties for the resources plus those properties pointing to them. This is related to the Details on Demand interaction pattern and can be implemented via a simple list of properties and values of the resource of interest or through a specific visualisation tailored to the kind of resource at hand, e.g. a map for geo-located resources.

The computational models generated to implement the Inforamtion Architecture components giving support to these tasks are detailed in Section 5.

4.3 Applications Domains

The main strengh of Rhizomer is that it is independent from the application domain. A basic set of components is generated automatically from the dataset structure, taking into account both ontologies, schemas or thesaurus and the actual triples in the dataset. This basic set of components allows exploring the data through overviews and facets that provide filtering and pivoting. Moreover, as it is described in the next section, Rhizomer provides extension mechanism like visualisation plugins that can be associated to specific resources, like a map view for any resource with the *lat* and *long* properties.

5. ARCHITECTURE AND KEY COMPONENTS OF THE SYSTEM

The Rhizomer information architecture components are organised in its user interface as highlighted in Fig. 2. In addition to this components, Rhizomer also features automatically generated site map, site index and treemap that support the overview task. For the details task, Rhizomer provides specialised visualisations that depend on the resources characteristics and that are automatically attached to them.

All the Information Architecture components giving support to the previous tasks are automatically generated based on the underlying semantic data structure the first time Rhizomer is launched on a dataset. For the overview components, a tree model is generated from the underlying ontologies and thesaurus structuring the data. Then, instantiation data is also gathered using SPARQL queries so the overview components reflect the actual data. For instance, less relevant or no navigation menu entries are generated for classes with little or no instances. The same applies for TreeMaps, where the size of each class depends on the number of instances.

Then, for each class, the properties used to describe it are analysed in order to generate the facets model. However, not just ontologies and schemas are considered, Rhizomer also analyses how data models are instantiated and how resources are actually related through properties. This makes the facets model more robust to datasets without strict domain and range restrictions, which is actually the most common situation. Consequently, the facets model is generated by querying for all instances of a given class, which are the properties actually used to describe them and the more common values.

Moreover, the facets model is also enriched with "range" information for each class and property to facilitate the pivoting operation. SPARQL queries are used to check the most specific class shared by the facet values so it is possible to switch to the corresponding class faceted view when pivoting through that property. For instance, the values of the *starring* property for instances of the DBpedia class *Film* are analysed and as it is detected that all them are instances of the class *Agent*, it is possible to pivot to the *Agent* facets view when pivoting through that property from *Film*. More deatils are available from [8].

Once users have filtered the resources they are interested in, they can obtain more details about them. For instance, there is a map visualisation that is associated to any resource featuring the *lat/long* or geographic point properties. Similarly, the timeline visualisation is associated to resources with a date, start date or end date. Finally, charts are available for any property with numeric values.

6. CONCLUSIONS AND FUTURE WORK

After some rounds of development and testing with end-users, Rhizomer is capable of publishing semantic data while facilitating user awareness of what information is contained in the dataset. Awareness is accomplished by components borrowed from the Information Architecture discipline and generated automatically from the dataset structure and ontologies.

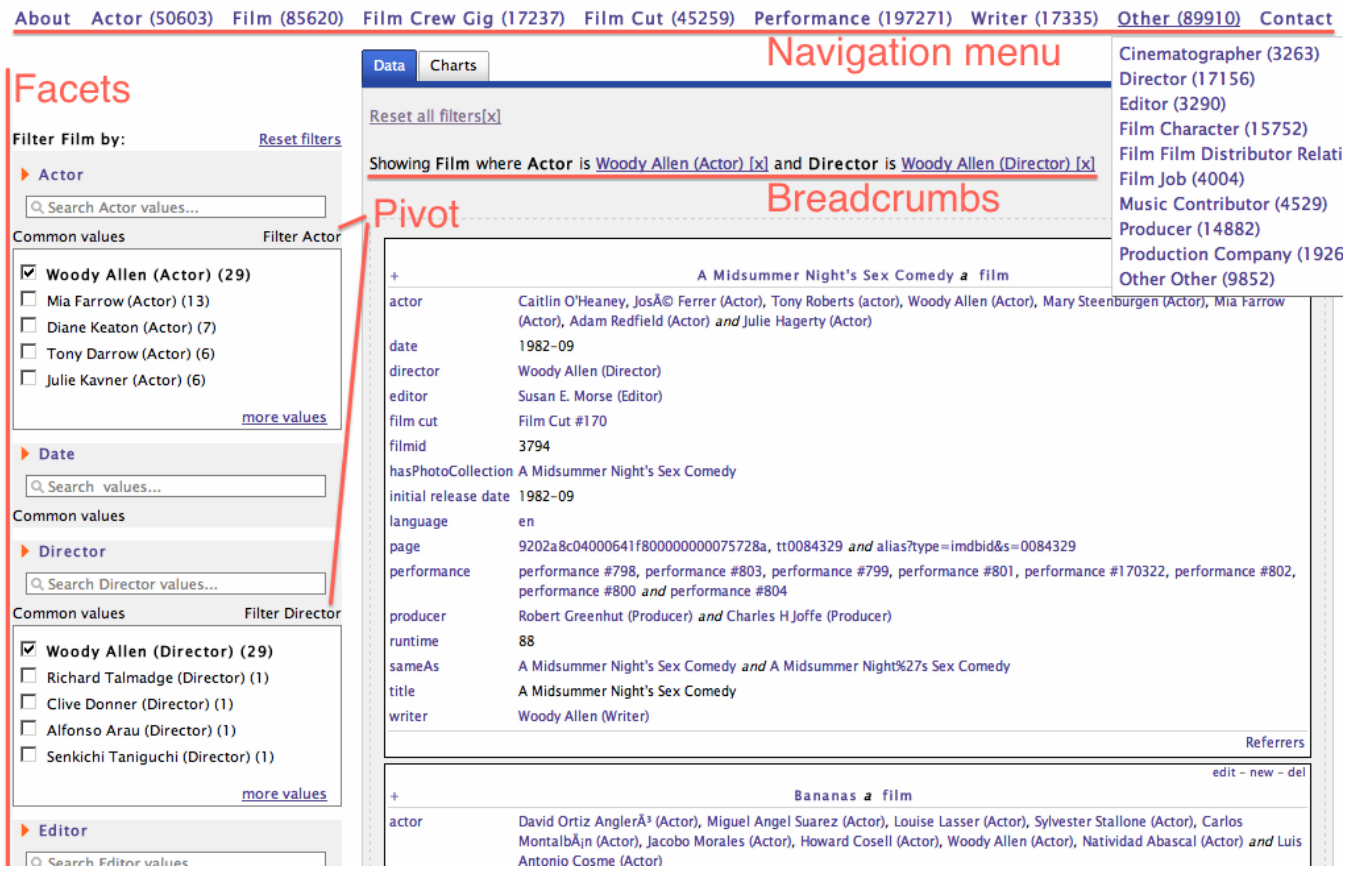

Figure 2: Information architecture components in the Rhizomer user interface.

Tests with users have shown that Rhizomer facilitates the exploration of the published datasets, and, among other quality in use factors, improves efficiency when compared to other semantic data exploration tools like SParallax and Virtuoso Facets. All them feature the pivoting operation, which empowers users to build much more complex queries to explore the data, without requiring knowledge of semantic query languages or the vocabularies used in the datasets.

Future work is to continue developing Rhizomer following the User Centred Design approach described. In addition, we plan to focus on improving the performance of the proposed approach, optimising the SPARQL queries and looking for alternatives like pre-computed indexes.

7. ACKNOWLEDGMENTS

Rhizomer has been partially supported by the research projects MediaMixer (EU FP7-318101) and InDAGuS (Spanish Government TIN2012-37826-C02).

8. REFERENCES

[1] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data - the story so far," *International Journal on Semantic Web and Information Systems*, vol. 5, no. 3, pp. 1–22, 2009.

[2] P. Morville and L. Rosenfeld, *Information Architecture for the World Wide Web*. Cambridge, MA: O'Reilly Media, 3rd ed. ed., 2006.

[3] J. L. González, R. García, J. M. Brunetti, R. Gil, and J. M. Gimeno, "SWET-QUM: a quality in use extension model for semantic web exploration tools," in *Proceedings of the 13th International Conference on Interacción Persona-Ordenador*, INTERACCION'12, (New York, NY, USA), pp. 15:1–15:8, ACM, 2012.

[4] J. M. Brunetti, R. García, R. Gil, J. M. Gimeno, and J. L. González, "Evaluating the quality in use of semantic web exploration tools featuring facets and pivoting," tech. rep., GRIHO Research Group, Universitat de Lleida, Lleida, Spain, 2012. `http://rhizomik.net/swet-qum/`.

[5] J. J. Garrett, *The Elements of User Experience: User-Centered Design for the Web and Beyond*. Indianapolis, IN, United States: New Riders, 2 ed., 2010.

[6] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," in *Proceedings of the IEEE Symposium on Visual Languages*, (Boulder, CO, USA), pp. 336–343, IEEE, 1996.

[7] B. Shneiderman, "Tree visualization with tree-maps: 2-d space-filling approach," *ACM Trans. Graph.*, vol. 11, no. 1, pp. 92–99, 1992.

[8] J. M. Brunetti, R. Gil, and R. García, "Facets and pivoting for flexible and usable linked data exploration," in *Interacting with Linked Data Workshop, ILD'12*, vol. 913, (Heraklion, Crete, Greece), pp. 22–35, CEUR Workshop Proceedings, 2012.