Interacting with Semantics: A User-Centered Visualization Adaptation Based on Semantics Data

Kawa Nazemi, Matthias Breyer, Jeanette Forster, Dirk Burkhardt, and Arjan Kuijper

Fraunhofer Institute for Computer Graphics Research, Fraunhoferstr. 5
64283 Darmstadt, Germany
{Kawa.Nazemi,Matthias.Breyer,Jeanette.Forster,
Dirk.Burkhardt,Arjan.Kuijper}@igd.fraunhofer.de

Abstract. Semantically annotated data gain more and more importance in future information acquiring processes. Especially the Linked Open Data (LOD) format has already experienced a great growth. However, the user-interfaces of web-applications mostly do not reflect the added value of semantics data. The following paper describes a new approach of user-centered data-adaptive semantics visualization, which makes use of the advantages of semantics data combined with an adaptive composition of information visualization techniques. It starts with a related work section, where existing LOD systems and information visualization techniques are described. After that, the new approach will bridge the gap between semantically annotated data (LOD) and information visualization and introduces a visualization system that adapts the composition of visualizations based on the underlying data structure. A case study of an example case will conclude this paper.

Keywords: linked open data, semantic visualization, semantic web.

1 Introduction

Semantic-Web and semantically annotated knowledge and information gain more and more importance in future information and knowledge acquiring processes. While formal descriptions of information, e.g. Ontology or Topic Map are still under the investigation of research institutions and enterprises, semantic structures based on the collective intelligence of web-users became silently an inherent part of the web. Especially the Linked Open Data (LOD) format has experienced a great growth in the open internet and became an established data model for conceptualizing knowledge entities and describing semantic relationships between knowledge entities and domains. The Linked Data format is not only used to model a specific domain by a small set of knowledge engineers, it is more a reflection of the knowledge interpretation of a whole community, which models domain-comprehensive knowledge for structure and disseminate it to a diversified audience. A single Linked Data database gains millions of knowledge entities per day and grows faster than expected.

Although the data in the LOD databases are semantically well-defined, the amount of data is more than sufficient and their structure provides the opportunity for the

M.J. Smith, G. Salvendy (Eds.): Human Interface, Part I, HCII 2011, LNCS 6771, pp. 239–248, 2011.

usage of alternative knowledge-acquisition and interaction with semantics, today's user interfaces of Linked Open Data do not really evince an added value to existing search-result user interfaces. The presentation of the information is often categorized in a concept hierarchy, but the presentation of information is principally text-based.

The already existing trend of alternative interaction ways with computer and information systems is not considered sufficiently in the presentation process of the underlying information. To provide a more natural interaction with the semantic knowledge in the structured LOD databases a graphical representation of the knowledge is valuable. Graphical objects are more similar to the objects of the real world and therefore provide a more intuitive interaction with the semantics structures.

Existing semantics visualization techniques do not consider the surpluses of the Linked Open Data structures, where the semantics structure has to be built-up with a routine of query requests. They focus on various but specific ontology characteristics, e.g. displaying the hierarchical inheritance structure, multiple inheritance or semantic relations between ontology entities. The complex structure of the Linked Data varies, based on the users' query on the data. The heterogeneity of the requested data should be exploited for the visualization and hence enable a more efficient interaction with the underlying semantics.

In this paper we describe a new approach of user-centred data-adaptive semantics visualization for Linked Open Data. The main innovation of our approach is the autonomous selection of different and various visualization types for the results of users' queries. Therefore the results of a user query are analyzed based on different criterions of the data characteristics, e.g. amount of result entities, amount of related entities, amount of related classes and domains or hierarchical structure of the underlying domains. We will show that each result of a semantic user query requires different types of visualization. Complementing our previous work the composition of different visualization types for the visualization of Linked Open Data will be introduced. A classification of the semantic search-result-types will open new possibilities for visualizing and interacting with semantics data. Based on the introduced classification different visualization types and their combination and composition respectively will be introduced. A main aspect of our approach is the interaction with graphical representation of the semantic entities and their structure, which supports the explorative information acquisition process of users and the interaction with alternative devices, especially gesture-based interaction systems.

The paper begins with a related work section, where existing Linked Open Data systems, the amount of their data and their interaction possibilities are analyzed. Supplementary existing semantics visualization techniques are introduced and the interaction possibilities and information acquisition processes are analyzed. In the following section a systematic classification of possible semantic search results will be proceed, where the different characteristics of Linked Open Data are classified especially for their visualization. In comparison to the classification of the semantic attributes the characteristics of information visualization techniques will be introduced and classified. The main section of this paper will bring the Linked Open Data and the visualization characteristics together and introduces a visualization system that adapts the visualization and user interface respectively based on the data under the consideration of users' information acquisition abilities. The paper concludes with an exemplary case study, where different semantic search queries and their visualization will argue the surpluses of our approach.

2 Related Work

The fast growth of the Linked Open Data community has led to a massive amount of interlinked, semantic annotated data on the web, which is called the Web of Data [1]. According to the 'Linked Data Principles' by Berners-Lee [2] a lot of open datasets developed and joined the LOD-Community. To use this available massive amount of semantic data, new powerful ways are needed to visualize the semantic information included in the Web of Data and to access the full potential of this new, machine readable data sources. To address this issue, we first give an overview of existing LOD-Systems and in the following introduce two existing semantics visualization systems and their capability to deal with the semantic information in LOD-Systems.

2.1 LOD Databases

This sections describes and analyze two of the most important community based LOD-Databases in the Linking Open Data Cloud[1]. First we will introduce DBpedia[2], a database which was generated using the open data from the web encyclopedia Wikipedia[3] and evolves into a central hub in the LOD-Cloud. In the following we will describe and analyze Freebase[4], a collaborative open database created as a wiki of structured data. To present an overview of existing LOD-Systems, we conclude with a table of common open Datasets, their amount of data and interlinking capability.

DBpedia. DBpedia is one of the most important and fastest growing LOD-Systems in the web of data. It is a community project to extract data from Wikipedia, structure it and make it accessible and interlinked with other datasets. Its great variety of topics led to intersects with many of the available open Datasets, so that it evolves to an excellent interlinking hub in the LOD-Cloud [3].

This fast growing is due to the close bonds between DBpedia and Wikipedia. While Wikipedia grows fast through the effort of its large community, DBpedia also benefits from this new data. Every new or updated article in Wikipedia can be used to extract structured data for the semantic database with help of the DBpedia knowledge extraction framework [4]. In January 2011 the DBpedia datasets consist of over 3.5 million entities which are connected by over 672 million RDF-Triples [5].

While the search-capability of Wikipedia is limited to full-text search, the semantic database DBpedia offers new ways to find useful information. The Faceted Wikipedia Search[5] Application, which builds upon the DBpedia database, can handle semantic search queries like: 'Actor of the movie The Lord of the Rings born in Berlin', by using the semantic structure of the data. This enables the user to specify his search

[1] http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
[2] http://dbpedia.org
[3] http://www.wikipedia.org
[4] http://www.freebase.com
[5] http://dbpedia.neofonie.de/browse

query to find exactly the information he is looking for without browsing through hundreds of search results.

Freebase is a collaboratively created open database, to structure information from general human knowledge collected by a large community. Unlike DBpedia, users can directly edit and change the data and the structuring schema [6]. This means users by themselves can develop new types, categories and overall domains to structure the data they supply. This freedom led to a large community and accordingly to an enormous grows of thousands of new facts a week. Freebase was founded by Metaweb Technologies, Inc. in 2007 and consists already of over 20 million entities[6]. 2010 Google acquired Metaweb and the Freebase-Project to improve Google-Search with better support of semantic user queries by taking advantage of the freebase semantic dataset [7].

Overview of LOD-Datasets. In the following table we show an overview of well-known semantic datasets in the LOD-Cloud. The data shows a comparison in amount of data as RDF-Triples, the interlinking (both in and out coming links) and the domain of the datasets. Even though DBpedia does not include the most data, you can clearly see that it takes the major role in the interlinking process. This table also shows how important LOD became and how much data has been collected already.

Table 1. Overview of LOD-Databases (2009-2010) [8, 9]. Interlinking: low >10.000, medium >1 million, central >1 billion extern RDF-links.

Databases	Amount of Data	Interlinking	Domain
LinkedGeoData	> 3000 million	low	geographic
DBpedia	> 400 million	central	multi-domain
GeneID	> 170 million	medium	bioinformatics
Freebase	> 100 million	medium	multi-domain
RDF Book Mashup	> 100 million	low	literature
Geonames	> 90 million	medium	geographic
Musicbrainz	> 60 million	medium	music

2.2 Semantics Visualization

Semantic visualization is not a new field, but dealing with the massive amount of semantic structure on the Web of Data raises new demands for visualizing the structure of the data, while facing the problems of information overload. In this section we introduce two semantic visualization systems, which both use the approach of combining different visualization techniques to improve overview of structure, relations and detail of the data and give a short introduction in our previous work of the combination of visualizations in a knowledge Cockpit [10, 11].

Knoocks: (Knowledge Blocks) [12, 13, 14] is a desktop-application for the visualization of ontologies, which combines several visualization techniques for semantic data. To display concepts of a specific ontology, Knoocks uses the so called 'Knowledge Block' approach. A block is used to visualize a concept and includes all of its

[6] According to the Freebase Explore-Page: http://www.freebase.com accessed: 01/2011.

subconcepts within the block-boundaries. Besides the block-visualization of the structure, you can also use the outlook-window and a node-linked visualization component to view the semantic relation of the data.

Thinkbase: [15, 16] is a visualization and exploration tool for the Freebase database, which uses the Freebase UI as part of its interface. It consists of two combined views of the same data, a graph-visualization of the relational structure and the text-based Freebase interface, which shows detailed properties of the topic of interest. As can be seen in figure 1, the graph-visualization window shows all instances, which are directly related to the topic of interest and combines all relations with the same role into so called 'aggregation nodes' [15] to provide a better overview. The second part of the interface is the Freebase-UI, which contains textual information, images and other properties and relations of the selected topic.

Also this combined visualization has some advantages in giving the user the possibility to understand the neighboring structure of the instance in focus and simultaneously provide information about its properties, the systems deals with problems in adaptability and lucidity. The user cannot choose to see more than the direct neighbors of the instance in focus and every time he navigates through the graph, it changes completely, so he has to adapt to the new structure in every step. Also the Freebase-UI is not designed to be a part of a combined visualization, so it shows too much data, needs a lot of scrolling and is not interacting well with the graph visualization.

Visualization Cockpit. As described in our previous work, the visualization Cockpit [10, 11] is a user-adaptable approach of visualization combination. It lets the user combine different visualization techniques and representation details (like color, icons, depth of relational structure) into a personal 'knowledge Cockpit'. Thus he can display the same data in different visualization techniques, which focus on different aspects of the data to improve a deeper understanding of the whole semantics structure. The user may for example combine the SeMap-Visualization [17] with a graph-like visualization to combine the advantages of the quick and easy overview capability of SeMap with the more detailed, but complex visualization of the graph.

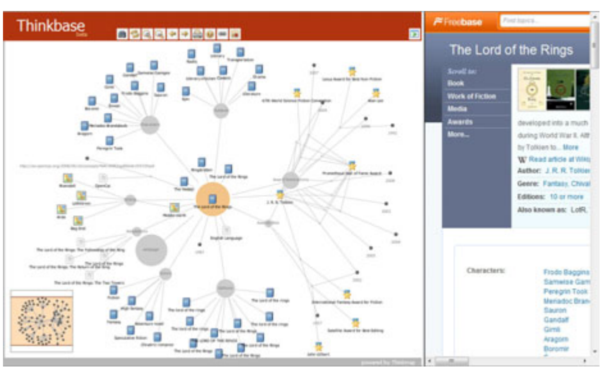

Fig. 1. Thinkbase visualization of the topic 'The Lords of the Rings'

3 Conceptual Model

After introducing related work on visualization of LOD, we will describe our approach to overcome some of the disadvantages of these technologies. First we will describe our approach to systematic classify search results in LOD-Databases. Supplementary we will point out the weak points of common visualization techniques and finally introduce the idea of combining visualizations according to the underlying data, which varies with the results of user queries.

3.1 Extraction and Organization of Linked Open Data

The semantic structure of linked open data is that of an ontology, which consist of concepts, sub-concepts and instances, commonly available through RDF-Triples. This structure provides useful information in grouping the data into categories and sub-categories. Our approach includes the improvement of user search by making use of this preexisting categorization, to help users define their search query more precisely.

When a user defines a search query, most of the time he uses only one or a few words. The result commonly consists of a large set of instances, which correspond to the users search query with different relevance. The common view of the user is a list of results sorted by relevance, visualized by their names, maybe a short description or picture. With help of semantics data, we can improve the visualization of search results to help the user to further specialize his query. The existing categorization of the data can be used to structure the search results within these categories, so that the user can choose high order categories to decrease the number of relevant search results.

To make use of the semantic categorization, the system has to perform a number of queries against the database, which are sketched out in Figure 2. First the user query is passed to the LOD-Database to get a number of relevant instances (likely this number will be very large due to the massive amount of data). If the number of instances is too big to show in a list, the system performs a query to get ancestor-concepts of these instances. This procedure can be repeated until a reasonable number of categories are found, a breakpoint is reached or the hierarchical structure ends. We will end up having a set of categories which cluster the search results in reasonable sized groups and support the user in finding the information he is looking for.

3.2 Visualization of Semantics Data

Semantics data consist of a set of different and varying components. We have a hierarchical structure, which we used in the previous section to generate a categorization of search results. The instances have relations between each other and contain additional information in form of properties. These properties consist of varying data types, which includes number, date, time, geological features, text and so on. Knowing this, it becomes clear that a visualization of all these different features will be a different task to perform.

Most visualization techniques specialize upon one feature of the above. This is because the visualization methods they use have advantages for a special feature, but disadvantages for others. We can easily show the relations between instances in

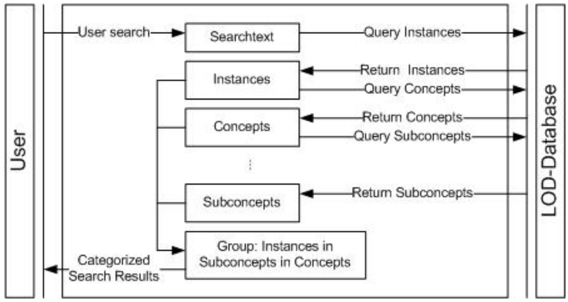

Fig. 2. Extraction of semantic data classification for search results

a graph-visualization, which provides interaction methods for expanding or collapsing a node to gain a better overview, but we can hardly display an article text, a picture or properties like geological or time related data. On the other hand can a map-based visualization be very comfortable to compare geological features of semantic data, but you cannot show the hierarchical structure of the data.

This raises the idea of combining different visualization techniques, as we pointed out in our previous work [10, 11]. To show all components of semantics data, a combination of visualizations is used to overcome the disadvantages of the individual systems by letting the user choose to look at the most suitable visualization for his means. Thinkbase also uses this promising approach by combining a graph-like view to show the structure of the data side by side with the Freebase-UI, to bring up the textual components. But there is still room for improvement. The graph-based view shows only the relations between instances, but cannot display the concept-hierarchy of the data, which can provide useful information in clustering and categorizing data. Also the animation of the graph is not suitable to let the user follow his navigation. As we mentioned before, the graph shows only the direct neighbors of an object and cannot vary the depth as the user likes. It also does not have additional visualization methods the user could choose from and the Freebase-UI is much too crowded to be one part of a combined visualization tool.

3.3 User Centered Adaption Based on Semantics Data

Our main approach is to combine the ideas of categorizing search results and combined visualizations for a more suitable representation of semantics data. As we pointed out in the previous sections, Thinkbase and Knoocks both follow the approach of having different visualizations combined. But both only provide static, predefined visualizations, which should cover the most common datasets. In our previous work we described the idea of letting the user select his combination of visualizations directly by himself [10, 11]. This approach requires a certain kind of understanding in the underlying data and may not be suitable for inexperienced users. The main improvement of our approach is the automatic, dynamic selection and combination of visualizations, which are most suitable for the dataset currently in focus. This means the system has to analyze the data, structure it and choose a promising set of visualization techniques, which present the elements of the data they are adapted to.

The first step to do this is structure the data which returns as result from a user query. In the previous section we showed a method of categorizing this data into their concepts and sub-concepts to provide a better overview of the returned results. But semantics data contain more than only hierarchical order. Semantic relations and properties of the instance in focus can easily be queried on demand.

After the data is structured, the system has to decide on the visualization composition. Following the processing of the user query, we give an overview of the returned instances. The visualization of the search results is dependent on the number of instances, the number of categories and the number of instances grouped in these categories. Commonly there will be a lot of instances which respond to the search query. In this case, an unstructured list of all results would be a poor choice. The system will instead choose to categorize the instances as we described above and add a simple hierarchical visualization like SeMap [17], which will provide easy to use and understand access to the categorized search results. In contrary if there are only a few instances, there is no need to further categorize the data and it can be shown as an ordered list.

The next step is to visualize the instance in focus. Before any user interaction, this will be the most relevant instance from the search results, which in many cases is the one the user is looking for. The system first has to analyze the information contained in the instance of focus, which could be number of semantic relations or the type of properties it contains. As an example the instance could be a famous person, so it may contain properties like age or weight, which will be displayed using a tabular structure. It may also contain textual descriptions, pictures or links to the Wikipedia entry, visualized in a textual content window. Time dependent data like birthday or death will be displayed in a timeline-view and any geographical data in a map-based visualization. The system will analyze all this information and combines the different visualization techniques into one (further adaptable) knowledge cockpit. According to the number of different visualizations the system may choose to display only the most relevant views and hide the others, giving the user the possibility to expand the visualization he needs.

4 Case Study

In this section, we will present a comparative view of an example search query and user interaction in Thinkbase and our data-adaptive system. We will show the advantages of our approach analyze the differences between the two systems and discuss the choices our system made to display data returned from the search query.

We start with the search query: ‘Moon’ with the intention to find out something about the novel ‘New Moon’ (assuming the user doesn’t know the correct title). Thinkbase first shows a list of the first 10 instances which correspond to the search query, ordered by relevance and displayed with their names and domain. The user has to scan through the list and expand it two times to find ‘New Moon (Book)’. In our system, the search results are categorized into not more than 10 domains, from which one is Literature and displayed using SeMap [17]. If the user expands Literature, he finds the instance ‘New Moon’ within a short list of books with moon in title (sorted by relevance).

The next step is the selection of the instance ‘New Moon’. Thinkbase builds up a graph-view of directly related instances and shows the ‘aggregation nodes’: Characters, Genres, Editions, Subjects and Webpage. You can also extract the

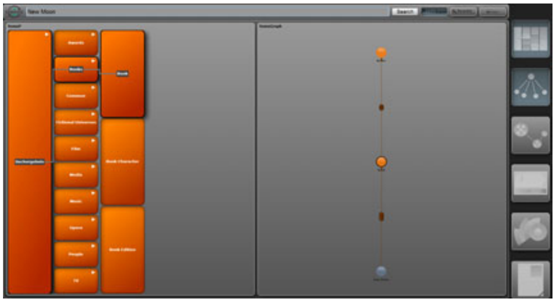

Fig. 3. Adaptive Visualization of Search Result

author, publication date and series the book belongs to. The Freebase-UI shows a short text and picture and contains as well all information, which is also part of the graph visualization (e.g. author, characters, editions …). Our data-adaptive system displays a similar graph-visualization to show the related objects, but offers the user the possibility to adjust the depth of the graph. It also adds a timeline-visualization, to show the publication date of the book alongside with the publication dates of the related books and release dates of the related films. Additionally, a text-based visualization shows an article text, a picture and links to Wikipedia.

In this step, you can clearly see the main difference between the two systems. While Thinkbase displays the same information in two different views, duplicating the majority of the data, our system clearly separates the information, displaying only the data the specific visualization is adapted to and combines them to an overall view of the semantic information.

5 Conclusion

Considering the fast development of the Web of Data it is clear that semantic annotated data is already an important part of the internet. While the amount of open data is steadily growing, the user interfaces still cannot reflect the whole power of semantics data. New visualization techniques are needed, to take advantage of the information, contained in semantic relations and structure.

In this paper we improved our previous work on user-centered combination of different visualization techniques to a new visualization system, which analyses, structures and visualizes semantic data automatically, according to the structure of the underlying data. We pointed out how search results can be categorized according to the preexisting hierarchical structure in Linked Open Data and how this could improve the information acquiring process of common users. To reduce the amount of information in a single visualization, the data is clearly separated into pieces and visualized by a combination of data-adaptive selected visualization methods.

To improve the user-centered design, further work should include user models, to adapt the combination of visualizations to users liking.

Acknowledgments. The here described work was developed as a part of the Core-Technology-Cluster for Innovative User Interfaces and Visualizations of the THESEUS Program, a 60-month program partially funded by the German Federal Ministry of Economics and Technology.

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story so Far. International Journal on Semantic Web and Information Systems 5(3), 1–22 (2009)
2. Berners-Lee, T.: Linked Data. International Journal on Semantic Web and Information Systems 4(2), W3C (2006)
3. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A Nucleus for a Web of Open Data. In: The Semantic Web, pp. 722–735. Springer, Heidelberg (2007)
4. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia - A crystallization point for the Web of Data. In: Web Semantics: Science, Services and Agents on the World Wide Web. Elsevier, Amsterdam (2009)
5. Bizer, C.: DBpedia 3.6 released (2011), `http://blog.dbpedia.org/2011/01/17/dbpedia-36-released` (online accessed: January 20, 2011)
6. Bollacker, K., Tufts, P., Pierce, T., Cook, R.: A Platform for Scalable, Collaborative, Structured Information Integration. In: Sixth International Workshop on Information Integration on the Web, AAAI (2007)
7. Menzel, J.: Deeper understanding with Metaweb (2010) `http://googleblog.blogspot.com/2010/07/deeper-understanding-with-metaweb.html` (online accessed: January 24, 2011)
8. Statistics on links between Data sets (2009), `http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/DataSets/LinkStatistics` (online accessed January 26, 2011)
9. Statistics on Data sets, `http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/DataSets/Statistics` (online accessed January 26, 2011)
10. Nazemi, K., Breyer, M., Burkhardt, D., Fellner, D.W.: Visualization Cockpit: Orchestration of Multiple Visualizations for Knowledge-Exploration. International Journal of Advanced Corporate Learning (IJAC) 3(4) (2010)
11. Nazemi, K., Breyer, M., Burkhardt, D., Fellner, D.W.: Semantic Visualization Cockpit: Adaptable Composition of Semanitcs-Visualization Techniques for Knowledge-Exploration. In: Conference ICL 2010 (2010)
12. Kriglstein, S., Motschnig-Pitrik, R.: Knoocks: New Visualization Approach for Ontologies. In: 12th International Conference on Information Visualisation, IV 2008, pp. 163–168. IEEE, Los Alamitos (2008)
13. Kriglstein, S., Wallner, G.: Knoocks - A Visualization Approach for OWL Lite Ontologies. In: 2010 International Conference on Complex, Intelligent and Software Intensive Systems (CISIS), pp. 950–955. IEEE, Los Alamitos (2010)
14. Kriglstein, S.: OWL Ontology Visualization: Graphical Representations of Properties on the Instance Level. In: 2010 14th International Conference on Information Visualisation (IV), pp. 92–97 (2010)
15. Hirsch, C., Grundy, J., Hosking, J.: Thinkbase: A Visual Semantic Wiki. In: Demo Session of the 7th International Semantic Web Conference (2008)
16. Hirsch, C., Hosking, J., Grundy, J.: Thinkbase: Interactive Visualization Tools for exploring the Semantic Graph of large Knowledge Spaces. In: Workshop on Visual Interfaces to the Social and the Semantic Web, VISSW 2009 (2009)
17. Nazemi, K., Breyer, M., Hornung, C.: SeMap: A concept for the visualization of semantics as maps. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5616, pp. 83–91. Springer, Heidelberg (2009)