

# Semantic Web and Information Visualization

Riccardo Albertoni, Alessio Bertone, Monica De Martino  
*Istituto di Matematica Applicata e Tecnologie Informatiche*  
*Consiglio Nazionale delle Ricerche*  
*{albertoni,bertone,demartino}@ge.imati.cnr.it*

## Abstract

*The paper augments the results achieved in the research of Semantic Web technologies for semantic search application, underlining the importance of integrating Information Visualization into semantic web technologies.*

*We present the results of a preliminary investigation of user needs in the activity of information search and an overview of the most used visualization tools in the web and tools for the ontology visualization. Our goal is to identify different types of information search problems and to demonstrate the advantages given by the adoption of information visualization.*

## 1. Introduction

The Semantic Web is an extension of the actual World Wide Web (WWW) that aims to make the web content machine understandable, allowing agents and applications to access a variety of heterogeneous resources. In particular, it has been proposed to deal with problems such as information overload and info-smog that are responsible for the “lost on the net” effect and make the web content inaccessible.

Semantic search is an application of the Semantic Web designed to improve traditional web searching. It is based on an explicit representation of semantics about web resources and real world object and aims to improve both the proportion of relevant material actually retrieved (recall of results) and the proportion of retrieved material that is actually relevant (precision of results).

Two of the main aspects, which need to be considered to make the semantic search effective, are the amount of resources contained in the Semantic Web and the needs of the seeker during the search process. Regarding the former, the user will rapidly

have to face with huge amount of data because the Semantic Web represents an extension of the actual WWW and it is supposed to grow up as fast as the web. Regarding the latter, a highly interactive process characterizes any search activity in the Web as well as Semantic Web. The seeker needs to alternate phases where he queries the Semantic Web to phases where he browses the Semantic Web content and he often needs to refine his selection criteria according to the obtained results.

Information Visualization joined to the Semantic Web technology may represent a solution to support the seeker both to manage the large amount of results and to perform the interactive process. As Chaomei Chen writes, “Semantic Web emphasizes that data should be machine readable, whereas information visualization aims to maximize our perceptual and cognitive abilities to make sense of visual-spatial representations”. However, due to the complementary natures of these two topic areas, there are still some problems to join the capabilities of Information Visualization and Semantic Web: on the one hand, the Semantic Web strives to make the information more structured and machine-like, while on the other, information visualization strives to make the information more accessible and less structured to improve usability. In the context of Semantic Web, the term “visualization” is often used in a broader sense: it mainly refers to the appearance of the textual information and it does not necessary include any information visualization techniques.

In this paper we want to argue the advantages coming from the adoption of Information Visualization to improve semantic search applications. In particular the paper intends:

- To introduce general issues in the information search activity and to highlight seeker needs in the search process.
- To provide an overview about web and ontology visualization.

- To summarize the limits of the existing approaches and to discuss the advantages coming from the adoption of Information Visualization techniques in the search process and in the Semantic Web context.

## 2. Seeker needs

This paragraph provides an overview of the problems that may occur when the user searches for information in the Web. The typical Information Retrieval (IR) systems now available are characterized by a representation of a request for information (query) and the system usually responds with a set of results which most closely matches the request. Whatever representation of a request the seeker has to formulate, he has often to face with problems related to the clear specification of his information needs.

The Semantic Web provides increasingly sophisticated and more powerful inferences to sift intelligently through this large information space. But although it provides great benefits to the seeker, it does not satisfy all his needs.

In the following we discuss some problems outlined by Spoerri in [1] which are intrinsically related to the information searching activity and problems which come up because the search is performed in an open environment<sup>1</sup> such as the WWW or the Semantic Web context.

### 2.1. Problems of information searching activity

There are some human factors that affect the results of information searching activity:

- Seeker knowledge. The seeker has often only a perception of his information needs: he has a limited knowledge of what he is looking for.
- Seeker and provider relationship. Seekers and information providers have different skill levels and different domain of knowledge. Moreover there is usually no direct interaction among them.
- Seeker anxiety. The gap between what the seeker understands and what he thinks he should understand generates anxiety. This happens whenever information does not fulfill his needs.

---

<sup>1</sup> The paper refers to the term open environment as a place where heterogeneous information are published by different providers without being coordinated by a central management service.

These factors may get seeker into some problems:

- Problem of *Query formulation/refinement*: “how to modify the query to find more relevant information?”. The limited knowledge the user has about his information needs has been widely discussed by Belkin and colleagues as a result of the Anomalous State of Knowledge (ASK) [2]. They argue that the need for new information arises whenever the knowledge needed to carry out an activity or to solve a problem is incomplete. In this case, the user decides to complete his knowledge by searching for new information. Although, in some cases, user knows what is looking for and he is able to specify the most suitable search criteria, in the most of the cases he is not able to precisely specify the information that he needs. The limits to properly specify the search criteria in the Anomalous State of Knowledge force the seeker to enter into dialogue with the IR system in order to find more relevant documents.
- Problem of *Vocabulary*: “Which term to use?”. The difference in terms of knowledge and perception between the information providers and the seeker has been modelled in terms of informative space and cognitive space. The former is defined as a set of object and relations among them held by the system whereas the latter is defined as a set of concepts and relations among them held by individual[3]. Information providers organize their resources according to their knowledge and to the vocabulary that concurs in building the “informative spaces”. If seekers have a different knowledge background, or a different purpose, then his cognitive space has a poor overlapping to the informative space. This make reasonable to assume they will use different terms to identify the same concept. So they have to discover which are the proper terms to express a query in the information space.

### 2.2. Problems in the open environment

The problems discussed above are common to all search activities. If the search is performed in a distributed and heterogeneous environment as web or Semantic Web, the search becomes harder: the seeker anxiety grows up according to the heterogeneity and the amount of information available in the Semantic Web. It generates the following problems:

- Problem of *database selection*: “which search engine to select?”. The problem is well known in the WWW because the actual search engines are able to cover a limited portion of the web

resources. The seeker has to decide which search engine to make use of. Also semantic web may be affected by similar problems since it will probably have different engines or web services, which will differ each other according to technology and conceptualization they rely on.

- Problem of *information overload*: “how to explore many retrieved documents?”. Semantic search promises to increase the quality of results by exploiting explicit representation of semantics about web resources and real world objects [4]. That will increase the number of relevant information sources returned by each query and will reduce the Info-smog problem. However, user still has to face with huge amount of candidates, which are all pertinent to what he is looking for. He needs to be supported in the analysis of heterogeneous information sources to be able to choose the most suitable ones for his purpose.
- Problem of *query coordination*. The seeker may need to be supported in the choice for queries. Human behavioral studies during the search activity have shown that the user is lazy and usually tends to create short queries and rarely adopts Boolean expression in his query criteria [5] [6]. Whenever the seeker needs information, which might seriously affect the results of his work, he is forced to a deeper search. Due to the ASK problem he needs to be supported both in the formulation of different queries and in their comparison.

In the following we illustrate the advantages given by Information Visualization in the solution of the mentioned problems.

### 3. Visualization for the Semantic Web

Information Visualization may support semantic search process providing functionalities to make the information more accessible and improving their usability.

Various visualizations have been developed to show the structure of an information space of the Web [1] and other more specific for the Semantic Web.

In the Web, Information Visualization provides visualization approaches to manage big amount of information in a summarized way and graphical interaction techniques to manipulate the search results. On the contrary, visualization in the Semantic Web is still an undefined concept: few works explicitly address the problem of visualization in the

Semantic Web, and most of them just address the ontology visualization. In the following an overview of some well-known visualization tools, which are available both for WWW visualization and Ontology Visualization, are illustrated. We aim to identify the most common functionalities and to argue how they support the seeker needs in the information search process.

#### 3.1. Visualization in the WWW

In this paragraph we want to show how the existing tools to browse the WWW adopt visualization to satisfy seeker needs. The study has been limited to some of the most well known tools such as Kartoo[7], Grokker[8], Web Theme[9], Aduna AutoFocus[10].

To achieve this purpose the following research activities have been performed:

- Identification of the main functionalities provided by these tools.
- Analysis both of the correlation among these functionalities and of the problems in the information search.

The results of these activities are summarised in Table 1 and Table 2.

Table 1 is the result of the first activity. It illustrates the association between the tools (columns) and some of their functionalities (rows). We have identified some heterogeneous functionalities: graphical visualization functionalities (Hierarchical Visualization, Clustering Visualization, Map Based Visualization), graphical interaction functionalities (Visualization Manipulation, Graphical Selection) and those functionalities that are a combination of them (Highlighting, Coloured Query Result, Filter Result Representation, “co-occurrent term interaction/visualization”). In the following, a description for each of them is provided:

- *Hierarchical Visualization*: the visualization represents its content according to different levels of granularity. This allows to browse the information at different levels of detail (as Grokker [8]).
- *Clustering Visualization*: the content is visualized (grouped) according to some similarity criteria. The groups can be obtained either by applying a clustering algorithm (galaxy view[9]) or according to properties specified by the user (cluster map [11]).
- *Map Based Visualization*: it imitates the geographical map appearance; the content is organized according to thematic terms or co-occurrence criteria, which are represented as

peaks in the map (i.e. Kartoo [7] represents the isograms and the name of the mountains respectively as concentric isolines and thematic terms on the top of them).

- *Visualization Manipulation*: the interaction between user and the graphical representation allows to re-organize the elements displayed, to move them and to add new ones (i.e. Grokker [8] and Kartoo [7], allow to add a new web site to the search and to insert it in the displayed graph according to user needs).
- *Graphical Selection*: the selection of a single (Grokker, Aduna AutoFocus, Kartoo) or many elements at a time allows the user to select different information source such as URI, PDF or DOC document in Grokker, Aduna Autofocus, Kartoo or data as in Web Theme[9].
- *Highlighting*: whenever an element of the visualization is selected, all the sources related to such element are highlighted too. Aduna AutoFocus and Kartoo allow to highlight the related co-occurrent terms, whereas Grokker permits the highlighting both of the related co-occurrent terms and of the related elements in the visualization.
- *Coloured Query Result*: Web Theme[9] allows to query the visualized data set and to set a particular colour to each result set. This facilitates the comparison among different queries (results).
- *Filter Results Representation*: some filters can be applied to the contents shown in the visualization. For instance, Grokker allows to filter on the rank, on the domain and on the source, whereas Kartoo allows to filter on the co-occurrent terms.
- *Co-Occurrent Terms Visualization*: As users tend to formulate their queries using common words, a statistical thesaurus expands these queries with other highly frequent terms that should help the user in discriminating relevant documents ([12]).

Table 2 is the result of the activities to identify the contribution of the functionalities to solve problems related to seeker needs. It is possible to argue that:

- Graphical visualization functionalities provide different results. They give a structured organization of information offering the user an overview of the available information relieving the *information overload problem*. They supports the query formulation/refinement: a correct and rapid understanding of search results is the prerequisite to have a successfully

query *refinement*. Graphical visualization functionalities provides useful hints to solve the *vocabulary problem* by map based and clustering visualizations. They show co-occurring terms as cluster representative or in map representation permitting to learn which terms belong to the information space and how terms are related to each other.

- Graphical interaction facilitates the information overload and query formulation /refinement problems: visualization manipulation supports in the analysis of results by modifying the layout, whereas graphical selection provides a visual and intuitive way to select results user is interested to.
- Finally, the functionalities based on the integration between interaction and visualization techniques support in the entire problem mentioned about. In particular, functionalities as Coloured Query Result allow to compare the results of different queries supporting in the *queries coordination problem*. Whenever the information about which search engines have found a result is maintained, such functionality can be exploited to compare the results coming from different search engines supporting the solution of *database problem*.

### 3.2. Ontology visualization

In this paragraph some of the tools proposed to visualize and interact with the ontology are analysed: OntoViz[13], TGViz[13], Jambalaya[14], Dope Browser[15].

The results of the analysis are illustrated in Table 3. Similarly to the previous discussions, the analysis is limited to a rough set of functionalities. Some of them are reprised from those previously given in Table 1 and Table 2, so in the following only a complete description of the new ones is provided:

- *Choice of the Hierarchy Level Shown*: the visualization layout can be modified according to the number of level to be displayed. TGViz allows to display from 0 to 10 levels of the hierarchy. This increases user knowledge during the ontology navigation, since he can explore the hierarchy at a different level of details.
- *Instance*: OntoViz, Jambalaya and Dope Browser visualize the instances of each selected class directly on the graph or on a side window. This feature is often related to the highlighting, that means that i.e. the graphical selection of an instance highlights both the instance and the related co-occurring terms.

		Grokker	Aduna AutoFocus	Kartoo	Web theme
Graphical Visualization	Hierarchical Visualization	☺	-	-	-
	Clustering Visualization	☺	☺	-	☺
	Map Based Visualization	-	-	☺	☺
Graphical Interaction	Visualization Manipulation	☺	☺	☺	-
	Graphical Selection	☺	☺	☺	☺
Interaction and Visualization	Highlighting	☺	-	-	-
	Colored Query Result	-	-	-	☺
	Filter Results Representation	☺	-	☺	-
	Co-Occurring Terms Interaction/Visualization	☺	☺	☺	-

**Table 1. Functionalities provided by some existing tools to browse the WWW**

		Database	Vocabulary	Query Formulation/ Refinement	Information Overload	Query Coordination
Graphical Visualization	Hierarchical Visualization	-	-	☺	☺	-
	Clustering Visualization	-	☺	☺	☺	-
	Map Based Visualization	-	☺	☺	☺	-
Graphical Interaction	Visualization Manipulation	-	-	☺	☺	-
	Graphical Selection	-	-	☺	-	-
Interaction and Visualization	Highlighting	-	-	-	☺	-
	Coloured Query Result	☺	-	-	-	☺
	Filter Results Representation	-	-	☺	☺	-
	Co-Occurring Terms Interaction/Visualization	-	☺	☺	-	-

**Table 2. Tools functionalities and how they satisfy seeker needs**

- *Ontology Graph Navigation*: an important characteristic when dealing with ontology is how to easily navigate the complicate structure it may assume. An animated navigation (Jambalaya) connected to the possibility to choose different layouts for the hierarchy, appears as a powerful solution to the problem. Other less animated solutions (Dope Browser, TGViz) are less impressive; nevertheless they reach their scope.
- *String Search*: The search for both a co-occurrent word and a string can help to navigate the hierarchy. All the selected tools but OntoViz provide this feature.  
In the following paragraph we will discuss how the provided considerations can improve the search for information in the Semantic Web.

		OntoViz	TGViz	Jambalaya	Dope Browser
Graphical Visualization	Hierarchical Visualization	☹	☹	☺	☺
	Clustering Visualization	-	-	☺	☺
Graphical Interaction	Visualization Manipulation	-	☺	☺	☺
	Graphical Selection	☺	☺	☺	☺
Interaction and Visualization	Highlighting	☺	☺	☺	☺
	Co-Occurring Terms Interaction/Visualization	-	☺	☺	☺
	Choice of Hierarchy Level shown	-	☺	-	-
	Instance	☹	☺	☺	☺
	Ontology Graph Navigation	-	☺	☺	☺
	String Search	-	☺	☺	☺

**Table 3. Functionalities provided by existing tools to browse ontology**

#### 4. Discussion and conclusion

We addressed our research activity to investigate how Information Visualization is joined to Semantic Web: our main goal is to design and realize a new Semantic Web tool based on visualization to support the seeker in the information search process.

The first activity to be performed to achieve this goal is to well identify seeker needs and information search problems and to have a clear overview of the main functionalities provided by existing tools based on Information Visualization paradigm both in the context of web browsing and in the ontology visualization.

This paper argues some problems occurring in information searching: the analysis of seeker needs shows that although semantic search promises an improvement in the quality of query results, it does not face with all of the problems. Some functionalities of Information Visualization tools applied to World Wide Web (Table 1) as well as to ontology visualization (Table 3) have been briefly analysed. The analysis demonstrates that Information Visualization helps the user in his search activity: for each mentioned problems there is at least an information visualization functionality designed to provide a support to the user (Table 2).

However, the application of Information Visualization techniques to facilitate information search in semantic web seems at the moment to be only marginal. Even if a more detailed analysis of ontology visualization tools is needed, it is possible to state that they mostly aim to show and edit the ontology structure rather than to support an intelligent browsing of the ontology instances. Although the Dope Browser differs from the others, since it focuses on the ontology instances and provides both an ontology structure and a thesaurus tool to refine the user search, it does not support all the mentioned problems (i.e. database problem).

A complete view of how to combine and adapt the information visualization techniques in the context of semantic web is still missing as well as how to take advantages of the semantic information in the usual information visualization techniques.

More detailed analysis of the available tools as well as their functionalities is work in progress. A homogeneous view of how to combine new Information Visualization techniques and Semantic Web technology to support a general user who does not have any knowledge about ontology and semantic web technological amenities should be developed.

## 5. References

- [1] A. Spoerri, "How Visual Query Tools Can Support Users Searching the Internet", *Proceedings of Eighth International Conference on Information Visualisation*, IEEE, 14-7-2004, pp. 329-334.
- [2] N. J. Belkin, N. R. Oddy, and M. H. Brooks, "ASK for Information Retrieval: Part I. Background and Theory", *Journal of Documentation*, vol. 38, no. 2, 1982, pp. 61-71.
- [3] G. B. Newby, "Cognitive space and information space", *Journal of the American Society for Information Science and Technology*, vol. 52, no. 12, 2001, pp. 1026-1048.
- [4] R. Guha, R. McCool, and E. Miller, "Semantic search", *Proceedings of the twelfth international conference on World Wide Web*, ACM Press, 2003, pp. 700-709.
- [5] A. Spink, D. Wolfram, M. B. J. Jansen, and T. Saracevic, "Searching the Web: the public and their queries", *J.Am.Soc.Inf.Sci.Technol.*, vol. 52, 2001, pp. 226-234.
- [6] N. J. Belkin and W. B. Croft, "Information filtering and information retrieval: two sides of the same coin?", *Commun.ACM*, vol. 35, 1992, pp. 29-38.
- [7] Kartoo, ver. 4, KARTOO S.A., <http://www.kartoo.com>, 30-10-2004.
- [8] Grokker, ver. 2.1, MIII Groxis, Inc., <http://www.grokker.com>, 2004.
- [9] M. A. Whiting and N. Cramer, "WebTheme (TM): Understanding Web information through visual analytics", *Lecture Notes in Computer Science*, vol. 2342, 2002, pp. 460-468.
- [10] AutoFocus, ver. 2004.1, ADUNA, <http://aduna.biz/products/autofocus/personal/index.html>, 2004.
- [11] C. Fluit and J. Wester, "Using visualization for information management tasks", *Proceedings of the Sixth International Conference on Information Visualisation*, IEEE, 2002, pp. 447-454.
- [12] H. J. Peat and P. Willett, "The Limitations of Term Cooccurrence Data for Query Expansion in Document-Retrieval Systems", *Journal of the American Society for Information Science*, vol. 42, no. 5, 1991, pp. 378-383.
- [13] OntoViz, Intelligence, Agents, Multimedia Electronics and Computer Science Dept. University of Southampton, <http://www.ecs.soton.ac.uk/~ha/TGVizTab/TGVizTab.htm>, 7-6-2004.
- [14] Jambalaya, ver. 2, CHISEL-Computer Human Interaction and Software Engineering Lab, University of Victoria, <http://www.thechiselgroup.org/jambalaya>, 11-6-2004.
- [15] H. Stuckenschmidt, F. van Harmelen, A. de Waard, T. Scerri, R. Bhogal, J. van Buel, I. Crowlesmith, C. Fluit, A. Kampman, J. Broekstra, and E. van Mulligen, "Exploring large document repositories with RDF technology: the DOPE project", *Intelligent Systems, IEEE*, vol. 19, no. 3, 2004, pp. 34-40.