Information Search: the Challenge of Integrating Information Visualization and Semantic Web

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 examines the potentialities offered by Information Visualization to improve information search in the Semantic Web. In particular, the paper aims to investigate user problems in the information searching activity and to demonstrate that Information Visualization is effective to solve these problems even if it has not been yet properly adopted in the Semantic Web.

1. Introduction

The Semantic Web (SW) is rising as an extension of the current Web to improve the accessibility of web content providing sophisticated and powerful inferences to sift intelligently through this large information space. In particular, Semantic Search [1] is emerging as the application of the SW is designed to improve the search in the Web, since it relies on an explicit representation of semantics about web resources and real world objects. Nevertheless Semantic Search does not exhaustively solve all the problems related to the search. Firstly, every searching activity in the WWW as well as in the SW is characterized by a highly interactive process. The seeker needs to refine his selection criteria according to the obtained results alternating phases where he queries the SW to phases where he browses the SW content. Secondly, since anyone can publish new information sources in the SW, the information sources resulted by a user's query might be redundant or represent different points of view. Even supposing that Semantic Search is able to improve the quality of results, there are stages in the search activity where only the seeker can decide which sources to discharge. In this context Information Visualization (IV) comes into the scene. According with the definition of Card in [2], "Information Visualization is the use of computer-supported, interactive visual representations of abstract data to amplify cognition". Information Visualization (IV) provides tools both to perform the interactive process and to deeply analyse the search results.

The paper analyses and argues how Information Visualization may aid SW in a number of fundamental issues concerning the information search. In the first part it presents an analysis of the problems the user has still to face with during the information search in the SW. In the second part an analysis of the most representative visualization-based tools available in the WWW and in the SW is performed: a classification of some functionalities implemented in the tools are identified and a synthesis of their potentiality to solve the mentioned problems in the WWW and in the SW is provided.

The paper does not want to provide a state of the art of the available tools, but it aims to underlay that the potentialities offered by Information Visualization in the WWW could be applied in the SW to improve information search.

2. Problems in information search

Information search cannot be performed without any involvement of individuals since the "searching skills" are strongly dependent on human factors as the seeker's anxiety whenever the query result does not fulfil his needs, the limited seeker's knowledge, lack of relationship between seeker and information providers. These different factors get the user into some problems even when he searches for information in the SW. For instance, the definition of the search criteria is strongly affected by the limited knowledge of the seeker. According to Belkin this is a result of the Anomalous State of Knowledge (ASK) [3]. They

Proceedings of the 16th International Workshop on Database and Expert Systems Applications (DEXA'05)
1529-4188/05 $20.00 © 2005 **IEEE**

argue that the need for new information arises whenever the knowledge needed to carry out an activity is incomplete. In this case, the user decides to complete his knowledge by searching for new information but he is not able to precisely specify the information that he needs. Independently of the considered environment (traditional information retrieval systems, WWW or SW), the ASK forces the seeker to enter into dialogue with the IR systems engaging in a query refinement process. Spoerri mentioned how the query refinement process is affected by some problems in WWW [4]. In the following, the paper discusses the problems mentioned by Spoerri and it argues their relevance in the SW context.

Problem of query formulation: "how to precisely communicate the query criteria to the system?" This problem is strongly related to the choice of the language adopted to query. The formal languages are usually precise but they result unfriendly for seekers who have an inappropriate background. On the other hand, the natural language is considered more friendly but it is often not precise. To make the SW successful it is necessary to provide a friendly mean to precisely express the selection criteria the user has in his mind.

Problem of vocabulary: "Which term to use?" The difference of knowledge and perception between the information providers and the seekers is modelled in terms of informative space and cognitive space. The former one is defined as a set of objects and relations among them held by the system, whereas the latter one is defined as a set of concepts and relations held by individual [5]. Information providers organize resources according to their knowledge and vocabulary concurring to built the "informative space". If the seeker has different knowledge background his cognitive space has a poor overlapping with the information space, and he will use different terms to identify the same concepts. This problem is still relevant in the SW. Even if the use of ontology and of lexical databases like WordNet can ease the vocabulary problem, a mapping between each pair of cognitive space and information space is hardly provided.

Problem of database selection: "which search engine to select?". The seeker has to decide which search engine to use. The problem is well known in the WWW because the actual search engines are able to cover a limited portion of the web resources. Also SW is affected by similar problems since it will probably have different engines or web services, which differ each other in the technology and in the conceptualisation they rely on.

Problem of retrieved resources exploration: "how to explore many retrieved documents?". Semantic Search [1] is based on an explicit representation of semantics about web resources and real world objects. It 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). However, because of the huge amount of resources that will be available in the SW, even adopting the Semantic Search the seeker will have to face with a huge amount of query results. Furthermore, due to the ASK the queries formulated by the seeker might not correspond to a proper representation of his needs, as a result the order induced by ranking measures might be misleading. The seeker needs to be supported in their analysis of results to choose the most suitable for his purpose.

Problem of query coordination: "How to query?". Human behavioural studies shows that the seeker is lazy, usually he tends to create short queries and rarely adopts boolean expression in his query criteria[6]. On the other hand, he is forced to a deeper search in the Web as well as in the SW whenever he is the only one who can define the searching criteria and the result of his search can seriously affect the success of his work.

3. Tools analysis

An analysis of some visual-based tools in the WWW and SW is illustrated: a comparison of the tools in terms of their functionalities is proposed to identify their complementarities and to underline the potentialities offered by IV to improve information search. The aim is to demonstrate the advantages of adopting IV in the SW rather than to provide a complete state of art.

3.1. Tools vs. functionalities

Some of the most representative tools available in the WWW and SW are considered. Concerning the SW there are no specific tools based on visualization designed for Semantic Search; anyway there are well known tools developed in the field of Ontology which could be extended to ease information search. The analysis concerns tools for the Web as Kartoo [7], Grokker [8], Web Theme[9], Aduna AutoFocus [10], MetaCrystal [11], and the tools for visualization and interaction with the ontology as OntoViz [12], TGViz [12], Jambalaya[13], Spectacle [14]. The authors are well aware that several other tools are available in the web, but they mainly differ in the implementation or in how they combine the functionalities. The choice of the tools has been performed giving priority to those ones that can be freely downloaded. Only two tools raise an exception: Spectacle and Metaviz. The reason is that they offer functionalities which are not provided by the other tools, and even if it was impossible to make a direct use of them, they had to be included.

Table 1 : Functionalities provided by Information Visualization tools in the WWW and SW

		World Wide Web					Semantic Web			
		Grokker	Aduna AutoFocus	Kartoo	Meta-Crystal	Web theme	OntoViz	TGViz	Jambalaya	Spectacle
Graphical Visualization	Hierarchical Visualization	☺	-	-	-	-	😐	😐	☺	😐
	Clustering Visualization	☺	☺	-	-	☺	-	-	-	☺
	Map Based Visualization	-	-	☺	-	☺	-	-	-	-
	Venn diagram representation	-	☺	-	☺	-	-	-	-	☺
Graphical Interaction	Visualization Manipulation	☺	☺	☺	-	-	-	☺	☺	☺
	Graphical Selection	☺	☺	☺	☺	☺	-	☺	☺	☺
Interaction and Visualization	Highlighting	☺	-	-	☺	-	😐	☺	☺	☺
	Co-Occurring Terms Interaction/Visualization	☺	☺	☺	-	-	-	☺	☺	☺
	Coloured Query Result	-	-	-	-	☺	-	-	-	-
	Filter Results Representation	☺	☺	😐	😐	-	-	-	-	-
	String Search	☺	-	-	-	😐	-	☺	☺	☺
	Choice of Hierarchy Level shown	☺	n.a.	n.a.	n.a.	n.a.	-	☺	-	-
	Ontology Instances	n.a.	n.a.	n.a.	n.a.	n.a.	😐	☺	☺	☺
	Ontology Graph Navigation	n.a.	n.a.	n.a.	n.a.	n.a.	-	☺	☺	😐

The analysis concerns the study of the tools to identify the most relevant functionalities useful in the information search. We have grouped them in three main categories: graphical visualization, graphical interaction, and a combination of them.

The following functionalities have been outlined:

- *Hierarchical Visualization* to visualize and browse the content according to different levels of granularity (e.g. Grokker).
- *Clustering Visualization* to visualize and group the content according to similarity criteria. The groups are obtained either by applying a clustering algorithm (galaxy view[9]) or according to properties specified by the user (cluster map [15]).
- *Map Based Visualization* to organize the content according to thematic terms or co-occurrence criteria as in the geographical map (e.g. Kartoo).
- *Venn diagram representation* to describe and compare the elements and characteristics of items and to quickly convey a compact view of data (e.g. MetaCrystal, Spectacle).
- *Visualization Manipulation* to re-organize, move and add graphical elements (e.g. Grokker allow to insert a new web site in the displayed graph).
- *Graphical Selection* to select different information sources such as URI, PDF or DOC document in Grokker, Aduna Autofocus, Kartoo or data as in Web Theme[9].
- *Highlighting* to visualize a selected element and all its related sources (e.g. Aduna AutoFocus, Kartoo and Spectacle allow to highlight the related co-occurring terms, in Grokker the keywords used in filters are highlighted in the web pages visualization).
- *Co-Occurring Terms Visualization* to visualize a statistical thesaurus to expand user queries with other highly frequent terms. They should help the user in discriminating relevant documents ([16]).
- *Coloured Query Result* to set different colours for the query results to facilitate their comparison (e.g. Web Theme [9]).
- *Filter Results Representation* to apply filters to the contents. For instance, Grokker allows to filter the rank, the domain and the source, whereas Kartoo allows to filter the co-occurring terms.
- *String Search*: to search for a co-occurring word and to navigate the ontology hierarchy.
- *Choice of the Hierarchy Level Shown* to choose the number of levels displayed in the hierarchy exploring it at different levels of details.
- *Ontology Instances* to visualize the instances of a selected class separately or directly in the ontology graph (e.g. OntoViz, Jambalaya and Spectacle).
- *Ontology Graph Navigation* to easily navigate the ontology graph structure (e.g. Jambalaya proposes different layouts and an animated navigation to

browse the hierarchy). Table 1 summarizes the results of the analysis. It shows the associations between the tools (columns) and their functionalities (rows), as following:

- "n.a." the functionality is not applicable;
- "-" the functionality is not implemented;
- "😐" the functionality is partially implemented or implemented in a trivial way;
- "☺" the functionality is fully implemented.

The idea of this table is to identify a set of conceptual functionalities implemented in the considered tools rather than to provide a complete state of art. The evaluation of a functionality for each tool is based on its description reported in the material related to the tools or on its direct usage whenever the tools were available. The choice between ☺ or 😐 is performed according to the authors' impression. In general, a functionality is classified as implemented in trivial way (😐) if its implementation in the considered tool appears less impressive than the implementations provided by the other tools.

3.2. Functionalities vs. problems

Table 2 shows the result of the analysis of the support provided by each functionality in solving the mentioned problems for the WWW as well as for the SW. Functionalities which are not yet implemented for the SW but potentially useful are also outlined. The table has the following legend:

- "n.a." the functionality is not applicable in the considered context (WWW or SW);
- "-" the functionality does not seem to help to solve the problem, independently from its implementation;
- "Empty cell" the functionality could help to solve the problem, but it is not provided by any tool;
- "😐" the functionality gives a partial support in the resolution of the problem;
- "☺" the functionality provides a satisfactory help in the problem solution.

For each problem the table shows the comparison between the contributions that each functionality provides in the WWW and in the SW, respectively represented in the first half-column and second half-column of each problem column. The functionality/problem evaluation is obtained according to the principle that exists at least one of the considered tools which provides the functionality able to solve the problem. Different faces "😐,☺" are assigned according to the quality of the support provided by the functionality to solve the problem. Analysing the Table 2 it is possible to state that:

- Some functionalities ease the problems only if conceptualised and implemented in the SW ("ligh-grey background");

Table 2: Information Visualization and information search problems in the WWW and in the SW.

		Database		Vocabulary		Query Formulation		Results Comprehension		Query Coordination	
		WWW	SW	WWW	SW	WWW	SW	WWW	SW	WWW	SW
Graphical Visualization	Hierarchical Visualization	-	-	😐	😐	😐	😐	☺	😐	-	-
	Clustering Visualization	-	-	😐		😐	😐	☺	😐	-	-
	Map Based Visualization	-	-	😐		😐		☺		-	-
	Venn diagram representation	☺		-	-	😐	😐	😐	😐		
Graphical Interaction	Visualization Manipulation	-	-	-	-	-	-	☺	😐	-	-
	Graphical Selection	-	-	-	-	☺	😐	-	-	-	-
Interaction and Visualization	Highlighting	-	-	-	😐	-	-	☺	😐	-	-
	Co-Occurring Terms Interaction/Visualization	-	-	☺	☺	☺	☺	-	-	-	-
	Coloured Query Result			-	-	-	-	-	-	☺	
	Filter Results Representation	-	-	-	-	-	-	☺	😐	-	-
	String Search	-	-	-	😐	-	-	-	-	-	-
	Choice of Hierarchy Level shown	-	-	-	😐	-	-	😐	😐	-	-
	Ontology Instances	n.a.	-	n.a.	-	n.a.	😐	n.a.	😐	n.a.	😐
	Ontology Graph Navigation	n.a.	-	n.a.	😐	n.a.	😐	n.a.	😐	n.a.	-

- Some functionalities are implemented both in the WWW and in the SW, but in the former one they provide better results than in the latter one ("middle-grey background);
- Some functionalities are implemented only in the WWW (represented in "dark-grey background").

In general, Information Visualization has the potentiality to help the user in the searching task: for each problem there is at least a functionality that provides a useful support in the Web. However as detailed in the following, IV seems still not properly developed for the SW.

Concerning the vocabulary problem, a partial support is already provided in the SW. It is important to note that all the comparisons highlighted by light-grey background are in the columns of vocabulary problem. It suggests how SW partially takes advantage from the use of ontologies. However, the IV tools in the SW are far from to completely solve the problem: ontologies contain information about the application domain but patterns induced by the use of domain are not made explicit. Novel IV techniques able to make explicit similarities and patterns among sources have to be developed.

Considering the columns related to the problems of query formulation and result comprehension, the middle-grey and the dark-grey background are mainly due to the fact that except for Spectacle the tool in SW are realized to support in the design of ontologies rather than in the search. A significant improvement can be obtained by focusing the same functionalities on the solution of the problem.

Regarding the database and query coordination problems, although a support is provided in the WWW, there are no "face" symbols (☺/😐) in correspondence of the columns related to the SW. The adoption of some functionalities coming from the Web (e.g. colour query results, filter results representation) may help both in the database and the query coordination problems.

4. Conclusions

The paper examines the potentiality of applying IV into SW to improve information search. Several seekers' problems are outlined and an analysis of their occurrence in the context of SW is provided. A set of visual searching tools are analysed to show that IV is able to face with these problems. Even if the analysis has been limited to a subset of tools, it has outlined that there are functionalities that according to our perception are able to ease in facing with all the mentioned problems. In spite of that, IV is largely applied in WWW but it is not yet completely exploited in the SW.

Future development will address how to adapt the existing IV techniques to the SW. It will pave the way for a conceptual framework which integrates Semantic Search, Ontology and Information Visualization to solve all searching problems.

Acknowledgment

This work has been partially developed within the Network of Excellence Aim@Shape, FP6 IST NoE 506766.

Bibliography

[1] R. Guha, R. McCool, and E. Miller, "Semantic search", Proceedings of the 12th international conference on World Wide Web,pp.700-709,ACM Press, 2003.

[2] S. K. Card, J. D. Mackinlay, and B. Shneiderman, Information visualization 1999, pp. 1-34.

[3] N.J.Belkin, N.R.Oddy, and M.H.Brooks, "ASK for Information Retrieval: Part I. Background and Theory", Jounal of Documentation, vol.38, no.2, pp.61-71, 1982.

[4] A. Spoerri, "How Visual Query Tools Can Support Users Searching the Internet", Proceedings of Eighth International Conference on Information Visualisation, pp. 329-334, IEEE, 2004.

[5] G.B.Newby,"Cognitive space and information space", Journal of the American Society for Information Science and Technology, vol.52, no.12, pp.1026-1048, 2001.

[6] 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 pp. 226-234, 2001.

[7] Kartoo, Ver . 4, 2004, http://www.kartoo.com.

[8] Grokker, Ver .2.1, 2004, http://www.grokker.com.

[9] M.A.Whiting and N.Cramer,"WebTheme (TM): Understanding Web information through visual analytics", Lecture Notes in Computer Science, vol. 2342 pp. 460-468, 2002.

[10] AutoFocus, Ver .2004.1, 2004, http://aduna.biz/products/autofocus/personal/index.html.

[11] A. Spoerri, "Coordinated views and tight coupling to support meta searching", Proceedings of Second International Conference on Coordinated and Multiple Views in Exploratory Visualization, pp. 39-48, 2004.

[12] TGVizTab, 2004, www.ecs.soton.ac.uk/~ha/TGVizTab/.

[13] Jambalaya, Ver .2, 2004, www.thechiselgroup.org.

[14] C. Fluit, M. Sabou, and F. v. Harmelen, "Ontology-based Information Visualization," Visualizing the Semantic Web 2002, pp. 36-48.

[15] C. Fluit and J. Wester, "Using visualization for information management tasks", Proceedings of the Sixth International Conference on Information Visualisation, pp. 447-454, IEEE, 2002.

[16] H. J. Peat and P. Willett, "The Limitations of Term Co-occurrence Data for Query Expansion in Document-Retrieval Systems", Journal of the American Society for Information Science, vol.42, no.5,pp.378-383,1991.