Chapter 7
Visualizing Search Results of Linked Open Data

Christian Stab, Dirk Burkhardt, Matthias Breyer, and Kawa Nazemi

Abstract Finding accurate information of high quality is still a challenging task particularly with regards to the increasing amount of resources in current information systems. This is especially true if policy decisions that impact humans, economy or environment are based on the demanded information. For improving search result generation and analyzing user queries more and more information retrieval systems utilize Linked Open Data and other semantic knowledge bases. Nevertheless, the semantic information that is used during search result generation mostly remains hidden from the users although it significantly supports users in understanding and assessing search results. The presented approach combines information visualizations with semantic information for offering visual feedback about the reasons the results were retrieved. It visually represents the semantic interpretation and the relation between query terms and search results to offer more transparency in search result generation and allows users to unambiguously assess the relevance of the retrieved resources for their individual search. The approach also supports the common search strategies by providing visual feedback for query refinement and enhancement. Besides the detailed description of the search system, an evaluation of the approach shows that the use of semantic information considerably supports users in assessment and decision-making tasks.

7.1 Introduction

Assessing information is a common task for decision makers. Especially in the area of policy modeling, analysts have to consider different perspectives when designing a new policy that impacts humans, economy and the environment. So homogeneous access possibilities as well as adequate representations for distributed data plays a major role for providing adequate tools that facilitate decision making. Thereby, semantic technologies provide adequate tools for linking heterogeneous data sources and for generating broader contexts that facilitate information access and enables

C. Stab (✉) · D. Burkhardt · M. Breyer · K. Nazemi
Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt, Germany
e-mail: christian.stab@igd.fraunhofer.de

T. Hussein et al. (eds.), *Semantic Models for Adaptive Interactive Systems*, Human–Computer Interaction Series, DOI 10.1007/978-1-4471-5301-6_7,

data exchange between different systems (Shadbolt et al. 2006). With the ongoing establishment of semantic technologies like the *Resource Description Framework* (RDF), the *Web Ontology Language* (OWL), semantic-oriented query languages like *SPARQL* and *Linked Open Data* (LOD) platforms like DBpedia (Auer et al. 2007) these developments are not only limited to specific domains but also adopted in daily search processes (Fernandez et al. 2008) and even for interlinking government data in so-called *Linked Government Data* (LGD) platforms (Ding et al. 2010; Wood 2011). For accessing these large amounts of interlinked data usually information retrieval methods are utilized that retrieve relevant resources by means of given search terms or keywords. The results of search processing are usually presented in sorted lists and in most cases the ordering of list entries represents the relevance of the results for the individual search according to various criteria (Cutrell et al. 2006). So the most relevant result is placed in the first row followed by less important ones. The semantic information of the resources that is used during the search result generation and the analysis of search terms remains hidden from the user in most cases, though this information considerably supports users in information-seeking tasks and selection of appropriate documents for further examination.

For designing search user interfaces efficient and informative feedback about the retrieved resources is critically important for the user to be able to assess the presented search results (Hearst 2009). In particular this includes feedback about query formulation and about reasons the results were retrieved from the information system. However, the use of additional relevance indicators in result lists besides relevance ordering such as numerical scores or special icons turned out to be not successful for supporting users in understanding the retrieved results because the meaning of the relevance score is opaque to the user (White et al. 2007). This is because the majority of existing relevance indicators only presents a single relevance per search result that summarizes all criteria instead of offering a more fine-grained insight to search result processing.

In contrast to common search result presentations the presented approach makes use of information visualizations for representing search results of semantically modeled data. In order to offer users an adequate tool for assessing the relevance of the retrieved search results the approach combines information visualization techniques with semantic information and different weights that emerge during the search process. The approach also visually supports the common search strategies by providing visual feedback for query evolution. The main contributions and benefits of the presented approach are:

- *Query-Result-Relations*: The semantic search processing of the approach analyzes the given queries and identifies relations between query terms and search results. The visualization of these relations offers a fine-grained overview of search result relevancies and facilitates information seeking and assessment tasks.
- *Relevance Assessment*: The inclusion of semantic information in search result presentation offers more transparency in search result generation and successfully supports user in assessing the relevance of the presented resources.
- *Visual Query Enhancement*: The visual recommendation of additional query terms related to the set of search results supports the common search strategy and

allows users to narrow current search results and to immediately receive visual feedback.

In the next section we give an overview of related work that influenced the development of our approach. Subsequently, we introduce our approach for processing and visualizing searches in semantic domains and give a detailed description of each component. We present the results of an evaluation comparing our approach to already existing solutions followed by an outlook of its application in the area of policy modeling.

7.2 Related Work

Semantic models are increasingly used for linking heterogeneous data sources as well as for generating broader contexts that facilitate information access (Shadbolt et al. 2006). Nowadays, these technologies are not only limited to specific domains but also adopted in daily search processes of web-based search engines (Fernandez et al. 2008). A commonly used and useful approach for representing search results is the term highlighting technique (Aula 2004) where the terms of a given query are highlighted in the surrogates of search result lists. This approach is also referred to as Keyword-In-Context (KWIC). For example the BioText System (Hearst et al. 2007) represents beside extracted figures from relevant articles, query terms highlighted in the title and boldfaced in the text excerpts for communicating reasons the particular results were retrieved. Even though term highlighting can be useful for improving search result list presentations, it does not reveal the semantic interpretation of search results and prevent users from scanning the whole result list for getting an overview.

Although the initial intention of semantic technologies was not focused on presenting semantics to end-users, there are several approaches that benefit thereby. SemaPlorer (Schenk et al. 2009) is an interactive application that makes use of multiple semantic data sources and allows users to visualize results of their search in various views. The user interface of SemaPlorer combines a geographic visualization and a media view for visualizing geospatially annotated data and picture galleries respectively. The approach also includes facettation of search results and focuses on combining search results from different heterogeneous knowledge bases. Another approach of presenting semantic information to the user is the Relfinder interface (Heim et al. 2010). It supports users in interactively discovering relations between resources in semantic knowledge spaces. Users can prompt two or more resources and the relations between them are shown in a graph-based visualization. Although this approach demonstrates the benefit of communicating semantic knowledge to users, it is strictly limited to relation discovery.

Other approaches utilize different information visualization techniques for improving search result presentations. To name only a few, the Microsoft Academic Search interface[1] incorporates geographic, graph-based and temporal visualization

[1] http://academic.research.microsoft.com.

techniques for searching and exploring publications or authors and offers also a stacked area chart for analyzing trends in the field of computer science. Skyline-Search (Stoyanovich et al. 2011) is a search interface that supports life science researchers in performing scientific literature search. It leverages semantic annotations to visualize search results in a scatterplot representing relevance against publication date. Even though semantic annotations are used for search processing and estimating relevance values, semantic knowledge is not directly presented to the user. The WebSearchViz (Nguyen and Zhang 2006) is an approach for visualizing web search results based on the metaphor of the solar system. It offers users the possibility to observe the relevance between a query and a web search result by the spatial proximity between objects. However, the system does not visualize semantic interpretations of search results or semantic structures.

The visual design of the presented visualization is based on a force-based visualization method similar to RadVis (Hoffman et al. 1997). Force-based visualization methods utilize physical laws (e.g. Hooke's Law) to locate each data record on a two dimensional screen by assigning different forces between the visual representations of each record. By adapting these forces in an iterative simulation, the physical system reaches a mechanical equilibrium resulting in an aesthetical layout of the given data. Originally, RadVis is a visualization technique for multivariate data. For each dimension this visualization locates an anchor point on a circle. Theses anchor points can be seen as fixed ends of springs. Each representation of a data record is attached to all related anchor points and the attraction force for each spring is weighted with the value of the record for the specific dimension. This results in different distances between the data records and the anchor points that represent the characteristics of each record according to the visualized dimensions. This approach is especially useful for identifying outliers in the data and for recognizing clusters.

Although there are different approaches that make use of semantic information for improving search results or result presentation and approaches that utilize information visualization for representing search results, there is still a gap in combining semantic search processing, information visualization and search user interfaces. The approach presented in this chapter aims at combining these three technologies into an interactive search user interface that facilitates information access and relevance assessment.

7.3 Search Procedure

For visualizing search results, there are also additional requirements for the search procedure and in particular the result generation. In contrast to commonly used textual list presentations that are based on ordered result lists, for the visualization of search results structured data is needed. For instance these additional structures may include clusters, relations, or labeled taxonomies that can be exploited for providing meaningful visual representations. The semantic information provided e.g. in Linked Open Data databases provides a useful starting point for extracting this structural information.

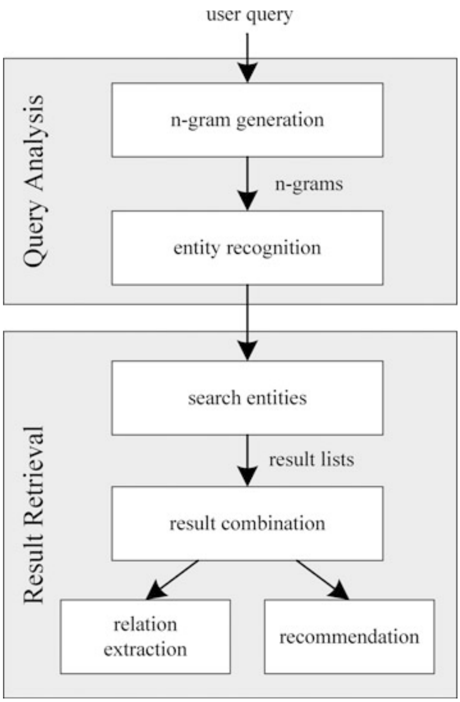

Fig. 7.1 Process for retrieving search results, query-result-relations and recommendations from semantic databases

The procedure used for retrieving the needed result structures from a semantic database consists of several steps at two different stages (Fig. 7.1). In the first stage the given user query is analyzed and dissembled into n-grams. Each of these n-grams is then mapped to an instance in the semantic data base e.g. by using an entity recognition approach like that described in Paulheim and Fümkranz (2012). Additionally, the type label of each recognized entity is retrieved by identifying the most specific concept in the semantic structure. So the result of the query analysis is a set of n-grams related to an instance and its most specific concept. We also considered different filtering strategies for discarding incomplete results. In particular there are two different filters: (1) an entity filter that filters all n-grams that could not be mapped to an entity and (2) a type filter that removes all entities whose concepts could not be retrieved. The search system can be parameterized to use one, both or none of these filters.

In the second stage the results for each recognized entity are retrieved and extended with additional structural information. In the first step each recognized entity that meets the given filter condition is searched in the semantic data base using an indexed search. The result lists are combined using two sorting fields. The first sorting field refers to the number of occurrences for each retrieved resource in the retrieved lists and the second sort field is the number of references that each retrieved resource has in the underlying semantics data base. So the first element is the element that is retrieved for most of the recognized query terms. Based on the combined result list, the search system extracts relations between the recognized query terms and results and retrieves additional recommendations as resources related to the result list (Fig. 7.1). The relation extraction creates for each resource a set of relations to related n-grams. These relations are weighted according to the rank of the resource in the result list. Thus the result of the relation extraction is a set of relations between query parts that are recognized as entities and the results in the combined result list. For providing additional information about the search, the procedure retrieves also resources that are related to the result list. These recommendations are extracted by querying all related entities of the resources in the combined result list from the semantic database. For each of these related resources the most specific concept is extracted. The result is a set of concepts with associated resources that are related to the search result list. These concepts are ordered according to the number of related resources. So each recommendation is a concept with a set of associated resources that are related to the search result list.

For our visualization approach we implemented the described search procedure as a web service using the DBpedia SPARQL-Endpoint. Besides the access to the semantics, the system also requires an indexed search for retrieving the result lists. For our prototype we used the DBpedia Lookup Service[2] that ranks the results according to the number of references (refCount). In the following section we describe how the results of the semantic search procedure are presented to the user and how the visualization interacts with the search system.

7.4 Visualizing Semantic Search Results

Our visualization approach for representing search results in semantic domains is based on the semantics visualization framework SemaVis[3] (Nazemi et al. 2013). The Framework includes several aspect-oriented visualization techniques (e.g. sunburst visualizations (Stab et al. 2010b), timelines (Stab et al. 2010a), map-based approaches (Nazemi et al. 2009), etc.) that can be combined to an application-specific visualization cockpit to represent different aspects of the underlying data (Nazemi et al. 2010). Several data providers for common semantic file formats and service APIs, a modular representation model for adapting the visual appearance (Nazemi et al. 2011) as well as a script-based configuration language called *Semantics Visualization Markup Language* (SVML) (Nazemi et al. 2013) are also included in the framework. In the following, we will first describe the basic concepts of the visualization component before introducing the details in succeeding sections.

The approach for visualizing semantic search results distinguishes between two different node types: (1) *Term Nodes* and (2) *Result Nodes*. Each node type is treated by different layout algorithms in the visualization and used to represent different information:

[2]For the prototype we use the DBpedia lookup service that is available at http://wiki.dbpedia.org/lookup/.

[3]SemaVis Framework: http://www.semavis.net.

- *Term Nodes* represent the terms recognized by the described search process. These nodes are placed by a concentric layout algorithm at the startup of the visualization. Users are also able to freely move and order them on the surface according to their individual preferences.
- *Result Nodes* represent the hits in the combined result list that are found for the given user query. These nodes are visually connected to related term nodes with directed edges and are treated by a force-based layout algorithm according to their weights to the surrounding term nodes.

The placement of the result nodes in the center of the visualization is done by a force-based layout algorithm. The algorithm positions nodes in a two-dimensional space by assigning different forces to the edges and the nodes of a graph. These forces are adapted during the layout process in an iterative simulation until the physical system reaches a mechanical equilibrium. Due to this layout technique, the overlapping of nodes and edges is prevented as far as possible and an aesthetical layout of the graph is achieved. Another interesting characteristic of force-based layout methods is that the forces between the nodes can be weighted with different similarity values. As a result of the weighting, different distances between the nodes are derived and the placement of the nodes is affected. To exploit this feature we assign different values to the visualized result nodes and their edges and utilize a model based on the weights emerged during the semantic retrieval process. In the representation, a result node is then placed nearer to similar term nodes and further away from terms that are less related to it. So the placement of the result nodes indicates the relevance of the results to the surrounding terms and users are able to distinguish different characteristics of retrieved hits.

Figure 7.2 shows an example of the visualization approach[4] for the query term "apple computer steve jobs". The query analysis stage of the search process identified the six terms "apple computer", "steve jobs", "apple", "computer", "steve" and "jobs" and for 4 of these terms the most specific concepts "company", "person", "plant" and "given name". Each of these identified terms is represented in a term node and the retrieved result nodes are attached to them using the weights identified during the relation extraction step. In the next section, we describe how this visualization metaphor can be used for providing more transparency in search processing and how it can be utilized for fostering search result comprehension.

7.4.1 Query-Result-Relations

Semantic technologies enable information retrieval systems to "understand" on the one hand the query terms given by the user (Sect. 7.3) and on the other hand different properties of the underlying resources. Based on the occurrences of the query terms

[4] A version using the DBpedia database and the described search processing is available at http://semanticsearch.semavis.net.

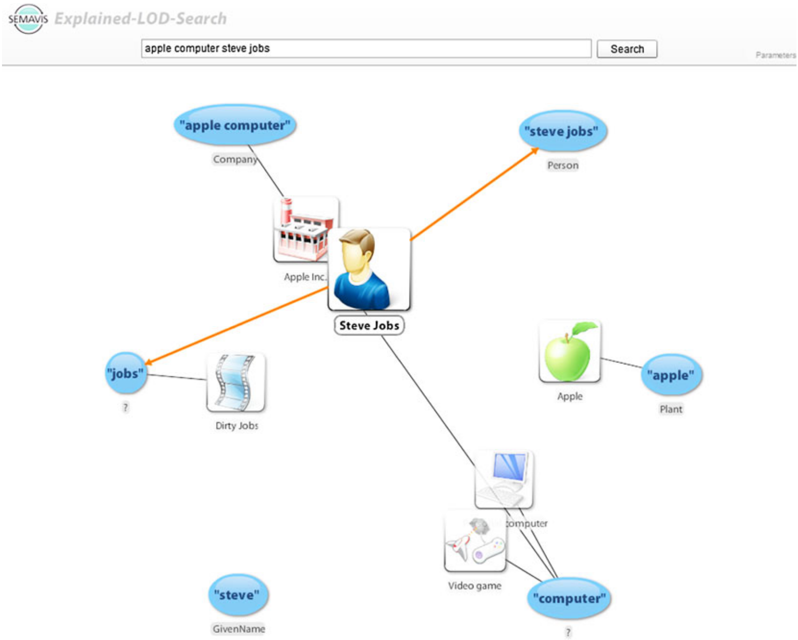

Fig. 7.2 Semantic search result visualization presenting the six most relevant search results for the query "apple computer steve jobs". Each result is attached between the terms recognized during the semantic query analysis and according to the retrieved query-result-relations

either in the content, in the semantic properties or even in the semantic neighborhood, the underlying resources are filtered and ranked using information retrieval methods for presenting the result resource to the end user. In this process semantic information improves the retrieval process by providing an extended feature space for each resource as well as by offering methods for disambiguating and interpreting the query terms given by the user. However, the semantic information utilized during the retrieval process remains in most cases hidden from the user though it provides useful feedback about the reasons the results were retrieved. For example the query term "ford" might be interpreted as the name-property of a car manufacturer, as the surname-property of the famous inventor or the title-property of an activity for crossing rivers. Each of these interpretations will deliver a completely different result set. So it is not sufficient to only present the relations between recognized query terms and results, but it is also necessary to point out the semantic interpretation of the given query terms to allow an unambiguous assessment of retrieved results.

To meet these demands and to provide an adequate tool that allows users to unambiguously determine the most relevant result for their individual search, our approach visualizes both query-result-relations and the interpreted semantic meaning of query terms. Therefore, each term of the given query is presented in a term node of the visualization. The interpreted semantic meaning emerged during search

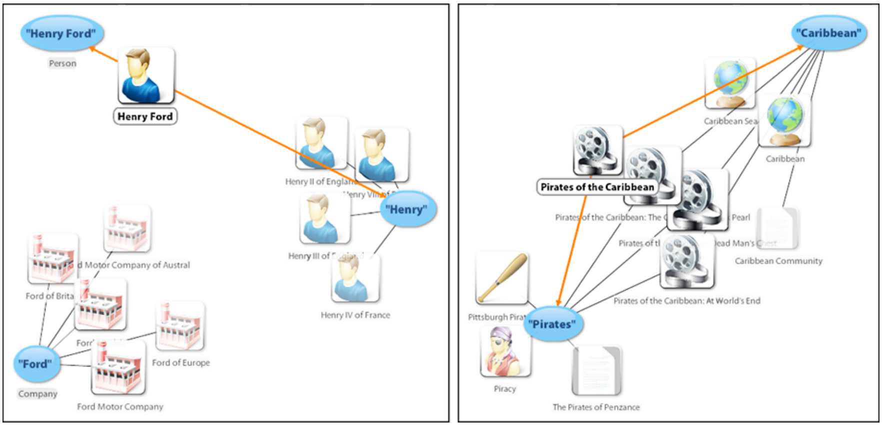

Fig. 7.3 *Left*: The visualization of query-result-relations reveals that only one of the ten results is related to the requested person name. *Right*: The visual representation of the results avoids mistakes in result assessment tasks

processing (the label of the most specific concept) is visible in the label of the term node. So for every possible interpretation a new node is created that represents the query term and its retrieved interpretation. The relations between search results and the term nodes are depicted as directed and weighted edges between term nodes and result nodes. As mentioned above, the weighting of a query-result-relation is derived according to the rank of the result in the individual result list. As consequence of this weighting, the results are placed nearer to more relevant entities and term nodes respectively.

Figure 7.3 shows two examples of the visualization approach representing two different queries and in each case the first ten hits of the result set. For the sake of clarity, we limited the number of simultaneously visible results to a fixed number and added common paging functions for switching between pages. The left example shows the visualization of the results for the query "Henry Ford" where the term "Henry Ford" is identified as person and the term "Ford" as company. Additionally, the term "Henry" is recognized by the query analysis step of the search processing but there is no specific type other than owl:Thing available. So the type for this query part is not evident. However, the visualization of the results reveals that only one of the results is related to the queried person whereas other results are related to the recognized company (Fig. 7.3 left).

The second example shows the visualization of the results for the query "Pirates of the Caribbean" (Fig. 7.3 right). For the given query the system identifies two entities "Pirates" and "Caribbean". By visualizing the connections between search results and query terms, users are able to recognize the four movies that are related to both search terms.

7.4.2 Mapping Results' Relevance to Visual Properties

When designing search result visualizations it is crucially important to provide fine-grained insights to search result processing so that the user is able to explore more details about the retrieved results. As mentioned in the introduction, the integration of additional relevance indicators in search result lists turned out to be not sufficient for supporting users in understanding the retrieved results because common relevance indicators only presents a single relevance per search result.

For improving the result presentation in the presented visualization approach different values that emerge during the result retrieval process are utilized and combined with the visualization of query-result-relations. In particular there are two different values that are mapped to visual properties of the visualization to indicate a more fine-grained relevance metrics:

- *Relation Weights* indicate the relevance between a query term and a retrieved search result. These weights are the result of the relation extraction step described in Sect. 7.3. With respect to the presented visualization, these weights correspond to a relevance metric between the result nodes and related term nodes.
- *Result Rank* indicates the overall relevance of a search result according to the given query. The result rank is equivalent to the rank identified during the result combination step of the described search process.

In order to make the optimum use of these values, each of them is mapped to specific visual properties like length, color and size that can be preattentively perceived (Ward et al. 2010). Thanks to the characteristic of preattentive perception, these relevance indicators can be perceived faster and easier by the users compared to common indicators that are often represented as textual percentages. To take these advantages, the result rank is used to adjust the size and the intensity of the result nodes. Thereby the resource that has the highest overall rank for a specific search query is presented most conspicuous whereas resources with minor rank are visualized less notable (Fig. 7.2). On the other hand the relation weights are used to adapt the weighting of edges between results and term nodes. This results in different lengths of the visible connections and indicates the relevance between specific query terms and search results.

7.4.3 Visual Support for Query Evolution

A search process of common users includes various search requests and queries until the needed information is found. Usually such a process starts with a general query that is revised in consecutive search queries until some resources for further exploration are found in the result set. This kind of search strategy was also revealed in several studies which showed that it is a common strategy for the user to first issue a general query, then review a few results, and if the desired information is not found, to reformulate or to enhance the query (Hearst 2009; Jansen et al. 2005, 2007). To support this behavior of common users when searching for information in query-based information retrieval systems, we integrated several features that enable users to interactively change their queries in the visualization.

In the presented visualization approach the query terms of the user are visualized in term nodes that are arranged in a circular form around the result nodes. According to this characteristic, the state of these visible term nodes and the included terms reflect the current search intention of the user. Transferred to the visualization approach, the before mentioned strategy of query evolution corresponds to a substitution, reassignment, creation or removal of term nodes. Hence, the change of the current state of the term nodes results in a new search condition that in turn changes the visible result set or the visible relations between result and term nodes. On the one hand, the creation of further terms defines a more specific search condition and on the other hand, the removal of term nodes results in wider-ranged search spaces. In contrast to commonly used search user interfaces, the influence of changing search conditions is immediately visible in the visualization. The representation of query-result-relations reveals which of the current search results fulfill new conditions (Fig. 7.4) and provides an immediate visual feedback of the users' query evolution.

To ensure that users are aware of additional terms and resources respectively, the visualization also represents the recommendations retrieved by the semantic search process for supporting users in finding needed information. These recommendations are visualized as additional term nodes that are labeled with a question mark to encourage users to instantiate them for narrowing their search. The size of the recommended term nodes is mapped to their influence to the current result set. So term nodes whose instantiation will cause major changes of the result set are represented larger than term nodes whose instantiation will only affect smaller parts. By selecting a specific recommendation, users are able to select different resources for instantiating the term node and narrowing their retrieved results (Fig. 7.4).

7.5 Evaluation of the Visualization Approach

For assessing the effectiveness of the presented approach we performed a user study in which we compared the visualization approach with a common list view that includes the identical information for each result resource. The main focus of the evaluation was to answer the question whether the presented visualization approach can support users in assessing search results and if our approach satisfies the needs of searchers. For verifying our assumption we investigated the task completion time and formulated the following hypothesis:

- *H1: There is a difference in task completion time between the list view and the visualization in assessing search results.*

In addition to the task completion time we considered several subjective criteria that were collected with additional questionnaires for each task and participants. In

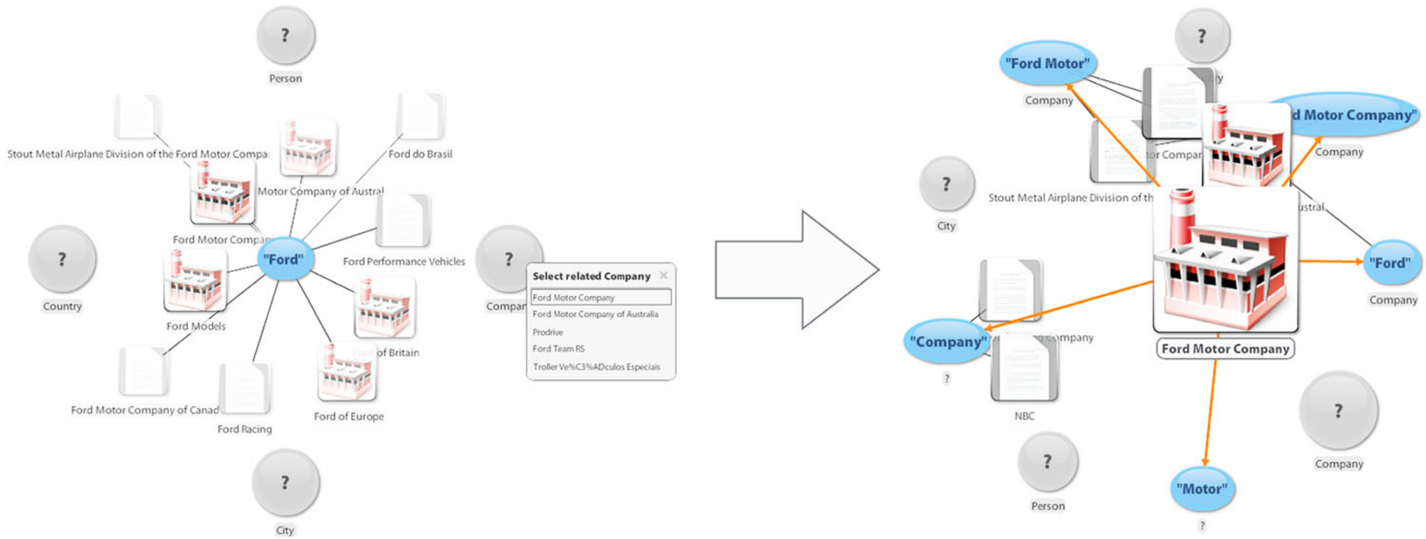

Fig. 7.4 The recommendation of additional terms supports users in their common search strategy and provides a visual tool for query evolution and narrowing search results. The impact of query changes is immediately visible in the visualization

particular, we captured the user satisfaction and the following three additional items for getting an impression of the user experience:

- *Q1: With the help of the system, I was able to quickly and effectively solve the given task.*
- *Q2: The system presented the information needed for answering the question clear and unambiguous.*
- *Q3: Would you use the system in the future for similar search tasks?*

Each of these items was rated by the participants on a five point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

7.5.1 Experimental Design

According to the hypothesis that contains one independent variable with two different conditions (list view and visualization) the design of our experiment is based on a basic design (Lazar et al. 2010). Additionally, we decided to use a within-group design for our experiment where each participant accomplishes the given tasks in each condition (in this case the different user interfaces). In contrast to between-group designed experiments, in within-group designs less participants are needed and individual differences between the participants are isolated more effectively (Lazar et al. 2010). Possible learn effects when switching between conditions are controlled by a systematic randomization of condition- and task-ordering. Furthermore, participants were advised to disregard the knowledge from previous conditions and to explicitly show the solution of tasks by means of elements in the user interface.

Altogether, the experiment contains three tasks that had to be accomplished from every participant with both conditions (list view and visualization). Because the focus of the evaluation is the comparison of two different user interfaces and not the investigation of the whole search process, we were able to pre-assign the query terms for every task. So every participant retrieves the same results for every task and thus also the same visual representation and the evaluation outcome is not influenced by other factors.

In the first task participants had to identify the relations between each search result and the recognized terms of the given query. The second task was of the same type as the first task with the difference that the result set contains more complex relations. In the third task participants had to identify the most relevant item for a specific search situation. To ensure that the solution could be found in each condition, we performed several pretests. We also ensured that each participant gets the same visual presentation for each task and condition. The time limit for each task was set to three minutes. If a wrong answer was given or a participant could not solve a task, the completion time of the task was also set to three minutes.

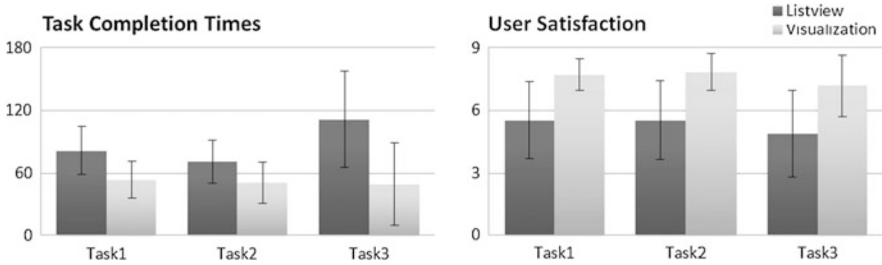

Fig. 7.5 Task completion time and user satisfaction

7.5.2 Procedure

17 participants, mainly graduates and students attended the evaluation. The participants were between 24 and 29 years old and mainly involved in computer science (M = 4.65; SD = 0.6).[5] After a general introduction to the user study and an explanation of the procedure and tasks, participants got a brief introduction to both systems in systematically randomized ordering. Both systems were queried with a reference query and participants had the chance to ask questions about the systems. After each task, participants had to rate their overall satisfaction with the system on a scale from 1 to 9 and the three before mentioned items concerning their subjective opinion about the system on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). After participants had completed all tasks, they had to answer a brief demographic questionnaire.

7.5.3 Results

Figure 7.5 shows the average task completion times for each of the three tasks and both conditions. The direct comparison of the average task completion times reveals that participants performed better with our visualization approach (avg(t) = 51.3 s; SD = 25.8) compared to the list view (avg(t) = 88.1 s; SD = 30.1).

A paired-samples t-test also suggests that there is a significant difference in the task completion time between the group who used the list view and the group who used our visualization approach ($t(50) = 7.8028$, $p < 0.05$). Hence, the null hypothesis is refuted and the alternative hypothesis confirmed. The comparison of means also indicates that users performed significantly faster with the visualization approach compared to the list presentation. So we can proceed from the assumption that visualizing search results taking semantic information into account has a positive effect on the efficiency when assessing the relevance of search results.

[5]Measured on a five point scale (5 = very much experience; 1 = very little experience) in the demographic part of the questionnaire.

Table 7.1 Results of the subjective ratings indicate that users prefer the visualization approach

Presentation	Q1	Q2	Q3
Visualization	4.33	4.37	4.25
Listview	3.15	3.03	3.14

The evaluation of satisfaction ratings indicates that participants feel more comfortable with our visualization approach instead of the commonly used list view. The list view obtained an average rating of 5.31 with a standard deviation of 1.91 whereas the visualization obtained an average rating of 7.57 and a standard deviation of 1.10. Additionally, the result of the subjective ratings (Table 7.1) and in particular question "Would you use the system in the future for similar search tasks?" confirms the assumption that users prefer the visualization to the list presentation (list: M = 3.14; SD = 0.87; visualization: M = 4.25; SD = 0.77).

7.6 Visualizations and Linked Data in the Policy Modeling Process

Retrieving and accessing information is a challenging task and crucially important in many different domains. This is especially true in the area of policy modeling where decisions impact humans, economy or environment. The creation of novel policies is a very complex task that requires several process steps (Macintosh 2004) to ensure validity and positive effects. It is easy to imagine that an insufficient analysis of the underlying problem and the consideration of all impact factors will result in a policy that fails the intended goals. Accurate decision making in this domain not only requires the consideration of diverse impact factors but also the inclusion of increasingly complex and dynamic scenarios. For improving this process and the quality of the achieved policies respectively, recent initiatives, like Open Government Data or Linked Government Data aim at publishing and interlinking vast amounts of data for enabling accurate decision support and innovative ICT solutions for fostering political decision making. However, the amount of available data that contains implicit and hidden information relevant for specific policy decisions and scenarios continuously increases and the access gets more and more complicated.

To improve the retrieval tools, not only the interlinking of open administrative data gains enormous importance for policy modeling but also the development of novel result presentations and exploration tools. Present systems for searching and accessing this information are currently limited to textual result presentations and require comprehensive knowledge about the domain for finding the information that fits to the specific case. Approaches, like the presented visualization technique, will on the one hand provide a homogeneous access for distributed and interlinked data and on the other hand enable political decision makers to identify unknown and hidden information. In particular during the information foraging step of the policy modeling process (Kohlhammer et al. 2012), visualization techniques will enable an optimal analysis of the need for a policy and accurate assessment of issues relevant to a specific scenario. In order to further develop this idea, we investigate different visualization approaches along each step of the policy modeling process in the European project FUPOL 287119: Future Policy Modeling, partially supported by the European Commission. The FUPOL project proposes a comprehensive and new governance model to support the policy design lifecycle. The innovations are driven by the demand of citizens and political decision makers to support the policy domains in urban regions with appropriate ICT technologies.

References

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: a nucleus for a web of open data. In *Proceedings of the 6th international semantic web and 2nd Asian semantic web conference*, ISWC'07/ASWC'07, Busan, Korea (pp. 722–735).

Aula, A. (2004). Enhancing the readability of search result summaries. In *Proceedings of HCI 2004, the 18th British HCI group annual conference* (Vol. 2, pp. 1–4).

Cutrell, E., Robbins, D., Dumais, S., & Sarin, R. (2006). Fast, flexible filtering with phlat. In *Proceedings of the SIGCHI conference on human factors in computing systems*, CHI '06, Montreal, Quebec, Canada (pp. 261–270).

Ding, L., DiFranzo, D., Graves, A., Michaelis, J. R., Li, X., McGuinness, D. L., & Hendler, J. A. (2010). TWC data-gov corpus: incrementally generating linked government data from data.gov. In *Proceedings of the 19th international conference on world wide web*, WWW '10. Raleigh, North Carolina, USA (pp. 1383–1386).

Fernandez, M., Lopez, V., Sabou, M., Uren, V., Vallet, D., Motta, E., & Castells, P. (2008). Semantic search meets the web. In *2008 IEEE international conference on semantic computing* (pp. 253–260).

Hearst, M. A. (2009). *Search user interfaces* (1st ed.). New York: Cambridge University Press.

Hearst, M. A., Divoli, A., Guturu, H., Ksikes, A., Nakov, P., Wooldridge, M. A., & Ye, J. (2007). BioText search engine: beyond abstract search. *Bioinformatics*, *23*(16), 2196–2197.

Heim, P., Lohmann, S., & Stegemann, T. (2010). Interactive relationship discovery via the semantic web. In *Proceedings of the 7th international conference on the semantic web: research and applications, Part I*, ESWC'10. Heraklion, Crete, Greece (pp. 303–317).

Hoffman, P., Grinstein, G., Marx, K., Grosse, I., & Stanley, E. (1997). DNA visual and analytic data mining. In *Visualization '97, proceedings* (pp. 437–441).

Jansen, B. J., Spink, A., & Pedersen, J. (2005). A temporal comparison of AltaVista web searching: research articles. *Journal of the American Society for Information Science and Technology*, *56*(6), 559–570.

Jansen, B. J., Spink, A., & Koshman, S. (2007). Web searcher interaction with the dogpile.com metasearch engine. *Journal of the American Society for Information Science and Technology*, *58*(5), 744–755.

Kohlhammer, J., Nazemi, K., Ruppert, T., & Burkhardt, D. (2012). Toward visualization in policy modeling. *IEEE Computer Graphics and Applications*, *32*(5), 84–89.

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research methods in human-computer interaction*. New York: Wiley.

Macintosh, A. (2004). Characterizing e-participation in policy-making. In *Proceedings of the 37th annual Hawaii international conference on system sciences, 2004*.

Nazemi, K., Breyer, M., & Hornung, C. (2009). SEMAP: a concept for the visualization of semantics as maps. In C. Stephanidis (Ed.), *Lecture notes in computer science*: Vol. 7. *HCI* (pp. 83–91). Berlin: Springer.

Nazemi, K., Breyer, M., Burkhardt, D., & Fellner, D. W. (2010). Visualization cockpit: orchestration of multiple visualizations for knowledge-exploration. *International Journal of Advanced Corporate Learning*, *3*(4), 26–34.

Nazemi, K., Stab, C., & Kuijper, A. (2011). A reference model for adaptive visualization systems. In *Proceedings of the 14th international conference on human-computer interaction: design and development approaches, Part I*, HCII'11. Orlando, FL (pp. 480–489).

Nazemi, K., Breyer, M., Burkhardt, D., Stab, C., & Kohlhammer, J. (2013). SemaVis—a new approach for visualizing semantic information. *Towards the internet of services: the Theseus project*.

Nguyen, T., & Zhang, J. (2006). A novel visualization model for web search results. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 981–988.

Paulheim, H., & Fümkranz, J. (2012). Unsupervised generation of data mining features from linked open data. In *Proceedings of the 2nd international conference on web intelligence, mining and semantics*, WIMS '12 (pp. 31:1–31:12).

Schenk, S., Saathoff, C., Staab, S., & Scherp, A. (2009). SemaPlorer—interactive semantic exploration of data and media based on a federated cloud infrastructure. *Journal of Web Semantics*, *7*(4), 298–304.

Shadbolt, N., Berners-Lee, T., & Hall, W. (2006). The semantic web revisited. *IEEE Intelligent Systems*, *21*(3), 96–101.

Stab, C., Nazemi, K., & Fellner, D. W. (2010a). SemaTime—timeline visualization of time-dependent relations and semantics. In *Proceedings of the 6th international conference on advances in visual computing, Part III*, ISVC'10. Las Vegas, NV, USA (pp. 514–523).

Stab, C., Breyer, M., Nazemi, K., Burkhardt, D., Hofmann, C. E., & Fellner, D. W. (2010b). SemaSun: visualization of semantic knowledge based on an improved sunburst visualization metaphor. In *ED-media 2010* (pp. 911–919). Chesapeake: AACE.

Stoyanovich, J., Lodha, M., Mee, W., & Ross, K. A. (2011). SkylineSearch: semantic ranking and result visualization for PubMed. In *Proceedings of the 2011 ACM SIGMOD international conference on management of data*, SIGMOD '11. Athens, Greece (pp. 1247–1250).

Ward, M., Grinstein, G., & Keim, D. (2010). *Interactive data visualization: foundations, techniques, and applications*. Natick: AK Peters.

White, R. W., Bilenko, M., & Cucerzan, S. (2007). Studying the use of popular destinations to enhance web search interaction. In *Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval*, SIGIR '07. Amsterdam, The Netherlands (pp. 159–166).

Wood, D. (2011). *Linking government data*. New York, Dordrecht, Heidelberg, London: Springer.