Visualization of Personalized Faceted Browsing

Michal Tvarožek and Mária Bieliková

Institute of Informatics and Software Engineering, Faculty of Informatics and Information technologies, Slovak University of Technology, Ilkovičova 3, 842 16 Bratislava, Slovakia, {tvarozek,bielik}@fiit.stuba.sk

Abstract: Current user needs and expectations increasingly shift focus away from simple information lookup towards more complex tasks collectively described as exploratory search. This requires the development of novel approaches and tools to search, navigation and processing of information, such as personalized faceted browsing. We describe a visualization approach for personalized faceted browsers suitable for exploratory search tasks in large information spaces, and apply it to the job offers, scientific publications and digital images domains.

1. Introduction

Contemporary applications and users deal daily with huge amounts of data on a daily basis, such as deep web (relational) databases and Semantic Web repositories, or miscellaneous web resources available via web search engines. While effective access to these resources via search and navigation is steadily improving, it can still be argued that the growing size and complexity of the available information space hampers overall user experience.

Moreover, growing user needs and expectations together with advances in information retrieval approaches shift the focus from simple lookup tasks to more complex tasks collectively described as *exploratory search* (Marchionini, 2006). These include *learning* (e.g., comprehension or comparison) and *investigation* (e.g., analysis, discovery, forecasting or transformation). This trend is further emphasized by the younger "Net generation" (i.e., people who grew up with the Web) as their needs, expectations, attitudes and information seeking behaviour are significantly different compared to the "pre-Web generation" (Oblinger, 2005).

In order to address issues of effective information access and processing, we propose information retrieval, human computer interaction, user interfaces, and initiatives such as the Semantic Web and the Adaptive Web as prime candidates for cross-fertilization of approaches for the creation of successful solutions. We

take advantage of semantic search in the Semantic Web environment (Guha, 2003) and provide users with an adaptive faceted browser interface as means for integrated search, navigation and exploration of the available information space.

2. Related Work

Exploratory search tasks involve a broader range of activities than traditional lookup tasks. Users thus need quick and easy access to more advanced methods, tools and features for information search, processing and understanding.

Advanced means for query construction and modification facilitate exploratory search, e.g., view-based search based on faceted browsers was proposed and successfully evaluated in various scenarios (Yee, 2003; Wilson, 2006).

Adaptive user interfaces dynamically adjust system behaviour to specific conditions, which might include device or environment properties, user characteristics or social relations. Typical applications include automatic user interface generation (Dakka, 2005; Oren, 2006) or personalization (Tvarožek, 2007), content adaptation or recommendation, e.g. in educational systems (Brusilovsky, 2004).

Authors in (Wilson, 2008) compare three major faceted browsers which allow users to formulate queries via navigation by successively selecting metadata terms in a set of available facets, and to interactively browse the corresponding search results. mSpace is a domain specific browser of RDF data, which provides users with a projection of high dimensional information spaces into a set of columns (filters), which can be manually added, rearranged or removed by users (Wilson, 2006). Flamenco stresses interface design and guides the user through the information seeking process from a high level overview through query refinement and results preview to the exploration of individual results (Yee, 2003). The BrowseRDF faceted browser supports facet generation from arbitrary RDF data and extends the expressiveness of faceted browsing by providing additional operators in addition to selection and intersection (Oren, 2006).

While in Flamenco the facets are static and predefined, users can manually adapt columns in mSpace to match their needs. BrowseRDF automatically identifies facets in source data based on several statistical measures, yet does not directly address issues of information overload or interface usability and adaptivity.

3. Visualization of Personalized Faceted Browsing

We proposed the extension of "classical" faceted browsers with personalization support based on user characteristics in (Tvarožek, 2007). Our goal was to improve navigation in large/complex information spaces by providing adaptive navigation guidance and orientation support aimed at the specific needs and requirements of individual users ultimately improving overall user experience.

We developed a personalized faceted browser – Factic, which allows users to navigate Semantic Web data collections represented in OWL repositories. Factic belongs to a larger platform aimed at personalization and presentation of semantically enriched data, acquired (semi)automatically from various web resources (Barla, 2007). Factic constructs the user interface based on a faceted classification extracted from the domain ontology and is thus effectively domain independent, provided that useful facets can be generated to populate the user interface.

To address exploratory search in the (Semantic) Web, our faceted browser was primarily designed for large information spaces containing millions of information artefacts with many available facets and restrictions imposing high requirements on the cognitive abilities of users. We assume high structural complexity – many different concepts with complex relations, and information space dynamics indicating frequent and/or unexpected changes in the data and the structure of the information space, also assuming missing or unknown data and metadata about items. We also address high user diversity – many different users with specific backgrounds, needs and levels of expertise.

The main focus of this paper lies in the visualization of personalization of facets and restrictions, which are the users' primary means of navigation in a faceted browser. Therefore, we employ existing successful faceted browser interface concepts such as the overall interface layout with facets on the left, the current query at the top and the search results in the centre (see Figure 1).

We use different kinds of facets based on the data types they operate on (e.g., date, numbers, enumerations, taxonomies) using these personalization steps:
1. *Select available facets* – evaluate the set of available facets and add or remove facets from the global facet pool to the individual user's facet pool, optionally constructing new facets from available metadata via dynamic facet generation.
2. *Determine facet types* – active facets are fully functional and their contents are visible, inactive facets are used for queries, yet their contents are hidden, disabled facets are not used for querying and their contents are hidden. However, inactive and disabled facets and their contents can be access on demand.
3. *Order facets and restrictions* – adapt the facet ordering based on descending relevance and usage statistics, orders the restrictions in facets either alphabetically, by relevance or by their size (number of instances they correspond to).
4. *Recommend facets and restrictions* – evaluate the relevance, expressiveness and "usefulness" of items, recommend the most suitable for further navigation.
5. *Annotate facets and restrictions* – add additional (domain specific) clues and information about individual items (e.g., the size of restrictions, explain the meaning of facets and restrictions, show examples of "what is behind them").

We distinguish individual facet types via background colour. Facets which were already used – a restriction was selected are shown in green, while blue facets are yet unused – without any selected restrictions. All restrictions in active facets are shown, while only the headers of inactive and disabled facets are visible. Two additional icons serve as buttons and indicators allowing users to activate/enable inactive and disabled facets respectively (see Figure 1, left).
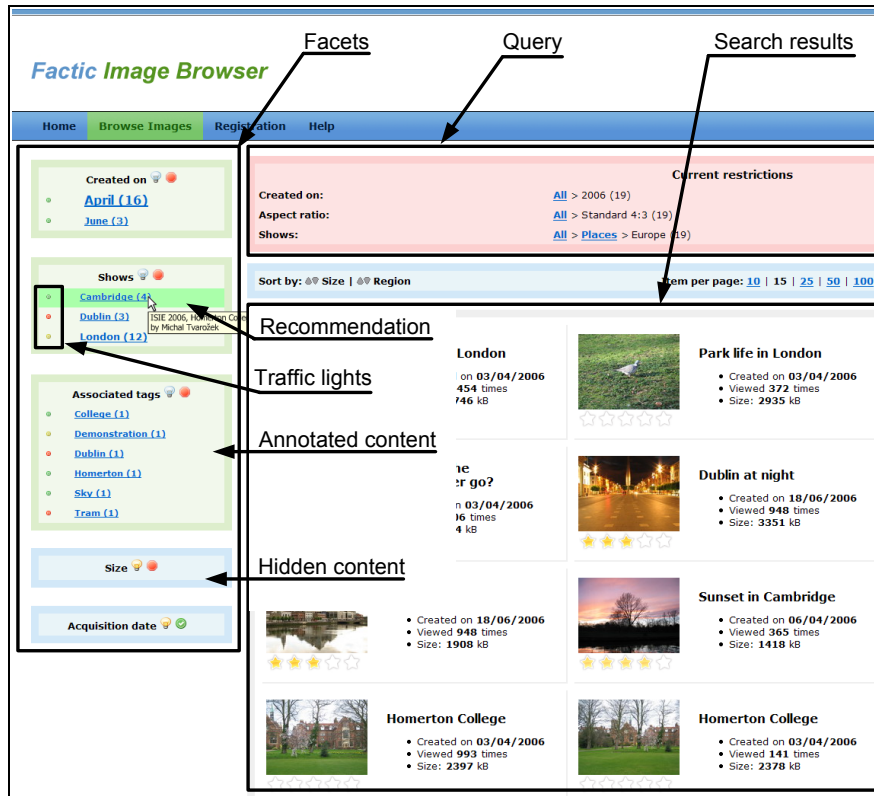
Figure 1 Example of our personalized faceted browser visualization in the digital image domain. Active/inactive facets are shown with different background colours; icons in facet headings indicate facet type and allow users to activate/enable facets on demand. The traffic light metaphor describes restriction relevance, font size indicates the relative number of instances that satisfy restrictions, and numbers indicate the absolute number of results. Restriction background colour denotes recommendations, while tooltips contain additional annotations (e.g., metadata descriptions or (personalized) content summarization).

While the visualization of restriction relevance via ordering was at first a tempting idea, our initial experiments revealed that it was not well accepted by users who were confused by the rather erratic ordering of items (e.g., due to weak relevance estimates) and inability to search them in a natural way.

Consequently, we decided to always sort restrictions alphabetically and display relevance and additional information via link annotation. We show *restriction relevance* using the traffic light metaphor, which has already been successfully used in several adaptive systems (Brusilovsky, 2004). Here, green corresponds to high relevance, yellow for medium relevance and red for low relevance (i.e., less suitable for further navigation). Similarly, in addition to the traditional number of

search results corresponding to individual restrictions, we present restrictions with progressively different font sizes based on the relative number of instances that satisfy them. Thus, the most "rich" restrictions are quickly visible somewhat corresponding to sorting based on restriction size (see Figure 1, left).

To compensate for deep and/or complex faceted classifications and to provide shortcuts to popular items we recommend specific restrictions to users based on their estimated characteristics and social groups. These are presented at the top of the restriction list in each facet with green background colour. We can recommend both restrictions at the current level or deeper in the restriction hierarchy. In order to allow users to search the restriction list while having all recommended restrictions at the top, we duplicate recommended restrictions if they are at the current level of the facet hierarchy. We annotate facets and restrictions via tooltips with their descriptions retrieved as metadata from the used domain ontology, while also summarizing the contents of restrictions, e.g. by presenting some attributes of the corresponding search results that satisfy a given restriction (see Figure 1, left).

Lastly, we provide users with personalized visualizations with additional annotations, which provide different levels of abstraction via hierarchical cluster views, present fewer or more attributes of individual instances in adaptive views, and provide varying levels of navigation guidance and orientation (Tvarožek, 2008). Search results are normally unordered, or ordered via specialized external tools for single-/multi-criteria ordering (Gurský, 2005). Alternatively, context-based (keyword) search can be used to order/search the entire collection as in (Návrat, 2007).

Furthermore, we use horizontal incremental graph navigation in/visualization of concepts, instances and relations (Bieliková, 2007). E.g., the size, relative layout and colour of clusters provides information about instance counts, similarity, relevance (e.g., via a user's social network) and overall suitability (e.g., via user characteristics).

3. Conclusions

We proposed the visualization of an adaptive faceted browser interface suitable for the browsing of large open information spaces and developed the faceted browser Factic for search and navigation in Semantic Web repositories.

We performed experiments in three application domains – job offers (project NAZOU, nazou.fiit.stuba.sk), and scientific publications and digital images (project MAPEKUS, mapekus.fiit.stuba.sk). E.g., our publication ontology was populated with metadata acquired from ACM DL, SpringerLink and DBLP totalling about 985,000 publications (~140k from ACM DL, ~35k from SpringerLink and ~810k from DBLP) which were conceptualized into about 390 classes.

We perform (semi)automatic user interface generation based on a supplied domain ontology describing both the structure of an information space (i.e., metadata ~ facets) and data (i.e., individual information artefacts ~ search results). We ad-

dress issues of large, open or inconsistent information spaces via adaptation employing usage statistics, user preferences and social navigation observations. Since both the visualization and the personalization approaches are effectively domain independent, our browser can be used in other application domains for which a domain description is available.

References

Barla, M., Bartalos, P., Bieliková, M., Filkorn, R., Tvarožek, M.: Adaptive portal framework for Semantic Web applications. In: Proc. of 2nd Int. Workshop on Adaptation and Evolution in Web Systems Engineering, ICWE 2007 Workshops, pp. 87-93 (2007).

Bieliková, M., Jemala, M.: Adaptive Incremental Browsing of Ontology Structure. In: Proc. of the 18th ACM Conf. on Hypertext and Hypermedia, UK, ACM Press, 143-144 (2007).

Brusilovsky, P.: Adaptive navigation support: From adaptive hypermedia to the adaptive Web and beyond. Psychology 2, 1:7-23 (2004).

Dakka, W., Ipeirotis, P.G., Wood, K.R.: Automatic Construction of Multifaceted Browsing Interfaces. In: Proc. of the 14th ACM Int. Conf. on Information and knowledge management, ACM Press, New York, NY, USA, 768-775 (2005).

Guha, R., McCool, R., Miller, E.: Semantic Search. In: Proc. of the 12th Int. Conf. on World Wide Web. ACM Press, New York, NY, USA, 700-709 (2003).

Gurský, P., Lencses, R., Vojtáš, P.: Algorithms for user dependent integration of ranked distributed information. In: Proc. of TED Conference on e-Government, 123-130 (2005).

Marchionini, G.: Exploratory search: from finding to understanding. Communications of the ACM 49, 4:41–46 (2006).

Návrat, P., Taraba, T.: Context Search. In: Y. Li et al.: Proc. of Int. Conf. on Web Intelligence and Intelligent Agent Technology (Workshops), IEEE CS, USA, 99-102 (2007).

Oblinger, D., Oblinger, J.: Is it age or IT: First steps toward understanding the Net generation. Educating the net generation. Educause, www.educause.edu/educatingthenetgen (2005).

Oren, E., Delbru, R., Decker, S.: Extending Faceted Navigation for RDF Data. In: Proc. of the 5th Int. Conf. on Semantic Web. LNCS 4273, Springer, Heidelberg, 559-572 (2006).

Tvarožek, M., Bieliková, M.: Personalized Faceted Navigation for Multimedia Collections. In: Proc. of SMAP 2007, 2nd Int. Workshop on Semantic Media Adaptation and Personalization. CS IEEE Press, 104-109 (2007).

Tvarožek, M., Bieliková, M.: Collaborative Multi-Paradigm Exploratory Search. In: Proc. of Web Science Workshop at Hypertext 2008, ACM Press, NY, USA, (2008), to appear.

Yee, K.P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, ACM Press, New York, NY, USA, 401-408 (2003).

Wilson, M.L., schraefel, m.c.: mspace: What do numbers and totals mean in a flexible semantic browser. In: 3rd Int. Semantic Web User Interaction Workshop at ISWC2006 (2006).

Wilson, M.L., schraefel, m.c., White, R.W.: Evaluating Advanced Search Interfaces using Established Information-Seeking Models. In: Journal of the American Society for Information Science and Technology (JASIST), (2008), to appear.