Personalized Faceted Browsing for Digital Libraries*

Michal Tvarožek, Mária Bieliková

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

Abstract. Current digital libraries and online bibliographies share several properties with the Web and thus also share some of its problems. Faceted classifications and Semantic Web technologies are explored as possible approaches to improving digital libraries and alleviating their respective shortcomings. We describe the possibilities of using faceted navigation and its personalization in digital libraries. We propose a method of faceted browser adaptation based on an automatically acquired user model with support for dynamic facet generation.

1 Introduction

Present day digital libraries (DL) and bibliographies enable users to browse, search in and view their contents via web browsers, and can thus be considered a specific subspace of the Web. Consequently, they share several of its properties:

- *Size* – contemporary DL present large information spaces.
- *Changeability* – DL are much less changeable than the Web. While new titles are continuously added, the content of existing ones and their metadata (e.g., author, title, editor) remain the same. However, a diverse range of different DL exists ranging from simple publications to archaeological artifacts.
- *Complexity* – varying degrees of complexity are common ranging from relatively simply structured to very complex information spaces.
- *User diversity* – varies based on the target audience and DL type, with also the individual users' interests changing over time.

Unlike the current Web, DL display relatively good availability of metadata, which however does not prevent DL users from navigation problems and high recursion rate of navigation, which are common in the general Web [1]. Many existing DL allow users to browse entries by subjects or titles and also support

* This work was partially supported by the Slovak Research and Development Agency under the contract No. APVT-20-007104 and the Scientific Grant Agency of Slovak Republic, grant No. VG1/3102/06.

search with limited personalization capabilities (e.g., ACM DL or IEEE DL). Others use simple faceted browsers that allow users to search for and navigate in available entries using multiple views (e.g., SpringerLink).

Since existing DL offer only limited support for advanced browsing and navigation in individual entries with minimal personalization features in terms of adaptation and user collaboration, new approaches were proposed to address information overload and insufficient search and navigation support. The Semantic Web incorporates shared semantics thus improving interoperability between systems [2], while faceted browsers take advantage of faceted classification and provide combined support for search and navigation as outlined in [3].

2 Adaptive Faceted Browsing for Digital Libraries

To address the disadvantages of faceted browsers insufficient personalization, difficult understanding of the size and content of the information domain, and access to popular topics, adaptation techniques were proposed in [4]. In this paper we explore and extend the possibilities of faceted navigation use in digital libraries and describe enhancements to faceted browsers by taking advantage of ontologies and adaptation based on an automatically acquired user model.

We first determine the relevance of facets and restrictions based on the in-session user behavior (i.e., user clicks), on the user model (i.e., user characteristics described by their *relevance* to the user and the *confidence* in their estimation) and based on global statistics (i.e., all user models). Next, we optionally generate new dynamic facets at run-time and lastly adapt the faceted browser interface in these steps:

1. *Facet ordering* – all facets are ordered in descending order based on their relevance with the last used facet always being at the top (i.e., most relevant).
2. *Active facet selection* – the number of active facets is reduced to 2 or 3 most relevant facets since many facets are potentially available. Inactive facts are used for queries but their contents are not restored, disabled facets are not used at all. Both inactive and disabled facets are still available on demand.
3. *Facet and restriction annotation* – active facets are annotated with tooltips describing the facet, numbers of instances satisfying each restriction and the relative number of instances satisfying each restriction via font size/type.
4. *Facet restriction recommendation* – the most relevant restrictions in a facet are marked as recommended (e.g., with background color or "traffic lights").

Adaptive views. Users can choose from several visualization options by selecting one of the available views – simple overview, extended overview or detailed view, which display increasingly more detailed information about individual search results. The attributes of the displayed instances are adaptively chosen based on their estimated relevance derived from the user model.

Information overload prevention. Based on facet and restriction relevance we reduce the total number of accessible items in order to allow users to find relevant facets and restrictions more efficiently without having to scroll several screens down. The selection of appropriate facet types and displayed restrictions is performed automatically based on their relevance in the user model and based on the current in-session user behavior so that it matches both long-term user interests and short-term user goals.

Query refinement. By using additional facets created by dynamic facet generation, users can refine their queries beyond what would by possible with statically defined facets. Furthermore, these are combined with additional functions often used in advanced search such as OR, NOT or braces. For example, if some users were interested in publications related to a given topic they would select that topic as the subject of the publication in a static facet resulting in publications which deal directly with the given topic. By using a dynamic facet generated from the domain ontology (i.e., one that was not anticipated by the system's administrator), users can instead select publications presented on conferences or in journals that deal with the given topic thus receiving a much broader set of publications which are still related to the given topic.

Orientation support. Since faceted classifications and large information spaces tend to be complex and hard to understand, we annotate facets and restrictions with additional information to aid users in orientation. Facet and restriction annotation includes the number of instances that satisfy a restriction and a textual description of their meaning. Individual restrictions can be further annotated with background color indicating e.g. their relation to users' field of work. Individual search results are annotated using background color, based on their relation to a given set of publications (e.g., already read or the author's own publications) by means of an external concept comparison tool [5].

Guidance support. The proposed method improves user guidance in several ways. First, we order the set of available facets based on their estimated user relevance thus recommending the most relevant facets. Next, we evaluate the relevance of individual restrictions and recommend the most relevant ones based on the user model, e.g. by means of background color. Moreover, we can recommend the most relevant search results by ordering them using external ordering tools [5, 6] to evaluate the relevance of the final search results against the current user model and query.

Social navigation and recommendation. Since the domain of DL is somewhat closely related to social networks of e.g. authors, we take advantage of other users' preferences in the evaluation of concept relevance. *Global relevance* describes the overall "popularity" of concepts while *cross relevance* also considers the similarity between users. Thus we can recommend a publication if it is relevant for many researchers in a particular field and the user is also interested in this field, or a generic publication that seems to be relevant for many users.

Visual navigation and presentation. In order to improve the understandability of the domain and the available data a visual presentation method may be more suitable than pure text. Visual navigation in clusters [7] provides users with the necessary "global" overview of the respective information subspace selected in a faceted browser. Likewise, a seamless transition between an adaptive textual view with support for faceted navigation and a visual view, representing the selected information subspace (e.g., based on clusters), with successive visual navigation can provide users with a more intuitive browsing experience.

3 Conclusions

We presented a novel method of faceted browser adaptation with dynamic facet generation. We evaluated selected part of the proposed method in the domain of scientific publications (project MAPEKUS, `mapekus.fiit.stuba.sk`) by experimenting with developed adaptive faceted browser – *Factic*. Factic is evaluated as part of the personalized presentation layer proposed in [8] where it is integrated with tools aimed at automatic user action logging and characteristics acquisition. Preliminary evaluation showed that the adaptation of facets alleviates some disadvantages of faceted classification, such as difficult access to popular items, and significantly improves overall efficiency by reducing information overload.

Future work will include the design and evaluation of additional method enhancements, especially dynamic facet generation, social navigation in the faceted browser and recommendation based on user relationships.

References

1. Lavene, M., Wheeldon, R.: Navigating the World-Wide-Web. In Lavene, M., Poulovassilis, A., eds.: Web Dynamics, Springer (2003)
2. Shadbolt, N., Berners-Lee, T., Hall, W.: The semantic web revisited. IEEE Intelligent Systems **21**(3) (2006) 96–101
3. Shen, R. et al.: Exploring digital libraries: integrating browsing, searching, and visualization. In: JCDL '06: Proc. of the 6th ACM/IEEE-CS Joint Conf. on Digital libraries, New York, NY, USA, ACM Press (2006) 1–10
4. Tvarožek, M., Bieliková, M.: Adaptive Faceted Browser for Navigation in Open Information Spaces. In: Proc. of WWW 2007, ACM Press, New York, USA (2007)
5. Návrat, P., Bartoš, P., Bieliková, M., Hluchý, L., Vojtáš, P., eds.: Tools for Acquisition, Organization and Presenting of Information and Knowledge, Research Project Workshop, Bystrá Dolina, Low Tatras, Slovakia (2006)
6. Horváth, T., Vojtáš, P.: Ordinal classification with monotonicity constraints. In Perner, P., ed.: Industrial Conf. on Data Mining. Volume 4065 of Lecture Notes in Computer Science., Springer (2006) 217–225
7. Frivolt, G., Bieliková, M.: Topology generation for web communities modeling. In: SOFSEM, Springer, LNCS 3381 (2005) 167–177
8. Tvarožek, M., Barla, M., Bieliková, M.: Personalized Presentation in Web-Based Information Systems. In J. van Leeuwen et al., ed.: Proc. of SOFSEM 2007, Springer, LNCS 4362 (2007) 796–807