Improving Semantic Search via Integrated Personalized Faceted and Visual Graph Navigation

Michal Tvarožek, Michal Barla, György Frivolt,
Marek Tomša, 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
{Name.Surname}@fiit.stuba.sk

Abstract. Growing need for information retrieval, information processing, and the associated need for navigation in existing information spaces resulted in several approaches that aim to improve efficiency of the respective user tasks. However, problems related to user navigation and orientation in large open information spaces still persist possibly due to increasing demands and the imperfections of individual approaches. We propose an integrated search and navigation solution that takes advantage of the faceted browsing paradigm and visual navigation in graphs both extended with support for automatic personalization based on user context also taking advantage of a user's social network. The proposed solution is primarily evaluated in the domain of scientific publications, i.e. digital libraries, with possible extensions to other application domains.

1 Introduction

Growing information processing requirements continuously challenge our information retrieval, navigation and visualization capabilities. Effective use of the available information is becoming ever more difficult as the size, changeability and complexity of the information space increase as do the diversity of users and their requirements.

Consequently, issues such as the infamous navigation problem, high navigation recurrence, information overload or insufficient personalization support arise and significantly decrease overall user productivity in demanding information processing tasks [1]. The typical steps performed by users when seeking information are the query, selection, navigation and query modification. However, most current approaches provide very little support for the crucial navigation step where users locate the desired information. For example, current web search engines are able to locate a set of good starting points for navigation yet offer users no navigation support that would help them in realizing their goals [1].

Moreover, several studies indicated that keyword-based queries tend to be short (two to three words) [2] and that advanced search interfaces of keyword-based search engines are impractical to use [3], while also being unsuitable for

browsing and exploratory tasks [4]. Thus more advanced search and navigation paradigms such as faceted and visual graph navigation were explored:

- *Faceted navigation* used in faceted browsers is based on the use of faceted classification – an orthogonal multidimensional classification of information artifacts, which was originally developed in library sciences [5]. Its basic principle lies in the use of facets, describing individual properties of instances in an information space, to specify the desired properties of instances in the visible information space. The final search query combines restrictions from individual facets via the logical AND function resulting in an unordered list of instances that satisfy all specified restrictions.
- *Graph navigation* is based on a graph representation of the information space. A graphical visualization of the respective (sub)graph of the information space is used, where individual nodes are annotated to aid user navigation. Two modes of operation can be used. Vertical navigation in the graph based on, for example hierarchical clustering of instances, where multiple clustering criteria can be used to select subspaces of the original information space similarly to faceted navigation. Horizontal navigation in the graph allows users to explore the relations between individual instances (nodes), typically employing either a local cluster, centered on a specific instance or a window, which displays instances and relations up to a certain distance.

In this paper we aim to improve existing navigation approaches by combining personalized faceted navigation with visual navigation in graphs, taking advantage of ontologies and users' social networks. Section 2 describes current approaches that improve upon existing navigation methods, while our extensions to faceted navigation and its corresponding extension with visual navigation in search results and associated instances are described in sections 3 and 4. We describe our evaluation in section 5 and summarize the contribution in section 6.

2   Related Work

The field of generic information retrieval has already been explored in much detail resulting in several methods and commercially successful search engines. However, the specific area of navigation also in Semantic Web data has not yet been sufficiently explored. Thus our focus is information navigation in ontological instances taking advantage of semantic metadata available in OWL ontologies.

In [6], the authors describe BrowseRDF – a faceted browser for RDF data, which generates the faceted browsing interface from the supplied RDF data. They extend the faceted query model with new types of queries (e.g., existential, inverse existential, inverse join) and define metrics and algorithms for automatic facet ranking based on statistics computed from the source RDF data to prevent information overload. However, since only RDF data without additional metadata are used, all properties are used as facets disregarding their semantics (e.g., numbers, dates, enumerations, hierarchies, direct/indirect properties

of the desired search results) making efficient use for large information spaces impractical.

Similarly, the Semantic Web portal tool OntoViews [7] publishes RDF content via a predefined faceted browser interface and provides the user with a content-based search engine and link generation/recommendation based on relationships between ontological concepts. OntoViews also presents results in clusters based on the semantic similarity of instances (i.e., sharing some common properties).

Authors in [4] stress the importance of user interface usability and divide the faceted search process into three phases. The *opening* gives users a broad overview of the scope, size and content of the collection, the *middle game* allows users to narrow down the result set by refining the query, the *end game* shows the final search result and allows users to navigate laterally through the collection.

Neither of these solutions however provides personalized features based on individual users' characteristics nor their context. Furthermore, the faceted classification of an information domain can be somewhat difficult to understand resulting in orientation problems, while the access to popular items possibly nested deep within a classification hierarchy can be tedious and impractical. Moreover, from the Semantic Web perspective, none of these approaches take advantage of semantic metadata (i.e., OWL), which in addition to instances also describe the structure of the information domain, thus being of only limited use for search and navigation in OWL ontologies.

Navigation in a graph heavily depends on the adequate visualization of the graph, i.e., on the use of suitable graph drawing styles. In *force-directed* algorithms nodes repel each other while edges are interpreted as springs exerting attracting forces on the nodes they connect. The algorithm then iteratively moves some nodes according to the overall force acting on each of them until a reasonably stable configuration is found [8]. *Layered drawing* is another drawing style where the majority of the edges follows some overall direction that shall be emphasized by drawing as many edges as possible in one specific direction, e. g., form top to bottom.

Drawing clustered graphs is more complicated because of the additional inclusion hierarchy. Possible solutions to use an additional dimension or to emphasize the underlying graph by drawing the inclusion tree as nested regions (inclusion edges are not drawn explicitly, but they are visualized through the geometric nesting of the corresponding regions – mostly rectangles or circles).

Another vital issue is the visualization of graph changes – how efficiently can users follow the expansion or contraction of the graph as they they interact with it. With poor visualization it takes the user considerable time to become familiar with the graph's drawing again. This means that searching for details of interest has to be performed without losing orientation within the graph. According to [9], this is one of the main problems, when exploring large structures. Therefore, often different representations for overview and detail are linked together in such a way that manipulating one view automatically updates the others. When the user searches for some structures of interest in one view, these structures are also accentuated in the other views, e.g. by highlighting or color-coding. Working

with different views at once, the user can effectively change the representations and choose the best for the task at hand.

In our approach, we build upon existing solutions – we assume an ontological domain and user model representation in accordance with the Semantic Web initiative [10], and propose an integrated navigation approach combining faceted navigation and visual navigation in graphs with support for personalization based on an automatically acquired user model.

3 Personalized Faceted Navigation

Faceted browsers provide graphical user interfaces which enable users to interactively select one or more restrictions in the set of available facets thus visually constructing (semantic) search queries via navigation instead of inventing and writing keywords. In practice, faceted browsers can be effectively used for faceted browsing of an information space without a specific goal, e.g. to get an overview of what information is available, or for faceted search if the specific properties of the desired search results are known, e.g. if journal papers on adaptive hypermedia and the Semantic Web not older than 3 years should be returned.

Some disadvantages of faceted browsers include a somewhat more complex user interface, the need to understand the used faceted classification and faceted browsing paradigm as well as the possibly difficult access to popular items, which might be nested deep in the classification hierarchy.

We proposed personalization, adaptation, and the use of ontologies to address the aforementioned shortcomings of "classical" faceted browsers, and to improve user orientation and guidance, and overall search efficiency [11]. The request processing of our adaptive faceted semantic browser, which allows users to navigate in a domain ontology (in OWL format) and adapts the navigation based on an automatically acquired user model [12], is shown in Figure 1.

- *Dynamic facet generation* takes advantage of metadata stored in the domain ontology to create at run-time new facet descriptions relevant to the user's information goals and characteristics in the user model (Figure 1, center).
- *Facet and restriction adaptation, annotation and recommendation* adapt the set of available facets and restrictions based on the in-session user behavior and based on more long term user characteristics stored in the user model also considering the characteristics of other users. Individual facets are hidden or disabled if they seem less relevant to the current user task, or reordered based on their relevance. Restrictions are annotated with additional information (e.g., instance count or relevance based on the user model) and/or recommended (e.g., shown with different background color) (Figure 1, left).
- *Search results annotation and adaptation* improve user orientation and guidance by providing additional information about individual search results (e.g., how close is the topic of a paper to the user's interests) and displaying only relevant attributes of instances. (Figure 1, right)
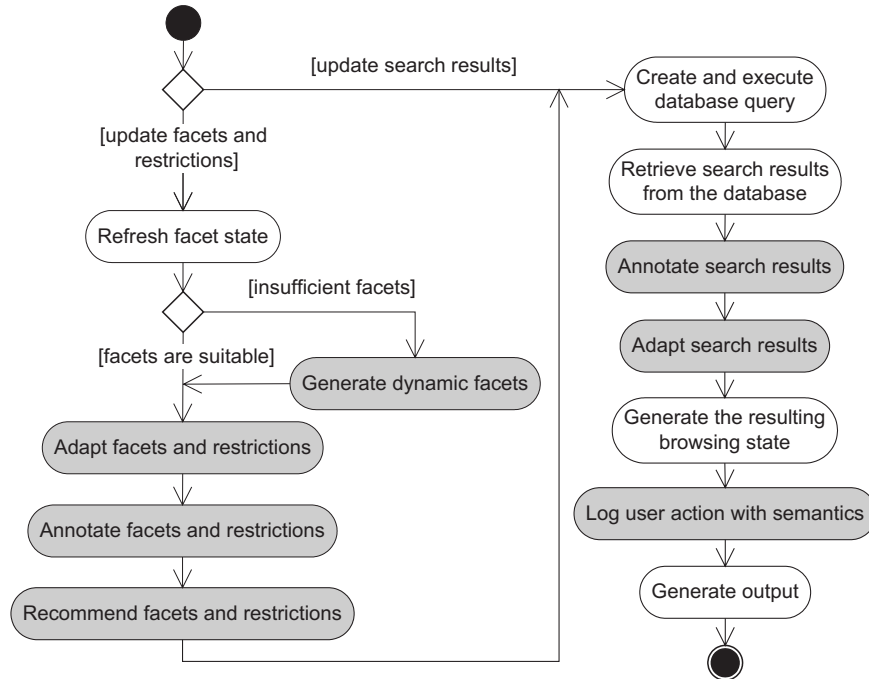
Fig. 1. Request handling of the adaptive faceted semantic browser, extensions in gray.

- *User action logging with semantics* facilitates automatic user characteristics acquisition by logging evidence of user actions with the associated semantics via a user modeling server (Figure 1, bottom right).

We also enhanced the faceted browser with support for nested facets for indirect properties of instances, derived from ontological metadata describing the structure of the domain ontology (i.e., classes), and additional facet combination functions (e.g., OR, NOT).

Since relations between users represented as social networks can significantly improve the user model and thus indirectly improve user experience, we extend the existing relevance evaluation model of the adaptive faceted semantic browser [12] with support for social networks of users who may be associated via different relationships (e.g., friend, coauthor, colleague).

We use social network data to:

- recommend and annotate search results based on social networks of individual users (e.g., sorting of results based on others' ratings);
- recommend and annotate restrictions based on social networks of individual users (e.g., highlighting restrictions popular with the user's colleagues);
- create new facets and restrictions based on social networks of individual users (e.g., a facet which restricts the information space to publications which were visited and/or rated by the user's colleagues).

4 Integrating Visual Navigation with Faceted Browsers

Faceted browsers typically present detailed information about search results in tables or simple textual overviews. However, this approach becomes ineffective for more complex information spaces, which consist of many tightly interconnected instances (e.g., data in OWL ontologies).

We employ visual navigation in graphs to present data about individual search results (ontological instances) and their respective associated instances. For example, in the domain of scientific publications, papers might represent search results, while authors, conferences, or individual research fields might represent associated instances.

Ontologies serve as initial data sources for both faceted navigation and navigation in graphs which use the semantics of the stored information. Furthermore, we take advantage of personalization based on the user context, which is automatically constructed, e.g., based on observed user behavior by specialized user modeling server and the corresponding user modeling agents, or based on the used device.

Our integrated navigation approach is outlined by these steps:

1. The user starts a navigation session by using the faceted browser interface and narrowing down the visible information space via the use of facets and restrictions.
2. The list of search results is displayed either in "classical" textual visualization or using a personalized hierarchically grouped textual or visual view based on a given grouping/clustering criterion. During this stage, the user can either further refine the query via facets or by selecting a group/cluster of search results.
3. Once the level of individual instances is reached, the user can view detailed information about specific instances and their associated instances using visual navigation in the graph starting from the selected search result instance. Alternatively, a (nested) "classical" textual table of attributes and their values can be used.

4.1 Hierarchical Search Results Refinement

Currently the most predominant browsing model is based on hierarchically organized information spaces. Multidimensional faceted navigation integrated with hierarchical search results refinement introduces new alternative views on the available information spaces. Moreover, a visual view of the information space restricted via facets, based on a cluster hierarchy, improves user orientation.

We visualize the internal structure, i.e. the relations between the nodes in clusters (see Figure 2) by showing related vertices closer to each other. Since we visualize only two levels of clustering simultaneously – the current level and the level directly below, multiple layers are not mixed together and thus do not confuse users. This is further supported by visualizing the relations between clusters on the current level only implicitly via the structure of the next level.
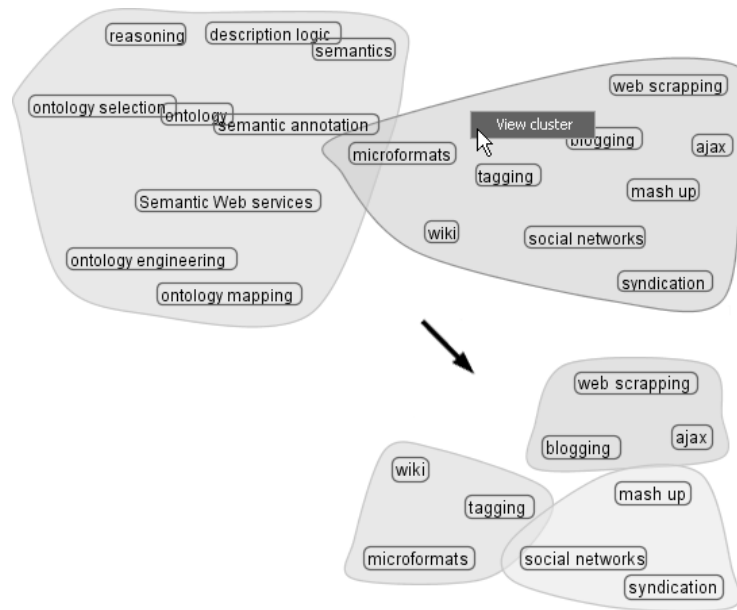
Fig. 2. Example of visual graph navigation in clusters. After selecting the right cluster (top), the view shows its internal structure divided into three smaller clusters (bottom).

Since users choose clusters for further navigation based on their annotations and the annotations of individual nodes, their proper choice is crucial for any successful visual navigation approach. Node and cluster annotations must be distinctive from other annotations of and should clearly describe the respective instance or instances (for a cluster).

In the domain of scientific publications, nodes are authors, which are connected by edges if a relation such as co-author, friend, colleague, or reviewer exists between them. The authors' publications are indexed using keywords and topics. Consequently, clusters are annotated with the keywords and topics of the respective authors' publications contained within a cluster.

These visualization techniques describing the features of nodes are used:

- *Tooltips*[1] provide extended information about nodes. In the domain of publications this information may include the most frequent authors or institutions, number of publications in the cluster or the most frequent conferences, workshops where the papers where presented.
- *Font size* emphasizes the size of the cluster (i.e., the number of instances).
- *Background color* used for personalization indicates suitability similarly as in [13]. For example, clusters often visited by related users can be recommended over less popular clusters.

---

[1] Tooltips are small boxes with supplementary information that appear when the user hovers the cursor over an item.

4.2 Horizontal Navigation in Instances

Horizontal navigation is traditionally considered as the main benefit of hypertext. Browsing instances/groups of instances visually by taking advantage of their similarity improves orientation and usability over the classical tabular view of faceted browsers.

In [14] we proposed incremental browsing of an ontology's structure. We use the same principle is for instances/group of instances visualized in a graph. In every moment only a specific part of the information space – a window is visible, based on the evaluation of the presented entities (e.g., ontology instances). The rest of the graph is visualized on demand when needed (e.g., based on further user clicks navigation). Entity evaluations are computed based on user interaction with the ontology (e.g., people browsing manually or applications).

4.3 User Modeling

The acquisition of user characteristics is fundamental for personalization and can be performed implicitly, e.g. by monitoring and analyzing an individual user's activity, by evaluating his or her activities in the context of social relations, or explicitly by getting feedback from users either by answering manually prepared or system generated questions based on a domain ontology, or by evaluating other forms of feedback (numeric content rating). All knowledge about the user is stored in a centralized user model separated from the rest of the adaptive system using a user modeling server.

Characteristics based on social networks. We consider social networks to be an interesting and valuable source of characteristics about users since we can infer information from the size and nature of the social network or import user characteristics from other known members of the network.

One important property of the scientific publications domain is that users who search for publications can also be authors. Furthermore, there are social relations between authors which can be reflected in a social relation model between users and used for personalization.

Present day social portals usually provide only a single type of relation – *a-friend-of*. However, it seems appropriate to take advantage of typed relations (e.g., *a colleague*, *close friend*, *co-author* and *friend*) for potential publication authors and use these typed relations to detect additional user interests or to make recommendations. Individual relations can be strongly-typed (i.e., defined by an ontology) or labeled as tags created by users.

Detection of social relations can be automated or performed manually using forms. Certainty can be achieved only if users provide explicit confirmation of detected relationships.

Authorship identification. A possible exception is the co-authorship relation, which can be reliably detected without its explicit definition. However, we use

explicit confirmation of publication authorship to eliminate cases of name coincidence or ambiguity.

Information about publication authorship is gained from users during their profile creation. If a user claims to be the author of at least one scientific publication he or she gets a list of publications from the information space he or she may have possibly authored. This list is obtained by selecting publications with author names similar to or derived from the name provided by the user, who then confirms publication authorship by selecting and checking publications from the offered list, thus updating the corresponding user model. If all selected publications share the same author instance, equality between the user identity and the author instance in the domain model can be assumed.

Usage of authorship information to solve the cold start problem. We derive initial weights for user interests from properties (keywords, index terms etc.) of publications for which a user confirms authorship. Furthermore, using co-authorship information we extend interests in the user model provided that people who co-authored some publication share some common interests.

Navigation-related characteristics The filling-in of questionnaires is often considered an annoying and time-consuming activity. Therefore, we minimize user involvement required for the user modeling process and employ automated means of user model construction based on the monitoring and evaluation of user behavior.

Our approach is similar to other user modeling server based approaches [15]. Each (adaptive) application (either the faceted browser, cluster navigator or their combination) sends records about user actions (evidence) performed within the application to a user modeling server. Each record contains information about the respective event, its attributes and the consequence of the event (change of the user interface) as well as the timestamp and user–session identification [16].

Our logging approach is unique as it preserves the semantics of user interaction and allows for better separation (and thus re-usability) of user modeling functionality from the rest of the system. Interaction records are processed by user modeling agents which perform knowledge inference based on the collected evidence and update the user model respectively. Our inference agent is driven by a set of rules [17] which were devised to capture various interesting navigational, domain and system dependent or independent patterns, and to connect them with a set of consequences on user characteristics stored in the user model. The example of a rule (heuristic) is e.g., *result browsing* when user starts to explore the details of displayed instances *without* changing any restrictions in the faceted browser. We can reason that the currently selected restrictions represent user interests and incorporate this knowledge into the user model.

5 Evaluation

We evaluate the proposed approach in the domain of scientific publications in project MAPEKUS (`mapekus.fiit.stuba.sk`), where we developed the *Publi-*

cation Presentation Portal – P3 based on the portal engine part of the Cocoon framework (`cocoon.apache.org`).

P3 provides its users with several options for exploring and navigating in the information space in accordance with the personalized presentation layer approach [18]. Users use either a special portlet for faceted browsing by means of the tool *Factic* (see Figure 3) or a portlet for visual graph navigation (tool *Cluster navigator*), which operates on domain clusters prepared by the *Clusterer* tool. Graph navigation is realized using Prefuse (`prefuse.org`) – a Java framework for graph visualization, which works as a Java applet. A third portlet allows users to use the combined approach of faceted and graph navigation. All presentation tools can take advantage of an ontology-based user model and contribute to it by logging user interaction via the user modeling server part of the system.

Fig. 3. Sample GUI of an enhanced faceted browser.

Our underlying publication ontology was populated with metadata acquired from two well-known digital libraries (ACM DL and SpringerLink) and the scientific publications portal DBLP. We wrapped metadata about 83,000 publications (A-Box) (49,000 from ACM DL, 35,000 from SpringerLink, and 48,000 from DBLP) which were conceptualized into 390 classes in our ontology (T-Box). We used Sesame (`openrdf.org`) as an RDF data store.

For graph navigation, we extract instances of the class `Author` from the domain ontology, while the edges between the authors are defined by the property `isAuthorOf`. As the graph is too large, we applied off-line clustering of the network. Raising the number of instances stored in Sesame led to serious degradation of user experience with the whole system since query response times increased exponentially. We consider the lack of proper and mature technologies for Semantic Web applications as a serious drawback for successful widespread acceptance of RDF/OWL and the Semantic Web.

6 Conclusions

We presented an original approach for personalized navigation via integrated faceted and visual graph navigation in Semantic Web data taking advantage of personalization together with a user's social network. Semantic Web technologies allowed us to enhance the traditional approaches either directly, for example by introducing features like dynamic facet generation or indirectly by enhancing the underlying data (e.g., intelligent clustering of semantically described data).

An important feature of our approach is adaptivity. Both navigation approaches log user interaction via the user modeling server and take advantage of the created user model to improve the user's browsing experience and offer navigation support and guidance.

References

1. Levene, M., Wheeldon, R.: Navigating the World-Wide-Web. In Levene, M., Poulovassilis, A., eds.: Web Dynamics: Adapting to Change in Content, Size, Topology and Use, Springer-Verlag, Berlin Heidelberg (2004) 117–151
2. Jansen, B.J., Spink, A., Bateman, J., Saracevic, T.: Real life information retrieval: a study of user queries on the web. SIGIR Forum **32**(1) (1998) 5–17
3. Markkula, M., Sormunen, E.: End-user searching challenges indexing practices inthe digital newspaper photo archive. Inf. Retr. **1**(4) (2000) 259–285
4. Yee, K.P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: CHI '03: Proc. of the SIGCHI conf. on Human factors in computing systems, New York, USA, ACM Press (2003) 401–408
5. Wynar, B.S., Taylor, A.G.: Introduction to Cataloging and Classification. Libraries Unlimited Inc. (1992)
6. Oren, E., Delbru, R., Decker, S.: Extending faceted navigation for rdf data. In Cruz I. et al., ed.: LNCS: The Semantic Web – ISWC 2006. Volume 4273., Springer-Verlag Berlin heidelberg (2006) 559–572

7. Mäkelä, E., Hyvönen, E., Saarela, S., Viljanen, K.: Ontoviews - a tool for creating semantic web portals. In McIlraith, S.A., Plexousakis, D., van Harmelen, F., eds.: Int. Semantic Web Conf. LNCS 3298, Springer (2004) 797–811

8. Raitner, M.: Efficient Visual Navigation of Hierarchically Structured Graphs. PhD thesis, University of Passau (2004)

9. Schulz, H.J., Schumann, H.: Visualizing Graphs - A Generalized View. Tenth International Conference on Information Visualisation (IV'06) **0** (2006) 166–173

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

11. Tvarožek, M., Bieliková, M.: Adaptive faceted browser for navigation in open information spaces. In: WWW'07: Proc. of the 16th Int. Conf. on World Wide Web, New York, NY, USA, ACM Press (2007) 1311–1312

12. Tvarožek, M., Bieliková, M.: Personalized faceted navigation for multimedia collections. In: SMAP '07: Proc. of the 2nd Int. Workshop on Semantic Media Adaptation and Personalization. (2007) Accepted.

13. Brusilovsky, P., Rizzo, R.: Map-based horizontal navigation in educational hypertext. In: HYPERTEXT'02: Proc. of the 13th ACM Conf. on Hypertext and Hypermedia, New York, NY, USA, ACM Press (2002) 1–10

14. Bielikova, M., Jemala, M.: Incremental visual browsing of ontology structure based on metadata evaluation and usage. In: HYPERTEXT'07: Proc. of the 13th ACM Conf. on Hypertext and Hypermedia, ACM Press (2007)

15. Yudelson, M., Brusilovsky, P., Zadorozhny, V.: A User Modeling Server for Contemporary Adaptive Hypermedia: An Evaluation of the Push Approach to Evidence Propagatation. In Conati, C., McCoy, K., Paliouras, G., eds.: UM 2007. LNAI 4511, Corfu, Greece (2007) 27–36

16. Andrejko, A., Barla, M., Bieliková, M., Tvarožek, M.: User Characteristics Acquisition from Logs with Semantics. In Kelemenová, A., Kolář, D., Meduna, A., Zendulka, J., eds.: ISIM 07: 10th Int. Conf. on Information System Implementation and Modeling., Hradec nad Moravicí, Czech Republic (2007) 103–110

17. Barla, M., Bieliková, M.: Estimation of User Characteristics using Rule-based Analysis of User Logs. In: Data Mining for User Modeling, Proc. of Workshop held at the Int. Conf. on User Modeling UM 2007, Corfu, Greece (2007) 5–14

18. 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