Personalized Exploratory Search in the Semantic Web

Michal Tvarožek
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava
Ilkovičova 3, 842 16 Bratislava, Slovakia
tvarozek@fiit.stuba.sk

Supervisor: Mária Bieliková
bielik@fiit.stuba.sk

Abstract. Effective access to information on the Web, which has become vital to many users and to the whole society, is being hampered by information overload, unavailability of information, navigation issues and user diversity. We aim to facilitate the slow adoption of the Semantic Web by devising an enhanced faceted semantic browser with support for multi-paradigm exploration, personalized recommendation and adaptive view generation. We employ facet and restriction selection, ordering and annotation to address information overload and user guidance, and adaptive view generation with incremental graph visualization to enable end-user grade exploration of semantic web content. We present the highly promising results of two user studies performed with our browser prototypes in the job offers and digital image domains, which confirm the viability and practicality of our approach in terms of improved task times and user understanding of the explored information space.

1. Problem & Motivation

The Web has become a global ubiquitous socioeconomic space providing information, services and facilitating both private and business communication with an estimated 1.8 billion users (survey by Miniwatts Marketing Group, 2010). Many use the Web daily as an integral part of their work, ultimately making the human society as we know it today dependent on the storage, availability and exchange of information on the Web. Due to the constant growth, complexity and size of the Web, several issues hamper online user experience:

- **information overload** (i.e., too much information being available),
- **the lack of information availability** (i.e., the required information is available somewhere in the Web, but unavailable to users who need it),
- **the navigation problem** (i.e., users losing track of their position in the information space effectively "getting lost in hyperspace"),
- **ignorance of user diversity** (i.e., the fact that sites are created to suit the "average user").

The *Semantic Web* aims to provide better search and browsing capabilities by enabling machine readability of information on the Web taking advantage of ontologies [8]. *Adaptive Web* approaches aim to learn about user preferences and adapt the user experience to the specific needs of individuals [3], while the *Social Web* aims to facilitate communication and collaboration on the Web by exploiting social wisdom to improve individual user experience [9]. The *Exploratory search* initiative aims to provide users with better tools for advanced information seeking tasks such as learning, investigation and analysis [6].

We believe that Semantic Web approaches would effectively address several of the aforementioned issues. However, despite continuous progress in semantic technologies, the original promise of the Semantic Web still remains unrealized mainly because there are few real-world applications that allow end-users to access, view and process Semantic Web information [8]. Since there has been little cross-fertilization between the aforementioned initiatives, our aim is to facilitate Semantic Web adoption by providing an *end-user grade exploratory browser for the Semantic Web* taking advantage of adaptive social and exploratory approaches.

2. Background & Related Work

Due to the strong multidisciplinary character of our work and the novelty of exploratory search, there are, to the best of our knowledge, no other approaches our work could be directly compared to. Thus we focus on partial related approaches from individual disciplines with respect to current challenges in *visualization* (raw data, no default visualization), *querying* (complexity of semantic queries) and *exploration* (complex highly interlinked graph of resources) of the Semantic Web.

We chose faceted browsers as the fundamental part of our approach as they have already been explored in various fields including HCI, Semantic Web and exploratory search as they integrate search and browsing, while providing support for query construction. Wilson and schraefel compared three prominent faceted browsers – Flamenco, mSpace and RelationBrowser++ [12], while Oren described the faceted browser BrowseRDF [7]. From these only mSpace and BrowseRDF use RDF data, but do not exploit the more expressive RDFS/OWL variants. Additionally, BrowseRDF provides elementary facet generation capability over simple RDF as it automatically identifies facets based on several statistical measures, but offers only very limited interaction options. None of these solutions support adaptation and thus user modeling and personalization in any way, except for manual reordering of columns in mSpace.
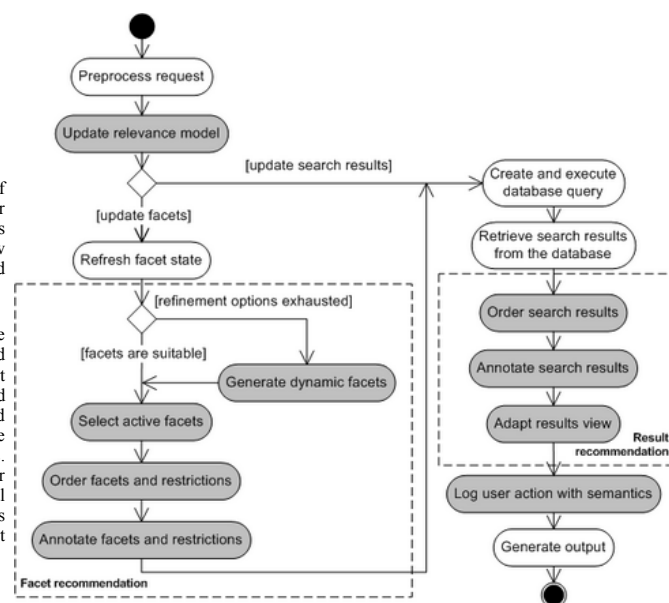

Figure 1. Request handling of the enhanced faceted browser, extensions shown in gray.

Neither BrowserRDF nor mSpace can be effectively used for complex interactive exploration of Semantic Web content, which in addition to faceted querying needs to support interactive information visualization and exploration of graphs (Semantic Web being a graph). The VisGets interface supports interactive/exploratory web search by querying it in three dimensions – time, location and topic, while also providing advanced visualization of search results [4]. However, VisGets is limited to the three predefined dimensions and does not account for information space changes and semantic content. These issues have been addressed by the generic semantic browser Tabulator [2], which allows the exploration of linked data, but again lacks the capability to perform complex search queries and adapt to individual users' needs, while being too complex to use by end-users due to its interface and installation issues.

Consequently, in our opinion, no existing approach has so far provided an effective *end-user grade visualization, querying and exploration* solution for (dynamic) semantic web content.

3. Uniqueness of the Approach

We devised a *comprehensive faceted exploration browser for the Semantic Web*, which acts as an integrated tool for search when it acts like a client-side semantic search engine

front-end, and for navigation when it supports navigation across a collection of "pure" information artifacts accessed via a semantic endpoint. Our browser is based on these principles:

- **Semantic information space representation** – we use a domain ontology (e.g., in RDFS or OWL as defined by W3C) to represent both the data and the metadata describing the structure of the information space. We also employ a user ontology which stores user models describing individual users' characteristics, and an event ontology which describes the events that occur in the browser and its states during user interaction so that they can be used for subsequent automated user characteristics acquisition.
- **Multi-paradigm exploration**, which integrates view-based faceted search with content-based (query-by-example) search and traditional keyword-based search to provide users with the most suitable means to create queries or navigate the information space. This also includes visualization and navigation options for the browsing of search results such as result lists, result attribute tables and attribute/thumbnail matrices, incremental graph visualization and history visualization for revisitation and orientation support.
- **Visual query construction**, which provides support during the construction and modification of complex semantic queries in an intuitive user friendly way.
- **Personalized recommendation** to address information overload, provide guidance during complex information seeking sessions and compute relevance for interface generation.
- **Adaptive view generation**, which facilitates the generation of user interfaces necessary for exploration including facets, result overviews and exploration views, accommodates for the dynamics of the information space and preferences of individual users.
- **Collaborative content/meta-data creation** to harness the power of social wisdom.

Our enhanced faceted semantic browser extends the typical request handling of faceted browsers with additional steps that perform specific tasks as outlined in Figure 1. Due to space constraints we focus on personalized recommendation and provide only a brief description of the remaining aspects of our solution.

3.1 Model for relevance evaluation

Our browser logs events that occurred as results of user interaction with the current state of the browser via a specialized *semantic logging* service which preserves the semantics of events as opposed to traditional web server logs, which store them only implicitly in request URLs (Figure 1, bottom right). The acquired events are processed by a separate user modelling back-end [1] and in turn retrieved as an updated user model, which drives our personalization engine (Figure 1, top left). Each logged event uses our event ontology to specify the semantics of the respective user action and also references the domain and user ontologies as required.

The user modelling back-end provides us with several sources of adaptation, which we employ with different weights depending on how closely related they are to the current user task:

1. **In-session user behaviour** – user navigation, facet and restriction selection during the current user session (i.e., user clicks). Frequent use of specific items indicates higher relevance to the current task and/or user interest in the corresponding domain concepts.
2. **Short/long term user model** – user characteristics acquired during multiple sessions described by their relevance to the user and the confidence in their estimation in the range <0,1>.
3. **Similar/related user models** are assumed to belong to users with similar needs and are thus used for relevance evaluation if user specific data is unavailable or has low confidence. Social user context can be exploited by assigning custom weights to specific relations between users resulting in social recommendation. Moreover, if usage data about other users are "publicly" available, users might directly browse the trails of their peers (e.g., see what images their friends viewed or what papers their colleagues downloaded).
4. **Global usage statistics** computed from the overall relevance and usage of individual domain concepts (e.g., facets, restrictions, target objects – be it images, publications or job offers) from all user models. The overall "popularity" of facets and restrictions increases the likelihood of their recommendation for a specific user, especially if his or her specific preferences are unknown or have low confidence.

Let $L_U(X) = relevance_U(X)$ be the local relevance of concept X from the domain ontology for user U. For example, X might be a facet, a restriction or a search results. We define $C_U(X)$ as the cross relevance of X determined as the average local relevance for all users V weighted by their similarity $sim(U,V)$ to user U. As an alternative and/or addition to cross relevance, we also use the weighted social relevance $\check{C}_U(X)$ if social network data for a specific relation $rel(U,V)$ are available. We evaluate user similarity $sim(U,V)$ from $<0,1>$ as the sum of square differences in concept relevance between users:

$$C_U(X) = \frac{\sum_{V \in users}(sim(U,V) * L_V(X))}{1 + \sum_{V \in users} sim(U,V)}, U \neq V \qquad \check{C}_U(X) = \frac{\sum_{rel(U,V) \in relations}(w(rel) * L_V(X))}{|rel(U,V)|} \qquad sim(U,V) = 1 - \frac{\sum_{X \in concepts}\left(L_U(X) - L_V(X)\right)^2}{|concepts|}$$

Let $G(X)$ be the global relevance of X defined as its mean local relevance for all users. Static relevance $S_U(X)$ defines the relevance of concept X based on the user model and the respective confidence in the relevance estimation. Total relevance $T_U(X)$ defines the total relevance of concept X based on the user model and the current in-session user behaviour derived from the total number of clicks on that concept type – e.g., a facet or a restriction:

$$G(X) = \frac{\sum_{V \in users} L_V(X)}{|users|} \qquad S_U(X) = L_U(X) * confidence_U(X) + (C_U(X) + G_U(X)) * (1 - confidence_U(X)) \qquad T_U(X) = S_U(X) + \frac{Clicks(X)}{1 + TotalClicks}$$

3.2 Personalized recommendation

With personalization, we empower users to make their own decisions more effectively via additional annotations while also providing sensible means of automatic adaptation. As opposed to most existing approaches, we perform *personalization primarily on the client side* (i.e., in the client browser), which has two benefits:

- **Personally sensitive data is kept entirely on the client side** thus preserving privacy. Optional server-side user modelling and statistics tracking can be enabled to further improve user models and provide social information to authorized users.
- **Server-side services need not have support for adaptation** as it is performed by our browser on the client side, thus providing personalization for all information resources at no additional cost.

Facet recommendation distinguishes three types of facets adapted at run-time to the specific needs of individual users – *active facets*, *inactive facets* and *disabled facets*. The adaptation process first determines the relevance of individual facets and restrictions in our relevance model and then uses it in these steps (see Figure 2):

1. **Active facet selection** – the total number of active facets is reduced to a relatively low number, e.g. 2 or 3 facets, since many facets are potentially available in complex information spaces. Active facets are selected based on relevance, recency and number of accesses. The rest of the facets is made inactive or left in disabled state.
2. **Facet and restriction ordering** – all facets are ordered in three groups (i.e., active, inactive, disabled) in descending order based on their relevance with the last used facet always being at the top. Restrictions are ordered alphabetically, since alternative orderings based on relevance or the number of matching search results were not well accepted by users as they made it difficult to search for specific items.
3. **Facet and restriction annotation** – active facet restrictions are annotated with the number of matching instances, the relative number of matching instances by means of font size/type, or directly recommended (e.g., with background colour or the "traffic lights" metaphor) effectively providing shortcuts to deeply nested restrictions. Additional tooltips can describe individual facet/restriction meanings (e.g., the *rdfs:comment* annotation in ontologies).

Search result recommendation extends the processing of search results with support for personalized result ordering, annotation and view adaptation (Figure 1, right). We employ

external tools that evaluate the relevance of individual search results, e.g., by means of concept comparison with the user model or via the evaluation of (explicit) user feedback. Subsequently, we reorder the search results or annotate them with additional information.

Figure 2. Example of facet adaptation, annotation and restriction recommendation showing active, and inactive facets (left), also showing a list view of search results with attributes and additional operations (right).

3.2 Adaptive view generation

3.2.1 Facet generation

During the facet identification stage, we examine metadata describing the information space, identify object and literal facet templates, and select either an enumeration or hierarchical restriction template to use based on ontological metadata. The facet construction stage determines the interaction mode based on the overall number of potential restrictions; *list mode* is used for a small number of predefined values (e.g., days of the week), *search mode* is used for large numbers of values (e.g., all cities on Earth). If an ordering of values is defined in the ontology for object values, we also create restriction intervals to cover continuous values (e.g., real numbers or dates). The last facet mapping stage selects a suitable user interface widget to render the generated facet in the faceted browser, and maps the constructed facet and restriction values onto the widget. The widget provides facet visualization (see Fig. 3) and handles user interaction forwarding events and facet metadata to the back-end search services, which provide the corresponding querying services for the generated facet. Thus facet generation defines these facet properties:

- A **facet template**, which corresponds to a pattern found in domain metadata and specifies the overall type and behavior of the facet.
- A **restriction template**, which defines how the individual restrictions in the facet are constructed and mapped onto the domain ontology.
- A **query template**, which defines how the back-end query engine creates database queries and maps them onto facet restrictions.
- A **visualization and interaction template** (i.e., the corresponding widget type), which binds the facet to the graphical user interface and handles user input.

Figure 3. (A) Generated facets with a list-based result overview showing all result properties (top left). (B) A matrix result overview with image thumbnails and the

corresponding generated annotation pane for collaborative content creation (bottom right).

3.2.2 Result overview generation

We generate two result overviews – the *ListView* shows thumbnails and properties of individual results (see Fig. 3, top left), while the *MatrixView* shows thumbnails, provides additional information in tooltips, and in addition offers a generated editing pane for modification of individual result attributes (see Figure 3, bottom right). *ListView* shows attributes of a specific result directly derived from the domain ontology visualized as *label-value* pairs. For multi-value properties such as *Type* in Figure 3, a column with all values is shown. We either show all result properties to maximize information or apply personalization to select only the most relevant properties. The annotation pane is generated separately; for each specific result type, we identify all applicable properties from the domain ontology metadata, construct editing widgets based on property types (e.g., text boxes with language selection or autocomplete combo boxes, with single/multi-value support). Properties with existing values are shown first, while properties without values are shown at the bottom (see Fig. 3, bottom right).

3.2.3 Graph exploration view generation

We generate the graph exploration view directly from the domain ontology showing resources and their relations, taking advantage of relevance evaluation from the personalization engine and filters manually enabled by the user (Figure 4). Relations are intentionally visualized as separate nodes connecting resources to reduce information overload when one relation can have multiple values and to improve graph layout. We employ a force-based layout algorithm, but also allow the user to fix and manually reposition nodes in the resulting graph.
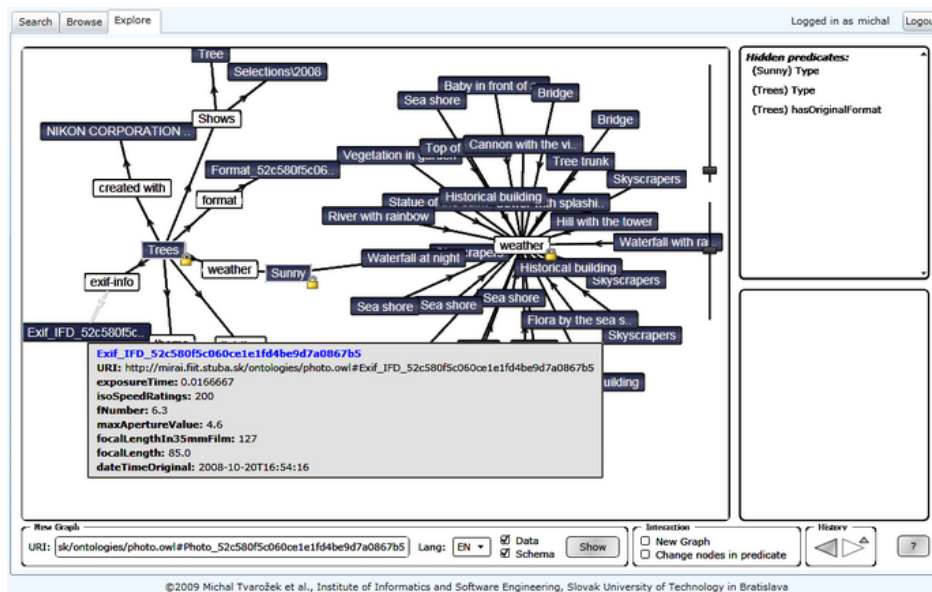

Figure 4. Example of our generated graph-view exploration interface. Dark nodes represent individual resources, white nodes correspond to relations (top). Hovering over nodes shows the attributes of a node (center); additional tools include zooming, spatial expansion, node hiding and history (right), with additional filtering options for languages and data/schema only visualization (bottom).

4. Results & Contributions

4.1 Validation

To validate our approach we performed experiments with two prototypes of our faceted semantic browser Factic. Since exact analytical validation of user-centered approaches is difficult if at all possible, also considering the novelty of the exploratory search field and immaturity of methodologies for task design and browser evaluation [5], our evaluation goals focused on user studies and proof of concept validation of our individual approaches.

We used our initial prototype to evaluate the usefulness our facet personalization approach in the job offers domain via a user study and to gather feedback on its design [11]. We then created a second, improved prototype, which extended the original functionality with support for facet generation, additional result visualization and the graph exploration view, while also addressing performance and usability issues of the first prototype [10]. We used the second prototype in the digital image domain to perform a proof of concept experiment with our facet generation approach, and to perform a user study with our graph exploration approach.

4.1.1 User study of facet personalization

We present some of the experimental results from our first study in the job offers domain, where our approach proved to be particularly suitable, since it is a very complex information space with several deep hierarchical classifications (e.g., regions or positions) and intricate concept relations. We compared three modes of operation – *without adaptation*, *with adaptation* including ordering and hiding of facets, and *with recommendation*, which included ordering, hiding of facets and restriction recommendation. Figure 5 illustrates the total user effort in time and number of user clicks necessary to complete a given scenario, i.e. to find a set of relevant job offer instances.

Our evaluation showed that adaptive selection of active facets (i.e., fully rendered) can significantly reduce information overload (i.e. the number of facets a user must examine) and thus total processing time which depends roughly linearly on the number of displayed facets. However, the number of clicks increased since the right facets were not always active and thus had to be manually enabled. This resulted in shorter refresh times and consequently shorter total task times.
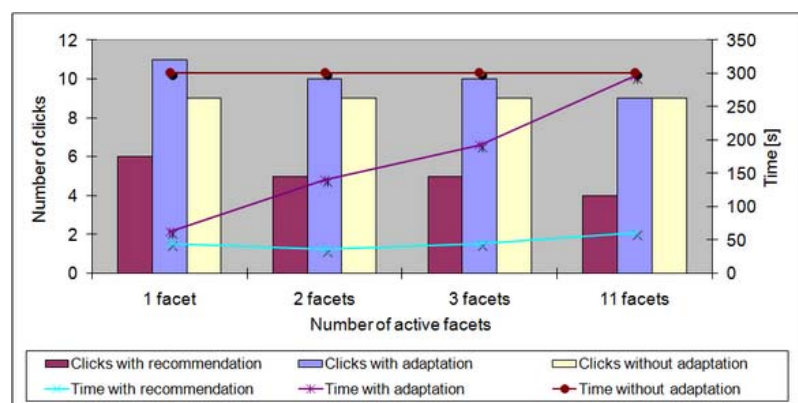

Figure 5. Experimental results for different adaptation modes and different numbers of simultaneously active facets.

Recommendation of suitable restrictions based on the user model further improved total task time and also decreased the number of necessary clicks due to the effective creation of navigational shortcuts that allowed users to skip several clicks by directly selecting suitable restrictions within a restriction hierarchy. As before, the number of clicks increased as the number of active facets decreased as more facets had to be manually activated.

Despite the very positive feedback and highly promising results, we encountered scalability issues with remote repositories due to repository querying limitations and network delays, which we addressed in our second prototype.

4.1.2 User study of graph exploration and facet generation

For this user study, we used our digital image domain ontology with about 150,000 facts describing about 8,000 manually and semi-automatically annotated images including EXIF metadata and additional annotations (e.g., author, object, place, weather). The user study was performed with 10 end-users aged between 20 and 25 years with an IT background, who completed a set of 5 tasks using the browser (e.g., finding a specific image, discovering image properties or getting a better understanding of the domain).

In the proof of concept experiment, we generated facets from the available data and examined how the browser behaved and whether the interface was still usable for its intended purpose in terms of usability and performance. The experiments proved that the approach was viable for interface generation with minimal performance impact. We successfully managed to distinguish facet and restriction templates, direct query templates, and construct and map facets to interface widgets and use them in our exploration interface without any significant negative impact over manually created facets due to facet generation. Note that it is not possible to quantitatively evaluate the "quality" of the generated set of facets, because there is no "best" set of facets.

The user study with the graph exploration interface showed that 9 out of 10 users managed to find the specified image, although the time required varied widely – 141 seconds and 8 clicks were required on average, although the fastest user needed less than 50 seconds while the slowest one required almost 5 minutes. Overall, the users managed to answer 75% of the questions correctly leaving 25% false answers (this also includes answers that were close to the correct ones, but not exactly right). Based on these results, we conclude that graph-based exploration is viable for Semantic Web browsing as most users were able to accomplish the given tasks despite having no prior experience with a similar interface. Still, improvements to layouting and node selection are necessary to improve understandability and task times, which was also confirmed by user feedback which indicates that non-expanded graphs are easy to understand (rating 4.5 on a 5 level Likert scale), while expanded graphs are less readable (rating 3.4).

4.2 Contributions

We devised and evaluated a novel *comprehensive faceted exploration approach for the Semantic Web* with specific contribution to:

- **Facet personalization** – devising a dynamic method of facet and restriction adaptation based on semantic logging of user action and continuous evaluation of the devised ontological user and relevance models.
- **Faceted GUI generation** – devising a method for facet identification in ontological metadata, its transformation into interface widgets and their mapping onto the ontological querying backend (e.g., semantic search engines).
- **Multi-paradigm exploration** – integrating view-based, content-based and keyword-based search with advanced adaptive visualizations and incremental graph exploration.

Our experiments in the job offers and digital image domains have proved the viability of the proposed approaches for their intended purposes in terms of their practicality (i.e., it can be done) and improved user experience (i.e., improved task times, better understanding of the information space, efficient resource revisitation). Consequently, we successfully addressed the original goals of our work:

- **Empowering end-users with access to semantic information spaces** by providing an end-user grade exploratory browser for the Semantic Web with interfaces for effective query formulation, result overview browsing and individual result exploration.
- **Facilitating the adoption of the Semantic Web** by enabling Adaptive Social Semantic Web exploration for end-users via our exploratory search browser.

We see the extension of our approach with support for legacy web content, i.e., for specific pages (e.g., personal browsing history) or for whole web sites (e.g., generating a faceted browsing interface for a typical corporate web site) as a possible direction for future work *ultimately providing a seamless search and browsing experience for both legacy web and semantic web content.*

I have already successfully presented partial results at several international venues during my master's study at Adaptive Hypermedia'06 and my current doctoral study at WWW'07 (poster), ICWE'07 (demo), SMAP'07, WWW'08 (poster), HCIS'08, Web Science Workshop at HT'08, Web Information Retrieval Support Systems Workshop at WI'09.

5. References

1. Barla, M., Tvarožek, M., Bieliková, M.: Rule-Based User Characteristics Acquisition from Logs with Semantics for Personalized Web-Based Systems. Computing and Informatics , 28 (4), 399-427, (2009).
2. Berners-Lee, T. et. al: Tabulator: exploring and analyzing linked data on the Semantic Web, In: Proceedings of the 3rd International Semantic Web User Interaction Workshop, (2006).
3. Brusilovsky, P., Kobsa, A., Nejdl,W., eds.: The Adaptive Web: Methods and Strategies of Web Personalization. LNCS, vol. 4321. Springer, Berlin (2007).
4. Dörk, M., Carpendale, S., Collins, C., Williamson, C.: Visgets: Coordinated visualizations for web-based information exploration and discovery. IEEE Transactions on Visualization and Computer Graphics, 14(6) 1205–1212, (2008).
5. Kules, B., Capra, R., Banta, M., and Sierra, T.: What do exploratory searchers look at in a faceted search interface?. In Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries. JCDL '09. ACM, New York, pp. 313-322, 2009.
6. Marchionini, G.: Exploratory search: from finding to understanding. Communications of the ACM, 49 (4), 41-46, (2006).
7. Oren, E., Delbru, R., Decker, S.: Extending Faceted Navigation for RDF Data. In Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P. et al. (eds.) ISWC 2006: Proc. of the 5th Int. Sem. Web Conf.. LNCS, vol. 4273, pp. 559-572. Springer, Heidelberg (2006).
8. Shadbolt, N., Berners-Lee, T., Hall, W.: The Semantic Web Revisited. IEEE Intelligent Systems, 21 (3), 96-101, (2006).
9. Staab, S., Domingos, P., Mika, P., Golbeck, J., Ding, L., Finin, T., et al.: Social Networks Applied. IEEE Intelligent Systems, 20 (1), 80-93, (2005).
10. Tvarožek, M., Bieliková, M.: Reinventing the web browser for the semantic web. In: WI-IAT '09: Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, IEEE CS, pp. 113–116, (2009).
11. Tvarožek, M., Bieliková, M.: Personalized Faceted Navigation in the Semantic Web. In L. Baresi, P. Fraternali, & G.-J. Houben (Ed.), ICWE 2007: Proceedings of the International Conference on Web Engineering. LNCS 4607, pp. 511-515. Springer, Heidelberg (2007).
12. Wilson, M. L., schraefel, M. C., and White, R. W.: Evaluating advanced search interfaces using established information-seeking models. In: Journal of the American Society for Information Science and Technology, 60 (7), 1407-1422, (2009).