SemWeB Semantic Web Browser – Improving Browsing Experience with Semantic and Personalized Information and Hyperlinks

Melike Şah, Wendy Hall and David C De Roure

Intelligence, Agents and Multimedia Group, School of Electronics and Computer Science, University of Southampton
{ms305r, wh, dder}@ecs.soton.ac.uk

Abstract. Imagine a Web browser that can understand the context of a Web page and recommends related semantic hyperlinks in any Web domain. In addition, imagine this browser also understands your browsing needs and personalizes information for you. The aim of our research is to achieve this in open Web environment using Semantic Web technologies and adaptive hypermedia techniques. In this paper, we discuss a novel Semantic Web browser, SemWeB, which utilizes linked data for context-based hyperlink recommendation and uses a behavior-based and an ontology-driven user modeling architecture for personalization on Web documents. The aim of this research is to bring the gap between the technology and user needs using Semantic Web technologies in Web browsing.

Keywords: Linked Data, Semantic Web, User Modeling, Ontology, Personalization, Natural Language Processing, Semantic Annotation.

1 Introduction

Imagine a Web browser that can guide users with relevant semantic hyperlinks and personalized contents. Such kind of guidance during Web browsing is increasingly getting essential because of the unstructured network structure of the WWW. As the size and complexity of the network is growing, it is getting more difficult to browse the non-linear document space efficiently with a standard Web browser. Therefore, most often users trust on search engines for finding information resources if they can formulate a formal search query. But this is only half of the story. When they found a good information resource, they have to read and understand the page content and often links to relevant pages are not provided (insufficient link problem). On the other hand, different users have diverse browsing needs and adaptive hypermedia systems try to tailor information/links to individual users for better information discovery. In our opinion, personalization is especially needed during Web browsing because of the enormous amount of information available to users. Our hypothesis is that browsing can be supported better by using Semantic Web [1] and adaptive hypermedia [2]. In this way, the deficiencies of the current Web technologies in Web browsing can be improved for benefits of users.

2 Contributions and Research Challenges

Our research investigates and uses different research techniques to provide context aware, personalized information and links to users on open corpus Web. To the best of our knowledge, none of the existing Semantic Web browsers support personalization. In addition, the majority of adaptive hypermedia systems are tested in educational hypermedia domain and their benefits are well demonstrated on closed corpus data sources. Our aim is to achieve open corpus adaptive hypermedia and dynamic linking on Web. Our contributions are to Semantic Web browsers and to open corpus adaptive hypermedia using Semantic Web technologies and linked data. In our approach, we have used linked data as a source of semantic metadata. Linked data is a new trend and growing practice in Semantic Web community. We wanted to illustrate the value of the linked data within everyday Web browsing.

The following research challenges are undertaken while implementing SemWeB: *natural language processing and semantic annotation* to understand context of a Web page, *an extension to a standard Web browser* for embedding dynamic links to Web documents without the control of the author of the page, *creating semantic hyperlinks* from the linked data, a *novel user model* for Web-based personalization, *user modeling architecture* for user profiling and *adaptation module* for personalization.

3 System Design

The system architecture of SemWeB is presented in figure 1. The communication between SemWeB sidebar extension and SemWeB server is based on Asynchronous Javascript and XML (AJAX). We briefly explain system components as follows:

SemWeB Sidebar Extension for Presenting Semantic and Personalized Information: We extended the Mozilla Firefox Web browser with a sidebar to provide context-based semantic hyperlinks and personalization in any Web domain. Users can use sidebar to choose a linked data domain (currently ECS, DBpedia and DBLP), annotate a Web page and add/highlight semantic hyperlinks on the Web page. Using the embedded hyperlinks, users can request semantic links. In figure 2 (a), embedded semantic links are shown.

Natural language processing and semantic annotation: In order to provide relevant semantic hyperlinks, the context of the Web page should be understood. We perform information extraction and semantic annotation according to a linked data domain (e.g. DBpedia, DBLP and ECS) using a modified GATE framework [3]. We extended GATE with a *lookup service* to match the found lexicons to resolvable URIs and an *annotation generation and storage unit*. For each linked data domain that we have tested (e.g. DBpedia, ECS and DBLP (few URI sets)), we analyzed the linked data URIs and their lexicons for information extraction and semantic annotation. Users can annotate a Web page from the SemWeB sidebar extension.

Creating Semantic Hyperlinks: Using the annotated Web page, users can request semantic hyperlinks by clicking on the embedded links on the Web page. We dynamically dereference a linked data URI over HTTP protocol and identify links and information using SPARQL queries. This information is sent to browser in XML

since communication is based on AJAX. We use a presentation vocabulary to communicate with the browser and the created semantic information and links are presented in a new Web page as illustrated in figure 2(b). If the user is logged in, then this information is personalized.

A Novel User Model for Web-based Personalization: We developed a new behavior-based and ontology-driven user model for adaptation, which can be applied to different domains. In the user model, we use seven categories: identification, preference, security, *browsing goal*, *interest*, *expertise* and *browsing behavior* (main contributions are in italic). More information about the user model can be found in [4] [5]. Our user model relates user profiles to dereferenceable linked data URIs, as a result user profiles are pointing to distributed resources on the Web and can be expanded with any semantic instance.

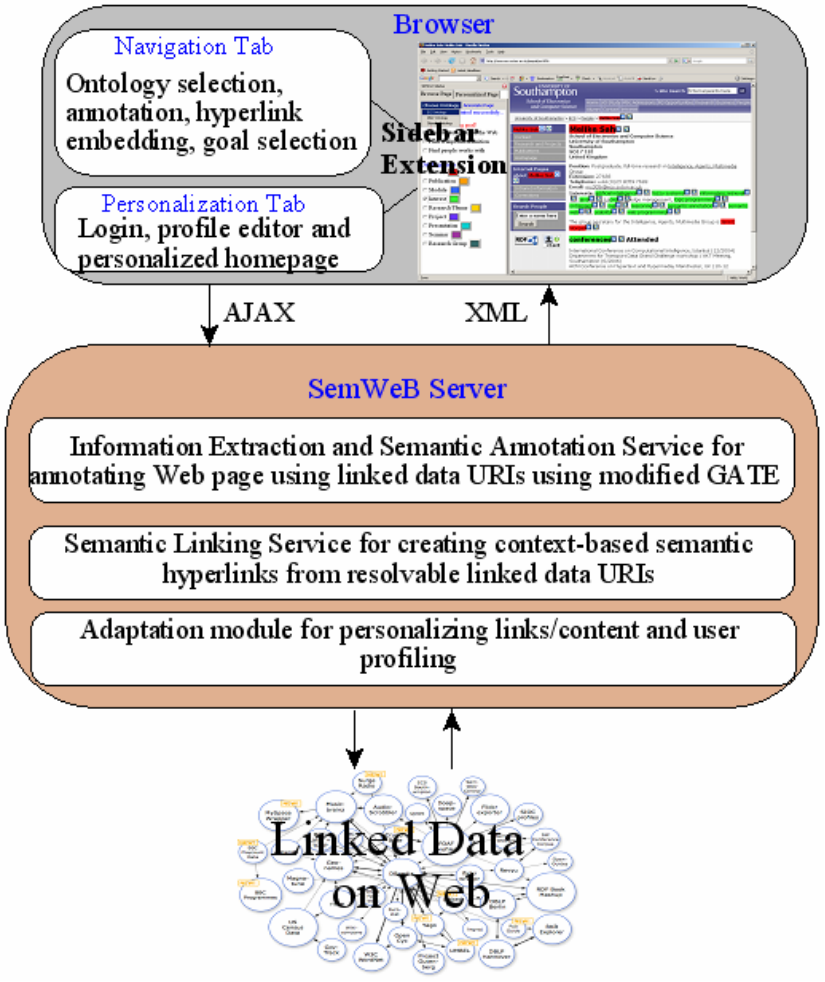
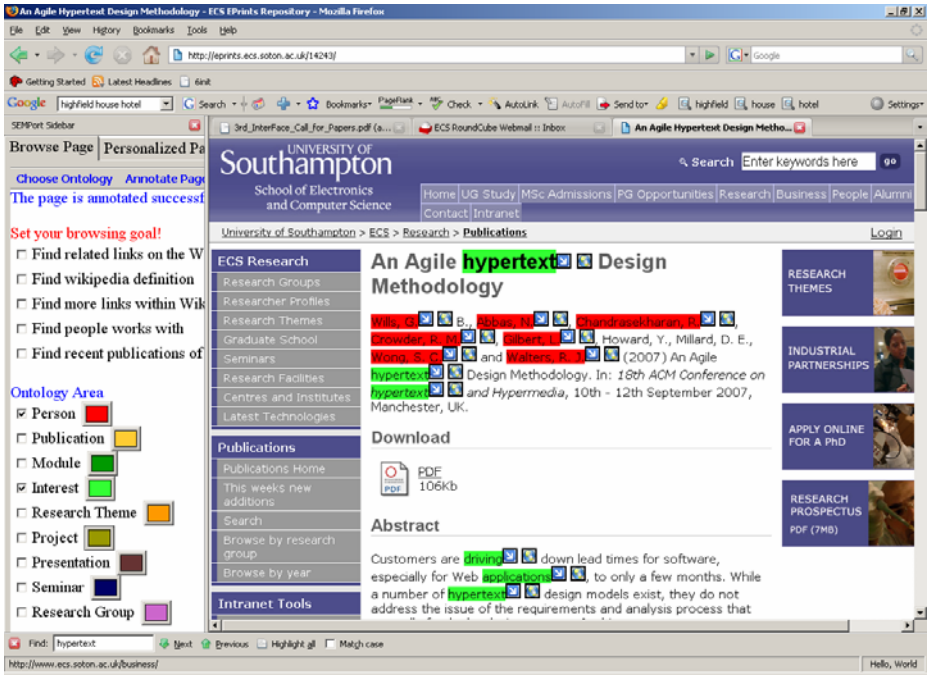


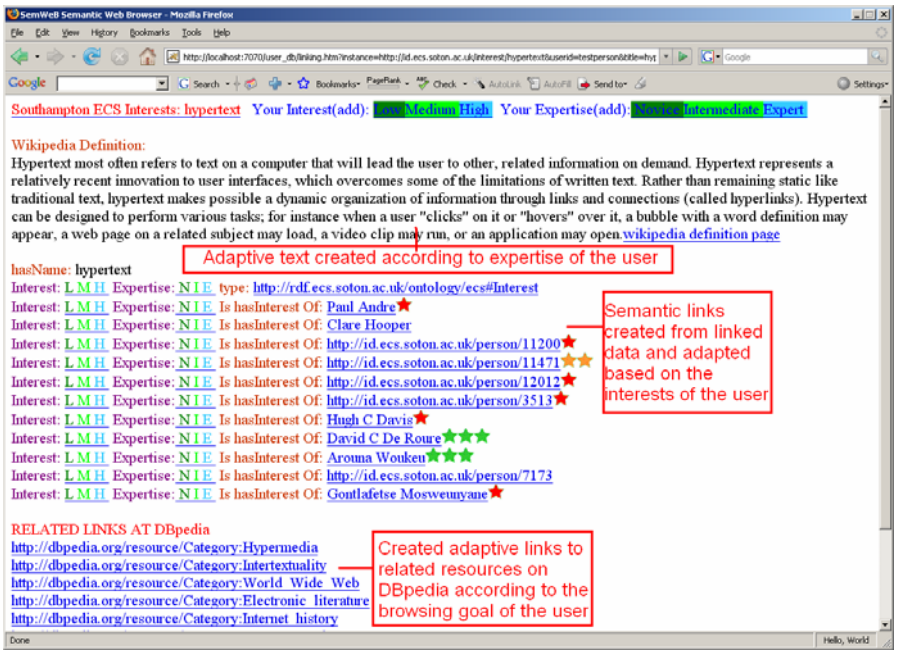
Fig. 1. SemWeB system architecture

User Profiling: Users need to register and login to SemWeB from their browser extension. We request user's name, address, language preferences and security password in the registration. Users can add interests to semantic instances recognized on Web pages and/or provide expertise values to them during Web browsing. For each linked data domain, pre-defined browsing goals are provided in the sidebar, such as shown in figure 2 (a). Users can add those goals into their profiles for adaptive information and links.

Adaptation: Different kinds of adaptation are provided. Based on browsing goals, information from linked data is adaptively retrieved. In figure 2 (b), related links from the DBpedia are searched and presented. Based on browsing interests, the presented semantic links are annotated with visual cues according to a novel semantic relatedness measure. For each presented semantic link, we measure the relatedness to the user profile and present links with different visual cues (see figure 2 (b)). Based on expertise values, we provide different adaptation: for novices and intermediate users on a topic, we show Wikipedia definition (query DBpedia) and for experts more links to Wikipedia pages are presented (DBpedia is queried with detailed SPARQL queries). A personalized homepage is also provided and can be accessed from sidebar.



(a) SemWeB sidebar extension, annotated Web page using the ECS linked data domain



(b) Created semantic links and personalized contents

Fig. 2. SemWeB in action

4 Conclusions and Future Work

Our research aim is to enrich user's Web browsing experience with coherent semantic hyperlinks and personalization. With our approach, we try to alleviate problems of the current Web technology with Semantic Web technologies and adaptive hypermedia. In future work, we will investigate the impact of SemWeB on users. For instance, we will test how useful are the presented semantic links and personalized contents and in which scenarios our approach can provide the full benefits to users in Web browsing. Profile security and encryption to trusted metadata will be also considered in future.

References

1. Berners-Lee, T., J. Hendler, and O. Lassila. The Semantic Web. Scientific American, 284(5), 34-43, (2001).
2. Brusilovsky, P. Methods and techniques of Adaptive Hypermedia. User Modeling and User-Adapted Interaction, 6(2-3), 87-129, (1996).
3. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications, (2002).
4. Şah, M., W. Hall and D. C. De Roure. Designing a Personalized Semantic Web Browser. International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, LNCS, 5149, 333-336, (2008).
5. Şah, M., W. Hall and D. C. De Roure. SemWeB: A Semantic Web Browser for Supporting the Browsing of Users using Semantic and Adaptive Links. International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems Doctoral Consortium, LNCS, 5149, 431-436, (2008).