

# SemWeB: A Semantic Web Browser for Supporting the Browsing of Users using Semantic and Adaptive Links

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.** Web browsing is a complex activity and in general, users are not guided during browsing. The aim of this research is to support the browsing of users using semantic and adaptive hyperlinks using Semantic Web technologies and personalization methods. In this paper, we propose a novel Semantic Web browser (SemWeB), which uses a behavior-based and an ontology-driven user modeling architecture. In our approach, semantic links and adaptive hypermedia can be achieved on different websites. In addition, user profiles can be easily updated with semantic metadata coming from the Semantic Web browser.

**Keywords:** Semantic Web Browser, Semantic Web, User Modeling, Ontology, Personalization

## 1 Introduction

Browsing is an important activity on the Web and in general users are not guided during browsing. Our hypothesis is that browsing can be supported better by using Semantic Web technologies and personalization methods. Semantic Web technologies supply powerful knowledge representation formalisms and inferencing mechanisms on the Web. Browsing can be enriched by using this power. Additionally, different users have different browsing needs and the page content and hyperlinks should be adapted accordingly. Adaptive hypermedia is a solution, where personalization mechanisms adapt information to the needs of the users.

On the other hand, browsing is a complex activity and it is important to understand its nature to better help users. Bawden [1] categorize browsing into three groups: purposive browsing (looking for a definite piece of information), capricious browsing (randomly examining material without a defined goal) and exploratory browsing (deliberately searching for inspiration). Cove and Walsh [2] also divide browsing into three categories: search browsing (searching for defined information), general purpose browsing (looking for items of interest) and serendipity browsing (random). Based on these definitions, we can say that browsing tends to be used in three broad senses: a purposeful activity (directed), searching for inspiration (semi-directed) and capricious behavior (undirected). In our opinion, to better support browsing, user profiles should contain information about the user's browsing behaviors.

## 2 Research Aim and Contributions

The aim of this research is investigating Semantic Web technologies and user modeling approaches to better support the user's browsing using adaptive hypermedia and enriched semantic links. From this point of view, first we attempt to create a personalized semantic portal with adaptive and enriched semantic hyperlinks [3], [4]. Although we achieved personalization and semantic linking on a static database, our main research question was how to achieve personalization and semantic linking on the Web scale. Therefore, this research question let us to think about existing Semantic Web architectures and user modeling approaches.

COHSE [5] and Magpie [6] are two systems that aim to provide useful browsing links using Semantic Web technologies. However, they both use static databases for linking. In addition, they paid little attention to the user's role and they do not supply adaptive links or contents. Existing personalization mechanisms on the Web require users to log in to multiple websites and the user profiles change from site to site. There is a need for generic user profiles and personalization architectures, which can achieve adaptive hypermedia on diverse websites. Our hypothesis is that Semantic Web technologies can offer the solution to these problems. Ontology-based user profiles are interoperable, and they can be easily extended and combined with semantic metadata on the Web. Therefore, user modeling approaches take advantage of semantic metadata; IMS LIP [7] and IEEE PAPI [8] are well known user modeling standards. Although these standards can be applied to any domain, they do not contain data about the user's browsing interests, browsing goals and browsing strategies.

In this research, we attempt to provide a personalized Semantic Web browser, which can be used on different domains for supporting browsing of users using semantic and adaptive links. Our main contribution is the integration of a behavior-based and an ontology-driven user modeling architecture into the Semantic Web browser. We also attempt to provide semantic links using Web as source for linking.

## 3 Results Achieved

This research can be divided into different stages: (1) *analysis*, (2) *system design*, (3) *information extraction*, (4) *creating semantic hyperlinks*, (5) *user modeling*, (6) *creating adaptive hyperlinks*, (7) *SemWeB evaluation*. Part 1 has finished. Parts 2, 3, 4, 5, 6 are ongoing and part 7 of this research will begin in May 2008.

### 3.1 Analysis

This part of research investigates existing Semantic Web enabled systems, semantic annotation platforms and user modeling approaches. Existing Semantic Web enabled systems, such as COHSE and Magpie do not supply adaptive links and contents. In user modeling, the browsing behaviors are not taken into account.

### 3.2 System Design

For understanding interactions of users with Web and enable adaptive hypermedia on different Web sites, we implemented SemWeB as a browser extension of the Mozilla

Firefox Web browser (Fig. 1). SemWeB extends the Web browser with a vertical sidebar. The sidebar has two tabs: the navigation tab and the personalization tab. The navigation tab is used for highlighting ontological concepts found on the page and adding semantic and adaptive hyperlinks (Fig. 2). The personalization tab is used for updating user profiles.

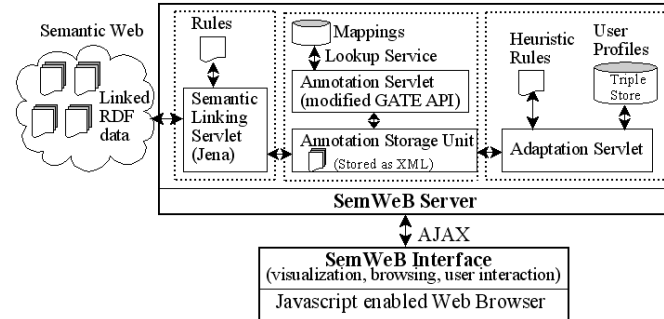


Fig. 1. The architecture of SemWeB

### 3.3 Information Extraction

For information extraction (IE), SemWeB uses ontologies and an ontology-driven lexicon based on a modified GATE [9] framework. GATE is a text engineering architecture for extracting named entities from text. To improve IE, we added new gazetteers and rules. In addition, we extend GATE with a lookup service and annotation storage unit. Lookup service returns the URIs of found concept instances and annotation storage unit stores created annotations as XML files at server-side. When the same page is requested again, the stored annotation is returned. In this way, we prevent undesired delays during semantic annotation. Because IE requires some pre-processing (creating mappings, etc.), SemWeB uses predefined ontologies, particularly ECS ontology [10] and instances. In ECS, every Web page has an RDF file associated to it. RDF files are crawled for the creation of gazetteers and mappings. SemWeB can also be adapted to different ontologies since interface, semantic and adaptive links are created independent of ontologies. GATE can also be adapted to different ontologies. In future, IE will be tested on large scale.

### 3.4 Creating Semantic Hyperlinks

Semantic links are added using the navigation tab. When user highlights a concept, SemWeB embeds an icon next to recognized instance on the page (Fig. 2). To prevent long delays, the semantic link injection is not automatic. The user is required to click the icon next to recognized instances. Once user clicks, URI of the recognized instance is sent to SemWeB server using AJAX. Service firstly dereferences URI over Web using HTTP content negotiation. Then parses RDF file using Jena and identifies more RDF link URIs. If necessary, more URIs are dereferenced. Possible link anchors and targets are identified and an XML response to client's browser is created. Links to open data sources (i.e. DBpedia, DBLP, etc.) are investigated. For this purpose, we are using Sindice Semantic Web search engine [11]. By using Sindice, we identify more related URIs on the Web. Related semantic hyperlinks are obtained by querying

SPARQL endpoints of data sources. For example, links to DBLP recent publications, and Wikipedia broader/narrower topic links are created. Finally, the semantic links are sent back to client as XML, and links are presented in a new Web page (Fig. 2).

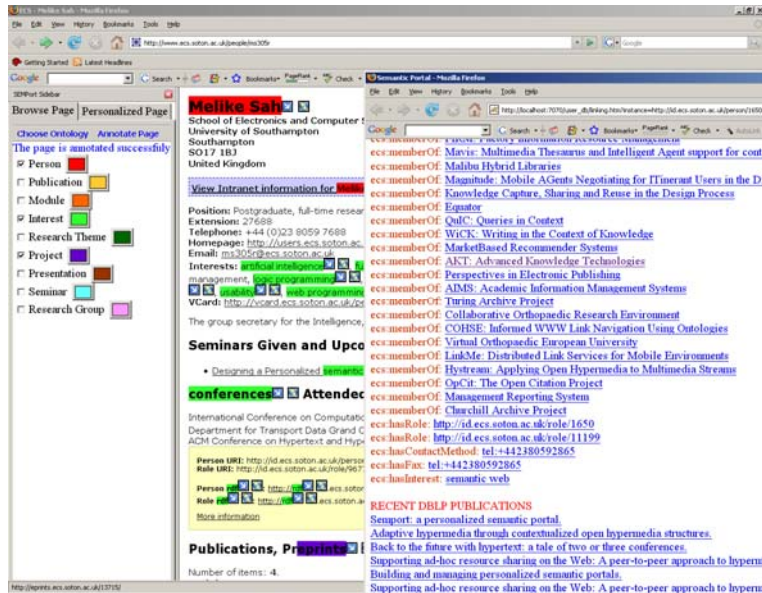


Fig. 2. SemWeb browser extension

### 3.5 User Modeling

For user modeling, we developed a new ontology-based user model, which uses the user's browsing behaviors for adaptation. The model can also be applied to different domains. In our model, currently we use seven categories: identification, preference, security, *browsing goal*, *interest*, *expertise* and *browsing behavior* (our main contributions are in *italic*). The browsing goal concept represents browsing aims of users and it is divided into two sub-concepts: short-term browsing goal and persistent browsing goal. The interest category represents browsing interests of users and it is divided into bookmark (interest to a page) and browsing interest (interest to a semantic instance). Users can also rate their interests (*low*, *medium*, *high*). The expertise category represents knowledge of users for a semantic instance (*novice*, *intermediate*, *expert*). Browsing Behavior category is used to understand the activities of users. The browsing behavior has *browsing\_level* and *browsing\_type* properties. *Browsing\_level* is the number of clicks made by a user in a browsing session. *Browsing\_type* indicates browsing aims of users as suggested by Cove and Walsh [1] and Bawden [2]. When the user has a short-term browsing goal, it is assumed that user is looking for a defined piece of information and *browsing\_type* is set to "*directed*". When the user has a browsing interest or has bookmarked current Web page, then it is assumed that the user is looking for items of interest and *browsing\_type* is set to "*semi-directed*". When the user does not have short-term browsing goals or browsing interests, *browsing\_type* is set to "*undirected*".

Users can log in and register to personalization from the personalization tab. User profiles are kept at a server-side triple store. Users can explicitly update profiles from their browsers using SemWeB interface. In our ongoing work, profiles will be implicitly updated based on interactions with SemWeB (i.e browsing behavior).

### 3.6 Creating Adaptive Links

In our ongoing work, adaptation will be achieved on the recommended semantic links by calculating semantic distance between user model and semantic hyperlinks based on the following conditions. If it is directed browsing, show related links according to short-term browsing goals. If it is semi-directed browsing, use most recently added interests to supply related links. If it is un-directed browsing, make use of semantics. When a link is requested by a novice user, provide Wikipedia links. When the user is an expert, provide detailed semantic links. Also, link sorting and link annotation can be done based on interest ratings, goal priorities, expertise and browsing levels.

### 3.7 SemWeB Evaluation

We are planning to perform a system-based evaluation. Adaptability of SemWeB to different URIs and ontologies will be tested. Adaptability of user profiles to different ontologies will be checked. Interoperability of SemWeB to diverse RDF metadata will be analyzed. Finally, scalability of IE and user profiles will be evaluated.

## 4 Conclusions

This research investigates Semantic Web technologies and user modeling approaches and tries to find new architectures and user models to better support the user's browsing using adaptive hypermedia and semantic links. In this paper, we have presented our ongoing research for personalized Semantic Web browser (SemWeB).

## References

1. Bawden, D.: Information Systems and the Stimulation of Creativity. *J. of Information Science*, vol. 12, pp. 203-216, (1986)
2. Cove, J., Walsh, B.: Online Text Retrieval via Browsing. *Information Processing and Management*, vol. 24, pp. 31-37, (1988)
3. Şah, M., Hall, W.: Building and Managing Personalized Semantic Portals. In *Proceedings of International World Wide Web Conference*, pp. 1227-1228, (2007)
4. Şah, M., Hall, W., Gibbins, N. M., De Roure, D. C.: SEMPort – A Personalized Semantic Portal. In *18th ACM Conference on Hypertext and Hypermedia*, pp. 31-32, (2007)
5. Carr, L., Hall, W., Bechhofer, S., and Goble, C.: Conceptual Linking: Ontology-based Open Hypermedia. *International World Wide Web Conference*, pp. 334-342, (2001)
6. Dzbor, M., Domingue, J., and Motta, E.: Magpie – towards a semantic web browser. *International Semantic Web Conference*, (2003)
7. IMS Learner Information Package Specification, <http://www.imsglobal.org/profiles/lipinfo01.html>
8. IEEE PAPI (Public And Private Information for Learners), <http://edutool.com/papi/>
9. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *ACL*, (2002)
10. ECS Ontology, <http://id.ecs.soton.ac.uk/docs/>
11. A Semantic Web Search Engine, <http://www.sindice.com/>