Adaptive Web based on Semantic Model

Radek Jun, Ivan Jelínek

Abstract: *This paper describes an interest in the usage of a semantic in adaptive web systems. All source data would be in a semantic model which could provide better adaptation. We would like to put together existing technologies to make a functional adaptive web system. Another interesting part is an aspiration look for suitable user interface. The main part covers an overview and analysis of existing technologies in semantic web, adaptive web, and visualization areas.*

Key words: *Semantic web, adaptive web, data mining, visualization.*

INTRODUCTION

We can see the growth of information sources on the Internet every day. Together with this growth, the user base and miscellaneous type of terminal devices also grows. In such a situation, the necessity of sorting information for specific group of users arises and the necessity of filtering output information for various types of devices (PDA, mobile phone, TV) becomes an issue. Filtering could reduce amount of received information, thereby making data searches easy.

Adaptive systems can enhance effective utilisation of huge information sources by reducing the range of information provided to a specific user need. However, the reduction has to be smart enough, neither too little nor too much. A user has to have the chance to change adapted results to non-adapted results and thus control the adaptation process. It should be possible to adapt navigation, content and presentation.

We would like to use semantic model for data source of an adaptive web system. In the HERA methodology [1] it is called *conceptual design*. The semantic model could enable better adaptation process. Results from present research in semantic web based on a semantic show that the related complications outweigh the benefits.

Our goal is to search for the right way to use semantics on the web in order to make implementation simple and provide a well arranged user interface. We would like to use the standardised formats, such as *XML*, *RDF*, *RDFS*, *OWL, which* could help share and reuse information sources. [1].

DOMAIN OF INTEREST

Some type of web sites can be recognised. *Search engines* with a simple user interface and an output list of results. *Category tree* sites where users can browse through the trees to obtain the required class of trees. For example news pages, where users can browse through sport or foreign news. And, lastly, web sites which we can imagine as *web applications*, for example mail client, calendar, forum and others.

We can apply an adaptation to all types of web sites to make them more comfortable for users. But the changes to the web system should converge in persistent state. The adaptation on each type of the web sites could be different, for

example, by changing a sequence of results, by remodelling a class hierarchy tree, up to change of the content range (e.g. abstract versus full article). The use of the semantic can improve the adaptation process.

Figure 1 shows various incoming devices which are sending some request (query) to a web application. This request is received by a controller (an agent processing incoming requests and deciding what to do). Important point is the knowledge of the sending device type (i.e. of the user who sent the request). With the new request we can update the user profile (notice new request to events list). Now the user profile, with information about user's goals and user's preferences is updated.

Requested data (list of results if application is search-like, or some specific page with some texts) are loaded from the database. The data and user profile are posted to an adaptor (part which makes the adaptation). In the adaptor the structure of content (determined by device type) and content (determined by user's goals) are filtered.

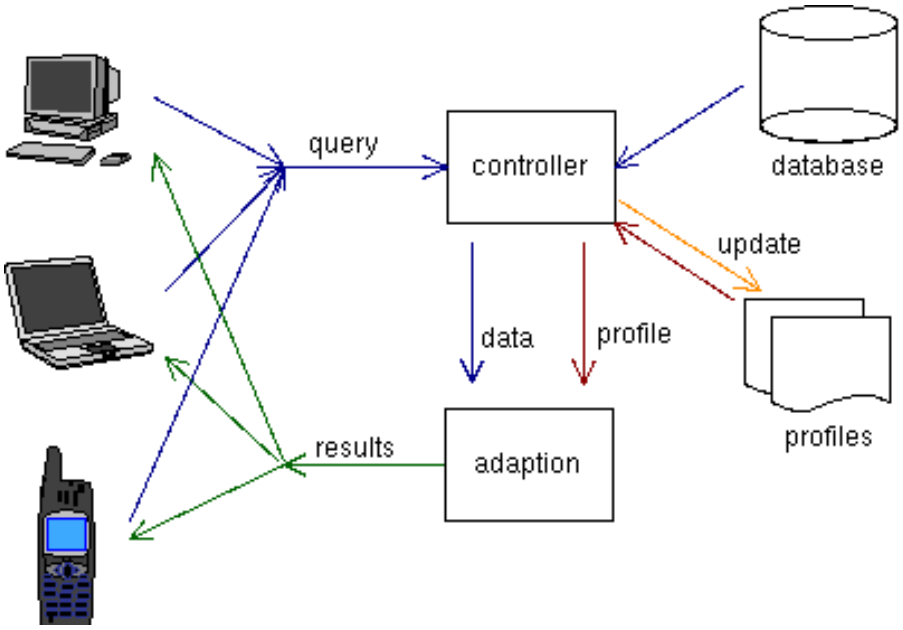

Figure 1: Adaptive system.

Another interesting part is the presentation of the web application. A uniform user interface could help the user to make the work easy. Especially, in search engines it is advantageous to have chance to give to the application the scale of user satisfaction. Graphical representation also allows displaying more information (binding between founded results and neighbourhood).

In the next section we provide an overview of used technologies – adaptive web, semantic web and visualization.

STATE OF ART

In this part is review of existing technologies, which are suitable for our research. First of all, we looked at the adaptive web, after that at the semantic web and the last section is about visualisation.

ADAPTIVE WEB

Adaptive web is a web site, which automatically improves presentation and navigation to user needs. Changes in navigation and presentation are evolved from user's activity history and user's preferences.

User profiles: If the exact user needs are known, we could return more relevant information. By knowing user goals, it is possible to predict user behaviour and accommodate system accordingly [2]. User goals are represented by user histories in log files. From the log files user's behaviour can be observed. Information from log files can be stored in *user profiles* [3].

Improving user's profiles in this way is called *implicit feedback*. The Advantage is that users need not be engaged in this process. Other option is user's preferences, where user can set his own specific preferences. This is not as comfortable for the user as is the *implicit feedback*, but it is very precise.

Adaptability and adaptivity: Adaptability enables viewing HTML documents in special devices such as PDA, mobile phones and etc. To describe, for example, device capabilities it is possible to use CC/PP[1] format [1]. The description may contain, for example, size of the screen, capability of picture display, CPU type and others.

Webs dynamically adapted to user needs, called adaptivity, are pursuant to user's history (user's profiles).

SEMANTIC WEB

Usage of semantic web has many advantages, but also many disadvantages. Advantages include machine understanding the data content resulting in better and more accurate searches. However, the disadvantages prevail, for example, standardised knowledge format, standardised vocabulary (it is important for share and reuse) [4] and readability of human documents (better is graphical representation).

Another problem with using semantics on the web can come into being while sharing semantics between similar web pages (web pages which have similar semantic structure). But several projects try to solve this problem - *OBSREVER*, *KRAFT*, *PICSEL* or *DWQ* [5].

Simple way of meta-data description is an *annotation*. Each part in the sentence can be appended with some annotation, for example, "`John`" can be annotated as "`person`". And also it is possible to make a connection, for example, "`has-`

1 http://www.w3.org/Mobile/CCPP/

`brother(John, Robert)`". If we store the annotated text in a structured database, then we are talking about *knowledge base* [4].

In the following sections we are talking about data harvesting (obtain information from web), Latent Semantic Indexing, Clustering and about External annotation. The last section is devoted to basic semantic format (format for exchange knowledge).

Data harvesting: Typical web sites are composed of (X)HTML documents which are without any meta-data description (without any semantics). For searching of something, user can use some of search engine (i.e. Google™) which return a list of "relevant" pages. User must visit all the pages and check if the content of a each page is relevant, because there is the problem with textual matching without meaning. For example, you can try find something about *LaTeX*, but not all of received results are only about typesetting system. Problem with textual matching causes that the search engine do not know meaning of the word I am looking for.

Adding extra information (meta-data) to the text could help to overlap problem with textual matching. Other possibility is usage of the semantic web which has structured data model with meaning. Upgrading of this structure is too complicated and it would be nice to have some tool which automatically transforms the text document to the text with the semantic.

Latent Semantic Indexing: We want to classify set of documents to some classes. To make this classification we have to use some representation of documents. We can try find frequently used words in the examined document. And if these words are not as frequently in the rest of the set of documents as in the examined document, then these words characterise the document. And this is what the method *Latent Semantic Indexing* [6] do.

Clustering: Clustering of (X)HTML documents is grouping similar documents in a group (cluster) and different to other groups (clusters). Documents which are in one the cluster have common set of characterise words. Also is possible to make hierarchic structure of clusters [4].

External annotation: An external annotation [7] is using of extra files (documents) with describe some content (another document). Benefits are in separating meta-data from a document data. To describe parts in documents are used *XPath2* and *XPointer3*. In annotation can be defined importance of a specific content part (specification can be provided by *XPath*). With this information is possible to filter a content to variously devices (Computer, PDA, Phone).

Basic Data Formats: Resource Description Framework (RDF)[4] is a format for encoding structured meta-data and is feasible for exchange and reuse it. RDF is based on XML[5] which is the popular format for storing and exchanging data. RDF enables describing resources by a *Resource* (anything that can have a URI), a

2 http://www.w3.org/TR/xpath
3 http://www.w3.org/TR/WD-xptr
4 http://www.w3.org/RDF/
5 http://www.w3.org/XML/

Property (resource that has a name) and a *Property value* (some string value). Combination of them make a *Statement* (also known as the subject, predicate and object)[8].

RDF Schema is extension of RDF which provides mechanisms for declaring new properties. RDF Schema is often called as a vocabulary for RDF.

VISUALIZATION

In adaptive systems is range of managed information bigger than in classical systems. Because there are necessity of administering user's profiles and noticing of user's behaviour and wishes. If user can enter relevance of results of some query, this can be stored to user's profile. This process should be called as learning. If we know, what user exactly wants then we can return more relevant data. But this expands a user interface and complicated control.

It is possible to graphically represent in a general graph [9]. In the graph can size of a node represents cardinality of a cluster and edges can represent relationships between clusters. In the graph is possible view whole situation. Also is possible use of 3D [10] representation, but it can be difficulty for manipulation and user's imagination.

The graphic representation can make easy navigation across the web site or simplify selection of results.

CONCLUSIONS AND FUTURE WORK

We would like to use existing technologies, such as XML, RDF, RDF(S), SVG and others, and join them to a functional unit. This unit should provide adaptability services on data which are in a semantic model. An advantages of usage the semantic model and the user profiles are a chance to return more relevant data (quality of the returned data is depend on a user profile quality) and grater possibility to filtered huge amount of information on the Internet.

Collaboration of the semantic webs can bring a big profit as an information wealth. But collaboration of the semantic webs supposed a highly similar semantic across all semantic webs. If it isn't true, then there is needed to make mapping between various semantic and that make one common semantic over all involved semantic webs. The problem of mapping various semantics to common one hasn't satisfying solution. We would like to find an appropriate semantic model which could help to solve this mapping problem.

Another challenge is searching of a suitable user interface which is intuitive and simply to use and still enough good for the adaptive systems. We trust that usage of a graphical interface instead of a textual interface can brings more opportunities in the controlling of a web site.

Results of our research are part of work of a special group WEBING[6].

6 http://webing.felk.cvut.cz/

ACKNOWLEDGEMENT

This research has been supported by MSMT under research program No. 6840770014. This research has been supported by the grant of the Czech Grant Agency No. 201/06/0648.

REFERENCES

[1] F. Frasincar, Geert-Jan Houben, "Hypermedia Presentation Adaptation on the Semantic Web", http://wwwis.win.tue.nl/~houben/respub/ah2002.pdf, 2002.

[2] J. Snopek, I. Jelínek. "PROCEEDINGS of the International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing", "Web Access Predictive Models", V.4., 2005.

[3] C. Danilowicz, H. C. Nguyen, N. T. Nguyen. "Knowledge-Based Information Retrieval And Filtering From the Web", "Model of User Profiles and Personaliyation for Web/Based Information Retrieval System", 121-136, 2003.

[4] B. Berendt, A. Hotho, D. Mladenic, M. van Someren, M. Spiliopoulou, G. Stumme. "A roadmap for web mining: From web to semantic web", "Web Mining: From Web to Semantic Web", 1-22, First European Web Mining Forum, 2003.

[5] H. Stuckenschmidt, F. van Harmelen. "Information Sharing on the Semanti Web", "Semantic Integration", 3-22, Springer, ISBN: 3-540-20594-2, 2005.

[6] S. Fronk, I. Jelínek. "PROCEEDINGS of the International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing", "Semantic Mining of Web Documents", V.2., 2005.

[7] M. Hori. "Spinning te Semantic Web", "Semantic Annotation for Web Content Adaptation", 403-430, Springer, 2003.

[8] K. Wecel. "Knowledge-Based Information Retrieval And Filtering From the Web", "Towards an Ontological Representation of Knowledge on the Web", 1-22, 2003.

[9] C. Fluit, M. Sabou, F. van Harmelen, "Visualizing the Semantic Web", "Ontology-based Informaction Visualization", 36-48, Springer, ISBN: 1-85233-576-9, 2004.

[10] S. Kimanim, T. Catarci, I. Cruz. "Visualizing the Semantic Web", "Web Rendering Systems: Techniques, Classifications, Crieteria and Challenges", 63-89, Springer, ISBN: 1-85233-576-9, 2004.

ABOUT THE AUTHORS

Ing. Radek Jun, Prof. Ivan Jelinek, Department of Computer Science and Engineering, Czech Technical University, Phone: +420 224 357 316,. E-mail: junr1@fel.cvut.cz.