An Architecture for Personal Semantic Web Information Retrieval System

Haibo Yu
Graduate School of Information Science and Electrical Engineering
Kyushu University
6-1 Kasuga-Koen, Kasuga
Fukuoka 816-8580, Japan
yu@al.is.kyushu-u.ac.jp

Tsunenori Mine
Faculty of Information Science and Electrical Engineering
Kyushu University
6-1 Kasuga-Koen, Kasuga
Fukuoka 816-8580, Japan
mine@al.is.kyushu-u.ac.jp

Makoto Amamiya
Faculty of Information Science and Electrical Engineering
Kyushu University
6-1 Kasuga-Koen, Kasuga
Fukuoka 816-8580, Japan
amamiya@al.is.kyushu-u.ac.jp

ABSTRACT

The semantic Web and Web service technologies have provided both new possibilities and challenges to automatic information processing. There are a lot of researches on applying these new technologies into current personal Web information retrieval systems, but no research addresses the semantic issues from the whole life cycle and architecture point of view. Web services provide a new way for accessing Web resources, but until now, they have been managed separately from traditional Web contents resources. In this poster, we propose a conceptual architecture for a personal semantic Web information retrieval system. It incorporates semantic Web, Web services and multi-agent technologies to enable not only precise location of Web resources but also the automatic or semi-automatic integration of hybrid Web contents and Web services.

Categories and Subject Descriptions: H.4.0 [Information Systems]: Information Systems Applications D.2.11 [Software Engineering]: Software Architectures

General Terms: Design, Management

Keywords: Semantic Web, Web Services, Web Portal, Information Retrieval System

1. INTRODUCTION

Since the information overload, Web information retrieval systems are facing new challenges for helping people not only locating relevant information precisely but also accessing and aggregating a variety of information from different resources automatically.

With semantic Web technology, Web information can be given well-defined meaning which can be understood and processed by machines. It provides new possibilities for automatic Web information processing.

Currently, there are a lot of researches such as TAP [1] and Haystack [2], which are trying to apply semantic Web technologies into personal Web information retrieval systems. However, they only addressed the problems concerning certain phases or certain aspects of what they involved, none of them addressed the semantic issues from the whole life cycle of information retrieval and architecture point of view.

The Web services mechanisms provide a good solution for application interoperability between heterogeneous environments. Though it is mainly used for business processes until now, we can predicate that Web services will soon be used by Web portals for information gathering, displaying and delivering, and will play a vital role on Web information retrieval activities. However, the traditional "Web contents" and new "Web services" have been managed separately for their publishing, discovering, accessing, and processing, and there isn't any personal Web information retrieval system managed them all together until now.

2. APPROACH

In this poster, we propose our conceptual architecture for a personal semantic Web information retrieval system based on the following three main ideas.

First, "all participants contribute for the semantic description consistently". As we know that efficient searching for high quality result is based on a pertinent matching between well-defined resources and user queries, where the matching reflects user preferences, all participants (consumer, provider, and mediator) need to contribute to the semantic description consistently. The user needs to describe their requirements and the provider needs to describe their capabilities precisely as well. The mediator needs to ensure that semantics are not lost sight of during the processing and guarantee the semantic interoperability by ontology mapping if necessary. A multi-agent system will function as the mediator in our architecture.

Second, "integrating Web contents with Web services". In the semantic Web, contents are given well-defined meaning, both Web contents and Web services can be consumed by machines, so they have the common ground to be managed together. On the other hand, Web users generally only care about final results, not realization details, hence an integrated or unified management of Web contents and Web services need to be carried out through different levels including the descripiton of capabilities and requirements, querying, discovering, accessing and aggregation.

Third, "providing a gateway to all information that the user is interested in". Users need to access and process a variety of internal and external information based on their

WWW 2005, May 10–14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

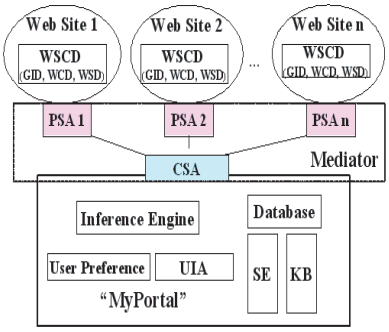

Figure 1: A Conceptual Architecture

preferences, not only the information from the Web but also the information created and managed by themselves. So a gateway to all relevant information is necessary. Web portals are trying to provide such functions as gateways. However, "no one size can fit all", and personalization for the specific user is strongly required. Currently many Web portals have provided personalization such as "my yahoo", "my AOL", but their functions are limited as they lack semantics and separation from the user's local information. Therefore a Web portal which is owned and managed by the specific user, located on the user's local desktop or local network, and constructed based on semantic Web technology is necessary for satisfying the requirements of a user to all his relevant information. That is what we called "Myportal" [4]. It is just like a compact portal with many more personalization functionalities.

3. A CONCEPTUAL ARCHITECTURE

Our conceptual architecture of personal semantic Web information retrieval system can be illustrated as shown in Figure 1.

Each provider describes their capabilities in what we call a WSCD (Web Site Capability Description) and is assigned a PSA (Provider Search Agent). Each consumer describes user's requirements including his preferences. It is assigned a consumer search agent (CSA) and also has a user interface agent (UIA) for providing an intelligent unified interface to the user. The CSA and PSA will function as mediators between the consumer and the provider as they communicate with each other to fulfill the searching and accessing task. The consumer is constructed as a "Myportal" providing a gateway to all relevant information.

3.1 Web site capability description (WSCD)

We describe the capabilities of a Web site by layers. First, we semantically describe the general capabilities of the Web site, and we call this a "general information description (GID)". Second, we give its Web contents capability description (WCD) and Web service capability description (WSD). There are links from GID to WCD and WSD in order to facilitate the further matching and use of Web contents and Web services. In order to semantically describe the capabilities and support the concrete realization of services, we express the service capability description in two layers: "semantic Web service description" and "concrete Web service description". This hierarchical capability-describing mechanism enables semantic capability-describing and matchmaking for different levels. For the detail of our Web site capability description mechanism, one can refer to document[3].

3.2 "Myportal"

"Myportal" is composed of three main functional components: core component, consumer component and provider component. The core component provides basic support for semantic technologies and information management. It consists of "Knowledge Base (KB)", "Query Engine (QE)" and "Inference Engine (IE)". As a consumer, it consists of a user interface, user profile, consumer search agent components and bring together a variety of necessary information from different resources for the user. The semantic Web contents and the outputs of Web services based on the common domain ontology will be integrated, saved and reused in "Myportal". As a provider, the contents and services of "Myportal" can be consumed by humans as well as machines. The human can be the user her/himself or other permitted persons, and the machine can be local or remote. The interfaces for browsing, searching and facilitating Web contents and services need to be provided. For the detail of our "Myportal" description, one can refer to document [4].

3.3 Mediator

The mediator consists of a consumer search agent (CSA) and provider search agents (PSA). The CSA receives the requirements from the user interface agent, breaks and transforms the requirements into formal queries based on user preferences, communicates with relevant agents, selects and invokes Web services, integrates the information and sends the results back to the user interface agent. The PSA receives queries from a CSA and returns matching results to the CSA based on different preferences and requirements. The information in GID and WCD such as a category can be used for Web services discovery and selection.

The semantic interoperability can be realized by using a common domain ontology or through ontology mapping. We think it is reasonable to define a common domain ontology for the communication between the CSA and the PSAs and enable the CSA and PSAs to do ontology mapping for a consumer and providers to the common ontology respectively if necessary.

4. CONCLUSION

In this poster, we addressed the main aspects of a semantic Web information retrieval system architecture trying to answer the requirements of the next-generation semantic Web user. In the future, we would like to further detail the components and interfaces, and implement a prototype to reveal the possibility and effectiveness of our proposed architecture.

5. REFERENCES

[1] R. Guha, R. McCool, and E. Miller. Semantic search. In *Proceedings of WWW2003*, pages 700–709, 2003.

[2] D. Quan and D. R. Karger. How to make a semantic web browser. In *Proceedings of WWW2004*, pages 255–265, 2004.

[3] H. Yu, T. Mine, and M. Amamiya. Towards automatic discovery of web portals -semantic description of web portal capabilities-. In *Semantic Web Services and Web Process Composition: First International Workshop, SWSWPC 2004, July 6, 2004, LNCS 3387/2005*, pages 124–136.

[4] H. Yu, T. Mine, and M. Amamiya. Towards a semantic myportal. In *The 3rd International Semantic Web Conference (ISWC 2004) Poster Abstracts*, pages 95–96, 2004.