# Towards a Semantic "Myportal"

**Haibo Yu, Tsunenori Mine, Makoto Amamiya**
Kyushu University
Fukuoka, Japan
{yu, mine, amamiya}@al.is.kyushu-u.ac.jp

## 1 Introduction

Web portals are information-rich sites trying to provide a single "one stop" for end-users to their necessary Web information. However, they have the following problems. First, it's difficult to locate relevant Web portals and their resources because of the information overload problem. Second, the current Web portals generally target human consumption not machine computing, so it's not easy to aggregate and reuse their resources automatically. Third, current Web portals are centralized resources using fixed organizational schema targeting at uniform access by large numbers of people. However, "no one size can fit all" and personalization functionality is strongly required. Fourth, centralized systems are a bottleneck for accessing traffic, are not robust for failure and their maintenance is also expensive. In this paper, we incorporate Semantic Web technology with Web services technologies to present our approach on constructing a fully personalized Semantic "Myportal". We also propose a mechanism for describing Web portal capabilities to enable precise and automatic location of relevant Web portals and the use of their services.

## 2 A Fully Personalized Semantic "Myportal"

"Myportal" is a "one stop" that links the user to all the information s/he needs. It is located on the user's own desktop or local server and is designed to satisfy a user's personal information requirements and to be mastered freely by the user her/himself. The information can be shared by others with proper authority.

### 2.1 The Characters of "Myportal"

The "Myportal" that we propose here is significantly different from other Web portals in the following respects.

**Personalization:** Finding relevant information depends on what is thought to be relevant. The system must reflect and adapt to user-specific characteristics and interests (preferences) in order to provide information relevant to that particular user. The reflection of user preferences should be an integral part of the whole life cycle of information usage from querying, searching, filtering, storing, managing, retrieving, and presenting to sharing.

**Localization:** As we tend to repeatedly and frequently use a certain amount of information from the Web but seldom or never use other information, it is essential to store frequently used information locally for the user with an efficient retrieval mechanism. This mechanism works like the cache memory in a personal computer. The most used information is cached locally, and external access only happens when the request cannot be satisfied locally. As the information that interests the user is a limited resource and external accessing time is decreased, the total retrieval time will be decreased significantly compared to a search of the vast open Web.

**Automation:** As the internet has changed the scale of available information, nobody can manage this information overload efficiently without help of a machine. In order to enable the machine to cooperate with people well, we must first transform the current Web information into machine understandable well-defined meaningful data. We use Semantic Web technologies to construct and manage the context information, and use semantic Web services to provide application interoperability between heterogeneous environments to realize the automation of information retrieval and management.

**Distribution:** The Web is an open environment, anybody can say anything at anytime from anywhere, so it's most natural to model the Web as a dynamic distributed system. "Myportals" can form the basic units of a distributed information resource. They are equal in position. Each one is a consumer and a provider simultaneously. A distributed system can avoid accessing bottlenecks and is much robust for failure than a centralized system.

**Dual Roles:** "Myportal" can act as a consumer as well as a provider simultaneously. It collects information from other "Myportals" and information-rich Web sites, and shares its own information with the others.

**Dual Supports:** It supports both human and machine consumption requirements. Because the contents of "Myportal" are constructed based on semantic Web technology, machines can understand and process the well-defined, meaningful data. "Myportal" provides not only browsing and searching functions for users, but also services which can be used as a computational program.

### 2.2 The Structure of "Myportal"

As we mentioned above, "Myportal" plays the role of consumer and provider simultaneously. Both functionalities and the core system supporting functionality need to be provided.

**Core Functionality:**
The core component provides basic support for Semantic technologies and information management. It consists of "Ontology", "Knowledge Base", and "Inference Engine" sub-components. The "Ontology" is the backbone of the core component. It includes all ontologies necessary for precisely describing the contents, services, user preferences, user behavior and "Myportal" capabilities. The "Knowledge Base" stores all information of interest to the user such as documents, Web links, search

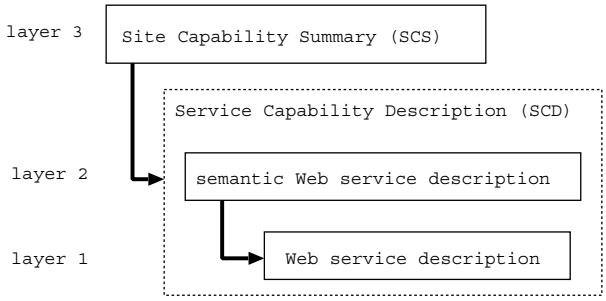


Figure 1: Structure of capability description.

results and information concerning user preferences. It is constructed gradually based on the ontology in use. We use the semi-structured RDF data model to provide easy multi-dimensional information presentation. The "Inference engine" will derive new facts from the ontology and knowledge base based on basic rules.

**Provider Functionality:** As a provider, the contents and services that "Myportal" provides can be consumed by humans as well as machines. The human can be the user her/himself or other permitted persons, the machine can be local or remote. So we need to consider all the possibilities. The interfaces for browsing, searching and facilitating services need to be provided.

**Consumer Functionality:** As a consumer, it will search and aggregate a variety of necessary information automatically or semi-automatically for the user through different technologies. It also needs to aggregate other information-rich sites of user interest with appropriate technologies.

## 3 The Description of Web Portal Capabilities

Because locating "Myportals" and their resources should be based on a matching between user requirements and the portal capabilities, a mechanism for expressing the capabilities of "Myportals" is required. We propose a mechanism for semantically describing the capabilities of "Myportal", enabling automatic discovery of "Myportal" as well as its services.

We describe the capabilities of the Web portal by layers. First, we semantically describe the general capabilities of the Web portal, and we call this a "site capability summary(SCS)". Second, we describe its "service capability description(SCD)". There is a link from SCS to SCD in order to facilitate the use of services. In order to semantically describe the capabilities and support the concrete realization of services, we express the service capability in two layers: "semantic Web service description" and "Web service description". So the structure of capability description can be illustrated as seen in Figure. 1.

This hierarchical capability-describing mechanism enables semantic capability-describing and matchmaking for different levels.

For the detail of our mechanism for describing Web portal capabilities, one can refer to document [Haibo Yu *et al.*, 2004].

## 4 Constructing a Research Community Information Retrieval System with "Myportal"

The location of Web portals, and the discovery and invocation of Web portal services can be realized in various ways. Research community members are generally loosely coupled people, distributed across many locations and organizations. They are the community resource consumers and providers simultaneously. The peer to peer model is very natural for modeling community information concerns. We use the KODAMA [Guoqiang Zhong *et al.*, 2002] multi-agent system to model the research community and construct the community information retrieval system. Each community member is modeled as a peer and there is one or more than one agents to serve each community member. They help the community member to locate the relevant resources and make use of the services that they need.

## 5 Related Work

Haystack [Dennis Quan *et al.*, 2003] investigated various mechanisms for realizing personalization, but has not been constructed from the Web portal point of view, and doesn't emphasize the support of machine interoperability between users in areas such as Web services functionalities. The existing Web portals based on Semantic Web technology [Ruben Lara Hernandez *et al.*, 2004] don't support Web services at present as far as we know. RSS [RSS, 2000] is a lightweight multipurpose extensible metadata description and syndication format can be used for summarizing Web site capability. But the resources of the summarized Web site can not be used as a computational part of the application.

## 6 Conclusion and Future Work

Our goal is to respond to the user requirements for Web information retrieval. The fully personalized "Myportal" powered by semantic Web and Web services technologies enables us to reach this goal. Our mechanism for describing Web portal capabilities not only enables precise and automatic discovery of Web portals, but also enables the application to use the Web portal resources after they are located. Since we use standard ontology language and Web services technology, common existing applications, tools and resources can be used. In the future, we would like to construct "Myportal" to form a multi-agent based P2P information retrieval system and evaluate it.

## References

[Dennis Quan *et al.*, 2003] Dennis Quan, David Huynh, and David R. Karger. Haystack: A Platform for Authoring End User Semantic Web Applications. In Proceedings of 2nd International Semantic Web Conference (ISWC 2003), LNCS 2870, pp.738-753, Sanibel Island, Florida, USA, October 20-23, 2003

[Guoqiang Zhong *et al.*, 2002] Guoqiang Zhong, Satoshi, Amamiya, Kenichi Takahashi, Tsunenori Mine and Makoto Amamiya. The Design and Implementation of KODAMA System. In IEICE Transactions on Information and Systems, E85-D(4):637-646, April 2002.

[Haibo Yu *et al.*, 2004] Haibo Yu, Tsunenori Mine, Makoto Amamiya. Towards Automatic Discovery of Web Portals - Semantic Description of Web Portal Capabilities. In Proceedings of First International Workshop on Semantic Web Services and Web Process Composition (SWSWPC 2004), San Diego, California, USA, July 6, 2004

[RSS, 2000] RDF Site Summary (RSS)1.0, http://web.resource.org/rss/1.0/

[Ruben Lara Hernandez *et al.*, 2004] Ruben Lara Hernandez, YingDing, Sung-Kook Han, Dieter Fensel. Semantic Web Portals - State of the Art Survey. DERI Technical Report 2004-04-03