SEMANTIC WEB POWERED PORTAL INFRASTRUCTURE

YING DING[1]
Digital Enterprise Research Institute
Leopold-Franzens Universität Innsbruck
Austria

DIETER FENSEL
Digital Enterprise Research Institute
National University of Ireland, Galway
Leopold-Franzens Universität Innsbruck, Austria

Introduction

The World Wide Web (WWW, or the Web for short), has made a huge amount of information electronically available, being an impressive success story in terms of both available information and the growth rate of human users. The Web has evolved from an in-house solution for around 1000 users in 1990 to more than 1 billion users and more than 3 billion pages nowadays, not only world-wide but also device-wide. This success has been based mainly on its simplicity, giving software developers, information providers and users easy access to new content. Nevertheless, the same simplicity that made the impressive expansion of the Web possible has brought important, and in some cases critical, drawbacks that are hampering a further development of the Web.

So far, various communities have taken advantage of the current Web functionalities to strengthen communication and information exchange not only within the community but also with external communities or

1 Author contact information: Technikerstraße 13, A-6020 Innsbruck, Austria, ying.ding@deri.org, dieter.fensel@deri.org

individual users. Miscellaneous web portals have appeared with the purpose of providing an open and effective communication forum for their members. In a prototypical case, a portal collects and presents relevant information for the community, and users can publish events or information to the community. Portals provide facilities for users to locate interesting information in the portal according to their personal preferences, topics, etc. In some cases, users with common interests can build their own specific community inside the general community to submit and share information about a given topic.

Nevertheless, current Web technology limits the efficiency for information access, such as searches being imprecise, often yielding matches to many thousands of hits. Moreover, users face the task of reading the documents retrieved in order to extract the information desired. These limitations naturally appear in existing Web portals based on the current technology, making information searching, accessing, extracting, interpreting and processing a difficult and time-consuming task.

The next generation web, Semantic Web, organizes the unstructured data in a structured way by adding metadata and semantics to the data. These semantics can be further processed and reasoned by machines. Semantic Web provides the flexible platform for data communication and collaboration (Davies, et al., 2002; Fensel, 2003. Fensel et al., 2003). Hyperlink structures provides interesting ways to identify the similarity of content. Peer-to-Peer communication echoes the essences of communication styles of virtual groups formed in cyberspace. Distributed component based structures are very important for community portals since the bring about the flexibility to reuse different components. Therefore, Semantic Web technologies can considerably improve the communication and collaboration by overcoming the problems of current web portals. In this sense, portals based on Semantic Web technologies represent the next generation of community portals (see Figure 1):

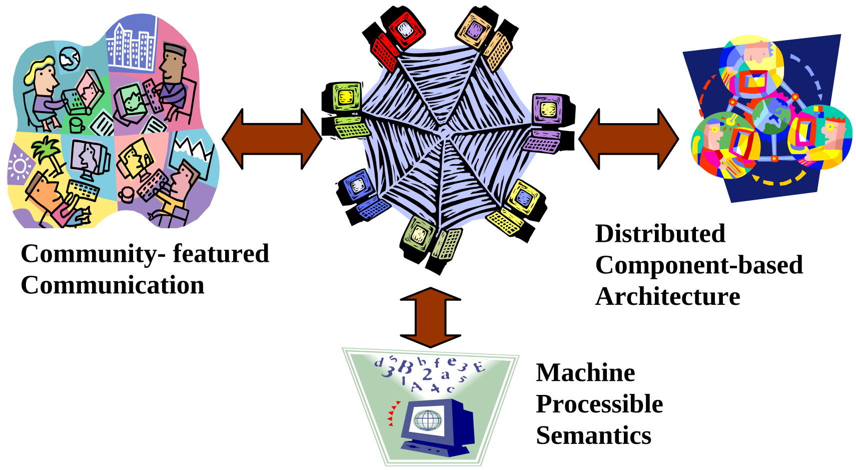

Figure 1: SWPi overview

- Community-featured communication: Community communication and collaboration are strengthened by efficiently using hyperlink structure[2] and peer-to-peer communication.
- Machine processable semantics: ontology and metadata add semantics to information, which can be further processed and understood by machines. Therefore, interoperability among the components of the portal itself and across different portals can be automated to some circumstance.
- Component-based architecture: Distributed components bring the flexibility for reusing, composing and decomposing different functionalities, such as distributed ontology management and evolution, automatic knowledge acquisition and metadata generation, personalized working environment. Furthermore, it also ensures the easy plug-in with other components.

[2] http://wiki.org/, http://www.foaf-project.org/

State of the Arts

Semantic web portal requires smooth integration of several advanced technologies. In this section, we give a brief introduction of these key technologies.

Semantic Web

The explicit representation of the semantics underlying data, programs, pages, and other web resources will enable a knowledge based web that provides a qualitatively new level of service. Automated services will improve their capacity to assist humans in achieving their goals by “understanding” the content on the web, and thus providing more accurate filtering, categorization, and searches of information sources. This process will ultimately lead to an extremely knowledgeable system that features various specialized reasoning services. These services will support us in nearly all aspects of our daily life, making access to information as pervasive, and necessary, as access to electricity is today.

The backbone technology for the semantic web is *ontologies*. Ontologies provide a shared understanding of certain domains that can be communicated between people and application systems. Ontologies are formal structures supporting knowledge sharing and reuse. They can be used to represent explicitly the semantics of structured and semi-structured information enabling sophisticated automatic support for acquiring, maintaining, and accessing information. As this is at the centre of recent problems in knowledge management, enterprise application integration, and e-commerce, increasing interest in ontologies is not surprising (Moeller, et al., 2004).

Natural Language Processing

NLP deals with deriving automatically the semantic meaning of texts. Advanced and robust NL processing technology can extract domain-relevant objects and relations, and develop general techniques for interfacing NLP annotations with Semantic Web annotations and knowledge processing techniques. Whilst the need for NLP processing for automation of Semantic Web annotation is widely acknowledged on the verge of being a truism – no general and far-reaching concept has, as of today, been presented to achieve integration of language and knowledge processing in an encompassing and general approach, or demonstrated proof-of-concept beyond toy systems or experimentation. NLP is currently playing a role in ontology learning and semantic annotation. However, current NLP methods have a closed architecture and cannot discover automatically and make use of other relevant services and resources (Mitchell, 1997).

Information Retrieval

Information retrieval deals with automatic document indexing and retrieving. It is a probabilistic based approach using statistical term frequency or inversed term frequency to identify relevant documents with ranking. User query can be further refined based on the searching result and feedback. Evaluation in information retrieval is rather straight forward by directly involving end users and in a rather qualitative way.

Hyperlink

Hyperlinks of the portal are built on W3C document annotation standards, in particular XPointer and XLink extensions of XML, which provide means for extended, relational, bi-directional and external hyperlinking. Hyperlinks added by the portal users reflect the social network features of communication and collaboration. Information entities (noun phrases, proper names, etc.) and their association can be detected by NLP tools and these entities and their relations can be further hyperlinked into documents. Automatic hyperlinking builds on mature NLP technology, including well-known statistical and rule-based procedures for detection of domain-relevant terms and expressions. Automatic hyperlinking ensures increased consistency, by use of conceptual linguistic databases providing synonyms and conceptual relations. The use of multilingual ontology technology will support linking of concepts independently from the actual linguistic realisation in documents of different source languages. Dynamic insertion of hyperlinks requires advanced contextual disambiguation techniques, to ensure that the automated search for linking targets is associated with appropriate semantic concepts, to ensure access to collections of contextually appropriate linking targets.

Peer-to-Peer

Emerging **peer-to-peer** solutions are particularly well suited to the increasingly decentralized nature of today's communication and collaboration, be it a single enterprise or a dynamic network of organizations. They make it possible for different participants (organizations, individuals, or departments within an organization) to maintain different views of the world while exchanging information. They also circumvent the bottlenecks associated with more traditional solutions, which rely on one or a small number of centralized servers. At the same time, because they rely on keyword search and rather simple knowledge representation techniques, today's peer-to-peer solutions are extremely limited. They cannot easily support the introduction of new concepts, make it difficult to determine whether two terms are equivalent, and generally can only support very limited levels of automation – all types of functionality, which Semantic Web technologies have been shown to support. The power of peer-to-peer computing and the Semantic Web can actually be combined to support decentralized environments where participants can maintain individual views of the world, while sharing knowledge in ways such that administration efforts are low, but knowledge sharing and finding is easy. Key to the success of combining peer-to-peer solutions with Semantic Web technologies is the use of ***Emergent Semantics***. Emergent semantics build on light-weight and/or heavy-weight ontologies that different individuals, departments, or organizations have created. It considers the overlap between ontology definitions and the use of concepts and relations with actual data in order to extract shared ontologies for sets of individuals or groups of people.

Methodology

A portal architecture should be component-based to make it easily extensible and provide flexibility for external plug-ins. It should also consider the community communication and collaboration features, which are the main synergy for the emergent semantics in the community. In order to realize the Semantic Web technology, the architecture should provide strong link between document management system and ontology management system. So in summary, an architecture includes the following components (see Figure 2):

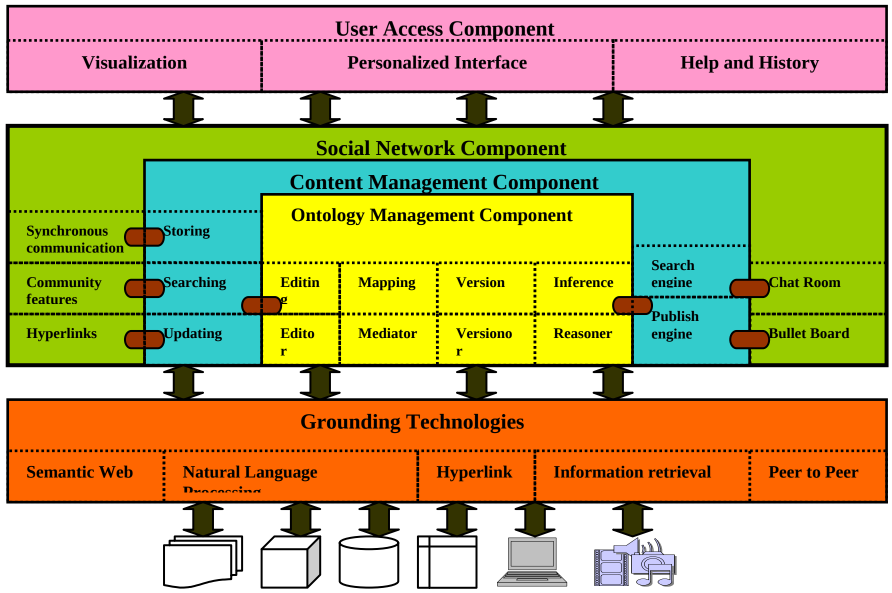

Figure 2: Components of a portal Architecture

<u>User access component</u>

This part deals with the interface between user and the portal. User can have their own personalized interface to access the portal. Information can be visualized for user to browse and search. Help is provided to the user by extensively documented help files or by FAQ. User searching activities can be documented and provided as history to help tracing back or reusing.

<u>Social Network component</u>

This part deals with community features, e.g.members of a community can go to a chat room to chat and discuss issues, they can also publish events and information to the bullet board. It should provide synchronous communication facilities, such as synchronous editing for ontology and content. Hyperlink structures should be efficiently positioned here to link various related piece of information in order to generate emergent semantics.

Content management component

This part deals with document management. It provides storage of information in order to make the search efficient. It provides various searching functionalities, such as keyword searching, ontology-based searching, browsing. Since the portal is a platform where social involvement of community members is crucial, it should provide an easy way for members to publish and update information (Stollberg, 2004).

Ontology management component

This part deals with semantic web technology part. It mainly focuses on the ontology management, which includes ontology editing (with ontology editor), ontology mapping and aligning (with ontology mediator), ontology versioning (with ontology versioner), and ontological reasoning (with ontology reasoner).

Conclusion

This paper has illustrated the architecture of the Semantic Web powered Portal infrastructure mainly for community portals (More information about current results can be found at: http://www.deri.at/research/projects/sw-portal/ and http://www.omwg.org). In the future, we would like to focus on using Semantic Web services (such as WSMO – http://www.wsmo.org) to realize some of the portal functions.

Acknowledgement

We would like to thank members of the **Semantic Web Portal** working group (http://www.deri.at/research/projects/sw-portal/) and **Ontology Management** working group (http://www.omwg.org). We specially thank the support of the Science Foundation Ireland under Grant No. SFI/02/CE1/I131.

Bibliography

DAVIS, J., FENSEL, D., and VAN HAMELEN, F. (eds.) (2002) *Towards the Semantic Web: Ontology-Driven Knowledge Management*, Wiley.

FENSEL, D. (2003) *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag, 2nd edition, Berlin.

FENSEL, D. et. al. (eds.) (2003) *Spinning the Semantic Web: Bringing the World Wide Web to its full potential*, MIT Press.

MITCHELL, T. (1997) *Machine Learning*, McGraw Hill.

MOELLER, K., PREDOIU, L. and BACHLECHNER, D. (2004) *Portal Ontology*. Deliverable 1, Semantic Web Portal Working Group, DERI. http://www.deri.at/research/projects/sw-portal/papers/deliverables/D1_v1.0PortalOntology.pdf

STOLLBERG, M. (2004) *Semanticweb.org – content analysis*. Deliverable 4, Semantic Web Portal Working Group, DERI. http://www.deri.at/research/projects/sw-portal/papers/deliverables/SemanticWebOrg-ContentAnalysis2-1.pdf