Semantic Portals

Brooke Abrahams
Victoria University, Australia

Wei Dai
Victoria University, Australia

INTRODUCTION

Web portals provide an entry point for information presentation and exchange over the Internet for various domains of interest. Current Internet technologies, however, often fail to provide users of Web portals with the type of information or level of service they require. Limitations associated with the Web affect the users of Web portals ability to search, access, extract, interpret, and process information. The Semantic Web (Berners-Lee, Hendler, & Lassila, 2001) enables new approaches to the design of such portals and has the potential of overcoming these limitations by enabling machines to interpret information so that it can be integrated and processed more effectively. The notion of semantic portals is that a collection of resources is indexed using a rich domain ontology (shared and formal description of domain concepts), as opposed to, say, a flat keyword list. Search and navigation of the underlying resources then occur by exploiting the structure of this ontology. This allows searches to be tied to specific facets of the descriptive metadata and to exploit controlled vocabulary terms, leading to much more precise searches (Reynolds, 2001). This article presents the state of the art application of semantic Web technologies in Web portals and the improvements that can be achieved by the use of such technologies. Four main areas are identified: the need for semantic portals, comparison with traditional portals, cross portal integration, and challenges and future trends. A prototype accommodation services portal is also presented toward the end of the article.

NEED FOR SEMANTIC PORTALS

Developers of Web portals are increasingly in need of more powerful technologies capable of collecting, interpreting, and integrating the vast amount of heterogeneous information available on the Web. This heterogeneity stems from the fundamental disparity of Web domains. In the tourism industry, for example, there are numerous portals containing vast amounts of information about accommodation, transportation, entertainment, and insurance. The information has severe limitations, however, because it is largely displayed in HTML, which is designed for humans to read rather than machines to interpret and automatically process. Consequently, current Web technology presents serious limitations to making information accessible to users in an efficient manner. These limitations are summarized in Lausen, Stollberg, Hernandez, Ding, Han, and Fensel (2003), who state that the main problem is that searches are imprecise, often yielding matches to many thousands of hits. Users face the task of reading the documents retrieved in order to extract the information desired. These limitations naturally appear in existing portals based on conventional technology, making information searching, accessing extracting, interpreting, and processing a difficult and time consuming task. What is needed is a system based on global schemas where information can be interpreted and exchanged by machines. The application of semantic Web technologies offers the tools and standardization of Web languages needed to achieve this goal, thus providing the opportunity for improved information accessibility.

The Semantic Web is an initiative by the W3C, in a collaborative effort with a number of scientists and industry partners, with the goal of providing machine readable Web intelligence that would come from hyperlinked vocabularies, enabling Web authors to explicitly define their words and concepts. The idea allows software agents to analyze the Web on our behalf, making smart inferences that go beyond the simple linguistic analysis performed by today's search engines (Alesso & Smith, 2004b, p. 166). The applications that deliver these online solutions are based on new Web markup languages such as Resource Description Framework (RDF) (Manola & Miller, 2004), Ontology Web Language (OWL) (McGuinness & Harmelen, 2004), and ontologies. RDF provides a simple way for descriptions to be made about Web resources using a set of triples based on description logic. RDF is limited to descriptions about individual resources and does not provide any modeling primitives for the development of ontologies. RDFS extends RDF by providing a vocabulary by which we can express classes and their subclass relationships, as well as define properties and associate them with classes. OWL builds on RDFS to provide more vocabulary for defining complex relationships between classes like disjointness, cardinality of properties, and richer semantic capability such as symmetry. As a result of this expressive power, Semantic Web languages are

able to facilitate inference and enhanced searching of Web content. In the tourism industry, for example, it becomes possible through the use of semantics to infer what attractions are associated with a particular resort based on the resort's location. It would also be possible to reclassify the location as a particular location type based on the accommodation, restaurants, and other activities that are in the vicinity. A tourism customer, for example, could then easily search for destinations that meet the domain rules specified for a backpacker classification.

COMPARISON WITH TRADITIONAL PORTALS

There are several advantages to using Semantic Web standards for information portal design compared to the use of traditional portals. The ability to infer knowledge as discussed in the previous section is obviously of major significance. So too is the decentralized nature of Semantic Web technologies, which makes it possible for the portal information to be an aggregation of a large number of small information sources instead of being a single central location to which people submit information. This reduces the complexity of managing and updating information sources. Reynolds (2001) explains that in this situation, central organization is still needed in the initial stages to provide the start-up impetus and ensure that appropriate ontologies and controlled vocabularies are adopted; however, once the system reaches a critical mass, information providers can take responsibility for publishing their own information provided it is annotated consistently with a relevant domain ontology. An example of this decentralized approach is the ARKive portal[1], which publishes multimedia objects depicting endangered species. ARKive just provides the backbone structure of resources by making their ontology available for use. Individual communities of interest then supply the additional classification and annotations to suit their needs. These types of portals can be reorganized to suit different user needs, while the domain indexes remain stable and reusable. Communities of interest can share access to the same underlying information using a completely different navigation structure, search facility, and presentation format. Semantic Web technologies also make it easier to aggregate information from separate portals into a single integrated portal by applying mapping and merging techniques to shared or compatible ontologies. Techniques for cross portal integration are discussed in detail in a later section. Table 1 summarizes the advantages of using semantic portals compared to traditional portal design.

CROSS PORTAL INTEGRATION

It is not realistic to assume that all information in a particular domain of interest will one day be annotated according to a single ontology. The reality is that there are many ways in which a domain can be modeled and individual organizations will for the most part choose to structure their information in a way that best suits their needs. Ontology merging and alignment techniques make it possible to integrate data across multiple portals, thus facilitating queries over federated data sources. Ontology merging can be defined as the process of generating a unique ontology from the original sources (Noy & Musen, 2002). Ontology mapping means establishing different kinds of mappings (or links) between two ontologies. This article will focus on ontology merging techniques.

Table 1. Comparison of traditional and semantic portals (Reynolds, 2001)

Traditional Design Approach	**Semantic Portals**
Search by free text and stable classification hierarchy.	Multidimensional search by means of rich domain ontology.
Information organized by structured records; encourages top-down design and centralized maintenance.	Information semistructured and extensible allows for bottom-up evolution and decentralized updates.
Community can add information and annotations within the defined portal structure.	Communities can add new classification and organizational schemas and extend the information structure.
Portal content is stored and managed centrally.	Portal content is stored and managed by a decentralized Web of supplying organizations and individuals. Multiple aggregations and views of the same data are possible.
Providers supply data to each portal separately through portal-specific forms. Each copy has to be maintained separately.	Providers publish data in reusable form that can be incorporated in multiple portals but updates remain under their control.
Portal aimed purely at human access. Separate mechanisms are needed when content is to be shared with a partner organization.	Information structure is directly machine accessible to facilitate cross-portal integration.