Semantic Web Implications for Web Portals

Pankaj Kamthan
Concordia University, Canada

INTRODUCTION

A Web portal is a gateway to the information and services on the Web, where its users can interchange and share information. In their brief lifetime, Web portals have benefited various sectors of the society and found widespread use (Jafari & Sheehan, 2003; Tatnall, 2005). By careful aggregation of information, Web portals simplify access, as well as decrease the time and effort of locating resources on topical themes. In doing so, they have created a sense of community with common interests.

It is crucial that a Web portal be able to capture, represent, and syndicate information adequately. To that regard, the Web portals today face the challenges of increasing amounts of information, diversity of users and user contexts, and ever-increasing variations in proliferating computing platforms. They need to continue being a successful business model for providers and continue to be useful to their user community in the light of these challenges.

This article discusses the potential of Semantic Web technologies in tackling the issues of agility, sustainability, and maintainability of the information architecture of domain-specific Web portals. The organization of the article is as follows. We first outline the background necessary for the discussion that follows and state our position. This is followed by a detailed treatment of social prospects and technical concerns pertaining to knowledge representation of integrating Semantic Web technologies in Web portals. Next, challenges and directions for future research are outlined and, finally, concluding remarks are given.

BACKGROUND

That the users are able to access relevant information in an efficient and precise manner is critical to the success of any Web portal. A special-purpose Web portal facilitates access to Web sites that are closely related: it addresses a specific *domain* of application, such as information on wine or on travel. To enable automated processing and reasoning by agents, this domain knowledge needs to be accurately *represented*. However, the technologies that are commonly used today for expressing information in a typical Web portal are insufficient.

It is common for Web portals to express information in the HyperText Markup Language (HTML) where, by static or dynamic means of generation, they can reach a broad demographic. Users find information on a Web portal with the help of navigation or via searching. Navigation is implemented via the hyperlinking mechanism, while searching is realized through a form-script-based scheme. However, the focus is mainly on the presentation, rather than on representation of information. Finding relevant documents by manually traversing the links has limited scalability, as the number of resources increase, including annotations in document headers provides a limited solution for searching, and searching is limited to keyword match.

The Semantic Web has recently emerged as an extension of the current Web that adds technological infrastructure for better knowledge representation, interpretation, and reasoning (Hendler, Lassila, & Berners-Lee, 2001). We formally define a *semantic portal* to be a product that results from the fusion of technologies inherent in the Semantic Web architecture into Web portals.

Semantic portals are beginning to appear in both educational (Hartmann & Sure, 2004) and commercial contexts (Lausen, Ding, Stollberg, Fensel, Hernández, & Han, 2005). An evaluation of Esperonto, OntoWeb, Empolis K42, and Mondeca ITM Semantic Portals has been given (Lausen, et al, 2005). At the core of these semantic portals is knowledge representation, the prospects and concerns of which we discuss next.

KNOWLEDGE REPRESENTATION IN A SEMANTIC PORTAL

Our discussion of semantic portals is based on the knowledge representation framework given in Table 1.

The first column addresses semiotic levels. Semiotics (Stamper, 1992) is concerned with the use of symbols to convey knowledge. From a semiotics perspective, a representation can be viewed on six interrelated levels: physical, empirical, syntactic, semantic, pragmatic, and social, each depending on the previous one in that order. The physical level is concerned with the representation of signs in hardware, and is not directly relevant here.

The second column corresponds to the Semantic Web "tower" that consists of a stack of technologies (Daconta,

Table 1. Knowledge representation tiers in a semantic portal

<table>
<tr><th>Semiotic Level</th><th>Semantic Web Technological Layer</th><th>Decision Support</th></tr>
<tr><td>Social</td><td>Trust</td><td rowspan="6">Feasibility</td></tr>
<tr><td>Pragmatic</td><td>Inferences</td></tr>
<tr><td>Semantic</td><td>Metadata, Ontology, Rules</td></tr>
<tr><td>Syntactic</td><td>Markup</td></tr>
<tr><td>Empirical</td><td>Characters, Addressing, Transport</td></tr>
<tr><td>Physical</td><td>Not Directly Applicable</td></tr>
</table>

Leo, Obrst, & Smith, 2003) that could be viewed as varying across the technical to social spectrum as we move from bottom to top, respectively. The definition of each layer in this technology stack depends upon the layers beneath it.

Lastly, in the third column, we acknowledge that there are time, effort, and budgetary constraints on producing a representation. We therefore include feasibility, a part of decision theory, as an all-encompassing factor on the layers to make the representation framework practical. There are various techniques for carrying out feasibility analysis, and further discussion of this aspect is beyond the scope of this article.

The architecture of a semantic portal is an extension of the architecture of a traditional Web portal on the server-side in the following manner: (a) by expressing information in a manner that focuses on *description* rather than presentation or processing of information; and (b) by associating with it a knowledge management system (KMS) consisting of one or more domain-specific ontologies and a reasoner that communicates with them and with the servers used by the portal if and when necessary.

We now turn our attention to the each of the levels in our framework for knowledge representation in semantic portals.

EMPIRICAL LEVEL OF THE SEMANTIC PORTAL

This layer is responsible for the communication properties of signs.

Among the given choices, the Unicode Standard provides a suitable basis for the signs themselves, and is character-by-character equivalent to the ISO/IEC 10646 Standard Universal Character Set (UCS). Unicode is based on a large set of characters that are needed for supporting internationalization and special symbols, which are necessary for universality of Web portals. For example, the Madiera Data Portal (Assini, 2005) provides a customizable multilingual user interface to a wide array of statistical datasets published by some of the major European social sciences data archives.

The characters must be uniquely identifiable and locatable, and thus addressable. The uniform resource identifier (URI), or its successor international resource identifier (IRI), serves that purpose.

Finally, we need a transport protocol, such as hypertext transfer protocol (HTTP) or the simple object access protocol (SOAP) to transmit data across networks.

SYNTACTIC LEVEL OF THE SEMANTIC PORTAL

This layer is responsible for the formal or structural relations between signs.

The Extensible Markup Language (XML) lends a suitable syntactical basis for expressing information that allows focusing on the content rather than processing or presentation. There are a number of ancillary technologies that strengthen XML and have matured over the years. The XML document type definition (DTD) and its successor, XML schema, provide means for expressing structural and data type constraints on the syntax and content of the elements and attributes in XML documents. Namespaces in XML is a mechanism for uniquely identifying XML elements and attributes of a markup language, thus making it possible to create *heterogeneous* documents that unambiguously mix elements and attributes from multiple different XML documents. The Extensible Stylesheet Language (XSL) is a stylesheet language for associating presentation semantics with arbitrary XML documents, while its companion XSL Transformations (XSLT) is a stylesheet language for transforming XML documents into other, including non-XML, documents. Support for querying XML documents is provided by XQuery and client- or server-side tree-based processing of XML documents is enabled by the document object model (DOM).

Related Content