Capitalising on richer Web data

January 4 2005

A quiet revolution is coming our way. Recent successful trials of European semantic-Web applications suggest that machine-readable data will soon usher in an **improved Web** that will facilitate information reuse, and provide for painless building and maintenance of community portals.
"Computers struggle to attach meaning to information written in common Web languages such as Hypertext Markup Language [HTML]," says Libby Miller, coordinator of the IST SWAD-Europe project. "One way to help computers is to add tags, in a language that looks a bit like HTML but is actually new. The result of these information tags is rich data, allowing people and computers to work better together."

Miller's project was led by the World Wide Web Consortium (W3C), and involved the University of Bristol, HP Labs, Rutherford-Appleton Laboratories and Stilo Ltd. The project's goal was to ensure that Semantic Web technologies become widely accepted in networked computing.

The Semantic Web explained

But what exactly is the Semantic Web? W3C describes it as a "common framework that allows data to be shared and reused across application, enterprise and community boundaries." According to Miller, "It's based on a family of standards developed by the W3C. The main one is the Resource Description Framework (RDF), which is a way of describing the world in a way that computers can do useful things with the information."

HTML, she underlines, is fine as a vocabulary for presenting a Web page or describing people, calendars and documents for humans. RDF, whose main syntax is XML (eXtensible Markup Language), goes a step further by describing these sorts of things for computers. “The tags provided by RDF are like glorified links, providing more information than HTML,” says Miller. “They enable computers to harvest, process and display data. Our project aimed to help Web developers to create and use this data.”

When the project began, says Miller: “Developers needed the tools to create, store and view Semantic Web data, but relatively few were available at that time. Today, the Semantic Web is a large community, with thousands of developers working on and using these tools. Along with other projects, SWAD-Europe has helped to build and maintain this community”

Environmental semantics

The project partners successfully demonstrated the Semantic Web’s capabilities through two demonstrators, created by HP Labs. The first, the Semantic Web Environmental Directory (SWED), is a prototype of a new kind of directory of environmental organisations and projects. “Easy to use and maintain, the directory keeps data creation close to the people who create the data,” says Miller. Rather than centralising the storage, management and ownership of the information, in SWED the organisations and projects themselves hold, and maintain their own information. The information is published on their websites.

“The site offers the software tools for people to write about their organisations in RDF format, information that SWED simply harvests to create the directory,” she adds. “The data itself is always kept separate from the portal. The idea is that if the portal is no longer funded the data is still there and a new portal can be created easily.”

The SWED browser, produced under SWAD-Europe, allows anyone to build their own directory pages. “The SWED browser lets them see available options or facets,” says Miller. “Each box in the form contains a facet, such as organization type or project area. This feature expands searching possibilities, while facilitating navigation through complex information spaces with rich data.” The SWED browser is one of a new breed of faceted browsers, which call on hierarchical faceted metadata to refine and expand search queries.

Blogging for librarians

The project’s second demonstrator involves semantic blogging in the bibliography management area. It was chosen because people working in this field need to share small items of information with a peer group in an easy and timely fashion. “The portal includes peoples’ Web diaries [blogs] with three semantic behaviours – view, navigation and query – built over the base,” says Miller. “If you are in a research reading group, you can call on a number of scripts to generate forms which tell others what the research paper discusses, and so on.” The enriched information generated by blogging and Web technologies makes it easier for computers to retrieve archived data.

SWAD-Europe received positive feedback at dissemination workshops held in eight European countries. “We targeted our work at Web developers, focusing on open-source tools to generate a critical mass,” says Miller. These events, she notes, helped create online communities of interested developers as well as Semantic Web advocates.

Commercial outlets

As to the future, the UK’s Natural History Museum is considering using SWED’s data exchange/aggregation capabilities to describe its own

collections. In addition, outside interest shown in the semantic blogging portal indicates a genuine market need in this area. The project partners have engaged in discussions with various organisations involved in bibliographic metadata standards.

“Though the project is over, some partners are continuing to work on Semantic Web best practices within the W3C’s working group in that area,” says Miller, who highlights the standardisation of ways of describing thesauri as an example.

Source: IST Results