Marketplace

WWW10 in Asia

With broadband access available to 95 percent of households and all businesses, Hong Kong seemed an appropriate choice for the first Asian host to the International World Wide Web conference series. WWW 10 started off with a roar as two lion dancers welcomed the 1,200 guests to the rhythm of drums and traditional instruments off stage. Over the days that followed, attendees had access to more than 300 presentations and papers.

The World Wide Web Consortium (W3C) has made much progress toward fleshing out the "semantic Web" since last May's Amsterdam meeting. In Hong Kong, working group members got to show off phase-next specs and tools and continue the ongoing dialogue.

The Road Ahead

To kick off his opening keynote, W3C director Tim Berners-Lee officially released the XML Schema recommendation. With that done, he turned to the crowd and asked, "Are we done yet?" as he followed a hyperlink in his presentation to a component diagram showing the W3C technology road map for reaching the full-fledged semantic Web (see opposite page). Many ellipses represented current, prototyped, or finished work, but just as many represented missing pieces. Clearly, the work is not yet through. (Try viewing the "road map" scalable vector graphic using W3C's Amaya browser, http://www.w3.org/Amaya/ or one of several plug-in viewers.)

The foundations of XML are in place now that namespaces and schemas are ready, but XML remains incomplete. Query functions are still needed, says Berners-Lee, and things like XPointer, XInclude, and XLink are still under development to fill in missing elements. XML also needs mechanisms for privacy, security, quality of service, and handling binary attachments. Moreover, he said, the current spec is complex, and elements in it could be interpreted in conflicting ways. He thus suggested revisiting the architecture and processing model to remove clutter so that XML can be implemented on PDAs, phones, and so on.

A lot was accomplished over the past year: XHTML is stable; SVG is making inroads in usage; and cascading style sheets (CSS) and extensible stylesheet language transformations (XSLT) are in place for presentation. Some convergence questions remain, however, regarding things like methodologies for mixing languages such as XHTML and MathML, or CSS and XSLT.

Infrastructure questions are also gaining importance as "always on" connectivity increases because of broadband and wireless access. What's more, W3C's concern with universal access issues pushes up against the so-called digital divide. Berners-Lee suggested that if the Internet were created again without the installed telecommunications infrastructure found in the West, the network might be a very different thing — that perhaps choices along the road thus far were not always optimal.

W3C's ability to solve these issues is limited by resources (their annual operating budget is about US$7 million), but some very real questions remain for the technical community as a whole. W3C specs should address such concerns, he said, but they can't solve the digital divide. Those issues are also discussed within the IETF and Internet Society (ISOC).

Berners-Lee closed his talk by congratulating W3C members for great progress toward establishing a solid foundation for the semantic Web — "the stuff of creating harmony and integration." But he finished with a warning about the dangers of "frivolous patents" that threaten open standards.

To help limit patent-related troubles, W3C asks its corporate members to waive royalties before they are allowed into working groups. By supporting open standards, Berners-Lee counseled, everyone will benefit far more from the expanded user base than from proprietary claims on technologies that are tied to specifications.

Semantic Overview

Many working group members were clearly excited about all the ongoing work surrounding the semantic Web activities, which were frequently described as being in the "fun stage" of development. As W3C metadata activity leader Ralph Swick said, "We want to actually build some stuff, like in the early days of the Web."

Not surprisingly, most sessions at WWW10 focused on technologies like RDF and the aspects of the XML puzzle that fill in the pieces of Berners-Lee's road map. W3C initiatives can be broadly divided into three areas:

- information management (sharing information in small units, including descriptions for privacy and so on),
- process and workflow (semantics and issues like who can do what within a process), and
- trust and proof (ensuring data integrity with digital signatures and other technologies).

See the sidebar, "Semantic Resources," for pointers to some of the tools and utilities for incorporating W3C specifications into development processes.

Device Independence

Web technologies are diverging in non-PC arenas with things like digital TV, the wireless application protocol (WAP), and NTT DoCoMo's iMode service, but W3C is trying to coordinate and cooperate with external standards organizations to minimize fragmentation of the Web.
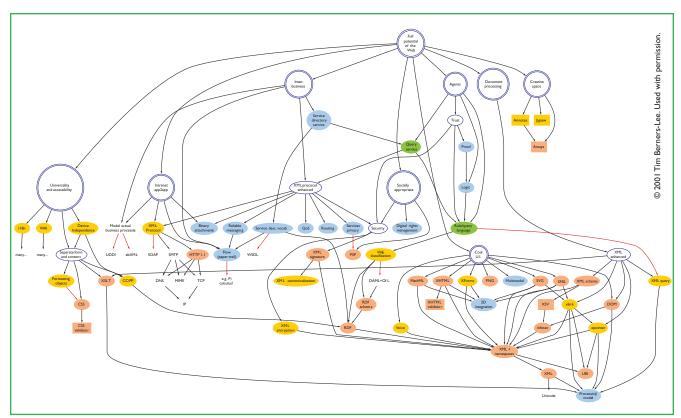
The W3C device independence (DI) activity wants to develop general frameworks for network access regardless of the devices in use. The activity is only a few months old now, but it includes a number of development efforts including composite capability/preference profiles (CC/PP) and voice browsers.
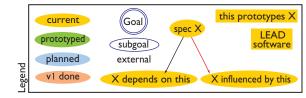
Profiles and Preferences

CC/PP will provide information to help adapt content to different devices and user needs. A CC/PP profile uses the RDF model and syntax to define device capabilities. RDF triples describe components and properties for hardware, software, and user preferences, and can describe proxy behavior as well.

The CC/PP WG created a working draft in March 2001 to define the structure and recommend vocabulary, but the working group is leaving full vocabulary definitions to groups like the WAP Forum, which adopted CC/PP in its user profiles in 1999, and the 3GPP, which incorporated the framework in 2000. Hidetaka Ohto, chair of the DI WG, said the current proposals aim to define a framework rather than a protocol, and the final details might actually be worked out by the IETF if it comes to a protocol-level answer. A birds-of-a-feather meeting at the 48th IETF meeting included mention of CC/PP.

© 2001 Tim Berners-Lee. Used with permission.

W3C technology road map. *In the end, all W3C activities are in service to the top-level goal of reaching the semantic Web's full potential. Arrows indicate "how" things are implemented; following them in reverse indicates "why" they exist (or should). View the full-size figure at http://w3.org/2001/04/roadmap/all.svg.*

Improving Web Search

The information retrieval community has been wrestling with search issues for the past 25 years or so, but the Web adds a new aspect over the relatively controlled corpus of information in a typical enterprise setting. On the Web, data is primarily unstructured, remote, hyperlinked, and often dynamic.

New Challenges

The "Beyond Keyword Search" panel discussion at the 10th World Wide Web Conference started with such existential questions as, "What is the value of Web search?" and turned to some technical fundamentals like, "Is search a computational problem?" In the end, the panel worked through to some research topics that can improve the search experience for frustrated users.

"Link analysis is a good start for bootstrapping semantics, but it is not enough," said moderator Prabhakar Raghavan (see the interview on enterprise portals with him in this issue of *IC Online* at http://computer.org/internet/v5n4/). Panelists agreed that semantic representations should be made in both natural language and machine-usable computational descriptions. That means understanding user needs as well as the relationships between documents, which reside in a "free-for-all" space on the Web.

New Techniques

Leveraging metadata is vital to the effort to mine the diffuse underlying structure because the keyword interface is too weak to leverage the inherent relational structure that exists on the Web.

AltaVista's Andrei Broder suggested that next-generation search techniques will use semantic context analysis and cross-information integration to help determine "the need behind the query."

John Lowe of AskJeeves believes advances in search will come from the query side. "Dialog management will bring dramatic improvements using linguistics and voice-recognition techniques," but task and content-dependent evaluations must first be standardized to enable conversational searches. ∎

Speech interfaces

The voice browser WG (http://w3.org/voice/) was founded in May 1999 and is due for rechartering later this year with a clarified IP policy. Current efforts focus on designing a speech interface framework that could be used by people with disabilities as well as by motorists and others who need hands-free network access. The working group has published requirements drafts for LexiconML, which aims at pronunciation extensibility, and Natural LanguageML, which aims to go beyond keywords to a richer natural language comprehension that can include sophisticated linguistic analysis techniques.

Dave Raggett, technical lead for W3C's work on voice browsers, says speech recognition technology now boasts greater than 90 percent accuracy, especially with constrained entry (as with lists of choices). Voice interfaces are currently used mainly for personal assistants, voice portals, and front-end call centers, but future usage will be multimodal — matching voice and graphical display — which first requires a dialogue-based interaction model.

VoiceXML is based on forms, but it includes tags for events and actions that can be used for voice interfaces to applications. The current view of the architecture includes a VoiceXML gateway between the PSTN/VoIP network and the Web site, which will also have a front end that feeds audio files and parses grammars.

W3C Town Meeting

On the final day of the regular conference, a panel of W3C staffers answered questions about W3C policies and activities at an open-microphone session. The discussion turned at one point to how W3C prioritizes its activities, and the panel admitted that there aren't really any set rules.

The activity process starts when the W3C team or members propose a topic. The member organizations vote on whether to commit resources to the topic, and the W3C director makes the final decision. Some in the audience suggested adding online voting forms.

In general, the ad hoc per-project approach no longer seems sufficient for managing the array of activities the W3C has going. HTML WG chair Steven Pemberton said he is preparing to propose a charter for a "horizontal" working group, like the Web accessibility initiative (WAI), to protect users' interests and help with interactivity coordination. He said NIST, IBM, and Sun have already expressed support for the idea.

When asked about conformance testing for XHTML and other specifications to ensure interoperable implementations and reduce fragmentation, Pemberton added that the consortium is about to launch a "Conformance and Quality Assurance Activity" (http://w3.org/QA/). The activity will focus on ensuring that W3C recommendations are correctly implemented. It should thus help address concerns on adding and improving test suites, which Pemberton said "continually haunt us," not least because even the WGs have trouble reaching consensus on implementation details.

There are already test suites for SVG, the synchronized multimedia integration language (SMIL), MathML, CSS, and the DOM activity is building them now for all three levels — although one audience member noted that different tests for DOM level 1 interpret things differently. The SVG conformance suite implementation, which actually came from outside the WG, was released at the same time as the original spec, and a new version is due out soon.

Conclusion

As modern living becomes ever more reliant on Internet-based applications, the specifications that rule the Web are increasingly important to the international community. W3C is reaching out to the developer community by publishing implementation reports and adding more public areas to its Web site.

For information on next May's conference in Honolulu, Hawaii, see http://www.www11.org/.

— Steve Woods

XML Resources Update

IC Online's XML resources page (http://computer.org/internet/xml/) has been updated to include the newest products, standards, and development utilities in the ever-expanding XML space – from graphical user interfaces (GUIs) and wireless technologies to Web services.

Web services are a new breed of self-describing, self-contained modular applications engineered to be platform-neutral, operate within heterogeneous environments, and interconnect to create products and services on demand. The applications are usually written in a platform-specific programming language and then networked so that they launch and coordinate with each other in real time.

Web services won't really take off until they are attached to an intuitive user interface and can be accessed across a wireless network. It's a bold vision and one that still has a ways to go before being realized, but the basic concepts are sound and the recent shift toward service-based architectures and lightweight, cross-platform clients has moved the Internet a few steps closer to ubiquity.

— *Lisa Rein*

Apache Batik 1.0

The Apache XML group has released the open-source Batik 1.0 (http://xml.apache.org/batik/) Java-based toolkit for scalable vector graphics (SVG). The toolkit comprises a set of modules for generating, manipulating, transcoding, and searching SVG images. For example, Java-based applications can export graphics as SVG using Batik's SVG generator.

The Apache team plans to update the toolkit to reflect modifications in SVG, which is currently a W3C Proposed Recommendation. Batik 1.0 supports most of the static features in the SVG basic effectivity test suite; future releases should support dynamic SVG behaviors, such as scripting and animation with the synchronized multimedia integration language (SMIL).

PDA Encryption

Certicom's new movianCrypt software uses a password-based user log-in system and 128-bit Advanced Encryption Standard (AES) to encrypt data stored on Palm PDAs running Palm OS 3.0 or higher and Handspring Visors running Palm OS 3.1 or higher. The company says applications run unmodified, but data is encrypted on the fly as it is stored and decrypted as it is accessed.

Certicom also shipped version 1.1 of its IPSec-based movianVPN client for handheld devices. The software supports two-factor authentication mechanisms – such as the RSA SecurID token cards used in gateways from Alcatel, Cisco Systems, and Nortel Networks – for devices using both Palm OS 3.5 and WinCE 3.0.

Further information is available from http://www.moviansecurity.com/.

Getting Smart

In response to the disproportionate number of fraudulent Internet-based versus offline sales, MasterCard has started several initiatives aimed at end-to-end protection. The company has introduced new merchant rules and authorization schemes that use the secure electronic transaction (SET) protocol and 3DSET, and users can even get a virtual card just for Internet commerce. Bundled with IBM's e-wallet, the virtual card includes a security code and card number but no physical card.

MasterCard has also initiated several smart-card programs throughout the Asia-Pacific region in the past five years. Most employ a public-key infrastructure (PKI) and the Multos operating system (http://www.multos.com/), which can run multiple applications on each card. Various cards are already being used by banks in 16 Chinese cities, and the Taipei city government and the National Interoperability Project in Australia are using them for identification, tracking, transportation, e-cash, and several other applications. ∎

Semantic Resources

Test Suites / Validators

Cascading Style Sheets
http://www.w3.org/Style/CSS/Test/
HTML Validation Service
http://validator.w3.org/
Mathematical Markup Language
http://w3.org/Math/testsuite/
Scalable Vector Graphics (SVG)
http://w3.org/Graphics/SVG/Test/
IBM XML Schema Quality Checker
http://alphaworks.ibm.com/tech/xmlsqc/

Web Tools

Amaya browser
http://w3.org/Amaya/
Annotea annotation application
http://w3.org/2001/Annotea/
DC-Dot Dublin Core editing tool
http://www.ukoln.ac.uk/metadata/dcdot/
HTML Tidy
http://w3.org/People/Raggett/tidy/
Jigsaw Web server
http://w3.org/Jigsaw/
Jena Java API for RDF
http://www-uk.hpl.hp.com/people/bwm/rdf/jena/index.htm
Redland RDF application framework (beta) http://purl.org/net/redland/

Other Resources

DARPA Agent Markup Language
http://www.daml.org/
Dave Beckett's RDF Resource Guide
http://ilrt.org/discovery/rdf/resources/
Dublin Core Metadata Initiative
http://dublincore.org/
IC's **XML Resources Pages**
http://computer.org/internet/xml/
Int'l Semantic Web Workshop
30-31 July 2001, Stanford University,
http://semanticweb.org/SWWS/
Ontology Inference Layer (OIL)
http://www.ontoknowledge.org/oil/
RDF Interest Group
http://www.w3.org/RDF/Interest/
RDF Interest Group "scratch pad"
http://rdfig.xmlhack.com/
RDF at Mozilla.org
http://www.mozilla.org/rdf/doc/
RDF Site Summary (RSS) 1.0
http://purl.org/rss/1.0/
XML Cover Pages
http://www.oasis-open.org/cover/