Report on WebDB'98: International Workshop on the Web and Databases

Paolo Atzeni
Università di Roma Tre
atzeni@dia.uniroma3.it

Alberto O. Mendelzon
University of Toronto
mendel@db.toronto.edu

The International Workshop on the Web and Databases was held in Valencia, Spain, on March 27 and 28, 1998, following the Sixth International Conference on Extending Database Technology (EDBT'98). The program committee, consisting of Paolo Atzeni, Sophie Cluet, Jon Kleinberg, Alon Levy, Udi Manber, Gianni Mecca, Alberto Mendelzon, Erich Neuhold, and Oded Shmueli, selected 16 papers out of 37 submissions. The international character of the workshop was much in evidence: 8 countries were represented in the authorship of the papers (Australia, France, Germany, Israel, Italy, Switzerland, United Kingdom, and the USA), and 11 countries in the list of 51 registered participants. Even though March 28 was a Saturday and sessions continued well into the late afternoon, the room was always full and discussions were lively.

The Paper Sessions

The first session, *Internet Programming: Tools and Applications* discussed four diverse topics: algorithms for caching web pages, implementation of transaction mechanisms on top of existing WWW protocols, the extension of the persistent programming paradigm to web applications, and a prototype application of the Araneus web site management system to a medical domain. The second one, *Integration and Access to Web Data*, centered on issues of how to query the heterogeneous and semistructured data that is found on the web. The third session, *Hypertext Views on Databases*, included papers from four different groups, all concerned with providing integrated web views over heterogeneous data sources. Finally, the session on *Searching and Mining the Web* presented an overview paper on a large cooperative project in the area of developing web data management tools, a pattern-based incremental approach to integrating widely scattered web information, a method for data mining web server logs, and a technique for finding duplicate or near-duplicate copies of documents on the web.

Preliminary proceedings were distributed at the workshop and are available in electronic form at
`http://poincare.dia.uniroma3.it:8080/webdb98/papers.html`.
A revised and refereed proceedings volume will be published by Springer-Verlag in their LNCS series, and is scheduled for the Fall of 1998.

The Panel

In addition to the paper sessions, a panel on "Databases and the Web: The Next Ten Years" was held midway through the Workshop. The panel was moderated by Alberto Mendelzon, with panelists Karl Aberer (IPSI-GMD), Serge Abiteboul (INRIA), Piero Fraternali (Politecnico di Milano), Tova Milo (Tel Aviv University) and Yannis Papakonstantinou (University of California at San Diego). As Piero pointed out, it is hard enough to predict the next six months, let alone the next ten years; but a number of common issues and concerns emerged from the discussion, including among others:

- Should database researchers participate more actively in web standardization activities such as XML and RDF? Piero and Keith Jeffery forcefully advocated such participation, given the dangers of leaving what are essentially data model and query language design activities to other communities; while Serge suggested there are more productive ways for researchers to spend their time. One way to increase the visibility of database researchers in the WWW community is to submit papers to the WWW conference; a call for papers for WWW8, to be held in Toronto in May of 1999, can be found in `http://www8.org`.

- "The Web is the database, and XML is its model." Yannis' dictum suggests a near future in which a mixture of structured, semistructured, and unstructured data is seamlessly queried and transformed by tools that adapt to the degree of structure present in the data. Such tools to be made possible by communities standardizing on metadata, with a little help from Microsoft.

- Electronic commerce: Karl and Serge emphasized the importance of research that addresses the infrastructure for it, including transactions, security and authentication, negotiation, verification.
- Site management and design: Piero predicted sites will evolve in the direction of putting more and more information into the database, including business rules and user profiles, and tools will be needed to deal with site development at higher levels of abstraction than current ones. Tova pointed out the lack of clear criteria for good site designs.

The Future

There are rumours of a second run of the WebDB workshop, perhaps in combination with SIGMOD '99; Sophie Cluet and Tova Milo are rumoured to be the organizers.

Appendix: List of Papers

Internet Programming: Tools and Applications

A Unified Algorithm for Replacement and Consistency in Web Proxy Servers
J. Shim, P. Scheuermann, R. Vingralek

Transactional Services for the Web
D. Billard

On the Unification of Persistent Programming and the World Wide Web
R. Connor, K. Sibson, P. Manghi

The Fenarete System: a Web based Clinical Protocol Management System
S. Puglia, G. Rumolo

Integration and Access to Web Data

Schema Based Data Translation
T. Milo, S. Zohar

Interactive Query and Search in Semistructured Databases
R. Goldman, J. Widom

Bringing Database Functionality to the WWW
D. Konopnicki, O. Shmueli

Fixpoint Path Queries
N. Bidoit, M. Ykhlef

Hypertext views on databases

A Language for Publishing Virtual Documents on the Web
F. Paradis, A. M. Vercoustre

Incremental Maintenance of Hypertext Views
G. Sindoni

Using YAT to Build a Web Server
J. Simeon, S. Cluet

Language and Tools to Specify Hypertext Views on Databases
G. Falquet, J. Guyot, L. Nerima

Searching and Mining the Web

WebSuite—A Tool Suite for Harnessing Web data
C. Beeri, G. Elber, T. Milo, Y. Sagiv, O. Shmueli, N. Tishby, Y. Kogan, D. Konopnicki, P. Mogilevski, N. Slonim

Extracting Patterns and Relations from the World Wide Web
S. Brin

WUM - A Tool for WWW Ulitization Analysis
M. Spiliopoulou, L. C. Faulstich

Finding Near-Replicas of Documents and Servers on the Web
N. Shivakumar, H. Garcia-Molina