

Guest Editors' Introduction

Arun K. Iyengar and David De Roure

WITHIN the past decade, the World Wide Web has emerged as a critically important mechanism for information dissemination, retrieval, and electronic commerce. Research and development of the Web has been occurring at a rate that has rarely, if ever, been matched in other technological fields. The rapid development of the Web has been enabled by continuous breakthroughs in several areas including information retrieval and searching techniques, Web browsers, languages for representing information, and security. Efficient techniques for serving Web data, caching, load balancing, and replication have allowed Web sites to handle ever increasing amounts of traffic with reasonably high availability.

This special section includes seven papers on research which is laying the foundations for the future of the Web. The papers included in this issue are enhanced versions of seven of the best papers from the Eleventh International World Wide Web Conference (WWW2002) held 7-11 May, 2002, in Honolulu, Hawaii. These papers were selected from 72 papers accepted by the main refereed paper track of the conference out of a total pool of 454 submissions.

The first paper, entitled "Specifying and Enforcing Application-Level Web Security Policies," addresses vulnerabilities inherent in the code of a Web application itself. Several examples of common application-level attacks are presented. The authors present a scalable structuring mechanism for abstracting security policies from large Web applications developed in heterogeneous environments. They also describe tools they have built for programmers to develop secure applications, which are resilient to a wide range of common attacks.

The second paper, entitled "Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search," is geared to improving the ranking of search-query results. The author proposes computing a set of PageRank vectors, biased using a set of representative topics, to capture importance more accurately with respect to a particular topic. This contrasts with previous approaches, which only compute a single vector. The author shows that using multiple vectors in this fashion generates more accurate results than using a single vector.

The third paper, entitled "The Yin/Yang Web: A Unified Model for XML Syntax and RDF Semantics," provides a

unified model for the Extensible Markup Language (XML) and the Resource Description Framework (RDF). These two standards from the World Wide Web Consortium underpin the Semantic Web: XML is used to write and exchange information, while RDF is used to describe the semantics of the information and to reason about it. The paper argues that the syntax and semantics of information need to work together in order to lead the Semantic Web to its full potential, and demonstrates the unified model through an information integration scenario.

The fourth paper, entitled "Scalable Consistency Maintenance in Content Distribution Networks Using Cooperative Leases," addresses cache consistency in content distribution networks. This important application area requires consistency maintenance across a large number of Web caches, with consistency guarantees that can be tailored to meet requirements. The paper introduces the notion of cooperative consistency, in which proxies cooperate with one another to reduce the overheads of consistency maintenance, and a single lease may be shared among multiple caches.

The fifth paper, entitled "Query Expansion by Mining User Logs," proposes a new method for query expansion based on user interactions recorded in the user logs. Queries to search engines are often too short to provide sufficient information for effectively selecting relevant documents, motivating query expansion. The authors' approach extracts correlations between query terms and document terms from user logs. The correlations are then used to select high-quality expansion terms for new queries.

The sixth paper, entitled "Managing and Sharing Servents' Reputations in P2P Systems," proposes an approach to peer-to-peer security in which servents can keep track of, and share with others, information about the reputation of their peers. Shared reputations are based on a distributed polling algorithm in which resource requestors can assess the reliability of providers before initiating downloads. The approach complements existing peer-to-peer protocols and maintains the current level of anonymity of requestors, providers, and other parties sharing views on reputations.

The seventh paper, entitled "Searching with Numbers," addresses an inadequacy in the handling of numbers in current search engines, which typically treat them as strings. It focuses on documents which largely consist of name-number pairs embedded in text, as exemplified by product information. The paper defines a notion called reflectivity, and shows that for low reflectivity data, it is possible to conduct an effective search even if the values in the data have not been assigned attribute names and the

- A. Iyengar is with the IBM T.J. Watson Research Center, PO Box 704, Yorktown Heights, New York. E-mail: aruni@us.ibm.com.
- D. De Roure is with the Department of Electronics and Computer Science, University of Southampton, Southampton, United Kingdom. E-mail: dder@ecs.soton.ac.uk.

For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number 118232.

user has omitted attribute names in the query. It also addresses techniques for high reflectivity data and validates the approach using real data sets.

We would like to thank the many authors, program committee members, and other organizers who contributed to the success of the Eleventh International World Wide Web Conference (WWW2002).

Arun K. Iyengar and David De Roure
Guest Editors



Arun K. Iyengar received the PhD degree in computer science from the Massachusetts Institute of Technology. He does research and development into Web performance, caching, storage allocation, and distributed computing at the IBM T.J. Watson Research Center. His work has improved performance at some of the most highly accessed Web sites hosted by IBM, and has been incorporated into several products including IBM's Websphere. He is the National Delegate representing the IEEE Computer Society to IFIP Technical Committee 6 on Communication Systems, the Chair of IFIP Working Group 6.4 on Internet Applications Engineering, and the Chair of the IEEE Computer Society's Technical Committee on the Internet. He has also earned the distinction of IBM Master Inventor. Dr. Iyengar was program cochair for the 2002 International World Wide Web Conference.



David De Roure received the PhD degree in computer science from the University of Southampton in 1990. He is a professor of computer science in the Department of Electronics and Computer Science at the University of Southampton, United Kingdom, where he is head of the Grid and Pervasive Computing research group. Professor De Roure has worked for many years with distributed information systems and has been active in the hypertext, multimedia, agent-based computing, and distributed computing communities. He is a member of the Advisory Committee of the World Wide Web Consortium and is involved in current activities including the OWL Web Ontology Language. His current research focus is the application of advanced knowledge technologies to distributed systems; his projects focus on semantic grid, knowledgeable devices, and the pervasive knowledge fabric. He is chair of the Semantic Grid Research Group in the Global Grid Forum.