WEB ENGINEERING: THE EVOLUTION OF NEW TECHNOLOGIES

The Web is now the most popular environment for circulating information as network-accessible data; the exponential growth in this information's dissemination is contributing to the evolution of research topics in the field of Web engineering. This issue of *CiSE* magazine focuses on understanding and emphasizing engineering topics as they're applied in today's Web environment and infrastructure.

Many engineers and researchers in the Web engineering community have proposed interesting methodologies, shown important results, and installed helpful implementations. The articles in this special issue cover a wide range of these topics, which fall in the broad categories of Web data representation, access, and effective information retrieval.

1521-9615/04/$20.00 © 2004 IEEE
Copublished by the IEEE CS and the AIP

ATHENA I. VAKALI AND GEORGIOS I. PAPADIMITRIOU
Aristotle University

Data Representation and Access

A primary issue in representing Web data is decid-

ing which standard to use to ensure that it is in a widely acceptable format. Most people have adopted XML, a flexible format that can represent many classes of data, including unstructured text, structured records such as those in relational databases, and semistructured data.[1] In "Managing XML Data: An Overview," Juliana Freire discusses the major issues involved in managing XML data and describes some existing solutions. Due to its flexibility, XML is a natural format for both exchanging and integrating data from diverse data sources (distributed over the Web).[2] But, as Freiere emphasizes, XML's flexibility also implies that we can't expect out-of-the-box solutions for every different application.

Once we have the appropriate representation format and an underlying topology, we need to guarantee a secure access infrastructure. In "Trust Negotiation: Concepts, Systems, and Languages," Elisa Bertino et al. outline trust management over networks and the Web in terms of the needed classes, requirements, and languages.

Effective Information Retrieval

Because the availability of information on the Web increases every day, user queries often produce unnecessary and sometimes misleading information. Emerging topics such as Web mining, Web agents, and refined Web searching aim to solve this problem.

Web caching and prefetching play a key role in information retrieval (IR). "Web Searching and Information Retrieval" by Jaroslav Pokorný overviews today's most important search-engine architectures and techniques. Although it's important to distinguish among the different challenges involved in searching the Web (because it's such a common practice),[6] simple Web searching doesn't always provide the best results. Pranam Kolari and Anupam Joshi thoroughly discuss this topic in "Web Mining: Research and Practice."

Moreover, dissemination is an underlying infrastructure issue; D. Zeinalipour-Yazti et al. present an infrastructure for network-based IR in "Information Retrieval in Peer-to-Peer Networks." This article includes a survey of search techniques for IR in popular peer-to-peer networks.[3,4]

The notion of agents has gained popularity, to better serve Web users' requests. In "Intelligent Agents on the Web: A Review," Mohammad Obaidat and N. Boudriga present a survey of intelligent agents that execute tasks on behalf of individuals and organizations. Goal-driven, autonomous agents that communicate with each other can retrieve, analyze, and evaluate information from heterogeneous online information sources.

In the same context of facilitating access to information, the article "Cache Management for Web Proxy Services" by Bo Li et al. discusses proxy caching, which is an effective technique for reducing the network resources that Web services consume.[7] This article describes the issues and challenges of deploying Web caching proxies over the Internet and focuses on cache management for typical Web objects.

From a technological viewpoint, this special issue presents a profound selection of important research efforts in the area of Web engineering and data management. In particular, it highlights the underlying topologies and environments that support these evolving topics and provides a guide for researchers, academics, and engineers. Such a guide should be beneficial in terms of understanding methodologies and theories involved in the Web engineering field, focusing on current topics and practices related to emerging Web engineering aspects, and adopting Web engineering methodologies and practices in current research and implementations related to the Web.

References

1. T. Bray, J. Paoli, and C.M. Sperberg-McQueen, "Extensible Markup Language (XML) 1.0," 3rd ed., W3C recommendation, Feb. 2004; www.w3.org/TR/REC-xml.
2. V. Christophides and J. Freire, eds., *Int'l Workshop on Web and Databases*, 2003; www.cse.ogi.edu/webdb03/submission.htm.
3. M. Bawa et al., "Make It Fresh, Make It Quick: Searching a Network of Personal Webservers," *Proc. 12th Int'l World Wide Web Conf.*, 2003; www2003.org/cdrom/index.html.
4. R.A. Baeza-Yates and B.A. Ribeiro-Neto, *Modern Information Retrieval*, Addison-Wesley Longman, 1999.
5. E. Bertino, E. Ferrari, and A. Squicciarini, "Trust-X: A Peer-to-Peer Framework for Trust Negotiations," to appear in *IEEE Trans. Knowledge and Data Eng.*, 2004.
6. A. Arasu et al., "Searching the Web," *ACM Trans. Internet Technology*, vol. 1, no. 1, 2001, pp. 2–43.
7. B. Li et al., "On the Optimal Placement of Web Proxies in the Internet," *Proc. Conf. Computer Comm.* (IEEE Infocom '99), IEEE CS Press, 1999, pp. 1282–1290.

Athena I. Vakali is an assistant professor in the Department of Informatics at Aristotle University, Greece. Her research interests include Web data management and data storage, caching, and clustering. She received a PhD in computer science from Aristotle University. She is a member of the ACM, the IEEE Computer Society, and the Usenix Association. Contact her at avakali@csd.auth.gr.

Georgios I. Papadimitriou is an assistant professor in the Department of Informatics at Aristotle University. His research interests include optical networks, wireless networks, high-speed LANs, and learning automata. He received his PhD in computer engineering from the University of Patras, Greece. He is a senior member of the IEEE. Contact him at gp@csd.auth.gr.