
Report on the Fifth ACM International Workshop on Web Information and Data Management (WIDM 2003)

Roger H. L. Chiang

College of Business
University of Cincinnati
Cincinnati, Ohio 45221, USA
Roger.Chiang@uc.edu

Alberto H. F. Laender

Department of Computer Science
Federal University of Minas Gerais
31270-901 – Belo Horizonte – MG, Brazil
laender@dcc.ufmg.br

Ee-Peng Lim

School of Computer Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 639798
aseplim@ntu.edu.sg

1 Introduction

The Fifth ACM International Workshop on Web Information and Data Management (WIDM 2003) was held at the Hotel Inter-Continental, in New Orleans, Louisiana, on November 7-8, 2003, in conjunction with the 12th ACM International Conference on Information and Knowledge Management (CIKM 2003). The main objective of the workshop was to bring together researchers, industrial practitioners and developers to discuss how Web information can be extracted, stored, analyzed and processed to provide useful knowledge to the end users for various advanced database and Web applications.

For the first time, since the first edition of the workshop in 1998, WIDM 2003 implemented a two-days program. WIDM 2003 received the submission of 54 papers from 18 countries, 15 of which were accepted for presentation as full papers and 10 as short papers. All 25 accepted papers were included in the proceedings published by ACM Press [1].

For WIDM 2003, a program committee larger than in previous years was assembled with some of the most prominent researchers in the area. This program committee participated actively in the paper reviewing process, being mostly responsible for the high quality of the workshop technical program.

2 Topics of Interest

As stated in the call-for-papers, the main topics of interest of the workshop were:

- **Web Mining:** Web Usage Mining, Web Classification, Web Clustering, Resource Discovery, Web Personalization, Web Data Extraction, Web Structure Mining;
- **Web Databases:** Web Data Models and Meta-data, Web Query Languages, Web Data Storage Systems, Web Indexing, Web Warehousing, Web Data Integration;
- **Semantic Web:** Web Annotation, Ontologies, Semantic Web Services;

-
- **Web Commerce:** Web Services, Intelligent Agents on the Web, Web Site Modeling and Design, Advanced XML Tools, Mobile Commerce;
 - **Advanced Web Applications:** Digital Libraries, Web Portals, Intelligent Web Search Engines, Web Visualization Tools, Web Information Filtering;
 - **Performance of Web Applications:** Web Caching, Web Monitoring.

Submissions covered most of the above topics with a predominance of papers on Web data extraction, Web data modeling and querying, Web clustering and usage mining, Web data integration, and advanced XML tools. These seem to be some of the “hot” topics the community has been working on these days.

3 Program Overview

All submitted papers were reviewed by three program committee members on a two round reviewing, the second one for resolving discrepancies. The 25 accepted papers were divided into five full paper sessions and two short paper sessions.

The first day program included two full paper sessions and one short paper session. The first full paper session, on *Web Data Extraction and Structure Mining*, included the papers “Schema-Guided Wrapper Maintenance for Web-Data Extraction”, by Xiaofeng Meng, Dongdong Hu and Chen Li from China and USA, “Data Rover: A Taxonomy Based Crawler For Automated Data Extraction From Data-intensive Websites”, by Hasan Davulcu, Sukumar Koduri and S. Nagarajan from USA, and “Fine-grain Web Site Structure Discovery”, by Valter Crescenzi, Paolo Merialdo and Paolo Missier from Italy. The second full paper session, on *XML Data Modeling and Storage*, comprised the papers “Local XML Functional Dependencies”, by Jixue Liu, Millist Vincent and Chengfei Liu from Australia, “A Temporal Data Model and Management System for Normative Texts in XML Format”, by Fabio Grandi, Federica Mandreoli, Paolo Tiberio and Marco Bergonzini from Italy, and “XVerter: Querying XML Data with OR-DBMS, by Humberto Vieira, Gabriela Ruberg and Marta Mattoso from Brazil. The short paper session focused on *Tools for Integrating and Querying Web Information* and included the following papers: “The Biological Integration System”, by Zoé Lacroix, Omar Boucelma and Mehdi Essid from USA and France, “KPSpotter: A Flexible Information Gain-based Keyphrase Extraction System”, by Min Song, Il-Yeol Song and Xiaohua Hu from USA, “MediSeek: A Web Based Diffusion System for Medical Visual Information”, by Silvio Antonio Carro, Jacob Scharcanski and José Valdeni de Lima from Brazil, “OntoKhoj: A Semantic Web Portal for Ontology Searching, Ranking and Classification”, by Chintan Patel, Kaustubh Supekar, Yugyung Lee and E. K. Park from USA, and “XWebSOGO: An Ontology-based Language to Describe and Query Web Sources”, by Edna Ruckhaus and Maria-Esther Vidal from Venezuela.

On the second day of the workshop, we had three additional full paper sessions and one short paper session. The first full paper session addressed the topic *Web Clustering and Usage Mining* and included the papers “Clustering Documents in a Web Directory”, by Giordano Adami, Paolo Avesani and Diego Sona from Italy, “Evaluation of Web Usage Mining Approaches for User's Next Requests”, by Mathias Géry and Hatem Haddad from Finland, and “Evaluating the Markov Assumption for Web Usage Mining”, by Søren Jespersen, Torben Bach Pedersen and Jesper Thorhauge from Denmark. The second full paper session, on *Intelligent Web Information Access*, included the papers “An Intelligent Search Agent System for Semantic Information Retrieval on the Internet”, by Carmine Cesarano, Antonio d’Acerno and Antonio Picariello from Italy, “Conceptual Modeling of Semantic Navigation Structures: The MoSeNa-Approach”, by Jörg Becker, Christian Brelage, Karsten Klose and Michael Thygs from

Germany, and “Answering Imprecise Database Queries: A Novel Approach”, by Ullas Nambiar and Subbarao Kambhampati from USA. The third full paper session was on *Query and View Processing* and included the papers “Query Rewriting Using Views in the Presence of Inclusion Dependencies”, by Qingyuan Bai, Jun Hong and Michael F. McTear from UK, “Generating Rules for Incremental Maintenance of XML View of Relational Data”, by Vânia M. P. Vidal, Marco A. Casanova and Valdiana da Silva Araújo from Brazil, and “Efficiently Supporting Order in XML Query Processing”, by Maged El-Sayed, Katica Dimitrova and Elke A. Rundensteiner from USA. Finally, the short paper session on this second day of the workshop addressed *XML and Information Integration* and included the following papers: “Finding Similar Identities among Objects from Multiple Web Sources”, by Joyce C. P. Carvalho and Altigran S. da Silva from Brazil, “Extracting Association Rules from XML Documents using XQuery”, by Jacky W.W. Wan and Gillian Dobbie from New Zealand, “Partial Answers in Information Integration Systems”, by Gösta Grahne and Victoria Kiricenko from Canada, “Conceptual Modeling of XML Schemas”, by Bernadette F. Lóscio, Ana C. Salgado and Luciano do Rêgo Galvão from Brazil, and “XPath Query Transformation based on XSLT Stylesheets”, by Sven Groppe and Stefan Böttcher from Germany

In addition to the full and short paper sessions, the workshop also included a keynote address on “Business-Aware Management: the Next Frontier for Web Services”, which was presented on the first day by Dr Fabio Casati from HP Labs, Palo Alto. A senior scientist at HP Labs, Dr Casati gave a comprehensive overview of the major issues involving today's Web services and discussed the impact of this new technology on the corporations. His talk was immediately followed by several questions from the workshop audience on the major research issues behind this technology and its future directions.

As for the social events, at the end of first day of the workshop, a happy hour session was organized for the participants at the hotel pub. Most of the participants present on that day turned up at the happy hour to enjoy the typical Louisiana finger food and drinks. Also, an authentic Cajun dinner was organized at the famous French Quarter after the last workshop session.

4 Conclusions

With its two-days program, WIDM 2003 provided a stimulating environment for discussing and disseminating ideas on important issues on Web information and data management. The high quality of the papers presented at WIDM 2003 encouraged the program committee co-chairs to organize again a special issue of the Data and Knowledge journal with selected papers from the final program (the Data and Knowledge Engineering WIDM 2002 special issue is forthcoming [2]). The authors of five of the best WIDM 2003 papers were invited to submit a revised and extended version of their original papers to this new special issue, which is expected to be published by the Springer in 2005.

The next WIDM workshop will be held together with CIKM 2004 (<http://ir.iit.edu/cikm2004>) in Washington, on November 12-13. A new workshop co-chair, Dongwon Lee, from Pennsylvania State University, has been recruited and will replace Roger Chiang and Ee-Peng Lim. WIDM 2004 call-for-paper announcement will be out soon. As for WIDM 2003, a two-day workshop program will be again implemented to encourage more paper submissions and participation.

5 References

- [1] Chiang, R.H.L., Laender, A.H.F. and Lim, E.-P. (eds.). *Proceedings of Fifth ACM CIKM International Workshop on Web Information and Data Management (WIDM 2003)*, November 7-8 2003, New Orleans, Louisiana, USA, ACM Press.
- [2] Chiang, R.H.L., Laender, A.H.F. and Lim, E.-P. (eds.). *Data and Knowledge Engineering, Special Issue: Fourth ACM CIKM International Workshop on Web Information and Data Management (WIDM 2002)*, 49(2), 2004.