

**Report on the 6th ACM International Workshop  
on Web Information and Data Management (WIDM 2004)  
held at CIKM 2004**

**Alberto H. F. Laender**

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

**Dongwon Lee**

School of Information Sciences and Technology  
Pennsylvania State University  
University Park, PA 16802, USA  
*dongwon@psu.edu*

## **1 Introduction**

The 6th ACM International Workshop on Web Information and Data Management (WIDM 2004) was held at the Hyatt Arlington Hotel, in Washington DC, on November 12-13, 2004, in conjunction with the 13th ACM International Conference on Information and Knowledge Management (CIKM 2004). 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. As for previous years, WIDM 2004 was sponsored by ACM SIGIR.

For the second time, since the first edition of the workshop in 1998, WIDM 2004 implemented a two-day program. WIDM 2004 received the submission of 61 papers from 20 countries around the world, 20 of which were accepted for presentation at the workshop and included in the proceedings published by ACM Press [1].

The WIDM 2004 program committee was composed of 38 members from 10 different countries (Australia, Brazil, Canada, China, France, Italy, Singapore, Switzerland, USA and Venezuela). This program committee, which included some of the most prominent researchers in the area, 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 mining and clustering, Web data modeling and querying, Web searching and exploration, Web personalization, 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 to four program committee members on a two round reviewing, the second one for resolving discrepancies. The 20 accepted papers were divided into seven technical sessions.

The first day program included three technical sessions. The first session, on *Web Crawling and Exploration*, included the papers “*A Two-Phase Sampling Technique for Information Extraction from Hidden Web Databases*,” by Yih-Ling Hedley, Muhammad Younas, Anne James and Mark Sanderson from UK, “*SmartCrawl: A New Strategy for the Exploration of the Hidden Web*,” by Augusto de Carvalho Fontes and Fábio Soares Silva from Brazil, and “*Probabilistic Models for Focused Web Crawling*,” by Hongyu Liu, Evangelos Milios and Jeannette Janssen from Canada. The second session, on *XML Processing*, comprised the papers “*Parsing Concurrent XML*,” by Ionut E. Iacob, Alex Dekhtyar and Kazuyo Kaneko from USA, “*A Comprehensive Solution to the XML-to-Relational Mapping Problem*,” by Sihem Amer-Yahia, Fang Du and Juliana Freire from USA, and “*Ctree: A Compact Tree for Indexing XML Data*,” by Qinghua Zou, Shaorong Liu and Wesley Chu from USA. The last session, on *XML and Semistructured Data Querying*, consisted of “*XPath Lookup Queries in P2P Networks*,” by Angela Bonifati, Alfredo Cuzzocrea and Ugo Matrangolo from Italy, and Mayank Jain from India, “*Measuring Similarity Between Collection of Values*,” by Carina F. Dorneles, Carlos A. Heuser, Andrei E. N. Lima, Altigran S. da Silva and Edleno S. de Moura from Brazil, and “*Querying Structured Data in an Unstructured P2P System*,” by Dimitrios Tsoumakos from USA, Verena Kantere from Greece and Nick Roussopoulos from USA.

On the second day of the workshop, we had four additional technical sessions. The first session addressed the topic *Web Personalization* and included the papers “*Web Personalization Integrating Content Semantics and Navigational Patterns*,” by Magdalini Eirinaki, Charalampos Lampos, Stratos Pavlakis and Michalis Vazirgiannis from Greece, “*Web Personalization Based on Static Information and Dynamic User Behavior*,” by Massimiliano Albanese, Antonio Picariello, Carlo Sansone and Lucio Sansone from Italy, and “*A Clickstream-Based Collaborative Filtering Personalization Model: Towards A Better Performance*,” by Dong-Ho Kim, Il Im, Nabil Adam, Vijayalakshmi Atluri, Michael Bieber and Yelena Yesha from USA. The second session, on *Web Searching*, comprised the papers “*Grouping Search-Engine Returned Citations for Person-Name Queries*,” by Reema Al-Kamha and David W. Embley from USA, “*WISE-Cluster: Clustering E-Commerce Search Engines Automatically*,” by Qian Peng, Weiyi Meng, Hai He and Clement Yu from USA, and “*Specialisation Dynamics in Federated Web Search*,” by Rinat Khossainov and Nicholas Kushmerick from Ireland. The third session, on *Web Mining and*

---

*Clustering*, included the following papers: “A Version Model for Supporting Adaptation of Web Pages,” by Rodrigo Giacomini Moro, Renata de Matos Galante and Carlos Alberto Heuser from Brazil, “FS-Miner: Efficient and Incremental Mining of Frequent Sequence Patterns in Web Logs,” by Maged El-Sayed, Carolina Ruiz and Elke A. Rundensteiner from USA, and “Stylistic and Lexical Co-training for Web Block Classification,” by Chee How Lee, Min-Yen Kan and Sandra Lai from Singapore. Finally, the last session, on User Interfaces and Services, consisted of only two papers: “User Evaluation of the NASA Technical Report Server Recommendation Service,” by Michael L. Nelson, Johan Bollen, JoAnne R. Calhoun and Calvin E. Mackey from USA, and “A Web Interface to Image-based Concurrent Markup Using Image Maps,” by Jerzy W. Jaromczyk from USA, Mirosław Kowaluk from Poland and Neil Moore from USA.

In addition to the technical sessions, the workshop also included a keynote address on “Next Generation CiteSeer”, which was presented on the first day by Dr. C. Lee Giles from Pennsylvania State University. A David Reese Professor at Pennsylvania State University, Dr. Giles gave a comprehensive (and witty) overview of the widely popular digital library system, CiteSeer, and discussed the challenging issues in building the next generation of the CiteSeer system. His talk was immediately followed by several questions from the workshop audience on the major research issues behind this technology and its future directions.

## 4 Conclusions

With its two-day program, WIDM 2004 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 2004 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 2003 special issue is forthcoming [2]). The authors of five of the best WIDM 2004 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 2005 (<http://www.tzi.de/CIKM2005/>) in Bremen, Germany, from October 31 to November 5. A new workshop co-chair, Angela Bonifati, from ICAR-CNR, Italy, has been recruited and will replace Alberto H. F. Laender. WIDM 2005 call-for-paper announcement will be out soon. Unlike WIDM 2004, a one-day workshop program will be implemented to accommodate feedbacks gathered from attendants and organizers of WIDM 2004.

## 5 References

- [1] Laender, A.H.F. and Lee, D. (eds.). *Proceedings of the Sixth ACM International Workshop on Web Information and Data Management (WIDM 2004)*, November 12-13 2004, Washington, DC, USA, ACM Press.
- [2] Chiang, R.H.L., Laender, A.H.F. and Lim, E.-P. (eds.). *Data and Knowledge Engineering, Special Issue: Fifth ACM International Workshop on Web Information and Data Management (WIDM 2003)*. (to be published)