

Report on the 7th ACM International Workshop on Web Information and Data Management (WIDM 2005)

Angela Bonifati

Icar CNR
Italian National Research Council
I-87036 – Rende – Italy
bonifati@icar.cnr.it

Dongwon Lee

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

1 Introduction

The 7th ACM International Workshop on Web Information and Data Management (WIDM 2005) was held at the Hilton Bremen Hotel in Bremen (Germany), on November 5, 2005, in conjunction with the 14th ACM International Conference on Information and Knowledge Management (CIKM 2005). 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. WIDM 2005 was sponsored by ACM SIGIR and by GI (Gesellschaft für Informatik) and was in cooperation with ACM SIGMOD.

WIDM 2005 implemented a one-day program as it happened in previous editions since 1998. WIDM 2005 received the submission of 44 papers from 15 countries around the world, 12 of which were accepted for presentation at the workshop and included in the proceedings published by ACM Press [1].

For the first time in WIDM tradition, we accepted 8 full (8 pages long) and 4 short (6 pages long) contributions.

The WIDM 2005 program committee was composed of 39 members from 14 different countries (Australia, Brazil, Canada, China, France, Germany, Greece, Ireland, Italy, Japan, 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 thus 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 and XML Data Management:** Web Data Models and Meta-data, XML Query Languages, XML Data Storage Systems, XML Indexing, XML Warehousing, XML 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;
 - **Web Exploration:** Web crawling, Intelligent Web Search Engines;
 - **Performance of Web Applications:** Web Caching, Web Monitoring.

Submissions covered most of the above topics with a predominance of papers on Web ranking and retrieval, XML data management and Web discovery, Web clustering and filtering. These seem to be some of the “hottest” topics that 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 12 accepted papers were divided into three technical sessions.

The first session, on *Web Ranking and Retrieval*, included the papers “*Web path recommendations based on page ranking and Markov models*” by Magdalini Eirinaki, Michalis Vazirgiannis and Dimitris Kapogiannis from Greece, “*Semantic Similarity Methods in WordNet and their Application to Information Retrieval on the Web*” by Giannis Varelas, Epimenidis Voutsakis, Paraskevi Raftopoulou, Euripides G.M. Petrakis and Evangelos Milios from Greece and Canada, and “*DirectoryRank: Bringing Order to Web Directories*” by Vlassis Krikos, Sofia Stamou, Alexandros Ntoulas, Pavlos Kokosis and Dimitris Christodoulakis from Greece and USA.

The second session, on *XML Data Management and Web Discovery* comprised the papers “*Exploiting Native XML Indexing Techniques for XML Retrieval in Relational Database Systems*,” by Felix Weigel, Klaus U. Schultz and Holger Meuss from Germany, “*Query Translation Scheme for Heterogeneous XML Data Sources*” by Cindy X. Chen, George A. Mihaila, Sriram Padmanabhan and Isabelle M. Rouvellou from USA, “*Impact of XML Schema Evolution on Valid Documents*” by Giovanna Guerrini, Marco Mesiti and Daniele Rossi from Italy, and “*A Framework for Semantic Web Services Discovery*” by Jyotishman Pathak, Neeray Koul, Doina Caragea and Vasant G. Honavar from USA.

The last session, on *Web Clustering, Filtering and Applications*, consisted of “*Narrative Text Classification for Automatic Key Phrase Extraction in Web Document Corpora*,” by Yongzheng Zhang, Nur Zincir-Heywood and Evangelios Milios from Canada, “*On improving Local Website Search Using Web Server Traffic Logs: A preliminary Report*,” by Qing Cui and Alex Dekhtyar from USA, “*Preventing Shilling Attacks in Online Recommender Systems*,” by Cristian Zamfir, Paul-Alexandru Chirita, and Wolfgang Nejdl from Germany, “*Looking at Both the Present and the Past to Efficiently Update Replicas of Web Content*” by Luciano Barbosa, Ana Carolina Salgado, Francisco de Carvahlo, Jacques Robin and Juliana Freire from USA and Brazil, and “*A Search Result Clustering Method using Informatively Named Entities*” by Hitoyuki Toda and Ryoji Kataoka from Japan.

In addition to the technical sessions, the workshop also included a keynote address on “*A Web of Data; New Architectures for New Technology?*”, which was presented by Prof. Donald Kossmann from ETH Zurich, Switzerland. The development and evolution that the Web has undergone this last

decade has not been coupled with an evolution of construction methods of Web applications. To such a purpose, imperative programming languages (e.g., Java or C#) and middleware architectures are still dominant in industry. His talk discussed why these old architectures are problematic and spurred interesting ideas for novel software architectures to build Web applications.

His talk was immediately followed by several questions from the workshop audience on the major research issues behind this technology and its future interesting directions.

4 Conclusions

WIDM 2005 provided a stimulating environment for discussing and disseminating ideas on important issues on Web information and data management. The next WIDM workshop will be held together with CIKM 2006 (<http://www.cikm.org/CIKM2006-CFP.pdf>) in Arlington (VA), USA on November 10, 2006. A new workshop co-chair, Iriini Fundulaki, from University of Edinburgh, UK, has been recruited and will replace Dongwon Lee. WIDM 2006 call-for-paper will be announced soon.

5 References

[1] Laender, A.H.F. and Lee, D. (eds.). *Proceedings of Sixth ACM International Workshop on Web Information and Data Management (WIDM 2005)*, November 5, 2005, Bremen, Germany, ACM Press.