



CIKM 2006 Workshops

Proceedings of
CAMA'06 • DOLAP'06
ACM GIS'06
HIKM'06 • P2PIR'06
TMBIO'06 • WIDM'06

November 6-11, 2006
Arlington, VA, USA

Sponsored by:
ACM SIGIR & ACM SIGWEB

ISBN:
1-59593-531-2
ACM Order No.
605069



**ACM
Fifteenth
Conference
on
Information
and
Knowledge
Management**



Workshops

Main Page

Copyright © 2006 by the Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission to republish from: Publications Dept., ACM, Inc.

P2PIR'06

Fax +1 (212) 869-0481 or permissions@acm.org

WIDM'06

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

TMBIO'06

ACM-GIS'06

HIKM'06

CAMA'06

Author Index

Notice to Past Authors of ACM-Published Articles

ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM. If you have written a work that has been previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform permissions@acm.org, stating the title of the work, the author(s), and where and when published.

ISBN: 1-59593-525-8

Additional copies may be ordered prepaid from:

ACM Order Department

PO Box 11405, New York, NY 10286-1405

ACM Order Number 605069



Please be sure you have the latest version and updates of Acrobat installed.

You can find it at:

<http://www.adobe.com/products/acrobat/update.html>

Printed & Made in the USA

Foreword

The *2006 International Workshop on Web Information and Data Management (WIDM 2006)* is the eighth in a series of workshops on Web Information and Data Management held in conjunction with the International Conference on Information and Knowledge Management (CIKM). The objective of the workshop is to bring together researchers, industrial practitioners, and developers to study how Web information can be extracted, stored, analyzed, and processed to provide useful knowledge to the end users for various advanced database applications. WIDM 2006 has received the sponsorship from ACM SIGIR. We hope that these proceedings will serve as a valuable reference for all experts in the field.

The call for papers resulted in the submission of 51 papers from Asia, Australia, Canada, Europe, South America, and the United States. Starting from last year, the workshop has a one-day schedule. All papers were thoroughly reviewed by the program committee and external reviewers. This year, the program committee accepted 11 papers, resulting in a competitive 21.5% acceptance rate. The 11 accepted papers were divided into 4 sessions: “XML Data Management and P2P,” “Web Ranking and Classification,” “Web Resource Crawling and Searching”, and “Web Organization”. In addition, the WIDM 2006 program also includes an invited talk on “The Power of Structured Data on the Web” by Sihem Amer-Yahia, from Yahoo! Research (USA).

The workshop would not be possible without the support from the School of Informatics of the University of Edinburgh, which provided both the manpower and computing resources to host the workshop Web Site, and from the Microsoft Conference Management Tool, which was used to run the review process.

We would like to thank the CIKM 2006’s main committee for its support. We are particularly thankful to Prof. Il-Yeol Song and to Prof. Eun-Kyo Park for being the CIKM 2006 workshop and organization chairs, respectively.

Finally we would like to thank authors, program committee members, external reviewers, and contributors for their effort in supporting the workshop.

Angela Bonifati and Irini Fundulaki
Program Committee Co-Chairs

WIDM 2006:
The Eighth International Workshop on Web information and Data Management

November 10,
2006



Main Page

[Angela Bonifati and Irini Fundulaki \(editors\)](#)

[WIDM 2006 Welcome](#)

DOLAP'06

[WIDM 2006 Workshop Organization](#)

WIDM'06

[Additional Reviewers](#)

[Sponsor](#)

TMBIO'06

Keynote Address

Session Chair: A. Bonifati (Icar CNR, Italy)

GIS'06

The Power of Structured Data on the Web ([Page 1](#))

S. Amer-Yahia (Yahoo! Research)

P2PIR'06

Session 1: XML Data Management and P2P

Session Chair: I. Fundulaki (University of Edinburgh)

Bridging Relational Database History and the Web: the XML Approach ([Page 3](#))

F. Wang (Siemens Corporate Research)

X. Zhou (NCR Corporation)

C. Zaniolo (University of California at Los Angeles)

Efficient Evaluation of n-ary Conjunctive Queries over Trees and Graphs ([Page 11](#))

F. Bry, T. Furche, B. Linse, A. Schroeder (University of Munich)

SRI: Exploiting Semantic Information for Effective Query Routing in a PDMS ([Page 19](#))

F. Mandreoli, R. Martoglia, S. Sassatelli (University of Modena and Reggio Emilia)

W. Penzo (University of Bologna)

Session 2: Web Ranking and Classification

Session Chair: P. Mitra (The Pennsylvania State University)

Ranking Target Objects of Navigational Queries ([Page 27](#))

L. Raschid, Y. Wu, W.-J. Lee (University of Maryland)

P. Tsaparas (University of Helsinki)

M. E. Vidal (Universidad Simón Bolívar)

P. Srinivasan, A. K. Sehgal (The University of Iowa)

Coarse-grained Classification of Web Sites by Their Structural Properties ([Page 35](#))

C. Lindemann, L. Littig (University of Leipzig)

Efficient, Automatic Web Resource Harvesting ([Page 43](#))

M. L. Nelson, J. A. Smith, I. G. del Campo (Old Dominion University)

H. Van de Sompel, X. Liu (Los Alamos National Laboratory)

Session 3: Web Resource Crawling and Searching

Session Chair: F. Geerts (University of Edinburgh)

Identifying Redundant Search Engines in a Very Large Scale Metasearch Engine Context ([Page 51](#))

R. Desai, W. Meng (SUNY Binghamton)

Q. Yang (University of Wisconsin-Platteville)

*Z. Wu (University of Louisiana at Lafayette)
C. Yu (University of Illinois at Chicago)*

Pushing Task Relevant Web Links down to the Desktop ([Page 59](#))
P.-A. Chirita, C. S. Firer, W. Nejdl (University of Hannover)

Lazy Preservation: Reconstructing Websites by Crawling the Crawlers
([Page 67](#))
*F. McCown, J. A. Smith, M. L. Nelson (Old Dominion University)
J. Bollen (Los Alamos National Laboratory)*

Session 4: Web Organization

Session Chair: F. McCown (Old Dominion University)

An Architecture for Creating Collaborative Semantically Capable Scientific Data Sharing Infrastructures ([Page 75](#))

A. R. Jaiswal, C. L. Giles, P. Mitra, J. Z. Wang (The Pennsylvania State University)

The GEON Portal: Accelerating Knowledge Discovery in the Geosciences ([Page 83](#))

*U. Nambiar, B. Ludaescher (University of California at Davis)
K. Lin, C. Baru (University of California at San Diego)*

[Return to Top](#)

WIDM 2006 Workshop Organization

Program Chairs: Angela Bonifati (*Icar CNR, Italy*)

Irini Fundulaki (*University of Edinburgh, UK*)

Program Committee: Bernd Amann (*Universite Pierre et Marie Curie, France*)

Denilson Barbosa (*University of Calgary, Canada*)

Daniele Braga (*Politecnico di Milano, Italy*)

Vanessa Braganholo (*Universidade Federal do Rio de Janeiro Brazil*)

Cindy Chen (*University of Massachussets at Lowell, USA*)

Sophie Cluet (*INRIA-Rocquencourt, France*)

Mariano Consens (*University of Toronto, Canada*)

Alin Deutsch (*University of California San Diego, USA*)

Ariel Fuxman (*University of Toronto, Canada*)

Floris Geerts (*University of Edinburgh, UK*)

Giorgio Ghelli (*Universita di Pisa, Italy*)

Anastasios Kementsietsidis (*University of Edinburgh, UK*)

Christoph Koch (*Saarland University, Germany*)

Georgia Koutrika (*University of Athens, Greece*)

Alexandros Labrinidis (*University of Pittsburgh, USA*)

Dongwon Lee (*Penn State University, USA*)

Daniel Lieuwen (*Bell Laboratories, USA*)

Sebastian Maneth (*National ICT, Australia*)

Amelie Marian (*Rutgers University, USA*)

Maarten Marx (*University of Amsterdam, The Netherlands*)

Elio Masciari (*Icar-CNR, Italy*)

Laurent Mignet (*IBM New Delhi, India*)

Dan Olteanu (*Saarland University, Germany*)

Dimitris Plexousakis (*University of Crete, Greece*)

Neoklis Polyzotis (*University of California, Santa Cruz, USA*)

Rachel Pottinger (*University of British Columbia, Canada*)

Ganesh Ramesh (*Microsoft, USA*)

Jerome Simeon (*IBM Watson, USA*)

Radu Sion (*State University of NY, StonyBrook, USA*)

Val Tannen (*University of Pennsylvania, USA*)

Zografoula Vagena (*IBM Almaden Research Center, USA*)

Maurice Van Keulen (*University of Twente, The Netherlands*)

Michalis Vazirgiannis (*Athens U. of Economics and Business, Greece*)

Yannis Velegrakis (*AT&T Labs Research, USA*)

Pierangelo Veltri (*University of Catanzaro, Italy*)

Anne-Marie Vercoustre (*INRIA-Rocquencourt, France*)

Ming Xiong (*Bell Laboratories, USA*)

External Reviewers:	Sadek Ali Ricardo Baeza-Yates Ioannis Chrysakis Camelia Constantin Magdalini Eirinaki Shenoda Guirguis Yaron Kanza Hima P Karanam	Edgar Meij Mirella Moura Moro Sylvain Renaudier Erik Tjong Kim Sang Andrea Tagarelli Paolo Trunfio Ranga Raju Vatsavai Jie Xu
----------------------------	--	--

