Welcome to the CIKM'07 Co-Located Workshops Proceedings on CD

Irini Fundulaki, Neoklis Polyzotis, Chris Jones, Ross Purves, Jian Pei, Aparna Varde, Torben Bach Pedersen, Il-Yeol Song, Leslie Carr, C. Lee Giles, Prasenjit Mitra, Editors

These CD-proceedings are a cross-platform medium which allows Windows and Mac users to share the same directory structure and access a common set of files. To navigate this CD a graphical web browser is required. The full-text content on this disk is in Adobe PDF format. A version of Adobe Acrobat Reader is required to view the content.

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

Sponsored by:

UNIVERSIDADE DE LISBOA

Copyright © 2007 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. Fax +1 (212) 869-0481 or permissions@acm.org

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.

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.


Additional copies may be ordered prepaid from: ACM Order Department General Post Office P.O. Box 30777, New York, NY 10087-0777

ACM Order Number 605079

Made in the USA
Foreword

The ACM CIKM07 Workshop on Web Information and Data Management (WIDM 2007) is the ninth 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. 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 80 papers from 32 countries: Australia, Austria, Belgium, Brazil, Canada, Chile, China, Cuba, Cyprus, Czech Republic, France, Germany, Greece, India, Iran, Israel, Italy, Japan, Mauritius, Mexico, Netherlands, Norway, Pakistan, Poland, Portugal, Singapore, South Korea, Sweden, Switzerland, Taiwan, United Kingdom and the United States.

Starting from 2005, 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 20 papers, resulting in a competitive 25% acceptance rate. The 20 accepted papers were divided into five sessions: “XML and Semi-Structured Data”, “P2P and System Design Issues”, “Personalization”, “Mining Knowledge from Web Data” and “Web Metadata and Search”.

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 EasyChair Conference System, which was used to run the review process.

We would like to thank the main committee of CIKM 2007 for its support. We are particularly thankful to Prof. Dongwon Lee for being the CIKM 2007 workshops chair.

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

Irini Fundulaki and Neoklis Polyzotis

WIDM 2007 Program Committee Co-Chairs
## WIDM 2007 Table of Contents

### Foreword

#### Organization

#### Session 1: XML and Semi-Structured Data

- **Session Chair: Maarten Marx (University of Amsterdam)**
- **XCentric: Logic Programming for XML Processing** (Page 1)
  - Jorge Coelho (Instituto Superior de Engenharia do Porto & LIACC)
  - Mario Florido (University of Porto)
  - Evaluation of Datalog Extended with an XPath Predicate (Page 9)
    - Royi Ronen (Technion - Israel Institute of Technology)
    - Oded Shmueli (Technion - Israel Institute of Technology)
  - An Approach to XML Path Matching (Page 17)
    - Alexander R. Vinson (UFRGS - II)
    - Carlos A. Heuser (UFRGS - II)
    - Altigran S. da Silva (UFAM - ICE)
    - Edleno S. de Moura (UFAM - ICE)
- **FleDEx: Flexible Data Exchange** (Page 25)
  - Filipe Mesquita (Federal University of Amazonas)
  - Denilson Barbosa (University of Calgary)
  - Eli Cortez (Federal University of Amazonas)
  - Altigran S. da Silva (Federal University of Amazonas)

### Session 2: P2P and System Design Issues

- **Session Chair: Neoklis Polyzotis (University of California-Santa Cruz)**
- **Data Allocation Scheme Based on Term Weight for P2P Information Retrieval** (Page 33)
  - Hisashi Kurasawa (The University of Tokyo)
  - Hiromi Wakaki (Toshiba R & D Center)
  - Atsuhiro Takasu (National Institute of Informatics)
  - Jun Adachi (National Institute of Informatics)
- **Distributed Monitoring of Peer to Peer Systems** (Page 41)
  - Serge Abiteboul (INRIA Saclay and University Paris Sud)
  - Bogdan Marinoiu (INRIA Saclay and University Paris Sud)
- **Self-optimizing Block Transfer in Web Service Grids** (Page 49)
  - Anastasios Gounaris (University of Cyprus)
  - Christos Yfoulis (ATEI of Thessaloniki)
  - Rizos Sakellariou (University of Manchester)
  - Marios D. Dikaiakos (University of Cyprus)
- **Load Balancing Distributed Inverted Files** (Page 57)
  - Mauricio Marin (Yahoo! Research)
  - Carlos Gomez (Yahoo! Research)

### Session 3: Personalization

- **Session Chair: Dongwon Lee (The Pennsylvania State University)**
- **Supporting Personalized Top-k Skyline Queries using Partial Compressed Skycube** (Page 65)
  - Jongwuk Lee (POSTECH)
  - Gae-won You (POSTECH)
  - IkChan Sohn (POSTECH)
  - Seung-won Hwang (POSTECH)
  - Kwangil Ko (Alticast Corp.)
  - Zino Lee (Alticast Corp.)
- **Toward Editable Web Browser: Edit-and-Propagate Operation for Web Browsing** (Page 73)
  - Satoshi Nakamura (Kyoto University)
  - Takehiro Yamamoto (Kyoto University)
  - Katsumi Tanaka (Kyoto University)
- **Mining User Navigation Patterns for Personalizing Topic Directories** (Page 81)
  - Theodore Dalamagas (National Technical University of Athens)
  - Panagiotis Bouros (National Technical University of Athens)
  - Theodore Galanis (University of Edinburgh)
  - Magdalini Eirinaki (San Jose State University)
  - Timos Sellis (National Technical University of Athens)
An Online PPM Prediction Model for Web Prefetching (Page 89)
Zhijie Ban (Beijing Institute of Technology)
Zhimin Gu (Beijing Institute of Technology)
Yu Jin (Beijing Institute of Technology)

Session 4: Mining Knowledge from Web Data
Session Chair: Magdalini Eirinaki (San Jose State University)

Extracting the Discussion Structure in Comments on News-Articles (Page 97)
Anne Schuth (University of Amsterdam)
Maarten Marx (University of Amsterdam)
Maarten de Rijke (University of Amsterdam)

Adaptive Web-page Content Identification (Page 105)
John Gibson (The MITRE Corporation)
Ben Wellner (The MITRE Corporation)
Susan Lubar (The MITRE Corporation)

Pattern Detection from Web using AFA Set Theory (Page 113)
Ikumi Horie (Dokkyo University)
Kazumori Yamaguchi (University of Tokyo)
Kenji Kashiwabara (University of Tokyo)

Web Based Linkage (Page 121)
Ergin Elmacioglu (The Pennsylvania State University)
Min-Yen Kan (National University of Singapore)
Dongwon Lee (The Pennsylvania State University)
Yi Zhang (University of California, Santa Cruz)

Session 5: Web Metadata and Search
Session Chair: Irini Fundulaki (University of Edinburgh)

Using Neighbors to Date Web Documents (Page 129)
Sérgio Nunes (Universidade do Porto)
Cristina Ribeiro (Universidade do Porto/INESC-Porto)
Gabriel David (Universidade do Porto/INESC-Porto)

Detecting Age of Page Content (Page 137)
Adam Jatowt (Kyoto University)
Yukiko Kawai (Kyoto Sangyo University)
Katsumi Tanaka (Kyoto University)

On Improving Wikipedia Search using Article Quality (Page 145)
Meiqun Hu (Nanyang Technological University)
Ee-Peng Lim (Nanyang Technological University)
Aixin Sun (Nanyang Technological University)
Ba-Quy Vuong (Nanyang Technological University)

SATYA: A Reputation-based Approach for Service Discovery and Selection in Service-Oriented Architectures (Page 153)
Alexandre G. Lages (Federal University of Rio de Janeiro)
Luci Pirmez (Federal University of Rio de Janeiro)
Paulo F. Pires (Federal University of Rio Grande do Norte)
Flávia C. Delicato (Federal University of Rio Grande do Norte)
WIDM 2007 Organization

Program Chairs:
Irini Fundulaki (University of Edinburgh, UK)
Neoklis Polyzotis (University of California-Santa Cruz, USA)

Program Committee:
Serge Abiteboul (INRIA-Futurs, France)
Ashraf Aboulnaga (University of Waterloo, Canada)
Pablo Barcelo (University of Chile, Chile)
Omar Benjelloun (Google Inc., USA)
Angela Bonifati (ICAR-CNR, Italy)
Chee-Yong Chan (National University of Singapore, Singapore)
Vassilis Christophides (University of Crete and ICS-FORTH, Greece)
Magdalini Eirinaki (San Jose State University, USA)
Tim Furche (Ludwig-Maximilians-University of Munich, Germany)
Floris Geerts (University of Edinburgh, UK)
Zoltan Gyongyi (Stanford University, USA)
Ihab Ilyas (University of Waterloo, Canada)
Benny Kimelfeld (Hebrew University of Jerusalem, Israel)
Rajasekar Krishnamurthy (IBM Almaden Research Center, USA)
Boon Thau Loo (University of Pennsylvania, USA)
Bertram Ludaescher (University of California-Davis, USA)
Qiong Luo (Hong Kong University of Science and Technology, Hong Kong)
Sebastian Maneth (National ICT and UNSW, Australia)
Ioana Manolescu (INRIA-Futurs, France)
Gerome Miklau (University of Massachusetts, Amherst, USA)
Tova Milo (Tel Aviv University, Israel)
Prasenjit Mitra (Penn State University, USA)
Michael Nelson (Old Dominion University, USA)
Alexandros Ntoulas (Microsoft Research, USA)
Fatma Ozcan (IBM Almaden, USA)
Evaggelia Pitoura (University of Ioannina, Greece)
Arnaud Sahuguet (Google Inc., USA)
Wang-Chiew Tan (University of California-Santa Cruz, USA)
Martin Theobald (Stanford University, USA)
Vasilis Vassalos (Athens University of Economics and Business, Greece)
Yannis Velegrakis (University of Trento, Italy)
Stratis Viglas (University of Edinburgh, UK)
Yi Zhang (University of California-Santa Cruz, USA)
Additional reviewers: Mehmet Altinel
                Andrei Arion
                Tawan Banchuen
                George Beskales
                Harrison Duong
                Olivier Gauwin
                Periklis Georgiadis
                Claudio Gutierrez
                Shu Huang
                Kazuo Inaba
                Yaron Kanza
                Gerwin Klein
                Joseph Kopena
                Aurelien Lemay

                Benedikt Linse
                Mengmeng Liu
                Shuai Ma
                Heiko Mueller
                Joachim Niehren
                Chichun Pan
                Jorge Perez
                Mohamed Soliman
                Bingjun Sun
                Qingzhao Tan
                Zografoula Vagena
                Xiao Zhang
                Steven Zhou

Sponsors:

Local Arrangements:

UNIVERSIDADE DE LISBOA