# Table of Contents

## Keynote

1 The New Economy — An Engineer’s Perspective  
D. Brown, Motorola, UK

## Session 1: Ontologies

Chair: P. Bouquet, University of Trento

3 Position Paper: A Comparison of Two Modelling Paradigms in the Semantic Web  
P. F. Patel-Schneider, Lucent Technologies  
I. Horrocks, University of Manchester

13 Web Ontology Segmentation: Analysis, Classification and Use  
J. Seidenberg, A. Rector, University of Manchester

23 Constructing Virtual Documents for Ontology Matching  
Y. Qu, W. Hu, G. Cheng, Southeast University

## Session 2: Adaptivity & Mobility

Chair: E. Brewer, Intel Research Berkeley

33 Browsing on Small Screens: Recasting Web-Page Segmentation into an Efficient Machine Learning Framework  
S. Baluja, Google, Inc.

43 Image Classification for Mobile Web Browsing  
T. Maekawa, T. Hara, S. Nishio, Osaka University

53 Fine Grained Content-based Adaptation Mechanism for Providing High End-User Quality of Experience with Adaptive Hypermedia Systems  
C. H. Muntean, J. McManis, Dublin City University

## Session 3: Fighting Search Spam

Chair: Z. Bar-Yossef, Technion

63 Topical TrustRank: Using Topicality to Combat Web Spam  
B. Wu, V. Goel, B. D. Davison, Lehigh University

73 Site Level Noise Removal for Search Engines  
A. L. da Costa Carvalho, Federal University of Amazons  
P.-A. Chirita, L3S and University of Hannover  
E. S. de Moura, Federal University of Amazons  
P. Calado, IST/INESC-ID  
W. Nejdl, L3S and University of Hannover

83 Detecting Spam Web Pages through Content Analysis  
A. Ntoulas, University of California at Los Angeles  
M. Najork, M. Manasse, D. Fetterly, Microsoft Research

## Session 4: XML

Chair: M. Solanki, De Montfort University

93 XML Screamer: An Integrated Approach to High Performance XML Parsing, Validation and Deserialization  
M. G. Kostoulas, M. Matsa, N. Mendelsohn, E. Perkins, A. Heifets, IBM Corporation  
M. Mercaldi, University of Washington

103 Symmetrically Exploiting XML  
S. Zhang, C. Dyreson, Washington State University

## Session 5: Developing Regions & Peer-to-Peer

Chair: K. Toyama, Microsoft Research India

113 FeedEx: Collaborative Exchange of News Feeds  
S. Jun, M. Ahamad, Georgia Institute of Technology

## Session 6: Browsers

Chair: K. Bharat, Google Inc.

123 Examining the Content and Privacy of Web Browsing Incidental Information  
K. Hawkey, K. M. Inkpen, Daffousie University

133 Off the Beaten Tracks: Exploring Three Aspects of Web Navigation  
H. Weinreich, H. Obendorff, University of Hamburg  
E. Herder, University of Twente  
M. Mayer, University of Hamburg

143 pTHINC: A Thin-Client Architecture for Mobile Wireless Web  
J. Kim, R. A. Baratto, J. Nieh, Columbia University

## Session 7: E-Communities

Chair: B. Ludaescher, University of California at Davis

153 Bringing Communities to the Semantic Web and the Semantic Web to Communities  
K. F. Lawrence, m. c. schraefel, University of Southampton

163 Invisible Participants: How Cultural Capital Relates to Lurking Behavior  
V. Soroka, S. Rafaeli, University of Haifa

173 Probabilistic Models for Discovering E-Communities  

## Session 8: User Interfaces: Semantic Tagging

Chair: Y. Maarek, IBM Research, Haifa

183 The Web Beyond Popularity: A Really Simple System for Web Scale RSS  
D. Gruhl, D. N. Meredith, J. H. Pieper, A. Cozzi, S. Dill, IBM Almaden Research Center

193 Visualizing Tags over Time  
M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, A. Tomkins, Yahoo! Research

203 Knowing the User’s Every Move – User Activity Tracking for Website Usability Evaluation and Implicit Interaction  
R. Atterer, M. Wnuk, A. Schmidt, University of Munich

## Session 9: Mining the Web

Chair: D. Pavlov, Yahoo! Research

213 Finding Advertising Keywords on Web Pages  
W.-t. Yih, J. Goodman, Microsoft Research  
V. R. Carvalhalo, Carnegie Mellon University
Table of Contents

223 Communities from Seed Sets
R. Andersen, University of California at San Diego
K. J. Lang, Yahoo! Research

M. Toyoda, M. Kitsuregawa, University of Tokyo

Keynotes
243 A Case for Software Assurance
M. A. Davidson, Oracle Corporation

245 ‘e-Science and Cyberinfrastructure: A Middleware Perspective’
T. Hey, Microsoft Corporation

Session 12: Correctness & Security
Chair: H. Halpin, University of Edinburgh
247 SecuBat: A Web Vulnerability Scanner
S. Kals, E. Kirda, C. Kruegel, N. Jovanovic, Technical University of Vienna

257 Access Control Enforcement for Conversation-based Web Services
M. Meccia, Università di Roma
M. Ouzzani, Purdue University
F. Pisci, Università degli Studi di Milano
E. Bertino, Purdue University

267 Analysis of Communication Models in Web Service Compositions
R. Kazhamiakin, M. Pistore, L. Santuari, University of Trento

Session 13: Search Engine Engineering
Chair: D. Lowe, University of British Columbia
277 Toward Tighter Integration of Web Search with a Geographic Information System
T. Tezuka, T. Kurashima, K. Tanaka, Kyoto University

287 Geographically Focused Collaborative Crawling
W. Gao, Genieknows.com
H. C. Lee, University of Toronto
Y. Miao, Genieknows.com

297 To Randomize or Not To Randomize: Space Optimal Summaries for Hyperlink Analysis
T. Sarlós, A. A. Benczúr, K. Csatógáy, Hungarian Academy of Sciences and Eötvös University
D. Fogaras, B. Réka, Hungarian Academy of Sciences and Budapest University of Technology and Economics

Session 14: E-Learning & Scientific Applications
Chair: W. Nejdl, L3S and University of Hannover
307 Addressing the Testing Challenge with a Web-Based E-Assessment System that Tutors as it Assesses
M. Feng, N. T. Heffeman, Worcester Polytechnic Institute
K. R. Koedinger, Carnegie Mellon University

317 Knowledge Modeling and its Application in Life Sciences: A Tale of Two Ontologies
S. S. Sahoo, C. Thomas, A. Sheth, W. S. York, S. Tartir, University of Georgia

327 Reappraising Cognitive Styles in Adaptive Web Applications
E. Brown, T. Brailsford, T. Fisher, A. Moore, H. Ashman, University of Nottingham

Session 15: High Availability & Performance
Chair: M. Ahamad, Georgia Institute of Technology
337 Cat and Mouse: Content Delivery Tradeoffs in Web Access
B. Krishnamurthy, AT&T Labs–Research
C. E. Wills, Worcester Polytechnic Institute

347 WAP5: Black-box Performance Debugging for Wide-Area Systems
P. Reynolds, Duke University
A. Vahdat, University of California at San Diego

357 WS-Replication: A Framework for Highly Available Web Services
J. Salas, F. Pérez-Sorrosal, M. Patiño-Martínez, R. Jiménez-Péris, Universidad Politécnica de Madrid

Session 16: Web Mining with Search Engines
Chair: S. Dumais, Microsoft Research
367 Random Sampling from a Search Engine's Index
Z. Bar-Yossef, M. Gurevich, Technion

377 A Web-based Kernel Function for Measuring the Similarity of Short Text Snippets
M. Sahami, T. D. Heilman, Google Inc.

387 Generating Query Substitutions
R. Jones, B. Rey, O. Madani, Yahoo! Research
W. Greiner, Los Angeles Software Inc.

Session 17: Social Networks
Chair: C. Goble, University of Manchester
397 POLYPHONET: An Advanced Social Network Extraction System from the Web
Y. Matsuo, National Institute of Advanced Industrial Science & Technology
J. Mori, University of Tokyo
M. Hamasaki, K. Ishida, National Institute of Advanced Industrial Science and Technology
T. Nishimura, AIST
H. Takeda, K. Hasida, M. Ishizuka, University of Tokyo
Table of Contents

407 Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection
B. Aleman-Meza, M. Nagarajan, C. Ramakrishnan, University of Georgia
L. Ding, P. Kolari, University of Maryland
A. P. Sheh, I. B. Aspinar, University of Georgia
A. Joshi, T. Finin, University of Maryland

417 Exploring Social Annotations for the Semantic Web
X. Wu, Shanghai Jiao Tong University
L. Zhang, IBM China Research Laboratory
Y. Yu, Shanghai Jiao Tong University

Session 18: Web Engineering: Validation
Chair: M. Gaedke, Universität Karlsruhe
427 Relaxed—on the Way Towards True Validation of Compound Documents
J. Kosek, P. Nálevka, University of Economics, Prague

437 Model-based Version and Configuration Management for a Web Engineering Lifecycle
T. N. Nguyen, Iowa State University

447 Model-directed Web Transactions under Constrained Modalities
Z. Sun, I. Mahmud, S. Mukherjee, I. V. Ramakrishnan, Stony Brook University

Session 19: New Search Paradigms
Chair: R. Kumar, Yahoo!, Inc.
457 Retroactive Answering of Search Queries
B. Yang, G. Jeh, Google, Inc.

467 CWS: A Comparative Web Search System
J.-T. Sun, Microsoft Research Asia
X. Wang, University of Illinois at Urbana-Champaign
D. Shen, Hong Kong University of Science and Technology
H.-J. Zeng, Z. Chen, Microsoft Research Asia

477 Searching with Context
R. Kraft, C. C. Chang, F. Maghoul, R. Kumar, Yahoo! Inc.

Keynotes
487 Keynote
R. Granger, NHS IT

489 Broken Links on the Web: Local Laws and the Global Free Flow of Information
D. Weitzner, W3C/Massachusetts Institute of Technology

Session 20: Semantic Web: Ontology Construction
Chair: P. Patel-Schneider, Bell Labs Research
491 Position Paper: Ontology Construction from Online Ontologies
H. Alani, University of Southampton

497 Position Paper: Towards the notion of gloss, and the adoption of linguistic resources in formal ontology engineering
M. Jarrar, Vrije Universiteit Brussel

505 Bootstrapping Semantics on the Web: Meaning Elicitation from Schemas
P. Bouquet, University of Trento
L. Serfini, Istituto per la Ricerca Scientifica e Tecnologica
S. Zanobini, University of Trento
S. Sceffer, Istituto per la Ricerca Scientifica e Tecnologica

Session 21: Security, Privacy & Ethics
Chair: A. Juels, RSA Laboratories
513 Designing Ethical Phishing Experiments: A study of (ROT13) onl features
M. Jakobsson, J. Ratkiewicz, Indiana University

523 Invasive Browser Sniffing and Countermeasures
M. Jakobsson, S. Stamm, Indiana University

Session 22: Data Mining
Chair: M. Sahami, Google Inc.
533 A Probabilistic Approach to Spatiotemporal Theme Pattern Mining on Weblogs
Q. Mei, C. Liu, University of Illinois at Urbana-Champaign
H. Su, Vanderbilt University
CX. Zhai, University of Illinois at Urbana-Champaign

543 Time-Dependent Semantic Similarity Measure of Queries Using Historical Click-Through Data
Q. Zhao, Nanyang Technological University
S. C. H. Hoi, The Chinese University of Hong Kong
T.-Y. Liu, Microsoft Research Asia
S. S. Bhowmick, Nanyang Technological University
M. R. Lyu, The Chinese University of Hong Kong
W.-Y. Ma, Microsoft Research Asia

553 Interactive Wrapper Generation with Minimal User Effort
U. Imak, T. Suel, Polytechnic University

Session 23: Semi-structured Semantic Data
Chair: S. Bechhofer, University of Manchester
565 Towards Content Trust of Web Resources
Y. Gil, D. Artz, University of Southern California

575 Supporting Online Problem-Solving Communities with the Semantic Web
A. Ankolekar, University of Karlsruhe
K. Sycara, J. Herbsleb, R. Kraut, Carnegie Mellon University
C. Welty, IBM Watson Research Center

585 Semantic Wikipedia
M. Völkel, M. Krötzsch, D. Vrandecic, H. Haller, R. Studer, University of Karlsruhe
## Table of Contents

### Session 24: Performance, Reliability & Scalability
Chair: C. Wills, Worcester Polytechnic Institute

<table>
<thead>
<tr>
<th>Paper Number</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>605</td>
<td>Selective Early Request Termination for Busy Internet Services</td>
<td>J. Zhou, T. Yang, Ask.com and University of California at Santa Barbara</td>
</tr>
<tr>
<td>615</td>
<td>SCTP: An innovative transport layer protocol for the web</td>
<td>P. Natarajan, J. R. Iyenger, P. D. Amer, University of Delaware</td>
</tr>
</tbody>
</table>

### Session 25: Data Mining Classification
Chair: S. Chakrabarti, IIT Bombay

<table>
<thead>
<tr>
<th>Paper Number</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>625</td>
<td>Improved Annotation of the Blogosphere via Autotagging and Hierarchical Clustering</td>
<td>C. H. Brooks, N. Montanez, University of San Francisco</td>
</tr>
<tr>
<td>633</td>
<td>Large-Scale Text Categorization by Batch Mode Active Learning</td>
<td>S. C. H. Hoi, The Chinese University of Hong Kong</td>
</tr>
<tr>
<td>643</td>
<td>A Comparison of Implicit and Explicit Links for Web Page Classification</td>
<td>D. Shen, Hong Kong University of Science and Technology</td>
</tr>
</tbody>
</table>

### Session 26: E-Commerce & E-Government
Chair: M. Manasse, Microsoft Research

<table>
<thead>
<tr>
<th>Paper Number</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>651</td>
<td>An e-Market Framework for Informed Trading</td>
<td>J. Debenham, S. Simoff, University of Technology, Sydney</td>
</tr>
<tr>
<td>659</td>
<td>The Impact of Online Music Services on the Demand for Stars in the Music Industry</td>
<td>I. P. Volz, Johann Wolfgang Goethe University</td>
</tr>
<tr>
<td>669</td>
<td>The Web Structure of E-Government-Developing a Methodology for Quantitative Evaluation</td>
<td>V. Petricek, UCL Computer Science</td>
</tr>
</tbody>
</table>

### Session 27: XML & Web Services
Chair: M. Mecella, Università di Roma "La Sapienza"

<table>
<thead>
<tr>
<th>Paper Number</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>H. S. Thompson, World Wide Web Consortium and University of Edinburgh</td>
</tr>
</tbody>
</table>

### Session 28: Improved Search Ranking
Chair: N. Craswell, Microsoft Research Cambridge

<table>
<thead>
<tr>
<th>Paper Number</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>707</td>
<td>Beyond PageRank: Machine Learning for Static Ranking</td>
<td>M. Richardson, A. Prakash, E. Brill, Microsoft Research</td>
</tr>
<tr>
<td>717</td>
<td>Optimizing Scoring Functions and Indexes for Proximity Search in Type-annotated Corpora</td>
<td>S. Chakrabarti, K. Puniyani, S. Das, IIT Bombay</td>
</tr>
<tr>
<td>727</td>
<td>Automatic Identification of User Interest for Personalized Search</td>
<td>F. Qui, J. Cho, University of California at Los Angeles</td>
</tr>
</tbody>
</table>

### Panel

<table>
<thead>
<tr>
<th>Paper Number</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>745</td>
<td>Meaning on the Web: Evolution vs Intelligent Design?</td>
<td>R. Brachman, Yahoo!</td>
</tr>
<tr>
<td>747</td>
<td>Identity Management on Converged Networks: A Reality Check</td>
<td>A. Sahuguet, Google Inc.</td>
</tr>
<tr>
<td>749</td>
<td>Phoiling Phishing</td>
<td>R. Dhamija, Harvard University</td>
</tr>
<tr>
<td>750</td>
<td>The Next Wave of the Web</td>
<td>N. Shadbolt, T. Berners-Lee, J. Hendler, C. Hart, R. Benjamins</td>
</tr>
</tbody>
</table>

### Special Interest Tracks and Posters

### Session 4: XML
Chair: M. Solanki, De Montfort University

<table>
<thead>
<tr>
<th>Paper Number</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>751</td>
<td>Compressing and Searching XML Data Via Two Zips</td>
<td>P. Ferragina, F. Luccio, University of Pisa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G. Manzini, University of Piamonte Orientale</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S. Muthukrishnan, Rutgers University</td>
</tr>
</tbody>
</table>
# Table of Contents

## Session 5: Developing Regions & Peer-to-Peer

Chair: K. Toyama, *Microsoft Research India*

<table>
<thead>
<tr>
<th>Session Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>761 Wake-on-WLAN</td>
<td>N. Mishra, K. Chebrolu, B. Raman, A. Pathak, <em>Indian Institute of Technology</em></td>
</tr>
</tbody>
</table>

## Session 10: Developing Regions 2

Chair: K. Toyama, *Microsoft Research India*

<table>
<thead>
<tr>
<th>Session Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>791 Designing an Architecture for Delivering Mobile Information Services to the Rural Developing World</td>
<td>T. S. Parikh, E. D. Lazowska, <em>University of Washington</em></td>
</tr>
<tr>
<td>801 WebKhoj: Indian Language IR from Multiple Character Encodings</td>
<td>P. Pingali, J. Jagarlamudi, V. Varma, <em>International Institute of Information Technology</em></td>
</tr>
</tbody>
</table>

## Session 11: Industrial Practice & Experience

Chair: M. Najork, *Microsoft Research*

<table>
<thead>
<tr>
<th>Session Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>819 Detecting Semantic Cloaking on the Web</td>
<td>B. Wu, B. D. Davison, <em>Lehigh University</em></td>
</tr>
</tbody>
</table>

## Posters

<table>
<thead>
<tr>
<th>Poster Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Rules for Mobile Web Personalization</td>
<td>M. Halvey, M. T. Keane, B. Smyth, <em>University College Dublin</em></td>
</tr>
<tr>
<td>Behavior-Based Web Page Evaluation</td>
<td>G. Velayathan, S. Yamada, <em>National Institute of Informatics</em></td>
</tr>
<tr>
<td>Using Web Browser Interactions to Predict Task</td>
<td>M. Kellar, C. Watters, <em>Dalhousie University</em></td>
</tr>
<tr>
<td>DemIL: an Online Interaction Language between Citizen and Government</td>
<td>C. Maciel, A. C. B. Garcia, <em>Universidade Federal Fluminense</em></td>
</tr>
<tr>
<td>Determining User Interests About Museum Collections</td>
<td>L. Rutledge, <em>Telematica Instituut</em> &amp; L. Aroyo, N. Stash, <em>Technische Universität</em></td>
</tr>
<tr>
<td>Graphical Representation of RDF Queries</td>
<td>A. Harth, S. R. Kruk, S. Decker, <em>National University of Ireland</em></td>
</tr>
<tr>
<td>Question Answering on Top of the BT Digital Library</td>
<td>P. Cimiano, P. Haase, Y. Sure, J. Völker, Y. Wang, <em>University of Karlsruhe</em></td>
</tr>
<tr>
<td>XPath Filename Expansion in a Unix Shell</td>
<td>K. Giger, E. Wilde, <em>ETH Zürich</em></td>
</tr>
<tr>
<td>SGSDesigner, a Graphical Interface for Annotating and Debugging Semantic Grid Services</td>
<td>A. Gómez-Pérez, R. González-Cabero, <em>Ontology Engineering Group</em></td>
</tr>
<tr>
<td>Status of the African Web</td>
<td>R. C. Caminero, <em>Nagaoka University of Technology</em> &amp; P. Zavarsky, <em>Concordia University</em> &amp; Y. Mikami, <em>Nagaoka University of Technology</em></td>
</tr>
</tbody>
</table>
875 Verifying Genre-based Clustering Approach to Content Extraction
S. Gupta, H. Becker, G. Kaiser, S. Stolfo, Columbia University

877 A Browser for Browsing the Past Web
A. Jatowt, National Institute of Information and Communications Technology and Kyoto University
Y. Kawai, S. Nakamura, Y. Kidawara, National Institute of Information and Communications Technology
K. Tanaka, National Institute of Information and Communications Technology and Kyoto University

879 Live URLs: Breathing Life into URLs
N. Kannan, T. Hussain, Infosys Technologies Limited

881 Structuring Namespace Descriptions
E. Wilde, ETH Zürich

H. Li, I. Councill, W.-C. Lee, C. L. Giles, The Pennsylvania State University

885 Tables and Trees Don’t Mix (very well)
E. Wilde, ETH Zürich

887 Robust Web Content Extraction
M. Kowalkiewicz, The Poznan University of Economics
M. E. Orlowska, The University of Queensland
T. Kaczmarek, W. Abramowicz, The Poznan University of Economics

889 Rapid Prototyping of Web Applications Combining Domain Specific Languages and Model Driven Design
D. A. Nunes, D. Schwabe, PUC-Rio

891 A Pruning-based Approach for Supporting Top-K Join Queries
J. Liu, L. Feng, Y. Xing, Chinese Academy of Sciences

893 Towards DSL-based Web Engineering
M. Nussbaumer, P. Freudenstein, M. Gaedke, University of Karlsruhe

895 Capturing the Essentials of Federated Systems
J. Meinecke, M. Gaedke, F. Majer, University of Karlsruhe
A. Brändle, Microsoft Research Cambridge

897 From Adaptation Engineering to Aspect-Oriented Context-Dependency
S. Casteleyn, Vrije Universiteit Brussel
Z. Fiala, Dresden University of Technology
G.-J. Houben, Vrije Universiteit Brussel and Technische Universiteit Eindhoven
K. van der Sluijs, Technische Universiteit Eindhoven

899 Living the TV Revolution: Unite MHP to the Web or Face IDTV Irrelevance!
S. Ferretti, M. Roccetti, J. Andrich, University of Bologna

901 Using Graph Matching Techniques to Wrap Data from PDF Documents
T. Hassan, R. Baumgartner, Vienna University of Technology

903 Requirements for Multimedia Document Enrichment
A. Chakravarthy, V. Lanfranchi, F. Ciravegna, University of Sheffield

905 DiTaBBu — Automating the Production of Time-based Hypermedia Content
R. Lopes, L. Carriço, C. Duarte, LaSIGE/University of Lisbon

907 Capturing RIA Concepts in a Web Modeling Language
A. Bozzon, S. Comai, P. Fraternali, G. T. Carughii, Politecnico di Milano

909 Generation of Multimedia TV News Contents for WWW
H. C. Fu, Y. Y. Xu, C. L. Tseng, National Chiao-Tung University

911 Proposal of Integrated Search Engine of Web and TV Contents
H. Miyamori, NICT
Z. Stejic, T. Araki, Ricoh Co., Ltd.
M. Minakuchi, Q. Ma, NICT
K. Tanaka, NICT and Kyoto University

913 Using Semantic Rules to Determine Access Control for Web Services
B. Shields, O. Molloy, G. Lyons, J. Duggan, National University of Ireland

915 Strong Authentication in Web Proxies
D. Rotiroti, ASPNET S.c.p.a.

917 Safeguard against Unicode Attacks: Generation and Applications of UC-SimList
A. Y. Fu, W. Zhang, X. Deng, L. Wenyin, City University of Hong Kong

919 Efficient Edge-Services for Colorblind Users
G. Iaccarino, D. Malandrino, M. Del Percio, V. Scarano, Università di Salerno

921 A User Profile-based Approach for Personal Information Access: Shaping Your Information Portfolio
L. K. Kan, X. Peng, I. King, The Chinese University of Hong Kong

923 Finding Visual Concepts by Web Image Mining
K. Yanai, The University of Electro-Communications
C. Castillo, Università di Roma "La Sapienza"

925 Deriving Wishlists from Blogs Show Us Your Blog: and We’ll Tell You What Books to Buy
G. Mishne, M. de Rijke, University of Amsterdam

927 Relationship Between Web Links and Trade
R. Baeza-Yates, Yahoo! Research Barcelona
C. Castillo, Università di Roma "La Sapienza"

929 System for Spatio-Temporal Analysis of Online News and Blogs
A. Dalli, University of Sheffield

931 Extracting News-Related Queries from Web Query Log
M. Maslov, A. Golovko, I. Segalovich, P. Braslavski, Yandex

933 Visually Guided Bottom-Up Table Detection and Segmentation in Web Documents
B. Kräpl, M. Herzog, Vienna University of Technology
935 Generating Maps of Web Pages using Cellular Automata
H. Azzag, Polytech'Tours
D. Ratsimba, Université de Lyon 2
Da Costa, G. Venturini, Polytech 'Tours
C. Guinot, C.E.R.I.E.S.

937 BuzzRank . . . And the Trend is Your Friend
K. Berberich, S. Bedathur, Max-Planck-Institut für Informatik
M. Vaziringinis, Athens University of Economics and Business
G. Weikum, Max-Planck-Institut für Informatik

939 Detecting Nepotistic Links by Language Model Disagreement
A. A. Benczár, I. Bíró, K. Csalogány, M. Uher, Hungarian Academy of Sciences and Eötvös University

941 The Distribution of PageRank Follows a Power-Law only for Particular Values of the Damping Factor
L. Becchetti, C. Castillo, Universita di Roma “La Sapienza”

943 Mining Related Queries from Search Engine Query Logs
X. Shi, C. C. Yang, The Chinese University of Hong Kong

945 Discovering Event Evolution Graphs from Newsfeeds
C. C. Yang, X. Shi, The Chinese University of Hong Kong

947 Mining Clickthrough Data for Collaborative Web Search
J.-T. Sun, Microsoft Research Asia
X. Wang, University of Illinois at Urbana-Champaign
D. Shen, Hong Kong University of Science and Technology
H.-J. Zeng, Z. Chen, Microsoft Research Asia

949 Background Knowledge for Ontology Construction
B. Fortuna, M. Grobelnik, D. Mladenić, Institute Jožef Stefan

951 Mining RDF Metadata for Generalized Association Rules: Knowledge Discovery in the Semantic Web Era
T. Jiang, A.-H. Tan, Nanyang Technological University

953 AutoTag: A Collaborative Approach to Automated Tag Assignment for Weblog Posts
G. Mishne, University of Amsterdam

955 Merging Trees: File System and Content Integration
E. Wilde, ETH Zürich

957 A Content and Structure Website Mining Model
B. Poblete, R. Baeza-Yates, University of Pompeu Fabra

959 Online Mining of Frequent Query Trees over XML Data Streams
H.-F. Li, National Chiao-Tung University
M.-K. Shan, National Chengchi University
S.-Y. Lee, National Chiao-Tung University

961 Using Proportional Transportation Similarity with Learned Element Semantics for XML Document Clustering
X. Wan, J. Yang, Peking University

963 Template Guided Association Rule Mining from XML Documents
R. AliMohmmadzadeh, S. Soltan, M. Rahgozar, University of Tehran

965 Automatic Geotagging of Russian Web Sites
A. Pyalling, M. Maslov, P. Braslavski, Yandex

967 Using Symbolic Objects to Cluster Web Documents
E. Meneses, Costa Rica Institute of Technology
O. Rodríguez-Rojas, University of Costa Rica

969 Estimating Required Recall for Successful Knowledge Acquisition from the Web
W. Gatterbauer, Vienna University of Technology

971 Text-based Video Blogging
N. Hamaguchi, M. Doke, M. Hayashi, N. Yagi, Japan Broadcasting Corporation

973 A Decentralized CF Approach Based on Cooperative Agents
B. M. Kim, Kunoh National Institute of Technology
Q. Li, Information & Communications University
A. E. Howe, Colorado State University

975 Adaptive Web Sites: User Studies and Simulation
D. Warner, S. D. Durbin, J. N. Richter, Z. Gedeon, RightNow Technologies

N/A On a Service-Oriented Approach for an Engineering Knowledge Desktop
S. C. Wong, R. M. Crowder, G. B. Wills, University of Southampton

979 Design and Development of Learning Management System at Universiti Putra Malaysia: A Case Study of e-SPRINT
S. H. A. Aziz, A. S. M. Yunus, K. A. Bakar, H. B. Meseran, Universiti Putra Malaysia

981 Providing SCORM with Adaptivity
M. Rey-López, A. Fernández-Vilas, R. Díaz-Redondo, J. Pazos-Arias, University of Vigo

983 A Framework for XML Data Streams History Checking and Monitoring
A. Campi, P. Spolletini, Politecnico di Milano

N/A The Credibility of the Posted Information in a Recommendation System Based on a Map
K. Yamamoto, D. Katagami, K. Nitta, Tokyo Institute of Technology
A. Aiba, H. Kuwata, Shibaura Institute of Technology

987 Archiving Web Site Resources: A Records Management View
M. Pennock, B. Kelly, University of Bath

989 Geographic Locations of Web Servers under African Domains
K. T. Nakahira, T. Hoshino, Y. Mikami, Nagaoka University of Technology

991 Why is Connectivity in Developing Regions Expensive: Policy Challenges More than Technical Limitations?
R. Tongia, Carnegie Mellon University

N/A Bilingual Web Page and Site Readability Assessment
T. P. Lau, I. King, The Chinese University of Hong Kong
Table of Contents

995 Mobile Web Publishing and Surfing Based on Environmental Sensing Data
D. Morikawa, M. Honjo, S. Nishiyama, M. Ohashi, KDDI Corporation

997 DoNet: A Semantic Domotic Framework
M. Attard, M. Montebello, University of Malta

999 Web Based Device Independent Mobile Map Applications. The m-HARTIS System
J. Garofalakis, T.-A. Michail, A. Plessas, University of Patras

N/A Context-Orientated News Filtering for Web 2.0 and Beyond
D. Webster, The University of Hull
W. Huang, Kingston University
D. Mundy, P. Warren, The University of Hull

1003 Efficient Search for Peer-to-Peer Information Retrieval Using Semantic Small World
H. Jin, X. Ning, H. Cten, Huazhong University of Science and Technology

1005 Semantic Link Based Top-K Join Queries in P2P Networks
J. Liu, L. Feng, C. He, Chinese Academy of Sciences

1007 Ontology-Based Legal Information Retrieval to Improve the Information Access in e-Government
A. Gómez-Pérez, F. Ortiz-Rodriguez, B. Villazón-Terrazas, Politecnico di Milano

N/A Oyster – Sharing and Re-using Ontologies in a Peer-to-Peer Community
R. Palma, Universidad Politécnica de Madrid
P. Haase, University of Karlsruhe
A. Gómez-Pérez, Universidad Politécnica de Madrid

1011 GoGetIt!: A Tool for Generating Structure-Driven Web Crawlers
M. L. A. Vidal, A. S. da Silva, E. S. de Moura, J. M. B. Cavalcanti, Universidade Federal do Amazonas

1013 Towards Practical Genre Classification of Web Documents
G. Ferizis, P. Bailey, CSIRO ICT Centre

1015 Do not Crawl in the DUST: Different URLs with Similar Text Extended Abstract
U. Schonfeld, Z. Bar-Yossef, I. Keidar, Technion

N/A Community Discovery and Analysis in Blogspace
Y. Zhou, J. Davis, The University of Sydney

1019 PageSim: A Novel Link-based Measure of Web Similarity
Z. Lin, M. R. Lyu, I. King, The Chinese University of Hong Kong

1021 Finding Specification Pages According to Attributes
N. Yoshinaga, Japan Society for the Promotion of Science and Japan Advanced Institute of Science and Technology
K. Torisawa, Japan Advanced Institute of Science and Technology

1023 Selective Hypertext Induced Topic Search
A. C. Awekar, North Carolina State University
P. Mitra, Indian Institute of Technology
J. Kang, North Carolina State University

N/A An Audio/Video Analysis Mechanism for Web Indexing
M. Furini, M. Aragone, University of Piemonte Orientale

1027 The SOWES Approach to P2P Web Search Using Semantic Overlays
C. Doukeridis, AUEB
K. Nørvang, NTNU
M. Vazirgiannis, AUEB

1029 Topic-Oriented Query Expansion for Web Search
S.-C. Wang, Y. Tanaka, Hokkaido University

1031 Predictive Modeling of First-Click Behavior in Web-Search
M. O’Brien, M. T. Keane, B. Smyth, University of College Dublin

N/A Proximity Within Paragraph: A Measure to Enhance Document Retrieval Performance
S. Palakvangna-Na-Ayudhya, J. A. Keane, The University of Manchester

1035 Finding Experts and Their Details in E-mail Corpora
K. Balog, M. de Rijke, University of Amsterdam

1037 Efficient Query Subscription Processing for Prospective Search Engines
U. Irmak, Polytechnic University
S. Milbaylov, University of Pennsylvania
T. Suel, Polytechnic University
S. Ganguly, R. Iznalov, NEC Laboratories America, Inc.

1039 Mining Search Engine Query Logs for Query Recommendation
Z. Zhang, O. Nasraoui, University of Louisville

1041 Effective Web-Scale Crawling through Website Analysis
I. Gonzalez, Carnegie Mellon University
A. Marcus, Rensselaer Polytechnic Institute
D. N. Meredith, L. A. Nguyen, IBM Almaden Research Center

N/A Focused Crawling: Experiences in a Real World Project
A. Badia, T. Muezzinoglu, O. Nasraoui, University of Louisville

1045 Image Annotation Using Search and Mining Technologies
X.-J. Wang, L. Zhang, F. Jing, W.-Y. Ma, Microsoft Research Asia

1047 Semantic Web Integration of Cultural Heritage Sources
P. Sinclair, P. Lewis, K. Martinez, University of Southampton
M. Addis, D. Prideaux, IT Innovation Centre

1049 The ODESeW 2.0 Semantic Web Application Framework
O. Corcho, University of Manchester
A. López-Cima, A. Gómez-Pérez, Universidad Politécnica de Madrid

1051 Visualizing An Historical Semantic Web with Heml
B. G. Robertson, Mount Allison University
Table of Contents

1053 Beyond XML and RDF: The Versatile Web Query Language Xcerpt
   S. Berger, F. Bry, T. Furche, B. Linse, A. Schroeder,
   University of Munich

1055 An Ontology for Internal and External Business Processes
   A. Haller, E. Oren, Digital Enterprise Research Institute
   P. Kotinurmi, Helsinki University of Technology

1057 Automatic Matchmaking of Web Services
   S. Agarwal, A. Ankolekar, University of Karlsruhe

1059 Adding Semantics to RosettaNet Specifications
   P. Kotinurmi, Helsinki University of Technology
   T. Vitvar, Digital Enterprise Research Institute

1061 Automatic Annotation of Content Units in TANGRAM
   J. Jovanović, University of Belgrade
   D. Gasević, Simon Fraser University Surrey
   V. Devedžić, University of Belgrade

1063 Logical Structure Based Semantic Relationship Extraction from Semi-Structured Documents
   Z. Kuo, W. Gang, L. JuanZi, Tsinghua University

1065 OWL FA: A Metamodeling Extension of OWL DL
   J. Z. Pan, University of Aberdeen
   I. Horrocks, University of Manchester

1067 Learning and Inferencing in User Ontology for Personalized Semantic Web Services
   X. Jiang, A.-H. Tan, Nanyang Technological University

1069 Upgrading Relational Legacy Data to the Semantic Web
   J. B. Rodríguez, A. Gómez-Pérez, Ontology Engineering Group

1071 How Semantics Make Better Wikis
   E. Oren, J. G. Breslin, S. Decker, National University of Ireland

1073 Integrating Ecoinformatics Resources on the Semantic Web
   C. S. Parr, A. Parafiyanyk, J. Sachs, L. Ding, S. Dombush,
   T. Finin, D. Wang, University of Maryland
   A. Hollander, University of California at Davis

1075 HTML2RSS: Automatic Generation of RSS Feed Based on Structure Analysis of HTML Document
   T. Nanno, M. Okumura, Tokyo Institute of Technology

1077 Path Summaries and Path Partitioning in Modern XML Databases
   A. Arion, INRIA Futurs
   A. Bonifati, ICAR CNR
   I. Manolescu, INRIA Futurs
   A. Pugliese, University of Calabria

1079 Evaluating Structural Summaries as Access Methods for XML
   M. M. Moro, University of California at Riverside
   Z. Vagena, IBM Almaden Research Center
   V. J. Tsotras, University of California at Riverside

1081 FLUX: Fuzzy Content and Structure Matching of XML Range Queries
   H.-G. Li, S. A. Aghili, D. Agrawal, A. El Abbadi, University of California at Santa Barbara

1083 Author Index