

# **INFORMATION PROCESSING & MANAGEMENT**

**An International Journal**

Editor-in-Chief: **Tefko Saracevic**

Email: [tefko@scils.rutgers.edu](mailto:tefko@scils.rutgers.edu)

Home page: <http://www.scils.rutgers.edu/~tefko>

*Rutgers University*

Volume 43 (2007) Contents and Author Index



# INFORMATION PROCESSING & MANAGEMENT

—An International Journal—  
(Incorporating *INFORMATION TECHNOLOGY*)

Email: [ipm@scils.rutgers.edu](mailto:ipm@scils.rutgers.edu)

Website: <http://www.elsevier.com/locate/infoproman/>

## Editor-in-Chief

### Tefko Saracevic

School of Communication, Information and Library Studies  
Rutgers University  
4 Huntington Street, New Brunswick, NJ 08901-1071, U.S.A.  
[tefko@scils.rutgers.edu](mailto:tefko@scils.rutgers.edu)  
<http://www.scils.rutgers.edu/~tefko>

## Associate Editor (Europe)

### C. J. van Rijsbergen

Department of Computing Science  
University of Glasgow  
8-17 Lilybank Gardens  
Glasgow, G12 8QQ Scotland, U.K.  
[keith@dcs.gla.ac.uk](mailto:keith@dcs.gla.ac.uk)

## Associate Editors

### Nicholas J. Belkin

School of Communication, Information and Library Studies  
Rutgers University  
4 Huntington Street  
New Brunswick, NJ 08901-1071, U.S.A.  
[nick@belkin.rutgers.edu](mailto:nick@belkin.rutgers.edu)

## Associate Editor (Book Reviews)

### Bernard J. Jansen

School of Information Science  
The Pennsylvania State University  
University Park, PA 16803, U.S.A.  
[jjansen@acm.org](mailto:jjansen@acm.org)

### W. Bruce Croft

Department of Computer and Information Science  
University of Massachusetts  
Lederle Graduate Research Center  
Amherst, MA 01003-4610, U.S.A.  
[croft@cs.umass.edu](mailto:croft@cs.umass.edu)

## Founding Editor

### Harold Borko

Graduate School of Education and Information Studies  
University of California  
102 South Hall  
Los Angeles, CA 90024-1520, U.S.A.  
[hborko@ucla.edu](mailto:hborko@ucla.edu)

## Editorial Board

### Maristella Agosti

Padova University, Padova, Italy  
[agosti@dei.unipd.it](mailto:agosti@dei.unipd.it)

### Tatiana Aparac

University of Osijek  
Zagreb, Croatia  
[taparac@pesos.hr](mailto:taparac@pesos.hr)

### Ricardo Baeza-Yates

Universitat Pompeu Fabra  
Barcelona, Spain  
[rbaeza@upf.edu](mailto:rbaeza@upf.edu)

### David C. Blair

University of Michigan  
Ann Arbor, MI, U.S.A.  
[dcl Blair@mac.com](mailto:dcl Blair@mac.com)

### Abraham Bookstein

University of Chicago  
Chicago, IL, U.S.A.  
[a-bookstein@uchicago.edu](mailto:a-bookstein@uchicago.edu)

### Pia Borlund

The Royal School of Library and Information Science  
Aalborg, Denmark  
[pbj@db.dk](mailto:pbj@db.dk)

### Michael Buckland

University of California  
Berkeley, CA, U.S.A.  
[buckland@sims.berkeley.edu](mailto:buckland@sims.berkeley.edu)

### Guoqing Chen

Tsinghua University  
Beijing, China  
[chengq@em.tsinghua.edu.cn](mailto:chengq@em.tsinghua.edu.cn)

### Yves Chiaramella

University of Grenoble  
Grenoble Cedex, France  
[chiara@imag.fr](mailto:chiara@imag.fr)

### Colleen Cool

Queens College  
Flushing, NY, USA  
[ccool@qc.edu](mailto:ccool@qc.edu)

### Fabio Crestani

University of Strathclyde  
Scotland, UK  
[F.Crestani@cis.strath.ac.uk](mailto:F.Crestani@cis.strath.ac.uk)

### J. Stephen Downie

University of Illinois at Urbana-Champaign  
Champaign, IL, U.S.A.  
[jdownie@uiuc.edu](mailto:jdownie@uiuc.edu)

### Susan Dumais

Microsoft Research  
Redmond, WA, U.S.A.  
[sdumais@microsoft.com](mailto:sdumais@microsoft.com)

### Raya Fidel

University of Washington  
Seattle, WA, U.S.A.  
[fidelr@u.washington.edu](mailto:fidelr@u.washington.edu)

### Edward A. Fox

Virginia Polytechnic Institute & State University  
Blacksburg, VA, U.S.A.  
[fox@cs.vt.edu](mailto:fox@cs.vt.edu)

### Jonathan Furner

University of California  
Los Angeles, CA, U.S.A.  
[jfurner@ucla.edu](mailto:jfurner@ucla.edu)

### Donna Harman

National Institute of Standards & Technology  
Gaithersburg, MD, U.S.A.  
[donna.harman@nist.gov](mailto:donna.harman@nist.gov)

### David J. Harper

The Robert Gordon University  
Aberdeen, U.K.  
[david.harper@smartweb.rgu.ac.uk](mailto:david.harper@smartweb.rgu.ac.uk)

### William Hersh

Oregon Health Sciences University  
Portland, OR, U.S.A.  
[hersh@ohsu.edu](mailto:hersh@ohsu.edu)

### Peter Ingwersen

The Royal School of Library and Information Science  
Copenhagen, Denmark  
[pi@db.dk](mailto:pi@db.dk)

### Tetsuya Ishikawa

University of Tokyo, HI  
Tokyo, Japan  
[ishikawa@hi.u-tokyo.ac.jp](mailto:ishikawa@hi.u-tokyo.ac.jp)

### Kalervo Järvelin

University of Tampere, Tampere, Finland  
[kalervo.jarvelin@uta.fi](mailto:kalervo.jarvelin@uta.fi)

### Noriko Kando

National Institute of Informatics  
Tokyo, Japan  
[kando@nii.ac.jp](mailto:kando@nii.ac.jp)

### Paul B. Kantor

Rutgers University  
New Brunswick, NJ, U.S.A.  
[kantor@scils.rutgers.edu](mailto:kantor@scils.rutgers.edu)

### Diane Kelly

University of North Carolina  
Chapel Hill, NC, USA  
[dianek@email.unc.edu](mailto:dianek@email.unc.edu)

### Mounia Lalmas

Queen Mary, University of London  
London, UK  
[mounia.lalmas@dcs.qmul.ac.uk](mailto:mounia.lalmas@dcs.qmul.ac.uk)

### Elisabeth D. Liddy

Syracuse University, Syracuse, NY, U.S.A.  
[liddy@mailbox.syr.edu](mailto:liddy@mailbox.syr.edu)

### Gary Marchionini

University of North Carolina  
Chapel Hill, NC, U.S.A.  
[march@ils.unc.edu](mailto:march@ils.unc.edu)

### Michel J. Menou

Consultant in Information Management  
Les Rosiers sur Loire, France  
[Michel.Menou@wanadoo.fr](mailto:Michel.Menou@wanadoo.fr)

### Sung Hyon Myaeng

Information and Communication University  
Daejeon, Korea  
[myaeng@icu.ac.kr](mailto:myaeng@icu.ac.kr)

### Dong Oard

University of Maryland  
College of Information Studies, MD, U.S.A.  
[oard@glue.umd.edu](mailto:oard@glue.umd.edu)

### Iadh Ounis

University of Glasgow, Glasgow, U.K.  
[ounis@dcs.gla.ac.uk](mailto:ounis@dcs.gla.ac.uk)

### Fausto Rabitti

Consiglio Nazionale delle Ricerche, Pisa, Italy  
[Fausto.Rabitti@isti.cnr.it](mailto:Fausto.Rabitti@isti.cnr.it)

### Edie M. Rasmussen

University of British Columbia  
Vancouver, BC, Canada  
[eddie.rasmussen@ubc.ca](mailto:eddie.rasmussen@ubc.ca)

### Stephen E. Robertson

Microsoft Research Ltd., Cambridge, England, U.K.  
[ser@microsoft.com](mailto:ser@microsoft.com)

### Mark Sanderson

University of Sheffield  
Sheffield, South Yorkshire, U.K.  
[m.sanderson@sheffield.ac.uk](mailto:m.sanderson@sheffield.ac.uk)

### Fabrizio Sebastiani

Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy  
[fabrizio.sebastiani@isti.cnr.it](mailto:fabrizio.sebastiani@isti.cnr.it)

### Alan F. Smeaton

Dublin City University, Dublin, Ireland  
[asmeaton@computing.dcu.ie](mailto:asmeaton@computing.dcu.ie)

### Amanda Spink

Queensland University of Technology  
Brisbane, Australia  
[ah.spink@qut.edu.au](mailto:ah.spink@qut.edu.au)

### Pertti Vakkari

University of Tampere, Tampere, Finland  
[pertti.vakkari@uta.fi](mailto:pertti.vakkari@uta.fi)

### Justin Zobel

RMIT University, Melbourne, Australia  
[jz@cs.rmit.edu.au](mailto:jz@cs.rmit.edu.au)

**Author enquiries:** For enquiries relating to the submission of articles (including electronic submission where available) please visit this journal's homepage at <http://www.elsevier.com/locate/infoproman>. You can track accepted articles at <http://www.elsevier.com/trackarticle> and set up e-mail alerts to inform you of when an article's status has changed. Also accessible from here is information on copyright, frequently asked questions and more.

Contact details for questions arising after acceptance of an article, especially those relating to proofs, are provided after registration of an article for publication

**Publication information:** *Information Processing & Management* (ISSN 0306-4573). For 2007, volume 43 is scheduled for publication. Subscription prices are available upon request from the Publisher or from the Regional Sales Office nearest you or from this journal's website: (<http://www.elsevier.com/locate/infoproman>). Further information is available on this journal and other Elsevier products through Elsevier's website: (<http://www.elsevier.com>). Subscriptions are accepted on a prepaid basis only and are entered on a calendar year basis. Issues are sent by standard mail (surface within Europe, air delivery outside Europe). Priority rates are available upon request. Claims for missing issues should be made within six months of the date of dispatch.

**Orders, claims, and journal enquiries:** please contact the Customer Service Department at the Regional Sales Office nearest you:

**Orlando:** Elsevier, Customer Service Department, 6277 Sea Harbor Drive, Orlando, FL 32887-4800, USA; phone: (877) 8397126 or (800) 6542452 [toll free numbers for US customers]; (+1) (407) 3454020 or (+1) (407) 3454000 [customers outside US]; fax: (+1) (407) 3631354 or (+1) (407) 3639661; e-mail: [usjcs@elsevier.com](mailto:usjcs@elsevier.com) or [elspscs@elsevier.com](mailto:elspscs@elsevier.com)

**Amsterdam:** Elsevier, Customer Service Department, PO Box 211, 1000 AE Amsterdam, The Netherlands; phone: (+31) (20) 4853757; fax: (+31) (20) 4853432; e-mail: [nlinfo-f@elsevier.com](mailto:nlinfo-f@elsevier.com)

**Tokyo:** Elsevier, Customer Service Department, 4F Higashi-Azabu, 1-Chome Bldg, 1-9-15 Higashi-Azabu, Minato-ku, Tokyo 106-0044, Japan; phone: (+81) (3) 5561 5037; fax: (+81) (3) 5561 5047; e-mail: [jp.info@elsevier.com](mailto:jp.info@elsevier.com)

**Singapore:** Elsevier, Customer Service Department, 3 Killiney Road, #08-01 Winsland House I, Singapore 239519; phone: (+65) 6349 0222; fax: (+65) 6733 1510; e-mail: [asiainfo@elsevier.com](mailto:asiainfo@elsevier.com)

© 2007 Elsevier Ltd. All rights reserved.

**USA mailing notice:** *Information Processing & Management* (ISSN 0306-4573) is published in 6 issues, January, March, May, July, September and November by Elsevier Ltd. (The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK). Annual subscription price in the USA US\$ 1656 (valid in North, Central and South America), including air speed delivery. Periodical postage paid at Rahway NJ and additional mailing offices.

**USA POSTMASTER:** Send change of address to *Information Processing & Management*, Elsevier, 6277 Sea Harbor Drive, Orlando, FL 32887-4800.

**AIRFREIGHT AND MAILING** in USA by Mercury International Limited, 365 Blair Road, Avenel, NJ 07001.

# LIST OF CONTENTS

## Number 1

		<i>Editorials</i>
<b>Gary Marchionini</b> <b>Tefko Saracevic</b> <b>Jack Carroll</b> <b>Donald Kraft</b> <b>William Hersh</b> <b>Josiane Mothe</b> <b>Justin Zobel</b> <b>Peter Hernon</b> <b>Candy Schwartz</b>	1	Reviewer merits
<b>Tefko Saracevic</b>	3	Honoring reviewers and reviewing
	4	Honor List of Referees for Information Processing & Management October 2004–May 2006
<b>Anita Komlodi</b> <b>Gary Marchionini</b> <b>Dagobert Soergel</b>	10	Search history support for finding and using information: User interface design recommendations from a user study
<b>Diane Kelly</b> <b>Xin Fu</b>	30	Eliciting better information need descriptions from users of information search systems
<b>Dania Bilal</b> <b>Imad Bachir</b>	47	Children’s interaction with cross-cultural and multi-lingual digital libraries: I. Understanding interface design representations
<b>Dania Bilal</b> <b>Imad Bachir</b>	65	Children’s interaction with cross-cultural and multi-lingual digital libraries. II. Information seeking, success, and affective experience
<b>Per Ahlgren</b> <b>Jaana Kekäläinen</b>	81	Indexing strategies for Swedish full text retrieval under different user scenarios
<b>Kazuaki Kishida</b>	103	Term disambiguation techniques based on target document collection for cross-language information retrieval: An empirical comparison of performance between techniques
<b>Seung-Hoon Na</b> <b>In-Su Kang</b> <b>Jong-Hyeok Lee</b>	121	Parsimonious translation models for information retrieval
<b>Shao Fen Liang</b> <b>Siobhan Devlin</b> <b>John Tait</b>	146	Investigating sentence weighting components for automatic summarisation
<b>Rey-Long Liu</b>	154	Dynamic category profiling for text filtering and classification

<b>Milad Shokouhi</b> <b>Justin Zobel</b> <b>Saied Tahaghoghi</b> <b>Falk Scholer</b>	169	Using query logs to establish vocabularies in distributed information retrieval
<b>Vanessa Murdock</b> <b>Diane Kelly</b> <b>W. Bruce Croft</b> <b>Nicholas J. Belkin</b> <b>Xiaojun Yuan</b>	181	Identifying and improving retrieval for procedural questions
<b>Fidel Cacheda</b> <b>Victor Carneiro</b> <b>Vassilis Plachouras</b> <b>Iadh Ounis</b>	204	Performance analysis of distributed information retrieval architectures using an improved network simulation model
<b>Kyoung-Min Kim</b> <b>Jin-Hyuk Hong</b> <b>Sung-Bae Cho</b>	225	A semantic Bayesian network approach to retrieving information with intelligent conversational agents
<b>Susumu Yata</b> <b>Masaki Oono</b> <b>Kazuhiro Morita</b> <b>Masao Fuketa</b> <b>Toru Sumitomo</b> <b>Jun-ichi Aoe</b>	237	A compact static double-array keeping character codes
<b>Quan Thanh Tho</b> <b>S.C. Hui</b> <b>A.C.M. Fong</b>	248	A citation-based document retrieval system for finding research expertise
<b>L. Egghe</b>	265	Existence theorem of the quadruple $(P, R, F, M)$ : Precision, recall, fallout and miss
<b>E. Sanz-Casado</b> <b>C. Suárez-Balseiro</b> <b>I. Iribarren-Maestro</b> <b>M. Ramírez-de Santa Pau</b> <b>J. de Pedro-Cuesta</b>	273	Bibliometric mapping of scientific research on prion diseases, 1973–2002
<b>Gheorghe Muresan</b>	285	<i>Book reviews</i> TREC – Experiment and Evaluation in Information Retrieval. Ellen M. Voorhees, Donna K. Harman (Eds.), The MIT Press, Cambridge, MA (2005). 462 pp., ISBN: 0-262-22073-3
<b>Madhu Reddy</b> <b>Joanna Abraham</b>	287	Handbook of Evaluation Methods for Health Informatics, Jytte Brender (2005). ISBN: 0-12-370464-2, £69.95
<b>Mohamed Taher</b>	288	Automated Web Site Evaluation: Researchers' and Practitioners' Perspectives (Kluwer International Series on Human Computer Interaction Vol. 4). Melody Y. Ivory, Kluwer Academic Publishers (2003). 200p, ISBN: 1402016727, \$94.00

- Wilfred Ng** 290 An Introduction to Search Engines and Web Navigation, Mark Levene, Addison Wesley Publisher (2006). 392 pp., ISBN: 0321306775, US\$97.5
- Athulang Mutshewa** 292 The Skeptical Business Searcher: The Information Advisors Guide to Evaluating Web, Data, Sites and Sources, Berkman, R. Medford: Information Today, Inc., 281 pp., ISBN: 0-910965-66-8, \$29.95
- Michael Pittarelli** 295 Uncertainty and Information: Foundations of Generalized Information Theory, George J. Klir, Wiley, New York, 2005, 499 pages + xv, ISBN: 0471748676, Cost: \$94.95
- Peng Liu** 296 Information Security – A Strategic Approach, Vincent Leveque. IEEE Computer Society & Wiley Interscience. 263 pp., ISBN: 0471736120, Cost: \$62

## Number 2

## SPECIAL ISSUE

## AIRS2005: Information Retrieval Research in Asia

- Gary Geunbae Lee** 297 *Editorial*  
Special issue on AIRS2005: Information retrieval research in Asia
- Seung-Hoon Na** 302 *IR theories and models: language modeling and document ranking*  
An empirical study of query expansion and cluster-based retrieval in language modeling approach
- In-Su Kang**  
**Ji-Eun Roh**  
**Jong-Hyeok Lee**
- Lingpeng Yang** 315 Document reranking by term distribution and maximal marginal relevance for chinese information retrieval
- Donghong Ji**  
**Munkew Leong**
- June-Jei Kuo** 327 *Text mining issues: classification, clustering, question answering, topic detection and tracking*  
Cross-document event clustering using knowledge mining from co-reference chains
- Hsin-Hsi Chen**
- Francois Paradis** 344 Contextual feature selection for text classification
- Jian-Yun Nie**
- Kyoung-Soo Han** 353 Answer extraction and ranking strategies for definitional question answering using linguistic features and definition terminology
- Young-In Song**  
**Sang-Bum Kim**  
**Hae-Chang Rim**

<b>Yun Jin</b> <b>Sung Hyon Myaeng</b> <b>Yuchul Jung</b>	365	Use of place information for improved event tracking
<b>Akinori Fujino</b> <b>Naonori Ueda</b> <b>Kazumi Saito</b>	379	A hybrid generative/discriminative approach to text classification with additional information
<b>Yongwook Yoon</b> <b>Gary Geunbae Lee</b>	393	Efficient implementation of associative classifiers for document classification
		<i>Web related issues: web mining, web search engine and FAQ retrieval</i>
<b>Kui-Lam Kwok</b> <b>Laszlo Grunfeld</b> <b>Peter Deng</b>	406	Employing web mining and data fusion to improve weak ad hoc retrieval
<b>Harksoo Kim</b> <b>Hyunjung Lee</b> <b>Jungyun Seo</b>	420	A reliable FAQ retrieval system using a query log classification technique based on latent semantic analysis
<b>Wei Lu</b> <b>Min-Yen Kan</b>	431	Supervised categorization of JavaScript™ using program analysis features
<b>Tao Qin</b> <b>Tie-Yan Liu</b> <b>Xu-Dong Zhang</b> <b>Guang Feng</b> <b>De-Sheng Wang</b> <b>Wei-Ying Ma</b>	445	Topic distillation via sub-site retrieval
		<i>Multimedia retrieval: image, video, audio, metadata and music recommendation</i>
<b>Tzu-Chuen Lu</b> <b>Chin-Chen Chang</b>	461	Color image retrieval technique based on color features and image bitmap
<b>Qing Li</b> <b>Sung Hyon Myaeng</b> <b>Byeong Man Kim</b>	473	A probabilistic music recommender considering user opinions and audio features
<b>Wen-Cheng Lin</b> <b>Yih-Chen Chang</b> <b>Hsin-Hsi Chen</b>	488	Integrating textual and visual information for cross-language image retrieval: A trans-media dictionary approach
<b>Seungji Yang</b> <b>Sang Kyun Kim</b> <b>Kyong Sok Seo</b> <b>Yong Man Ro</b> <b>Ji-Yeon Kim</b> <b>Yang Suk Seo</b>	503	Semantic categorization of digital home photo using photographic region templates
<b>Jean-Pierre Chevallet</b> <b>Joo-Hwee Lim</b> <b>Mun-Kew Leong</b>	515	Object identification and retrieval from efficient image matching. Snap2Tell with the STOIC dataset

		<i>Experimentation: test collection and evaluation methodology</i>
<b>Tetsuya Sakai</b>	531	On the reliability of information retrieval metrics based on graded relevance

## Number 3

**SPECIAL ISSUE**  
**Heterogeneous and Distributed IR**

		<i>Guest Editorial</i>
<b>Fabrizio Silvestri</b>	549	Special issue on heterogeneous and distributed IR
		<i>Heterogeneous Data Management</i>
<b>Henrik Nottelmann</b> <b>Umberto Straccia</b>	552	Information retrieval and machine learning for probabilistic schema matching
<b>P. Veltri</b> <b>M. Cannataro</b> <b>G. Tradigo</b>	577	Sharing mass spectrometry data in a grid-based distributed proteomics laboratory
		<i>Distributed IR</i>
<b>C.S. Badue</b> <b>R. Baeza-Yates</b> <b>B. Ribeiro-Neto</b> <b>A. Ziviani</b> <b>N. Ziviani</b>	592	Analyzing imbalance among homogeneous index servers in a web search system
<b>B. Barla Cambazoglu</b> <b>Evren Karaca</b> <b>Tayfun Kucukyilmaz</b> <b>Ata Turk</b> <b>Cevdet Aykanat</b>	609	Architecture of a grid-enabled Web search engine
		<i>Peer-to-Peer IR</i>
<b>Henrik Nottelmann</b> <b>Gudrun Fischer</b>	624	Search and browse services for heterogeneous collections with the peer-to-peer network Pepper
<b>Josiane Xavier Parreira</b> <b>Sebastian Michel</b> <b>Gerhard Weikum</b>	643	p2pDating: Real life inspired semantic overlay networks for Web search
<b>Fabrizio Falchi</b> <b>Claudio Gennaro</b> <b>Pavel Zezula</b>	665	Nearest neighbor search in metric spaces through Content-Addressable Networks
		<i>General Research Papers</i>
<b>Ryen W. White</b> <b>Gary Marchionini</b>	685	Examining the effectiveness of real-time query expansion

<b>Ahmed Abdelali</b> <b>Jim Cowie</b> <b>Hamdy S. Soliman</b>	705	Improving query precision using semantic expansion
<b>Jimmy Lin</b>	717	User simulations for evaluating answers to question series
<b>Zheng-Yu Niu</b> <b>Dong-Hong Ji</b> <b>Chew Lim Tan</b>	730	Using cluster validation criterion to identify optimal feature subset and cluster number for document clustering
<b>Muriel Foulonneau</b>	740	Information redundancy across metadata collections
<b>Aurora Pons-Porrata</b> <b>Rafael Berlanga-Llavori</b> <b>José Ruiz-Shulcloper</b>	752	Topic discovery based on text mining techniques
<b>Joaquín Adiego</b> <b>Gonzalo Navarro</b> <b>Pablo de la Fuente</b>	769	Using structural contexts to compress semistructured text collections
<b>Seikyung Jung</b> <b>Jonathan L. Herlocker</b> <b>Janet Webster</b>	791	Click data as implicit relevance feedback in web search
<b>Carole L. Palmer</b> <b>Melissa H. Cragin</b> <b>Timothy P. Hogan</b>	808	Weak information work in scientific discovery
		<i>Book reviews</i>
<b>James D. Anderson</b>	821	The Turn: Integration of Information Seeking and Retrieval in Context, Peter Ingwersen, Kalervo Järvelin. Springer, Dordrecht, Netherlands (c2005). xiv, 448 p. (The Kluwer International Series on Information Retrieval). \$79.95, ISBN: 13 978-1-4020-3850-1 (HB)
<b>Kwong Bor Ng</b>	822	Information Retrieval Design, James D. Anderson, Jose Perez-Carballo. Ometeca Institute, Inc., East Brunswick, NJ (2005). 617 pp., \$35.00, ISBN: 0-9763547-0-5
<b>Carlos Pampulim Caldeira</b>	824	Introduction to Information Systems: Supporting and Transforming Business. R. Rainer, E. Turban, R. Potter. John Wiley & Sons Inc. (2007). €52.50, 432p., ISBN: 0-471-73636-8
<b>Rabia Ramlogan</b>	825	Guide to Library and Information Agency Management, Charles Curran, Lewis Miller. Scarecrow Press (2005). \$45, ISBN: 0810851156
<b>Bernard J. Jansen</b>	827	The Craft of Research, 2nd edition (Chicago Guides to Writing, Editing, and Publishing) (Paperback) by Wayne C. Booth, Joseph M. Williams, Gregory G. Colomb. Paperback: 336 pages. University of Chicago Press; 2nd edition. ISBN: 0226065685



<b>Michael Buckland</b>	828	Memory Practices in the Sciences, Geoffrey C. Bowker. The MIT Press (2005). xi, 261 pp., \$34.95, ISBN: 0-262-02589-2
<b>Bernard J. Jansen</b>	830	Effective Expert Witnessing, fourth ed., Matson, Jack V., Daou, Suha F., Soper, Jeffrey G. CRC, 160 p., \$99.95, ISBN: 0849313015
<b>Mubasher Baig</b>	831	Cryptography, Information Theory, and Error-Correction a Handbook for the 21st Century, Aiden A. Bruen, Mario A. Forcinito. John Wiley & Sons, Inc. (2005). \$95, ISBN: 0-471-65317-9
		<i>Errata</i>
<b>Michel Zitt</b> <b>Elise Bassecoulard</b>	834	Erratum to “Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences” [Information Processing and Management 42 (2006) 1513–1531]
<b>Yeohoon Yoon</b> <b>Choong-Nyoung Seon</b> <b>Songwook Lee</b> <b>Jungyun Seo</b>	835	Erratum to “Unsupervised word sense disambiguation for Korean through the acyclic weighted digraph using corpus and dictionary” [Information Processing and Management 42 (2006) 710–722]
<b>Yeohoon Yoon</b> <b>Choong-Nyoung Seon</b> <b>Songwook Lee</b> <b>Jungyun Seo</b>	836	Unsupervised word sense disambiguation for Korean through the acyclic weighted digraph using corpus and dictionary
		Number 4
<b>Olga Vechtomova</b> <b>Murat Karamuftuoglu</b>	849	Query expansion with terms selected using lexical cohesion analysis of documents
<b>J. Bhogal</b> <b>A. Macfarlane</b> <b>P. Smith</b>	866	A review of ontology based query expansion
<b>Seung-Hoon Na</b> <b>In-Su Kang</b> <b>Jong-Hyeok Lee</b>	887	Adaptive document clustering based on query-based similarity
<b>Kyung-Soon Lee</b> <b>Kyo Kageura</b>	902	Virtual relevant documents in text categorization with support vector machines
<b>Tai-Yue Wang</b> <b>Huei-Min Chiang</b>	914	Fuzzy support vector machine for multi-class text categorization
<b>Jung-Mn Kim</b> <b>Hyopil Shin</b> <b>Hyoung-Joo Kim</b>	930	Schema and constraints-based matching and merging of Topic Maps

<b>Octavio Santana Suárez</b> <b>Francisco J. Carreras Riudavets</b> <b>Zenón Hernández Figueroa</b> <b>Antonio C. González Cabrera</b>	946	Integration of an XML electronic dictionary with linguistic tools for natural language processing
<b>Robert M. Losee</b>	958	Decisions in thesaurus construction and use
<b>Zhou GuoDong</b> <b>Zhang Min</b>	969	Extracting relation information from text documents by exploring various types of knowledge
<b>Filipe Mesquita</b> <b>Altigran S. da Silva</b> <b>Edleno S. de Moura</b> <b>Pável Calado</b> <b>Alberto H.F. Laender</b>	983	LABRADOR: Efficiently publishing relational databases on the web by using keyword-based query interfaces
<b>Anni Järvelin</b> <b>Antti Järvelin</b> <b>Kalervo Järvelin</b>	1005	<i>s</i> -grams: Defining generalized <i>n</i> -grams for information retrieval
<b>Robert M. Losee</b>	1020	Percent perfect performance (PPP)
<b>Thomas Krichel</b>	1030	Information retrieval performance measures for a current awareness report composition aid
<b>Anselm Spoerri</b>	1044	Examining the Authority and Ranking Effects as the result list depth used in data fusion is varied
<b>Anselm Spoerri</b>	1059	Using the structure of overlap between search results to rank retrieval systems without relevance judgments
<b>Arthur R. Taylor</b> <b>Colleen Cool</b> <b>Nicholas J. Belkin</b> <b>William J. Amadio</b>	1071	Relationships between categories of relevance criteria and stage in task completion
<b>David Nicholas</b> <b>Paul Huntington</b> <b>Hamid R. Jamali</b> <b>Tom Dobrowolski</b>	1085	Characterising and evaluating information seeking behaviour in a digital environment: Spotlight on the ‘bouncer’
<b>Lena Veiga e Silva</b> <b>Marcos André Gonçalves</b> <b>Alberto H.F. Laender</b>	1103	Evaluating a digital library self-archiving service: The BDBComp user case study
<b>William H. Walters</b>	1121	Google Scholar coverage of a multidisciplinary field
<b>Katie M. Lever</b> <b>James E. Katz</b>	1133	Cell phones in campus libraries: An analysis of policy responses to an invasive mobile technology
<b>Stephen Edward Bales</b>	1140	<i>Book reviews</i> Messages, Meaning, and Symbols: The Communication of Information. Charles T. Meadow, Scarecrow Press,

- Lanham, MD (2006). 256 pp., \$40, ISBN: 978-0-8108-5271-6
- Mohamed Taher** 1142 Project Management: Tools and Techniques for Today's LIS Professional, Barbara Allan. Neal-Schuman Publishers, New York (2004). Published by FACET – imprint of the Chartered Institute of Library and Information Professionals, \$75.00
- Rabia Ramlogan** 1143 Teams in Library Technical Services, Rosann Bazirjian, Rebecca Mugridge (Eds.), The Scarecrow Press (2006). \$40, ISBN: 978-0-8108-5294-5
- Rabia Ramlogan** 1145 The Nextgen Librarian's Survival Guide, Rachel Singer Gordon. Information Today, Inc. (2006). xiv, 208 pp., \$29.50, ISBN: 1-57387-256-3
- Bernard J. Jansen** 1147 The Long Tail: Why the Future of Business is Selling Less or More, Chris Anderson. Hyperion, New York (2006). \$24.95, ISBN: 1-4013-0237-8

## Number 5

**SPECIAL ISSUE****Patent Processing**

- Guest Editorial*
- Atsushi Fujii** 1149 Introduction to the special issue on patent processing  
**Makoto Iwayama**  
**Noriko Kando**
- Sumio Fujita** 1154 Technology survey and invalidity search: A comparative study of different tasks for Japanese patent document retrieval
- In-Su Kang** 1173 Cluster-based patent retrieval  
**Seung-Hoon Na**  
**Jungi Kim**  
**Jong-Hyeok Lee**
- Yaoyong Li** 1183 Advanced learning algorithms for cross-language patent retrieval and classification  
**John Shawe-Taylor**
- Jae-Ho Kim** 1200 Patent document categorization based on semantic structural information  
**Key-Sun Choi**
- Yuen-Hsien Tseng** 1216 Text mining techniques for patent analysis  
**Chi-Jen Lin**  
**Yu-I Lin**

		<i>General Research Papers</i>
<b>Changki Lee</b> <b>Gary Geunbae Lee</b> <b>Myunggil Jang</b>	1249	Dependency structure language model for topic detection and tracking
<b>Lois Rigouste</b> <b>Olivier Cappé</b> <b>François Yvon</b>	1260	Inference and evaluation of the multinomial mixture model for text clustering
<b>Hyoungdong Han</b> <b>Youngjoong Ko</b> <b>Jungyun Seo</b>	1281	Using the revised EM algorithm to remove noisy data for improving the one-against-the-rest method in binary text classification
<b>Ben He</b> <b>Iadh Ounis</b>	1294	Combining fields for query expansion and adaptive query expansion
<b>H.C. Wu</b> <b>Robert W.P. Luk</b> <b>K.F. Wong</b> <b>K.L. Kwok</b>	1308	A retrospective study of a hybrid document-context based retrieval model
<b>Yewei Xue</b> <b>Yunhua Hu</b> <b>Guomao Xin</b> <b>Ruihua Song</b> <b>Shuming Shi</b> <b>Yunbo Cao</b> <b>Chin-Yew Lin</b> <b>Hang Li</b>	1332	Web page title extraction and its application
<b>George Pallis</b> <b>Lefteris Angelis</b> <b>Athena Vakali</b>	1348	Validation and interpretation of Web users' sessions clusters
<b>Jeroen Kraaijenbrink</b>	1368	Engineers and the Web: An analysis of real life gaps in information usage
<b>Sujin Kim</b> <b>John Gilbertson</b>	1383	Information requirements of cancer center researchers focusing on human biological samples and associated data
<b>Keiko Kurata</b> <b>Mamiko Matsubayashi</b> <b>Shinji Mine</b> <b>Tomohide Muranushi</b> <b>Shuichi Ueda</b>	1402	Electronic journals and their unbundled functions in scholarly communication: Views and utilization by scientific, technological and medical researchers in Japan
<b>Marcos André Gonçalves</b> <b>Bárbara L. Moreira</b> <b>Edward A. Fox</b> <b>Layne T. Watson</b>	1416	“What is a good digital library?” – A quality model for digital libraries

		<i>Book review</i>
<b>Hwan-Kyu Park</b>	1438	Multimedia Content and the Semantic Web: Methods, Standards and Tools, G. Stamou, S. Kollias (Eds.), John Wiley & Sons, Ltd., West Sussex, England (2005)
Number 6		
<b>SPECIAL ISSUE</b>		
<b>Text Summarization</b>		
		<i>In Memoriam</i>
<b>Stephen Robertson</b> <b>John Tait</b>	1441	In Memoriam. Karen Spärck Jones
<b>Donna Harman</b>	1445	Dedication
		<i>Guest Editorial</i>
<b>Donna Harman</b>	1446	Editorial
<b>Karen Spärck Jones</b>	1449	Automatic summarising: The state of the art
<b>Stacy President Hobson</b> <b>Bonnie J. Dorr</b> <b>Christof Monz</b> <b>Richard Schwartz</b>	1482	Task-based evaluation of text summarization using Relevance Prediction
<b>Jonas Sjöbergh</b>	1500	Older versions of the ROUGEeval summarization evaluation system were easier to fool
<b>Paul Over</b> <b>Hoa Dang</b> <b>Donna Harman</b>	1506	DUC in context
<b>Tsutomu Hirao</b> <b>Manabu Okumura</b> <b>Norihito Yasuda</b> <b>Hideki Isozaki</b>	1521	Supervised automatic evaluation for summarization with voted regression model
<b>R. Soricut</b> <b>D. Marcu</b>	1536	Abstractive headline generation using WIDL-expressions
<b>David Zajic</b> <b>Bonnie J. Dorr</b> <b>Jimmy Lin</b> <b>Richard Schwartz</b>	1549	Multi-candidate reduction: Sentence compression as a tool for document summarization tasks
<b>Tadashi Nomoto</b>	1571	Discriminative sentence compression with conditional random fields

<b>Daniel M. Dunlavy</b> <b>Dianne P. O’Leary</b> <b>John M. Conroy</b> <b>Judith D. Schlesinger</b>	1588	QCS: A system for querying, clustering and summarizing documents
<b>Lucy Vanderwende</b> <b>Hisami Suzuki</b> <b>Chris Brockett</b> <b>Ani Nenkova</b>	1606	Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion
<b>Sanda Harabagiu</b> <b>Andrew Hickl</b> <b>Finley Lacatusu</b>	1619	Satisfying information needs with multi-document summaries
<b>Shiren Ye</b> <b>Tat-Seng Chua</b> <b>Min-Yen Kan</b> <b>Long Qiu</b>	1643	Document concept lattice for text understanding and summarization
<b>Josef Steinberger</b> <b>Massimo Poesio</b> <b>Mijail A. Kabadjov</b> <b>Karel Jeřek</b>	1663	Two uses of anaphora resolution in summarization
<b>Bonnie J. Dorr</b> <b>Terry Gaasterland</b>	1681	Exploiting aspectual features and connecting words for summarization-inspired temporal-relation extraction
<b>Gonenc Ercan</b> <b>Ilyas Cicekli</b>	1705	Using lexical chains for keyword extraction
<b>Alberto Díaz</b> <b>Pablo Gervás</b>	1715	User-model based personalized summarization
<b>Dou Shen</b> <b>Qiang Yang</b> <b>Zheng Chen</b>	1735	Noise reduction through summarization for Web-page classification
<b>Marie-Francine Moens</b>	1748	Summarizing court decisions
<b>Lawrence H. Reeve</b> <b>Hyoil Han</b> <b>Ari D. Brooks</b>	1765	The use of domain-specific concepts in biomedical text summarization
<b>Xu Ling</b> <b>Jing Jiang</b> <b>Xin He</b> <b>Qiaozhu Mei</b> <b>Chengxiang Zhai</b> <b>Bruce Schatz</b>	1777	Generating gene summaries from biomedical literature: A study of semi-structured summarization

<b>Uddin Md. Sharif</b> <b>Elmarhomy Ghada</b> <b>Elsayed Atlam</b> <b>Masao Fuketa</b> <b>Kazuhiro Morita</b> <b>Jun-ichi Aoe</b>	1793	<i>General Research Papers</i> Improvement of building field association term dictionary using passage retrieval
<b>M. Mercedes Martínez-González</b> <b>Pablo de la Fuente</b>	1808	Introducing structure management in automatic reference resolution: An XML-based approach
<b>Kshitij Grover</b>	1833	<i>Book review</i> Management Information System – Managing the Digital Firm, ninth ed., Kenneth C. Laudon, Jane P. Laudon, Prentice-Hall, New Jersey (2005). ISBN: 0-131-53841-1
	I	Volume Contents and Author Index

## AUTHOR INDEX

- Abdelali, A., 705  
Abraham, J., 287  
Adiego, J., 769  
Ahlgren, P., 81  
Amadio, W. J., 1071  
Anderson, J. D., 821  
Angelis, L., 1348  
Aoe, J.-i., 237, 1793  
Atlam, E., 1793  
Aykanat, C., 609
- Bachir, I., 47, 65  
Badue, C. S., 592  
Baeza-Yates, R., 592  
Baig, M., 831  
Bales, S. E., 1140  
Bassecoulard, E., 834  
Belkin, N. J., 181, 1071  
Berlanga-Llavori, R., 752  
Bhogal, J., 866  
Bilal, D., 47, 65  
Brockett, C., 1606  
Brooks, A. D., 1765  
Buckland, M., 828
- Cacheda, F., 204  
Calado, P., 983  
Caldeira, P. C., 824  
Cambazoglu, B. B., 609  
Cannataro, M., 577  
Cao, Y., 1332  
Cappé, O., 1260  
Carneiro, V., 204  
Carreras Riudavets, F. J., 946  
Carroll, J., 1  
Chang, C.-C., 461  
Chang, Y.-C., 488  
Chen, H.-H., 327, 488  
Chen, Z., 1735  
Chevallet, J.-P., 515  
Chiang, H.-M., 914  
Cho, S.-B., 225  
Choi, K.-S., 1200  
Chua, T.-S., 1643  
Cicekli, I., 1705  
Conroy, J. M., 1588  
Cool, C., 1071  
Cowie, J., 705  
Cragin, M. H., 808  
Croft, W. B., 181
- Dang, H., 1506  
da Silva, A. S., 983  
de la Fuente, P., 769, 1808  
de Moura, E. S., 983  
de Pedro-Cuesta, J., 273  
Deng, P., 406  
Devlin, S., 146  
Díaz, A., 1715  
Dobrowolski, T., 1085  
Dorr, B. J., 1482, 1549, 1681  
Dunlavy, D. M., 1588
- Egghe, L., 265  
Ercan, G., 1705
- Falchi, F., 665  
Feng, G., 445  
Fischer, G., 624  
Fong, A. C. M., 248  
Foulonneau, M., 740  
Fox, E. A., 1416  
Fu, X., 30  
Fujii, A., 1149  
Fujino, A., 379  
Fujita, S., 1154  
Fuketa, M., 237, 1793
- Gaasterland, T., 1681  
Gennaro, C., 665  
Gervás, P., 1715  
Ghada, E., 1793  
Gilbertson, J., 1383  
Gonçalves, M. A., 1103, 1416  
González Cabrera, A. C., 946  
Grover, K., 1833  
Grunfeld, L., 406
- Han, H., 1281, 1765  
Han, K.-S., 353  
Harabagiu, S., 1619  
Harman, D., 1445, 1446, 1506  
He, B., 1294  
He, X., 1777  
Herlocker, J. L., 791  
Hernández Figueroa, Z., 946  
Hernon, P., 1  
Hersh, W., 1  
Hickl, A., 1619  
Hirao, T., 1521  
Hobson, S. P., 1482  
Hogan, T. P., 808  
Hong, J.-H., 225  
Hu, Y., 1332  
Hui, S. C., 248  
Huntington, P., 1085
- Iribarren-Maestro, I., 273  
Isozaki, H., 1521  
Iwayama, M., 1149
- Jamali, H. R., 1085  
Jang, M., 1249  
Jansen, B. J., 827, 830, 1147  
Järvelin, A., 1005  
Järvelin, K., 1005  
Ježek, K., 1663  
Ji, D., 315  
Ji, D.-H., 730  
Jiang, J., 1777  
Jin, Y., 365  
Jung, S., 791  
Jung, Y., 365
- Kabadjov, M. A., 1663  
Kageura, K., 902  
Kan, M.-Y., 431, 1643
- Kando, N., 1149  
Kang, I.-S., 121, 302, 887, 1173  
Karaca, E., 609  
Karamuftuoglu, M., 849  
Katz, J. E., 1133  
Kekäläinen, J., 81  
Kelly, D., 30, 181  
Kim, B. M., 473  
Kim, H., 420  
Kim, H.-J., 930  
Kim, J., 1173  
Kim, J.-H., 1200  
Kim, J.-M., 930  
Kim, J.-Y., 503  
Kim, K.-M., 225  
Kim, S., 1383  
Kim, S.-B., 353  
Kim, S. K., 503  
Kishida, K., 103  
Ko, Y., 1281  
Komlodi, A., 10  
Kraaijenbrink, J., 1368  
Kraft, D., 1  
Krichel, T., 1030  
Kucukyilmaz, T., 609  
Kuo, J.-J., 327  
Kurata, K., 1402  
Kwok, K.-L., 406  
Kwok, K. L., 1308
- Lacatusu, F., 1619  
Laender, A. H. F., 983, 1103  
Lee, C., 1249  
Lee, G. G., 297, 393, 1249  
Lee, H., 420  
Lee, J.-H., 121, 302, 887, 1173  
Lee, K.-S., 902  
Lee, S., 835, 836  
Leong, M., 315  
Leong, M.-K., 515  
Lever, K. M., 1133  
Li, H., 1332  
Li, Q., 473  
Li, Y., 1183  
Liang, S. F., 146  
Lim, J.-H., 515  
Lin, C.-J., 1216  
Lin, C.-Y., 1332  
Lin, J., 717, 1549  
Lin, W.-C., 488  
Lin, Y.-I., 1216  
Ling, X., 1777  
Liu, P., 296  
Liu, R.-L., 154  
Liu, T.-Y., 445  
Losee, R. M., 958, 1020  
Lu, T.-C., 461  
Lu, W., 431  
Luk, R. W. P., 1308
- Ma, W.-Y., 445  
Macfarlane, A., 866  
Marchionini, G., 1, 10, 685  
Marcu, D., 1536



- Martínez-González, M. M., 1808  
 Matsubayashi, M., 1402  
 Mei, Q., 1777  
 Mesquita, F., 983  
 Michel, S., 643  
 Mine, S., 1402  
 Moens, M.-F., 1748  
 Monz, C., 1482  
 Moreira, B. L., 1416  
 Morita, K., 237, 1793  
 Mothe, J., 1  
 Muranushi, T., 1402  
 Murdock, V., 181  
 Muresan, G., 285  
 Mutshewa, A., 292  
 Myaeng, S. H., 297, 365, 473  
  
 Na, S.-H., 121, 302, 887, 1173  
 Navarro, G., 769  
 Nenkova, A., 1606  
 Ng, K. B., 822  
 Ng, W., 290  
 Nicholas, D., 1085  
 Nie, J.-Y., 344  
 Niu, Z.-Y., 730  
 Nomoto, T., 1571  
 Nottelmann, H., 552, 624  
  
 Okumura, M., 1521  
 Oono, M., 237  
 O'Leary, D. P., 1588  
 Ounis, I., 204, 1294  
 Over, P., 1506  
  
 Pallis, G., 1348  
 Palmer, C. L., 808  
 Paradis, F., 344  
 Park, H.-K., 1438  
 Parreira, J. X., 643  
 Pittarelli, M., 295  
 Plachouras, V., 204  
 Poesio, M., 1663  
 Pons-Porrata, A., 752  
  
 Qin, T., 445  
 Qiu, L., 1643  
  
 Ramírez-de Santa Pau, M., 273  
 Ramlogan, R., 825, 1143, 1145  
  
 Reddy, M., 287  
 Reeve, L. H., 1765  
 Ribeiro-Neto, B., 592  
 Rigouste, L., 1260  
 Rim, H.-C., 353  
 Ro, Y. M., 503  
 Robertson, S., 1441  
 Roh, J.-E., 302  
 Ruiz-Shulcloper, J., 752  
  
 Saito, K., 379  
 Sakai, T., 531  
 Santana Suárez, O., 946  
 Sanz-Casado, E., 273  
 Saracevic, T., 1, 3  
 Schatz, B., 1777  
 Schlesinger, J. D., 1588  
 Scholer, F., 169  
 Schwartz, C., 1  
 Schwartz, R., 1482, 1549  
 Seo, J., 420, 835, 836, 1281  
 Seo, K. S., 503  
 Seo, Y. S., 503  
 Seon, C.-N., 835, 836  
 Sharif, U. Md., 1793  
 Shawe-Taylor, J., 1183  
 Shen, D., 1735  
 Shi, S., 1332  
 Shin, H., 930  
 Shokouhi, M., 169  
 Silvestri, F., 549  
 Sjöbergh, J., 1500  
 Smith, P., 866  
 Soergel, D., 10  
 Soliman, H. S., 705  
 Song, R., 1332  
 Song, Y.-I., 353  
 Soricut, R., 1536  
 Spärck Jones, K., 1449  
 Spoerri, A., 1044, 1059  
 Steinberger, J., 1663  
 Straccia, U., 552  
 Suárez-Balseiro, C., 273  
 Sumitomo, T., 237  
 Suzuki, H., 1606  
  
 Tahaghoghi, S., 169  
 Taher, M., 288, 1142  
 Tait, J., 146, 1441  
  
 Tan, C. L., 730  
 Taylor, A. R., 1071  
 Tho, Q. T., 248  
 Tradigo, G., 577  
 Tseng, Y.-H., 1216  
 Turk, A., 609  
  
 Ueda, N., 379  
 Ueda, S., 1402  
  
 Vakali, A., 1348  
 Vanderwende, L., 1606  
 Vechtomova, O., 849  
 Veiga e Silva, L., 1103  
 Veltri, P., 577  
  
 Walters, W. H., 1121  
 Wang, D.-S., 445  
 Wang, T.-Y., 914  
 Watson, L. T., 1416  
 Webster, J., 791  
 Weikum, G., 643  
 White, R. W., 685  
 Wong, K. F., 1308  
 Wu, H. C., 1308  
  
 Xin, G., 1332  
 Xue, Y., 1332  
  
 Yang, L., 315  
 Yang, Q., 1735  
 Yang, S., 503  
 Yasuda, N., 1521  
 Yata, S., 237  
 Ye, S., 1643  
 Yoon, Y., 393, 835, 836  
 Yuan, X., 181  
 Yvon, F., 1260  
  
 Zajic, D., 1549  
 Zezula, P., 665  
 Zhai, C., 1777  
 Zhang, M., 969  
 Zhang, X.-D., 445  
 Zhou, G., 969  
 Zitt, M., 834  
 Ziviani, A., 592  
 Ziviani, N., 592  
 Zobel, J., 1, 169