

Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries

Vancouver, British Columbia, Canada

June 18-23, 2007



Sponsors:

ACM SIG on Information Retrieval

ACM SIG on Hypertext, Hypermedia and the Web

IEEE Technical Committee for Digital Libraries

In cooperation:

American Society for Information Science & Technology

Coalition for Networked Information



Advancing Computing as a Science & Profession

Building & Sustaining the Digital Environment

JCDL 2007

Ray Larson, Edie Rasmussen, Shigeo Sugimoto and Elaine Toms, Editors

Table of Contents

JCDL 2007 Conference Organization	xii
---	-----

JCDL 2007 Sponsors	xiv
--------------------------	-----

Session A: Visualization

- **World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-Referenced Collections** 1
S. Ahern, M. Naaman, R. Nair, J. Yang (*Yahoo! Research, Berkeley*)
- **Categorization and Analysis of Text in Computer Mediated Communication Archives using Visualization** 11
A. Abbasi, H. Chen (*University of Arizona*)
- **Delineating the Citation Impact of Scientific Discoveries** 19
C. Chen, Jian Zhang, W. Zhu, M. Vogeley (*Drexel University*)

Session B: Digital Curation and Preservation

- **How to Choose a Digital Preservation Strategy: Evaluating a Preservation Planning Procedure** 29
S. Strodl, C. Becker, R. Neumayer, A. Rauber (*Vienna University of Technology*)
- **Factors Affecting Website Reconstruction from the Web Infrastructure** 39
F. McCown, N. Diawara, M. L. Nelson (*Old Dominion University*)
- **Defining What Digital Curators Do and What They Need to Know: The DigCCurr Project** 49
C. A. Lee, H. R. Tibbo, J. C. Schaefer (*University of North Carolina at Chapel Hill*)
- **Generating Best-Effort Preservation Metadata for Web Resources at Time of Dissemination** 51
J. A. Smith, M. L. Nelson (*Old Dominion University*)

Session C: Information Extraction 1

- **Document Clustering using Small World Communities** 53
B. Chee, B. Schatz (*University of Illinois at Urbana-Champaign*)
- **Efficient Summarization-Aware Search for Online News Articles** 63
W. Dakka, L. Gravano (*Columbia University*)
- **Integrating Data and Text Mining Processes for Digital Library Applications** 73
R. Sanderson, P. Watry (*University of Liverpool*)

Session D: Panel 1

- **The OAI-ORE Effort: Progress, Challenges, Synergies** 80
C. Lynch (*Coalition for Networked Information*), S. Parastatidis (*Microsoft*), N. Jacobs (*JISC*),
H. Van de Sompel (*Los Alamos National Laboratory*), C. Lagoze (*Cornell University*)

Session E: Information Extraction 2

- **SlideSeer: A Digital Library of Aligned Document and Presentation Pairs** 81
M.-Y. Kan (*National University of Singapore*)
- **TableSeer: Automatic Table Metadata Extraction and Searching in Digital Libraries** 91
Y. Liu, K. Bai, P. Mitra, C. L. Giles (*The Pennsylvania State University*)
- **CiteSearch: Next-Generation Citation Analysis** 101
K. Yang, L. Meho (*Indiana University*)

- **Retrieval Effectiveness of Table of Contents and Subject Headings** 103
Y. Choi, I. Hsieh-Yee, B. Kules (*Catholic University of America*)
- **Mining a Digital Library for Influential Authors** 105
D. Mimno, A. McCallum (*University of Massachusetts at Amherst*)

Session F: Social Networks

- **Can Social Bookmarking Enhance Search in the Web?**..... 107
Y. Yanbe, A. Jatowt, S. Nakamura, K. Tanaka (*Kyoto University*)
- **Task-based Interaction with an Integrated Multilingual, Multimedia Information System: A Formative Evaluation** 117
P. Zhang, L. Plettenberg, J. L. Klavans, D. W. Oard, D. Soergel (*University of Maryland*)
- **Modeling Personal and Social Network Context for Event Annotation in Images** 127
B. Shevade, H. Sundaram (*Arizona State University*), L. Xie (*IBM T.J. Watson Research Center*)
- **Longitudinal Study of Changes in Blogs** 135
P. L. Bogen II (*Texas A&M University*), L. Francisco-Revilla (*The University of Texas*),
R. Furuta, T. Hubbard, U. P. Karadkar, F. Shipman (*Texas A&M University*)

Session G: Systems

- **SearchGen: A Synthetic Workload Generator for Scientific Literature Digital Libraries and Search Engines** 137
H. Li, W.-C. Lee, A. Sivasubramaniam, C. L. Giles (*The Pennsylvania State University*)
- **A Retrospective Look at Greenstone: Lessons from the First Decade** 147
I. H. Witten, D. Bainbridge (*University of Waikato*)
- **A Unified Platform for Archival Description and Access** 157
C. J. Prom, C. A. Rishel, S. W. Schwartz, K. J. Fox (*University of Illinois at Urbana-Champaign*)

Session H: Educational Digital Libraries

- **Children's Interests and Concerns When Using the International Children's Digital Library: A Four-Country Case Study**..... 167
A. Druin, A. Weeks, S. Massey, B. B. Bederson (*University of Maryland at College Park*)
- **Digital Library Education in Computer Science Programs** 177
J. Pomerantz, S. Oh, B. M. Wildemuth (*University of North Carolina at Chapel Hill*),
S. Yang, E. A. Fox (*Virginia Tech*)
- **A Study of How Online Learning Resource Are Used** 179
M. Recker (*Utah State University*), S. Giersch (*National Science Digital Library*),
A. Walker, S. Halioris, X. Mao, B. Palmer (*Utah State University*)
- **Standards or Semantics for Curriculum Search?** 181
B. Marshall, R. F. Reitsma (*Oregon State University*), M. N. Cyr (*Worcester Polytechnic Institute*)
- **Information Behavior of Small Groups: Implications for Design of Digital Libraries** 183
N. Zhou, G. Stahl (*Drexel University*)

Session I: Information Retrieval and Extraction 1

- **Adaptive Sorted Neighborhood Methods for Efficient Record Linkage** 185
S. Yan, D. Lee (*The Pennsylvania State University*), M.-Y. Kan (*National University of Singapore*),
C. L. Giles (*The Pennsylvania State University*)
- **Distributed Web Search Efficiency by Truncating Results**..... 195
C. T. Fallen, G. B. Newby (*University of Alaska at Fairbanks*)
- **Adaptive Graphical Approach to Entity Resolution** 204
Z. Chen, D. V. Kalashnikov, S. Mehrotra (*University of California at Irvine*)

Session J: Panel 2

- **Panel: Cyberinfrastructure for the Humanities and Social Sciences — Advancing the Humanities Research Agenda**214
J. Ray (*Institute of Museum and Library Services*), C. Lynch (*Coalition for Networked Information*),
B. Bobley (*National Endowment for the Humanities*), G. Crane (*Tufts University*),
S. Wheatley (*American Council of Learned Societies*)

Session K: Information Extraction 3

- **FLUX-CiM: Flexible Unsupervised Extraction of Citation Metadata**215
E. Cortez, A. S. da Silva, M. A. Gonçalves, F. Mesquita, E. S. de Moura (*Universidade Federal do Amazonas*)
- **Measuring Conference Quality by Mining Program Committee Characteristics**225
Z. Zhuang, E. Elmacioglu, D. Lee, C. L. Giles (*The Pennsylvania State University*)
- **Toward Alternative Measures for Ranking Venues:
A Case of Database Research Community**235
S. Yan, D. Lee (*The Pennsylvania State University*)

Session L: Models and Case Studies

- **A Model for Inclusive Design of Digital Libraries**245
S. Chandrashekar, N. Caidi (*University of Toronto*)
- **Representing Aggregate Works in the Digital Library**247
G. Buchanan (*Swansea University*), J. Gow, A. Blandford (*UCL Interaction Centre*),
J. Rimmer, C. Warwick (*UCL SLAIS*)
- **StoryBank: An Indian Village Community Digital Library**257
M. Jones, W. Harwood, G. Buchanan (*Swansea University*), M. Lalmas (*Queen Mary University*)
- **The Gray Lady Gets a New Dress: A Field Study of the Times News Reader**259
C. C. Marshall (*Microsoft Corporation*)

Session M: Architecture and Ontologies

- **Drowning in Data: Digital Library Architecture to Support Scientific
Use of Embedded Sensor Networks**269
C. L. Borgman, J. C. Wallis, M. S. Mayernik, A. Pepe (*University of California at Los Angeles*)
- **A Practical Ontology for the Large-Scale Modeling
of Scholarly Artifacts and Their Usage**278
M. A. Rodriguez, J. Bollen, H. Van de Sompel (*Los Alamos National Laboratory*)
- **A Dynamic Ontology for a Dynamic Reference Work**288
M. Niepert, C. Buckner, C. Allen (*Indiana University*)

Session N: Music Digital Libraries

- **Preparing Resource Discovery for Digitized Music — An Analysis
of an Australian Application**298
J. Thomas, M. Middleton (*Queens University of Technology*), M. Warren (*State Library of Queensland*)
- **Goal-Directed Evaluation for the Improvement
of Optical Music Recognition on Early Music Prints**303
L. Pugin, J. A. Burgoyne, I. Fujinaga (*McGill University*)
- **Annotation Functionality for Digital Libraries Supporting
Collaborative Performance: An Example of Musical Scores**305
M. Winget (*University of Texas at Austin*)
- **Toward an Understanding of Similarity Judgments
for Music Digital Library Evaluation**307
J. S. Downie, J. H. Lee, A. A. Gruzdz, M. C. Jones (*University of Illinois at Urbana-Champaign*)

Session O: User Studies and User Interfaces

- **Agreeing to Disagree: Search Engines and Their Public Interfaces**.....309
F. McCown, M. L. Nelson (*Old Dominion University*)
- **Static Reformulation: A User Study of Static Hypertext for Query-Based Reformulation**319
M. Huggett, J. Lanir (*University of British Columbia*)
- **A Rich OPAC User Interface with AJAX**329
J. P. Gozali, M.-Y. Kan (*National University of Singapore*)
- **Constructing Digital Library Interfaces**.....331
D. M. Nichols, D. Bainbridge (*University of Waikato*),
M. B. Twidale (*University of Illinois at Urbana-Champaign*)

Session P: Information Retrieval and Extraction 2

- **Retrieval in Text Collections with Historic Spelling using Linguistic and Spelling Variants**333
A. Ernst-Gerlach, N. Fuhr (*University of Duisburg-Essen*)
- **Efficient Topic-based Unsupervised Name Disambiguation**.....342
Y. Song, J. Huang, I. G. Council, J. Li, C. L. Giles (*The Pennsylvania State University*)
- **Using Bilingual ETD Collections to Mine Phrase Translations**352
R. Richardson, E. A. Fox (*Virginia Tech*)
- **Evaluation of Kernel-based Link Analysis Measures on Research Paper Recommendation**354
M. Shimbo, T. Ito, Y. Matsumoto (*Nara Institute of Science and Technology*)

Session Q: Large-Scale Collections

- **A New Generation of Textual Corpora — Mining Corpora from Very Large Collections**.....356
G. Stewart (*Harvard University*), G. Crane, A. Babeu (*Tufts University*)
- **Subject Metadata Enrichment using Statistical Topic Models**366
D. Newman (*University of California at Irvine*), K. Hagedorn (*University of Michigan*),
C. Chemudugunta, P. Smyth (*University of California at Irvine*)
- **Organizing the OCA: Learning Faceted Subjects from a Library of Digital Books**376
D. Mimno, A. McCallum (*University of Massachusetts at Amherst*)

Session R: Metadata

- **Trends in Metadata Practices: A Longitudinal Study of Collection Federation**386
C. Palmer, O. Zavalina, M. Mustafoff (*University of Illinois at Urbana-Champaign*)
- **Induced Tagging: Promoting Resource Discovery and Recommendation in Digital Libraries**.....396
J. A. Sánchez, A. Arzamendi-Pétriz, O. Valdiviezo (*Universidad de las Américas Puebla*)
- **Standards Alignment for Metadata Assignment**.....398
A. R. Diekema, O. Yilmazel, J. Bailey, S. C. Harwell, E. D. Liddy (*Syracuse University*)
- **Identifying Personal Photo Digital Library Features**.....400
S. J. Cunningham, M. Masoodian (*University of Waikato*)

Session S: Historical Digital Libraries

- **Locating Thematic Pinpoints in Narrative Texts with Short Phrases: A Test Study on Don Quixote**402
J. Deng, R. Furuta, E. Urbina (*Texas A&M University*)

- **Digital Donne: Workflow, Editing Tools, and the Reader's Interface of a Collection of 17th-Century English Poetry**411
C. Monroy, R. Furuta, G. Stringer (*Texas A&M University*)
- **A Multilingual Approach to Technical Manuscripts: 16th and 17th-Century Portuguese Shipbuilding Treatises**413
C. Monroy, R. Furuta, F. Castro (*Texas A&M University*)
- **First Class Objects and Indexes for Chant Manuscripts**415
L. W. G. Barton, P. G. Jeavons, J. A. Caldwell (*University of Oxford*)
- **Recommending Related Papers Based on Digital Library Access Records**.....417
S. Pohl, F. Radlinski, T. Joachims (*Cornell University*)

Session T: Automatic Classification

- **Automatic Patent Classification using Citation Network Information: An Experimental Study in Nanotechnology**419
X. Li, H. Chen, Z. Zhang, J. Li (*University of Arizona*)
- **Collaborative Classifier Agents: Studying the Impact of Learning in Distributed Document Classification**428
W. Ke, J. Mostafa, Y. Fu (*Indiana University*)
- **UpdateNews: A News Clustering and Summarization System Using Efficient Text Processing**438
T. Takeda (*The Graduate University for Advanced Studies*), A. Takasu (*National Institute of Informatics*)
- **Automatic Syllabus Classification**440
X. Yu, M. Tungare, W. Fan, M. Perez-Quinones, E. A. Fox (*Virginia Tech*),
W. Cameron, GF. Teng, L. Cassel (*Villanova University*)

Session U: Search Behavior and Personalization

- **Effects of Structure and Interaction Style on Distinct Search Tasks**442
R. Capra, G. Marchionini, J. S. Oh, F. Stutzman, Y. Zhang (*University of North Carolina at Chapel Hill*)
- **Towards Automatic Conceptual Personalization Tools**452
F. Ahmed, S. de la Chica (*University of Colorado*), K. Butcher (*University of Pittsburgh*),
T. Sumner, J. H. Martin (*University of Colorado*)
- **Mobile G-Portal Supporting Collaborative Sharing and Learning in Geography Fieldwork: An Empirical Study**462
Y.-L. Theng, K.-L. Tan, E.-P. Lim, Jun Zhang, D. H.-L. Goh (*Nanyang Technological University*),
K. Chatterjea, C. H. Chang (*National Institute of Education*),
A. Sun, Y. Han, N. H. Dang (*Nanyang Technological University*),
Y. Li (*National Institute of Education*), M. C. Vo (*Nanyang Technological University*)

Posters

- **Highly Structured Scientific Publications**472
R. B. Allen (*Drexel University*)
- **Cooperative Collection Building in NSDL MatDL Pathway through iVia Data Fountains**473
L. M. Bartolo, C. S. Lowe (*Kent State University*),
J. Ruscheinski, D. Bisom (*University of California at Riverside*)
- **MESUR: Usage-based Metrics of Scholarly Impact**474
J. Bollen, M. A. Rodriguez, H. Van de Sompel (*Los Alamos National Laboratory*)
- **A Publisher of Last Resort: Enduring Document Access**475
G. Buchanan (*Swansea University*)
- **Educational Application Integration with Digital Repository**476
R. Chavez, A. Kumar, N. Schwertner (*Tufts University*)

• Blogger Perceptions on Digital Preservation	477
C. Hank, S. Choemprayong, L. Sheble (<i>University of North Carolina at Chapel Hill</i>)	
• “Evolution of a Data Archive”	478
J. Crabtree, D. Sheaves (<i>University of North Carolina at Chapel Hill</i>)	
• Examining Perception of Digital Information Space	479
J. D’Ignazio, J. Ryan, S. Harwell, A. Diekema, E. Liddy (<i>Syracuse University</i>)	
• Tagging Video: Conventions and Strategies of the YouTube Community	480
G. Geisler, S. Burns (<i>The University of Texas at Austin</i>)	
• DRIADE: A Data Repository for Evolutionary Biology	481
J. Dube, S. Carrier, J. Greenberg (<i>University of North Carolina at Chapel Hill</i>)	
• AlouetteCanada Metadata Toolkit	482
M. Jordan (<i>Simon Fraser University</i>)	
• Building a Digital Library of Traditional Mongolian Historical Documents	483
G. Khaltarkhuu, A. Maeda (<i>Ritsumeikan University</i>)	
• Evaluating Digital Libraries with Webmetrics	484
M. Khoo (<i>NSDL Core Integration</i>), R. A. Donahue (<i>WGBH Educational Foundation</i>)	
• Tagging for Health Information Organisation and Retrieval	485
M. E. I. Kipp (<i>University of Western Ontario</i>)	
• Augmenting OAI-PMH Repository Holdings Using Search Engine APIs	486
M. Klein, M. L. Nelson (<i>Old Dominion University</i>), J. Z. Pao (<i>NASA Langley Research Center</i>)	
• The Cyberinfrastructure for Scholars Project: Componentized Architecture for Sustainable Scholarly Portals	487
A. Krowne, S. Martin, U. Gadi, M. Wedemeyer, M. Halbert (<i>Emory University</i>)	
• Social Bookmarking in Digital Library Systems: Framework and Case Study	488
F. Puspitasari, E.-P. Lim, C.-H. Chang, Y.-L. Theng, D. H.-L. Goh, K. Chatterjea, Jun Zhang, A. Sun, Y. Li (<i>Nanyang Technological University</i>)	
• Automated Collection Strength Analysis	489
C. Llewellyn (<i>University of Liverpool</i>), B. Rea (<i>University of Manchester</i>), R. Sanderson (<i>University of Liverpool</i>)	
• Digital Library Education: Some International Course Structure Comparisons	490
Y. Ma (<i>Victoria University of Wellington</i>), A. O’Brien (<i>Loughborough University</i>), W. Clegg (<i>Victoria University of Wellington</i>)	
• PIM through a 5S Persepective	491
Y. Ma, E. A. Fox (<i>Virginia Tech</i>), M. A. Gonçalves (<i>Federal University of Minas Gerais</i>)	
• Use vs. Access: Design and Use in Educational Digital Libraries	492
F. McMartin, B. Muramatsu (<i>Utah State University</i>)	
• What Do Faculty Need and Want from Digital Libraries?	493
F. McMartin (<i>Broad-based Knowledge</i>), A. Wolf (<i>University of Wisconsin at Madison</i>), G. Morgan (<i>George Mason University</i>), E. Iverson, C. Manduca (<i>Carleton College</i>), J. Morrill (<i>Morrill Solutions</i>)	
• Building Cross-Browser Interfaces for Digital Libraries with Scalable Vector Graphics (SVG)	494
F. Molina, B. Sweeney, T. Willard (<i>AAAS Project 2061</i>), A. Winter (<i>carto.net</i>)	
• Understanding Target Users of a Digital Reference Library	495
D. Rosner, J. M. Josling, A. Moed, E. Oreglia (<i>University of California at Berkeley</i>)	
• Capturing Relevant Information for Digital Curation	496
C. Shah, G. Marchionini (<i>University of North Carolina at Chapel Hill</i>)	
• Merging the Norwegian Gazetteer with the ADL Gazetteer	497
Ø. Vestavik, I. T. Sølvsberg (<i>Norwegian University of Science and Technology</i>)	
• Information System Media Education (ISM): Cooperating for Media Literacy	498
H. vom Orde (<i>International Central Institute for Youth and Educational Television (IZI)</i>)	

- **Digitizing & Providing Access to Contextual Cultural Materials:
The Liner Notes Digitization Project** 499
M. Winget (*University of Texas at Austin*)
- **Use of Online Digital Learning Materials and Digital Libraries:
Comparison by Discipline** 500
A. Wolf (*University of Wisconsin at Madison*), E. Iverson, C. Manduca (*Carleton College*),
G. Morgan (*George Mason University*), F. McMartin (*Broad-based Knowledge*), J. Morrill (*Morrill Solutions*)

Demos: Demonstrations

- **XML as the Articulation between Information Retrieval
and Multimedia in a Musical Heritage Dissemination** 501
R. Bailly (*Cité de la musique*)
 - **Lightweight Realistic Books: The Greenstone Connection** 502
V. Liesaputra, I. H. Witten, D. Bainbridge (*University of Waikato*)
 - **Rapid Document Navigation for Information Triage Support** 503
G. Buchanan (*Swansea University*)
 - **Fluid Interaction for the Document in Context** 504
P. Cubaud, J. Dupire, A. Topol (*Conservatoire National des Arts et Métiers (CNAM)*)
 - **Demonstrating the Semantic GrowBag:
Automatically Creating Topic Facets for FacetedDBLP** 505
J. Diederich, W.-T. Balke, U. Thaden (*Leibniz Universität Hannover*)
 - **The David L. Bassett Stereoscopic Atlas of Human Anatomy:
Developing a Specialized Collection
within the Stanford MediaServer Digital Library** 506
J. C. Durack, S.-Y. Kung, R. A. Chase, A. L. Ladd, M. Krebs, P. Dev (*Stanford University*)
 - **Evalutron 6000: Collecting Music Relevance Judgments** 507
A. A. Gruz, J. S. Downie, M. C. Jones, J. H. Lee (*University of Illinois at Urbana-Champaign*)
 - **VCenter: A Digital Video Management System with Mobile Search Service** 508
J.-H. Hsiao, Y.-Z. Wang (*Academia Sinica*)
 - **Creativity Support: The Mixed-Initiative Composition Space** 509
A. Kerne, E. Koh (*Texas A&M University*)
 - **Visual Understanding Environment** 510
A. Kumar (*Tufts University*)
 - **Mobile Digital Libraries for Geographical Education** 511
M.-C. Vo, F. Puspitasari, E.-P. Lim, C.-H. Chang, Y.-L. Theng, D. H.-L. Goh,
K. Chatterjea, Jun Zhang, A. Sun, Y. Li (*Nanyang Technological University*)
 - **The Internet Public Library: An Online Learning Laboratory for Digital Libraries** 512
L. Mon, L. Dennis, K. Kim (*Florida State University*)
 - **5SQual — A Quality Assessment Tool for Digital Libraries** 513
B. L. Moreira, M. A. Gonçalves, A. H. F. Laender (*Federal University of Minas Gerais*),
E. A. Fox (*Virginia Tech*)
 - **ContextMiner: A Tool for Digital Library Curators** 514
C. Shah, G. Marchionini (*University of North Carolina at Chapel Hill*)
 - **From Kinescope to Rich Media: 50 Years (ago) with Mike Wallace** 515
Q. Stewart (*University of Texas at Austin*), G. Pasch (*Uni. Fco. Marroquín*), R. Arias (*nTropic, Inc.*)
- Author Index** 516