

6th ACM/IEEE-CS Joint Conference on Digital Libraries 2006

Opening Information Horizons

Sponsored by:

ACM SIG on Information Retrieval

ACM SIG on Hypertext, Hypermedia and the Web

IEEE Technical Committee for Digital Libraries



June 11-15, 2006 • Chapel Hill, NC, USA

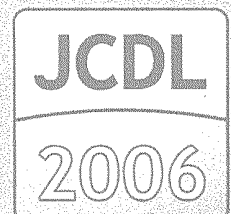


Table of Contents

JCDL 2006 Organization	xii
JCDL 2006 Program Committee	xiii
JCDL 2006 Sponsors & Supporters	xvi

Session 1: Visualization for Libraries

- **Exploring Digital Libraries: Integrating Browsing, Searching, and Visualization** 1
R. Shen, N. S. Vemuri, W. Fan (*Virginia Tech*), R. da S. Torres (*State University of Campinas*),
E. A. Fox (*Virginia Tech*)
- **combinFormation: A Mixed-Initiative System for Representing Collections
as Compositions of Image and Text Surrogates**..... 11
A. Kerne, E. Koh, B. Dworaczyk, J. M. Mistrot, H. Choi, S. M. Smith, R. Graeber,
D. Caruso, A. Webb, R. Hill, J. Albea (*Texas A&M University*)
- **InfoGallery: Informative Art Services for Physical Library Spaces**.....21
K. Grønbaek (*University of Aarhus*), A. Rohde, BS. Sundararajah (*Alexandra Institute Ltd.*),
S. Bech-Petersen (*Municipal Libraries in Aarhus*)

Session 2: Named Entities 1

- **The Challenge of Virginia Banks: An Evaluation of Named Entity Analysis
in a 19th-Century Newspaper Collection**..... 31
G. Crane, A. Jones (*Tufts University*)
- **Learning to Deduplicate**.....41
M. G. de Carvalho, M. A. Gonçalves, A. H. F. Laender (*Federal University of Minas Gerais*),
A. S. da Silva (*Federal University of Amazonas*)
- **An Effective Approach to Entity Resolution Problem Using Quasi-Clique
and its Application to Digital Libraries**..... 51
B.-W. On, E. Elmacioglu, D. Lee (*The Pennsylvania State University*), J. Kang (*NCSU*),
J. Pei (*Simon Fraser University*)
- **Also By The Same Author: AKTiveAuthor, a Citation Graph Approach
to Name Disambiguation** 53
D. M. McRae-Spencer, N. R. Shadbolt (*University of Southampton*)

Session 3: Classification and Links

- **Probabilistic, Object-oriented Logics for Annotation-based
Retrieval in Digital Libraries** 55
I. Frommholz, N. Fuhr (*University of Duisburg-Essen*)
- **Bibliometric Impact Measures Leveraging Topic Analysis** 65
G. S. Mann, D. Mimno, A. McCallum (*University of Massachusetts at Amherst*)
- **A Comparative Study of Citations and Links in Document Classification** 75
T. Couto, M. Cristo, M. A. Gonçalves (*Federal University of Minas Gerais*),
P. Calado (*IST/NESC-ID*), N. Ziviani (*Federal University of Minas Gerais*),
E. Moura (*Federal University of Amazonas*),
B. Ribeiro-Neto (*Federal University of Minas Gerais and Google Engineering*)

Session 4: Panel

- **Augmenting Interoperability Across Scholarly Repositories**85
T. Hey (*Microsoft*), D. Waters (*Andrew W. Mellon Foundation*), C. Lynch (*Coalition for Networked Information*),
H. Van de Sompel (*Los Alamos National Laboratory*), C. Lagoze (*Cornell University*)

Session 5: Digital Preservation

- **Quantifying Software Requirements for Supporting Archived Office Documents using Emulation** 86
T. Reichherzer, G. Brown (*Indiana University*)
- **Building a Research Library for the History of the Web** 95
W. Y. Arms, S. Aya, P. Dmitriev, B. J. Kot, R. Mitchell, L. Walle (*Cornell University*)
- **The Processing of Digitized Works** 103
J. Borbinha, J. Gil, G. Pedrosa, J. Penas (*INESC-ID - Instituto de Engenharia de Sistemas e Computadores*)
- **Document Level Interoperability for Collection Creators** 105
D. Bainbridge (*University of Waikato*), K.-Y. Ke (*DL Consulting*),
I. H. Witten (*University of Waikato*)
- **Repository Software Evaluation using the Audit Checklist for Certification of Trusted Digital Repositories** 107
J. S. Kaczmarek, T. G. Habing, J. Eke (*University of Illinois at Urbana-Champaign*)

Session 6: Document Analysis

- **A Hierarchical, HMM-based Automatic Evaluation of OCR Accuracy for a Digital Library of Books** 109
S. Feng, R. Manmatha (*University of Massachusetts at Amherst*)
- **Combining DOM Tree and Geometric Layout Analysis for Online Medical Journal Article Segmentation** 119
J. Zou, D. Le, G. R. Thoma (*National Library of Medicine*)
- **Automatic Categorization of Figures in Scientific Documents** 129
X. Lu, P. Mitra, J. Z. Wang, C. L. Giles (*The Pennsylvania State University*)
- **XML Views for Electronic Editions** 139
I. E. Iacob, A. Dekhtyar (*University of Kentucky*)

Session 7: Time and Space

- **Exploring Erotics in Emily Dickinson's Correspondence with Text Mining and Visual Interfaces** 141
C. Plaisant, J. Rose (*University of Maryland*), B. Yu, L. Auvil (*University of Illinois*),
M. G. Krischenbaum, M. N. Smith, T. Clement, G. Lord (*University of Maryland*)
- **Time Period Directories: A Metadata Infrastructure for Placing Events in Temporal and Geographic Context** 151
V. Petras, R. R. Larson, M. Buckland (*University of California, Berkeley*)
- **ETANA-ADD: An Interactive Tool for Integrating Archaeological DL Collections** 161
N. S. Vemuri, R. Shen, S. Tupe, W. Fan, E. A. Fox (*Virginia Tech*)
- **Enabling Exploration: Travelers in the Middle East Archive** 163
L. Spiro, M. Wise, G. Henry, C. Bearden, S. Byrd, E. Garza (*Rice University*),
M. Decker (*University of South Florida*)

Session 8: Digital Library Curriculum

- **Digital Library Education: The Current Status** 165
Y. Ma, W. Clegg (*Victoria University of Wellington*), A. O'Brien (*Loughborough University*)
- **Curriculum Development for Digital Libraries** 175
J. Pomerantz, B. M. Wildemuth (*University of North Carolina at Chapel Hill*),
S. Yang, E. A. Fox (*Virginia Tech*)

- **Learning by Building Digital Libraries**..... 185
D. M. Nichols, D. Bainbridge (*University of Waikato*),
J. S. Downie, M. B. Twidale (*University of Illinois at Urbana-Champaign*)

- **What Do Digital Librarians Do?** 187
Y. Choi (*State University of New York at Oswego*), E. Rasmussen (*The University of British Columbia*)

Session 9: Panel

- **The NDIIPP Preservation Network: Progress, Problems, and Promise** 189
L. E. Campbell (*Library of Congress*), H. R. Tibbo, P. Leousis (*University of North Carolina at Chapel Hill*)

Session 10: Images and Sound

- **Windowing Time in Digital Libraries** 190
M. G. Christel (*Carnegie Mellon University*)

- **Concept Maps to Support Oral History Search and Use** 192
R. W. White, H. Song, J. Liu (*University of Maryland*)

- **Facilitating Access to Large Digital Oral History Archives through Informedia Technologies** 194
M. G. Christel (*Carnegie Mellon University*), J. Richardson (*The HistoryMakers*),
H. D. Wactlar (*Carnegie Mellon University*)

- **Review Mining for Music Digital Libraries: Phase II**..... 196
J. S. Downie, X. Hu (*University of Illinois at Urbana-Champaign*)

- **Looking for a Picture: An Analysis of Everyday Image Information Searching**..... 198
S. J. Cunningham, M. Masoodian (*University of Waikato*)

- **Image-based Evaluation of Video-acquired Research Skills** 200
U. P. Karadkar, M. Nordt, R. Furuta, C. Lee, C. Quick (*Texas A&M University*)

Session 11: Information Retrieval 1

- **Keyword Extraction-Based Query Expansion in Digital Libraries** 202
M. Song (*Temple University*), I.-Y. Song, R. B. Allen (*Drexel University*), Z. Obradovic (*Temple University*)

- **Categorizing Web Search Results into Meaningful and Stable Categories Using Fast-Feature Techniques** 210
B. Kules, J. Kustanowitz, B. Shneiderman (*University of Maryland*)

- **A Comprehensive Comparison Study of Document Clustering for a Biomedical Digital Library MEDLINE** 220
I. Yoo, X. Hu (*Drexel University*)

Session 12: Supporting Education

- **Metadata Aggregation and “Automated Digital Libraries”: A Retrospective on the NSDL Experience** 230
C. Lagoze, D. Krafft, T. Cornwell, N. Dushay, D. Eckstrom, J. Saylor (*Cornell University*)

- **Using Resources Across Educational Digital Libraries** 240
M. Recker, B. Palmer (*Utah State University*)

- **Template-based Authoring of Educational Artifacts** 242
S. Davis, P. Bogen, L. Cifuentes (*Texas A&M University*), L. Francisco-Revilla (*University of Texas at Austin*),
R. Furuta, T. Hubbard, U. P. Karadkar, D. Pogue, F. Shipman (*Texas A&M University*)

- **EcoPod: A Mobile Tool for Community Based Biodiversity Collection Building** 244
YY. Yu, J. A. Stamberger, A. Manoharan, A. Paepcke (*Stanford University*)

- **Factors Motivating Use of Digital Libraries** 254
F. McMartin (*Broad Based Knowledge*), E. Iverson, C. Manduca (*Carleton College*),
A. Wolf (*University of Wisconsin*), G. Morgan (*California State University*)

Session 13: Metadata in Action

- **Scaffolding the Infrastructure of the Computational Science Digital Library**256
D. Tanase (*Shodor Education Foundation*), M. Bruce (*Ohio Supercomputer Center*),
J. Stuart-Moore (*Shodor Education Foundation*), D. A. Joiner (*Kean University*)
- **Dynamic Generation of OAI Servers**258
J. A. Sánchez, A. Razo, J. M. Córdova, A. Villegas (*Universidad de las Américas Puebla*)
- **FRBR: Enriching and Integrating Digital Libraries**260
G. Buchanan (*University of Wales*)
- **Learning from Artifacts: Metadata Utilization Analysis**270
W. E. Moen, S. D. Miksa, A. Eklund, S. Polyakov, G. Snyder (*University of North Texas*)
- **Looking Back, Looking Forward: A Metadata Standard
for LANL's aDORe Repository**272
B. Goldsmith, F. Knudson (*Los Alamos National Laboratory*)

Session 14: Information Retrieval 2

- **Measuring Inter-Indexer Consistency Using a Thesaurus**274
O. Medelyan, I. H. Witten (*University of Waikato*)
- **Learning Metadata from the Evidence in an On-Line Citation Matching Scheme**276
I. G. Councill, H. Li, Z. Zhuang, S. Debnath, L. Bolelli, W.-C. Lee,
A. Sivasubramaniam, C. L. Giles (*The Pennsylvania State University*)
- **Using Controlled Query Generation To Evaluate Blind
Relevance Feedback Algorithms**286
C. Jordan, C. Watters, Q. Gao (*Dalhousie University*)
- **Thesaurus Based Automatic Keyphrase Indexing**296
O. Medelyan, I. H. Witten (*University of Waikato*)

Session 15: Usage and Relationships

- **An Architecture for the Aggregation and Analysis of Scholarly Usage Data**298
J. Bollen, H. Van de Sompel (*Los Alamos National Laboratory*)
- **An Experimental Framework for Comparative Digital Library Evaluation:
The Logging Scheme**308
C.-P. Klas, N. Fuhr, S. Kriewel (*University of Duisburg-Essen*),
H. Albrechtsen (*Institute of Knowledge Sharing*), G. Tsakonas, S. Kapidakis, C. Papatheodorou (*Ionian University*),
P. Hansen (*Swedish Institute of Computer Science*), L. Kovacs, A. Micsik (*SZTAKI*), E. Jacob (*Indiana University*)
- **Insights into Collections Gaps through Examination
of Null Result Searches in DLESE**310
B. DeFelice (*Dartmouth College*), K. A. Kastens (*Columbia University*),
C. Rinaldo (*Harvard University*), J. Weatherley (*UCAR*)
- **The Social Life of Books in the Humane Library**312
Y. Chisik, N. Kaplan (*University of Baltimore*)

Session 16: Named Entities 2

- **Search Engine Driven Author Disambiguation**314
Y. F. Tan, M.-Y. Kan (*National University of Singapore*), D. Lee (*The Pennsylvania State University*)
- **Tagging of Name Records for Genealogical Data Browsing**316
M. Perrow, D. Barber (*IDIAP Research Institute*)
- **Automatic Feature Thesaurus Enrichment — Extracting Generic Terms
from Digital Gazetteer**326
J. Wang, N. Ge (*Peking University*)

Posters

- **An Assessment of Access and Use Rights for Licensed Scholarly Digital Resources**.....334
K. R. Eschenfelder, I. Benton (*University of Wisconsin-Madison*)
- **Information Seeking in Academic Learning Environments:
An Exploratory Factor Analytic Approach to Understanding Design Features**335
S.-S. Lee, Y.-L. Theng, D. H.-L. Goh, S. S.-B. Foo (*Nanyang Technological University*)
- **A Performance Support Systems Approach to Digital Publishing in Libraries**336
C. Thomas (*Florida Center for Library Automation*), R. H. McDonald (*Florida State University Libraries*)
- **Selecting Books: A Performance-based Study**337
N. Wacholder, L. Liu, Y.-H. Liu (*Rutgers University*)
- **User Perceptions of a Federated Search System**338
I. Hsieh-Yee, R. Tang, S. Zhang (*Catholic University of America*)
- **Automatic Extraction of Table Metadata from Digital Documents**339
Y. Liu, P. Mitra, C. L. Giles, K. Bai (*The Pennsylvania State University*)
- **A Tool for Teaching Principles of Image Metadata Generation**.....341
P. Achananuparp, K. W. McCain, R. B. Allen (*Drexel University*)
- **Evaluating the National Science Digital Library**342
M. Khoo (*NSDL Core Integration*)
- **Multi-linguistic Collaborative Distance Learning:
From Information Translation to Knowledge Translation**343
X. Mu (*University of Wisconsin at Milwaukee*)
- **Metadata Data Dictionary for Analog Sound Recordings**344
C. Lai, I. Fujinaga (*McGill University*)
- **Feasibility of Developing Curriculum Standards Metadata**345
R. T. Brown (*University of North Carolina at Chapel Hill*),
S. W. Bowers (*NASA Langley Research Center*)
- **On-demand Metadata Extraction Network (OMEN)**346
I. Fujinaga, D. McEnnis (*McGill University*)
- **Managing Intellectual Property Issues in a Commons of Geographic Data**.....347
J. Campbell, M. Lutz, D. McCurry, H. Onsrud, K. Williams (*University of Maine*)
- **Scientific Research Groups, Digital Libraries, & Education:
Metadata from Nanoscale Simulation Code**348
L. M. Bartolo, C. S. Lowe (*Kent State University*),
S. C. Glotzer, C. R. Iacovella (*University of Michigan*)
- **Interface Design for Browsing Faceted Metadata**.....349
J. Stuart-Moore, M. Evans, P. Jacobs (*Shodor Education Foundation*)
- **Developing a Metadata Schema for CSERD:
A Computational Science Digital Library**350
D. E. Swain (*North Carolina Central University*),
J. Wagy (*University of North Carolina at Chapel Hill*),
M. McClelland (*North Carolina Central University*), P. Jacobs (*Shodor Foundation*)
- **Creating a Multi Disciplinary Digital Library in the 5S Framework**.....351
M. Drutar, C. Coleman (*The SAS Institute Inc.*), E. Fox (*Virginia Tech*)
- **An Analysis of the Bid Behavior of the 2005 JCDL Program Committee**.....352
M. A. Rodriguez, J. Bollen, H. Van de Sompel (*Los Alamos National Laboratory*)

• Supporting Biological Information Work: Research and Education for Digital Resources and Long-lived Data	353
C. L. Palmer, M. H. Cragin, P. B. Heidorn (<i>University of Illinois at Urbana-Champaign</i>)	
• Finding a Metaphor for Collecting and Disseminating Distributed NSDL Content and Communications	354
C. M. Morris, H. Hembrooke, L. Rayle (<i>Cornell University</i>)	
• A Curated Harvesting Approach to Establishing a Multi-Protocol Online Subject Portal	355
R. Sanderson, J. Harrison, C. Llewellyn (<i>University of Liverpool</i>)	
• Incorporating Computational Science Activities in High School Algebra	356
J. DeLuca, D. A. Joiner (<i>Kean University</i>)	
• Adapting Peer Verification, Validation and Accreditation Processes for Digital Libraries	357
L. Schmalbeck, J. Stuart-Moore, M. Evans (<i>Shodor Education Foundation</i>)	
• Quantifying the Accuracy of Relational Statements in Wikipedia: A Methodology	358
G. Weaver, B. Strickland, G. Crane (<i>The Perseus Project</i>)	
• The Ingest and Maintenance of Electronic Records: Moving from Theory to Practice	359
K. L. Glick (<i>Yale University</i>), E. Wilczek, R. Dockins (<i>Tufts University</i>)	
• Technical Architecture Overview: Tools For Acquisition, Packaging and Ingest of Web Objects into Multiple Repositories	360
S. Rani, J. Goodkin, J. Cobb (<i>Online Computer Library Center</i>), T. Habing, R. Urban, J. Eke (<i>University of Illinois at Urbana-Champaign</i>), R. Pearce-Moses (<i>Arizona State Library</i>)	
• Browsing Affordance Designs for the Human-Centered Computing Education Digital Library	361
E. Clarkson, J. D. Foley (<i>Georgia Institute of Technology</i>)	
• Indexing Institutional Data to Promote Library Resource Discovery	362
T. Sierra (<i>North Carolina State University</i>)	
• Keeping the Context: An Investigation in Preserving Collections of Digital Video	363
C. A. Lee, H. R. Tibbo, D. Howard, Y. Song, T. Russell, P. Jones (<i>University of North Carolina at Chapel Hill</i>)	
• Cloudalicious: Folksonomy Over Time	364
T. Russell (<i>University of North Carolina at Chapel Hill</i>)	
• Apparatus and Methods for Production of Printed Aromatic and Gustative Information	365
B. P. Hyacinthe (<i>Florida State University</i>)	
• Teaching Box Builder: Customizing Pedagogical Contexts for Use of Digital Library Resources in Classrooms	366
H. J. Khan, K. E. Maul (<i>University of Colorado</i>)	
• ClaimID: A System for Personal Identity Management	367
F. Stutzman, T. Russell (<i>University of North Carolina at Chapel Hill</i>)	
• Pathways Core: A Data Model for Cross-Repository Services	368
J. Bekaert (<i>Ghent University</i>), X. Liu, H. Van de Sompel (<i>Los Alamos National Laboratory</i>), C. Lagoze, S. Payette, S. Warner (<i>Cornell Information Sciences</i>)	
• Pilot Testing the DigiQual™ Protocol: Lessons Learned	369
M. Kyrillidou, S. Giersch (<i>Association of Research Libraries</i>)	
• Using Citations for Ranking in Digital Libraries	370
B. Larsen, P. Ingwersen (<i>Royal School of Library and Information Science</i>)	

Demos

• LemiURL Web Link Analysis for Digital Libraries	371
A. Zuccala, M. Thelwall (<i>University of Wolverhampton</i>)	
• Exploring Content-Actor Paired Network Data Using Iterative Query Refinement with NetLens	372
H. Kang, C. Plaisant (<i>Maryland Institute for Advanced Computer Studies</i>), B. Lee, B. B. Bederson (<i>University of Maryland</i>)	
• A Content-based Video Browsing System Based on Visual Neighbor Similarity	373
X. Mu (<i>University of Wisconsin at Milwaukee</i>)	
• OpenArXiv = arXiv + RDBMS + Web Services	374
J. Fisher, H. Kil, D. Lee (<i>The Pennsylvania State University</i>)	
• Real-time Collaboration through Visual Search and Voice-over-IP	375
C. Hoare, H. Sorensen (<i>University of College Cork</i>)	
• Demonstrating the Use of a SenseCam in Two Domains	376
S. Yang, B. Congleton, G. Luc, M. A. Pérez-Quiñones, E. A. Fox (<i>Virginia Tech</i>)	
• SIMPEL: A Superimposed Multimedia Presentation Editor and Player	377
U. Murthy, K. Ahuja (<i>Virginia Tech</i>), S. Murthy (<i>Portland State University</i>), E. A. Fox (<i>Virginia Tech</i>)	
• Extended XQuery for Digital Libraries	378
A. Dekhtyar, I. E. Jacob, K. Kiernan, D. C. Porter (<i>University of Kentucky</i>)	
• ETANA-GIS: GIS for Archaeological Digital Libraries	379
D. Gorton, R. Shen, N. S. Vemuri, W. Fan, E. A. Fox (<i>Virginia Tech</i>)	
• MANGAS Infrastructure	380
H. Manguinhas, J. Borbinha (<i>INESC-ID - Instituto de Engenharia de Sistemas e Computadores</i>)	
• How Science Web Sites are Leveraging DLESE Search Web Services to Extend Value to their Users	381
L. Davies, J. Weatherley (<i>University Corporation for Atmospheric Research</i>)	
• Unsupervised Structure Discovery for Biodiversity Information	382
H. Cui (<i>University of Western Ontario</i>), R. M. McCourt (<i>Academy of Natural Sciences</i>), M. Feist (<i>Université Montpellier II</i>)	
• Visualizing an Enterprise Social Network from Email	383
W. Zhu, C. Chen, R. B. Allen (<i>Drexel University</i>)	
Author Index	384