Curriculum Vitae

JIE LU

12/01/2006

jielu@cs.cmu.edu
Tel: (412) 268-9515
Fax: (412) 268-6298
http://www.cs.cmu.edu/~jielu

Language Technologies Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

RESEARCH INTEREST

My general research interests lie in the use and development of information retrieval and machine learning techniques for effective and efficient adaptive information access and mining. I am particularly interested in content management and search in distributed environments such as enterprise networks, peer-to-peer file-sharing networks, and social networks. My other areas of interest include personalized search and domain-dependent (e.g., legal, medical) or task-specific (e.g., retail, travel) information processing and knowledge discovery.

EDUCATION

Ph.D., **Language and Information Technologies**, expected May 2007
Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
Advisor: Jamie Callan
Thesis committee members: Jamie Callan (Chair), Jaime Carbonell, Christos Faloutsos, Norbert Fuhr
Ph.D. Dissertation: Full-text federated search in peer-to-peer networks

Master of Science, **Language Technologies**, May 2004
Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

Master of Science, **Computer Science**, May 2000
Department of Computer Science, Yale University, New Haven, CT, USA

Bachelor of Science and Engineering, **Computer Science**, July 1998
Department of Computer Science and Technology, Tsinghua University, Beijing, China

PUBLICATIONS

Journal Paper

- Jie Lu and Jamie Callan (2006)
  Full-text federated search of text-based digital libraries in peer-to-peer networks.
  Journal of Information Retrieval, Volume 9, Number 4, 2006.

Conference Papers

- Jie Lu and Jamie Callan (submitted)
  Content-based peer-to-peer network overlay for full-text federated search.
  8th RIAO Conference on Large-Scale Semantic Access to Content (RIAO '07).

- Jie Lu and Jamie Callan (2006)
  User modeling for full-text federated search in peer-to-peer networks.
  29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06).

- Luo Si, Jie Lu and Jamie Callan (2006)
  Combining multiple resources, evidences and criteria for genomic information retrieval.
  15th Text REtrieval Conference (TREC '06).

- Jie Lu and Jamie Callan (2005)
  Federated search of text-based digital libraries in hierarchical peer-to-peer networks.
  27th European Conference on Information Retrieval Research (ECIR '05).

- Jie Lu and Jamie Callan (2004)
  Merging retrieval results in hierarchical peer-to-peer networks.
  27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '04).

- Jie Lu and Jamie Callan (2003)
  Content-based retrieval in hybrid peer-to-peer networks.
  12th ACM International Conference on Information and Knowledge Management (CIKM '03).

- Jie Lu and Jamie Callan (2003)
  Reducing storage costs for federated search of text databases.
  3rd Annual National Conference on Digital Government Research (dg.o '03).

- Luo Si, Jie Lu and Jamie Callan (2003)
  Distributed information retrieval with skewed database size distributions.
  3rd Annual National Conference on Digital Government Research (dg.o '03).

- Jie Lu and Jamie Callan (2002)
  Pruning long documents for distributed information retrieval.
  11th ACM International Conference on Information and Knowledge Management (CIKM '02).

Workshop and Technical Reports

- Jie Lu (2005)
  Full-text federated search in peer-to-peer networks.
  CMU-LTI-05-197, Language Technologies Institute, Carnegie Mellon University.

- Jie Lu and Jamie Callan (2004)
  Federated search of text-based digital libraries in hierarchical peer-to-peer networks.
  Workshop on Peer-to-Peer IR at the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '04).

Datasets

- Jie Lu and Jamie Callan (2003, 2006)
  Peer-to-peer testbed definitions.
  http://www.cs.cmu.edu/~callan/Data/

RESEARCH EXPERIENCE

- Research Assistant, Carnegie Mellon University, September 2001 – Present
  Federated search in distributed and peer-to-peer environments using an integrated framework of network overlay, search, and evolution models to provide effective and efficient full-text ranked retrieval with the support of dynamic self-organization in contents and connections.
  Genomic information retrieval combining multiple resources, evidences, and criteria for query expansion, retrieval and result ranking.

- Research Assistant, Yale University, June 1999 – May 2000
  Three-dimensional model construction from two-dimensional images under various illumination conditions.

RECENT PROJECTS

- Research project, Carnegie Mellon University (in collaboration with the University of Duisburg-Essen, Germany), January 2003 – Present
  PEPPER: Peer-to-peer architectures for federated search of complex digital libraries.
  http://www.is.informatik.uni-duisburg.de/projects/pepper/index.html.en

- Research project, Carnegie Mellon University (in collaboration with Purdue University), May 2006 – July 2006
  TREC 2006 Genomics Track (the 3rd ranked group in performance among 30 participating groups).

- Course project, Dynamic Organizations and Networks, Carnegie Mellon University, Fall 2004
  Using regular equivalence to measure similarity for collaborative filtering.

- Course project, Advanced IR Seminar and Lab, Carnegie Mellon University, Fall 2003
  Duplicate detection in large public comment datasets.

TEACHING EXPERIENCE

- Teaching Assistant, Information Retrieval, Carnegie Mellon University, Spring 2006

- Teaching Assistant, Human Language Technologies, Carnegie Mellon University, Fall 2005

- Teaching Assistant, Introduction to Programming, Yale University, Fall 1999

PROFESSIONAL ACTIVITIES

- Program committee member for the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2007.

- Referee for European Conference on Information Retrieval Research (ECIR), 2007.

- Referee for Human Language Technology Conference (NAACL-HLT), 2007.

- Organizing committee member for the Workshop on Information Retrieval in Peer-to-Peer Networks (P2PIR 2006) at the 15th ACM International Conference on Information and Knowledge Management (CIKM '06), 2006.

- Referee for ACM Transactions on Information Systems (TOIS), 2004 – 2006.

- Referee for Journal of Computer Science and Technology (JCST), 2006.

- Referee for the ACM International Conference on Information and Knowledge Management (CIKM), 2004.

- Invited talk at Baidu.com, Inc., Beijing, P.R.China, April 2005.
  Federated search in distributed and peer-to-peer environments.

- Invited talk at Department of Computer Science and Technology, Tsinghua University, Beijing, P.R.China, March 2005.
  Research at the Language Technologies Institute and the Distributed Information Retrieval Group.

- Invited talk at Department of Informatics and Interactive Systems, University of Duisburg-Essen, Duisburg, Germany, March 2004.
  Federated search of text-based digital libraries in hierarchical peer-to-peer networks.

- Invited talk at NEC Labs China, Beijing, P.R.China, December 2003.
  Content-based retrieval in hybrid peer-to-peer networks.

AWARDS & FELLOWSHIPS

- August 2001 – Present, Research Fellowship, Carnegie Mellon University
- June 1999 – May 2000, Research Scholarship, Yale University
- August 1998 – May 1999, University Fellowship, Yale University
- September 1996 – July 1998, HSBC Scholarship, Tsinghua University
- September 1993 – July 1996, First Class Scholarship for Outstanding Undergraduate, Tsinghua University

PROFICIENCY

- C, C++, Java, Perl, MATLAB
- Windows, UNIX, Linux