

Department of Computer Science  
University of Illinois at Urbana-Champaign  
201 N Goodwin Ave  
Urbana, IL 61801, USA

office: +1 (217)244-4026  
cell: +1 (217)778-7649  
email: jiang4@cs.uiuc.edu  
www: sifaka.cs.uiuc.edu/jiang4

# Jing Jiang

## RESEARCH INTERESTS

Natural Language Processing, Information Retrieval, Information Extraction, Machine Learning, Biomedical Text Mining.

## EDUCATION

- 2003–2008 (expected)      **University of Illinois at Urbana-Champaign**  
Ph.D. Candidate, Computer Science. Advisor: Prof. ChengXiang Zhai.
- 2002–2003                      **Stanford University**  
M.S., Computer Science.
- 1999–2002                      **Stanford University**  
B.S. (with Distinction), Computer Science. Minor in Mathematics.

## RESEARCH EXPERIENCE

- Aug 2003–Present              **Department of Computer Science**  
**University of Illinois at Urbana-Champaign**  
Ph.D. Candidate. Advisor: Prof. ChengXiang Zhai.
- Investigated the problem of adapting statistical classifiers to new domains in natural language processing. Proposed two frameworks based on instance weighting and feature selection. Evaluated the frameworks on a number of tasks including named entity recognition, part-of-speech tagging and spam filtering with promising results.
  - Designed a unified graphic feature representation for relation extraction from text. Systematically explored a large spectrum of features.
  - Investigated the use of hidden Markov models for automatically extracting relevant passages of variable length from text.
- Jan 2005–May 2007              **Institute for Genomic Biology**  
**University of Illinois at Urbana-Champaign**  
Research Assistant. Advisors: Prof. ChengXiang Zhai, Prof. Bruce Schatz.
- Developed a maximum entropy Markov model based gene and protein name recognizer. The program is now integrated into BeeSpace, an NSF funded project.
  - Studied the effectiveness of various kinds of tokenization strategies for biomedical information retrieval. Implemented a tokenizer for biomedical text.
  - Helped develop an automatic gene summarization system using various information retrieval and information extraction techniques.
- May 2005–Jul 2005              **School of Computing**  
**National University of Singapore**  
Research Intern. Supervisor: Prof. Tat-Seng Chua.

## RESEARCH EXPERIENCE (continued)

- Investigated the use of semantic role labels for question answering. Designed and implemented a semantic passage re-ranking module using a semantic parser and WordNet. The module was incorporated into the NUS QA system for TREC 2005.

Mar 2002–Dec 2002

**Department of Computer Science  
Stanford University**

Research Assistant. Advisors: Dr. Chris Olston, Prof. Jennifer Widom.

- Built a simulated distributed network traffic monitoring system. Implemented Chris Olston’s adaptive filtering algorithm, and tested its effectiveness in the simulated system.
- Built a Java-based user interface for the purpose of visualizing the adaptive filtering algorithm.

## TEACHING EXPERIENCE

Jun 2007–Jul 2007

**Department of Computer Science  
University of Illinois at Urbana-Champaign**

Student mentor for the Data Science Summer Institute of the Multimodal Information Access and Synthesis Center. Supervised seven undergraduate and master students to develop a named entity recognition tool with a graphic user interface.

Aug 2003–Dec 2004

**Department of Computer Science  
University of Illinois at Urbana-Champaign**

Teaching assistant for “Introduction to Artificial Intelligence” for three semesters. Designed homework assignments and machine problems, held office hours, and graded homework and exams.

## PUBLICATIONS

### REFEREED JOURNAL ARTICLES

1. X. Ling, **J. Jiang**, X. He, Q. Mei, C. Zhai and B. Schatz, “Generating gene summaries from biomedical literature: A study of semi-structured summarization,” *Information Processing & Management (IP&M)*, 43(6):1777-1791, November 2007.
2. **J. Jiang** and C. Zhai, “An empirical study of tokenization strategies for biomedical information retrieval,” in *Information Retrieval*, 10(4–5):341–363, October 2007.
3. **J. Jiang** and C. Zhai, “Extraction of coherent relevant passages using hidden Markov models,” *ACM Transactions on Information Systems (TOIS)*, 24(3):295-319, July 2006.

### REFEREED CONFERENCE PAPERS

1. **J. Jiang** and C. Zhai, “A two-stage approach to domain adaptation for statistical classifiers,” in *Proceedings of the ACM 16th Conference on Information and Knowledge Management (CIKM)*, pages 401–410, 2007. (17% acceptance)
2. **J. Jiang** and C. Zhai, “Instance weighting for domain adaptation in NLP,” in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 264–271, 2007. (22% acceptance)
3. **J. Jiang** and C. Zhai, “A systematic exploration of the feature space for relation extraction,” in *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, pages 113–120, 2007. (24% acceptance)

## PUBLICATIONS (continued)

4. **J. Jiang** and C. Zhai, “Exploiting domain structure for named entity recognition,” in *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pages 78–81, 2006. (25% acceptance)
5. X. Ling, **J. Jiang**, X. He, Q. Mei, C. Zhai and B. Schatz, “Automatically generating gene summaries from biomedical literature,” in *Proceedings of the Pacific Symposium on Biocomputing (PSB)*, pages 40–51, 2006.
6. **J. Jiang** and C. Zhai, “Accurately extracting coherent relevant passages using hidden Markov models,” in *Proceedings of the ACM 14th Conference on Information and Knowledge Management (poster session) (CIKM)*, pages 289–290, 2005.
7. C. Olston, **J. Jiang** and J. Widom, “Adaptive filters for continuous queries over distributed data streams,” in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 563–574, 2003. (15% acceptance)

## OTHER PUBLICATIONS

1. **J. Jiang**, X. He and C. Zhai, “Robust pseudo feedback estimation and HMM passage extraction: UIUC at TREC 2006 Genomics Track,” in *Proceedings of the 15th Text REtrieval Conference (TREC)*, (selected for oral presentation), 2007.
2. **J. Jiang** and C. Zhai, “An empirical study of tokenization strategies for biomedical information retrieval,” technical report UIUCDCS-R-2006-2733, University of Illinois at Urbana-Champaign, May 2006.
3. R. Sun, **J. Jiang**, Y-F. Tan, H. Cui, T-S. Chua and M-Y. Kan, “Using syntactic and semantic relation analysis in question answering,” in *Proceedings of the 14th Text REtrieval Conference (TREC)*, 2006.
4. **J. Jiang** and C. Zhai, “UIUC at HARD 2004—passage retrieval using HMMs,” in *Proceedings of the 13th Text REtrieval Conference (TREC)* (selected for oral presentation), 2005.

## INVITED TALKS

“Machine Learning in Information Extraction”, Nanyang Technology University, Singapore, July 2007.  
“Two Perspectives on Domain Adaptation in Natural Language Processing,” National University of Singapore, Singapore, July 2007.

## AWARDS AND HONORS

Recipient of the ACL Don and Betty Walker International Student Fund, 2007.  
Member of Tau Beta Pi, the Engineering Honor Society.  
Recipient of the Keith Van Patten Scholarship, Stanford University, 1999–2002.

## PROFESSIONAL SERVICES

Program Committee Member: AIRS 2008, NLDB 2008  
Reviewer: EMNLP-CoNLL 2007, TOIS  
External reviewer: CIKM 2005, ICML 2006, CIKM 2006

## REFERENCES

Available upon request.