

James J. Fan

T. J. Watson Research Lab
IBM Corp.
19 Skyline Dr.
Hawthorne, NY 10532

Phone: 914.784.7161
Cell: 512.784.4028
fanj@us.ibm.com
<http://www.cs.utexas.edu/~jfan/>

Research Interests

Knowledge based systems, natural language processing, question answering, web services, machine learning, data mining and neural evolution.

Education

Ph.D. in Computer Sciences, 2006
University of Texas, Austin, TX
Thesis: Interpreting Loosely Encoded Knowledge
Advisor: Professor Bruce Porter

M.S. in Computer Sciences, 2001
University of Texas, Austin, TX

B.S. in Computer Sciences, *highest honor*, 1996
University of Texas, Austin, TX

Research Experience

IBM T. J. Watson Research Research Staff Member	Hawthorne, NY 2006–now
--	---------------------------

University of Texas at Austin Graduate Research Assistant As a graduate student, I contributed to several projects in artificial intelligence at the University of Texas at Austin. 1999-present.	Austin, TX 2000–2006
---	-------------------------

Early in my graduate study, I developed an open-source software, *MC*, that rapidly creates vector-space models from text documents to be used for text mining application. *MC* is available from GNU software directory (<http://directory.fsf.org/>), and it is described in detail in *Data Mining for Scientific and Engineering Applications*.

I've also explored neural networks and genetic algorithms under the guidance of professor Risto Miikkulainen. I co-developed *RESP*, a neuroevolution algorithm that utilizes prior symbolic knowledge [*ICML 2003*].

My dissertation work is done under the supervision of professor Bruce Porter. During the Rapid Knowledge Formation project, I designed the representational scheme for role concepts [*KCAP 2001*]. I've identified the problem of loose speak, the misalignments between a user unfamiliar with the ontology and the idiosyncratic requirements of the ontology, and

I designed and implemented an interpreter for loose speak [IJCAI 2003, AAAI 2004]. The loose-speak interpreter has been used in Project Halo and other applications [KCAP 2005], and it is expected to be used in other future knowledge base system projects.

Industry Experience

Dell Inc. Roundrock, TX
Senior Programmer Analyst 1997–2000
Participated in the development of a multi-tiered customer relation management software that replaced the legacy mainframe system. Supervisor: Emilio Griman.

Macromedia Inc. Richardson, TX
Software Engineer 1996–1997
Developed text tools for Japanese version of FreeHand 7, a vector graphics tool for desktop publishing professionals. Supervisor: Keith Bottner.

Papers in Refereed Conferences

- J. Dolby, J. Fan, A. Fokoue, A. Kalyanpur, A. Kershenbaum, L. Ma, W. Murdock, K. Srinivas and C. Welty, *Scalable Cleanup of Information Extraction Data Using Ontologies* In Proceedings of the sixth International Semantic Web Conference (ISWC 2007)
- P. Clark, S. Chaw, K. Barker, V. Chaudhri, J. Fan, J. Thompson, P. Harrison, B. John, B. Porter, A. Spaulding and P. Yeh *Capturing and Answering Questions Posed to a Knowledge-Based System* In Proceedings of the Fourth International Conference on Knowledge Capture (KCAP 2007)
- S. Chaw, J. Fan, D. Tecuci and P. Yeh *Capturing a Taxonomy of Failures During Automatic Interpretation of Questions Posed in Natural Language* In Proceedings of the Fourth International Conference on Knowledge Capture (KCAP 2007)
- K. Barker, B. Agashe, S. Chaw, J. Fan, N. Friedland, M. Glass, J. Hobbs, E. Hovy, D. Israel, D. S. Kim, R. Mulkar, S. Patwardhan, B. Porter, D. Tecuci and P. Yeh. *Learning by Reading: A Prototype System, Performance Baseline and Lessons Learned* Proceedings of the Twenty-Second National Conference on Artificial Intelligence 2007 (AAAI 2007).
- J. Fan, K. Barker and B. Porter. *Indirect Anaphora Resolution as Semantic Path Search* Proceedings of the Third International Conference on Knowledge Capture 2005 (KCAP 2005).
- J. Fan and B. Porter. *Interpreting Loosely Encoded Questions* Proceedings of the Nineteenth National Conference on Artificial Intelligence 2004 (AAAI 2004).
- K. Barker, S. Chaw, J. Fan, B. Porter, D. Tecuci, P. Yeh, V. Chaudhri, D. Israel, S. Mishra, P. Romero, P. Clark. *A question-answering system for AP Chemistry: Assessing KR&R Technologies* Proceedings of the Ninth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2004).
- J. Fan, K. Barker and B. Porter. *The Knowledge Required to Interpret Noun Compounds* Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence 2003 (IJCAI 2003).

- J. Fan, R. Lau and R. Miikkulainen *Utilizing Domain Knowledge in Neuroevolution* Proceedings of the Twentieth International Conference on Machine Learning 2003 (ICML 2003)
- J. Fan, K. Barker, B. Porter and P. Clark. *Representing Roles and Purpose* Proceedings of First International Conference on Knowledge Capture 2001 (KCAP 2001)

Papers in Magazines or Journals

- N. Friedland, P. Allen, P. Matthews, M. Witbrock, D. Baxter, J. Curtis, B. Shepard, P. Miraglia, J. Angele, S. Staab, E. Moench, H. Oppermann, D. Wenke, D. Israel, V. Chaudhri, B. Porter, K. Barker, J. Fan, S. Chaw, P. Yeh, D. Tecuci & P. Clark. *Project Halo: Towards a Digital Aristotle* AI Magazine Winter 2004.
- A. Zuluaga, U. Utzinger, A. Durkin, H. Fuchs, J. Fan, A. Gillenwater and R. Richards-Kortum. *Fluorescence Excitation-Emission Matrices of Human Tissue: A system for in-vivo measurement and method of data analysis* Applied Spectroscopy, Vol. 53, No. 3 (March 1999) pp 302-311
- J. Fan. *Implementing a Request Distribution Server with DCOM and ATL C/C++* Users Journal, June 1999

Book Chapter

- C. Welty, J. W. Murdock and J. Fan *SemanticClean: Cleaning Noisy Data Using Semantic Technology* In Language Technology, Ontologies and the Semantic Web, N. Ide, D. Tufis and D. Cristea (Eds), Springer, forthcoming
- I. Dhillon, J. Fan and Y. Guan. *Efficient Clustering of Very Large Document Collections* In Data Mining for Scientific and Engineering Applications, R. Grossman, C. Kamath, P. Kegelmeyer, Vipin Kumar, Raju R. Namburu (Eds), Kluwer Academic Publishers, 2000

Workshop and Symposium

- B. Andrews, J. Fan, J. W. Murdock and C. Welty. *COALA: A Tool for Inter-Document Coreference Resolution Evaluation* In AAAI Spring Symposium Series, 2007
- B. Porter, K. Barker, J. Fan, P. Navratil, D. Tecuci and P. Yeh. *Mining Answers from Text and KBs: Our Position* In AAAI Spring Symposium Series, 2002

Tech Report

- J. Fan, K. Barker and B. Porter. *The Knowledge Required to Interpret Noun Compounds* UTCS Technical Report UT-AI-TR-03-301

Awards and Honors

- 2003: Travel scholarship, The Twentieth International Conference on Machine Learning.
- 2001: Travel grant, First International Conference on Knowledge Capture.
- 1996: Member of Phi Beta Kappa Honor Society.
- 1996: Bachelor of Science with highest honor and special departmental honor.

Professional Activities

- **Program committee member:** The Fourth International Conference on Knowledge Capture 2007 (KCAP 2007)
- **Program committee member:** The Twenty-Second National Conference on Artificial Intelligence (AAAI 2007)
- **Reviewer:** International Journal of Computational Intelligence Theory and Practice.
- **Reviewer:** IEEE Transactions on Systems, Man, and Cybernetics.
- **Reviewer:** The First International Conference on Knowledge Capture.
- **Member:** American Association of Artificial Intelligence.
- **Software:** open-source software, *MC*, that creates vector-space models from text documents that can be used for text mining application. Available at <http://www.cs.utexas.edu/users/jfan/dm>

Personal Information

- **Citizenship:** USA
- **Languages:** English and mandarin Chinese

References

Prof. Bruce Porter, Professor
Department of Computer Sciences
University of Texas at Austin
1 University Station C0500
Austin, TX 78712-0233
porter@cs.utexas.edu
512.471.9565

Dr. Peter Clark, Research Scientist
Boeing Research and Technology
PO Box 3707, m/s 7L-66
Seattle, WA 98124-2207
peter.e.clark@boeing.com
425.865.4304

Dr. Chris Welty, Research Staff Member
IBM Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
welty@watson.ibm.com
914.784.7055

Prof. Risto Miikkulainen, Professor
Department of Computer Sciences
University of Texas at Austin
1 University Station C0500
Austin, TX 78712-0233
risto@cs.utexas.edu
512.471.9571