# David A. Cohn

**Contact information:**
1347 Malvern Avenue
Pittsburgh, PA 15217

email: David.Cohn@acm.org
web: http://www.cs.cmu.edu/~cohn
phone: (412) 654-3070

## Expertise

*Information Discovery and Retrieval* – learning, navigating and manipulating structure and information in unstructured or semi-structured document bases.

## Patents

A Knowledge-based Information Retrieval System with H. Fujisawa, A. Hatakeyama, and I. Kiuchi: US5555408, US5404506. **Japanese patents** 63-2609, 62-297568

## Experience

**2001-present: Chief Technology Officer, DigitalMC, Pittsburgh, PA.**

DigitalMC performs digital signal analysis and datamining for a variety of products in areas including audio fingerprinting, audio content retrieval, listener preference modeling, music recommendation, and statistical market research.

- Developed core algorithms for audio fingerprinting, audio retrieval and music recommendation (patents pending).

- As chief architect and acting manager of technical staff, directed development of product offerings by team of eight researchers and software developers.

- Identified, evaluated and pursued technical opportunities for DigitalMC in digital content delivery, media indexing and retrieval markets.

**2000-2001: Principal Scientist, Burning Glass Technologies, Pittsburgh, PA.**

Burning Glass Technologies is a leading provider of information extraction and predictive analytics software infrastructure for the Human Capital Management market, applying statistical natural language processing and information retrieval to predict the degree of match between a job candidate and the requirements of a position.

- Lead development of probabilistic model for inferring skills and requirements from stated skills/requirements in job postings and resumés.

- Developed joint probabilistic model for mapping resumése ("facilitated workplace hygiene") and postingese ("must be neat") to common conceptual framework.

- Led development of techniques for learning taxonomies of competencies from a corpus of examples, adaptive spelling correction and identification of canonical forms from free text.

**1999-2000: Research Scientist, [Justsystem Pittsburgh Research Center](#), Pittsburgh, PA.**

Research arm of [Justsystem Corporation](#), a leading vendor of document processing, knowledge management and information retrieval software in Japan.

- Developed algorithms for learning to find answers to natural language queries in large matched question-answer corpora.

- Developed techniques for learning, navigating and manipulating the structure of otherwise unstructured document bases.

- Developed algorithms for efficiently generating personalized models of document authority and relevance from few examples.

- Developed probabilistic bibliometrics, and the first published algorithm for reasoning about document contents and links using a unified probabilistic model. Applications in adaptive web spidering, clustering, cross-language retrieval and dynamic hypertext generation.

**1995-1998: Senior Research Scientist, Adaptive Systems Group, [Harlequin, Inc.](#), Cambridge, MA and Menlo Park, CA.**

Acquired by [Global Graphics](#); a provider of digital pre-press document workflow solutions, software tools, and information management and visualization systems (spun off as [Xanalys](#)).

- Core Architect on adaptive document workflow optimization system for digital pre-press industry. Led design of scheduling and resource estimation components, which used machine learning and real-time adaptive scheduling. Managed project involvement of 5 Ph.D.-level members of the Adaptive Systems Group.

- Initiated and served as Lead Designer on project to develop document/information management and retrieval system. Designed, implemented and demonstrated several simpler prototype document retrieval/clustering systems as testbeds.

- Invented, and led design and implementation of adaptive memory management system for dynamic memory allocation.

**1989: [IBM Thomas Watson Research Center](#), Yorktown Heights, NY.**

**1988: Visiting Researcher, [Hitachi Central Research Laboratory](#), Tokyo.**

**1986: [Research Institute for Advanced Computer Science](#), NASA Ames, Moffett Field, CA.**

## Academic Summary

**Adjunct Research Scientist, Carnegie Mellon University** - Robotics Institute (1999-present)

**Postdoctoral Associate, Massachusetts Institute of Technology** - Center for Biological and Computational Learning (1992-1995)

**Ph.D., University of Washington** - Dept. of Computer Science - (1992) [Separating Formal Bounds from Practical Performance in Learning Systems](#)

**A.B., Dartmouth College** - Computer Science and Physics (1985). [Creating `Havoc': a threaded, compiled, real-time laboratory-oriented programming language](#)

## Publications

**Recent Research (available at http://www.cs.cmu.edu/~cohn/papers.html)**

- David Cohn and Thomas Hofmann (2001). The Missing Link - A Probabilistic Model of Document Content and Hypertext Connectivity, in T. Leen et al, eds, Advances in Neural Information Processing Systems 13.

- Adam Berger, Rich Caruana, David Cohn, Dayne Freitag, and Vibhu Mittal (2000). Bridging the lexical chasm: Statistical approaches to answer-finding, Proceedings of the 23rd Annual Conference on Research and Development in Information Retrieval (ACM SIGIR). Athens.

- David Cohn and Huan Chang (2000). Probabilistically Identifying Authoritative Documents, Proceedings of the Seventeenth International Conference on Machine Learning. Stanford, CA.

- Huan Chang, David Cohn and Andrew McCallum (2000). Creating Customized Authority Lists, Proceedings of the Seventeenth International Conference on Machine Learning. Stanford, CA.

- Greg Schohn and David Cohn (2000). Less is More: Active Learning with Support Vector Machines, Proceedings of the Seventeenth International Conference on Machine Learning.

- Brigham Anderson, Andrew Moore and David Cohn (2000). A Nonparametric Approach to Noisy and Costly Optimization, Proceedings of the Seventeenth International Conference on Machine Learning. Stanford, CA.

**Journal and Book Publications**

- Michael Kearns, Sara Solla and David Cohn, eds. (1999). Advances in Neural Information Processing Systems 11, MIT Press.

- David Cohn, Zoubin Ghahramani, and Michael Jordan. (1997). Active learning with mixture models, in R. Murray-Smith and T. Johansen, eds., Multiple Model Approaches to Modeling and Control, Taylor and Francis, London.

- David Cohn, Zoubin Ghahramani, and Michael Jordan. (1996). Active learning with statistical models, *Journal of Artificial Intelligence Research*, (4): 129-145.

- David Cohn. (1996). Neural network exploration using optimal experiment design, *Neural Networks* (9)6: 1071-1083. Available online as AI Lab Memo 1491.

- David Cohn, Les Atlas and Richard Ladner. (1994) Improving generalization with active learning, *Machine Learning* 15(2):201-221.

- David Cohn, Eve Riskin and Richard Ladner. (1994) The theory and practice of vector quantizers trained on small training sets, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(1):54-65.

- D. Cohn and G. Tesauro. (1992) How tight are the Vapnik-Chervonenkis bounds? *Neural Computation* 4(2):249-269.

**Other Representative Publications**

- Satinder Singh and David Cohn. (1998) How to dynamically merge Markov decision processes, in M Jordan et al, eds, Advances in Neural Information Processing Systems 10.

- Satinder Singh, Peter Norvig and David Cohn. (1997) Agents and Reinforcement Learning, *Dr. Dobb's Journal* March 1997.

- Peter Norvig and David Cohn. (1997) Adaptive Software, PC AI Magazine, Jan 1997.

- David Cohn. (1997) Minimizing Statistical Bias with Queries, in M. Mozer et al., eds, Advances in Neural Information Processing Systems 9. Appears as AI Lab Memo 1552.

- David Cohn and Satinder Singh. (1997) Predicting lifetimes in dynamically allocated memory, in M. Mozer et al, eds, Advances in Neural Information Processing Systems 9.

- Stephen Atkins, William Hall, Paul DeBitetto and David Cohn (1995) The MIT/Draper Laboratory Autonomous Helicopter, Technical Report to the 1995 International Aerial Robotics Competition. Award for best paper.

- David Cohn. (1994) Robot Learning: Exploration and Continuous Domains (workshop summary) in J. Cowan et al, eds, Advances in Neural Information Processing Systems 6.

- D. Cohn, L. Atlas, R. Ladner, M. El-Sharkawi, R. Marks, M. Aggoune and D. Park. (1990) Training Connectionist Networks with Queries and Selective Sampling, in D. Touretzky, ed., Advances in Neural Information Processing Systems 2.

- M. Aggoune, L. Atlas, D. Cohn, M. Damborg, M. El-Sharkawi and R. Marks. (1989) Artificial Neural Networks for Power System Static Security Assessment, IEEE Proceedings, International Symposium on Circuits and Systems.

- D. Cohn, H. Fujisawa and I. Kiuchi. (1988) The Use of 'Familiarity' in Semantic Interpretation, Proceedings of Japanese Information Processing Society.

## Professional Activities

- **Cofounder and Managing Editor,** Journal of Machine Learning Research

- **Reviewer for** Journal of Artificial Intelligence Research, Machine Learning, Neural Computation, IEEE Transactions on Computers, IEEE Transactions on Pattern Analysis and Machine Intelligence, Neural Information Processing Systems, AAAI, IJCAI, SIGIR, ICML

**System Proficiencies**: C/C++, Perl, Python, Tk, Linux, Windows, MySQL, HTML, XML
**Language Proficiencies**: English, French, Japanese