

# Jan Pedersen

## Objective

Bring new information access technologies to product

## Education

Phd in Statistics                      **Stanford University** 1990

BA in Statistics                      **Princeton University** 1981

## Professional experience

2009 –                                      **Microsoft**                                      Mountain View, Ca

### **Chief Scientist Core Search**

I have recently joined Microsoft's Online Services Division in the role of Chief Scientist for Core Search.

2008 – 2009                                      **A9 an Amazon Company**                                      Palo Alto, Ca

### **Chief Scientist**

I worked briefly at A9 between Yahoo and Microsoft. A9 provides product search for Amazon.com, the foremost e-commerce site. While I was at A9, I participated in the development of the 2009 operating plan for both product search and A9's ClickRiver performance marketing product. In addition, with a few key architects, I articulated a self-service architecture for the A9 search technology that leveraged Amazon's EC2 and S3 utility computing infrastructure and developed a position paper describing A9 quality measurement and improvement strategy.

2003 – 2008                                      **Yahoo! Inc**                                      Sunnyvale, Ca

Yahoo! is the foremost Internet portal and one of the original Internet giants. Beginning in late 2002 Yahoo! re-entered the search technology space by first purchasing Inktomi (Dec 2002), a high-quality OEM algorithmic search service, and then purchasing Overture (Oct 2003), the founder of sponsored search advertising. Overture brought with it two additional search engines: Alta Vista and FAST Internet (Trondheim Norway). The search engines (and the teams) were integrated to form what is now called YST (Yahoo! Search Technology).

### **Chief Scientist and VP, Search and Advertising Technology Group**

- Initiated work on marketplace simulation and auction design for the sponsored search product. Facilitated the hiring of economists and other experts to deepen our understanding of this technology area. Co-chaired the Marketplace Design Group who defined the marketplace rules for the Panama relaunch of the sponsored search product.
- Co-developed Strategy for the search team: how to grow search share in the face of a strong category leader with a powerful brand. Suggested shift in focus from commoditize and distribute to differentiation.

- Initiated Next Generation Search program focused on intent-based search. The program combines higher level query analysis, rich content analysis and multi-phase ranking to sharpen results for specific user intents.

**Chief Scientist, Search and Marketplace**

- Instituted the Relevance group, a science team devoted to search algorithm development, from the various advanced development groups brought in by acquisition.
- Scaled the Relevance group from 20 to 90 scientists and engineers.
- Lead the development of several key Web Search Technologies:
- Machine Learned Ranking: a methodology for disciplined continuous improvement of search relevance ranking
- Query Speller: a text mining system that derives from the analysis of query logs very accurate run-time spelling corrections
- Guided the development of technology roadmaps for the five Relevance groups and participated in the development of search mission and strategy as part of the senior management team.

2002 – 2003

**AltaVista**

Palo Alto, Ca

AltaVista was at one point search share leader but its fortunes rapidly declined after the bursting of the internet bubble. I joined AltaVista as part of the turn-around team (with Jim Barnett and John Ellis). The goal was to stabilize search share and solidify the search technology and search operations in order to find a suitable buyer. This was accomplished when AltaVista was purchased by Overture in 2003. Overture was in turn purchased by Yahoo! a few months later

**Chief Scientist**

- As the most senior technical contributor, I worked closely with the VP engineering (John Ellis) to set technical direction for the staff of 100 engineers.
- Helped establish and guide an advanced development group of 10 scientists who acted as a reservoir of specialized skills for the engineering organization.
- Chaired a Scientific Advisory Board including Jerry Friedman (Stanford), Hector Garcia Molina (Stanford) and Marti Hearst (UC Berkeley)
- Participated as a member of the senior team in setting company strategy. In particular participated in the acquisition and disclosure discussions with Overture and Yahoo!
- Lead the AltaVista IP effort and initiated the filing of several additional patents covering ranking and other search technologies.

2000 – 2002

Various Startups

**Chief Scientist: Enkata Systems (2002)**

Enkata is a CRM analytic firms who specializes in deriving actionable insights from patterns in customer interactions. We developed Text mining technology for use as derived attributes.

**Engineering Director: Centrata (2001)**

Centrata is a KP-backed startup whose original business plan was to build a p2p

infrastructure platform. I was hired to build an Internet search application on this platform. Centrata's business plan shifted in January 2001 to Datacenter process automation.

**VP Engineering: Open Grid (2000)**

OpenGrid was a Motorola-backed startup developing an Internet-based application sharing technology similar to Zaplets. Unfortunately, the attempted extension of this technology to the wireless Internet was premature.

1998 – 2000 (Jan)                      **Infoseek/Go Network**                      Sunnyvale, Ca  
Infoseek was one of the first wave Internet search engine companies. I joined Infoseek after it had gone public and experienced its transformation into the Go Network subsequent to the Disney acquisition. I had two roles at Infoseek: the first was Advanced Technology Director reporting to Steve Kirsch the Infoseek founder and Chairman. The second was Director of the core Internet Search and Spidering service for Go Network.

**Director, Advanced Technology**

- Developed and prototyped several approaches to economically scaling the Infoseek Search Service index through distributed spidering and search. Transferred the resulting technology (code name BFI) to the Search Service organization.

**Director, Search and Spidering**

- Subsequent to the Disney acquisition of Infoseek, I was responsible for design, engineering, product management and operations of the core Infoseek Search Service within Go Network; an Internet product with an annual budget of \$6M, \$40M in revenues and 5.3Billion page views.
- Managed four groups with a total staff of 20 and an annual budget of \$6M.

1996 - 1998                              **Verity Inc**                              Sunnyvale, Ca  
Verity is the leading vendor of text retrieval software toolkits. I had two roles at Verity: I was hired as manager of the Advanced Technology group. Subsequently I was director of the Server Products group.

**Manager, Advanced Technology Group**

- Managed a group of 5 phd-level engineers. Responsible for the integration of new component technologies into the core Verity search engine product: clustering, summarization and QBE
- Managed the creation of the Knowledge Organizer (Yahoo! in a box) text categorization product concept.

**Director, Server Group**

- Managed 12 engineers. Responsible for the Information Server/Agent Server products: release 3.1 and service packs, rearchitected spider, rearchitected push component.
- Responsible for new Knowledge Organizer product development: integrated search and text categorization

• 1987 - 1996                              Xerox Palo Alto Research Center                      Palo Alto, CA  
Xerox PARC is one of the corporate research centers for Xerox corporation. I first

became affiliated with PARC in graduate school when I worked there as a student consultant (that work later formed the basis for my thesis). I had multiple roles at Xerox and at PARC. I was first hired as a system software developer by Xerox AIS, a business unit attempting to commercialize Xerox Interlisp-D. Later as I worked at PARC as a researcher under the aegis of the new Information Access Initiative. Finally I was a research Area Manager responsible for stewarding the Quantitative Content Analysis Group.

### Member of the Research Staff

- One of two authors of the Lisp-based TDB text retrieval system:  
70,000 lines of code
- Contributed to a high-performance Lisp-based finite-state calculus package
- Developed the PARC Hidden Markov Model-based part-of-speech tagger

### Area Manager, Quantitative Content Analysis

- Managed research into information access technologies
- Area output included 30 patent applications and technology transfer to Xsoft: TDB text retrieval engine and trainable document summarizer
- Developed the Scatter/Gather cluster-based document browsing paradigm
- Developed statistical classifier technology for text categorization: participated in TREC4
- Managed a group of 8 scientists

- 1986 - 1987                      **Xerox AI Systems**                      Palo Alto, CA
- Xerox AIS was a PARC spinout devoted to commercializing Xerox Interlisp-D.

### Senior Member of the Technical Staff

- Contributed to the Lyric and Medley release of Xerox Common Lisp
- Responsible for arrays, arithmetic and sequence functions

## Patents and publications

Fourteen issued patents.

For a complete list see <http://www.uspto.gov/patft/index.html> (search for “in/((jan and o) and pedersen)”).

Over thirty refereed publications on information access topics.

For a complete list see [http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/p/Pedersen:Jan\\_O=.html](http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/p/Pedersen:Jan_O=.html) and <http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/p/Pedersen:Jan.html> or search <http://citeseer.ist.psu.edu/cs> for “jan w/2 pedersen or j w/2 pedersen”

Top cited publications include:

Cutting, D., J. Kupiec, J. Pedersen and P. Sibun. 1992. *A Practical Part-of-Speech Tagger*. Proc. 3rd ANLP, Trento, Italy, pp. 133-140.

Julian Kupiec, Jan Pedersen, and Francine Chen. 1995 *A Trainable Document Summarizer*. Research and Development in Information Retrieval.

D. R. Cutting, D. R. Karger, J. O. Pedersen and J. W. Tukey. 1992 *Scatter/Gather: a cluster-based approach to browsing large document collections*. SIGIR'92.

Yang, Y., **Pedersen, J.O.**, 1997 *A Comparative Study on Feature Selection in Text Categorization*, ICML97.

M. A. Hearst and **J. O. Pedersen**. 1996 *Reexamining the cluster hypothesis: Scatter/Gather on retrieval results*. ACM SIGIR'96.

D. R. Cutting, D. R. Karger and **J. O. Pedersen**. 1993 *Constant interaction-time Scatter/Gather browsing of large document collections*. ACM SIGIR'93.

H. Schutze, D. Hull, **J. Pedersen**, 1995 *A Comparison of Classifiers and Document Representations for the Routing Problem*, ACM SIGIR'95.

Ramana Rao, **Jan O. Pedersen**, Marti A. Hearst, Jock D. Mackinlay, Stuart K. Card, Lar Masinter, Per-Kristian Halvorsen, and George G. Robertson, 1995 "Rich Interaction in the Digital Library." In Communications of the ACM.

**Additional professional activities**

Program committee member for SIGIR, WWW and CIKM conferences, most recently for SIGIR 2008, WWW 2008, EC 2008 and CIKM 2008

**Professional memberships**

ACM Distinguished Scientist

**Hobbies**

Reading, walking, cooking.