

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

Microsoft Corp 2009 – current Mountain View, Ca

The Applications Services Group (ASG) of Microsoft includes Bing (Web Search), AdCenter (Advertising), Skype, and Office (Software Productivity tools). I work with the Bing Core Search team.

Distinguished Engineer: 2012 – current

- I lead the Bing Core Relevance team, which is responsible for all aspects of Bing algorithmic search quality. This includes query completion, query understanding and query rewriting, core ranking, contextual ranking and personalization, whole page optimization, metrics, data mining and experimentation infrastructure. In addition, the team is responsible for publisher-side monetization for the Bing search results page.
- I manage over 350 developers at four sites; Bellevue WA, Mountain View CA, London UK and Munich GR. In addition the Core Relevance team has a close collaboration with associated teams in Beijing and Hyderabad.

Partner Development Manager: 2010 – 2012

- I led the Whole Page Relevance Team; one of three relevance teams at Bing. This team was responsible for late-stage content organization and display organization, including Web result caption generation, federated answer blending, publisher-side monetization and whole page metrics development and operation. Deployed features included rich captions, entity-attribute answers, quality improvements in federated answer triggering and blending, principled page weight reduction, and ad annotations.
- I managed 85 developers distributed over three sites; Bellevue WA, San Francisco CA (the old Powerset team), and Mountain View CA.
- As a remote site GM, I have also managed a small program management (PM) and test group.

Chief Scientist, Core Search: 2009 – 2010

- I co-authored two long-range planning documents for Bing Relevance. These include 3 to 5 year plans for new relevance initiatives that improve current search quality as well as create new opportunities.

A9

2008 – 2009

Palo Alto, Ca

A9 provides product search for Amazon.com.

Chief Scientist

I participated in the development of the 2009 operating plan for both product search and A9's ClickRiver performance marketing product. In addition, in collaboration with a few key architects, I articulated a self-service architecture for the A9 search stack that leverages Amazon's EC2 and S3 utility computing infrastructures and a strategy for improving A9's metrics and tuning methodologies. I moved to MSFT when my former manager at Yahoo, Qi Lu, became President of Microsoft's Online Service Division.

Yahoo! Inc

2003 – 2008

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.
- Led 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.

AltaVista**2002 – 2003****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. Yahoo! in turn purchased Overture 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.

Various Startups 2000 – 2002**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.

Infoseek/Go Network 1998 – 2000 (Jan) 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 as 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 5.3 Billion page views.
- Managed four groups with a total staff of 20 and an annual budget of \$6M.

Verity Inc

1996 - 1998

Sunnyvale, Ca

Verity was a 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

Xerox PARC

1987 - 1996

Palo Alto, CA

Xerox PARC was 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.

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

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

Xerox AI Systems

1986 - 1987

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

Fifteen 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 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, Larry Masinter, Per-Kristian Halvorsen, and George G. Robertson, 1995 *"Rich Interaction in the Digital Library."* In Communications of the ACM.

Additional professional activities

Frequent program committee member for the SIGIR, WSDM, WWW and CIKM conferences

Professional memberships

ACM Distinguished Scientist

Hobbies

Reading, walking, cooking.