

## Employment

---

*Mar 2014–now* **Google Inc., Mountain View, CA**  
*Senior Staff Research Scientist*

- Since Jan. 2015, part of the “Strategic Technologies” team. Managing a group of 20 researchers.
- From Mar. 2014 until Dec. 2014, part of the “Personal Search Infrastructure” team.

*Oct 2001–Mar 2014* **Microsoft Corporation, Microsoft Research Silicon Valley, Mountain View, CA**  
*Principal Researcher (Mar 2006–Mar 2014)*  
*Senior Researcher (Oct 2005–Mar 2006)*  
*Researcher (Oct 2001–Sep 2005)*

- Collaborated with Bing on various facets of social search.
- Explored link-based ranking techniques for web search results.
- Developed the Scalable Hyperlink Store, a specialized database giving extremely fast access to nodes and edges of the web graph induced by the Bing corpus.
- Explored techniques for identifying spam web pages.
- Consulted on large-scale web crawling for Bing.
- Principal contributor to PageTurner, a large-scale study of the evolution of web pages.
- Contributed to Boxwood, a distributed, scalable, and reliable B-Tree system.
- Four Microsoft Gold Star awards; three Microsoft Research Tech Transfer Awards; Microsoft Corporate Bench Program class of 2005; promoted to Partner level in 2009.

*Oct 1993–Sep 2001* **Digital Equipment Corporation (Compaq since 1998), Systems Research Center, Palo Alto, CA**  
*Manager, Programming Technology (Jun 2001–Sep 2001)*

- Managed a group of five Ph.D.-level researchers. Responsible for four projects.

*Senior Member of Technical Staff (Sep 1999–Jun 2001)*

- Principal contributor to Mercator, an extensible, high-performance web crawler. Mercator formed the web crawling component of AltaVista’s *Search Engine 3* product, which sold to over 1,200 enterprise customers, and it became the standard web crawler of the various AltaVista sites.

*Software Principal Engineer (Oct 1993–Sep 1999)*

- Main contributor to JCAT, a Java-based algorithm animation system.
- Worked on tools and techniques for building distributed, collaborative, web-based applications.
- Designed and implemented Obliq-3D, a fast-turnaround, interactive 3D animation environment.

## Education

---

*January 1994* **Ph.D. in Computer Science, University of Illinois at Urbana-Champaign**

- Dissertation “*Programming in Three Dimensions*” supervised by Prof. Simon Kaplan.

*May 1989* **Diplom-Wirtschaftsinformatiker, Technical University of Darmstadt, Germany**

- Program covers Computer Science, Mathematics, Business Administration, Economics, and Law.

## Honors

---

ACM Distinguished Scientist, since 2008.

IEEE Senior Member, since 2012.

## Professional Service

---

*ACM Transactions on the Web (TWEB)*

Editor in Chief, 2011–2015  
Associate Editor, 2005–2011

*Communications of the ACM (CACM)*

Co-Chair of News Board, 2008–2014

*Journal of Visual Languages and Computing (JVLC)*

Associate Editor, 2001–2011  
Visual Software Tools Editor, 1996–2001

*ACM International Conference on Web Search and Data Mining (WSDM)*

Steering Committee Member, since 2010  
Conference Chair, 2008  
Tutorials Co-Chair, 2011  
Senior Program Committee Member, 2009, 2012, 2015, 2016, 2017, 2018  
Program Committee Member, 2010

*International World Wide Web Conference (WWW)*

Program Co-Chair, 2004  
Co-Chair of “Posters” track, 2011  
Chair of “Tutorials and Workshops” track, 2007  
Co-Chair of “Search” track, 2010  
Co-Chair of “Industrial Practice & Experience” track, 2006, 2009  
Vice Chair of “Browsers and User Interfaces” track, 2003  
Deputy Chair of “Browsers and User Interfaces” track, 2002  
Deputy Chair of “Browsers and Tools” track, 2001  
Senior Program Committee Member, 2016  
Program Committee Member, 1999, 2000, 2005, 2012

*International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*

Area Chair, 2011, 2012, 2013  
Program Committee Member, 2005, 2006, 2008

*ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*

Senior Program Committee Member, 2010  
Program Committee Member, 2008

*IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*

Program Committee Member, 2017

*ACM Conference on Information and Knowledge Management (CIKM)*

Co-chair of Industry Track, 2015.  
Program Committee Member, 2008

*Conference on Design of Experimental Search and Information Retrieval Systems (DESIRES)*

Steering Committee Member, since 2017

*BigData Innovators Gathering (BIG)*

Program Co-Chair, 2017

*International Workshop on Adversarial Information Retrieval on the Web (AIRWEB)*

Program Co-Chair, 2006  
Program Committee Member, 2005, 2007, 2008, 2009

*International Conference on Internet Computing (IC)*

Program Committee Member, Section on Search and Information Retrieval for the Web, 2003

*Workshop on Algorithms and Models for the Web Graph (WAW)*

Program Committee Member, 2003

*International Workshop on Visual Languages and Computing (VLC)*

Program Committee Member, 2003, 2004, 2006, 2007, 2008

*IEEE Symposium on Visual Languages (VL)*  
Video Chair, 1999  
Program Committee Member, 1996, 1998, 1999, 2000, 2005, 2006

---

## Patents

- Issued* US Patent 6,263,364. *Web crawler system using plurality of parallel priority level queues having distinct associated download priority levels for prioritizing document downloading and maintaining document freshness* (with A. Heydon and J. Wiener).
- US Patent 6,301,614. *System and method for efficient representation of data set addresses in a web crawler* (with A. Heydon).
- US Patent 6,321,265. *System and method for enforcing politeness while scheduling downloads in a web crawler* (with A. Heydon).
- US Patent 6,351,755. *System and method for associating an extensible set of data with documents downloaded by a web crawler* (with A. Heydon).
- US Patent 6,377,984. *Web crawler system using parallel queues for queing data sets having common address and concurrently downloading data associated with data set in each queue* (with A. Heydon).
- US Patent 6,594,694. *System and method for near-uniform sampling of web page addresses* (with A. Heydon, M. Mitzenmacher and M. Henzinger).
- US Patent 6,910,077. *System and method for identifying cloaked web servers.*
- US Patent 6,952,730. *System and method for efficient filtering of data set addresses in a web crawler* (with A. Heydon).
- US Patent 7,007,027. *Algorithm for tree traversals using left links* (with C. Thekkath).
- US Patent 7,072,904. *Deletion and compaction using versioned nodes* (with C. Thekkath).
- US Patent 7,082,438. *Algorithm for tree traversals using left links* (with C. Thekkath).
- US Patent 7,139,747. *System and method for distributed web crawling.*
- US Patent 7,340,467. *System and method for maintaining a distributed database of hyperlinks.*
- US Patent 7,627,777. *Fault tolerance scheme for distributed hyperlink database.*
- US Patent 7,680,785. *Systems and methods for inferring uniform resource locator (URL) normalization rules.*
- US Patent 7,739,281. *Systems and methods for ranking documents based upon structurally interrelated information.*
- US Patent 7,783,671. *Deletion and compaction using versioned nodes* (with C. Thekkath).
- US Patent 7,792,854. *Query dependent link-based ranking.*
- US Patent 7,818,334. *Query dependent link-based ranking using authority scores.*
- US Patent 7,962,510. *Using content analysis to detect spam web pages* (with D. Fetterly, M. Manasse and A. Ntoulas).
- US Patent 8,209,305. *Incremental update scheme for hyperlink database.*
- US Patent 8,392,366. *Changing number of machines running distributed hyperlink database.*
- US Patent 8,666,920. *Estimating shortest distances in graphs* (with S. Gollapudi, R. Panigrahy and A. Das Sarma).
- US Patent 8,856,112. *Considering document endorsements when processing queries* (with R. Panigrahy and R. Shenoy).
- US Patent 8,949,232. *Social network recommended content and recommending members for personalized search results* (with T. Harrington, R. Panigrahy and R. Shenoy).

*Published* For a list of non-issued published patent applications please visit my home page <http://marc.najork.org>

## Publications

---

- Book Chapters* “High-Performance Web Crawling” (with A. Heydon). Chapter 2 in J. Abello et al. (editors), *Handbook of Massive Data Sets*, Kluwer Academic Publishers, 2002.
- “Algorithm Animation Using Interactive 3D Graphics” (with M. Brown). Chapter 9 in J. Stasko et al. (editors), *Software Visualization — Programming as a Multimedia Experience*, MIT Press, 1998.
- Encyclopedia Entries* “Web Crawler Architecture”. In *Encyclopedia of Database Systems*, Springer, 2009.
- “Web Search Relevance Ranking” (with H. Zaragoza). In *Encyclopedia of Database Systems*, Springer, 2009.
- “Web Spam Detection”. In *Encyclopedia of Database Systems*, Springer, 2009.
- Journal Articles* “Web Crawling” (with C. Olston). *Foundations and Trends in Information Retrieval* **4**, 3 (2010), 175–246.
- “Adversarial Information Retrieval on the Web (AIRWeb 2006)” (with B. Davison and T. Converse). *SIGIR Forum* **40**, 2 (December 2006), 27–30.
- “On the Evolution of Clusters of Near-duplicate Web Pages” (with D. Fetterly and M. Manasse). *Journal of Web Engineering* **2**, 4 (October 2004), 228–246.
- “A Large-Scale Study of the Evolution of Web Pages” (with D. Fetterly, M. Manasse and J. Wiener). *Software: Practice & Experience* **34**, 2 (February 2004), 213–237.
- “Performance Limitations of the Java Core Libraries” (with A. Heydon). *Concurrency: Practice & Experience* **12**, 6 (May 2000), 363–373.
- “Mercator: A Scalable, Extensible Web Crawler” (with A. Heydon). *World Wide Web* **2**, 4 (December 1999), 219–229.
- “Collaborative Active Textbooks” (with M. Brown). *Journal of Visual Languages and Computing* **8**, 4 (August 1997), 453–486.
- “Programming in Three Dimensions”. *Journal of Visual Languages and Computing* **7**, 2 (June 1996), 219–242.
- “Obliq-3D: A High-Level, Fast-Turnaround 3D Animation System” (with M. Brown). *IEEE Transactions on Visualization and Computer Graphics* **1**, 2 (June 1995), 175–193.
- “Roles and their role in posing recursive queries” (with R. John, S. Kuck and A. Lewe). *Information Systems* **15**, 2 (1990), 173–186.
- Keynotes and Invited Talks* “Using Machine Learning to Improve the Email Experience”. In *25th ACM International Conference on Information and Knowledge Management (October 2016)*, 891.
- “Social Search”. *14th International Conference on Web Engineering* (July 2014), 571-572.
- “Querying the Web Graph”. *16th International Symposium on String Processing and Information Retrieval* (October 2010), 1–12.
- “Web-Based Algorithm Animation”. *38th Design Automation Conference* (June 2001), 506–511.
- Conference Papers* “Quick Access: Building a Smart Experience for Google Drive” (with S. Tata, A. Popescul, M. Colagrosso, J. Gibbons, A. Green, A. Mah, M. Smith, D. Garg, C. Meyer, and R. Kan). To appear in *23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (August 2017)*.
- “Email Category Prediction” (with A. Zhang, L. Garcia-Pueyo, J.B. Wendt and A. Broder). *26th International World Wide Web Conference, Companion Volume* (April 2017), 495-503.
- “Learning from User Interactions in Personal Search via Attribute Parameterization” (with X. Wang, M. Bendersky and D. Metzler). *10th ACM International Conference on Web Search and Data Mining* (February 2017), 791-799.

“Learning to Rank with Selection Bias in Personal Search” (with X. Wang, M. Bendersky and D. Metzler). *39th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (July 2016), 115-124.

“Debugging a Crowdsourced Task with Low Inter-Rater Agreement” (with O. Alonso and C. Marshall). *15th ACM/IEEE-CS Joint Conference on Digital Libraries* (June 2015), 101-110.

“A Human-Centered Framework for Ensuring Reliability on Crowdsourced Labeling Tasks” (with O. Alonso and C. Marshall). *1st AAAI Conference on Human Computation and Crowdsourcing – Works in Progress* (November 2013).

“Are Some Tweets More Interesting Than Others? #HardQuestion” (with O. Alonso and C. Marshall). *7th Annual Symposium on Human-Computer Interaction and Information Retrieval* (October 2013).

“Boot-Strapping Language Identifiers for Short Colloquial Postings” (with M. Goldszmidt and S. Paparizos). *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (September 2013), 95-111.

“Robust Query Rewriting using Anchor Data” (with N. Craswell, B. Billerbeck and D. Fetterly). *6th ACM International Conference on Web Search and Data Mining* (February 2013), 335-344.

“Detecting Quilted Web Pages at Scale”. *35th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (August 2012), 385-394.

“How User Behavior is Related to Social Affinity” (with R. Panigrahy and Y. Xie). *5th ACM International Conference on Web Search and Data Mining* (February 2012), 713-722.

“Of Hammers and Nails: An Empirical Comparison of Three Paradigms for Processing Large Graphs” (with D. Fetterly, A. Halverson, K. Kenthapadi and S. Gollapudi). *5th ACM International Conference on Web Search and Data Mining* (February 2012), 103-112.

“Microsoft Research at TREC 2011 Web Track” (with B. Billerbeck, N. Craswell and D. Fetterly). *20th Text Retrieval Conference* (November 2011).

“The Power of Peers” (with N. Craswell and D. Fetterly). *33rd European Conference on Information Retrieval* (April 2011), 497-502.

“Microsoft Research at TREC 2010 Web Track” (with N. Craswell and D. Fetterly). *19th Text Retrieval Conference* (November 2010).

“A Sketch-Based Distance Oracle for Web-Scale Graphs” (with A. Das Sarma, S. Gollapudi and R. Panigrahy). *3rd ACM Intl. Conference on Web Search and Data Mining* (February 2010), 401-410.

“Microsoft Research at TREC 2009” (with N. Craswell, D. Fetterly, S. Robertson and E. Yilmaz). *18th Text Retrieval Conference* (November 2009).

“The Scalable Hyperlink Store”. *20th ACM Conference on Hypertext and Hypermedia* (June 2009), 89-98.

“Less is More: Sampling the Neighborhood Graph Makes SALSA Better and Faster” (with S. Gollapudi and R. Panigrahy). *2nd ACM International Conference on Web Search and Data Mining* (February 2009), 242-251.

“Efficient and Effective Link Analysis with Precomputed SALSA Maps” (with N. Craswell). *17th International Conference on Information and Knowledge Management* (October 2008), 53-61.

“Computing Information Retrieval Performance Measures Efficiently in the Presence of Tied Scores” (with F. McSherry). *30th European Conference on Information Retrieval* (April 2008), 414-421.

“Using Bloom Filters to Speed Up HITS-like Ranking Algorithms” (with S. Gollapudi and R. Panigrahy). *5th Workshop On Algorithms And Models For The Web-Graph* (December 2007), 195-201.

“Comparing the Effectiveness of HITS and SALSA”. *16th Intl. Conference on Information and Knowledge Management* (November 2007), 157-164.

“HITS on the Web: How does it Compare?” (with H. Zaragoza and M. Taylor). *30th Annual Intl. ACM SIGIR Conference on Research and Development in Information Retrieval* (July 2007), 471-478.

“Detecting Spam Web Pages through Content Analysis” (with A. Ntoulas, M. Manasse and D. Fetterly). *15th International World Wide Web Conference* (May 2006), 83–92.

“Detecting Phrase-Level Duplication on the World-Wide Web” (with D. Fetterly and M. Manasse). *28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (August 2005), 170–177.

“Boxwood: Abstractions as the Foundation for Storage Infrastructure” (with J. MacCormick, N. Murphy, C. Thekkath and L. Zhou). *6th Symposium on Operating Systems Design and Implementation* (December 2004), 105–120.

“Spam, Damn Spam, and Statistics: Using Statistical Analysis to Locate Spam Web Pages” (with D. Fetterly and M. Manasse). *7th International Workshop on the Web and Databases* (June 2004), 1–6.

“On the Evolution of Clusters of Near-Duplicate Web Pages” (with D. Fetterly and M. Manasse). *1st Latin American Web Congress* (November 2003), 37–45.

“A Large-Scale Study of the Evolution of Web Pages” (with D. Fetterly, M. Manasse and J. Wiener). *12th International World Wide Web Conference* (May 2003), 669–678.

“Efficient URL caching for World Wide Web crawling” (with A. Broder and J. Wiener). *12th International World Wide Web Conference* (May 2003), 679–689.

“Breadth-First Search Crawling Yields High-Quality Pages” (with J. Wiener). *10th International World Wide Web Conference* (May 2001), 114–118.

“On Near-Uniform URL Sampling” (with M. Henzinger, A. Heydon and M. Mitzenmacher). *9th International World Wide Web Conference* (May 2000), 295–308.

“Performance Limitations of the Java Core Libraries” (with A. Heydon). *ACM 1999 Java Grande Conference* (June 1999), 35–41.

“Measuring Index Quality Using Random Walks on the Web” (with M. Henzinger, A. Heydon and M. Mitzenmacher). *8th International World Wide Web Conference* (May 1999), 213–225.

“A Java-Based Implementation of Collaborative Active Textbooks” (with M. Brown and R. Raisamo). *IEEE Symposium on Visual Languages* (September 1997), 372–379.

“Distributed Applets” (with M. Brown). *CHI’97 Conference Companion* (March 1997), 204–205.

“Collaborative Active Textbooks: A Web-based Algorithm Animation System for an Electronic Classroom” (with M. Brown). *IEEE Symposium on Visual Languages* (September 1996), 266–275.

“Distributed Active Objects” (with M. Brown). *5th International World Wide Web Conference* (May 1996), 1037–1052.

“A Library for Visualizing Combinatorial Structures” (with M. Brown). *IEEE Visualization ’94* (October 1994), 164–171.

“Algorithm Animation Using 3D Interactive Graphics” (with M. Brown). *ACM Symposium on User Interface Software and Technology* (November 1993), 93–100.

“Cube: Eine dreidimensionale visuelle Programmiersprache” (with S. Kaplan). *Informatik, Wirtschaft, Gesellschaft* (September 1993), 340–345.

“Specifying Visual Languages with Conditional Set Rewrite Systems” (with S. Kaplan). *IEEE Symposium on Visual Languages* (August 1993), 12–18.

“A Prototype Implementation of the Cube Language” (with S. Kaplan). *IEEE Workshop on Visual Languages* (September 1992), 270–272.

“The Cube Language” (with S. Kaplan). *IEEE Workshop on Visual Languages* (October 1991), 218–224.

“Enhancing Show-and-Tell with a polymorphic type system and higher-order functions” (with E. Golin). *IEEE Workshop on Visual Languages* (October 1990), 215–220.

*Popular  
Magazines*

“Distributed Active Objects” (with M. Brown)  
*Dr. Dobb’s Journal* (March 1997), 34–41.

“Visual Programming in 3D”.  
*Dr. Dobb’s Journal* (December 1995), 18–31.

*Technical  
Reports*

“Crowdsourcing a Subjective Labeling Task: A Human-Centered Framework for Ensuring Reliable Results” (with O. Alonso and C. Marshall). *MSR-TR-2014-91, Microsoft Research* (June 2014).

“High-Performance Web Crawling” (with A. Heydon).  
*SRC Research Report 173, Compaq Systems Research Center, Palo Alto* (September 2001).

“Three-Dimensional Web-Based Algorithm Animations” (with M. Brown).  
*SRC Research Report 170, Compaq Systems Research Center, Palo Alto* (July 2001).

“Focus + Context Displays of Web Pages: Implementation Alternatives” (with M. Brown, H. Marais and W. Weihl). *SRC Technical Note 1997-010, DEC Systems Research Center, Palo Alto* (May 1997).

“Collaborative Active Textbooks: A Web-based Algorithm Animation System for an Electronic Classroom” (with M. Brown). *SRC Research Report 142, DEC Systems Research Center, Palo Alto* (May 1996).

“Distributed Active Objects” (with M. Brown).  
*SRC Research Report 141 (paper & video), DEC Systems Research Center, Palo Alto* (May 1994).

“Obliq-3D Tutorial and Reference Manual”.  
*SRC Research Report 129, DEC Systems Research Center, Palo Alto* (December 1994).

“A Library for Visualizing Combinatorial Structures” (with M. Brown).  
*SRC Research Report 128 (paper & video), DEC Systems Research Center, Palo Alto* (September 1994).

“Programming in Three Dimensions”.  
*Technical Report UIUCDCS-R-93-1838, Dept. of Computer Science, Univ. of Illinois* (October 1993).

“Algorithm Animation Using 3D Interactive Graphics” (with M. Brown).  
*SRC Research Report 110 (paper & video), DEC Systems Research Center, Palo Alto* (September 1993).

“Funktionale, logik-basierte und objektorientierte Sprachstile und Wege zur Vereinheitlichung”.  
*Thesis, Fachbereich Informatik, Technical University of Darmstadt, Germany* (1989).

“Enhanced ER-Easy: A Database Scheme Designer”.  
*Technical Report UIUCDCS-R-88-1464, Dept. of Computer Science, Univ. of Illinois* (May 1988).

“Roles and their role in posing recursive queries over the universal relation” (with R. John, S. Kuck and A. Lewe). *Technical Report UIUCDCS-R-88-1463, Dept. of Computer Science, Univ. of Illinois* (May 1988).