

Dennis Fetterly

CONTACT INFORMATION

2240 Semeria Ave
Belmont, CA 94002 USA

Phone: (650) 260-5529
E-mail: dennis@fetterly.org
Web: <http://www.fetterly.org/~dennis/>

RESEARCH INTERESTS

Distributed systems, information retrieval

EXPERIENCE

Microsoft Research Silicon Valley, Mountain View, California

Researcher

Senior Research Software Development Engineer

Technologist

February 2010 to present
October 2005 to February 2010
May 2003 to October 2005

- Contributions to the Dryad project to build a data parallel execution engine, which has seen production use inside of Microsoft since 2007.
- Significant contributions to the DryadLINQ project, which enables programmers to write large scale distributed application using SQL like constructs in C#.
- Lead the TidyFS project, which is a distributed filesystem designed for data-parallel, scalable systems such as Dryad.
- Perform the initial port of Dryad and DryadLINQ to Hadoop YARN so it could be open sourced.
- Develop techniques for identification of spam web pages which are used in the Bing search engine.
- Build a scalable, distributed web crawler used within Microsoft and licensed externally.
- Cluster similar web pages together and evaluate the evolution of those clusters over time.
- Co-mentor an intern investigating techniques for harvesting idle cycles within data-centers.

Compaq Computer Corporation/Hewlett-Packard Company, Palo Alto, California

Research Support Engineer, HP Labs

Systems/Software Engineer, Compaq Systems Research Center

July 2002 to May 2003
September 1999 to July 2002

- Conduct experiment exploring the rate of change of the World Wide Web.
- Consult with research staff to determine project requirements, and configure the computing environment to meet those requirements.
- Co-manage Compaq Research's Palo Alto Internet Gateway, including BGP routers, DNS servers with over 2000 zones, two publicly available NTP clocks, and a well known public ftp server.
- Negotiate service levels and support for Compaq Corporate Research with Corporate IT group.

Center for Coastal Studies, Scripps Institution of Oceanography, University of California, San Diego

System Administrator

February 1995 to August 1999

- Install, configure, and maintain web, email, and DNS servers including periodic security audits.
- Develop Perl scripts to automate system administration duties.
- Develop a Java applet to show near real time wind and current velocities in the Santa Barbara Channel.
- Design web and Java based user interfaces to system administration software.

United States Marine Corps

Corporal

December 1990 to September 1994

- Honorable discharge. Decorations included the Navy Achievement Medal awarded for performance of duties as the Company Training Non-Commissioned Officer.

PERSONAL
PROJECTS

In 2011, as a result of overcrowding in my local elementary school district, I proposed a distance based student assignment system that utilizes a minimum cost maximum flow algorithm to assign students to one of six elementary schools within the district based on the schools' capacity and the walking distances for each student to each of the six elementary schools. As a result of my proposal, I became a member of the committee tasked with identifying a solution and shepherded my proposal through the process. I provided the district with an implementation of this algorithm which has been used for the last three school years as well as a website to calculate and display walking distances from an address to each of the six schools.

EDUCATION

University of California, San Diego, La Jolla, California USA

B.A., Computer Science, 1999

HONORS AND
AWARDS

Jay Lepreau Award for Best Paper at OSDI'08

PUBLICATIONS

BOOK CHAPTERS

Mihai Budiu, Dennis Fetterly, Michael Isard, Frank McSherry, and Yuan Yu, Large-scale Machine Learning using DryadLINQ, in *Scaling Up Machine Learning*, R. Bekkerman, M. Bilenko, and J. Langford (editors), Cambridge University Press, 2011.

ENCYCLOPEDIA
ENTRIES

Web Characteristics and Evolution. *Encyclopedia of Database Systems 2009*: 3460-3462.

JOURNAL ARTICLES

Dennis Fetterly, Mark Manasse, and Marc Najork, On the Evolution of Clusters of Near-Duplicate Web Pages, in *Journal of Web Engineering*, vol. 2, no. 4, pp. 228-246, Institute of Electrical and Electronics Engineers, Inc., October 2004.

Dennis Fetterly, Mark Manasse, Marc Najork, and Janet Wiener, A Large-Scale Study of the Evolution of Web Pages, in *Software: Practice & Experience*, vol. 34, no. 2, pp. 213-237, Wiley, February 2004.

CONFERENCE
PAPERS

Omar Alonso, Dennis Fetterly, and Mark Manasse, Duplicate News Story Detection Revisited, in *The Ninth Asia Information Retrieval Societies Conference*, Springer Verlag, 9 December 2013.

Christopher J. Rossbach, Yuan Yu, Jon Currey, Jean-Philippe Martin, and Dennis Fetterly, Dandelion: a Compiler and Runtime for Heterogeneous Systems, *SOSP'13: The 24th ACM Symposium on Operating Systems Principles*, 3 November 2013.

Nick Craswell, Bodo Billerbeck, Dennis Fetterly, and Marc Najork, Robust Query Rewriting using Anchor Data, in *6th ACM International Conference on Web Search and Data Mining (WSDM)*, ACM, February 2013.

Marc Najork, Dennis Fetterly, Alan Halverson, Krishnaram Kenthapadi, and Sreenivas Gollapudi, Of Hammers and Nails: An Empirical Comparison of Three Paradigms for Processing Large Graphs, in *5th ACM International Conference on Web Search and Data Mining (WSDM)*, ACM, February 2012.

Bodo Billerbeck, Nick Craswell, Dennis Fetterly, and Marc Najork, Microsoft Research at TREC 2011 Web Track, in *Proc. of the 20th Text Retrieval Conference (TREC)*, National Institute of Standards and Technology, November 2011.

Dennis Fetterly, Maya Haridasan, Michael Isard, and Swaminathan Sundararaman, TidyFS: A Simple and Small Distributed File System, in *Proceedings of the USENIX Annual Technical Conference (USENIX'11)*, USENIX, June 2011.

Nick Craswell, Dennis Fetterly, and Marc Najork, The Power of Peers, in 33rd European Conference on IR Research (ECIR), Springer Verlag, April 2011.

Nick Craswell, Dennis Fetterly, and Marc Najork, Microsoft Research at TREC 2010 Web Track, in Proc. of the 19th Text Retrieval Conference (TREC), National Institute of Standards and Technology, November 2010.

Dennis Fetterly and Frank McSherry, A Data-Parallel Toolkit for Information Retrieval, in Proceedings of SIGIR, Association for Computing Machinery, Inc., July 2010.

Nick Craswell, Dennis Fetterly, Marc Najork, Stephen Robertson, and Emine Yilmaz, Microsoft Research at TREC 2009: Web and Relevance Feedback Tracks, in Proc. of the 18th Text Retrieval Conference (TREC), National Institute of Standards and Technology, November 2009.

Dennis Fetterly, Nick Craswell, and Vishwa Vinay, The Impact of Crawl Policy on Web Search Effectiveness, in Proceedings of the 32nd annual international ACM SIGIR conference on Research and development in information retrieval, Association for Computing Machinery, Inc., July 2009.

Dennis Fetterly, Nick Craswell, and Vishwa Vinay, Measuring the Search Effectiveness of a Breadth-First Crawl, in Proceedings of the 31st European Conference on Information Retrieval (ECIR), Springer Verlag, April 2009.

Yuan Yu, Michael Isard, Dennis Fetterly, Mihai Budiu, Úlfar Erlingsson, Pradeep Kumar Gunda, and Jon Currey, DryadLINQ: A System for General-Purpose Distributed Data-Parallel Computing Using a High-Level Language, in OSDI'08: Eighth Symposium on Operating System Design and Implementation, USENIX, December 2008.

Dennis Fetterly, Nick Craswell, and Vishwa Vinay, Search Effectiveness with a Breadth-First Crawl, in Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, July 2008.

Michael Isard, Mihai Budiu, Yuan Yu, Andrew Birrell, and Dennis Fetterly, Dryad: Distributed Data-parallel Programs from Sequential Building Blocks, in Proceedings of the 2007 Eurosys Conference, Association for Computing Machinery, Inc., March 2007.

Alexandros Ntoulas, Marc Najork, Mark Manasse, and Dennis Fetterly, Detecting Spam Web Pages Through Content Analysis, in 15th International World Wide Web Conference (WWW), Association for Computing Machinery, Inc., May 2006.

Dennis Fetterly, Mark Manasse, and Marc Najork, Detecting Phrase-Level Duplication on the World Wide Web, in 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Association for Computing Machinery, Inc., August 2005.

Dennis Fetterly, Mark Manasse, and Marc Najork, Spam, Damn Spam, and Statistics: Using statistical analysis to locate spam web pages, in 7th International Workshop on the Web and Databases (WebDB), Association for Computing Machinery, Inc., June 2004.

Dennis Fetterly, Mark Manasse, and Marc Najork, On the Evolution of Clusters of Near-Duplicate Web Pages, in Proceedings of the 1st Latin American Web Congress (LA-WEB), IEEE Computer Society, November 2003.

Dennis Fetterly, Mark Manasse, Marc Najork, and Janet Wiener, A large-scale study of the evolution of web pages, in Proceedings of the 12th International World Wide Web Conference (WWW), ACM, May 2003.

TECHNICAL
REPORTS

Omar Alonso, Dennis Fetterly, and Mark Manasse, Duplicate News Story Detection Revisited, no. MSR-TR-2013-60, May 2013.

Dennis Fetterly, Maya Haridasan, Michael Isard, and Swaminathan Sundararaman, TidyFS: A Simple and Small Distributed Filesystem, no. MSR-TR-2010-124, October 2010.

Yuan Yu, Michael Isard, Dennis Fetterly, Mihai Budiu, Úlfar Erlingsson, Pradeep Kumar Gunda, Jon Currey, Frank McSherry, and Kannan Achan, Some sample programs written in DryadLINQ, no. MSR-TR-2009-182, December 2009.

Muthukarrupan Annamalai, Andrew Birrell, Dennis Fetterly, and Ted Wobber, Implementing Portable Desktops: a New Option and Comparisons, no. TR-2006-151, October 2006.

PATENTS

ISSUED

US Patent 7,565,350. Identifying a web page as belonging to a blog (with S. Chien).

US Patent 7,962,510. Using content analysis to detect spam web pages (with M. Najork, A. Ntoulas, and M. Manasse).

US Patent 8,346,763. Ranking method using hyperlinks in blogs (with S. Chien).

PENDING
(PUBLISHED)

Application number 12/368231. General Purpose Distributed Data Parallel Computing using a High Level Language (with Y. Yu, M. Isard, Ú. Erlingsson, and M. Budiu).

Application number 11/537506. Distributed parallel computing (with M. Isard, A. Birell, M. Budiu and Y. Yu).

Application number 10/956228. Content evaluation (with M. Najork and M. Manasse).

PROFESSIONAL
SERVICE

AAAI Conference on Artificial Intelligence

- Program Committee Member, AI and the Web Special Track, 2007

ACM International Conference on Web Search and Data Mining

- Program Committee Member, 2009, 2010, 2011, 2012, 2015
- Workshops Program Committee Member, 2013, 2014

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

- Program Committee Member, 2012, 2013, 2014

ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

- Program Committee Member, 2010

International World Wide Web Conference

- Co-Chair of Demo track, 2011
- Co-Chair of Application, Experience, and Demo track, 2010
- Program Committee Member, 2008, 2009, 2012, 2013, 2015

International Workshop on Adversarial Information Retrieval on the Web

- Program Co-Chair, 2008, 2009
- Program Committee Member, 2005, 2006, 2007, 2011, 2012, 2013

ACM Conference on Information and Knowledge Management

- Program Committee Member, 2008, 2009, 2011, 2012