

Marti A. Hearst

Curriculum Vitae

Contact Information

102 South Hall, Berkeley CA 94720
510-642-8016
hearst@berkeley.edu
<http://www.ischool.berkeley.edu/~hearst>

Education

Ph.D., Computer Science, University of California, Berkeley, May 1994.
Advisor: Prof. Robert Wilensky.
Dissertation: *Context and Structure in Automated Full-Text Information Access*.
M.S., Computer Science, University of California, Berkeley, May 1989.
Advisor: Prof. Michael Stonebraker.
B.A., Computer Science, University of California, Berkeley, Dec. 1985.
Graduated with distinction and honors in Computer Science.

Professional Employment History

2008–present: Full Professor, School of Information, University of California, Berkeley.
2015–present: Full Professor, EECS Department, University of California, Berkeley
2009–2011: On leave from UCB, working in the U.S. Federal Government.
2002–2008: Associate Professor, School of Information, University of California, Berkeley.
1997–2002: Assistant Professor, School of Information, University of California, Berkeley.
2000 (Spring): Visiting Professor, Harvard University.
1994–1997: Member of the Research Staff, Xerox Palo Alto Research Center (PARC).
1987–1994: Research Assistant, University of California, Berkeley.
1989–1991 (Summers): Research Intern, Xerox Palo Alto Research Center.
1992 (Summer): Research Intern, AT&T Bell Laboratories, Murray Hill, NJ.
1986–1987: Software Engineer, Hewlett Packard, Cupertino, CA.
1985 (Summer): Intern, International Business Machines, Manassas, VA.
1984 (Summer): Intern, Lawrence Berkeley Laboratory, Berkeley, CA.

Awards and Honors

2015 Excellence in Teaching Award (School of Information)
2014 Excellence in Teaching Award (School of Information)
2013 Named a Fellow of the ACM
2002 Excellence in Teaching Award (School of Information)

2002 IBM Faculty Fellowship (\$40,000)
2000 NSF Career Grant, Principal Investigator (\$300,000)
2000 Hellman Faculty Fund award (\$30,000)
1999 Excellence in Teaching Award (School of Information)
1998 Okawa Foundation Fellow (\$10,000)

Most Cited Papers

CHI 2006, most cited paper
PSB 2003, most cited paper (most cited paper since 2003 for that conference)
CHI 2003, most cited paper
Infoviz 2001, most cited paper
ACL 1999, most cited paper
Computational Linguistics Journal, most cited article of 1997
COLING 1992, most cited paper of all time for that conference.
(Statistics obtained on 9/11/2015 from Aminer.org and Google Scholar)

Grants and Research Gifts

2014 Gift from Tableau (\$50,000)
2013 Research Grant from Google (\$200,000)
2013 MOOCLab Grant, UC Berkeley (\$75,000)
2012 NEH Grant, Principle Investigator (\$217,000)
2007 Research Grant from Google (\$50,000)
2006 NSF Grant, Principle Investigator (\$580,000)
2006 IBM Innovation Award (\$12,000)
2006 Gift from FXPal (\$10,000)
2004 Gift from Genentech (\$40,000)
2003 NSF Grant, Principle Investigator (\$840,000)
2003 Gift from Genentech (\$20,000)
2003 Gift from Oracle Corporation (\$30,000)
2002 Gift from Genentech (\$20,000)
2002 Gift from Oracle Corp. (\$20,000)
2001 ARDA AQUAINT Award, Principal Investigator (\$1,000,000)
2000 ARDA Award, Principal Investigator (\$200,000)
2000 Gift from Microsoft Research (\$30,000)
2000 Gift from Oracle Corporation (\$15,000)
1998 Sun Academic Equipment Grant (valued at \$50,000)

1998 Gift from Adobe Systems (\$7,000)

1994 NSF support AAAI for Fall Symposium on Improving Instruction of Artificial Intelligence (\$9,800)

1989 California Legislative Grant

Program Chair and Organizational Activities

Vice President, Association for Computational Linguistics, 2017.

Co-organizer, IEEE Infviz workshop on Pedagogy in Information Visualization, 2016.

Program Co-Chair, First Annual ACM Conference on Learning at Scale, 2014.

Steering Committee, ACM Learning at Scale Conference, 2014-present.

Steering Committee, ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR), 2014-present.

Steering Committee, AAAI International Conference on Weblogs and Social Media (ICWSM), 2013-2015.

General Chair, Fourth AAAI International Conference on Weblogs and Social Media (ICWSM), 2010.

Co-organizer, Workshop on Search and Social Media, WSDM Workshop, 2010.

Co-organizer, Workshop on Search and Social Media, SIGIR Workshop, 2009.

Program Co-Chair, First Annual Human Languages Technologies / NAACL, 2003

Program Co-Chair, ACM SIGIR 1999.

Co-Organizer and Program Co-chair of the 1996 AAAI Spring Symposium on Machine Learning in Information Access.

Organizer and Program Chair of the 1994 AAAI Fall Symposium on Improving Instruction of Artificial Intelligence.

Editorial Positions

Current:

Associate Editor, *ACM Transactions on Computer-Human Interaction* (TOCHI).

Former:

Associate Editor, *ACM Transactions on the Web* (TWeb).

Associate Editor, *ACM Transactions on Information Systems* (TOIS).

Editorial Board, *Computational Linguistics*.

Editorial Board and Feature Editor, *IEEE Intelligent Systems*.

Advisory Boards

Member of the Language Usage Panel for the *American Heritage Dictionary* (2008-)

CACM Web Revamp Advisory Board (2007-)

Advised Google LEAD educational program, curriculum development (2009-2011)

NSF/CISE Advisory Committee (2004-2006)

Yahoo Search Scientific Advisory Board (2004-2006)

Altavista Scientific Advisory Board (2002-2004)

Instruction

Courses

2016, 2015, 2014, 2008, 2005, 2004, 2002, 2000, 1998 Information Visualization and Presentation, UC Berkeley.

2015, 2014, 2013, 2006, 2004 Applied Natural Language Processing, UC Berkeley

2014 Seminar on Exploratory Data Analysis, UC Berkeley

2013 Proseminar on Digital Humanities, UC Berkeley

2012 Analyzing Big Data with Twitter, UC Berkeley

2007, 2005 Search Engines: Society, Technology, and Business, UC Berkeley

2007, 2006, 2005, 2004, 2003, 2002, 2001, 1999 User Interface Design and Development, UC Berkeley.

2012, 2002, 2001 Foundations of Software Design / Data Structures, UC Berkeley

2001 – 1997 Information Organization and Retrieval, UC Berkeley (with Ray Larson).

2007 Seminar on Computational Analysis of Social Media and Search, UC Berkeley

2000 Information Visualization and Information Retrieval, Harvard University (with Stuart Shieber).

1999 Seminar on Text Data Mining, UC Berkeley.

1998 Seminar on Current Topics in Information Access, UC Berkeley.

1997 Seminar on Information Use (Informavores), UC Berkeley.

Tutorials

Tutorial: Designing for Search and Finding, Russell, Morris, CHI, Teevan, Hearst, *ACM SIGCHI*, April 2015 (delivered remotely).

Tutorial: Text, Tags, and Thumbnails: Latest Trends in Bioscience Literature Search, *Pharmaceutical and Health Division of the SLA, Spring Meeting*, March 2009.

Course on Faceted Metadata for Information Architecture and Search (with Preston Smalley and Corey Chandler), *ACM SIGCHI 2007*, San Jose, CA, May 2007.

Course on Faceted Metadata for Information Architecture and Search (with Preston Smalley and Corey Chandler), *ACM SIGCHI 2006*, Montreal, Canada, March 2006.

Tutorial on User Interfaces and Visualization for Search, *ACM SIGIR 2004*, Sheffield, England, July 2004.

Tutorial on Information Visualization, *ACM SIGCHI 2003*, Fort Lauderdale, FL, April 2003.

Tutorial on Designing Information Architectures for Search, *ACM SIGIR 2002*, August 2002, Tampere, Finland.

Tutorial on Designing Search-Centric Information Architectures, *ACM SIGIR 2001*, New Orleans, September 2001.

Tutorial on User Interfaces for Information Retrieval, *NSF Summer Workshop on Language Engineering*, Johns Hopkins University, July 2000.

Tutorial on User Interfaces for Information Retrieval, *ACM SIGIR 2000*, Athens, Greece, July 2000.

Tutorial on Advanced Techniques for Information Access (with Michael Pazzani), *AAAI-98*, Madison, WI, July 1998.

Publications

Books

Hearst, M. *Search User Interfaces*, Cambridge University Press, 2009.

Edited Collections

Fox, A, Hearst, M.A., and Chi, M. (Eds.) *Proceedings of the First Annual Conference of ACM Learning At Scale*, Atlanta, GA, March 5, 2014.

Hearst, M.A., Cohen, W., and Gosling, S. (Eds.), *Proceedings of the Fourth AAI International Conference on Weblogs and Social Media*, Washington, D.C., May 23, 2010.

Hearst, M. and Ostendorf M. (Eds.), *Proceedings of HLT-NAACL, joint conference of Human Language Technologies and the North American Association for Computational Linguistics*, Edmonton, Canada, May 2003.

Hearst, M., Tong, R., and Gey, F. (Eds.), *Proceedings of ACM SIGIR'99, the 22nd International Conference on Research and Development in Information Retrieval*, Berkeley, CA, August 1999.

Hearst, M. and Hirsh, H. (Eds.), *Working notes of the AAI Spring Symposium on Machine Learning in Information Access*, March 1996, AAI Press.

Hearst, M. (Ed.), *Proceedings of the 1994 AAI Fall Symposium on Improving Instruction of Introductory AI*, New Orleans, LA, November 1994.

Book Chapters

Hearst, M., "User Interfaces for Search" in *Modern Information Retrieval, Second Edition*, edited by Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Pearson Publishing Company, 2010.

Hearst, M., "User Interfaces and Visualization," in *Modern Information Retrieval*, edited by Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison-Wesley Publishing Company, 1999.

Hearst, M., "The Use of Categories and Clusters in Information Access Interfaces", in *Natural Language Information Retrieval*, Strzalkowski (ed.), Kluwer Academic Publishers, 1999.

Hearst, M., "Automated Discovery of WordNet Relations," in *WordNet: An Electronic Lexical Database*, Christiane Fellbaum (ed.), MIT Press, 1998.

Chen, F., Hearst, M., Kimber, D., Kupiec, J., Pedersen, J., Wilcox, L. "Metadata For Mixed-Media Access," in Sheth and Klas (eds), *Multimedia Data Management: Using Metadata to Integrate and Apply Digital Data*, McGraw Hill, 1998.

Hearst, M. and Schuetze, H. "Customizing a Lexicon to Better Suit a Computational Task," *Corpus Processing for Lexical Acquisition*, Bran Boguraev and James Pustejovsky (eds.), MIT Press, 1996. (Also appeared in the *Proceedings of the ACL SIGLEX Workshop on Lexical Acquisition*, Columbus OH, June 1993.)

Hearst, M. "Direction Based Text Interpretation as an Information Access Refinement," Paul Jacobs (Ed.), *Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval*, Lawrence Erlbaum Associates, 1992.

Invited Articles and Essays

Hearst, M. "Can Natural Language Processing Become Natural Language Coaching?", in the *Proceedings of ACL 2015: the 53rd Annual Meeting of the Association for Computational Linguistics*, Beijing, July 2015.

Hearst, M.A., What's Missing from Collaborative Search?, *IEEE Computer*, Special Issue on Collaborative Information Seeking, 47 (3), March 2014.

Hearst, M. "Why Programs Have Bugs," in *The Edge Annual Question*, in response to the question "What Is Your Favorite Deep, Elegant, or Beautiful Explanation?", 2012.

Hearst, M. "Findex," in *The Edge Annual Question*, in response to the question "What Scientific Concept Would Improve Everybody's Cognitive Toolkit?", 2011.

Hearst, M. "I Now Expect To Hear What Other People Think," in *The Edge Annual Question*, in response to the question "How is the Internet Changing the Way You Think?", 2010.

Hearst, M. "Setting the Right Conditions for Collaboration," *GSA Intergovernmental Solutions Newsletter*, special issue on Government by Collaboration, Spring 2010, p. 16-17.

Hearst, M. "The Decline of Text," in *The Edge Annual Question*, in response to the question "What Technology Will Change Things?", 2009.

Hearst, M. "Computational Analysis of Language Requires Understanding Language," in *The Edge Annual Question*, in response to the question "What Have You Changed Your Mind About? Why?", 2008.

Hearst, M. "The Rise of Usability" in *The Edge Annual Question*, in response to the question "What Are You Optimistic About?", 2007.

Hearst, M. "The Search Problem Is Solvable" in *The Edge Annual Question*, in response to the question "What Do You Believe Is True Even Though You Cannot Prove It?", 2005.

Hearst, M. "Hearst's Law" in *The Edge Annual Question*, in response to the question "What Is Your Law?", 2004.

Hearst, M. Clustering versus Faceted Categories for Information Exploration, in *Communications of the ACM*, 49(4), April, 2006.

Hearst, M., English, J., Sinha, R., Swearingen, K., and Yee, K-P., Finding the Flow in Web Site Search, *Communications of the ACM*, 45 (9), September 2002.

Hearst, M., "Next Generation Web Search: Setting Our Sites," *IEEE Data Engineering Bulletin*, 23 (3), invited paper for section on Next Generation Search, September 2000.

Hearst, M., "Untangling Text Data Mining," in the *Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics*, University of Maryland, June 20-26, 1999. (Reprinted in *The Oxford Handbook of Computational Linguistics*, Ruslan Mitkov (Ed.))

Hearst, M., "When Information Technology 'Goes Social,'" in *IEEE Intelligent Systems*, Trends & Controversies feature, 14 (1), January-February 1999.

Hearst, M., "Interfaces for Searching the Web," *Scientific American*, March 1997.

Hearst, M., "Research in Support of Digital Libraries at Xerox PARC. Part I: The Changing Social Roles of Documents," *D-Lib Magazine*, May 1996. <http://www.dlib.org/dlib/may96/05hearst.html>

Hearst, M., Kopec, G., and Brotsky, D. "Research in Support of Digital Libraries at Xerox PARC. Part II: Digital and Paper Documents," *D-Lib Magazine*, June 1996. <http://www.dlib.org/dlib/june96/06hearst.html>

Hearst, M. "Overview of the Symposium 'On the Instruction of Introductory AI'," Special issue of *SIGART Bulletin*, 6 (2): 3-7, April 1995.

Chen, F., Hearst, M., Kupiec, J., Pedersen, J., and Wilcox, L., "Metadata in Mixed-Media Access," *SIGMOD*

Record, 23 (4):64-71, special issue on Metadata, December, 1994.

Stonebraker, M., Hearst, M., and Potamianos, S. "A Commentary on the POSTGRES Rules System," *SIG-MOD Record*, 18 (3):5-11, September 1989.

Stonebraker, M. and Hearst, M. "Future Trends in Expert Database Systems," *Proceedings from the Second International Conference on Expert Database Systems*, L. Kerschberg (ed.), Benjamin/Cummings Publishing Company, Inc., 1989.

Refereed Journal Articles

Riedl, C., Zanibbi, R., Hearst, M.A., Zhu, S., Menietti, M., Crusan, J., Metelsky, I., and Lakhani, K.R., Detecting Figures and Part Labels in Patents: Competition-Based Development of Graphics Recognition Algorithms, *International Journal on Document Analysis and Recognition*, Feb 20, 2016.

Nakov, P., and Hearst, M., Semantic Interpretation of Noun Compounds Using Verbal and Other Paraphrases, *ACM Transactions on Speech and Language Processing*, special issue on Multiword Expressions, 10 (3), 2013.

Muralidharan, A. and Hearst, M., "Supporting Exploratory Text Analysis in Literature Study", *Literary and Linguistic Computing*, 27 (4), Dec 24, 2012.

Nakov, P., Divoli, A., and Hearst, M., "Do Peers See More in a Paper than its Authors?", *Advances in Bioinformatics*, Volume 2012, Article ID 750215.

Hearst, "Natural Search User Interfaces," *Communications of the ACM*, 54(11), p. 60-67, Nov 2011.

Divoli, A., Wooldridge, M., and Hearst, M., "Full Text and Figure Display Improves Bioscience Literature Search", *PLoS ONE* 5(4): e9619, April 2010.

Hearst, M., Divoli, A., Guturu, H, Ksikes, A., Nakov, P., Wooldridge, M., and Ye, J. "BioText Search Engine: beyond abstract search," *Bioinformatics* 23(16):2196-2197, June 2007.

Oliver, D., Bhalotia, G., Schwartz, A. S., Altman, B., Hearst, M. "Tools for Loading MEDLINE into a Local Relational Database," in *BMC Bioinformatics* 5 (146), Oct 2004.

Ivory, M. and Hearst, M., "Improving Web Site Design," *IEEE Internet Computing, Special Issue on Usability and the Web*, 6 (2), March/April 2002, 56-63.

Pevzner, L., and Hearst, M., "A Critique and Improvement of an Evaluation Metric for Text Segmentation," *Computational Linguistics*, 28 (1), March 2002, pp. 19-36.

Elliott, A., and Hearst, M., "A Comparison of the Affordances of a Digital Desk and Tablet for Architectural Image Use Tasks," in the *International Journal of Human-Computer Studies*, 56 (2), February 2002, 173-197.

Ivory, M. and Hearst, M., "The State of the Art in Automated Usability Evaluation of User Interfaces," in *ACM Computing Surveys*, 33 (4), December 2001, 470-516.

Hearst, M., "TextTiling: Segmenting Text into Multi-Paragraph Subtopic Passages," *Computational Linguistics*, 23 (1), pp. 33-64, March 1997.

Palmer, D., and Hearst, M., "Adaptive Multilingual Sentence Boundary Disambiguation," *Computational Linguistics*, 23 (2), 241-267, June 1997.

Rao, R., Pedersen, J. O., Hearst, M. A., et al., "Rich Interaction in the Digital Library," *Communications of the ACM*, 38 (4), 29-39, April 1995.

Larson, R. L., Plaunt, C., Woodruff, A. G., and Hearst, M.A., "The Sequoia 2000 Electronic Repository," *Digital Technical Journal*, 7 (3), pp. 50-65, 1995.

Refereed Proceedings and Periodical Publications

Hearst, M.A., Laskowski, P., and Silva, L., Evaluating Information Visualization via the Interplay of Heuristic Evaluation and Question-Based Scoring, *Proceedings of CHI*, 2016.

Head, A., Appachu, C., Hearst, M.A., Hartmann, B. Tutorons: Generating Context-Relevant, On-Demand Explanations and Demonstrations of Online Code, *Proceedings of VL/HCC*, October 2015.

Coetzee, D., Lim, S., Fox, A., Hartmann, B., and Hearst, M.A. Structuring Interactions for Large-Scale Synchronous Peer Learning, *Proceedings of ACM CSCW*, March 2015.

Alspaugh, S., Chen, B., Lin, J., Ganapathi, A., Hearst, M.A., Katz, R., *Analyzing Log Analysis: An Empirical Study of User Log Mining*, USENIX LISA 2014.

Stephens-Martinez, K., Hearst, M.A., and Fox, A., Monitoring MOOCs: Which Information Sources Do Instructors Value?, *Proceedings of ACM Learning At Scale*, March 2014.

Coetzee, D., Fox, A., Hearst, M.A., and Hartmann, B. Chatrooms in MOOCs; All talk and no action, *Proceedings of ACM Learning at Scale*, March 2014.

Kong, N., Hearst, M.A., and Agrawala, M., Extracting References Between Text and Charts via Crowdsourcing, *Proceedings of CHI*, May 2014.

Coetzee, D., Fox, A., Hearst, M.A., and Hartmann, B. Should your MOOC Forum Use a Reputation System? *Proceedings of ACM CSCW*, February 2014.

Hearst, M.A., and Degler, D., Sewing the Seams of Sensemaking: A Practical Interface for Tagging and Organizing Saved Search Results, *Proceedings of the 7th Annual Symposium on Human-Computer Interaction and Information Retrieval (HCIR 2013)*, Vancouver, Canada, October 2013.

Luo, J. and Wing, C., and Yang, H., and Hearst, M.A., The Water Filling Model and The Cube Test: Multi-Dimensional Evaluation for Professional Search, *Proceedings of CIKM 2013*, Burlingame, CA, October 2013.

Robson, C., and Hearst, M.A., Comparing the use of Social Networking and Traditional Media Channels for Promoting Citizen Science, *Proceedings of the 16th ACM Computer Supported Cooperative Work*, San Antonio, TX, 2013.

Kaisser, M., Hearst, M.A., and Lowe, J.B., Evidence for Varying Search Results Summary Lengths, in the *Proceedings of ACL/HLT'08*, Columbus, OH, 2008.

Nakov, P., and Hearst, M.A., Solving Relational Similarity Problems Using the Web as a Corpus, in the *Proceedings of ACL/HLT'08*, Columbus, OH, 2008.

Fiore, A., and Taylor, L.S., and Mendelsohn, G., and Hearst, M.A., Assessing Attractiveness in Online Dating Profiles, in the *Proceedings of CHI 2008*, Florence, Italy, 2008.

Hearst, M.A., and Rosner, D., Tag Clouds: Data Analysis tool or Social Signaller?, in *Proceedings of IEEE HICSS 2008, Social Spaces minitrack*.

Divoli, A., Hearst, M.A., Wooldridge, M.A., Evidence for Showing Gene/Protein Name Suggestions in Bio-science Literature Search Interfaces, in *Proceedings of PSB 2008*, Kona, HI, 2008.

Schwartz, A., Divoli, A., and Hearst, M.A., Multiple Alignment of Citation Sentences with Conditional Random Fields and Posterior Decoding, in the *Proceedings of EMNLP-CoNLL-2007*, Prague, June 2007.

Stoica, E., Hearst, M., and Richardson, M., Automating Creation of Hierarchical Faceted Metadata Structures, in the *Proceedings of NAACL/HLT 2007*, Rochester, NY, April 2007.

Boshernitsan, M., Graham, S.L., and Hearst, M., Aligning Development Tools with the Way Programmers Think About Code Changes, in the *Proceedings of the Conference on Human Factors in Computing Systems (CHI2007)*, San Jose, April 2007.

- Nakov, P. and Hearst, M., Using Verbs to Characterize Noun-Noun Relations, in the *Proceedings of AIMSA 2006*, Bulgaria, September 2006.
- Yee, K-P., Wagner, D., Hearst, M., Bellovin, S. Prerendered User Interfaces for High-Assurance Electronic Voting. In *Proceedings of the USENIX/ACCURATE Electronic Voting Technology Workshop*, Vancouver, July, 2006.
- Stoica, E. and Hearst, M., Predicting Gene Functions from Text Using a Cross-Species Approach, in the *Pacific Biocomputing Symposium (PSB'06)*, Maui, HI, January, 2006.
- Dhamija, R., Tygar, J. D., and Hearst, M., "Why Phishing Works," in the *Proceedings of the Conference on Human Factors in Computing Systems (CHI2006)*, Montreal, April 2006.
- Rosario, B. and Hearst, M. "Multi-way Relation Classification: Application to Protein-Protein Interaction," in *HLT-NAACL'05*, Vancouver, 2005.
- Nakov, P., and Hearst, M., "Using the Web as an Implicit Training Set: Application to Structural Ambiguity Resolution," in *HLT-NAACL'05*, Vancouver, 2005.
- Nakov, P. and Hearst, M., "A Study of Using Search Engine Page Hits as a Proxy for n-gram Frequencies," in *RANLP'05*, Borovets, Bulgaria, 2005.
- Aragon, C., and Hearst, M., "Improving Aviation Safety with Information Visualization: A Flight Simulation Study," in *Proceedings of ACM CHI*, April 2005.
- Nakov, P., and Hearst, M. "Search Engine Statistics Beyond the n-gram: Application to Noun Compound Bracketing," in the *Proceedings of CoNLL-2005, Ninth Conference on Computational Natural Language Learning*, Ann Arbor, MI, June 2005.
- Rosario, B. and Hearst, M. "Classifying Semantic Relations in Bioscience Text," in the *Proceedings of ACL 2004*, Barcelona, July 2004.
- Yee, K-P., Swearingen, K., Li, K., and Hearst, M., Faceted Metadata for Image Search and Browsing, in *Proceedings of ACM CHI 03 Conference on Human Factors in Computing Systems*, April, 2003.
- Schwartz, A., and Hearst, M., "A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text," in the *Proceedings of the Pacific Symposium on Biocomputing (PSB 2003)*, Kauai, Jan 2003.
- Rosario, B., Hearst, M., and Fillmore, C., "The Descent of Hierarchy, and Selection in Relational Semantics," *Proceedings of the Association for Computational Linguistics, (ACL-02)*, July 2002.
- Ivory, M. and Hearst, M., "Statistical Profiles of Highly Rated Web Sites," *Proceedings of ACM CHI 02 Conference on Human Factors in Computing Systems*, April, 2002.
- Yee, P., Fischer, D., Dhamija, R., and Hearst, M. "Animated Exploration of Graphs with Radial Layout," in the *Proceedings of the IEEE Information Visualization Symposium*, San Diego, CA, October 2001.
- Ivory, M., Sinha, R., and Hearst, M. "Empirically Validated Web Page Design Metrics" in *CHI 2001, ACM Conference on Human Factors in Computing Systems*, 2001.
- Rosario, B., and Hearst, M. "Classifying the Semantic Relations in Noun Compounds via a Domain-Specific Lexical Hierarchy," in the *Proceedings of 2001 Conference on Empirical Methods in Natural Language Processing (EMNLP 2001)*, Pittsburgh, PA, June 2001.
- Sinha, R., Hearst, M., Ivory, M., and Draisin, M. "Content or Graphics? An Empirical Analysis of Criteria for Award-Winning Websites," in the *Proceedings of 7th Conference on Human Factors and the Web*, June 2001.
- Ivory, M., Sinha, R., and Hearst, M. "Preliminary Findings on Quantitative Measures for Distinguishing Highly Rated Information-Centric Web Pages" in the *Proceedings of 6th Conference on Human Factors and the Web*, Austin, Texas, 2000.
- Chen, M., Hearst, M., Hong, J., and Lin, J., "Cha-Cha: A System for Organizing Intranet Search Results," in

the *Proceedings of the 2nd USENIX Symposium on Internet Technologies & Systems (USITS)*, Boulder, CO, October 11-14, 1999.

Pratt, W., Hearst, M., and Fagan, L. "A Knowledge-Based Approach to Organizing Retrieved Documents," In *AAAI-99: Proceedings of the Sixteenth National Conference on Artificial Intelligence*, 1999.

Hearst, M. and Karadi, C., "Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy" in the *Proceedings of the 20th Annual International ACM/SIGIR Conference*, Philadelphia, PA, July 1997.

Hearst, M. and Pedersen, J., "Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results," in the *Proceedings of the 19th Annual International ACM/SIGIR Conference*, Zurich, August 1996.

Sahami, M., Hearst, M., and Saund, E., "Applying the Multiple Cause Mixture Model to Text Categorization," in the *Proceedings of the 13th International Conference on Machine Learning (ICML)*, Bari (Italy), July 3-6th, 1996.

Hearst, M., "Improving Full-Text Precision on Short Queries using Simple Constraints," *Proceedings of the Fifth Annual Symposium on Document Analysis and Information Retrieval (SDAIR)*, Las Vegas, NV, April 1996.

Pirolli, P., Schank, P., Hearst, M., and Diehl, C., "Scatter/Gather Browsing Communicates the Topic Structure of a Very large Text Collection," *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, May 1996.

Hearst, M., Pedersen, J., and Karger, D., "Scatter/Gather as a Tool for the Analysis of Retrieval Results," Working Notes of the *AAAI Fall Symposium on AI Applications in Knowledge Navigation*, November 1995, Cambridge, MA.

Hearst, M. "TileBars: Visualization of Term Distribution Information in Full Text Information Access," *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 59-66, Denver, CO, May 1995.

Hearst, M. "Using Categories to Provide Context for Full-Text Retrieval Results," *Proceedings of RIAO '94: Intelligent Multimedia Information Retrieval Systems and Management*, pp. 115-130, Rockefeller University, NY, October 1994.

Palmer, D., and Hearst, M. "Adaptive Sentence Boundary Disambiguation," *Proceedings of the 4th Conference on Applied Natural Language Processing*, pp. 78-83, Stuttgart, Germany, October 1994.

Chen, F., Hearst, M., Kupiec, J., Pedersen, J., and Wilcox, L., "Mixed-Media Access," *Proceedings of Digital Libraries '94*, College Station, TX, June 1994.

Hearst, M. "Multi-Paragraph Segmentation of Expository Text," *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, June 1994.

Hearst, M. and Plaunt, C. "Subtopic Structuring for Full-Length Document Access," in the *Proceedings of ACM/SIGIR*, Pittsburgh, PA. 1993.

Resnik, P. and Hearst, M. "Structural Ambiguity and Conceptual Relations," in the *Proceedings of the ACL Workshop on Very large Corpora*, Columbus, OH. June 1993.

Hearst, M. "Cases as Structured Indexes for Full-Length Documents," *Proceedings of the 1993 AAAI Spring Symposium on Case-based Reasoning and Information Retrieval*, Stanford, CA, 1993.

Hearst, M. "Automatic Acquisition of Hyponyms from Large Text Corpora," *Proceedings of the Fourteenth International Conference on Computational Linguistics*, Nantes, France, July 1992.

Hearst, M. "Noun Homograph Disambiguation using Local Context in large Text Corpora," *The Proceedings of the 7th Annual Conference of the UW Centre for the New OED and Text Research: Using Corpora*, Oxford, pp 1-22, October 1991.

Jaramillo, N. and Hearst, M. "Acquiring the Semantics of Simple Phrasal Patterns Using COBUILD", *Ma-*

chine Learning of Natural Language and Ontology: Proceedings from the AAAI Spring Symposium, Stanford, CA, March 1991.

Hearst, M. "A Hybrid Approach to Restricted Text Interpretation," AAAI Spring Symposium on Text Based Intelligent Systems, Stanford University, March 1990.

Refereed Short Papers, Workshop Papers, Demonstrations, and Videos

Booten, K., and Hearst, M.A., Patterns of Wisdom: Discourse Level Style in Multi-Sentence Quotations, *Proceedings of NAACL 2016*, June 2016.

Gagliano, A., Paul, E., Booten, K., and Hearst, M.A., Intersecting Word Vectors to Take Figurative Language to New Heights, *NAACL Workshop on Computational Linguistics for Literature*, June 2016.

Milli, S., and Hearst, M.A., Augmenting Course Material with Open Access Textbooks, *11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2016)*, June, 2016.

Hearst, M.A., Adar, E., Kosara, R., Munzner, T. Schwabisch, J., and Shneiderman, B. Vis, The Next Generation: Teaching Across the Researcher-Practitioner Gap, *Panel for IEEE Infoviz*, October 2015.

Appleman, Gupta, Rajagopal, Shishido, Hearst, Exploring Data For Fun And Profit: Case Study of Jeopardy! Poster, *IEEE Infoviz*, October, 2015.

Anand, D., DeSouza, A., Shreyas, Sharma, S., Starostenko, V., Ryokai, K., Hearst, M.A., Wordcraft: Playing with Sentence Structure, *Work in Progress, ACM CHI Play*, 2015.

Hearst, M.A., Coetzee, D., Fox, A., and Hartmann, B., All It Takes Is One: Evidence for a Strategy for Seeding Large Scale Peer Learning Interactions, *ACM Learning at Scale 2015*, Work in Progress Track.

Muralidharan, A., and Hearst, M.A., Improving the Recognizability of Syntactic Relations for Information Retrieval Interfaces, *Proceedings of ACL 2014*, Short Paper Track.

Ginosar, S. and Hearst, M.A., A Study of the Use of Current Speech Recognition in an Information-Intensive Task, *CHI 2014 Workshop on Designing Speech and Language Interactions*, 2014. (Refereed but not officially published.)

Lim, S., Coetzee, D., Hartmann, B., Fox, A., and Hearst, M.A., Initial Experiences with Small Group Discussions in MOOCs, *Proceedings of ACM Learning at Scale*, Work in Progress Track, 2014.

Muralidharan, A., Hearst, M.A., and Fan, C., WordSeer: A Knowledge Synthesis Environment for Textual Data, *CIKM 2013 Demo Track*, Burlingame, CA, October 2013.

Alspaugh, S., Ganapathi, A., Hearst, M.A., and Katz, R. Building Blocks for Exploratory Data Analysis, *KDD Workshop on Interactive Data Exploration and Analysis (IDEA)*, KDD, Chicago, IL, Aug 2013.

Muralidharan, A. and Hearst, M.A., Finding Literary Themes with Relevance Feedback, Poster paper in the Proceedings of the *Workshop on Human-computer Interaction and Information Retrieval (HCIR 2012)*, Cambridge, MA, Oct 2012.

Muralidharan, A. and Hearst, M.A., A sensemaking environment for literary text, Works in Progress, *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (ACM CHI)*, 2012.

Muralidharan, A. and Hearst, M.A., WordSeer: Exploring Language Use in Literary Text, Poster paper in the Proceedings of the *Workshop on Human-computer Interaction and Information Retrieval (HCIR 2011)*, Mountain View, CA, Oct 2011.

Hearst, M.A. and Stoica, E., NLP Support for Faceted Navigation in Scholarly Collections, *ACL 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries (NLPIR4DL)*, 2009.

Hearst, M.A. and Dumais, S. Blogging Together: An Examination of Group Blogs, *3rd International AAAI Conference on Weblogs and Social Media (ICWSM 2009)*, San Jose, CA 2009.

Nagarajan, M. and Hearst, M.A., An Examination of Language Use in Online Dating Profiles, *3rd International AAAI Conference on Weblogs and Social Media (ICWSM 2009)*, San Jose, CA 2009.

Hearst, M., Hurst, M., and Dumais, S. What should blog search look like? Workshop on Search and Social Media, ACM CIKM 2008.

Hearst, M. UIs for Faceted Navigation: Recent Advances and Remaining Open Problems. Workshop on Computer Interaction and Information Retrieval, HCIR 2008, Redmond, WA, October 2008.

Hearst, Divoli, Wooldridge, Ye, Exploring the efficacy of caption search for bioscience journal search interfaces, in the *Proceedings of BioNLP 2007, ACL 2007 Workshop*.

Hearst, M., "Design Recommendations for Hierarchical Faceted Search Interfaces," at the *ACM SIGIR Workshop on Faceted Search*, Seattle, WA, August, 2006.

Stoica, E., and Hearst, M., "Demonstration: Using WordNet to Build Facet Hierarchies," at the *ACM SIGIR Workshop on Faceted Search*, Seattle, WA, August 2006.

Snydal, J. and Hearst, M. "ImprovViz: Visual Explorations of Jazz Improvizations," in the *Proceedings of ACM CHI 2005 Conference Companion*, Portland, OR, April 2005.

Nakov, P., Schwartz, A., Wolf, B., and Hearst, M. "Scaling Up BioNLP: Application of a Text Annotation Architecture to Noun Compound Bracketing," in *ACL/ISMB BioLINK SIG: Linking Literature, Information and Knowledge for Biology*, Detroit, MI, June 2005.

Nakov, P., Schwartz, A., Wolf, B., Hearst, M., "Supporting Annotation Layers for Natural Language Processing," in the *ACL 2005 Poster/Demo Track*, Ann Arbor, MI, June 2005.

Hearst, M., "Teaching Applied Natural Language Processing: Triumphs and Tribulations," in the *ACL 2005 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing And Computational Linguistics*, Ann Arbor, MI, June 2005.

Hearst, M., "The HCI Emphasis at UC Berkeley's SIMS," in the *CHI 2005 workshop on Graduate Education in Human-Computer Interaction*, Portland, OR, April 2005.

Kee, K-P, and Hearst, M., "Content-Centered Discussion Mapping," in *Online Deliberation 2005 / DIAC-2005*, Palo Alto, CA, June 2005.

Nakov, P., Schwartz, A., and Hearst, M., "Citances: Citation Sentences for Semantic Analysis of Bioscience Text," in the *SIGIR'04 Workshop on Search and Discovery in Bioinformatics*, Sheffield, July 2004.

Stoica, E. and Hearst, M. Nearly-Automated Metadata Hierarchy Creation, in the *Companion Proceedings of HLT-NAACL'04*, Boston, May 2004.

Nakov, P. and Hearst, M., "Category-based Pseudowords," in the *Companion Volume of the Proceedings of HLT-NAACL'03*, May 2003.

English, J., Hearst, M., Sinha, R., Swearingen, K., and Yee, P., "Hierarchical Faceted Metadata in Site Search Interfaces," *Proceedings of ACM CHI 02 Conference on Human Factors in Computing Systems Conference Companion*, April 2002.

Elliott, A., and Hearst, M. "How Large Should a Digital Desk Be?," *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), Conference Companion*, The Hague, Netherlands, April 2000.

Glaser, D. and Hearst, M. "Space Series: Simultaneous display of spatial and temporal data," In the *Proceedings of the IEEE Symposium on Information Visualization (late-breaking results)*, October 1999.

Chen, M. and Hearst, M., "Presenting Web Site Search Results in Context: A Demonstration," in the *Proceedings of the 21st Annual International ACM/SIGIR Conference*, Melbourne, Australia, August 1998.

Hearst, M. and Karadi, C., "Searching and Browsing Text Collections with Large Category Hierarchies," in the *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), Conference*

Companion, Atlanta, GA, March 1997.

Sahami, M., Hearst, M., and Saund, E., "Applying the Multiple Cause Mixture Model to Text Categorization," in the *AAAI Spring Symposium on Machine Learning in Information Access* Stanford, CA, 1996.

Hearst, M. and Pedersen, J., "Visualizing Information Retrieval Results: A Demonstration of the TileBar Interface," Video track in the *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, May 1996. Derived from video of IJCAI 95.

Hearst, M. and Pedersen, J., "Revealing Collection Structure through Information Access Interfaces," Video track in the *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Montreal, Canada, Volume II, 2047-2048, 1995.

Hearst, M., and Grefenstette, G. "Refining Automatically-Discovered Lexical Relations: Combining Weak Techniques for Stronger Results," in Carl Weir (ed.), *Statistically-Based Natural Language Programming Techniques: Papers from the 1992 Workshop* Technical Report W-92-01, AAAI Press, Menlo Park, CA, 1992.

Non-refereed Conference Papers and Reports

Agichtein, E., Hearst, M.A., Soboroff, I., The Search and Social Media Workshop at SIGIR 2009, SIGIR Forum, 2010.

Nakov, P., and Hearst, M. UCB System Description for the WMT 2007 Shared Task, in *Proceedings of Second Workshop on Statistical Machine Translation*, co-located with ACL-2007, Prague, June 23, 2007.

Nakov, P., and Hearst, M. UCB System Description for SemEval Task #4, in *Proceedings of SemEval-2007 Workshop*, co-located with ACL-2007, Prague, June 23-24, 2007.

Hersh, W. and Cohen, A. and Yang, J. and Bhupatiraju, R.T. and Roberts, P. and Hearst, M., TREC 2005 Genomics Track Overview, Proceedings of TREC 2005, Gaithersburg, MD, 2005.

Nakov, P, Schwartz, A. S., Stoica, E., Hearst, M., BioText Team Experiments for the TREC 2004 Genomics Track *Proceedings of TREC 2004* Gaithersburg, MD, 2005.

Bhalotia, G., Nakov, P., Schwartz, A. S, and Hearst, M.A., BioText Team Report for the TREC 2003 Genomics Track, *Proceedings of TREC 2003*, Gaithersburg, MD, November 2004.

Hearst, M., Pedersen, J., Pirolli, P., Schuetze, H., Grefenstette, G. and Hull, D., "Four TREC-4 Tracks: the Xerox Site Report," *Proceedings of the Fourth Text REtrieval Conference (TREC-4)*, Nov 1-3, Arlington, VA, 1996.

Schuetze, H., Pedersen, J., and Hearst, M., "Xerox TREC-3 Report: Combining Exact and Fuzzy Predictors", *Proceedings of the Text REtrieval Conference (TREC-3)*, Nov 2-4, Arlington, VA, 1994.

Edited Features

Hearst, M., et al., "The Debate on Automated Essay Grading," *IEEE Intelligent Systems*, Trends & Controversies feature, 15 (5), September-October, 2000.

Hearst, M., and Pazzani, M. "Knowledge Discovery from Data?" *IEEE Intelligent Systems*, Trends & Controversies feature, 15 (2), March-April, 2000.

Hearst, M., and Hirsh, H. (Eds.) "AI's Greatest Trends & Controversies," *IEEE Intelligent Systems*, Trends & Controversies feature, 15 (1), January-February, 2000.

Hearst, M., et. al., "Mixed-Initiative Interaction," *IEEE Intelligent Systems*, Trends & Controversies feature, 14 (5), September-October 1999.

Hearst, M., et. al., "Building Intelligent Systems One E-Citizen at a Time," *IEEE Intelligent Systems*, Trends

& Controversies feature, 14 (3), May-June 1999.

Hearst, M., et. al., “The Changing Relationship between Information Technology and Society,” *IEEE Intelligent Systems*, Trends & Controversies feature, 14 (1), January-February 1999.

Hearst, M., et. al., “Information Integration,” *IEEE Intelligent Systems*, Trends & Controversies feature, 13 (5), September-October 1998.

Hearst, M., et. al., “Support Vector Machines,” *IEEE Intelligent Systems*, Trends & Controversies feature, 13 (4), July-August 1998.

Hearst, M., et. al., “Sketching Intelligent Systems,” *IEEE Intelligent Systems*, Trends & Controversies feature, 13 (3), May-June 1998.

Hearst, M., et. al., “Innovations in Electronic Academic Publishing,” *IEEE Intelligent Systems*, Trends & Controversies feature, 13 (1), January-February 1998.

Hearst, M., et. al., “Banter on Bayes: debating the usefulness of Bayesian approaches to solving practical problems,” *IEEE Intelligent Systems*, Trends & Controversies feature, 12 (6), November-December, 1997.

Hearst, M., et. al., “Dissonance on Audio Interfaces,” *IEEE Intelligent Systems*, Trends & Controversies feature, 12 (5), September-October, 1997.

Hearst, M., et. al., “Evolution-based Computational Models of Cognition,” *IEEE Intelligent Systems*, Trends & Controversies feature, 12 (4), July-August, 1997.

Hearst, M., et. al., “The Role of Aesthetics in Intelligent Systems,” *IEEE Intelligent Systems*, Trends & Controversies feature, 12 (3), May-June, 1997.

Invited Book Reviews

Hearst, M. “Review of ‘Empirical Methods for Artificial Intelligence’,” *Computational Linguistics* 22 (2), 1996.

Hearst, M., “Review of *Using Large Corpora*,” *Journal of Information Processing and Management*, 31 (5), 1995.

Hearst, M. “Review of ‘Statistical Language Learning,’” *AI Magazine*, 15 (3):88-89, Fall 1994.

Hearst, M. “Review of ‘Text and Texture’,” *Computational Linguistics*, 20 (1), March 1994.

Invited Talks

Keynote Talks – Conferences

Keynote speaker, ACL 2015, *Can Natural Language Processing Become Natural Language Coaching?*, Beijing, July 2015.

Keynote speaker, ECIR 2015, *Still Haven’t Found What I’m Looking For; Ideas for Search Research*, Vienna, Austria, March 2015.

Keynote speaker, Society for Computers in Psychology (SCiP), *Towards Collaborative Learning At Scale*, Long Beach, CA, November 2014.

Keynote speaker, ACM SIGIR, *Seeking Simplicity in Search User Interfaces*, Gold Coast, Australia, July 2014.

Keynote speaker, IS&T/SPIE Electronic Imaging Conference, *Ideas for Future Search Interfaces for Multimedia*, South San Francisco, CA, February 2013.

Keynote speaker, HCIR, *Human Computer Information Retrieval Ideas for Research*, Cambridge MA, Octo-

ber 2012.

Keynote speaker, *Blame it on the VTR: When Will Video and Audio Replace Text?*, Mobile Voice Conference, San Francisco, CA, March 2012.

Keynote speaker, *Future Trends in Search User Interfaces*, ACM Hypertext, Eindhoven, Netherlands, June 2011.

Keynote speaker, *Designing Search for Humans*, Usability Professionals' Association DC Chapter Conference, Oct 2010.

Keynote speaker, *The Future of Search Interfaces*, iKnow conference, Graz, Austria, September 2010.

Keynote speaker, *Designing Search for Humans*, Enterprise Search Summit, NY, NY, May 2010.

Keynote speaker, *IT Research to Support Open Government*, iSchools iConference' 10, University of Illinois, February 2010.

Keynote speaker, *Social Technology*, NASA CIDE conference, Mountain View, CA, October 2009.

Keynote speaker, *Some Thoughts on Social Tagging*, Taxonomy Bootcamp, San Jose, November 2007.

Keynote speaker, *The Interplay of Content Analysis and Visualization for Improving Search and Text Mining*, NVAC Consortium Meeting, Stanford, CA, May 23, 2006.

Keynote speaker, *Metadata in Search Interfaces*, AAAI conference, Pittsburgh, PA, July 2005.

Keynote speaker, *Untangling Text Data Mining*, the annual meeting of the Association for Computational Linguistics, July 1999.

Keynote Talks – Workshops

Keynote speaker, *Still Haven't Found What I'm Looking For; Ideas for Search Research*, 1st International Workshop on Novel Web Search Interfaces, affiliated with CIKM, (delivered remotely), October 22, 2015.

Keynote speaker, *Can Natural Language Processing be Natural Language Coaching*, Workshop on the Corpus of Contemporary Chinese in Taiwan, (delivered remotely), September 24, 2015.

Keynote speaker, 6th Forum for Information Retrieval Evaluation (FIRE) *Seeking Simplicity in Search User Interfaces*, Bangalore, India, (delivered remotely), December 2014.

Keynote speaker, Knowledge Management in Organizations / Learning Technologies for Education, *Towards Collaborative Learning at Scale*, Santiago Chile, (delivered remotely), September 2014.

Keynote speaker, 4th Workshop on Controlled Natural Language (CNL 2014), *Exploratory Text Analysis at the Middle Distance*, Galway Ireland (delivered remotely), August 2014.

Keynote speaker, 8th Russian Summer School in Information Retrieval *Seeking Simplicity in Search User Interfaces*, Nizhny Novgorod, Russia (delivered remotely), September 2014.

Keynote speaker, ACL Workshop on Interactive Language Learning, Visualization, and Interfaces, *The HCI Angle* Baltimore, June 2014.

Keynote speaker, CIKM ESAIR Workshop, *How Can Search Interfaces Enhance the Value of Semantic Annotations (and Vice Versa)?*, Burlingame CA, October 2013.

Keynote speaker, KDD IDEA Workshop, *Exploratory Text Analysis and the Middle Distance*, Chicago, IL, August 2013.

Keynote speaker, *Analyzing Text at the Middle Distance between the Close Read and Culturomics*, Language in Social Media Workshop at NAACL-HLT, Montreal, June 2012.

Keynote speaker, *Bringing HCI to the U.S. Federal Government*, ACM CHIMIT (Computer Human Interaction for the Managing of Technology), Boston, MA, December 2011.

Keynote speaker, *Future Trends in Search User Interfaces*, PSU 2011 IST Graduate Student Symposium, March 2011.

Keynote speaker, *Thoughts on Patent Search*, The 3rd International Workshop on Patent Information Retrieval, CIKM Conference, Toronto, October 2010.

Distinguished Lecture Series

Microsoft Computing in the 21st Century conference, *Learning at Scale as a Driver of Innovation*, Seoul, South Korea, Nov 2016.

Yahoo! Research Big Thinker talk series, *MOOCs in Progress*, Sunnyvale, CA, May 25, 2016.

University of Washington Computer Science Department's Distinguished Lecture Series, *Seeking Simplicity in Search User Interfaces*, Seattle, WA, Oct 30, 2104.

Cray Distinguished Speaker Series, *Exploratory Text Analysis at the Middle Distance*, University of Minnesota, April 2014.

Distinguished Lecture Series, University of Toronto, *Unambiguous + Unlimited = Unsupervised*, January 2006.

Distinguished Seminar series, program in Computational Biology and Chemistry at Carnegie Mellon University, *The Berkeley BioText Project*, April 2004.

Columbia University Distinguished Lectures in Computer Science, *Using Words to Search a Thousand Images*, October 2002.

Featured Panelist

Panel organizer and moderator, *The Changing Role of the Professor in the Digital World*, Online Learning Summit, Stanford, CA, April 16, 2016.

Invited panelist, *Computer-Aided Personalized Education*, CRA Computing Community Consortium Meeting, Washington DC, November 12, 2015.

Panelist, *Crowdsourcing and the Academy, Exploring promotes and problems of collective intelligence methods in humanities and social science research*, UC Berkeley, Nov 6, 2015.

Panelist, *Vis, The Next Generation: Bridging the Researcher-Practitioner Gap*, IEEE Infoviz Conference, October 29, 2015.

Invited panelist, *Edcast Future Learning 2020 Summit*, Stanford University, May 29-30, 2015.

Invited panelist, *Data Science Education*, UC Santa Cruz sponsored meetup, Palo Alto, CA, December 2014.

Invited panelist, *CS Research on MOOCs and Online Education*, CRA Snowbird Conference, July 2014.

Invited moderator, *The Online Revolution, Education for Everyone* with Andrew Ng, ACM Learning Webinar, April 2014.

Invited panelist, *Big Data*, CIKM Conference, Burlingame CA, October 2013.

Invited panelist, *Next Generation Information Experiences*, Microsoft Research Faculty Summit, Redmond WA, July 2012.

Invited panelist, *Open Innovation*, NITRD/NARA Summit on R&D for Open Government, Washington DC, March 2011.

Invited panelist, *Beyond the Text Box*, Bigthink Conference, San Francisco, Feb 2011.

Invited panelist, *Search is Dead, Long Live Search*, WWW conference, Raleigh NC, April 2010.

Invited panelist, *The Decline of Text*, Transparent Text Symposium, IBM Cambridge, October 2009.

Invited panelist, *Searching for our Future*, iConference, U Michigan, October 2006.

Invited panelist, *The Emerging Law of Search Engines*, Pike & Fischer audio conference, March, 2006.

Invited panelist, *User Interfaces for the Developing World*, Berkeley in Silicon Valley (Alumni event), Santa Clara, CA, April 2004.

Invited panelist, *Literature-Based Discovery: New Trends and Techniques*, ASIST, October 2003.

Invited panelist, *Measuring Information Architecture Quality – Prove it or not*, ACM SIGCHI Conference, March 2001.

Invited panelist, *Web Data Mining*, KDD Conference, August 1997.

Other Invited Talks and Lectures

Invited talk, *Using Information Graphics*, Lawrence Berkeley Laboratory Communications Council, Sep 26, 2016.

Invited talk, *Teaching Information Visualization*, Berkeley Communications Conference, UC Berkeley, June 1, 2016.

Invited talk, Workshop II: Culture Analytics and User Experience Design, *Student-Powered Cross-Checking of Visualization Designs*, UCLA, Los Angeles, CA, April 14, 2016.

Invited talk, Stanford PCD Seminar, *Teaching as Coaching and Learning at Scale*, Jan 22, 2016.

Invited talk, SFSU Big Data Graduate Seminar Colloquia, *Seeking Simplicity in Search User Interfaces*, November 18th, 2015.

Invited talk, UC Berkeley Haas MORS group, *Towards Collaborative Learning At Scale*, November 3, 2015.

Invited talk, Open Table Open Science Event, *Teaching Infoviz: Teaching As Coaching*, San Francisco, September 1, 2015

Invited talk, Data Dialogs, UC Berkeley iSchool, *Discovering Best Practices in Visual Exploratory Data Analysis*, November 14, 2014.

Invited talk, Digital Humanities Seminar, UC Berkeley, *Exploratory Text Analysis and the Middle Distance*, November 11, 2014.

Invited talk, University of Washington *Towards Collaborative Learning At Scale*, Oct 31, 2014.

Invited speaker, Big Data and Psychology Workshop, *Towards Collaborative Learning at Scale*, UC Berkeley, October 10-11, 2014.

Invited talk, MIT HCI Seminar *Towards Collaborative Learning At Scale*, April 11, 2014.

Guest lecture, Cogsci 1, UC Berkeley, *Natural Language Processing*, CogSci 1, March 18, 2014.

Invited talk, Science @ Cal, *Natural Search User Interfaces*, UC Berkeley, February 15, 2014.

Invited talk, UC Berkeley Graduate School of Education Colloquium Series, *Towards Collaborative Learning at Scale*, UC Berkeley, February 10, 2014.

Invited talk, Google Social Interactions PI Meeting, *Towards Collaborative Learning at Scale*, Lake Tahoe, CA, January 30, 2014.

Invited talk, UC San Diego Design At Large Seminar, *Towards Collaborative Learning At Scale*, San Diego, CA, Dec 5, 2013.

Invited talk, Prismatic, *Is Text Use Declining?*, San Francisco, CA, June 26, 2013.

Invited talk, STEM Native American High School Students, UC Berkeley, *Social Media Research*, June 26, 2013.

Invited talk, UC Santa Cruz Future of Games group, *Exploratory Text Analysis and the Middle Distance*, June 7, 2013.

Invited talk, Google headquarters, *HCIR Keynote Reprise*, May 24, 2013.

Invited panel, Women in CS Stanford/Berkeley Meeting, Berkeley, CA, March 9, 2013.

Invited talk, University of Michigan School of Information Yahoo! Lecture Series, *An HCI Researcher Goes to Washington: My Experience Helping Build Government IT 2.0*, March 29, 2013.

Invited talk, UC Berkeley Townsend Center for the Humanities Brown Bag Lunch Series, *Building Text-Analysis Tools for Literary Study*, March 18, 2013. With Bryan Wagner.

Invited talk, Learning Mode: Critical Issues in Online Education, *The Potential: New Methods of Learning, More People Learning*, UC Berkeley, March 15, 2013.

Invited talk, CITRIS Data Innovation Day (Streams, Gardens and Clouds), UC Berkeley, *The Middle Distance, Digital Humanities, and Visualization*, January 24, 2013.

Invited talk, Leonardo Art/Science Evening Rendevous (LASERs), 'Natural' Search User Interfaces, University of San Francisco, November 12, 2012.

Invited talk, Ecolunch, UC Berkeley, *The Biotext Search Project*, October 22, 2012.

Invited talk, Microsoft Research/Bing, *HCIR Keynote Reprise*, October 19, 2012.

Invited talk, Introspective Inc, *Designing Search for Humans*, April 2012.

Invited Talk, ACM SF Bay Professional Chapter Meeting, *Natural Search User Interfaces*, March 2012.

Invited Talk, University of Pittsburgh iSchool Colloquium Series, *Emerging Trends in Search User Interfaces*, March 2012.

Invited speaker, *Searching for the Future of Big Data*, O'Reilly Strata Big Data Conference, Feb 2012.

Invited Talk, UC Berkeley Center for Technology Research in the Interest of Society, Feb 2012, *The Untold Story of How Billions in Government Savings are Likely to Result from Current Administration IT Reforms*

Future Trends in Search User Interfaces, University of Maryland, Baltimore Campus, November 2011.

Future Trends in Search User Interfaces, University of Maryland, College Park, November 2011.

Future Trends in Search User Interfaces, University of Konstanz, June 2011.

Social Technology, invited talk, Johns Hopkins University, February 2010.

IT Research to Support Open Government, invited talk, University of Maryland, February 2010.

Faceted Navigation and Search, invited talk, MITRE, December 2009.

Search User Interfaces, invited speaker, Google Tech Talk, Google, November 2009.

Social Technology, invited talk, George Washington University, November 2009.

Best Practices for Search, invited lecture, U.S. Federal Web Managers University, November 2009.

Ideas for USA.gov, USA.gov project team, October, 2009.

PubMed Should Search Full Text and Show Figures in Search Results, National Library of Medicine, Bethesda, MD, June 2009.

Social Technology, National Science Foundation, CISE, March 2009.

Semi-Automated Creation of Hierarchical Faceted Metadata, Lawrence Livermore National Laboratory, February, 2009.

Tricks for Statistical Semantic Knowledge Discovery: A Selectionally Restricted Sample, NSF Symposium on Semantic Knowledge Discovery, Organization and Use, NYU, November 2008.

Unambiguous + Unlimited = Unsupervised, Google Seattle, September 2008.

Faceted Navigation, Earley & Associates Webinar, July 2008.

Some Thoughts on Social Tagging, DUB seminar, University of Washington, July 2008.

Unambiguous + Unlimited = Unsupervised, Computer Science Colloquium, Brigham Young University, December 2007.

Some Thoughts on Social Tagging, eBay, August 2007.

Some Thoughts on Social Tagging, Yahoo Berkeley Research Center, June 2007.

Some Thoughts on Social Tagging, MIT, HCI Seminar, April 2007.

Some Thoughts on Social Tagging, IBM USER Group, April 2007.

Unambiguous + Unlimited = Unsupervised, International Computer Science Institute, Berkeley, CA, April 2007.

Faceted Metadata in Search User Interfaces, Computer Science Colloquium, University of Illinois, Champaign, February 2007.

New Search Tools for Full Text BioScience Journal Articles, Bioinformatics Seminar, University of Illinois, Champaign, February 2007.

Semi-automated Creation of Facet Hierarchies, Invited Speaker, Taxonomy Bootcamp, San Jose, November 2006.

Unambiguous + Unlimited = Unsupervised, Neyman Seminar, Statistics Department, UC Berkeley, October 2006.

How to Group Search Results: Conclusions from Usability Studies, Google, September, 2006.

Unambiguous + Unlimited = Unsupervised, PARC, Palo Alto, CA, August 2006.

Metadata in Search Interfaces, invited talk, University of Buffalo, January 2006.

Semantic Relation Detection in BioScience Text, invited talk, Information Sciences Institute/University of Southern California, November 2004

Semantic Relation Detection in BioScience Text, invited talk, San Diego State University, November 2004

Semantic Relation Detection in BioScience Text, invited talk, Stanford University, November 2004

Hierarchical Faceted Metadata in Browsing and Search interfaces, invited talk, California Digital Library offices, November 2004.

Semantic Relation Detection in BioScience Text, invited talk, Johns Hopkins University, October 2004.

The State of the Art in Bioscience Text Mining, invited talk, Genentech CIT Offsite, March 2004.

Using Words to Search a Thousand Images: Hierarchical Faceted Metadata in Search Interfaces, invited talk, Boeing, September 2003.

Using Words to Search a Thousand Images: Hierarchical Faceted Metadata in Search Interfaces, invited talk, University of Washington, September 2003.

Hierarchical Faceted Metadata in Search, invited talk, Dublin Core 2003 Metadata and Search Workshop, September 2003.

Using Words to Search a Thousand Images, Microsoft Research, Search Engine Summer Institute, July 2003.

Using Words to Search a Thousand Images, Google Inc, June 2003.

Using Words to Search a Thousand Images: Hierarchical Faceted Metadata in Search Interfaces, Spotlight Speaker, User Interface 7 West conference, South San Francisco, March 2003.

Using Words to Search a Thousand Images, Stanford Seminar on People, Computers, and Design, Stanford University, January 2003.

Using Words to Search a Thousand Images, University of Arizona, November 2002.

Automating Web Design Evaluation, invited talk, IBM Make IT-Easy conference, San Jose, CA, June 2002.

Incorporating Metadata into Site Search, invited talk, Infonortics Search Engines Conference, San Francisco, CA, March 2002.

Automating Web Design Evaluation, invited talk, BayCHI-Web, Redwood City, CA, November, 2001.

When Information Technology 'Goes Social', SFBay Regional Chapter of the Special Libraries Association, San Francisco, CA, October 2001.

Incorporating Metadata into Search, University of Maryland, College Park, September 2001.

Incorporating Metadata into Search, Ariba, Sunnyvale CA, July 2001.

Metadata: Promise and Pitfalls, invited talk to Committee on Internet Navigation and the Domain Name System of the Computer Science and Telecommunications Board, National Research Council, July 2001.

A Statistical Approach for Classifying the Semantic Relations in Noun Compounds using MeSH, Stanford University, School of Medical Informatics, May 2001.

Incorporating Metadata into Search, Yahoo Corp., San Jose, CA, May 2001.

Incorporating Metadata into Search, Argus Inc., March 2001.

Automating Web Site Usability, University of Michigan, School of Information, March 2001.

Criteria and Metrics for Evaluating Architectures, invited talk, ASIS&T Information Architecture Summit, February 2001.

Incorporating Metadata into Search, Razorfish Inc, San Francisco, CA, November 2000.

Automating Web Site Usability, IBM Almaden Research Laboratories, San Jose, CA, October 2000.

Text Data Mining, Agyinc Corp., South San Francisco, August 2000.

Automating Web Site Usability, invited talk, NEC Research Institute Annual Symposium, Stanford University, June 2000.

Text Data Mining: Tools for Scientific Discovery, Central Intelligence Agency, Reston, VA, June 2000.

Text Data Mining: Tools for Scientific Discovery, invited workshop, plenary talk, Institute for Mathematical Sciences, University of Minnesota, April, 2000.

Text Data Mining, invitational workshop participation, Institute for Mathematics and Its Applications Workshop on Text Mining, Minneapolis MI, April 2000.

Incorporating Metadata into Search, Inktomi Corp., San Mateo, CA, April 2000.

Cha-Cha: A System for Organizing Intranet Search Results, Stanford University Digital Libraries Seminar, June 1999.

Untangling Text Data Mining, the Southern California chapter of Academic Business Librarians Exchange (ABLE), USC, April 1999.

When Information Technology "Goes Social,", Human-Centered Computing Seminar, UC Berkeley, October 1998.

Text Data Mining: Issues, Techniques, and the Relationship to Information Access, Stanford University Digital Libraries Seminar, April 1998.

Integrating Browsing and Search using a Large Category Hierarchy, Microsoft Research Laboratories, April 1998.

Text Data Mining: Issues, Techniques, and the Relationship to Information Access, Bay Area Information Systems Research Symposium, Fisher Center for Management & Information Technology, November, 1997.

An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy, IBM Almaden Research Laboratories, September 1997.

Text Data Mining: Issues, Techniques, and the Relationship to Information Access, University of Washington/Microsoft Research Invitational Workshop on Data Mining, July 1997.

Categories and Clusters in Information Access Interfaces, University of Chicago, May 1997.

Categories and Clusters in Information Access Interfaces, Rutgers University, April 1997.

Categories and Clusters in Information Access Interfaces, Stanford University, April 1997.

Categories and Clusters in Information Access Interfaces, MIT, April 1997.

TileBars, IEEE Symposium on Information Visualization, Burlingame, October 1996.

Multi-Paragraph Text Segmentation, SRI, October 1996.

What Happens After the Search?, Apple Computer, August 1996.

The TREC-4 Interactive Track, Digital Libraries Seminar, UC Berkeley Computer Science Department, March 1996. Repeated at UC Berkeley School of Information Management & Systems, April 1996.

Revealing Document Collection Structure Through Intelligent Information Access Interfaces, Section on Medical Informatics, Stanford University School of Medicine, December 1995.

Revealing Document Collection Structure Through Intelligent Information Access Interfaces, Artificial Intelligence Seminar, Carnegie Mellon University, November 1995.

Revealing Document Collection Structure Through Intelligent Information Access Interfaces, Artificial Intelligence Seminar, MITRE Corp., November 1995.

Adaptive Sentence Boundary Disambiguation, Johns Hopkins University Speech Research Symposium, June 1995.

Context and Structure in Full-Text Information Access, Terry Winograd's PCD Seminar, Stanford University, January 1995.

Context and Structure in Full-Text Information Access, Harvard University, December 1994.

New Ideas in Full-Text Information Access, Georgia Institute of Technology, February 1994.

New Ideas in Full-Text Information Access, University of Central Florida, February 1994.

New Ideas in Full-Text Information Access, University of Florida, February 1994.

New Ideas on Full-Text Access, Sequoia Seminar, University of California, Berkeley, February 1993.

Lexical Connectivity in Large-Grain Discourse Segmentation, presentation at the Meeting of the Society for Text and Discourse; Boulder, CO, June 1993.

Subtopic Structuring for Full-Length Document Access, Cornell University, May 1993.

Appearances in the Media (Selected)

"Allen Institute for AI Eyes the Future of Scientific Search" by Cade Metz, Wired Magazine, Nov 11, 2016.

"The Bot Versus the Bard: Researchers Teach a Computer to Write Poetry", by Katia Savchuk, California Magazine, Aug 30, 2016

“What’s missing from media claims linking Google Trends to Brexit vote,” by Allison Graves, *politifact*, July 1, 2016.

“The Science of Visualization,” by Jen Christiansen, *Scientific American*, Nov 3, 2015. <https://blogs.scientificamerican.com/sa-visual/the-science-of-visualization/>

“Celebrating Ada Lovelace Day: The O’Reilly community shares stories of inspiring women in tech”, Oct 14, 2014. <http://radar.oreilly.com/2014/10/celebrating-ada-lovelace-day.html>

“Five Questions: Marti Hearst, ‘Big Data’ pioneer,” Christopher Yee, *Daily Cal*, Nov 13, 2012, <http://www.dailycal.org/2012/11/13/five-questions-marti-hearst-big-data-pioneer/>.

“Keen On Big Data, Why UC Berkeley Might Have an Edge Over Stanford” *Tech Crunch TV*, May 30, 2012. <https://techcrunch.com/2012/05/30/keen-on-big-data-why-uc-berkeley-might-have-an-edge-over-stanford-tctv/>

2011 HCIR Workshop at Google, *New York Times beta620*, Brad Stenger, Nov 3, 2011, <http://open.blogs.nytimes.com/2011/11/03/2011-hcir-workshop-at-google/>

A New Perspective on Search Results, by Kate Greene, *MIT Technology Review*, February 6, 2008, <http://www.technologyreview.com/news/409495/a-new-perspective-on-search-results/>

Ballot Roulette, by Peter Weiss, in *Science News* 170(19), November 4, 2006.

In Search of a Better Search, in *DevSource.com’s Great Minds in Development* series, interview by Mike Elgan, Ziff Davis, June 26, 2006.

Science Puts Enron E-Mail to Use, by Ryan Singel, in *Wired News*, January 30, 2006. <http://www.wired.com/science/discoveries/news/2006/01/70100?currentPage=2>

Death to Folders! in *The Economist*, September 15th, 2005, <http://www.economist.com/node/4368267>

Revenge of the Nerds – Again, by Ben Elgin, in *BusinessWeek*, August 8, 2005. http://www.businessweek.com/magazine/content/05_32/b3946043_mz011.htm

Making a Web Search Feel Like a Stroll in the Library, by Anne Eisenberg, *The New York Times*, August 19, 2004, <http://www.nytimes.com/2004/08/19/technology/what-s-next-making-a-web-search-feel-like-a-stroll-in-the-library.html>

Next-generation search tools to refine results, by Michael Kanellos, *CNet*, August 9, 2004 (mentions Flamenoco), http://news.cnet.com/2100-1025_3-5299239.html

Eternal Search For Wisdom Generates Laws Unto Themselves, *The Courier Mail*, Queensland, Australia, By Michael Duffy, Jan 31, 2004. <http://www.edge.org/documents/press/courier-mail.1.31.04.html>

Digging For Nuggets Of Wisdom, *New York Times*, by Lisa Guernsey, October 16, 2003, <http://www.nytimes.com/2003/10/16/technology/digging-for-nuggets-of-wisdom.html>

Text Mining, *Future Tense*, on Minnesota Public Radio, interviewed by Jon Gordon, October 20, 2003.

The Search Engine that Could, by Spencer Michels, *Lehrer News Hour*, PBS, November 29, 2002 http://www.pbs.org/newshour/bb/business/july-dec02/google_11-29.html

Data Mining, KALX Radio Interview (Berkeley Groks Science Topics), February 2001.

Incredibly Strange Databases, by Annalee Newitz, in the *San Francisco Chronicle*, November 27, 2001, <http://www.sfgate.com/news/article/Incredibly-Strange-Databases-Unusual-2849319.php>

Back to the Drawing Board, *Interactive Week*, by Rory O’Connor February 26, 2001.

Assembling the World’s Biggest Library on Your Desktop, *Science*, 281 (5384), pp. 1784 - 1786, by Joseph Alper, September 18, 1998

<http://www.sciencemag.org/content/281/5384/1784.full>

Mining in Textual Mountains, *Mappa Mundi Magazine*, by Marty Lucas, November, 1999.

Reviewing

Journal Reviewing

2002-2016 ACM Transactions on Computer-Human Interaction (TOCHI) (Associate Editor)

2015, 2007, 2005, 2004, 2002 Communications of the ACM

2015 JCRR

2014 ACM TWEB

2014 ACM TIST

2014, 2013 ACM TOIS

2013 Proceedings of the National Academy of Sciences

2013 Information Systems Journal

2012 Human Technology Journal

2012 Foundations and Trends in Information Retrieval

2011 ACM Computing Surveys

2011, 2008, 1997, 1996 Computational Linguistics (Reviewer)

2006-2012 ACM Transactions on the Web TWEB (Associate Editor)

1997-2005 ACM Transactions on Information Systems (TOIS) (Associate Editor)

1999-2002 Computational Linguistics (Associate Editor)

2001 International Journal of Human Computer Studies (IJHCS)

2001 IEEE Internet

1996 Machine Learning

1995 International Journal on Artificial Intelligence Tools

Conference Reviewing

(Asterisk indicates member of the main program committee.)

2017*, 2016*, 2015*, 2014* ACM Learning@Scale Conference

2016* CHIIR

2016, 2015, 2014, 2013 KDD IDEA Workshop

2017, 2016*, 2014, 2013*, 2012, 2011, 2010, 2008, 2007*, 2006*, 2005, 2004, 2002, 2001, 2000*, 1999*, 1998, 1997, 1996 ACM SIGCHI (Human Factors in Computing Systems)

2016 EMNLP workshop on NLP and Computational Social Sciences

2016, 2013, 2012, 2010, 2008, 2007, 2006, 2004, 2002, 2001*, 2000, 1999*, 1998, 1997, 1996 ACM UIST (User Interfaces Software and Technology)

2014* ACL Interactive Visualization of Language Workshop

2016 IEEE Infoviz Workshop on Pedagogy of Data Visualization (co-organizer)

2014, 2011, 2010, 2008, 2006, 2005, 2004, 2001, 1999, 1998*, 1997*, 1996*, 1995* IEEE Infovis (Symposium on Information Visualization)

2016 CSCW

2012, 2008, 2007, 2006, 2005, 2004, 2002, 2001*, 1999*, 1998*, 1997* ACM SIGIR (Information Retrieval)

2011, 2008, 2007, 2005*, 2004, 2002, 2001*, 2000, 1999 ACL/NAACL (Association for Computational Linguistics)

2013 IUI

2012, 2009, 2007, 2006, 2003*, 2001* HLT-NAACL/ HLT (Human Language Technologies)

2007, 2004, 2002*, 2001*, 1999, 1997, 1996, 1995 ACL EMNLP / VLC (Empirical Methods in NLP)

2006, 2002, 2000, 1999, 1998, 1997 AAAI/IJCAI (American Association for Artificial Intelligence)

2012, 2011, 2009, 2006*, 2005, 2004, 2002, 1999 World Wide Web Conference

2012, 2011, 2007 VAST

2012, 2008 WSDM (Web Search and Data Mining)

2008, 2006 CoNLL

2010, 2008, 2006 EACL/LREC/SME workshop

2008, 2007, 2006, 2005 ACL/ISMB SIG Workshop / BioNLP

2006 CIKM

2008, 2007, 2006 PSB (Pacific Symposium on Biocomputing)

1996, 1995 SDAIR (Symposium on Document Analysis and Information Retrieval)

2002 SigGraph

2000* ACM E-Commerce

Morgan Kaufman Publishers

MIT Press

Cambridge University Press

Research Grant Reviewing

2011 National Endowment for the Humanities (NEH)

2007, 2003 National Institutes of Health (NIH)

2016, 2015, 2012, 2007, 2006, 2005, 2004, 2003, 2002, 2000, 1999, 1998, 1996, 1995, 1994 National Science Foundation (NSF)

1996 Research Council of Canada

Patents

“Interactive Interface for Specifying Searches,” Hearst, M., April 24, 2001, USP 6,223,145. (Patent owned by Xerox.)

“Interactive interface for viewing retrieval results,” Hearst, M. and Karadi, C., October 2, 2001, USP 6,297,824.

(Patent owned by Xerox.)

“Method for Learning to Infer the Topical Content of Documents Based upon their Lexical Content,” Saund, E. and Hearst, M., November 11, 1997, USP 5,687,364. (Patent owned by Xerox.)

“Method and apparatus for inferring the topical content of a document based upon its lexical content without supervision,” Saund, E. and Hearst, M., August 19, 1997, USP 5,659,766. (Patent owned by Xerox.)