

Brian D. Davison, Ph.D.

Associate Professor

Department of Computer Science and Engineering, Lehigh University
113 Research Drive, Bethlehem, PA 18015 USA

davison@cse.lehigh.edu — <http://www.cse.lehigh.edu/~brian/>

Office: 610-758-3453

Areas of Interest Information Search, Ranking, Retrieval and Prediction; Data Science; World Wide Web; Text Mining; Online Social Networks; Adversarial Information Retrieval

Professional Experience DEPT. OF COMPUTER SCIENCE & ENGINEERING, LEHIGH UNIVERSITY Bethlehem, PA
Associate Professor (2009 – present), Assistant Professor (2002 – 2009), Instructor (2001 – 2002)

FACEBOOK Menlo Park, CA
Visiting Professor, Core Data Science group, while on sabbatical from Lehigh (2013-2014)

Education RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY New Brunswick, NJ
Ph.D., Computer Science, October 2002
M.S., Computer Science, May 1995

BUCKNELL UNIVERSITY Lewisburg, PA
Major: Computer Engineering
Bachelor of Science in Engineering, cum laude with dept. honors, May 1991

Current Research Projects DATASET DISCOVERY, ACCESS, AND RETRIEVAL FOR CROSS-DISCIPLINARY RESEARCH
2017 – present *PI: Brian D. Davison; Co-PIs: J. Heflin and H. Jia*

With society's increasing ability to sense, capture and store data, along with open data initiatives and increased funding agency requirements for sharing data, more data are now publicly available. Researchers, professionals and decision-makers are working with data in innovative ways, not only at unprecedented scales, but more importantly across different fields and domains. It is difficult, however, for domain non-experts such as data journalists to know about all the datasets available in a given field, and impossible even for experts to know about datasets in other fields, which is becoming essential in interdisciplinary work. We have combined our strengths in data science, information retrieval, federating heterogeneous data sources, management of structured and semi-structured data, and user studies to create novel approaches to dataset search. Existing efforts are domain-specific, or are shallow, searching only the descriptions of the data, and so there are significant research opportunities to answer questions on how to generalize and scale such efforts. In this project, we learn from dataset searchers, propose novel interfaces, and create a useful prototype that indexes descriptions and content of datasets. This work was initially funded by an internal Lehigh seed research grant, and is now funded by NSF award IIS-1816325.

PROBABILISTIC RESILIENCE ASSESSMENT OF INTERDEPENDENT SYSTEMS (PRAISYS)
2015 – present *PI: P. Bocchini, Co-PIs include B.D. Davison*

After a disruptive extreme event, such as an earthquake or severe storm, the socio-economic recovery of the affected region depends on the recovery of its infrastructure systems. Lifelines, such as power and water distribution systems, transportation networks, communication systems, and critical buildings have a primary role in disaster response, management, and long-term recovery. The failure to rapidly restore the services required for personal, social, and commercial activities leads

to continued socio-economic losses and progressive depopulation. This collaborative project brings together scholars in Civil Engineering, Systems Engineering, Computer Science, Economics, Urban Planning, and Policy Making. Its purpose is to establish and demonstrate a comprehensive framework that combines models of individual infrastructure systems with models of their interdependencies for the assessment of interdependent infrastructure system resilience for extreme events under uncertainty. The “PRAISys” platform (Probabilistic Resilience Assessment of Interdependent Systems) will emphasize a probabilistic approach that permeates all aspects of the models, including the interdependencies. Some types of uncertainties that were not considered before (e.g., the possibility of using contingency plans that provide services without functioning infrastructure) will be classified; while mathematical and computational tools will be devised to capture their characteristics. PRAISys will enable better management and design of next generation infrastructure, more resilient to extreme events and to component failures under normal conditions. This will reduce the likelihood of extreme events becoming catastrophic in terms of casualties and injuries, long-lasting socio-economic losses, and environmental impact. This work has been funded by NSF award ENG-1541177.

SIGNALS FOR IP GEOLOCATION

2014 – present

IP geolocation databases map IP addresses to their geographical locations. These databases are important for several applications such as local search engine relevance, credit card fraud protection, geotargetted advertising, and online content delivery. While they are the most popular method of geolocation, they can have poor accuracy at the city level. In this work we evaluate and improve IP geolocation databases using a variety of signals, such as those collected from search engine logs and reverse DNS hostnames. We generate a large ground-truth dataset using real time global positioning data extracted from search engine logs. We show that incorrect geolocation information can have a negative impact on implicit user metrics. We investigate techniques to improve upon geolocation databases, and validate our approaches with large scale ground truth data.

**Selected
Past
Research
Experience**

DETECTING BIASES IN THE MEDIA

2013 – 2016

PI: Brian D. Davison

The news media have considerable power to educate and thus to influence opinions in media consumers. News stories typically contain more information than a simple list of facts; instead, they are told through the lens of the author or news outlet publishing the story. Often, news articles contain biases—intentional or unintentional inclinations towards a certain viewpoint—injected into the article by the author or that author’s news organization. Such biases, such as sensationalism, negativity, politics and religion, may have a significant effect on the way a news story impacts the reader. We wish to recognize viewpoints presented in stories and make any underlying biases clear. This project collected evidence to test for the existence of various biases through human judgments on collections of articles written about the same story. Ultimately, bias detection techniques could enable media customization and recommendations to match or complement reader-preferred biases in online news, and inform readers (and writers!) of the inherent biases discovered. This work was funded in part through an internal Lehigh research grant.

ARCHIVAL SEARCH

2008 – 2012

PI: Brian D. Davison; Co-PI: T. Suel (NYU)

Historical collections of web content are of tremendous value in preserving records of significant aspects of modern society. This research project investigated efficient and effective approaches to store, index, mine and retrieve web content from large-scale historical archives. This work was collaborative with Torsten Suel (NYU Polytechnic) and the Internet Archive and was funded by NSF award IIS-0803605.

CONTEXTUAL LINK ANALYSIS
2005 – 2012

PI: Brian D. Davison

A web link or scholarly citation is never created in a vacuum. Yet traditional link analysis (such as PageRank) calculates estimates of authority regardless of the context of the link. As a result, a page that is highly ranked for one topic may also be highly ranked for another topic in which it is not authoritative. In this work we incorporate document context (such as topicality) to “color” the links and thus calculate the context-specific reputation for a document. This work has been funded by an NSF CAREER award IIS-0545875, and by an equipment grant from Sun Microsystems.

SEARCH ENGINE SPAM IDENTIFICATION AND ELIMINATION
Summer 2004 – 2011

PI: Brian D. Davison

In this effort we worked to identify and/or reduce the effects of various kinds of search engine spam. By analyzing the link graph of the Web, we can identify particular spamming techniques and eliminate their effect on link analysis algorithms. We also examined the prevalence of other kinds of spamming techniques. This work has been funded in part by Microsoft Live Labs.

UNDERSTANDING AND ENHANCING QUERIES
Summer 2002 – Summer 2008

PI: Brian D. Davison

Studied the relationships between queries and their search engine results in order to better understand aggregate query behavior. We are able to recommend queries for searchers and for content providers, and to find related Web pages. This work was funded by the National Science Foundation under grant IIS-0328825.

EDIFY: DISRUPTION AND DELAY TOLERANT NETWORKING
Spring 2005 – Fall 2006

PI: M.C. Chuah; Co-PIs: B.D. Davison and L. Cheng

In this DARPA-funded project, we designed a store-and-forward messaging architecture that is tolerant of delays, disruptions, and supports node mobility over an existing network infrastructure.

WEB PREFETCHING AND EVALUATION, RUTGERS UNIV. & LEHIGH UNIV.
Fall 2000 – Spring 2004

Rutgers PI: Haym Hirsh; Lehigh PI: Brian D. Davison

Investigated potential improvements to web caching performance through intelligent action prediction for application to interactive prefetching. Research issues include the design and implementation of a caching and network effects simulator, and the proposal of a new method for the evaluation of prefetching proxy caches, integrating multiple sources of information (historical, textual, and community-based), and HTTP log analysis. This research was funded under NSF grant ANI-9903052 to Rutgers, and via a subcontract to Lehigh University.

DISCOWEB SEARCH ENGINE PROJECT, RUTGERS UNIV.
Spring 1999 – Spring 2000

PI: Apostolos Gerasoulis

Worked to develop a system for web analysis that discovers *communities* of web sites with related views on search query topics. Research issues include the exploitation of web page interconnections for capturing flow of influence, text analysis for clustering and page ranking, and search engine technologies for retrieving pages and building graphs of page relationships. This project was primarily funded through DARPA research grants. It was later spun out as an Internet startup called Teoma, and purchased in 2001 by Ask Jeeves.

**Research
Grants
Awarded**

“Domain-Agnostic Dataset Search”, National Science Foundation, \$515,770, 2018-2020. B. Davison, principal investigator. Co-PI: J. Heflin and H. Jia.

“Intelligent and Scalable Systems” REU Site, National Science Foundation, \$359,959, 2018-2021. B. Davison, principal investigator. Co-PI: M. Spear. With E. Baumer, Y. Cao, L. Cheng, M. C. Chuah, J. Heflin, H. Muñoz-Avila, R. Palmieri, J. Spletzer, T. Wang, and S. Xie.

“Cross-browser Fingerprinting: Attacks, Dynamics, and Detection”, Amazon, \$80,000 (plus \$20,000 in AWS credits), 2018. PI: Y. Cao. B. Davison, co-PI.

“Dataset Discovery, Access, and Retrieval for Cross-Disciplinary Research”, Internal Lehigh CORE Funding, \$58,771, 2017-2018. B. Davison, principal investigator. Co-PI: J. Heflin and H. Jia.

“IoT sensor testing for sleep analytics”, Internal Lehigh Faculty Research Grant, \$6,000, 2016-2017. B. Davison, principal investigator. With L. Napper and P. Costa.

“Collaborative Research: CRISP Type 2: Probabilistic Resilience Assessment of Interdependent Systems - PRAISys”, National Science Foundation, \$1,903,209, 2015-2018. PI: Paolo Bocchini. With A. Lamadrid, R. Sause, and L. Snyder. Collaborative with D. Mitsova-Boneva (Florida Atlantic Univ.), A. Sapat (Florida Atlantic Univ.), and A.-M. Esnard (Georgia State Univ.).

“The Lehigh Smart Spaces Project”, REU Site, National Science Foundation, \$340,000, 2014-2016. PIs: J. Spletzer and M. C. Chuah. With B. Chen, L. Cheng, J. Heflin, X. Huang, D. Lopresti, H. Munoz-Avila, M. Spear, and G. Tan.

“Detecting Biases in Online News Media”, Internal Lehigh FIG Funding, \$23,371, 2013-2014, B. Davison, principal investigator.

“III:RI:Small:RUI:Tracking Predators and Bullies via Chat Log Transcripts”, National Science Foundation, \$498,428, 2009-2012, A. Kontostathis (PI, Ursinus) and L. Edwards (Co-PI, Ursinus). Subcontract to Lehigh: \$181,597, B. Davison, principal investigator.

“Enhancing Productivity at Air Products Using Information and Systems Technology”, Pennsylvania Infrastructure Technology Alliance (PITA), \$36,850, plus funding from Air Products of \$55,000, 2009, R. Nagel, principal investigator, S. Goldman, B. Davison and E. Zimmers, co-principal investigators.

“Efficient and Effective Search Services Over Archival Webs”, National Science Foundation, \$900,000, 2008-2011, B. Davison, principal investigator, T. Suel (NYU Polytechnic), co-principal investigator. In collaboration with the Internet Archive.

“Social Networks and Virtual Reality as Environments for Business Development and Globalization”, Pennsylvania Infrastructure Technology Alliance (PITA), \$56,500, plus industrial and non-profit sponsors, \$87,000, 2008, R. Nagel, principal investigator, B. Davison and E. Zimmers, co-PIs.

“CAREER: Contextual Link Analysis”, National Science Foundation, \$500,000, 2006-2011, B. Davison, principal investigator.

“Incorporating Trust into Web Authority”, Microsoft Research / Live Labs “Accelerating Search”, \$34,898, 2006-2007, B. Davison, principal investigator.

“Unsupervised Continuous Adaptation for Accented Speech Recognition”, Johns Hopkins (sub-contract of an NSF award), \$25,000, 2004-2005, B. Davison, principal investigator, for D. W. Kirsch, student researcher.

“Enhanced Disruption/Fault Tolerant Bundle Delivery System (EDIFY)”, DARPA, \$348,000 (Phase I, 2005-2006), \$456,634 (Phase II, 2006-2007), M. Chuah, principal investigator, B. Davison, and L. Cheng, co-principal investigators.

“Understanding and Enhancing Queries”, National Science Foundation, award IIS-0328825, \$255,000, 2004-2007, B. Davison, principal investigator.

“REU Supplement: Understanding and Enhancing Queries”, National Science Foundation, award IIS-0328825, \$6,000, 2005, B. Davison, principal investigator.

“Intelligent Web Prefetching to Reduce Client Latencies”, National Science Foundation, award ANI-9903052, \$438,020, 2000-2004, H. Hirsh, principal investigator. Subcontract to Lehigh, \$297,241, B. Davison, principal investigator.

**Other
Grants**

“Computer Science and Engineering Undergraduate Sandbox Lab Expansion”, P.C. Rossin College of Engineering and Applied Science, \$51,683, June 2018. Y. Cao, B. Davison, B. Hodgson, J. Loew, D. Lopresti, R. Palmieri, M. Spear.

“Primary Undergraduate Labs Enhancements and Upgrades”, P.C. Rossin College of Engineering and Applied Science, \$32,097, January 2015, B. Chen, B. Davison, B. Hodgson, H. Korth, M. Spear and G. Tan.

“Computer Science & Engineering Undergraduate Sandbox Lab Upgrade”, P.C. Rossin College of Engineering and Applied Science, \$66,000, December 2010, B. Chen, B. Davison, B. Hodgson, H. Korth, D. Lopresti, M. Spear, and J. Spletzer.

“CSE Unix Lab Upgrade and Expansion”, P.C. Rossin College of Engineering and Applied Science, \$44,000, October 2008, H. Korth, B. Hodgson and B. Davison.

“Academic Excellence Grant”, Sun Microsystems, \$40,358, 2008 (equipment grant), B. Davison.

“Learning to Label a Graph: Application to Web Spam Detection”, EU Network of Excellence PASCAL, 12,000 Euros, 2007 (grant to provide competition support), R. Baeza-Yates, C. Castillo, B. Davison, and P. Gallinari.

“Undergraduate Laboratory Enhancement for Computer Science and Engineering”, P.C. Rossin College of Engineering and Applied Science, \$70,000, August 2006, H. Korth, H. Munoz-Avila, D. Hillman, B. Hodgson, B. Davison, L. Cheng, J. Spletzer, and M. Arnold.

“CS&E Undergraduate Labs Upgrade and Expansion”, P.C. Rossin College of Engineering and Applied Science, \$55,766, 2002, M. Arnold, L. Cheng, B. Davison, B. Hodgson, E. Kay, and J. Vollmer.

**Edited Books
and
Proceedings**

K. Collins-Thompson, Q. Mei, B. D. Davison, Y. Liu, and E. Yilmaz, eds. *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. July 2018, ACM Press. 978-1-4503-5657-2.

B. D. Davison, T. Suel, N. Craswell, and B. Liu, eds. *Proceedings of the Third ACM International Conference on Web Search and Data Mining*. February 2010, ACM Press. 978-1-60558-889-6.

C. Castillo, B. D. Davison, L. Denoyer, and P. Gallinari, eds. *Proceedings of the Graph Labelling Workshop and Web Spam Challenge*. September 2007, at ECML/PKDD in Warsaw, Poland.

C. Castillo, K. Chellapilla, and B. D. Davison, eds. *Proceedings of the Third International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*. ACM International Conference Proceedings Series; Vol. 215. May 2007, ACM Press. 978-1-59593-732-2.

B. D. Davison, M. Najork, and T. Converse (eds). *Proceedings of the Second International Workshop on Adversarial Information Retrieval on the Web*, August 2006, Technical Report LU-CSE-06-027, Department of Computer Science and Engineering, Lehigh University.

B. D. Davison (ed.). *Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web*, May 2005, Technical Report LU-CSE-05-030, Department of Computer Science and Engineering, Lehigh University.

F. Douglis and B. D. Davison (eds). *Web Content Caching and Distribution: Proceedings of the 8th International Workshop*. Kluwer, July, 2004. 1-4020-2257-3.

Journal Articles

W. Sun, P. Bocchini, and B. D. Davison. **Resilience metrics and measurement methods for transportation infrastructure: The state of the art.** *Sustainable and Resilient Infrastructure*, 2018. In press. DOI: 10.1080/23789689.2018.1448663

S. Kuang and B. D. Davison. **Learning word embeddings with Chi-square weights for health-care tweet classification.** *Applied Sciences*, Volume 7, Issue 8, Article 846, MDPI, August 2017.

O. Dan and B. D. Davison. **Measuring and Predicting Search Engine Users' Satisfaction.** *ACM Computing Surveys*, Volume 49, Issue 1, Article 18, July 2016.

Z. Yang, L. Hong, D. Yin and B. D. Davison. **Ranking Authors with Learning-to-Rank Topic Modeling.** *International Journal of Innovative Computing, Information and Control (IJICIC)*, Volume 11, Issue 4, pages 1295-1316, August 2015.

N. Dai and B. D. Davison. **Topic-Sensitive Search Engine Evaluation.** *Online Information Review*, Volume 35, Issue 6, pages 893-908, 2011.

C. Castillo and B. D. Davison. **Adversarial Web Search.** *Foundations and Trends in Information Retrieval*, Volume 4, Number 5, pages 377-486, 2010. Now Publishers.

Q. Ye, L. Cheng, M. C. Chuah, and B. D. Davison. **Performance Comparison of Different Multicast Routing Strategies in Disruption Tolerant Networks.** *Computer Communications*, Volume 32, Number 16, pages 1731-1741, October 2009.

X. Qi and B. D. Davison. **Web Page Classification: Features and Algorithms.** *ACM Computing Surveys*, Volume 41, Number 2, February 2009.

B. D. Davison, W. Zhang, and B. Wu. **Connecting P2P to the Web: Lessons from a Prototype Gnutella-WWW Gateway.** *Internet Research*, Volume 18, Number 3, pages 336-356, 2008. **Highly Commended Award Winner.**

B. D. Davison. **A Web Caching Primer.** *IEEE Internet Computing*, Volume 5, Number 4, July/August, pages 38-45, 2001.

B. D. Davison and V. Liberatore. **Pushing Politely: Improving Web Responsiveness One Packet at a Time.** *Performance Evaluation Review*, Volume 28, Number 2 (September), pages 43-49, 2000.

H. Hirsh, C. Basu, and B. D. Davison. **Learning to Personalize.** *Communications of the ACM*, Volume 43, Issue 8 (August), pages 102-106, 2000.

J. S. Hall, L. Steinberg, and B. D. Davison. **Combining Agoric and Genetic Methods in Stochastic Design.** *Nanotechnology*, Vol. 9, No. 3 (September), pages 274-284., 1998.

Encyclopedia Articles and Book Chapters

G. M. Weiss and B. D. Davison. **Data Mining.** In H. Bidgoli (ed.), *The Handbook of Technology Management*, Volume 2, Chapter 39, Wiley, January 2010.

S. L. Goldman, R. N. Nagel, B. D. Davison, and P. D. Schmid. **Next Generation Agility: Smart Business and Smart Communities.** In P.H.M. Vervest, D.W. van Liere, and L. Zheng (eds.), *The Network Experience: New Value from Smart Business Networks*, Chapter 4, pp. 49-55. Springer-Verlag, 2009.

B. Wu and B. D. Davison. **Web Spam Identification and Blocking.** In *The McGraw-Hill 2008 Yearbook of Science and Technology*, pp. 370-372, 2008.

**Refereed
Publications
in High
Impact
Conferences**

B. D. Davison. **World wide web search engines**. In *The McGraw-Hill Yearbook of Science and Technology*, pp. 347-349, 2006.

B. D. Davison. **Search engine**. In *The World Book Encyclopedia*, pp. 268-269, 2005.

A. Kontostathis, W. M. Pottenger, and B. D. Davison. **Identification of critical values in latent semantic indexing**. In T. Y. Lin, S. Ohsuga, C.J. Liao, X. T. Hu, and S. Tsumoto (eds.), *Foundations of Data Mining and Knowledge Discovery*, Springer-Verlag, 2005.

B. D. Davison. **Learning Web Request Patterns**. In M. Levene and A. Poulouvassilis, eds, *Web Dynamics: Adapting to Change in Content, Size, Topology and Use*, pp. 435-459, Springer-Verlag, 2004.

O. Dan, V. Parikh and B. D. Davison. **Improving IP Geolocation using Query Logs**. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 347-356, San Francisco, CA, February 2016.

Z. Yang, D. Yin and B. D. Davison. **Recommendation in Academia: a Joint Multi-Relational Model**. In *Proceedings of the IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM 2014)*, pp. 566-571, Beijing, China, August 2014.

D. Yin, B. Cao, J.-T. Sun and B. D. Davison. **Estimating Ad Group Performance in Sponsored Search**. In *Proceedings of the 7th International ACM Conference on Web Search and Data Mining (WSDM)*, pp. 143-152, New York City, February 2014.

D. Yin, S. Mei, B. Cao, J.-T. Sun and B. D. Davison. **Exploiting Contextual Factors for Click Modeling in Sponsored Search**. In *Proceedings of the 7th International ACM Conference on Web Search and Data Mining (WSDM)*, pp. 113-122, New York City, February 2014.

Z. Yang and B. D. Davison. **Academic Network Analysis: A Joint Topic Modeling Approach**. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 324-333, Niagara Falls, Canada, August 2013.

L. Hong, A. Doumith and B. D. Davison. **Co-Factorization Machines: Modeling User Interests and Predicting Individual Decisions in Twitter**. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 557-566, Rome, Italy, February 2013. **One of eight papers nominated for best paper award.**

D. Yin, S. Guo, B. Chidlovskii, B. D. Davison, C. Archambeau and G. Bouchard. **Connecting Comments and Tags: Improved Modeling of Social Tagging Systems**. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 547-556, Rome, Italy, February 2013.

L. Hong, R. Bekkerman, J. Adler, and B. D. Davison. **Learning to Rank Social Update Streams**. In *Proceedings of the 35th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 651-660, Portland, OR, August 2012.

L. Hong, D. Yin, J. Guo and B. D. Davison. **Tracking Trends: Incorporating Term Volume into Temporal Topic Models**. In *Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 484-492, San Diego, August 2011.

D. Yin, L. Hong, Z. Xue and B. D. Davison. **Temporal Dynamics of User Interests in Tagging Systems**. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, pp. 1279-1285, San Francisco, August 2011.

N. Dai, M. Shokouhi and B. D. Davison. **Learning to Rank for Freshness and Relevance**. In *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, pp. 95-104, Beijing, China, July 2011.

D. Yin, Z. Xue, L. Hong, and B. D. Davison. **A Probabilistic Model for Personalized Tag Prediction.** In *Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 959-968, Washington, DC, July 2010.

N. Dai and B. D. Davison. **Freshness Matters: In Flowers, Food, and Web Authority.** In *Proceedings of the 33rd Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 114-121, Geneva, Switzerland, July 2010.

L. Hong and B. D. Davison. **A Classification-based Approach to Question Answering in Discussion Boards.** In *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, pp. 171-178, Boston, July 2009.

X. Qi and B. D. Davison. **Classifiers Without Borders: Incorporating Fielded Text From Neighboring Web Pages.** In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, pp. 643-650, Singapore, July 2008.

L. Nie and B. D. Davison. **Separate and Inequal: Preserving Heterogeneity in Topical Authority Flows.** In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, pp. 443-450, Singapore, July 2008.

L. Nie, B. D. Davison, and B. Wu. **From Whence Does Your Authority Come? Utilizing Community Relevance in Ranking.** In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, pp. 1421-1426, Vancouver, Canada, July 2007.

X. Qi and B. D. Davison. **Knowing a Web Page by the Company It Keeps.** In *Proceedings of the Fifteenth ACM Conference on Information and Knowledge Management (CIKM)*, pp. 228-237, Arlington, VA, November 2006.

L. Nie, B. D. Davison, and X. Qi. **Topical Link Analysis for Web Search.** In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 91-98, Seattle, WA, August 2006.

B. Wu, V. Goel, and B. D. Davison. **Topical TrustRank: Using Topicality to Combat Web Spam.** In *Proceedings of the 15th International World Wide Web Conference*, pp. 63-72, Edinburgh, Scotland, May 2006.

B. Wu and B. D. Davison. **Detecting Semantic Cloaking on the Web.** In *Proceedings of the 15th International World Wide Web Conference*, pp. 819-828, Edinburgh, Scotland, May 2006.

B. Wu and B. D. Davison. **Identifying Link Farm Spam Pages.** In *Proceedings of the 14th International World Wide Web Conference*, pp. 820-829, Chiba, Japan, May 2005.

B. D. Davison. **Topical Locality in the Web.** In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, pp. 272-279, July, 2000.

S. A. Macskassy, A. Banerjee, B. D. Davison, and H. Hirsh. **Human Performance on Clustering Web Pages: A Preliminary Study.** In *Proceedings of The Fourth International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 264-268, New York City: AAAI Press, August, 1998.

L. Steinberg, J. S. Hall, and B. D. Davison. **Highest Utility First Search Across Multiple Levels of Stochastic Design.** In *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI)*, July 26-30, pp. 477-484, Madison, WI: AAAI Press, 1998.

Y. Yi, Z. Chen, J. Heflin, and B. D. Davison. **Recognizing Quantity Names for Tabular Data.** In *Joint Proceedings of the First International Workshop on Professional Search (ProfS2018); the*

Second Workshop on Knowledge Graphs and Semantics for Text Retrieval, Analysis, and Understanding (KG4IR); and the International Workshop on Data Search (DATA:SEARCH'18), pages 68-73. Presented at DATA:SEARCH'18: International Workshop on Searching Data on the Web, at SIGIR 2018, Ann Arbor, MI, July 2018.

W. Sun, P. Bocchini and B.D. Davison. **A Bottom-Up Simulation Model for Evaluating the Impact of Interdependencies on System Resilience**. Abstract presented at Engineering Mechanics Institute Conference 2018, Cambridge, MA, May 2018.

Z. Chen, H. Jia, J. Heflin, and B.D. Davison. **Generating Schema Labels through Dataset Content Analysis**. In *Companion Proceedings of the The Web Conference (WWW '18)*, pp. 1515-1522. Presented at the International Workshop on Profiling and Searching Data on the Web (Profiles & Data:Search'18), Lyon, France, April 2018. **Best paper award**.

S. Kuang and B. D. Davison. **Class-specific Word Embedding through Linear Compositionality**. Presented at the *International Workshop on Mining Actionable Insights from Social Networks (MAISoN'17)*, Cambridge, UK, February 2017.

S. Kuang and B. D. Davison. **Semantic and Context-aware Linguistic Model for Bias Detection**. In *Natural Language Processing meets Journalism IJCAI-16 Workshop*, pp. 57-62, New York, NY, July 2016.

Z. Yang and B. D. Davison. **Writing with Style: Venue Classification**. In *Proceedings of the 11th International Conference on Machine Learning and Applications (ICMLA)*, pp. 250-255, Boca Raton, FL, December 2012.

Z. Yang, B. D. Davison and N. Dai. **Venue Recommendation: Submitting your Paper with Style**. In *Proceedings of the 11th International Conference on Machine Learning and Applications (ICMLA)*, pp. 681-686, Boca Raton, FL, December 2012.

L. Hong, A. Doumith and B. D. Davison. **Personalized Retweet Prediction in Twitter**. Presented at the 4th Workshop on Information in Networks (WIN), NYU Stern School of Business, September 2012.

Z. Yang and B. D. Davison. **Distinguishing Venues by Writing Styles**. In *Proceedings of the 12th ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pp. 371-372, Washington, DC, June 2012.

D. Yin, L. Hong and B. D. Davison. **Structural Link Analysis and Prediction in Microblogs**. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM)*, pp. 1163-1168, Glasgow, October 2011.

X. Qi and B. D. Davison. **Hierarchy Evolution for Improved Classification**. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM)*, pp. 2193-2196, Glasgow, October 2011.

Z. Xue, D. Yin and B. D. Davison. **Normalizing Microtext**. In *Proceedings of the AAAI-11 Workshop on Analyzing Microtext*, pp. 74-79, San Francisco, CA, August 2011.

D. Yin, L. Hong, X. Xiong and B. D. Davison. **Link Formation Analysis in Microblogs**. In *Proceedings of the 34th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1235-1236, Beijing, China, July 2011.

Z. Yang and B. D. Davison. **Award Prediction with Temporal Citation Network Analysis**. In *Proceedings of the 34th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1203-1204, Beijing, China, July 2011.

N. Dai, M. Shokouhi and B. D. Davison. **Multi-Objective Optimization in Learning to Rank**. In *Proceedings of the 34th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1241-1242. Beijing, China, July 2011.

O. Dan, J. Feng and B. D. Davison. **A Bootstrapping Approach to Identifying Relevant Tweets for Social TV**. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pp. 462-465, Barcelona, Spain, July 2011.

N. Dai, X. Qi and B. D. Davison. **Bridging Link and Query Intent to Enhance Web Search**. In *Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia (HT)*, pp. 17-26, Eindhoven, The Netherlands, June 2011.

L. Hong, O. Dan and B. D. Davison. **Predicting Popular Messages in Twitter**. In *Companion Proceedings of the 20th International World Wide Web Conference*, pp. 57-58, Hyderabad, India, March 2011. **Best poster paper award**.

D. Yin, L. Hong and B. D. Davison. **Exploiting Session-like Behavior in Tag Prediction**. In *Companion Proceedings of the 20th International World Wide Web Conference*, pp. 167-168, Hyderabad, India, March 2011.

N. Dai, X. Qi and B. D. Davison. **Enhancing Web Search with Entity Intent**. In *Companion Proceedings of the 20th International World Wide Web Conference*, pp. 29-30, Hyderabad, India, March 2011.

O. Dan, J. Feng, and B. D. Davison. **Filtering microblogging messages for Social TV**. In *Companion Proceedings of the 20th International World Wide Web Conference*, pp. 197-200, Hyderabad, India, March 2011.

X. Qi, D. Yin, Z. Xue and B. D. Davison. **Choosing Your Own Adventure: Automatic Taxonomy Expansion to Permit Many Paths**. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management (CIKM)*, pages 1853-1856, Toronto, Canada, October 2010.

L. Hong and B. D. Davison. **Empirical Study of Topic Modeling in Twitter**. In *Proceedings of the Social Media Analytics (SOMA) Workshop at KDD 2010*, Washington, DC, July 2010.

N. Dai and B. D. Davison. **Capturing Page Freshness for Web Search**. In *Proceedings of the 33rd Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 871-872, Geneva, Switzerland, July 2010.

Z. Yang, L. Hong, and B. D. Davison. **Topic-driven Multi-type Citation Network Analysis**. In *Proceedings of the 9th International RIAO Conference: Adaptivity, Personalization and Fusion of Heterogeneous Information*, Paris, France, April 2010.

N. Dai and B. D. Davison. **Mining Anchor Text Trends for Retrieval**. In *Proceedings of the 32nd Annual European Conference on Information Retrieval (ECIR)*, LNCS Vol. 5993, pp. 127-139, Milton Keynes, UK, March 2010.

N. Dai, B. D. Davison, and Y. Wang. **Mining Neighbors' Topicality to Better Control Authority Flow**. In *Proceedings of the 32nd Annual European Conference on Information Retrieval (ECIR)*, LNCS Vol. 5993, pp. 653-657, Milton Keynes, UK, March 2010.

L. Hong and B. D. Davison. **Wanted: A Unified Model for Search in Social Media**. Presented in the *WSDM 2010 Workshop on Search in Social Media (SSM)*, New York City, February 2010.

N. Dai and B. D. Davison. **Vetting the Links of the Web**. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM)*, pp. 1745-1748, Hong Kong, November 2009.

J. Wang, L. Hong and B. D. Davison. **RSDC'09: Tag Recommendation Using Keywords and Association Rules.** In *Proceedings of the ECML/PKDD 2009 Discovery Challenge Workshop*, held at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Bled, Slovenia, September 2009.

L. Hong, Z. Yang and B. D. Davison. **Incorporating Participant Reputation in Community-driven Question Answering Systems.** In *Proceedings of the Symposium on Social Intelligence and Networking (SIN09, at SocialCom09)*, CSE, Volume 4, pp. 475-480, Vancouver, August 2009.

J. Wang and B. D. Davison. **Counting Ancestors to Estimate Authority.** In *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, pp. 658-659, Boston, July 2009.

N. Dai, B. D. Davison and X. Qi. **Looking into the Past to Better Classify Web Spam.** In *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*, pp. 1-8, Madrid, Spain, April 2009.

J. Wang and B. D. Davison. **Explorations in Tag Suggestion and Query Expansion.** In *Proceedings of the CIKM 2008 Workshop on Search in Social Media (SSM)*, pp. 43-50, ACM Press, October 2008.

Q. Ye, L. Cheng, M. C. Chuah, and B. D. Davison. **SHIM: A Scalable Hierarchical Inter-domain Multicast Approach for Disruption Tolerant Networks.** In *Proceedings of the International Wireless Communications and Mobile Computing Conference*, pp. 170-175, Honolulu: ACM Press, August 2007.

L. Nie, B. D. Davison, and B. Wu. **Ranking by Community Relevance.** In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval*, pp. 873-874, Amsterdam, July 2007.

L. Nie, B. Wu, and B. D. Davison. **Winnowing Wheat from the Chaff: Propagating Trust to Sift Spam from the Web.** In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval*, pp. 869-870, Amsterdam, July 2007.

X. Qi, L. Nie, and B. D. Davison. **Measuring Similarity to Detect Qualified Links.** In *Proceedings of the Third International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*, pp. 49-56, Banff, Canada, May 2007.

L. Nie, B. Wu, and B. D. Davison. **A Cautious Surfer for PageRank.** In *Proceedings of the 16th International World Wide Web Conference*, pp. 1119-1120, Banff, Canada, May 2007.

B. Wu, V. Goel, and B. D. Davison. **Propagating Trust and Distrust to Demote Web Spam.** In *Proceedings of Models of Trust for the Web*, held at WWW 2006, Edinburgh, Scotland, May 2006.

Q. Ye, L. Cheng, M. C. Chuah, and B. D. Davison. **OS-multicast: On-demand Situation-aware Multicasting in Disruption Tolerant Networks.** In *Proceedings of the IEEE 63rd Vehicular Technology Conference (VTC 2006-Spring)*, Vol. 1, pp. 96-100, Melbourne, Australia, May 2006.

M. C. Chuah, P. Yang, B. D. Davison, and L. Cheng. **Store-and-Forward Performance in a Disruption Tolerant Network.** In *Proceedings of the IEEE 63rd Vehicular Technology Conference (VTC 2006-Spring)*, Vol. 1, pp. 187-191, Melbourne, Australia, May 2006.

B. Wu and B. D. Davison. **Undue Influence: Eliminating the Impact of Link Plagiarism on Web Search Rankings.** In *Proceedings of the ACM Symposium on Applied Computing (SAC)*, pp. 1099-1104, Dijon, France, April 2006.

M. C. Chuah, B. D. Davison, and V. Goel. **Performance evaluation of mobility management scheme in DTN**. In *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 539–544, Las Vegas, NV, April 2006.

M. C. Chuah, L. Cheng, and B. D. Davison. **Enhanced Disruption and Fault Tolerant Network Architecture for Bundle Delivery (EDIFY)**. In *Proceedings of the IEEE Global Telecommunications Conference (GlobeCom)*, Vol. 2, pp. 807–812, St. Louis, MO, November 2005.

B. Wu and B. D. Davison. **Cloaking and Redirection: A Preliminary Study**. In *Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*, held at WWW 2005, Chiba, Japan, May 2005.

B. D. Davison and B. Wu. **Implementing a Web Proxy Evaluation Architecture**. In *Proceedings of the 30th International Conference for the Resource Management and Performance Evaluation of Enterprise Computing Systems (CMG2004)*, pp. 703–714, Las Vegas, December, 2004. **Best paper award**.

B. D. Davison. **The potential of the metasearch engine**. In *Proceedings of the American Society for Information Science & Technology Annual Meeting*, Vol. 41, pp. 393–402, Providence, RI, November, 2004.

H. S. Baird, D. Lopresti, B. D. Davison, and W. M. Pottenger. **Robust Document Image Understanding Technologies**. In *Proceedings of the First ACM Workshop on Hardcopy Document Processing*, pp. 9–14, Washington, DC, November, 2004.

A. Kontostathis, W. M. Pottenger, and B. D. Davison. **Assessing the impact of sparsification on LSI Performance**. In *Proceedings of the 2004 Grace Hopper Celebration of Women in Computing Conference*, Chicago, IL, October, 2004.

B. D. Davison, W. Zhang, B. Wu. **Lessons from a Gnutella-Web Gateway**. In *Alternate Track Papers and Posters Proceedings of the Thirteenth International World Wide Web Conference*, pp. 502–503, ACM Press, New York City, May, 2004.

B. D. Davison. **Unifying Text and Link Analysis**. In *Proceedings of the IJCAI-03 Workshop on Text-Mining and Link-Analysis (TextLink)*, Acapulco, August, 2003.

B. D. Davison. **Toward a Unification of Text and Link Analysis**. In *Proceedings of the 26th Annual International ACM SIGIR Conference*, Toronto, pp. 367–368, July, 2003.

B. D. Davison, D. G. Deschenes, and D. B. Lewanda. **Finding Relevant Website Queries**. In *Poster Proceedings of the Twelfth International World Wide Web Conference*, Budapest, Hungary, May, 2003.

B. D. Davison, C. Krishnan, and B. Wu. **When does a hit = a miss?** In *Proceedings of the Seventh International Workshop on Web Content Caching and Distribution (WCW-2002)*, pp. 91–102, Boulder, CO, August, 2002.

B. D. Davison. **Predicting Web Actions from HTML Content**. In *Proceedings of the Thirteenth ACM Conference on Hypertext and Hypermedia (HT'02)*, pp. 159–168, College Park, MD, June, 2002.

B. D. Davison. **HTTP Simulator Validation Using Real Measurements: A Case Study**. In *Proceedings of the Ninth International Symposium on Modeling, Analysis, and Simulation on Computer and Telecommunication Systems (MASCOTS)*, pp. 389–396, Cincinnati, August, 2001.

B. D. Davison. **Assertion: Prefetching with GET is Not Good**. In *Proceedings of the Sixth International Workshop on Web Caching and Content Distribution (WCW)*, pp. 203–215, Elsevier, Boston, June, 2001.

B. D. Davison. **Recognizing Nepotistic Links on the Web**. Presented at the AAAI-2000 Workshop on Artificial Intelligence in Web Search, Austin, TX, and published in *Artificial Intelligence for Web Search*, Technical Report WS-00-01, pp. 23-28, AAAI Press, July, 2000.

K. Rasheed and B. D. Davison. **Effect of Global Parallelism on the Behavior of a Steady State Genetic Algorithm for Design Optimization**. In *Proceedings of the Congress on Evolutionary Computation (CEC)*, Vol. 1, pp. 534-541, Washington, DC, July, 1999.

B. D. Davison, A. Gerasoulis, K. Kleisouris, Y. Lu, H. Seo, W. Wang, and B. Wu. **DiscoWeb: Applying Link Analysis to Web Search**. In *Poster Proceedings of the Eighth International World Wide Web Conference*, pp. 148-149, Toronto, May, 1999.

B. D. Davison. **Simultaneous Proxy Evaluation**. In *Proceedings of the Fourth International Web Caching Workshop (WCW99)*, pp. 170-178, San Diego, March, 1999.

B. D. Davison. **A Survey of Proxy Cache Evaluation Techniques**. In *Proceedings of the Fourth International Web Caching Workshop (WCW99)*, pp. 67-77, San Diego, March, 1999.

B. D. Davison. **Web Traffic Logs: An Imperfect Resource for Evaluation**. In *Proceedings of the Ninth Annual Conference of the Internet Society (INET'99)*, San Jose, June, 1999.

B. D. Davison and H. Hirsh. **Predicting Sequences of User Actions**. In *Predicting the Future: AI Approaches to Time Series Problems*, Technical Report WS-98-07, pp. 5-12, AAAI Press. Proceedings of the 1998 AAAI/ICML Workshop, July, 1998.

B. D. Davison and H. Hirsh. **Toward An Adaptive Command Line Interface**. In *Advances in Human Factors/Ergonomics Volume 21B - Design of Computing Systems: Social and Ergonomic Considerations*, pp. 505-508: Elsevier Science, August 1997.

J. S. Hall, L. Steinberg, and B. D. Davison. **Rational Control of Multi-level Stochastic Design**. *Proceedings of the IASTED International Conference on Modelling and Simulation*, Pittsburgh, PA, 1997.

H. Hirsh and B. D. Davison. **An Adaptive UNIX Command-Line Assistant**. In *Proceedings of the First International Conference on Autonomous Agents*, pp. 542-543, Marina del Ray, CA: ACM, January, 1997.

Other Articles

D. Yin, Z. Xue, X. Qi and B. D. Davison. **Diversifying Search Results with Popular Subtopics**. In *Proceedings of the Eighteenth Text REtrieval Conference (TREC 2009)*, Gaithersburg, MD: NIST, November 2009.

R. N. Nagel, S. L. Goldman, and B. D. Davison. **Migrating to Agility 2.0: Social Computing Enables Organizational Growth and Innovation**. White paper sponsored by the PA Infrastructure Technology Alliance, March 2009.

B. D. Davison, M. Lalmas and D. D. Lewis. **ACM SIGIR Annual Business Meeting 2008: Secretary's Notes**. *SIGIR Forum*, 42(2):3-8, December 2008.

C. Castillo, K. Chellapilla, and B. D. Davison. **Adversarial Information Retrieval on the Web (AIRWeb 2007)**. *SIGIR Forum*, 42(1):68-72, June 2008.

M. Najork and B. D. Davison. **Introduction to Special Section on Adversarial Issues in Web Search**. In *ACM Transactions on the Web*, 2(1), February, 2008.

B. D. Davison, M. Najork, and T. Converse. **Adversarial Information Retrieval on the Web (AIRWeb 2006)**. *SIGIR Forum*, 40(2), December 2006.

B. D. Davison. **Is Your Web Site Cache Friendly?** Invited article in *Web Techniques* magazine, Vol. 5, No. 3 (March), pp. 101-103, 2000.

Book Review Q. Wang and B. D. Davison. **Review: Web Caching and Replication.** *SIGMOD Record*, Volume 32, Number 4, December 2003, pp. 107-108.

Patent L. Steinberg, J. S. Hall, and B. D. Davison. **Method of producing optimized designs using computer systems and designs produced therefrom.** United States Patent 6,367,052. Granted 2 April 2002.

Invited Talks *Keynote from Academia: Revamping Student Education with Real-World Data.* OSISoft PI World Conference, Academic Symposium, April 2018.

Introduction to Time-Series Analysis with PI System and R. With M. Takac. OSISoft PI World Conference, Analytics Track, April 2018.

Probabilistic Restoration Function of Critical Infrastructure Systems After an Extreme Event. With W. Sun, P. Bocchini. MOPTA, Lehigh University, August 2017.

Useful Web Mining with R, Course for the International Summer School on Web Science and Technology (WebST 2016), Bilbao, Spain, July 18-22, 2016.

The Times of Our Lives, Keynote for the Workshop on Temporal, Social and Spatially-aware Information Access (TAIA 2015), Santiago, Chile, August 13, 2015.

The Potential for Research and Development in Adversarial Information Retrieval

- Universite de Paris-Sud XI, Orsay, France, November 30, 2010.
- AIRWeb 2009 Keynote, Madrid, Spain, April 21, 2009.

Managing and Mining Information in Microblogs, Georgetown University, Washington, DC, October 22, 2010.

Searching Archival Webs Efficiently and Effectively, International Internet Preservation Consortium, Ottawa, Canada, May 4, 2009.

Contextual Link Analysis for Web Search

- Microsoft Research, Cambridge, UK, July 2008.
- Polytechnic University, Brooklyn, NY, April 2007.
- Stern School of Business, New York University, New York, NY, March 2007.

Mining the Web for Fool's Gold

- Lafayette College, Easton, PA, February 2008
- Ursinus College, Collegeville, PA, September 2007
- High Performance Computing Day at Lehigh University, Bethlehem, PA, April 2007

Incorporating Trust into Web Authority, Microsoft Research Search Summit, Redmond, WA, July 2007

Search Engine Spam: Significance and Some Solutions

- Bucknell University, Lewisburg, PA, October 2006.
- Fordham University, New York, NY, October 2006.

Application-Layer Networking, Wireless Day, Lehigh University, December 2003.

Understanding and Enhancing Search Engine Queries, Department of Computer Science, Rutgers University, New Brunswick, NJ, October 2003.

Organizing information so that it can be found, 4th Annual ECCMA Membership Conference, Bethlehem, PA, October 2003.

The Design and Evaluation of Web Cache Pre-Loading Techniques, Spring 2001.

- Lehigh University, Bethlehem, PA
- George Mason University, Fairfax, VA
- Northwestern University, Evanston, IL
- Northeastern University, Boston, MA
- Pennsylvania State University, State College, PA
- Purdue University (CS & ECE), Lafayette, IN
- University of Pittsburgh (CS & DIST), Pittsburgh, PA
- Johns Hopkins University, Baltimore, MD

Web Caching: Balancing User Needs and Network Utilization, Lucent Bell Labs, Murray Hill, NJ, September 1999.

**Graduate
Students
Advised**

Current students:

- Zhiyu Chen (doctoral student)
- Ovidiu Dan (doctoral candidate)
- Sicong Kuang (doctoral candidate)
- Dan Luo (doctoral student)
- Dr. Wenjuan Sun (postdoctoral advisee)

Graduated from Lehigh University:

- William West (MS, 2014)
- Zaihan Yang (PhD, 2014)
- Dawei Yin (PhD, 2013)
- Liangjie Hong (PhD, 2013)
- Na Dai (PhD, 2013)
- Jennifer Bayzick (MS, 2013)
- Xiaoguang Qi (PhD, 2012)
- Zhenzhen Xue (MS, 2010)
- Jian Wang (MS, 2009)
- Lan Nie (PhD, 2008)
- YaoShuang Wang (MS, 2008)
- Shruti Bhandari (MS, 2008)
- Baoning Wu (PhD, 2007)
- Vinay Goel (MS, 2006)
- Wei Zhang (MS, 2004)
- Yonghui Mou (MS, 2004)

- Kiran K. Komaravolu (MS, 2004)
- Kalyan C. Boggavarapu (MS, 2004)
- David B. Lewanda (MS, 2003)
- David G. Deschenes (MS, 2003)

Instructional Experience

DEPT. OF COMPUTER SCIENCE & ENGINEERING, LEHIGH UNIV.
Served as Instructor for:

Bethlehem, PA

- Undergraduate course in data science (Fall '15, Spring '16, Fall '16, Spring '17, Fall '17, Spring '18, Fall '18)
- Undergraduate course on computers, internet and society (Fall '18)
- Undergraduate course in system administration (Spring '04, Spring '06, Spring '08, Spring '10, Spring '12, Fall '14, Spring '16)
- Undergraduate course in C and UNIX programming (Spring '05, Spring '07, Spring '09, Spring '11, Spring '13)
- Undergraduate course introducing programming (Fall '12)
- Undergraduate course in data structures and Java (Fall '10)
- Undergraduate course in system administration (Spring '04, Spring '06, Spring '08, Spring '10, Spring '12, Fall '14)
- Undergraduate course in networking (Spring '02, Fall '03, Fall '04, Fall '05, Fall '06, Fall '07, Fall '08, Fall '09)
- Undergraduate course in search engine marketing (Fall '06)
- Graduate course in research methods (Fall '14, Fall '15, Fall '16, Fall '17)
- Graduate/undergraduate course in WWW search engines (Fall '02, Fall '03, Fall '04, Spring '07, Spring '09, Spring '15, Spring '17)
- Graduate/undergraduate course in data mining (Spring '15)
- Graduate course in advanced programming techniques (Fall '12)
- Graduate seminar in network mining (Spring '10)
- Seminar in technology behind social networking (Spring '09)
- Graduate seminar in Web mining (Fall '05, Spring '08)
- Seminar in business applications of social networking (Spring '08)
- Graduate seminar in advanced networking topics (Spring '03)
- Graduate course in Web technologies and performance (Fall '01)

DEPARTMENT OF COMPUTER SCIENCE, RUTGERS UNIVERSITY
Served as Instructor for:

New Brunswick, NJ

- Second-semester undergraduate course in data structures (Spring '98)
- Undergraduate course in computer architecture (Summer '92)

Served as teaching assistant for:

- Graduate course in artificial intelligence (Fall '98)
- Advanced undergraduate course in digital system design (Spring '93)
- Undergraduate course in software engineering (Fall '92)
- Undergraduate course in computer architecture (Fall '91, Spring '92)

**Other
Experience**

VARIOUS LAW FIRMS
September 2002 – Present **Technical Consulting Expert**

Provided technical expertise on caching, content delivery, and search technologies for patent attorneys in multiple cases.

UNITED ONLINE New York City
August 2003 **Consultant**

Consulted on Web prefetching techniques and simulation thereof.

WWW.WEB-CACHING.COM
July 1998 – Present **Creator and Maintainer**

A leading Web resource on web caching and content delivery networks providing news, tools, references, and related resources.

DAVISON COMPUTER SERVICES Lehigh Valley, PA
May 1992 – Present **Founder and Owner**

While originally started and operated to provide local (NJ) Internet access (1992-1996), DCS now serves as a side business for occasional consulting, and served as parent for the now defunct *Brian's Books* online business of selling discounted computer books (operational 1995-2004).

PRINCIPIA PRODUCTS, A DIV. OF COMPUCON SERVICES CORP. West Chester, PA
Summer 1991 – Summer 1994 **Application Developer**

Developed interface and library routines for a Microsoft Windows-based application to read standardized forms.

GENERAL ELECTRIC AEROSPACE Valley Forge, PA
Summers 1989, 1990 **Summer Intern**

Wrote network monitoring scripts and reporting modules.

**Community
Service**

BETHLEHEM AREA MORAVIANS Bethlehem, PA
February 2017 – Present **Secretary/Treasurer**
February 2016 – Present **Member, Board of Directors**

EAST HILLS MORAVIAN CHURCH Bethlehem, PA
January 2015 – Present **President, Church Council**
January 2008 – August 2013 **Member, Board of Elders**

**Honors and
Awards**

Best paper award at the International Workshop on Profiling and Searching Data on the Web (Profiles & Data:Search'18, co-located with The Web Conference), Lyon, France, for Generating Schema Labels through Dataset Content Analysis. April 2018.

Outstanding Senior PC Member for the Twenty-fifth International Joint Conference on Artificial Intelligence (IJCAI), New York City, July 2016.

Outstanding Reviewer for the Eighth International Conference on Web Search and Data Mining, WSDM 2015.

Senior member, IEEE, promoted August 2011.

Best poster paper award at the 20th International World Wide Web Conference, Hyderabad, India, for *Predicting Popular Messages in Twitter*. March 2011.

Highly Commended Award Winner from *Internet Research* for our 2008 article on “Connecting P2P to the web: Lessons from a prototype Gnutella-WWW gateway”, 2009.

National Science Foundation Faculty Early Career Development (CAREER) Award, 2006.

Microsoft Live Labs “Accelerating Search” award, 2006.

Best paper award at the 30th International Conference for the Resource Management and Performance Evaluation of Enterprise Computing Systems (CMG2004) for *Implementing a Web Proxy Evaluation Architecture*. December 2004.

Sigma Xi. Elected associate member, 1991. Promoted to full member, 2004.

Rutgers University Dean’s Award for Graduate Research. Spring, 2001.

AT&T Student Research Symposium (a regional ACM Student Research Competition). Awarded *First Place, PhD Division* for “Measuring the Performance of Prefetching Proxy Caches.” November 1998.

ACM International Student Research Competition. Awarded *Third Place, Graduate Category* for the same work. March 1999.

Tau Beta Pi. Elected member, 1991.

National Merit Scholar. 1987.

Professional Memberships

Association for the Advancement of Artificial Intelligence (AAAI)
American Society for Information Science and Technology (ASIS&T)
Association for Computing Machinery (ACM)
Institute of Electrical and Electronics Engineers (IEEE)
Internet Society (ISOC)

Editorial Service

Editor-in-Chief, *ACM Transactions on the Web*, 2015–present. (Joint with M. Winslett 2015-2017).
Associate Editor, *Frontiers in Big Data: Data Mining and Management*, 2018–present.
Associate Editor, *ACM Transactions on the Web*, 2013–2014.
Special Section Co-editor, with M. Najork, *ACM Transactions on the Web*, Adversarial Issues in Web Search, February 2008.

University Service

Undergraduate Data Science minor (2017-present, director)
Data Governance Executive Committee (2017-present, member)
LTS Data Architect Recruiting Committee (2017-2018, member)
University Strategic Analytics Planning Committee (2016-2017, faculty advisor)
Middle States Accreditation Self-Study Working Group on Support for Student Experience (2016-2017, co-chair)
Ad hoc committee for data sciences curriculum development (2016-present, chair)
Lehigh Advocates and Allies for Gender Equity (part of ADVANCE) (2016-present, member)
LTS ACIS - Advisory Council for Information Services (Fall 2015, member)

Faculty Committee on Student Life (2009-2011 member; 2011-2012, chair, 2013 past chair)

Faculty Steering Committee (2011-2012, member while chair of FCSL)

Research Computing Steering Committee (2004-2008, 2017-present member; 2008-2009, chair; 2009-2013, 2014-2017, co-chair)

Faculty Advisory Council for the Institute for Data, Intelligent Systems, and Computation (2018-present, member)

Energy Systems Engineering Institute Strategic Advisory Council (2018-present, member)

RCEAS Faculty Council for an Institute for Data, Intelligent Systems and Computation (2017-2018, member)

RCEAS Dean Search Committee (2014-2015, member)

Mountaintop Campus Designer Selection Committee (2013, member)

RCEAS Task Force on First Year Engineering (2003-2004, 2009-2012, member)

Faculty Search Committees (2018-present, co-chair, RCEAS Fundamentals of Data Science and Analytics interdisciplinary committee 2017-2018, member, RCEAS Data Science and Analytics Interdisciplinary committee 2016-2017, member, Data X / BIS Faculty committee; 2015-2016, member, ISE POP committee; 2015-2016, co-chair, Data X / Marketing committee; 2014-2015, member, BIS committee, member, CSE POP committee; 2012-2013, chair, CSE committee; 2009-2011, member, ISE committee; 2002-2003, 2007-2009, member, CSE committee; 2001-2002, member, ECE committee.

High Performance Computing subcommittee of the Research Advisory Council for Engineering (2008-2013, member)

Lehigh Grants Internal Review Committee (2012-present, member)

Student life leadership awards selection committee (2011, member)

CSE Computer Facilities Committee (2001-2009, 2010-2011, 2017-present chair; 2009-2010, 2014-2017, member)

CSE Graduate Admissions Committee (2014-2016 chair)

CSE Graduate Program Advisor (2012-2013, Master's program advisor; 2014-2016, PhD program advisor)

CSE curriculum committee (2001, 2005-2009, 2011-2012, member)

CSE curriculum revision committee (2014-2017, member)

CSE Publicity and Web Committee (2016-2017, chair)

CSE Building C Rules and Procedures Committee (2016-present, member)

Lehigh in Ireland summer program (2015-present, departmental advisor)

**Conference
Leadership
Service**

Co-Program Chair, *Forty-first International ACM SIGIR Conference on Research and Development in Information Retrieval*, July 2018, Ann Arbor, Michigan.

Tutorial Co-chair, *Tenth ACM International Conference on Web Search and Data Mining (WSDM)*, February 2017, Cambridge, UK.

Doctoral Consortium Co-chair, *Thirty-ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, July 2016, Pisa, Italy.

Best Paper Selection Committee Chair, *Ninth ACM International Conference on Web Search and Data Mining (WSDM)*, February 2016, San Francisco, CA.

Doctoral Consortium Co-chair, *Thirty-eighth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, August 2015, Santiago, Chile.

Workshop Chair, *Seventh ACM International Conference on Web Search and Data Mining (WSDM)*, February 2014, New York City.

Co-organizer, *Workshop on Appification of the Web (AppWeb 2012)*, April 2012, Lyon, France.

Co-chair, *First WSDM Doctoral Consortium*, February 2012, Seattle.

Co-chair, *Third ACM International Conference on Web Search and Data Mining (WSDM)*, February 2010, New York City.

Co-organizer, *ECML/PKDD Graph Labelling Workshop and Web Spam Challenge*, September 2007, in Warsaw, Poland.

Co-organizer, *The Web Spam Challenge*, part of the EU PASCAL Network of Excellence Challenge Program, 2007.

Co-organizer and Program Co-chair, *3rd International Workshop on Adversarial Information Retrieval on the Web (AIRWeb 2007)*, May 2007, Banff, Canada.

Co-organizer and Program Chair, *2nd International Workshop on Adversarial Information Retrieval on the Web (AIRWeb '06)*, August 2006, Seattle, WA.

Organizer and Program Chair, *1st International Workshop on Adversarial Information Retrieval on the Web (AIRWeb '05)*, May 2005, Chiba, Japan.

General Chair, *8th Int'l Workshop on Web Content Caching and Distribution*, September/October 2003, Hawthorne, NY.

**Program
Committees**

Twenty-eighth International World Wide Web Conference (WWW), May 2019, San Francisco, CA. Modeling, Personalization, and Experience track. *Senior Program Committee Member*.

Twelfth ACM International Conference on Web Search and Data Mining (WSDM), February 2019, Melbourne, Australia. *Senior Program Committee Member*.

Sixth IEEE International Conference on Big Data, December 2018, Seattle, WA.

First ACM/IEEE Workshop on Hot Topics on Web of Things (HotWoT2018), collocated with ACM/IEEE SEC, October 2018, Bellevue, WA.

Twenty-seventh International World Wide Web Conference (WWW), April 2018, Lyon, France. *Senior Program Committee Member for the Web Search and Mining track*.

Eleventh ACM International Conference on Web Search and Data Mining (WSDM), February 2018, Los Angeles, CA. *Senior Program Committee Member*.

Fifth IEEE International Conference on Big Data, December 2017, Boston, MA.

Fifth IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb), October 2017, San Jose, CA.

Fortieth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 2017, Tokyo, Japan. *Senior Program Committee Member*.

Twenty-sixth International World Wide Web Conference (WWW), April 2017, Perth, Australia. User Modeling, Recommendation and Personalization track.

Tenth ACM International Conference on Web Search and Data Mining (WSDM), February 2017, Cambridge, UK. *Senior Program Committee Member*.

Twenty-fifth ACM International Conference on Information and Knowledge Management (CIKM 2016), October 2016, Indianapolis, IN.

Fourth IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb 2016), October 2016, Washington, DC.

Fifth-fourth Annual Meeting of the Association for Computational Linguistics (ACL 2016), August 2016, Berlin Germany. Document Analysis track.

Twenty-fifth International Joint Conference on Artificial Intelligence (IJCAI), July 2016, New York City. *Senior Program Committee Member for AI and the Web*.

Twenty-fifth International World Wide Web Conference (WWW), April 2016, Montreal, Canada. *Senior Program Committee Member for Behavioral Analysis and Personalization*.

Ninth ACM International Conference on Web Search and Data Mining (WSDM), February 2016, San Francisco, CA. *Senior Program Committee Member*.

IEEE/WIC/ACM International Conference on Web Intelligence, Special Session on Quality of User Generated Content, December 2015, Singapore.

Twenty-fourth ACM International Conference on Information and Knowledge Management (CIKM), October 2015, Melbourne, Australia. *Senior Program Committee Member for Information Retrieval*.

Twenty-first ACM SIGKDD Conference on Knowledge Discovery and Data Mining, August 2015, Sydney, Australia.

SIGIR 2015 Best Paper Award Committee, August 2015, Santiago, Chile.

Twenty-fourth International World Wide Web Conference (WWW), May 2015, Florence, Italy. *Web Mining Area Co-Chair*.

Eighth ACM International Conference on Web Search and Data Mining (WSDM), February 2015, Shanghai, China.

Twenty-third ACM International Conference on Information and Knowledge Management (CIKM), November 2014, Shanghai, China. *IR Area Chair for Adversarial IR and Spam*.

Twenty-eighth AAAI Conference on Artificial Intelligence, August 2014, Quebec City, Canada. *Senior Program Committee Member*.

Thirty-seventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2014, Gold Coast, Australia. *Area chair.*

Twenty-third International World Wide Web Conference (WWW), April 2014, Seoul, Korea. Web mining track.

Fourth Joint WICOW/AIRWEB Workshop on Web Quality (WebQuality 2014), April 2014, Seoul, Korea.

Seventh ACM International Conference on Web Search and Data Mining (WSDM), February 2014, New York City.

WSDM 2014 Tutorial Selection Committee, February 2014, New York City.

WSDM 2014 Doctoral Consortium Selection Committee, February 2014, New York City.

Nineteenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining, August 2013, Chicago, IL.

Seventh International AAAI Conference on Weblogs and Social Media (ICWSM), July 2013, Boston, MA.

Thirty-sixth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2013, Dublin, Ireland. *Area chair.*

Twenty-second International World Wide Web Conference (WWW), May 2013, Rio de Janeiro, Brazil. Content analysis track.

Sixth ACM International Conference on Web Search and Data Mining (WSDM), February 2013, Rome.

Demonstrations at the Twenty-first ACM International Conference on Information and Knowledge Management (CIKM), October 2012, Maui, Hawaii.

Thirty-fifth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 2012, Portland, Oregon.

Second Workshop on Social Media Analytics (SOMA), August 2012, Beijing, China.

Sixth International AAAI Conference on Weblogs and Social Media (ICWSM), June 2012, Dublin, Ireland.

Twenty-first International World Wide Web Conference (WWW), April 2012, Lyon, France. Web mining track.

Joint WICOW/AIRWEB Workshop on Web Quality (WebQuality 2012), April 2012, Lyon, France.

First International Workshop on Large Scale Network Analysis (LSNA 2012), April 2012, Lyon, France.

Fifth ACM International Conference on Web Search and Data Mining (WSDM), February 2012, Seattle.

WSDM 2012 workshop selection committee. February 2012, Seattle.

Workshop on Web Search Click Data at WSDM, February 2012, Seattle.

Eighteenth International Symposium on String Processing and Information Retrieval (SPIRE), October 2011, Pisa, Italy.

Seventeenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining, August 2011, San Diego, CA.

Twenty-fifth AAAI Conference on Artificial Intelligence, August 2011, San Francisco, CA. Senior Program Committee Member of AI and the Web track.

Fifth International AAAI Conference on Weblogs and Social Media (ICWSM), July 2011, Barcelona, Spain.

Thirty-fourth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2011, Beijing, China. Also served on best paper award committee.

Twentieth International World Wide Web Conference (WWW), March 2011, Hyderabad, India. Chair of Abuse, Security, and Privacy Track.

Joint WICOW/AIRWEB Workshop on Web Quality, March 2011, Hyderabad, India.

Fourth ACM International Conference on Web Search and Data Mining (WSDM), February 2011, Hong Kong. Also served on workshops selection committee and best paper award committee.

NIPS 2010 Workshop on Machine Learning for Social Computing, December 2010, Whistler, BC, Canada.

Thirty-third Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2010, Geneva, Switzerland.

SIGIR 2010 Workshop on Feature Generation and Selection for Information Retrieval (FGSIR), July 2010, Geneva Switzerland.

SIGIR 2010 Web N-gram Workshop, July 2010, Geneva, Switzerland.

Sixteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 2010, Washington, DC. Senior Program Committee Member.

SIGKDD Workshop on Social Media Analytics (SOMA), July 2010, Washington, DC.

Fourth International AAAI Conference on Weblogs and Social Media (ICWSM), May 2010, Washington, DC.

Nineteenth International World Wide Web Conference, April 2010, Raleigh, NC. Search track.

IEEE International Conference on Data Mining (ICDM), December 2009, Miami.

CIKM 2009 Workshop on Complex Networks in Information & Knowledge Management (CNIKM), November 2009, Kong Kong.

Thirty-second Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2009, Boston. Also served as a mentor.

Fifth International Workshop on Adversarial Information Retrieval on the Web (AIRWeb), April 2009, Madrid, Spain.

Eighteenth International World Wide Web Conference, April 2009, Madrid, Spain. Search track.

ECIR Workshop on Information Retrieval over Social Networks, April 2009, Toulouse, France.

Second ACM International Conference on Web Search and Data Mining (WSDM), February 2009, Barcelona, Spain. Main track and late-breaking papers track.

Workshop on Web Search Click Data (WSCD09) at WSDM, February 2009, Barcelona, Spain.

Wikis, Blogs, Bookmarking Tools - Mining the Web 2.0 Workshop at ECML/PKDD, September 2008, Antwerp, Belgium.

Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2008, Las Vegas.

First Workshop on Information Retrieval for Advertising, July 2008, Singapore.

Thirty-first Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2008, Singapore.

Seventeenth International World Wide Web Conference, April 2008, Beijing, China. Search track.

Fourth International Workshop on Adversarial Information Retrieval on the Web (AIRWeb), April 2008, Beijing, China.

Workshop on Targeting and Ranking for Online Advertising (TROA), April 2008, Beijing, China.

First ACM International Conference on Web Search and Data Mining (WSDM), February 2008, Stanford, CA.

Encuentro Internacional de Ciencias de la Computación: ENC-IR 2007, September 2007, Mexico.

Thirtieth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2007, Amsterdam.

Sixteenth International World Wide Web Conference, May 2007, Banff, Alberta, Canada. Search track and posters.

Document Recognition and Retrieval XIV (IS&T/SPIE Annual Symposium on Electronic Imaging), January 2007, San Jose, CA.

WebKDD 2006: SIGKDD Workshop on Web Mining and Usage Analysis, August 2006, Philadelphia, PA.

Twenty-ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 2006, Seattle, WA.

Fifteenth International World Wide Web Conference, May 2006, Edinburgh, Scotland. Search track.

Document Recognition and Retrieval XIII (IS&T/SPIE Annual Symposium on Electronic Imaging), January 2006, San Jose, CA.

Fourteenth International World Wide Web Conference, May 2005, Chiba, Japan. Search track.

Second International Workshop on Human Interactive Proofs, May 2005, Bethlehem, PA.

Thirteenth International World Wide Web Conference, May 2004, Budapest, Hungary. Search track.

Third International Workshop on Web Dynamics, May 2004, Budapest, Hungary.

IUI Workshop on Behavior-based User Interface Customization, January 2004, Island of Madeira, Portugal.

Twentieth International Conference on Machine Learning, August 2003, Washington, DC.

Practical Aspects of Performance Analysis Workshop, June 2002, Marina Del Rey, CA.

Research Grant Reviewing	NSF Computer & Information Science & Engineering Research Grants Council of Hong Kong Dutch National Science Foundation (NWO) Canadian Network of Centres of Excellence IBM Far-Reaching Research Vienna Science and Technology Fund WWTF
Article Reviewing Service	<i>ACM Computing Surveys</i> <i>ACM Crossroads</i> <i>ACM Transactions on Information Systems</i> <i>ACM Transactions on Intelligent Systems and Technology</i> <i>ACM Transactions on Internet Technology</i> <i>ACM Transactions on Knowledge Discovery from Data</i> <i>ACM Transactions on the Web</i> <i>Communications of the ACM</i> <i>Computers & Informatics</i> <i>Computer Networks</i> <i>Data Mining and Knowledge Discovery, IEEE Internet Computing</i> <i>IEEE Transactions on Computers</i> <i>IEEE Transactions on Knowledge and Data Engineering</i> <i>IEEE Transactions on Neural Networks</i> <i>IEEE Transactions on Systems, Man, and Cybernetics</i> <i>Information Processing & Management</i> <i>Information Retrieval</i> <i>Information Systems</i> <i>International Journal of Systems Architecture</i> <i>Internet Research</i> <i>Journal of Artificial Intelligence Research</i> <i>Journal of American Society for Information Science and Technology</i> <i>Journal of Systems and Software</i> <i>Journal of Web Engineering</i> <i>Machine Learning Journal</i> <i>Web Intelligence and Agent Systems (WIAS): An International Journal</i> <i>World Wide Web: Internet and Web Information Systems</i> Adaptive Hypermedia and Adaptive Web-based Systems conference International Joint Conference on Artificial Intelligence (IJCAI) International Conference on High Performance Computing (HiPC) International Wireless Communications and Mobile Computing Conference (IWCMC) ACM SIGMOD Conference IEEE International Conference on Pervasive Computing and Communications IEEE Wireless Communications and Networking Conference IEEE INFOCOM Conference
Book Reviewing Service	<i>Addison Wesley</i> <i>Cambridge University Press</i> <i>McGraw-Hill</i> <i>Morgan Kaufmann</i> <i>Pearson UK</i> <i>Prentice Hall</i> <i>Wiley</i>

Other Reviewing Dissertation reviewer for McGill University, Univ. Paris-Sud XI, and the Australian National University
External dissertation proposal reviewer for University of Chile
External evaluator for tenure and promotion at two US universities
External evaluator for promotion at a non-US university

Dissertation Title *The Design and Evaluation of Web Prefetching and Caching Techniques*

Dissertation Abstract User-perceived retrieval latencies in the World Wide Web can be improved by pre-loading a local cache with resources likely to be accessed. A user requesting content that can be served by the cache is able to avoid the delays inherent in the Web, such as congested networks and slow servers. The difficulty, then, is to determine what content to prefetch into the cache.

This work explores machine learning algorithms for user sequence prediction, both in general and specifically for sequences of Web requests. We also consider information retrieval techniques to allow the use of the content of Web pages to help predict future requests. Although history-based mechanisms can provide strong performance in predicting future requests, performance can be improved by including predictions from additional sources.

While past researchers have used a variety of techniques for evaluating caching algorithms and systems, most of those methods were not applicable to the evaluation of prefetching algorithms or systems. Therefore, two new mechanisms for evaluation are introduced. The first is a detailed trace-based simulator, built from scratch, that estimates client-side response times in a simulated network of clients, caches, and Web servers with various connectivity. This simulator is then used to evaluate various prefetching approaches. The second evaluation method presented is a novel architecture to simultaneously evaluate multiple proxy caches in a live network, which we introduce, implement, and demonstrate through experiments. The simulator is appropriate for evaluation of algorithms and research ideas, while simultaneous proxy evaluation is ideally suited to implemented systems.

We also consider the present and the future of Web prefetching, finding that changes to the HTTP standard will be required in order for Web prefetching to become commonplace.

Dissertation Advisor Haym Hirsh (Professor and former Dean of the Faculty of Computing and Information Science at Cornell University; former Professor and department chair, Dept. of Computer Science, Rutgers University; former Director of the Division of Information and Intelligent Systems at the National Science Foundation)

Nationality U.S. Citizen

References and extended curriculum vitae available upon request.