

Lidan Wang

CONTACT INFORMATION 3126 A.V. Williams Building Phone: (608) 334-0690
Computer Science Department Email: lidan@cs.umd.edu
University of Maryland Web: <http://www.cs.umd.edu/~lidan>
College Park, MD 20742

RESEARCH INTERESTS Information Retrieval, Data and Text Mining, Machine Learning, Cloud Computing
Large-Scale Text and Data Management

EDUCATION **University of Maryland, College Park, Maryland, USA**
Ph.D., Computer Science, Expected May 2012
Thesis advisor: Prof. Jimmy Lin
Thesis: Learning to Efficiently Rank

University of Wisconsin, Madison, Wisconsin, USA
M.S., Computer Science, May 2006
Advisor: Prof. Mark Craven
Thesis: Support Vector Machines for Gene Regulatory Network Prediction

University of Florida, Gainesville, Florida, USA
B.S., Computer Science, May 2003

AWARDS Best paper award runner up, CIKM 2010 (acceptance rate 13.4%; top 1%).
Co-authored a funded NSF grant proposal *Learning to Efficiently Rank with Cascades* (IIS-1144034) based on my Ph.D. thesis topic (2011).
Best results, TREC 2009, Web Track Category B.
Dean's Fellowship from the Dept. of Computer Science, Univ. of Maryland (2006-2008)
AIAA Scholarship (awarded to 25 undergraduate students in the United States) (2000).

PUBLICATIONS **Refereed Conference Papers**
Lidan Wang, Paul N. Bennett, Kevyn Collins-Thompson. Robust Ranking Models via Risk-Sensitive Optimization. *To Appear, Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR 2012). [Acceptance Rate: 20%]
Lidan Wang, Jimmy Lin, Donald Metzler. A Cascade Ranking Model for Efficient Ranked Retrieval. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR 2011), pages 105-114, 2011. [Acceptance Rate: 19%]
Lidan Wang, Donald Metzler, Jimmy Lin. Ranking under Temporal Constraints. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management* (CIKM 2010), pages 79-88, 2010. [Acceptance Rate: 13.4%. **Best Paper Award Finalist; top 1%**].
Lidan Wang, Jimmy Lin, Donald Metzler. Learning to Efficiently Rank. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR 2010), pages 138-145, 2010. [Acceptance Rate: 16.7%]

Lidan Wang, Douglas Oard. Context-based Message Expansion for Disentanglement of Interleaved Text Conversations. In *Proceedings of the 10th Annual Conference of the North American Chapter of the Association for Computational Linguistics, Human Language Technologies (NAACL-HLT 2009)*, pages 200-208, 2009. [Acceptance Rate: 29%]

Lidan Wang, Amol Deshpande. Predictive Modeling-based Data Collection in Sensor Networks. In *Proceedings of the Fifth European Conference on Wireless Sensor Networks (EWSN 2008)*, pages 34-51, 2008. [Acceptance Rate: 21%; **Best Paper Award**].

Jun Gong, **Lidan Wang**, Douglas Oard. Matching Person Names Through Name Transformation. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*, pages 1875-1878, 2009.

Jun Gong, **Lidan Wang**, Douglas Oard. Personal Name-Matching Through Name Transformation. In *Proceedings of the 9th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2009)*, June 2009.

Pin Wang, Chia-jung Tsui, Kenneth Fleischmann, Douglas Oard, **Lidan Wang**. Understanding IT Innovations through Discourse Analysis. In *Proceedings of the iConference*, Chapel Hill, NC, February 2009.

Workshop Publications and Technical Reports

Jimmy Lin, Donald Metzler, Tamer Elsayed, **Lidan Wang**. Of Ivory and Smurfs: Loxodontan MapReduce Experiments for Web Search. In *Proceedings of the 18th Text REtrieval Conference (TREC 2009)*, Gaithersburg, Maryland. November 2009.

Lidan Wang, Douglas Oard. Query Expansion for Noisy Legal Documents. In *Proceedings of the 17th Text REtrieval Conference (TREC 2008)*, Gaithersburg, Maryland.

Lidan Wang, Chloe Schulze. Badger: An Entropy-based Web Search Clustering System with Randomization and Voting. Technical Report 1537, Computer Science Department, University of Wisconsin, Madison, 2005

Lidan Wang, Jeffrey Freschl. Aggregate Sharing in Stream Databases. Technical Report 1536, Computer Science Department, University of Wisconsin, Madison, 2005

Full Papers Under Submission

Lidan Wang, Donald Metzler, Jimmy Lin. A Cascade-Based Approach for Ranking with Temporal Constraints. In this paper, I describe an improved method for returning ranking results within time budgets by using a cascade ranking model.

RESEARCH EXPERIENCE

Research Assistant, CS Dept., Univ. of Maryland, College Park. Spring 2009-present

• **Topic (Dissertation research):** *Web data mining, Information retrieval, Machine learning*

Proposed *Learning to Efficiently Rank*, a new machine learning-based framework for enabling efficient and scalable Web-scale data mining and ranking. Designed new learning to rank algorithms and scalable retrieval models for improving ranking effectiveness *without* incurring high run-time latency for Web-scale documents. Several methods are proposed in this framework – including ranking under time constraints, optimally trading off model effectiveness and run-time latency, and jointly improving result effectiveness and efficiency via a cascade ranking model. This research led to a NSF research grant (IIS-1144034).

Research Assistant, CS Dept., Univ. of Maryland, College Park. Fall 2008

- **Topics:** *Information Retrieval, Natural Language Processing*

Worked on information extraction and topic detection for text data generated from online user communities. Developed algorithms for overcoming data sparsity issues and improving the robustness of information extraction and topic detection over sparse text data by mining and utilizing temporal and spatial contexts.

Research Assistant, CS Dept., Univ. of Maryland, College Park. Fall 2006-Fall 2007

- **Topics:** *Databases, Machine Learning*

Distributed and streaming data exists in many areas such as the web, and sensing/tracking systems. The data's distributed and streaming nature poses many challenges in answering queries, and collecting and analyzing data in real-time. I designed methods based on statistical graphical models for performing real-time data extraction and analysis over sensor networks by modeling and exploiting rich temporal and spatial correlations existed among sensor data nodes.

Research Assistant, CS Dept., Univ. of Wisconsin, Madison. Fall 2005-Spring 2006

- **Topics:** *Bioinformatics, Machine Learning*

Developed machine learning and data mining techniques for bioinformatics problems of gene regulatory network prediction and expression profile analysis.

TEACHING
EXPERIENCE

University of Maryland, College Park, MD

- CMSC 131 Object-Oriented Programming, Spring 2007, Spring 2008.
Weekly lab sessions for students, designing homework assignments, and evaluation.

University of Wisconsin, Madison, WI

- CS 776 Advanced Bioinformatics, Spring 2005
 - CS 576 Introduction to Bioinformatics, Fall 2004
- Duties involved participation in the designing of the course, holding weekly office hours for students, and developing search and alignment homework and solutions.

INDUSTRIAL
EXPERIENCE

Research Intern, Microsoft Research, Redmond, WA Summer 2011

- **Topics:** *Search Personalization, Information Retrieval, Machine Learning*

Mentors: Paul N. Bennett, Kevyn-Collins Thompson; Group Manager: Susan T. Dumais. Designed methods for learning large-scale robust search personalization models. Created new optimization objectives and learning techniques for personalization models that can improve robustness and consistency of result quality while minimizing performance variance across individual user queries.

Research & Design Intern, IBM Silicon Valley Lab, CA Summer 2006

- **Topics:** *Distributed Resource Management*

Developed a framework for managing application offloading onto the storage servers to utilize distributed system resources more effectively. Also developed programs to trace and manage offloaded application threads.

Software Engineering Intern, IBM, Lenexa, KS Summer 2005

• **Topics:** *Databases*

Worked on quality assurance knowledge base and developed methods for automatically identifying targeted test cases for new and changed code features.

PROFESSIONAL
SERVICES

Program Committee

- The 35th International ACM SIGIR Conference on Research and Development for Information Retrieval. SIGIR 2012
- The 34th annual European Conference on Information Retrieval. ECIR 2012.
- The 13th Annual Conference of the North American Chapter of the Association for Computational Linguistics, Human Language Technologies. NAACL-HLT 2012
- The 34th International ACM SIGIR Conference on Research and Development for Information Retrieval. SIGIR 2011 (poster)

Reviewer

- ACM Transactions on Information Systems (TOIS). 2011.

TALKS

Summer internship talk at Microsoft Research, Redmond. Learning Robust Ranking Models. August 2011.

Dean's Fellowship talk, Department of Computer Science, University of Maryland, College Park. Predictive Modeling-based Data Collection in Sensor Networks. October 2007

Conference talks: SIGIR 2011, CIKM 2010, SIGIR 2010, NAACL-HLT 2009, EWSN 2008.

RESEARCH
SOFTWARE

Ivory (ivory.cc), an open-source Web-scale search engine that incorporates many aspects of my thesis research.

PATENT

Paul N. Bennett, Kevyn Collins-Thompson, Lidan Wang. Optimizing A Ranker For A Risk-Oriented Objective. United States. Filed December 2011.

REFERENCES

Professor Jimmy Lin
University of Maryland, College Park
Hornbake Building #2117F, College Park, MD 20742
Phone: 617-905-7535; Email: jimmylin@umd.edu

Professor Hal Daume
University of Maryland, College Park
3227 A. V. Williams Building, College Park, MD 20742
Phone: 301-405-1073; Email: hal@umiacs.umd.edu

Dr. Donald Metzler
Yahoo! Research
Santa Clara, CA 95054
Phone: 413-218-4877; Email: metzler@gmail.com

Dr. Paul N. Bennett
Microsoft Research, Redmond
One Microsoft Way, Redmond WA 98052
Phone: 425-421-7378; Email: paul.n.bennett@microsoft.com

CITIZENSHIP

United States