

# Curriculum Vitae

December 1, 2011

## Le Zhao

Language Technologies Institute  
School of Computer Science  
Carnegie Mellon University  
Pittsburgh, PA 15213  
lezhao@cs.cmu.edu  
412-651-5976  
<http://www.cs.cmu.edu/~lezhao>

### Research Interests

Information retrieval, probabilistic retrieval models, structured documents and queries, large scale retrieval and text mining.

Computer facilitated human problem identification and solving, proactive search need prediction.

### Education

**Ph.D. Language and Information Technologies.** School of Computer Science, Carnegie Mellon University, Pittsburgh, 2006 – 2012 (expected).

**Advisor:** Professor Jamie Callan.

**Ph.D. Thesis:** Modeling and Predicting Term Mismatch for Full-text Retrieval.

**M.S. Language Technologies.** Carnegie Mellon University, Pittsburgh, 2008.

**M.E. Computer Science and Technologies.** Tsinghua University, Beijing, 2006.

**Master's Thesis:** The Computational Nature of Novelty Detection (**Honored**).

**B.E. Computer Science and Technologies.** Tsinghua University, Beijing, 2003.

### Professional Experience

**Graduate Research Assistant,** Language Technologies Institute, School of Computer Science, Carnegie Mellon University. 2006 – present.

**Research Intern,** Microsoft Research Redmond. URL indexing selection. Jun. – Sep., 2010.

**Software Intern,** Sogou.com. Relevance ranking for Chinese Web documents. Mar. – Jun., 2006.

**Graduate Research Assistant,** Department of Computer Science and Technologies, Tsinghua University. 2003 – 2006.

### Teaching Experience

**Teaching Assistant,** Information Retrieval (graduate), Language Technologies Institute, School of Computer Science, Carnegie Mellon University. Redesigned homeworks to do sharded indexing and structured querying in Map/Reduce (Hadoop), and produced lecture slides and documentation to help students quickly learn to use Hadoop. Fall, 2009.

**Teaching Assistant,** Information Retrieval (graduate), Language Technologies Institute, School of Computer Science, Carnegie Mellon University. Spring, 2009.

**Teaching Assistant,** Data Structures (university-wide undergraduate), Department of Computer

Science and Technology, Tsinghua University. Spring, 2004.

**Teaching Assistant**, Artificial Intelligence (undergraduate), Department of Computer Science and Technology, Tsinghua University. Fall, 2003.

### Refereed Conference and Journal Publications

1. L. Zhao and J. Callan. "P( $t$  | R) guided query term diagnosis and expansion". (*In preparation*). *The annual international ACM SIGIR conference on Research and development in information retrieval*. 2012b.
2. L. Zhao and J. Callan. "P( $t$  | R) variation and efficient prediction". (*In preparation*). *The annual international ACM SIGIR conference on Research and development in information retrieval*. 2012a.
3. W. Xu, R. Grishman and L. Zhao. "Passage retrieval for information extraction using distant supervision". In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011)*. 2011b.
4. W. Xu, J. Tetreault, M. Chodorow, R. Grishman and L. Zhao. "Exploiting syntactic and distributional information for spelling correction with Web-scale n-gram models". In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*. 2011a.
5. L. Zhao and J. Callan. "Term necessity prediction". In *Proceedings of the 19th ACM Conference on Information and Knowledge Management (CIKM 2010)*. Toronto. 2010.
6. L. Zhao and J. Callan. "Effective and efficient structured retrieval" (poster description). In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*. Hong Kong. 2009a.
7. L. Zhao and J. Callan. "A generative retrieval model for structured documents". In *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM 2008)*, Napa valley, USA. 2008.
8. L. Zhao, M. Zhang, S. Ma. "The nature of novelty detection". *Journal of Information Retrieval*, 9(5): pp. 521-541, November 2006.
9. L. Zhao, M. Zhang, S. Ma. "The empirical impact of the nature of novelty detection" (poster description). In *Proceeding of the 2nd Asian Information Retrieval Symposium (AIRS 2005)*, Springer LNCS Vol. 3689/2005 pp. 497-502, Jeju Island, Korea. October 13-15, 2005.

### Other Publications

1. L. Zhao and J. Callan. "How to make manual Conjunctive Normal Form queries work in patents search". In *Proceedings of the Twentieth Text REtrieval Conference (TREC 2011)*. Gaithersburg, Maryland. November 2011.
2. L. Zhao. "Modeling and predicting term mismatch for full-text retrieval". *PhD Thesis proposal*. Carnegie Mellon University. 2011.
3. L. Zhao and J. Callan. "Formulating simple structured queries using temporal and distributional cues in patents". In *Proceedings of the Eighteenth Text REtrieval Conference (TREC 2009)*. Gaithersburg, Maryland. November 2009. 2009b.
4. L. Zhao, C. Liang and J. Callan. "Extending relevance model for relevance feedback". In *Proceedings of the Seventeenth Text REtrieval Conference (TREC 2008)*. Gaithersburg, Maryland. November 2008.
5. M. Bilotti, L. Zhao, E. Nyberg, and J. Callan. "Focused retrieval over richly-annotated collections". *SIGIR 2008 Workshop on Focused Retrieval*. Singapore. 2008.

6. M. Heilman, L. Zhao, J. Pino, and M. Eskenazi. "Retrieval of reading materials for vocabulary and reading practice". *The 3rd Workshop on Innovative Use of NLP for Building Educational Applications*. Association for Computational Linguistics. 2008
7. Y. Zhu, L. Zhao and J. Callan. "Structured queries for legal search". In *Proceedings of the Sixteenth Text REtrieval Conference (TREC 2007)*. Gaithersburg, Maryland. November 2007.
8. L. Zhao, R. Ceng, M. Zhang and S. Ma, "THUIR at TREC 2005 Terabyte track". In *Proceedings of the 14th Text REtrieval Conference (TREC 2005)*, Gaithersburg, Maryland. November 2005.
9. L. Ru, L. Zhao, M. Zhang and S. Ma, "THUIR at TREC 2004 Novelty track". In *Proceedings of the 13th Text REtrieval Conference (TREC 2004)*, Gaithersburg, Maryland. November 2004.
10. M. Zhang, C. Lin, Y. Liu, L. Zhao and S. Ma, "THUIR at TREC 2003: Novelty, Robust and Web". In *Proceedings of the Twelfth Text REtrieval Conference (TREC 2003)*, Gaithersburg, Maryland. November 2003.
11. M. Zhang, R. Song, C. Lin, S. Ma, Z. Jiang, Y. Jin, Y. Liu and L. Zhao, "THU TREC 2002: novelty track experiments". In *Proceedings of the Eleventh Text REtrieval Conference (TREC 2002)* Gaithersburg, Maryland. November 2002.

## Professional Service

- **Program Committee:**

AAAI, 2012.

ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR), 2011 (poster).

International Symposium on String Processing and Information Retrieval (SPIRE), 2011.

Patent Information Retrieval workshop (PaIR), 2011.

SIGIR Query Representation and Understanding workshop (QRU), 2011.

- **Reviewer:**

ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR), 2010.

Asian Information Retrieval Symposium (AIRS), 2009.

European Conference on Information Retrieval (ECIR), 2009, 2011.

International World Wide Web Conference (WWW), 2008.

The Handbook of Technology Management. John Wiley & Sons, Inc., 2007.

## Honors and Awards

Outstanding Thesis Award, Masters, Tsinghua University, 2006.

Lab Award of Outstanding Academic Organization of LITS, Tsinghua University, 2004 – 2005.

Excellent Research Scholarship of LITS lab, Tsinghua University, 2003 – 2004.

The Third-Class Excellent Academic Scholarship, Tsinghua University, 2000 – 2001.

The Third-Class Award Physics Competition of Beijing College Students, 2000.

The Third-Class Excellent Academic Scholarship, Tsinghua University, 1999 – 2000.

The Second-Class Excellent Freshman Scholarship, Tsinghua University, 1999.