

# Curriculum Vitae

Oct 31, 2010

## 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, Retrieval models, Probabilistic language modeling; Structured document retrieval, Structured queries; Large scale information retrieval tasks

### Education

**Ph.D. candidate Language and Information Technologies.** School of Computer Science, Carnegie Mellon University, Pittsburgh, 2006 – present.

**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.

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

### Research Projects

2006–present, **Search engine support for human language technology:** working on a variety of issues related to creating text structure (via annotations such as semantic role labels, part of speech, named entities and semantic representations such as discourse representation theory), and using text structure (annotations, but also document fields and document metadata such as date, document reading level, topic category) in Question Answering, XML element retrieval, and intelligent tutoring; also working on automatic structured query formulation from unstructured keyword queries, to improve the matching between the query and relevant documents, especially for long queries or short documents.

2008–2009, **ClueWeb09 dataset creation:** a 1 billion page, PageRank prioritized crawl of the Web. I worked on PageRank calculation and language identification in Map/Reduce (Hadoop). The dataset is being used by 60+ research groups worldwide, and 4 tracks in TREC 2009.

2006–2008, **Reader specific lexical practice for improved reading comprehension:** developed software to gather and filter material for use in a computer-assisted language learning system

(REAP). I worked on large scale harvesting of Web material, filtering according to reading level, text quality etc. to fit tutoring needs, and selecting on topic categories to interest the students. The system is implemented in both condor and Hadoop (Map/Reduce), and the REAP tutor is used by language labs at the University of Pittsburgh's English Language Instruction (ELI) program.

### Other Research Projects

2009 summer, TREC chemical track – patent retrieval: formulating structured query based on patent content and patent date information.

2008 summer, TREC relevance feedback track: combining relevance and pseudo relevance information in language modeling.

2007 summer, TREC legal track (**2<sup>nd</sup> place**): worked on expert created Boolean queries, and relevance feedback using structured query terms.

2005 summer, TREC terabyte track: dependency based high accuracy models on half a terabyte data.

2002–2004 summer, TREC novelty detection track: characterizing the task for designing effective novelty detection algorithms, designing different novelty detection measures (**1<sup>st</sup> place** in 2002 and 2003).

### Industrial Experience

2006.3–2006.6 Internship at Sogou.com, worked on relevance ranking for Chinese web documents.

2010.6–2010.9 Internship at Microsoft Research Redmond on URL indexing selection. Mentors: Chao Liu @MSR, Xiaodong Fan and Yan Ke @Bing.

### Teaching Experience

- Fall 2009, **Information Retrieval** ([15-493](#) and [11-741](#), off-cycle teaching assistant), 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.
- Spring 2009, **Information Retrieval** ([11-741](#), teaching assistant), Language Technologies Institute, School of Computer Science, Carnegie Mellon University. Given a lecture on **Probabilistic retrieval models**.
- Spring 2004, **Data Structures** (teaching assistant), Department of Computer Science and Technology, Tsinghua University.
- Fall 2003, **Artificial Intelligence** (teaching assistant), Department of Computer Science and Technology, Tsinghua University.

### Refereed Conference and Journal Publications

1. L. Zhao and J. Callan. "Term necessity prediction". In *Proceedings of the 19th ACM Conference on Information and Knowledge Management (CIKM 2010)*. Toronto.

2. 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.
3. 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
4. L. Zhao, M. Zhang, S. Ma. "The nature of novelty detection". *Journal of Information Retrieval*, 9(5): pp. 521-541, November 2006.
5. 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. "Formulating simple structured queries using temporal and distributional cues in patents". *Proceedings of the Eighteenth Text REtrieval Conference (TREC 2009)*. Gaithersburg, Maryland, November 2009.
2. L. Zhao, C. Liang and J. Callan. "Extending relevance model for relevance feedback". *Proceedings of the Seventeenth Text REtrieval Conference (TREC 2008)*. Gaithersburg, Maryland, November 2008.
3. M. Bilotti, L. Zhao, E. Nyberg, and J. Callan. "Focused retrieval over richly-annotated collections". *SIGIR 2008 Workshop on Focused Retrieval*. Singapore. 2008.
4. 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
5. 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.
6. L. Zhao, R. Ceng, M. Zhang, S. Ma, "THUIR at TREC 2005 terabyte track". In *Proceedings of the 14th Text REtrieval Conference (TREC 2005)*, Gaithersburg, Maryland, November 2005.
7. L. Ru, L. Zhao, M. Zhang, S. Ma, "THUIR at TREC 2004 novelty track". In *Proceedings of the 13th Text REtrieval Conference (TREC 2004)* Gaithersburg, Maryland, November 2004.
8. M. Zhang, C. Lin, Y. Liu, L. Zhao, S. Ma, "THUIR at TREC 2003: novelty, robust and web". In *Proceedings of the Twelfth Text REtrieval Conference (TREC 2003)*, Gaithersburg, Maryland, November 2003.
9. M. Zhang, R. Song, C. Lin, S. Ma, Z. Jiang, Y. Jin, Y. Liu, L. Zhao, "THU TREC 2002: novelty track experiments". In *Proceedings of the Eleventh Text REtrieval Conference (TREC 2002)* Gaithersburg, Maryland, November 2002.

### Professional Service

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

- Reviewer, European Conference on Information Retrieval (ECIR), 2009
- Reviewer, Asian Information Retrieval Symposium (AIRS), 2009
- Reviewer, International World Wide Web Conference (WWW), 2008
- Reviewer, The Handbook of Technology Management. John Wiley & Sons, Inc., 2007

### **Honors and Awards**

- |           |   |
|-----------|---|
| 2006      | Outstanding Thesis Award, Masters, Tsinghua University                      |
| 2004-2005 | Lab Award of Outstanding Academic Organization of LITS, Tsinghua University |
| 2003-2004 | Excellent Research Scholarship of LITS lab, Tsinghua University             |
| 2000-2001 | The Third-Class Excellent Academic Scholarship, Tsinghua University         |
| 2000      | The Third-Class Award Physics Competition of Beijing College Students       |
| 1999-2000 | The Third-Class Excellent Academic Scholarship, Tsinghua University         |
| 1999      | The Second-Class Excellent Freshman Scholarship, Tsinghua University        |