

# Li Wang

## EDUCATION

- Doctor of Philosophy** [Mar 2010 -- Present]  
School of Engineering, The University of Melbourne, Australia  
•Achieved an average score of 92.5 out of 100 in coursework study
- Master of Information Technology** [Feb 2008 -- Dec 2009]  
School of Engineering, The University of Melbourne, Australia  
•Achieved an average score of 86.7 out of 100 in coursework study
- Bachelor of Engineering in Computer Science and Technology** [Sep 2001 -- Jun 2005]  
Computer School, Wuhan University, Hubei, China  
•Achieved an average score of 78.1 out of 100 in coursework study

## ACHIEVEMENTS

- Google Plenary Highlight Paper Award** [2011]  
in Conference on Empirical Methods in Natural Language Processing 2011
- Department Winner** [2010]  
of the inaugural University of Melbourne 3 Minute Thesis Competition  
(Thesis Title: Making Sense of Forums)
- Winning Team** [2010]  
of Australasian Language Technology Programming Competition 2010
- 2011 Melbourne School of Engineering Conference Travel Scholarship** [2011]
- 2011 EII PhD School Scholarship** [2011]
- 2010 EII PhD School Scholarship** [2010]
- HCSNet 2010 WinterFest Scholarship** [2010]
- Melbourne International Research Scholarship (MIRS)** [2010 -- 2013]
- Melbourne International Fee Remission Scholarship (MIFRS)** [2010 -- 2013]
- NICTA Top-up Scholarship** [2010 -- 2013]

## PUBLICATIONS

Li Wang, Su Nam Kim and Timothy Baldwin (to appear) The Utility of Discourse Structure in Forum Thread Retrieval, In *Proceedings of the Ninth Asian Information Retrieval Societies Conference (AIRS 2013)*, Singapore.

Timothy Baldwin, Paul Cook, Marco Lui, Andrew MacKinlay and Li Wang (to appear) How Noisy Social Media Text, How Diffrent Social Media Sources?, In *Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013)*, Nagoya, Japan.

Li Wang, Su Nam Kim and Timothy Baldwin (2012) [The Utility of Discourse Structure in Identifying Resolved Threads in Technical User Forums](#), In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, Mumbai, India, pp. 2739–2756.

Li Wang, Diana McCarthy and Timothy Baldwin (2011) [Predicting Thread Linking Structure by Lexical Chaining](#), In *Proceedings of the 2011 Australasian Language Technology Workshop (ALTW 2011)*, Canberra, Australia, pp. 76–85.

Li Wang, Marco Lui, Su Nam Kim, Joakim Nivre and Timothy Baldwin (2011) [Predicting Thread Discourse Structure over Technical Web Forums](#), In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, Edinburgh, UK, pp. 13–25. **(Google Plenary Highlight Paper Award)**

Li Wang, Su Nam Kim and Timothy Baldwin (2010) [Thread-level Analysis over Technical User Forum Data](#), In *Proceedings of the 2010 Australasian Language Technology Workshop (ALTW 2010)*, Melbourne, Australia, pp. 27–31.

Su Nam Kim, Li Wang and Timothy Baldwin (2010) [Tagging and Linking Web Forum Posts](#), In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL 2010)*, Uppsala, Sweden, pp. 192–202.

Timothy Baldwin, David Martinez, Richard Penman, Su Nam Kim, Marco Lui, Li Wang and Andrew MacKinlay (2010) [Intelligent Linux Information Access by Data Mining: the ILIAD Project](#), In *Proceedings of the NAACL 2010 Workshop on Computational Linguistics in a World of Social Media: #SocialMedia*, Los Angeles, USA, pp. 15–16.

## PROFESSIONAL ACTIVITIES

**Paper Reviewer** [2013]

of the 2013 Conference on Empirical Methods in Natural Language Processing

**Paper Reviewer**

of Knowledge and Information Systems

**Paper Reviewer** [2011]

of the 34th Annual International ACM SIGIR Conference

**Student Representative** [2011 – 2012]

of Australasian Language Technology Association (ALTA) Executive Committee

Responsibilities:

- Maintain ALTA student webpages.
- Communicate with Australasian students on language technology related issues.
- Attend monthly ALTA executive meetings on the behalf of Australasian students.
- Help take meeting notes.

## RESEARCH AND TEACHING EMPLOYMENT

**Research Intern, Microsoft Research Asia** [Feb 2013 -- May 2013]

This internship involves two related projects:

- **Entity linking:** research over entity linking literature for the purpose of optimizing an entity linking system.
- **Internet Explorer browsing trail analysis:** research over huge volume of Internet Explorer browsing data.

**Tutor of subject “Knowledge Technologies”, The University of Melbourne** [Jul 2012 -- Nov 2012]

This subject focuses on fundamentals of knowledge systems, including data acquisition and aggregation, knowledge extraction, text retrieval, machine learning and data mining.

Responsibilities:

- Prepare and plan tutorials.
- Deliver tutorials and organize discussions.
- Help students via web-based learning.
- Mark exam papers.

## RESEARCH PROJECTS

### **Knowledge Discovery and Extraction of Domain-Specific Web Data** [Mar 2010 -- Present]

This is my PhD research topic. My current research focus is on improving information access over troubleshooting-oriented technical user forums, by utilizing and combining Machine Learning, Natural Language Processing and Information Retrieval technologies.

### **Predicting Thread Linking Structure by Lexical Chaining** [Aug 2011 -- Dec 2009]

Automatic prediction of links between posts in a thread by using lexical chaining, a technique for identifying lists of related words (lexical chains) within a given discourse.

### **Post Classification and Thread Discourse Analysis over Technical User Forum Data** [Aug 2009 -- Dec 2009]

In-depth experiments and analysis on the task of thread discourse structure parsing.

### **Thread and Post Classification over Technical User Forum Data** [Feb 2009 -- Jul 2009]

This project focuses on two main tasks in the domain of troubleshooting-oriented web user forums. The first one is the automatic identification of Solution Type and Problem Source of forum threads. The second one is the initial exploration of forum thread discourse structure parsing.

## PROFESSIONAL SKILLS SUMMARY

### **Leadership**

- Selected as the student representative of ALTA for 2011-2012 term.
- Student representative of university subjects such as "Algorithms and Complexity" and "Machine Learning".
- Planned and trained computer science students without dancing experience for a successful dancing show.

### **Communication**

- Communicated with many researchers face-to-face or remotely for successful collaborations.
- Presented in various conferences, group meetings and competitions in front of 10 -- 500 people.
- Interacted effectively with audience in different conference and group presentations.
- Excellent writing skills, especially for research papers and reports.

### **Team Work**

- Participated in and won the 2010 Australasian Language Technology Programming Competition.
- Participated in and/or led various team projects of university subjects and achieved high scores.
- Collaborated with researchers to produce optimal research outcomes.

### **Problem-solving**

- Capable of solving research problems, by communicating with other researchers, team work and self-learning.
- Capable of solving complex problems in real time while doing tutorials.

## **TECHNICAL SKILLS**

### **Programming Language:**

Python, Java, SQL, Javascript, MATLAB, R

### **Software:**

LaTeX, NLTK, MALLET, Weka, Indri, Terrier, MaltParser, CRFSGD, CRF++, SVMrank

### **Operating Systems:**

Power user of Mac OS X and Microsoft Windows, system administration and extensive development within Linux environment