Wei-Yun Ma

PhD Candidate at Columbia University
wm2174@columbia.edu
http://www.cs.columbia.edu/~ma/
212-939-7121

EDUCATION

Columbia University, New York City, U.S.A. 09/2008-05/2014(expected)
Doctor of Philosophy in Computer Science
Advisor: Dr. Kathleen McKeown
Thesis: "Multi-Engine Machine Translation"

Columbia University, New York City, U.S.A. 09/2006-05/2008
Master of Science in Computer Science

National Chiao Tung University, Hsin-Chu, Taiwan 09/1997-06/1999
Master of Science in Computer Science and Information Engineering
Thesis: "Speed Improvement for Continuous Speech Recognition"

Yuan Ze University, Jung-Li, Taiwan 09/1993-06/1997
Bachelor of Engineering in Computer Science and Engineering

ACADEMIC ACHIEVEMENTS

- Passed PhD candidacy exam by an oral presentation of the topic - "Multi-Engine Machine Translation" 02/2012
  Committee:
  Dr. Michael Collins, Dr. Nizar Habash and Dr. Kathleen McKeown
- Designed a game agent of Penta, ranking top1 out of 30 participators in a tournament held in AI class at Columbia University 10/2006
- Designed a Chinese word segmentation system, ranking top1 in a contest held by ACL SIGHAN workshop 03/2003
- Designed a speech recognition system on 8051, wining the Quality Award given by Industrial Technology Research Institute, Taiwan, and sold to a domestic semi-conductor company in 1996. 12/1996
- Rank 1/120 in department (undergraduate junior) 09/1995 – 6/1996

RESEARCH EXPERIENCES

Microsoft Research 06/2013 – 09/2013
Research Intern

- Developed novel techniques for Language Model Adaptation.

NLP group at Columbia University 09/2009 – present
Research Assistant

- Participated in STAGES (Statistical Translation And GEneration using Semantics), which is a joint machine translation project, worked with Colorado U, USC, Rochester U. Colorado U and Brandeis U. My work is to fuse the output from the statistical MT systems based on syntactic and semantic analysis of the source and target language sentences. The results were published in four conference papers (NAACL, AMTA, MT Summit and ROCLING) and one journal paper (IJCLCLP)

NLP group at Columbia University 09/2006 – 05/2009
Research Assistant

- Participated in GALE (Global Autonomous Language Exploitation), which is a question-answering project. My work focuses on MT automatic postediting and exploting multiple MT engines to provide eventual responses. The results were published in two conference papers (ACL).

Institute of Linguistics, Academia Sinica, Taiwan 01/2006 –07/2006
Research Assistant

- Worked with an UK company, LexMC., to develop a Chinese Word Sketch Engine, which could automatically generate a one-page, corpus-derived summary of a word's grammatical and collocational behavior. The results were published in one conference paper (PACLIC).

Institute of Information Science, Academia Sinica, Taiwan 01/2000 – 02/2004 &
Research Assistant 09/2005 – 01/2006

- Developed techniques for Chinese word identification. The results were published in one conference paper (COLING), two workshop papers (SIGHAN) and one journal paper (IJALP).
- Designed uniform and effective tagging techniques for a heterogeneous Giga-word Corpus. The results were published in one conference paper (LREC).
- Designed a mechanism for Chinese corpus construction. The results were published in one conference paper (ROCLING).

Industrial Technology Research Institute, Taiwan 09/1995 – 6/1996
Research Intern

- Developed a spoken command recognition system working on 8051 IC. The system got the Quality Award in 1996 and sold to a domestic semi-conductor company.

PUBLICATIONS

Journal Articles

- **Wei-Yun Ma** and Kathleen McKeown. 2012. "Detecting and Correcting Syntactic Errors in Machine Translation Using Feature-Based Lexicalized Tree Adjoining Grammars". *International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP), Vol 17, No. 4, pp. 1-14.*
- **Wei-Yun Ma** and Keh-Jiann Chen. 2004. "Design of CKIP Chinese Word Segmentation System". *International Journal of Asian Language Processing (IJALP), Vol 14, No. 3, pp. 235-249.*

Book Chapters

- Kristen Parton, **Wei-Yun Ma**, Kathleen McKeown, and James Allan. 2010. "Using Query Time Information to Improve Multilingual Search and Response Generation". *In Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation. Joseph Olive (ed.)*

Conference Papers

- **Wei-Yun Ma** and Kathleen McKeown. 2013. "Using a Supertagged Dependency Model to Select a Good Translation in System Combination". *In Proceedings of NAACL-HLT*
- **Wei-Yun Ma** and Kathleen McKeown. 2012. "Phrase-level System Combination for Machine Translation Based on Target-to-Target Decoding". *In Proceedings of the 10th Biennial Conference of the Association for Machine Translation in the Americas (AMTA), San Diego, CA.*
- **Wei-Yun Ma** and Kathleen McKeown. 2012. "Detecting and Correcting Syntactic Errors in Machine Translation Using Feature-Based Lexicalized Tree Adjoining Grammars". *In Proceedings of Conference on Computational Linguistics and Speech Processing (ROCLING)*
- **Wei-Yun Ma** and Kathleen McKeown. 2011. "System Combination for Machine Translation Based on Text-to-Text Generation". *In Proceedings of Machine Translation Summit XIII*
- **Wei-Yun Ma** and Kathleen McKeown. 2009. "Where's the Verb Correcting Machine Translation During Question Answering". *In proceedings of ACL-IJCNLP*
- Kristen Parton, Kathleen R. McKeown, Bob Coyne, Mona T. Diab, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Heng Ji, **Wei-Yun Ma**, Adam Meyers, Sara Stolbach, Ang Sun, Gokhan Tur, Wei Xu and Sibel Yaman. 2009. "Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task". *In Proceedings of ACL-IJCNLP*
- Chu-Ren Huang, **Wei-Yun Ma**, Yi-Ching Wu, and Chih-Ming Chiu. 2006. "Knowledge-Rich Approach to Automatic Grammatical Information Acquisition:Enriching Chinese Sketch Engine with a Lexical Grammar". *In Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation (PACLIC)*
- **Wei-Yun Ma** and Chu-Ren Huang. 2006. "The Identification of Nominalizations in Mandarin Chinese Using Corpus-based Models (in Chinese)". *In Proceedings of Conference on Computational Linguistics and Speech Processing (ROCLING)*
- **Wei-Yun Ma** and Chu-Ren Huang. 2006. "Uniform and Effective Tagging of a Heterogeneous Giga-word Corpus". *In Proceedings of Language Resources and Evaluation Conference (LREC)*
- Keh-Jiann Chen and **Wei-Yun Ma**. 2002. "Unknown Word Extraction for Chinese Documents". *In Proceedings of COLING*
- **Wei-Yun Ma** and Chen Keh-Jiann. 2001. "Construction and Management for Chinese Corpus (in Chinese)". *In Proceedings of Conference on Computational Linguistics and Speech Processing (ROCLING)*

Workshop Papers

- Jia-Fei Hong, Chu-Ren Huang, and **Wei-Yun Ma**. 2006. "Corpus-based Extraction of Cross-strait Corresponding Words (in Chinese)". *In Proceedings of the seventh Chinese Lexical Semantics Workshop (CLSW)*
- **Wei-Yun Ma** and Keh-Jiann Chen. 2003. "A Bottom-up Merging Algorithm for Chinese Unknown Word Extraction". *In Proceedings of the second SIGHAN Workshop on Chinese Language Processing*
- **Wei-Yun Ma** and Keh-Jiann Chen. 2003. "Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff". *In Proceedings of the second SIGHAN Workshop on Chinese Language Processing*

PROFESSIONAL ACTIVITIES

- Reviewer : The 2012 Conference on Computational Linguistics and Speech Processing (ROCLING 2012)

TEACHING EXPERIENCES

- Teaching assistant for the course - “Natural Language Processing” Fall 2010
- Lecturer in Training Workshop of National Digital Archives Program 09/2002 & 09/2003
- Tutor in Junior High School Mathematics 06/2002 – 09/2003
- Teaching assistant for the course - “Network Security” Spring 1998
- Teaching assistant for the course - “Engineering Mathematics” Fall 1997

COMPUTER SKILLS

- **Languages**: Perl, Java, C, Matlab, SQL, mySQL, Assembly
- **Softwares**: Moses, Weka
- **Operating Systems**: Windows, Linux, Unix

REFERENCE

Kathleen McKeown (PhD advisor)
Professor in the Department of Computer Science at Columbia University
kathy@cs.columbia.edu

Xiaodong He (Internship mentor)
Researcher at Microsoft Research
xiaohe@microsoft.com

Y. C. Ju (Internship mentor)
Senior RSDE at Microsoft Research
yuncj@microsoft.com