

JAE DONG KIM

560 Bigelow St., Pittsburgh, PA 15207
Home: 412-422-2128, Mobile: 412-999-9035
Email: jdkim@cs.cmu.edu

EDUCATION

- **Carnegie Mellon University**, Pittsburgh, USA.
Ph.D. In Language Technologies, August 2010 (expected date).
- **Carnegie Mellon University**, Pittsburgh, USA.
M.S. In Language Technologies, May 2005.
- **Korea Advanced Institute of Science and Technology**, Daejeon, S. Korea.
B.Sc. With Honors in Computer Science, August 1996.

WORK EXPERIENCE

- **Internship - June 2009 ~ September 2009**
Y! Labs, Sponsored Search Dept., Santa Clara, USA.
 - ◆ Predicting Term Importance in Queries for Improved Query-Ad Relevance Prediction
 - ✓ Supervised Machine Learning approach to learn context-sensitive ad term importance
 - ✓ Query-Ad Relevance Prediction improved by 3%
 - ◆ Adding Translation Likelihood features to Query-Ad Relevance Prediction System
 - ✓ Added Symmetric Probabilistic Alignment scores between Query-Title, Query-Description, Query-Bid Phrases.
 - ✓ Query-Ad Relevance Prediction improved by 3%
 - ◆ Synonym dictionary extraction
 - ✓ English-English translation dictionary using doc-doc pairs instead of query-doc pairs from Yahoo web search engine results
- **System Integration - June 2008 ~ May 2009**
Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, USA.
 - ◆ Karen Refugee Help System
 - ✓ Message distribution system through phone call
 - ✓ Integration of Machine Translation and Speech Synthesis with VoIP and MySQL
 - ✓ English text input → translation to Karen text → Karen speech synthesis → delivery through Public Switched Telephone Network
 - ✓ Written in Python
- **Research Assistant - August 2003 ~ Present**
Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, USA.
 - ◆ Conducting research on Example-Based Machine Translation (EBMT)
 - ✓ Development of a Phrase-to-phrase alignment algorithm SPA(Symmetric Probabilistic Alignment) using statistical information drawn from training data
 - ✓ Implementation/Investigation of clustering algorithms to build

- Word Equivalence Classes
 - ✓ Incorporating Statistical Machine Translation (SMT) Techniques into Example-Based Machine Translation
 - ✓ Investigating a Chunk-Based EBMT system including automatic chunk boundary detection using statistical information such as IDF and Entropy, semi-supervised chunk boundary detection using CRF, chunk projection, and chunk alignment exploiting IBM models.
 - ✓ Internal evaluations show that the chunk-based EBMT system combined with SPA is comparable to the state-of-the-art SMT system (Moses) in translation quality
- ◆ Data mining projects
 - ✓ Accident prediction using shipping information
 - ✓ Accident prediction from construction(+assessment) information
 - ✓ Used Machine Learning techniques such as Random Forest, SVM, etc
- **Teaching Assistant**
Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, USA.
 - ◆ *Inventing the Future of Services* taught by Anatole Gershman, Fall 2008
 - ◆ *Search Engines and Data Mining using Machine Learning Techniques* taught by Jaime Carbonell, Spring 2009
- **Programmer - June 2006 ~ Present**
Wordwords, Inc., Seoul, S. Korea.
 - ◆ Joined in developing MOHANA (a Korean Morphological Analysis System)
 - ✓ Dictionary & additional information compiler in C++
 - ◆ Developed TAHANA (Korean Part-of-speech Tagger) using Maximum Entropy model in C++
- **Programmer - October 2001 ~ August 2003**
SearchLine, Inc., Seoul, S. Korea.
 - ◆ Development of a web/enterprise search engine "Condor" with Carnegie Mellon University, Pittsburgh, USA.
 - ✓ Worked at CMU
 - ✓ Main Developer, implemented in C
 - ✓ Co-work with Prof. Jaime Carbonell and Prof. Jaime Callan
 - ◆ Development of supporting applications for Condor
 - ✓ Cache Server
 - ✓ Gateways to File systems/Oracle(OCI)/ODBC
- **Programmer/Technical Manager - February 2000 ~ October 2001**
AidTech, Inc., Seoul, S. Korea.
 - ◆ Maintenance of a search engine "AidSearch"
 - ✓ The engine was deployed at www.empas.com
 - ✓ The engine was deployed at www.naver.com
 - ◆ Modification of the engine to fit enterprise environment
 - ✓ Gateways to data repositories
 - ◆ Development of a document categorizer 'AidCategorizer'
 - ◆ Deployment of the search engine to customer sites
 - ✓ Korea Supreme Court, Auction, Inc., Hankyore Newspaper, etc.
 - ◆ Maintenance of UNIX/Linux servers & PCs
- **Programmer - February 1997 ~ February 2000**
3Soft, Inc., Seoul, S. Korea.
 - ◆ Localization of "Search 97"
 - ✓ "Search 97" of Verity, Inc. USA.
 - ◆ Deployment of the search engine to customer sites
 - ✓ Korea National Assembly, LG-EDS, The Korea Intellectual Property Office, Ministry of Culture and Tourism, Korea Britannica, etc.
 - ◆ Technical Supporter

- **Student Research – August 1995 ~ December 1995**
Korea Advanced Institute of Science and Technology, Daejeon, S. Korea.
 - ◆ The performance comparison of micro processors
- **Computer Languages**
C/C++, Java, Perl, Python, LISP, PHP, Visual Basic, Matlab, SQL, UML, Shell scripts for SH/BASH/CSH

RECENT PUBLICATIONS

- Violetta Cavalli-Sforza, Ralf D. Brown, Jaime G. Carbonell, Peter J. Jansen, and **Jae Dong Kim**. 2004. "Challenges in Using an Example-Based MT System for a Transnational Digital Government Project". In *Proceedings of the Ninth Workshop of the European Association for Machine Translation* (EAMT-04), pp. 33-42. University of Malta, April 26-27, 2004.
- **Jae Dong Kim**, Ralf D. Brown, Peter J. Jansen, and Jaime G. Carbonell. 2005. "Symmetric Probabilistic Alignment for Example-Based Translation". In *Proceedings of the Tenth Workshop of the European Association for Machine Translation* (EAMT-05), pp. 153-159. Pazmany Peter Catholic University, Budapest, Hungary, May 30-31, 2005
- Ralf D. Brown, **Jae Dong Kim**, Peter J. Jansen, and Jaime G. Carbonell. 2005. "Symmetric Probabilistic Alignment". In *Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond - Proceedings of the Workshop*, p. 87-90. Ann Arbor, Michigan, June 29-30, 2005.
- Seung-hyun Seo, In-ho Kang and **Jae Dong Kim**. 2006. "moHANA: Morphological Hangul Analyzer using Multi-Dimensional Analysis Dictionary", *19th Annual Conference on Human and Cognitive Language Technology*, October, 2007.
- **Jae Dong Kim**, and Stephan Vogel. 2007. "Iterative Refinement of Lexicon and Phrasal Alignment". In *Machine Translation Summit XI*, Copenhagen, Denmark, September 10-14, 2007.
- **Jae Dong Kim**, Ralf D. Brown, and Jaime G. Carbonell. 2010. "Chunk-Based EBMT". In *Proceedings of the 14th Workshop of the European Association for Machine Translation* (EAMT-10), to appear. Saint Raphael, France, May 27-28, 2010