Jia Jianfeng

jianfeng.jia@gmail.com (949) 678-9893

JOB OBJECTIVE

Seeking a *Summer Internship* position.

EXPERIENCE

Research Associate	Sogou.com	Jul.2008- Jul.2012
	Beijing, China	

Working at *Chinese Input Method Editor* (IME) Research Group, responsible for improving the precision of IME product using Natural Language Processing approaches based on terabyte dataset.

EDUCATION

Ph.D student,Computer Science	University of California,Irvine	Sept.2012 - Present
	Irvine, CA	

Research interest: *Parallel Computing, Data Mining,Natural Language Processing*

M.S. Computer Science	Xiamen University	Sept.2005 - Jul.2008
	Xiamen,China	

Thesis:*The Application of Dependency Grammar in Statistical Machine Translation*

B.S. Computer Science	Xiamen University	Seps.2001 - Jul.2005
	Xiamen,China	

COMPUTER SKILLS

Programming skills: C++,Python,Java;
Rich experiences on large scale data processing using Hadoop/Hbase/Pig.

PROJECTS

Feedback Data flow System of IME using HBase and Pig   Sogou.com

Building the automatic feedback data processing system using HBase to store 30G data per day, 15T data totally. Using Pig to analyze and explore the global user behavior and also keeping track of single user's daily statistical features.

Large-scale Language Model for Cloud IME   Sogou.com

Building the automatic process of LM updating weekly from the 500G new corpus using Hadoop platform.
Building the decoder for trigram LM and the re-rank model to improve the precision, which is 3% higher than competitors' products.

Automatic New Word Detection   Sogou.com

Devising a novel approach of New Word Detection which is based on Entropy-loss theory on Hadoop platform. This system do not need any Linguistics dictionary, fully base on the statistical information of corpus itself. The LM building on those new words improved 1% precision in desktop IME comparatively to those on normal linguistic words.

Dependency Treelet Based Chinese-to-English SMT System   Xiamen University

Devising two dependency structure based statistical machine translation models.

Shift-Reduced Dependency Parser   Xiamen University

Developing an action sequence based deterministic parser. We achieved dependence arc marker accuracy rate (LAS) 76.36% on Chinese and 82.93% on the English on the benchmark set in CoNLL2007.

HONORS AND AWARDS

Rhinoceroses prize for building the trigram and re-rank model for cloud IME, Sogou Research, 2010
Rhinoceroses prize for innovation on improve 20% precision rate on the long sentence test set of desktop IME, Sogou Research,2009
First-Class Scholarship, Xiamen University, 2005
First-Class Scholarship, Xiamen University, 2004

PUBLICATIONS

Jianfeng Jia, *The Application of Statistical Language Model in Sogou Pinyin Input Method Editor*, Journal of Chinese Association for Artificial Intelligence 2011.vol.1 (4).
Jianfeng Jia, Xiaodong Shi, Xingbang Lai, *HMM-based Chinese Pinyin Input Method*, J. Modern Computer 2008. (4) 4-6.
Xiaodong Shi, Yidong Chen, **Jianfeng Jia**, *Dependency-Based Chinese-English Statistical Machine Translation*, Conference on Intelligent Text Processing and Computational Linguistics (CICLing) 2007, Mexico City, Mexico.
Jianfeng Jia, Xiaodong Shi, Yu Chen, *Shift-Reduce Algorithm and Structure Model Based Dependency Statistical Parser*, International Chinese Computing Conference (ICCC) 2007, Wuhan, China.