Jia Jianfeng

ADDRESS

Donald Bren Hall, Room 2062
University of California,Irvine
Irvine,CA 92967

CONTACT

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

JOB OBJECTIVE

Seeking a *Summer Intern* position on program development.

EXPERIENCE

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

Leading the Desktop *Chinese Input Method Editor* (IME) Research Group, improving the precision of IME product.
In charge of backend service of Sogou Cloud IME

EDUCATION

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

Research interest: *Database, Large Scale Data Processing, Paralleling solution for Natrual Language Processing*

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

Thesis:*The Application of Dependency Grammar in Chinese-to-English Statistical Machine Translation*

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

COMPUTER SKILLS

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

HONORS AND AWARDS

Rhinoceroses price for building the trigram and re-rank model for cloud IME, Sogou Research, 2010
Rhinoceroses price 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.

PATENTS

A Input Method In The Hardware Device, **Jianfeng Jia**, Yanfeng Wang, Yang Zhang, CN102087550A public on Jun.8, 2011

A Method And A System Providing New Words And Hot Words, **Jianfeng Jia**, Yang Zhang, Yanfeng Wang, CN102163198A, public on Aug.24, 2011

SELECTED PROJECTS

Feedback Dataflow System of IME using HBase and Pig   Sogou.com

Building the automatic feedback data processing system on top of HBase. Using Pig to analyze the global user behavior and also keeping track of single user's daily precision.
Through this analyse tool, we found some certain flaw of the product and provided the solution to improve the precision.

Big Language Model for Cloud IME   Sogou.com

Building the automatic process of LM updating weekly from the new corpus using Hadoop platform. Building the decoder for trigram LM and the re-rank process to improve the precision of 1%, which is 3% higher than competitors' products.

Automatic New Word Detection   Sogou.com

Devising a novel aproach of New Word Detection which is based on Entropy-loss theory.
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 Universiy

Devising two dependency structure based statistical machine translation models.
Model 1 was completely lexicalized; we extract the treelet structure in the source language side and the continuous corresponding string of words in the target language side
Model 2 applied the generalization to summarize the learned lexical template. Different from the before systems, we applied grammar labels to constrain the generalized template. And the Bleu score of Model 2 was apparent higher than the Model1 and at the same level with the Pharaoh.

Shift-Reduced Dependency Parser   Xiamen Universiy

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.