Jia, Jianfeng

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

WORK EXPERIENCE

Software Engineer Intern SRCH2.com, Irvine, US Jun.2014 - Sept.2014
Managed the Android SDK develpment group. Implemented a server mode Android Search SDK.

Software Engineer Intern SRCH2.com, Irvine, US Jun.2013 - Sept.2013
Implemented a Java SDK for the C++ search engine library. Implemented a Android Search app.

Research Group Leader Sogou.com, Beijing, China Jul.2008 - Jul.2012
Lead a research group of 5 engineerers to improve the user experience on the Sogou Chinese Input Method product which was used by 300 million people.

EDUCATION

Ph.D student University of California, Irvine Sept.2012 - Present
Research interest: *Large scale data processing, Big data mangement*

M.S. Computer Science Xiamen University, China Sept.2005 - Jul.2008
B.S. Computer Science Xiamen University, China Seps.2001 - Jul.2005

SKILLS

C++, Java, Android, Ruby, Python, Hadoop, Pig.

PROJECTS

Android SDK for SRCH2 C++ library SRCH2.com
Developed the local and the server version of the Android Search SDK which used the SRCH2 C++ search engine library under the hood. We implemented an error-tolerant search app using this SDK.

Graph Computing System Comparison UCI
Designed the benchmark to evaluate the performance of popular Graph Computing Systems. Evaluated the performance number of Preglix, Giraph, GraphLab, GraphX, Hama.

Genome Assembling using Hyracks UCI
Used the Hyracks platform(Parallel data processing system competing with Hadoop) to build the genome graph which contains billions of nodes. And we achieved 30% performance improvement compare to Hadoop solutions.

Feedback data flow System of IME using HBase and Pig Sogou.com
Built a automatic feedback data processing system which used HBase to store 30G data per day, 15T data in total. Used Pig to analyze and explore the global user behavior and we also could keep track of the single user's statistical input patterns.

Large-scale Language Model(LM) for Cloud IME Sogou.com
Built an automatic module for the LM building process which was updated weekly from the 500G corpus.The system was built by Hadoop platform. Built a decoder using trigram LM and the re-rank model, which precision was 3% higher than competitors' products.

Automatic New Word Detection Sogou.com
Developed a New Word Detection system which was implemented on Hadoop platform.

Dependency Treelet Based Chinese-to-English SMT System Xiamen University
Implemented a dependency grammar structure based statistical machine translation system.

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

PUBLICATIONS

Yingyi Bu, Vinayak Borkar, **Jianfeng Jia**, Michael J. Carey, Tyson Condie, *Pregelix: Big(ger) Graph Analytics on A Dataflow Engine*, VLDB 2015.

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.