Jia, Jianfeng

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

EXPERIENCE

Ph.D. student UC Irvine Sep.2012 - Present

Researching on large scale data analytics and visualization platforms. Committer of Cloudberry, Apache AsterixDB, Hyracks, and Pregelix projects.

Software Engineer Intern Sumo Logic, Redwood City,US Jun.2015 - Sept.2015

Designed and implemented the lookup operator for a scalable streaming platform.

Software Engineer Intern SRCH2.com, Irvine, US Jun.2014 - Sept.2014

Led the Android development group.

IME Research Group Sogou.com, Beijing, China Jul.2008 - Jul.2012

Led a research group of 5 engineers to improve the precision of the Sogou Chinese Input Method product which was used by 300 million people.

EDUCATION

Ph.D. candidate University of California, Irvine Sept.2012 - Present

Research topic: *Large scale data analytics and visualization, Big data management*

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

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

SKILLS

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

PROJECTS

Cloudberry system, http://cloudberry.ics.uci.edu UCI

Developed a general purpose middleware system on top of a parallel database system (AsterixDB) to support efficient real-time analytics and visualization on very large data sets with various attributes.

Lookup operator for Streaming data Sumo Logic

Developed a lookup operator that can join the streaming data with the static data in databases.

Big-Object-Aware Memory manager for run-time operators Apache AsterixDB

Designed a big-object-aware memory manager that can adjust the page size based on the incoming record size. Rewrite the run-time operators(sort, group-by, join, etc) using this memory manager to support the Big-Object feature in AsterixDB.

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. 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.

Genome Assembling using Hyracks UCI

Used the Hyracks platform to build the genome graph composed of 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 an automatic feedback data processing system that used HBase to store 15T data. Used Pig to analyze and explore the global user behavior and also keep track of the individual's statistical input patterns.

Large-scale Language Model(LM) for Cloud IME Sogou.com

Built an automatic module for the LM building process updated weekly from the 500G corpus. The system was built on top of 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 top of Hadoop platform.

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

Implemented a dependency grammar structure based statistical machine translation system.

PUBLICATIONS

Towards Interactive Analytics and Visualization on One Billion Tweets,**J.Jia**, C.Li, X.Zhang, C.Li, M.Carey and S.Su, ACM SIGSPATIAL 2016.

Pregelix: Big(ger) Graph Analytics on A Dataflow Engine, Y.Bu, V.Borkar, **J.Jia**, M.Carey, T.Condie, VLDB 2015.

Dependency-Based Chinese-English Statistical Machine Translation, X.Shi, Y.Chen, **J.Jia**, CICLing 2007.