WENHAI ZHU

1302 West 23rd Street, Los Angeles, CA 90007 • (323) 614-5697
wenhaizh@usc.edu • http://wenhaizhu.github.io

EDUCATION

UNIVERSITY OF SOUTHERN CALIFORNIA, Los Angeles, CA 08/2013 – 05/2015
- Master of Science in Computer Science, GPA: 3.76/4.0

XIAMEN UNIVERSITY, Xiamen, Fujian, China
- Master of Engineering in Computer Applied Technology 09/2010 – 07/2013
- Bachelor of Engineering in Computer Science and Technology 09/2006 – 07/2010

TECHNICAL SKILLS

- JAVA, C, PHP, Python, SQL, Matlab, MapReduce, Hadoop, Pig, AWS, Storm
- Web: HTML, CSS, JavaScript, jQuery, Ajax, DOM, JSON, XML, Servlet, Struts, Hibernate
- Platforms & Tools: Windows, Linux, Android, Oracle, MySql, Eclipse, Apache, Tomcat, SVN, SPSS
- Experience with Natural Language Processing, Machine Learning, Statistics, Perl, pThread, OpenMP, MPI

WORK EXPERIENCE

Software Developer Intern @ Pricegrabber, Los Angeles, CA 06/2014 – Present
- Developed 7 new testing tools for bug monitoring, pixel tracking, AJAX information capture, statistics analysis, page loading time, dashboard, and event verification, which saved 30 mins every day for testing.
- Enhanced and maintained 15 PHP and Perl based web applications. Fixed at least 20 bugs.
- Used PHP, Perl, JavaScript, CSS, XML, jQuery, AJAX, MySQL, Linux, VIM, and SVN.

Web Developer Intern @ West Straits Communication Engineering Center, Xiamen, Fujian, China 07/2011 – 08/2011
- Participated in the development of Timecard Manage System using Java, JUnit and SSH2 framework.
- Completed bulletin-board module, including database design, announcement creation, edit, and deletion.

PROJECT EXPERIENCE

Employment Search Engine 09/2014 – 12/2014
- Led a group of 4 people to develop a job search engine based on Apache Tika and Solr.
- Implemented a de-duplication algorithm based on rules and Simhash algorithm in Java, and reduced 98% jobs.
- Implemented the link based ranking algorithm based on PageRank in Java, and used it to rank 261,711 jobs.

Data Science Mini-Project 04/2014 – 12/2014
- **Pig:** Analyzed 500 GB data using Pig/Hadoop on EMR, EC2, and S3 of AWS.
- **Machine Learning:** Analyzed 4 real word datasets using kNN, Naive Bayes, Decision Tree, Logistic Regression, SVM, k-means, and PCA in Matlab.
- **Pick Up the Nickel:** Took videos for 68 hours, collected 1086 people's 11 attributes, and built a model with 3 variables to predict whether people will pick up a nickel when they saw it.
- **Governor Race Prediction:** Conducted data scraping, wrangling, analysis and visualization of governor race with Numpy, Pandas, Matplotlib, Requests, and Pattern library in Python.
- **Click Through Rate Prediction (Kaggle competition):** Reduced the dimension of the train dataset (5.9G) from 22 features to 15 features with PCA and cross-validation. Built a logistic regression model with MapReduce in Matlab, and tested the model on a test dataset (673M).

Stock Information Search 04/2014 – 04/2014
- Built a stock information search website using PHP, Servlet, jQuery, JSON, Ajax, JavaScript, and CSS.
- Deployed PHP file on AWS Elastic Beanstalk and developed an Android application of this website.

Twitter Data Stream Analysis 08/2013 – 12/2013
- Organized a group of 3 people to analyze hot topics in real-time tweets with Storm, Twitter4J, and Java.
- Researched 78 related paper, introduced and designed the analysis process.
- Implemented the data export module and space-saving algorithm.

Research on Chinese Natural Language Interface to Database 09/2011 – 06/2013
- Designed an interactive question-answering system in Java, using anaphora resolution and decision tree.
- Selected 13 attributes and built a semantic network, based on HowNet, to get the attribute values.