4972 Hildasue Terrace
Fremont, CA 94555

Rui Bi

xmruibi.jobs@gmail.com
609-906-0778

EDUCATION BACKGROUND

M.S in Information Science and Technology, Information School **Aug 2013 -- May 2015**

School: University of Pittsburgh, Pittsburgh, PA, United States, GPA 3.85/4.0

Coursework: Web Application Development, Web Service and Distributed System, Information Retrieval, Cloud Computing.

B.S.E in Computer Science and Engineering, School of Computer Science and Software Engineering **Sept 2009 – June 2013**

School: Tianjin Polytechnic University, Tianjin City, Tianjin, China, GPA 88.6/100

Coursework: Algorithm, Data Structure, Software Design, Database Management etc.

Honors: First-Class Scholarship (University level); University Merit Student

PROFESSIONAL EXPERIENCES

Research Assistant, Information Retrieval, Integration and Synthesis Lab in University of Pittsburgh **Sep 2014 – Dec 2014**

Independent Study Project: Query Classifier for Collaborative Search Engine (Java)

- Designed a system for classifying query into its related topic. (<100ms for retrieval time and >83% precision)
- Implemented keywords extraction from searching tasks statements (large text) by using Stanford NLP Parser.
- Designed and implemented a ranking mechanism with Dirichlet Prior Smoothing algorithm and TFIDF VSM Similarity score.
- Designed user profile & action driven pattern to loaded partial data from database for reducing 500% retrieval time

Software Engineer Intern, Yonyou Software Co., Ltd, Xiamen **Jun 2012 – Sep 2012**

Java Developer for Customized UAP-NC ERP Platform

- Worked on Merchandise Search Module by working within a 5-person team.
- Responsible for back-end (Spring MVC, Hibernate) development and front end (JavaScript, jQuery, JSON) development
- Conducted Test-driven development using JUnit and version control via GIT.
- Applied session, privilege control, and SQL injection prevention to the system.

ACADEMIC PROJECTS

Scalable Website Practicum – Innovative Music Search Engine (Java) **May 2015 – Present**

- Responsible for website architecture design and backend services implementation on Linux environment.
- Configured and applied media file storage on AWS S3 and MongoDB clusters on EC2 instances as website database.
- Applied Elastic Search cluster on EC2 instances as data indexing engine and implemented by Spring-Data-ElasticSearch API.
- Designed RESTful web service architecture and implemented by Spring Boot App with Spring REST API.
- Applied Cookie to record user actions for user profile and implemented Item-based Collaborative Filter to recommend music

Stock Recommendation and Price Prediction with Big Data and Machine Learning (Java) **Feb 2015 – Apr 2015**

- Configured Sharding MongoDB Cluster across 15 ports and Hadoop Cluster across three remote Linux VMs.
- Implemented Map-Reduce Stock Crawler based on Hadoop cluster r/w data with MongoDB Connector (>4GB stock dataset).
- Implemented Blizzard MPQ algorithm to hash Stock historical quote object for more efficient and distributed shards load.
- Applied Mahout API to predict possible market trends with binary (Up/Down) forecast (AUC: 0.60).

PITT Course Registration System (Java Web Service) **Oct 2014 – Dec 2014**

- Redesigned current course registration system with SOA by using JAX-WS and JAXB as XML parser for SOAP message.
- Applied JMS Queues to address service request and response and P2P message transmission.
- Applied Message-Driven Bean to listen for request and response with Java Mail service to send email automatically.
- Designed the site front-end by Bootstrap as it client user interface (Course Enrollment module).

Tweets Retrieval, Micro-blog Track, Text REtrieval Conference (TREC), NIST **Aug 2014**

- Designed an Information Retrieval model for retrieving most related tweet by query from 243 million tweets collection.
- Implemented distributed Map-Reduce inverted indexing with Hadoop clusters on three Amazon EC2 instances
- Adopted WordNet API to fetch synonyms for queries expansion and evaluate similar tweets by TFIDF as document expansion.
- Analyzed each tweet sentiment by Stanford NLP package and use the sentimental element as a part of the tweets similarity score.
- Applied BM25 algorithm on VSM for retrieval score improved precision (R-Precision= 0.51, P@30 = 0.65).

Col*fusion (Collaborative Data Fusion) Project **Sep 2013 – Dec 2013**

- Designed a Java RESTful Service between PHP server-side layer and Java back-end to reduce data preview delay.
- Implemented large files (>500MB) uploading with previewing and basic data fusion function on data submission subproject.
- Implemented uploading data onto different databases (MySQL&Neo4j) or Excel files with storing as unified manageable format.

TECHNICAL SKILLS

Programming Languages: Java , JavaScript, HTML/CSS, jQuery, Linux Shell, Python **Cloud Techs:** Hadoop, AWS EC2

Web Development: Hibernate, Spring, Bootstrap.js, JSP, Servlet, JAX-WS, JMS, RESTful, Node.js **Machine Learning:** R

Database: MySQL/MSSQL, Derby, Mongo DB **Information Retrieval:** ElasticSearch, Lucene

Personal Blog: http://xmruibi.github.io Projects code on: github.com/xmruibi