853 Commodore Dr Apt 410,
San Bruno, CA 94066

YAYANG TIAN

www.linkedin.com/in/yayangtian

(215) 450-4186
yayangtian@gmail.com

EDUCATION

University of Pennsylvania Philadelphia, PA
- **Master of Science in Engineering, Computer and Information Science** May 2013
- **Courses**: Analysis of Algorithms, Internet & Web Systems, Machine Learning, Natural Language Processing, Computational Linguistics,
- Data Mining Thesis, Database Systems, Software Engineering, Networked Systems, Network Security, Android Programming, Architecture.

Shanghai Jiao Tong University Shanghai, China
- **Bachelor of Engineering, Information Engineering, Top Honor Class** July 2011
- **Courses**: C & C++ Programming, Data Structure, Algorithms, Operating System, Object-Oriented Design, Database, Networks, Assembly.

SOFTWARE SKILLS

- **Languages**: Java, JavaScript, Python, MySQL, C/C++
- **Web**: NodeJS, ReactJS, Flux, Spring, Play, Django
- **Backend**: Hadoop, MongoDB, HBase, Cassandra, Couchbase, Berkeley DB
- **Mobile:** Android, iOS, React Native, Windows Phone

EMPLOYMENT

Software Engineer III – **Walmart Labs, Sunnyvale, California , CA** Nov 2013 – Present
- Personalized Walmart.com **homepage, category and item pages** by implementing **web services** using Java and Spring.
- Developed NodeJS web apps for Walmart's Savings Catcher which helped resolve **100,000+** customer support tickets.
- Drove **weights feedback** and **analytics dashboards** pipelines for machine learning models of Walmart's personalization engine.

Software Engineer – Microsoft, Boston, MA Sept 2013 – Nov 2013
- Implemented a Windows Phone application in C# .NET that could predict user habits based on motion statistics from Nokia sensors.
- Gathered large amount of sensor data, ran Hadoop MapReduce in Amazon Web Service and built machine learning models.

Research Assistant – Natural Language Processing Research, University of Pennsylvania June 2012 – May 2013
- Performed **sentiment analysis** and opinion mining on Google NGram, and New York Times which increased prediction accuracy by 6.6%.
- Developed a web app summarizing readers' emotion towards Wall Street Journal which helped win EMNLP12 Google Best Paper Award.

SIDE PROJECTS

Mini Airbnb (React, NodeJS, MongoDB, ElasticSearch) Spring 2015
- Built a search engine, NoSQL database, and a rental website and scrapped and indexed millions of data into ElasticSearch.

WalmartLabs Hackathon (React, NodeJS, MongoDB, ElasticSearch) Spring 2015
- Developed a web platform for customers and buyers for suggesting missing inventory items using Facebook's React framework.

PROJECTS @ UPENN **github.com/upennyayang**

Distributed Web Search Engine – MiniGoogle (Java, AWS, Hadoop, Berkeley DB) Spring 2013
- Developed a **cloud computing** search engine running **Hadoop MapReduce** on Amazon EC2 consisting of crawler, indexer, PageRank, UI.
- Built a load-balanced **web crawler** over FreePastry **distributed hash tables** that gathered **430,000** pages in 4 hours in Amazon S3.
- Built a **TF-IDF indexer** for **information retrieval** and a **Google PageRank** engine for link analysis on Amazon Elastic MapReduce.
- Improved **search relevance**, using **machine learning** to weight and rank 10 features such as feedback, proximity, bigrams, and metadata.
- Built a **front-end** for relevant results**,** with **spell suggestion** and **recommendations** of Yelp, YouTube, Flickr, Twitter, Amazon and GMap.

Distributed P2P Search Engine – PennSearch (C++, NS-3) Spring 2012
- Developed a **DHT** based search engine over implementation of **Chord distributed hash table** with high availability for nodes failure.
- Implemented distant vector **routing protocols,** keyword based **information retrieval** and **multicast** algorithm. (20,000 lines of code).

Distributed YouTube Caching System (Java, REST) Spring 2013
- Built a **decentralized** caching system**,** storing YouTube search results over FreePastry **distributed hash tables** and key-based routing.
- Built a **RESTFul web services** based web interface that queried, received and parsed **JSON** messages between client and caching system.

Distributed Web Crawler with XPath Engine (Java, Berkeley DB) Spring 2013
- Built a RSS aggregator and an XPath Engine, which traversed the web and looked for **XML** documents matching topics defined by XPaths.
- Built a servlet **web interface**, allowing users to subscribe RSS, manage topics they like, and display XML (**XSLT**) stored in **Berkeley DB**.

Scalable Multithreaded Web Server – MiniTomcat (Java, Servlet) Spring 2013
- Created a **thread pool** based **HTTP** server from scratch that could process various requests, run Java servlets and render dynamic pages.
- Tested on ApacheBench and handled 50,000 requests with 1000 requests **concurrently** for HTML, CSS, images, cookies and sessions.

Amazon Reviews Data Mining (Python, Matlab) Fall 2011
- Developed a rating **prediction system,** trained from **100,000** reviews on Amazon using **machine learning** and got **top performance**.
- Implemented PCA, Naïve Bayes, boosting, kernels that increased accuracy from **40.1%** to **81.3%** and dropped RMSE from **1.460** to **0.853**.

Twitter Sentiment Analysis and Opinion Mining (Python, Django, SQLite) Spring 2013
- Innovated a **tweets prediction** system based on **42,400** tweets we crawled, including a six-class SVM model and a Django web front-end.
- Compared various **machine learning** methods and beat baseline by **21.197%** using features like emoticons, smileys, WordNet, N-grams.

The New York Times Document Summarization (Python, NLTK) Fall 2012, Fall 2011
- Performed **natural language processing** and **data mining** on the New York Times, and evaluated using ROUGE-1 in DUC2004 dataset.
- Implemented an automatic multi-document summarizer**,** using LexRank and MMR based **sentence selection**, **clustering** and **ordering**.