

EDUCATION

University of Pennsylvania

Philadelphia, PA

- **Master of Science in Engineering, Computer and Information Science**, GPA: 3.70/4.00 May 2013
- **Courses:** Analysis of Algorithms, Internet & Web Systems, Database Systems, Networked Systems, Machine Learning, Data Mining Thesis,
- Software Engineering, Natural Language Processing, Advanced NLP, Android Programming, Network Security, Computer Architecture.

Shanghai Jiao Tong University

Shanghai, China

- **Bachelor of Engineering, Information Science**, GPA 3.62/4.00 (top 10%) July 2011
- **Courses:** C & C++ Programming, Data Structure, Algorithms, Operating System, Object-Oriented Design, Database, Networks, Assembly.

SOFTWARE SKILLS

- **Programming Languages:** Java, Python, C++, C, SQL, PHP, HTML, CSS, JavaScript/jQuery, XQuery
- **Operating Systems:** Linux/Unix
- **Technologies:** Hadoop, MapReduce, REST/SOAP, Amazon Web Service(EC2, S3, EMR), Berkeley DB, AJAX, JSP, Servlet, MVC, JUnit, GIT

EMPLOYMENT

Research Assistant – Natural Language Processing Research, University of Pennsylvania

Mar 2012 - Present

- Performed **sentiment analysis** and **opinion mining** in **Python** on Google, New York Times, Switchboard, and Penn Discourse Treebank.
- Introduced an **emotion classification** method by integrating **ranking** and **machine learning** which got 6.6% improvement (NSF funded).

Software Developer – Computer Science Department, University of Pennsylvania

July – Oct 2012

- Built a website in **PHP** and MySQL mining readers' interest on Wall Street Journal which helped win EMNLP **Google Best Paper Award**.
- Developed a web app in **Java** Servlet summarizing users' emotional orientation from context of news headlines on Google News and CNN.

Software Engineer Intern – Cisco Systems, Inc., Shanghai

June - Aug 2010

- Provided **Cisco digital video** solution in Shanghai China by developing **Java** value-added products on **set-top box** multi-media terminals.
- Implemented video on demand and digital video recorder for interactive services, developed under MIDP 2.0 and deployed in Shanghai.

PROJECTS @ UPENN

github.com/upennyayang

Distributed Web Search Engine – MiniGoogle (Java, Hadoop, Berkeley DB)

Spring 2013

- Collaborated and built a **cloud computing** search engine running **Hadoop MapReduce** on Amazon EC2(crawler, indexer, PageRank, UI).
- Took in charge of **scoring**, **weighting** and **ranking**, TF-IDF **information retrieval**, **search relevance**, **web services**, and **web interface**.
- Presented a ranking algorithm comprised of ten features, with online learning and SVM based **relevance feedback** to re-rank results.
- Implemented features for spell suggestion, proximity, summary, preview, Yahoo, Amazon, Yelp, YouTube, Twitter, and Flickr integration.

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 XPath.
- 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 **load balanced** web server like Tomcat based on thread pool, which could run Java servlets and render dynamic web 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, Matplotlib)

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 classification** system based on **42,400** tweets we crawled, including a six-class **SVM** model and a web front-end.
- Outperformed baseline by **21.197%** using features like emoticons, smileys, WordNet, 8000 unigrams and 16000 bigrams with highest IG.

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**.

Database Web Application - GameMonopoly (HTML/CSS, Java, MySQL, jQuery)

Fall 2012

- Innovated a **multiplayer game** in which 4 players throw dice and buy lands to get rich, based on huge XML dataset from 2012 Olympics.
- Designed and implemented the **website** and game logic, the communication/synchronization using AJAX, and the animation using jQuery.

PUBLICATION & HONORS

- [1] Yayang Tian, Ani Nenkova, B-SWB: An Emotion Classifier Based on Unsupervised Binomial Ranking Method, to SemEval-2013
- [2] Peter Febernek, Yayang Tian, Ani Nenkova, Clustering and Ordering of Sentence for News Summarization, to NACCL, HLT-2013
- 1st Prize, Alibaba Internet **Cloud Computing** Finals, 2010
- 1st Prize, Chinese **Physics Olympiad** (CPO), Ranked **3/426673**
- Academic Scholarship (top **10%**), Outstanding Graduate
- 12th International Band Music Competition Outstanding Performance Award