3900 Chestnut Street, Apt 301
Philadelphia, PA 19104

YAYANG TIAN

cis.upenn.edu/~yaytian

(215) 450-4186
yaytian@cis.upenn.edu

EDUCATION

University of Pennsylvania — Philadelphia, PA
- **Master of Science in Engineering, Computer and Information Science**, GPA: 3.55/4.00 — May 2013
- Courses: Algorithms, Internet and Web Systems, Machine Learning, Natural Language Processing, Database Systems, Data Mining Thesis, Computational Linguistics, Software Engineering, Android Programming, Networked Systems, Computer Security, Computer Architecture.

Shanghai Jiao Tong University — Shanghai, China
- **Bachelor of Engineering, Information Science,** GPA 3.52/4.00 (Ranked **top 5%** of EECS Dept.) — July 2011
- Courses: Programming, C++, Data structure, Algorithms, Operating System, Database, Object-Oriented Design, Network, Graph Theory.

SOFTWARE SKILLS

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

EMPLOYMENT

Web Developer – Institute for Research in Cognitive Science, University of Pennsylvania — July – Oct 2012
- Built a **website** (PHP, MySQL) mining readers' interest on Wall Street Journal which helped win EMNLP 2012 **Google Best Paper Award**.
- Developed a **web app** (Java Servlet) summarizing users' emotional orientation from context of news headlines on Google News, and CNN.

Summer Research Intern – NLP@Penn Research Group, University of Pennsylvania — May - Sept 2012
- Performed **sentiment analysis** and opinion mining in Python on Google, New York Times, Switchboard, and Penn Discourse Treebank.
- Introduced a novel **emotion recognition** approach by integrating **ranking** and **machine learning**, which got 6.6% improvement (**NSF**).

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.

Software Developer – Intel Labs SJTU, Shanghai — June - Sept 2009
- Developed a visual **database management system** based on Borland **C++** Builder VCL framework using **object-oriented** programming.
- Managed **software lifecycle** including requirements analysis, database design, implementation, testing, deployment and maintenance.

PROJECTS @ UPENN – github.com/upennyayang

MiniGoogle - Cloud Computing Search Engine (Java, Hadoop) — Spring 2013
- Built a **large-scale distributed** search engine, computing **Hadoop MapReduce** on **Amazon EC2** including crawler, indexer, PageRank, UI.
- Developed a **load-balanced crawler** over **FreePastry DHT** that gathered **430,000 web pages** in 4 hours and stored in **Amazon S3**.
- Developed a **TF-IDF indexer** for **information retrieval** and a **Google PageRank** engine for link analysis on **Amazon Elastic MapReduce**.
- Built a **front-end,** integrating MiniGoogle results with Yahoo, Amazon, Yelp, YouTube, Twitter, Flickr, EBay, Wiki using **RESTful APIs**.
- **Weighted ten ranking scores** to improved **search relevancy**. Incorporated **proximity**, **summary, feedback, and spell suggestion.**

Twitter Sentiment Analysis and Opinion Mining (Python, Django, SQLite) — Spring 2013
- Innovated **a tweets classification system,** including a **six-class SVM model** and a **web front-end,** based on **42,400** tweets we **crawled**.
- Outperformed baseline by **21.197%** using features like emoticons, smileys, WordNet, 8000 unigrams and 16000 bigrams with highest **IG**.

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**.

YouTube P2P Caching System and RESTful Web Services (Java, Servlet) — Spring 2013
- Built a decentralized **caching system,** storing YouTube search results over FreePastry **key-based routing** and **distributed hash tables.**
- Built a **web services** based web application that queried and received REST(**JSON**)/SOAP messages between client and caching system.

The New York Times Summarization (Python, NLTK) — Fall 2011, Fall 2012
- 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**.

PennSearch - Distributed Search Engine (C++, NS-3) — Spring 2012
- Implemented a **peer-to-peer** search engine over implementation of **Chord distributed hash table** with high **availability** of nodes failure.
- Implemented **information retrieval** algorithm to distributed and search inverted index for keyword and URLs across 6 ENIAC machines.

Scalable Web and Application Server (Java, Servlet) — Spring 2013
- Created a high **load-balanced** and **availability** HTTP server like Tomcat, capable of running 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.

RSS Aggregator with XPath Engine (Java, Berkeley DB) — Spring 2013
- Built a RSS aggregator and an **XPath Engine**, which traverses the web and looks 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**.

GameMonopoly - Web Application (Java, MySQL, jQuery) — Fall 2012
- Innovated a **multiplayer game** on Google App Engine consisting of a fancy **web UI**, multithreaded server, AJAX interaction, and **database**.
- Utilized **XQuery** to parse 2012 Olympics XML files into database and performed **indexing** and **query optimization** to enhance efficiency.

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
- Outstanding Performance, **International Band Music** Competition