

## Frank F. Xu

---

CONTACT INFORMATION	SEIEE Bldg. 3-341 Department of Computer Science & Engineering Shanghai Jiao Tong University 800 Dongchuan Road Shanghai 200240, China	Phone: (+86) 188-1827-5852 E-mail: <a href="mailto:frankxu@sjtu.edu.cn">frankxu@sjtu.edu.cn</a> URL: <a href="http://frankxfz.me">frankxfz.me</a> Skype: <a href="https://www.skype.com/user/frankxu2004">fzxu2004</a> Github: <a href="https://github.com/frankxu2004">github.com/frankxu2004</a>
RESEARCH INTERESTS	Natural Language Processing, Text Mining, Knowledge Discovery, Computational Social Science, Urban Computing, Machine Learning, Artificial Intelligence	
EDUCATION	<b>Shanghai Jiao Tong University</b> , Shanghai, China M.S. Candidate, Computer Science, September 2016 (expected graduation date: May 2019) <ul style="list-style-type: none"><li>• Research Field: Natural language processing, text mining, knowledge discovery, spatial-temporal data mining</li><li>• Advisor: Prof. Kenny Q. Zhu</li></ul> B.Eng., Computer Science (IEEE Honored Class), June 2016 <ul style="list-style-type: none"><li>• An elite program at SJTU which aims to nurture scientists in computer science, electrical and electronic technology, and information science based on MITs educational model.</li><li>• Thesis Topic: “Traffic Prediction for Urban Planning”</li><li>• Advisor: Prof. Kenny Q. Zhu</li></ul>	
HONORS AND AWARDS	Best Thesis Award in Bachelor’s Thesis Defense (Top 1%), Shanghai Jiao Tong University, 2016 Second Place in EMC Smart Campus Open Data Contest, Shanghai Jiao Tong University, 2015	
ACADEMIC EXPERIENCE	<b>Shanghai Jiao Tong University</b> , Shanghai, China <i>Graduate Student and Research Assistant at ADAPT Lab</i> <b>September, 2016 - present</b> <a href="http://adapt.seiee.sjtu.edu.cn/">http://adapt.seiee.sjtu.edu.cn/</a> Working at ADAPT Lab, led by Prof. Kenny Q. Zhu, collaborate with Bill Y. Lin, work includes current Master research, coursework and research or developing projects.  <i>Teaching Assistant</i> <b>September, 2016 - January, 2017</b> Various duties on undergraduate level course for Computer Science (IEEE Honored Class). Shared responsibility for tutorials, exams, homework assignments, and grades. This is an English-only class. <ul style="list-style-type: none"><li>• SE305 Database System Technology, Fall 2016.</li></ul> <i>Research Intern at SJTU IIOT Lab</i> <b>January, 2015 - July, 2015</b> <a href="http://iiot.sjtu.edu.cn">http://iiot.sjtu.edu.cn</a> Worked at SJTU IIOT Lab, led by Prof. Xinbing Wang. Joining a team on AceMap: Academic Search and Paper Recommendation System and mainly contributed to academic crawler, paper topic model, search algorithms and infrastructure.  <i>Research Intern at OMNILAB</i> <b>June, 2014 - December, 2014</b> <a href="http://omnilab.sjtu.edu.cn">http://omnilab.sjtu.edu.cn</a> Worked at OMNILAB, led by Prof. Yaohui Jin, located at Network and Information Center in SJTU. Contributed to open data platform and data mining projects from web crawlers to data analysis and visualizations.	

**Automatic Extraction of Commonsense LocatedNear Relationship from Text**

LocatedNear relation describes two typically co-located objects, which is a useful commonsense knowledge. We propose to automatically extract such relation through a sentence level classification problem and the aggregation of instances detected from large amount of sentences. To enable research of these tasks, we release two benchmark datasets, one containing 5,000 annotated sentences extracted from the Gutenberg corpus; the other containing 500 pairs of physical objects and whether they are commonly located nearby. We also propose some baseline methods for the tasks and compare the results with a state-of-the-art general purpose relation classifier. We have submitted this work to EMNLP 2017

**Socio-linguistic Bilingual Word Representations for Computing Cross-Cultural Differences**

Capturing cross-cultural differences between terms is an important challenge in bilingual text understanding and machine translation. This paper presents a novel framework for obtaining bilingual word representations from social media by leveraging socio-linguistic vocabularies. Such representations can act as a building block for cross-lingual and cross cultural studies in computational social science. We evaluated our framework on two such tasks: detection of cross-cultural differences in named entities and bilingual lexicon induction for Internet slang. Experimental results show that methods based on our proposed word representations outperform a number of baseline methods by substantial margins. We have submitted this work to IJCAI 2017.

**Cross-region Traffic Prediction with Transfer Learning**

This is an interdisciplinary study of computer science, transportation and urban planning. We built a system to learn a prediction model from graphical traffic condition data, and thus everyone can predict the traffic conditions with nearly 90% accuracy in Shanghai, China, even if no historical traffic data is available for that area. This novel system is useful in urban planning, transportation dispatching as well as personal travel planning. We have published this work in 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science. We are still improving it with many innovative ideas. We propose a framework to predict traffic flow on road segments with no historical traffic data by transferring the knowledge from existing traffic data in probably another city. This work is still under progress and will be submitted to ICDM 2017.

**AceMap: Academic Search and Paper Recommendation System**

<http://acemap.sjtu.edu.cn>

A team work led by Prof. Xinbing Wang. Our goal is to build an academic search engine which can return paper search results based on topic similarity with user's query, analyze the latent topic distribution and topic development over time, visualize the "topic tree" starting from a particular paper and more. It is like an academic social network with features like search, academic maps, paper network, etc. We are currently implementing machine learning like LDA and HLDA, NLP algorithms and network analysis algorithms for processing and analyzing paper database which is crawled from various sources, and expose the search feature with the help of the robust Apache Solr with many customizations. I have also been working on Web development, data visualization with D3.js, some Apache Solr configuration, and paper citation analysis of conference datasets and clustering algorithms for Academic Map feature.

**Automatic Extraction of Commonsense LocatedNear Relationship from Text.**

**Frank F. Xu\***, Bill Y. Lin\*, Kenny Q. Zhu.

*To be submitted to Conference on Empirical Methods in Natural Language Processing, 2017 (\* indicates equal contribution)*

**Bilingual Word Representations with Cross-cultural Socio-linguistic Features.**

**Frank F. Xu\***, Bill Y. Lin\*, Hanyuan Shi, Kenny Q. Zhu and Seung-won Hwang.

*Under review of Twenty-sixth International Joint Conference on Artificial Intelligence, 2017 (\* indicates equal contribution)*

**Mining Cultural Differences of Named Entities from Large Text.**

Hanyuan Shi, Kenny Q. Zhu, Bill Y. Lin, **Frank F. Xu** and Seung-won Hwang.

*Submitted to Thirty-First AAAI Conference on Artificial Intelligence, 2017*

**Cross-region Traffic Prediction for China on OpenStreetMap.**

**Frank F. Xu**, Bill Y. Lin, Qi Lu, Yifei Huang and Kenny Q. Zhu.

*In the Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science, 2016*

**Traffic Prediction for Urban Planning**

**Frank F. Xu**

*Bachelor's Thesis (Best Thesis Award, Top 1%), 2016*

INDUSTRY  
EXPERIENCE

**Baixing.com**, Shanghai, China

*Software Developer Intern*

**August, 2015 - May, 2016**

Developed open APIs, crawlers, backend tools, and some modules to prevent spam and inappropriate posts, like posts filtering and classification, image identification. Mainly used Python 3, Tornado, MongoDB and Redis.

PROGRAMMING  
PROJECTS

**EMC Smart Campus Open Data Contest**

<https://github.com/frankxu2004/emc-contest-data-visualization>

Our team, who mined and analyzed the data of wireless network in the SJTU campus, came the second in this contest. We implemented features like "Are You Really Studying", finding friends based on interests, flow and crowd mapping and statistics. Cluster analysis, trajectory analysis and many other types of machine learning and data mining algorithms like SVM, random forest are used. I was mainly responsible for data processing, statistics and data visualization. The web platform is built using Python Flask and SQLite as backend Restful API and AngularJS and Highcharts as frontend Web App.

**OMNILab Open Data Platform**

<http://data.sjtu.edu.cn>

<https://github.com/frankxu2004/AQICrawler>

An open data platform based on the open-source CKAN data platform, is built by our lab OMNILab in Shanghai Jiaotong University led by Professor Yaohui Jin of Network and Information Center. We made many modifications to the original CKAN project, from backend data processing API and talking to PostgreSQL, to front-end data visualization, geo-information and search. I myself wrote a crawler named AQICrawler for air quality data from various sources like *pm25.in*, *aqicn.org* and government site and upload the data to our open data platform through APIs.

**DivineMove**

<https://github.com/frankxu2004/DivineMove>

An AI for Go Game team project, written in C with an open protocol for Computer Go, GTP, and using effective, proven algorithms like Monte-Carlo simulation and UCT tree search. We made several optimizations and modifications to the original algorithms presented in papers such as cutting branches in UCT tree, reusing tree structure, "filling water" method for fast observation of game situation and better Monte-Carlo simulations.

**PKULaw Crawler**

<https://github.com/frankxu2004/pkulaw>

It is a crawler for downloading law cases and texts on pkulaw.cn, a widely used online law document database in China, I did this because people doing text mining in legal related fields could find useful while collecting large corpus of law texts.

### **WeChatCrawler**

<https://github.com/frankxu2004/WeixinCrawler>

A WeChat public account crawler written in Python. It fetches articles of interested public accounts from *weixin.sogou.com* and stores them into MySQL database. The API for generating RSS feed for the articles and a simple page for adding new accounts are written in PHP.

### **LyricsSearch**

<https://github.com/frankxu2004/lyrics-search>

An music search engine by lyrics with PyLucene and web.py with some level of fuzzy query. The data is crawled with a Python crawler from Baidu MP3 for songs and lyrics and then processed and indexed by Lucene for search.

### **Feedy**

<https://github.com/SJTUCat/Feedy>

A Chrome extension developed by me and Easton Wang. It is a feed notifier and reader for custom RSS or Atom feeds, using Google Feed API.

- COMPUTER SKILLS
- Machine Learning and Deep Learning Packages: Scipy, scikit-learn, TensorFlow, Theano, Keras
  - Natural Language Processing Tools: Stanford CoreNLP, NLTK
  - Distributed Platforms: Hadoop, Spark
  - Databases: MySQL, PostgreSQL, neo4j
  - Languages: Python, Java, C/C++, PHP, Javascript/Node.js, HTML5, some use of Linux shell scripts, CUDA/MPI parallel processing library
  - Applications: GNUPlot, L<sup>A</sup>T<sub>E</sub>X, common spreadsheet and presentation software
  - Operating Systems: Unix/Linux, Windows