

Frank Lin

lin.frank@gmail.com · 1 412 223 7789 · San Francisco, CA, USA · www.cs.cmu.edu/~frank

- Education**
- Carnegie Mellon University**, Language Technologies Institute, School of Computer Science
Ph.D. in Language Technologies, August 2012 (QPA 3.97)
Thesis: *Scalable Methods for Graph-Based Unsupervised and Semi-Supervised Learning*
- Carnegie Mellon University**, Language Technologies Institute, School of Computer Science
Masters in Language Technologies, May 2005 (QPA 3.92)
- University of Arizona**, Computer Science Department
Bachelor of Science, August 1999 (GPA 3.70 Major GPA 3.92)
- Work Experience**
- Co-founder + CTO** July 2013 - Present
Enfind, Inc., San Mateo, CA
- enfind.com: relevant, unintrusive information lookup and ads right in your webpage
 - Front-end contextual knowledge/ad/link widget and back-end machine learning algorithms and service
- Data Scientist + Software Engineer** October 2012 - June 2013
Twitter, Inc., San Francisco, CA
- Project: user modeling, topic modeling, and analysis tools for large social networks
- Research Assistant** October 2007 - August 2012
Carnegie Mellon University, Pittsburgh, PA
- Research: efficient graph-based semi-supervised learning and clustering methods
 - Maintained MinorThird - an information extraction and text classification software package
- Course Instructor** January 2006 - June 2006
Carnegie Mellon University Qatar, Doha, Qatar
- Developed lectures, homework assignments, and other course materials for Introductory Programming
- Teaching Assistant** September 2005 - December 2005
Carnegie Mellon University Qatar, Doha, Qatar
- Teaching assistant for Introductory Programming and Fundamental Data Structures and Algorithms
- Research Assistant** September 2003 - September 2007
Carnegie Mellon University, Pittsburgh, PA
- Research: cross-lingual information retrieval, extraction, and question answering
 - Developer for JAVELIN - a multilingual question-answering system
- Book Translator** August 2002 - August 2003
Living Stream Ministry, Anaheim, CA
- Translated 5 books from Chinese to English, published by Living Stream Ministry
- Summer Jobs & Internships**
- Intern** June 2011 - September 2011
Twitter, Inc., San Francisco, CA
- Built a system for fast real-time clustering of personalized social networks.
 - Built fast interactive visualizations of personalized social networks.
- Intern** June 2009 - September 2009
Alibaba Clouding Computing, Alibaba Group, Hangzhou, China
- Analyzed historic and daily transactions data from taobao.com and alibaba.com.
 - Built large-scale analysis and visualization tools using MapReduce-based systems.
- Research Assistant** June 1999 - September 1999
Computer Science Department, University of Arizona, Tucson, AZ
- Research programmer for the Active Networks project.
- Software Developer** June 1998 - September 1998
Astronomy Department, University of Arizona, Tucson, AZ
- Developed web-based educational software for the University of Arizona Planetarium.

Selected Publications

Scalable Methods for Graph-Based Unsupervised and Semi-Supervised Learning

[Frank Lin](#). **PhD Thesis**, Pittsburgh, PA, USA.

A General and Scalable Approach to Mixed Membership Clustering

[Frank Lin](#) and William W. Cohen. **ICDM 2012**, Brussels, Belgium.

Adaptation of Graph-Based Semi-Supervised Methods to Large-Scale Text Data

[Frank Lin](#) and William W. Cohen. **KDD 2011 Workshop**, San Diego, California, USA.

Node Clustering in Graphs: An Empirical Study

Ramnath Balasubramanian, [Frank Lin](#) and William W. Cohen. **NIPS 2010 Workshop**, Vancouver, B.C., Canada.

Personalized Email Prioritization Based on Content and Social Network Analysis

Yiming Yang, Shinjae Yoo, [Frank Lin](#) and Il-Chul Moon. **IEEE Intelligent Systems: Special Issue on Social Learning**, Vol. 25(4), pp 12-18, July/August 2010.

A Very Fast Method for Clustering Big Text Datasets

[Frank Lin](#) and William W. Cohen. **ECAI 2010**, Lisbon, Portugal.

Semi-Supervised Classification of Network Data Using Very Few Labels

[Frank Lin](#) and William W. Cohen. **ASONAM 2010**, Odense, Denmark.

Power Iteration Clustering

[Frank Lin](#) and William W. Cohen. **ICML 2010**, Haifa, Israel.

Mining Social Networks for Personalized Email Prioritization

Shinjae Yoo, Yiming Yang, [Frank Lin](#) and Il-Chul Moon. **KDD 2009**, Paris, France.

From Episodes to Sagas: Understanding the News by Identifying Temporally Related Story Sequences

Ramnath Balasubramanian, [Frank Lin](#), William W. Cohen, Matthew Hurst and Noah A. Smith. **ICWSM 2009 (Poster)**, San Jose, California, USA.

The MultiRank Bootstrap Algorithm: Semi-Supervised Political Blog Classification and Ranking Using Semi-Supervised Link Classification

[Frank Lin](#) and William W. Cohen. **ICWSM 2008 (Poster)**, Seattle, Washington, USA.

JAVELIN III: Cross-Lingual Question Answering from Japanese and Chinese Documents

Teruko Mitamura, [Frank Lin](#), Hideki Shima, Mengqiu Wang, Jeongwoo Ko, Justin Betteridge, Matthew Bilotti, Andrew Schlaikjer and Eric Nyberg. **NTICIR-6**, 2007, Tokyo, Japan.

Keyword Translation Accuracy and Cross-Lingual Question Answering in Chinese and Japanese

Teruko Mitamura, Mengqiu Wang, Hideki Shima, [Frank Lin](#). **EACL 2006 Workshop on MLQA**, Trento, Italy.

Modular Approach to Error Analysis and Evaluation for Multilingual Question Answering

Hideki Shima, Mengqiu Wang, [Frank Lin](#) and Teruko Mitamura. **LREC 2006**, Genoa, Italy.

JAVELIN I and II Systems at TREC 2005

Eric Nyberg, Robert Frederking, Teruko Mitamura, Matthew Bilotti, Kerry Hannan, Laurie Hiyakumoto, Jeongwoo Ko, [Frank Lin](#), Lucian Lita, Vasco Pedro, and Andrew Schlaikjer. **TREC 2005**.

CMU JAVELIN System for NTCIR5 CLQA1

[Frank Lin](#), Hideki Shima, Mengqiu Wang, Teruko Mitamura. **NTCIR-5**, 2005, Tokyo, Japan.

Keyword Translation from English to Chinese for Multilingual QA

[Frank Lin](#) and Teruko Mitamura. **AMTA 2004**, Washington D.C., USA.

Relevant Coursework

Information Extraction

Information Retrieval

Machine Translation

Natural Language Processing

Software Engineering for Information Technology

Language and Statistics

Speech II: Phonetics, Prosody, Perception, Synthesis

Grammar and Lexicon

Human Language Technologies

Algorithms

Principles of Operating Systems

Principles of Computer Networking

Parallel and Distributed Programming

Comparative Programming Languages

Foundations of Computer Programming

Object Oriented Programming and Design

Data Structures and Algorithms

Minds, Brains, and Computers

Discrete Mathematics

Languages English, Mandarin; Basic, C, C++, Java, JavaScript, JSP, HTML, Lisp, MATLAB, MIPS ASM, Pig, Ruby, Scala

Updated 2014.01.30. References available upon request.