Fangzheng (Frank) Xu

5568 Forbes Ave, Pittsburgh, PA, USA 15217

(+1) 412-482-8915 | fangzhex@cs.cmu.edu | frankxfz.me

Research Interests

Natural language processing, code generation, data mining, human-computer interaction.

Education

Language Technology Institute, Carnegie Mellon University   *Pittsburgh, Pennsylvania*

Ph.D. in Computer Science   *Sep. 2019 - Present*

- Advisor: Prof. Graham Neubig
- Research Interest: Computational systems that could extract, understand, and reason with procedural knowledge expressed in natural language. Applications include semantic parsing and natural language to code generation.

Shanghai Jiao Tong University   *Shanghai, China*

M.Sc. in Computer Science   *Sep. 2016 - Mar. 2019*

- Thesis: *Indirect Supervision for Information Extraction*
- Research Interest: Natural language processing, text mining, knowledge graph, spatial-temporal data mining
- Advisor: Prof. Kenny Q. Zhu

Shanghai Jiao Tong University   *Shanghai, China*

B.Eng. in Computer Science, Honor Degree (IEEE Honored Class)   *Sep. 2012 - Jun. 2016*

- Thesis: *Traffic Prediction for Urban Planning*
- **Best Bachelor's Thesis Award (Top 1% in SJTU & Top 1 in the CSE Department)**
- Advisor: Prof. Kenny Q. Zhu

Honors & Awards

2018	Best Poster Award Honorable Mention, WWW	*Lyon, France*
2017	2nd in emerging and rare entity recognition shared task, WNUT@EMNLP	*Copenhagen*
2016-18	SJTU Graduate Student Scholarship	*SJTU*
2016	Best Bachelor's Thesis Award (Top 1% in SJTU & Top 1 in the CSE Department)	*SJTU*

Academic & Professional Experience

NeuLab, Carnegie Mellon University   *Pittsburgh, PA*

Research Assistant   *Aug. 2019 - Present*

- Advisor: Prof. Graham Neubig
- Working on natural language processing, procedural knowledge, code generation.

Microsoft Research Asia   *Beijing, China*

Research Intern   *May. 2019 - Aug. 2019*

- Mentor: Dr. Lei Ji, Dr. Nan Duan
- Worked on structured multimodal procedural knowledge acquisition and instruction summarization from narrative videos and texts.

HCII, Carnegie Mellon University   *Pittsburgh, PA*

Visitor   *Nov. 2018 - May. 2019*

- Advisor: Prof. Jeffrey P. Bigham
- Worked on AI and human collaboration, enabling human by incorporating multimodal ML techniques, including NLP and CV into daily life. Help blind people perceive the world, construct knowledge and understand natural language queries with state-action graphs of user interface, help memorize and organize events from users' daily conversations, etc.

ADAPT Lab, Shanghai Jiao Tong University *Shanghai, China*

RESEARCH ASSISTANT *Oct. 2015 - Mar. 2019*

- Advisor: Prof. Kenny Q. Zhu
- Working on natural language processing, computational social science, information extraction, text mining, with papers published in various venues. I used to work on data mining, traffic prediction and urban computing as well.

Baixing.com *Shanghai, China*

SOFTWARE DEVELOPER INTERN *Aug. 2015 - May. 2016*

- Developed open APIs, crawlers, backend code with mainly Python and PHP.
- Designed ML algorithm modules to prevent spam and inappropriate posts, e.g. posts filtering and classification, image identification.

OMNILab, Shanghai Jiao Tong University *Shanghai, China*

UNDERGRADUATE RESEARCH ASSISTANT *Jun. 2014 - Dec. 2014*

- Advisor: Prof. Yaohui Jin
- Worked on open data platform for urban air quality, human mobility and social media, ranging from web crawlers to data analysis and visualizations.

Teaching Experience

2016   **Teaching Assistant,** Database System Technology (SE305), in English *SJTU*

Publication

* indicates equal contribution

CONFERENCE PAPERS

In-IDE Code Generation from Natural Language: Promise and Challenges
Frank F. Xu, Bogdan Vasilescu, Graham Neubig.
arXiv: 2101.11149

Learning Structural Edits via Incremental Tree Transformations
Ziyu Yao, Frank F. Xu, Pengcheng Yin, Huan Sun, Graham Neubig.
In International Conference on Learning Representations (ICLR), 2021

Data-to-Text Generation with Style Imitation
Shuai Lin, Wentao Wang, Zichao Yang, Xiaodan Liang, Frank F. Xu, Eric Xing, Zhiting Hu.
In Findings of the Association for Computational Linguistics: EMNLP, 2020

A Benchmark for Structured Procedural Knowledge Extraction from Cooking Videos
Frank F. Xu, Lei Ji, Botian Shi, Junyi Du, Graham Neubig, Yonatan Bisk, Nan Duan.
In Proceedings of the First International Workshop on Natural Language Processing Beyond Text @ EMNLP, 2020

Minimally Supervised Categorization of Text with Metadata
Yu Zhang, Yu Meng, Jiaxin Huang, Frank F. Xu, Xuan Wang, Jiawei Han.
In Proceedings of 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2020

Incorporating External Knowledge through Pre-training for Natural Language to Code Generation
Frank F. Xu*, Zhengbao Jiang*, Pengcheng Yin, Bogdan Vasilescu, Graham Neubig.
In Proceedings of 58th Annual Meeting of the Association for Computational Linguistics (ACL), 2020

How Can We Know What Language Models Know?
Frank F. Xu*, Zhengbao Jiang*, Jun Araki, Graham Neubig.
In Transactions of the Association for Computational Linguistics (TACL), 2020

A Benchmark for Structured Procedural Knowledge Extraction from Cooking Videos
Frank F. Xu, Lei Ji, Botian Shi, Junyi Du, Graham Neubig, Yonatan Bisk, Nan Duan.
arXiv: 2005.00706

Parsimonious Morpheme Segmentation with an Application to Enriching Word Embeddings
Ahmed El-Kishky, Frank F. Xu, Aston Zhang, and Jiawei Han.
In Proceedings of IEEE International Conference on Big Data, 2019

HiGitClass: Keyword-Driven Hierarchical Classification of GitHub Repositories
Yu Zhang, Frank F. Xu, Sha Li, Yu Meng, Xuan Wang, Qi Li, Jiawei Han.
In Proceedings of IEEE International Conference on Data Mining (ICDM), 2019

StateLens: A Reverse Engineering Solution for Making Existing Dynamic Touchscreens Accessible

Anhong Guo, Junhan Kong, Michael Rivera, <u>Frank F. Xu</u>, Jeffrey P. Bigham.
In Proceedings of 32nd Annual ACM Symposium on User Interface Software and Technology (UIST), 2019

AlpacaTag: An Active Learning-based Crowd Annotation Framework for Sequence Tagging

Bill Y. Lin*, Dong-Ho Lee*, <u>Frank F. Xu</u>, Ouyu Lan, Xiang Ren.
In Proceedings of 57th Annual Meeting of the Association for Computational Linguistics (ACL Demo), 2019

Mining Cross-Cultural Differences and Similarities in Social Media

<u>Frank F. Xu</u>*, Bill Y. Lin*, Kenny Q. Zhu, Seung-won Hwang.
In Proceedings of 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018

Automatic Extraction of Commonsense LocatedNear Knowledge

<u>Frank F. Xu</u>*, Bill Y. Lin*, Kenny Q. Zhu.
In Proceedings of 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018

Toward Unsupervised Text Content Manipulation

Wentao Wang, Zhiting Hu, Zichao Yang, Haoran Shi, <u>Frank F. Xu</u>, Eric Xing.
arXiv: 1901.09501

ExtRA: Extracting Prominent Review Aspects from Customer Feedback

Zhiyi Luo, Shanshan Huang, <u>Frank F. Xu</u>, Bill Yuchen Lin, Hanyuan Shi, Kenny Zhu.
In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018

Empower Sequence Labeling with Task-Aware Neural Language Model

Liyuan Liu, Jingbo Shang, <u>Frank F. Xu</u>, Xiang Ren, Huan Gui, Jian Peng, Jiawei Han.
In Proceedings of Thirty-second AAAI Conference on Artificial Intelligence (AAAI), 2018

ReQuest: Indirect Supervision for Relation Extraction using Question-Answer Pairs

Zeqiu Wu, Xiang Ren, <u>Frank F. Xu</u>, Ji Li, Jiawei Han.
In Proceedings of The 11th ACM International Conference on Web Search and Data Mining (WSDM), 2018

Mining Cross-Cultural Differences of Named Entities: A Preliminary Study

<u>Frank F. Xu</u>*, Bill Y. Lin*, Kenny Q. Zhu.
Presented on First Workshop on Knowledge Base Construction, Reasoning and Mining (KBCOM) at WSDM, 2018

Entropy-Based Subword Mining with an Application to Word Embeddings

Ahmed El-Kishky, <u>Frank F. Xu</u>, Aston Zhang, Stephen Macke, Jiawei Han.
In Proceedings of the Second Workshop on Subword/Character LEvel Models (SCLeM) at NAACL, 2018

Open Information Extraction with Global Structure Constraints

Qi Zhu, Xiang Ren, Jingbo Shang, Yu Zhang, <u>Frank F. Xu</u>, Jiawei Han.
In Proceedings of The Web Conference (WWW Poster), 2018 **(Best Poster Award Honorable Mention)**

Commonsense LocatedNear Relation Extraction

<u>Frank F. Xu</u>*, Bill Y. Lin*, Kenny Q. Zhu.
Presented on 6th Workshop on Automated Knowledge Base Construction (AKBC) at NIPS, 2017

Multi-channel BiLSTM-CRF Model for Emerging Named Entity Recognition in Social Media

<u>Frank F. Xu</u>*, Bill Y. Lin*, Zhiyi Luo, Kenny Q. Zhu.
In Proceedings of the 3rd Workshop on Noisy User-generated Text (W-NUT) at EMNLP, 2017 **(2nd place in the shared task)**

Transfer Learning for Traffic Speed Prediction: A Preliminary Study

<u>Frank F. Xu</u>*, Bill Y. Lin*, Eve Q. Liao, Kenny Q. Zhu.
In Proceedings of SmartIoT 2018 Workshop: AI enhanced IoT data processing for Intelligent Applications at AAAI, 2018

Cross-region Traffic Prediction for China on OpenStreetMap.

<u>Frank F. Xu</u>, Bill Y. Lin, Qi Lu, Yifei Huang and Kenny Q. Zhu.
In Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science, 2016

Thesis

Indirect Supervision for Information Extraction

<u>Frank F. Xu</u>
Master's Thesis, Shanghai Jiao Tong University, 2019

Traffic Prediction for Urban Planning

<u>Frank F. Xu</u>
*Bachelor's Thesis **(Best Thesis Award, Top 1 in CSE)**, Shanghai Jiao Tong University*, 2016

Professional Service

Program Committee

First Workshop on Knowledge Base Construction, Reasoning and Mining at WSDM 2018 (KBCOM)

Conference Reviewer

AAAI Conference on Artificial Intelligence (AAAI 2016, 17)

ACM International Conference on Information and Knowledge Management (CIKM 2017)

ACM International Conference on Web Search and Data Mining (WSDM 2018)

The Web Conference (WWW 2018)

ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018)

The Annual Meeting of the Association for Computational Linguistics (ACL 2018, NAACL 2019, ACL 2019)