

Yu Su

783 Drees Laboratories, Columbus, OH 43210, USA

✉ su.809@osu.edu | 🌐 http://ysu1989.github.io | 📱 ysu1989

Research Interests

At the intersection of natural language processing, machine learning, and data mining, my research aims to develop an **AI-powered knowledge engine** to bridge the gap between human and data by innovations along the following dimensions:

- **Problem Areas:** Language Interfaces (Semantic Parsing, Question Answering, Dialogue Systems), Information Extraction, Knowledge-based Reasoning
- **Methodologies:** Deep Learning, Knowledge Representation, Human-Machine Collaboration, Transfer Learning
- **Applications:** Biomedicine, Healthcare, Internet of Things

Education

University of California, Santa Barbara

Santa Barbara, CA, USA

PH.D. IN COMPUTER SCIENCE

2012 - 2018

- Advisor: Xifeng Yan
- Dissertation: *Towards Democratizing Data Science with Natural Language Interfaces*
- *Outstanding Dissertation Award*

Tsinghua University

Beijing, China

B.ENG. IN COMPUTER SCIENCE AND TECHNOLOGY

2008 - 2012

- *Outstanding Graduate Award*

Professional Experience

Department of Computer Science & Engineering, The Ohio State University

Columbus, OH, USA

ASSISTANT PROFESSOR

2020 - Present

Microsoft Semantic Machines

Berkeley, CA, USA

SENIOR RESEARCHER

2018 - Present

Microsoft Research

Redmond, WA, USA

RESEARCH INTERN

2016, 2017

U.S. Army Research Laboratory

Adelphi, MD, USA

VISITING RESEARCHER

2015, 2016

IBM T.J. Watson Research Center

Yorktown Heights, NY, USA

RESEARCH INTERN

2015

Professional Service

CO-ORGANIZER

- Faculty Advisor to Student Research Workshop, NAACL 2021.
- First Workshop on Natural Language Processing for Programming, Bangkok, Thailand, 2021 (co-located with ACL).
- First Workshop on Natural Language Interfaces (NLI), Seattle, United States, 2020 (co-located with ACL).
- First Workshop on Knowledge Base Construction, Reasoning and Mining (KBCOM), Los Angeles, United States, 2018 (co-located with WSDM).

PROGRAM COMMITTEE MEMBER/REVIEWER

- Annual Meeting of the Association for Computational Linguistics (**ACL**): 2017, 2018, 2019, 2020
- Conference on Empirical Methods in Natural Language Processing (**EMNLP**): 2017, 2018, 2019, 2020

- Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL**): 2019, 2021
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD**): 2018, 2019 (**session chair**), 2020 (**session chair**)
- AAAI Conference on Artificial Intelligence (**AAAI**): 2019, 2021
- The International World Wide Web Conference (**WWW**): 2018
- SIAM International Conference on Data Mining (**SDM**): 2019
- ACM International Conference on Information and Knowledge Management (**CIKM**): 2017
- Conference on Automated Knowledge Base Construction (**AKBC**): 2019, 2020
- Conference on Computational Natural Language Learning (**CoNLL**): 2018
- CCF Conference on Natural Language Processing and Chinese Computing (**NLPCC**): 2017, 2018, 2020 (**area chair**)
- IEEE Transactions on Neural Networks and Learning Systems (**TNNLS**)
- IEEE Transactions on Knowledge and Data Engineering (**TKDE**)
- ACM Transactions on Knowledge Discovery from Data (**TKDD**)

OUTREACH ACTIVITY

- March 2018. Introduce data science and AI to high school students from Laguna Blanca School, Santa Barbara, CA.
- March 2018. “Afternoon with Academia,” UCSB chapter of Society of Asian Scientists and Engineers.

Teaching

2020	Instructor , CSE5539: Cutting-Edge Topics in Natural Language Processing (Undergraduate/Graduate)	OSU
2020	Instructor , CSE5243: Introduction to Data Mining (Undergraduate/Graduate)	OSU
2016	Guest Lecturer , CS273: Data and Knowledge Bases (Graduate)	UCSB
2016	Guest Lecturer , CS291K: Advanced Data Mining (Graduate)	UCSB
2015	Teaching Assistant , CS290D: Advanced Data Mining (Graduate)	UCSB
2013	Teaching Assistant , CS32: Object-Oriented Design and Implementation (Undergraduate)	UCSB
2013	Teaching Assistant , CS176B: Network Computing (Undergraduate)	UCSB
2012	Teaching Assistant , CS176A: Introduction to Computer Communication Networks (Undergraduate)	UCSB

Honors & Awards

2019	Best of IEEE ICDM 2019 Selection	IEEE
2019	Outstanding Dissertation Award of Computer Science	UCSB
2018	Best Distinguished Graduate Student Lecture	UCSB
2016	SIAM International Conference on Data Mining Travel Award	NSF
2012	Outstanding Graduate Award	Tsinghua University
2009 - 2011	Merit-based Fellowships	Tsinghua University
2008	Outstanding Freshman Fellowship	Tsinghua University

Publications

UNDER REVIEW

- [30] **Beyond I.I.D.: Three Levels of Generalization for Question Answering on Knowledge Bases**
Yu Gu, Sue Kase, Michelle Vanni, Brian Sadler, Percy Liang, Xifeng Yan, and Yu Su
- [29] **Compositional Generalization for Neural Semantic Parsing via Span-level Supervised Attention**
Pengcheng Yin, Hao Fang, Graham Neubig, Adam Pauls, Emmanouil Antonios Platanios, Yu Su, Sam Thomson, and Jacob Andreas

REFEREED CONFERENCE PAPERS

- [28] **An Imitation Game for Learning Semantic Parsers from User Interaction**
Ziyu Yao, Yiqi Tang, Scott Wen-tau Yih, Huan Sun, and Yu Su
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020
- [26] **KGPT: Knowledge-Grounded Pre-Training for Data-to-Text Generation**
Wenhu Chen, Yu Su, Xifeng Yan, and William Yang Wang
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020
- [27] **Document Classification for COVID-19 Literature**
Bernal Jimenez Gutierrez, Jucheng Zeng, Dongdong Zhang, Ping Zhang, and Yu Su
Findings of EMNLP, short paper, 2020
- [25] **Task-Oriented Dialogue as Dataflow Synthesis**
Semantic Machines, Jacob Andreas, John Bufe, David Burkett, Charles Chen, Josh Clausman, Jean Crawford, Kate Crim, Jordan DeLoach, Leah Dorner, Jason Eisner, Hao Fang, Alan Guo, David Hall, Kristin Hayes, Kellie Hill, Diana Ho, Wendy Iwaszuk, Smriti Jha, Dan Klein, Jayant Krishnamurthy, Theo Lanman, Percy Liang, Christopher H. Lin, Ilya Lintsbakh, Andy McGovern, Aleksandr Nisnevich, Adam Pauls, Dmitriy Petters, Brent Read, Dan Roth, Subhro Roy, Jesse Rusak, Beth Short, Div Slomin, Ben Snyder, Stephon Striplin, Yu Su, Zachary Tellman, Sam Thomson, Andrei Vorobev, Izabela Witoszko, Jason Wolfe, Abby Wray, Yuchen Zhang, and Alexander Zotov
Transactions of the Association for Computational Linguistics (TACL), 2020
Highlighted by Microsoft CEO Satya Nadella at Microsoft Build 2019
- [24] **Logical Natural Language Generation from Open-Domain Tables**
Wenhu Chen, Jianshu Chen, Yu Su, Zhiyu Chen, and William Yang Wang
Annual Conference of the Association for Computational Linguistics (ACL), 2020
- [23] **Model-based Interactive Semantic Parsing: A Unified Formulation and A Text-to-SQL Case Study**
Ziyu Yao, Yu Su, Huan Sun, and Scott Wen-tau Yih
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019
- [22] **HierCon: Hierarchical Organization of Technical Documents based on Concepts**
Keqian Li, Shiyang Li, Semih Yavuz, Hanwen Zha, Yu Su, and Xifeng Yan
IEEE International Conference on Data Mining (ICDM), 2019
Best of ICDM 2019 Selection
- [21] **Global Textual Relation Embedding for Relational Understanding**
Zhiyu Chen, Hanwen Zha, Honglei Liu, Wenhu Chen, Xifeng Yan, and Yu Su
Annual Conference of the Association for Computational Linguistics (ACL), short paper, 2019
- [20] **How Large A Vocabulary Does Text Classification Need? A Variational Approach on Vocabulary Selection**
Wenhu Chen, Yu Su, Yilin Shen, Zhiyu Chen, Xifeng Yan and William Yang Wang
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019
- [19] **Learning to Compose Topic-Aware Mixture of Experts for Zero-Shot Video Captioning**
Xin Wang, Jiawei Wu, Da Zhang, Yu Su, William Yang Wang
AAAI Conference on Artificial Intelligence (AAAI), 2019
- [18] **Concept Mining via Embedding**
Keqian Li, Hanwen Zha, Yu Su, Xifeng Yan
IEEE International Conference on Data Mining (ICDM), 2018
- [17] **XL-NBT: A Cross-lingual Neural Belief Tracking Framework**
Wenhu Chen, Jianshu Chen, Yu Su, Xin Wang, Dong Yu, Xifeng Yan and William Yang Wang
23th Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018
- [16] **What It Takes to Achieve 100% Condition Accuracy on WikiSQL**
Semih Yavuz, Izzeddin Gur, Yu Su, Xifeng Yan
23th Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018

- [15] **DialSQL: Dialogue Based Structured Query Generation**
Izzeddin Gur, Semih Yavuz, Yu Su, Xifeng Yan
56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018
(Acceptance rate: 258/1018=25.3%)
- [14] **Natural Language Interfaces with Fine-Grained User Interaction: A Case Study on Web APIs**
Yu Su, Ahmed H. Awadallah, Miaosen Wang, Ryan White
41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2018
(Acceptance rate: 86/409=21.0%)
- [13] **Global Relation Embedding for Relation Extraction**
Yu Su*, Honglei Liu*, Semih Yavuz, Izzeddin Gur, Huan Sun, Xifeng Yan (*: equal contribution)
16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2018
(Acceptance rate: 207/665=31.1%)
- [12] **Unsupervised Neural Categorization for Scientific Publications**
Keqian Li, Hanwen Zha, Yu Su, Xifeng Yan
18th SIAM International Conference on Data Mining (SDM), 2018
(Acceptance rate: 87/375=23.2%)
- [11] **Building Natural Language Interfaces to Web APIs**
Yu Su, Ahmed H. Awadallah, Patrick Pantel, Michael Gamon, Madian Khabisa, Mark Encarnacion
26th ACM International Conference on Information and Knowledge Management (CIKM), 2017
(Acceptance rate: 171/820=20.9%)
- [10] **Cross-domain Semantic Parsing via Paraphrasing**
Yu Su, Xifeng Yan
22th Conference on Empirical Methods in Natural Language Processing (EMNLP), 2017
(Acceptance rate: 216/836=25.8%)
- [9] **An End-to-End Deep Framework for Answer Triggering with a Novel Group-Level Objective**
Jie Zhao, Yu Su, Ziyu Guan, Huan Sun
22th Conference on Empirical Methods in Natural Language Processing (EMNLP), short paper, 2017
(Acceptance rate: 107/582=18.4%)
- [8] **Recovering Question Answering Errors via Query Revision**
Semih Yavuz, Izzeddin Gur, Yu Su, Xifeng Yan
22th Conference on Empirical Methods in Natural Language Processing (EMNLP), short paper, 2017
(Acceptance rate: 107/582=18.4%)
- [7] **On Generating Characteristic-rich Question Sets for QA Evaluation**
Yu Su, Huan Sun, Brian Sadler, Mudhakar Srivatsa, Izzeddin Gur, Zenghui Yan, Xifeng Yan
21th Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016
(Acceptance rate: 177/687=25.8%)
- [6] **Improving Semantic Parsing via Answer Type Inference**
Semih Yavuz, Izzeddin Gur, Yu Su, Mudhakar Srivatsa, Xifeng Yan
21th Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016
(Acceptance rate: 177/687=25.8%)
- [5] **A Fast Kernel for Attributed Graphs**
Yu Su, Fangqiu Han, Richard E. Harang, Xifeng Yan
16th SIAM International Conference on Data Mining (SDM), 2016
(Acceptance rate: 96/371=25.9%)
- [4] **Table Cell Search for Question Answering**
Huan Sun, Hao Ma, Xiaodong He, Wen-Tau Yih, Yu Su, Xifeng Yan
25th International Conference on World Wide Web (WWW), 2016
(Acceptance rate: 115/727=15.8%)

- [3] **Visual Graph Query Formulation and Exploration: A New Perspective on Information Retrieval at the Edge**
Sue Kase, Michelle Vanni, Joanne Knight, Yu Su, Xifeng Yan
SPIE 9851, Next-Generation Analyst IV (SPIE Defense+Security), 2016
- [2] **Exploiting Relevance Feedback in Knowledge Graph Search**
Yu Su, Shengqi Yang, Huan Sun, Mudhakar Srivatsa, Sue Kase, Michelle Vanni, Xifeng Yan
21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2015
(Acceptance rate: 160/819=19.5%)
- [1] **On the Validity of Geosocial Mobility Traces**
Zengbin Zhang, Lin Zhou, Xiaohan Zhao, Gang Wang, Yu Su, Miriam Metzger, Haitao Zheng, and Ben Y. Zhao
12th ACM Workshop on Hot Topics in Networks (HotNets), 2013
(Acceptance rate: 26/110=23.6%)

CONFERENCE TUTORIALS

- [2] **Scalable Construction and Querying of Massive Knowledge Bases**
Xiang Ren, Yu Su, Pedro Szekely, Xifeng Yan
The Web Conference (formerly WWW), 2018
- [1] **Construction and Querying of Large-Scale Knowledge Bases**
Xiang Ren, Yu Su, Xifeng Yan
26th ACM International Conference on Information and Knowledge Management (CIKM), 2017

PATENTS

- [2] **Natural Language to API Conversion (co-inventor)**
US Patent Pending 15/919,209
A framework for building interactive natural language interface to web APIs.
- [1] **Natural Language Interface to Web API (co-inventor)**
US Patent Granted 10496452B2
A framework for building natural language interface to web API via crowdsourcing with a hierarchical probabilistic model and optimization algorithm for crowdsourcing.

Grants

- [1] **(Sole PI) Fujitsu Lab of America: “Large-scale, Ontology-preserving Knowledge Graph Embedding”**
Grant: \$50,000, 02/2020-02/2021.

Invited Talks

- [16] Task-Oriented Dialogue as Dataflow Synthesis
Faculty-in-Focus Seminar, CSE@OSU. November 09, 2020. Columbus, OH, United States.
- [15] Towards Democratizing Data Science with AI-powered Knowledge Engines
Stanford NLP Seminar. February 07, 2019. Stanford, CA, United States.
- [14] Bridging the Gap between Human and Data with AI
Microsoft Semantic Machines. August 13, 2018. Berkeley, CA, United States.
- [13] Bridging the Gap between Human and Data with AI
Fudan University. July 24, 2018. Shanghai, China.
- [12] Bridging the Gap between Human and Data with AI
Tsinghua University. June 21, 2018. Beijing, China.
- [11] Bridging the Gap between Human and Data with AI
Fujitsu Laboratories of America. March 9, 2018. Sunnyvale, California, United States.

- [10] Bridging the Gap between Human and Data with AI
College of Information and Computer Sciences, University of Massachusetts. February 15, 2018. Amherst, Massachusetts, United States.
- [9] Bridging the Gap between Human and Data with AI
Dept. of Biomedical Informatics, OSU. January 26, 2018. Columbus, Ohio, United States.
- [8] Bridging the Gap between Human and Data with AI
Dept. of Computer Science and Engineering, OSU. January 25, 2018. Columbus, Ohio, United States.
- [7] On Building Natural Language Interface to Data and Services
Toutiao AI Lab. October 17, 2017. Beijing, China.
- [6] On Building Natural Language Interface to Data and Services
Tsinghua University. October 16, 2017. Beijing, China.
- [5] On Building Natural Language Interface to Data and Services
Fudan University. October 10, 2017. Shanghai, China.
- [4] Building Natural Language Interfaces to Web APIs
Microsoft Research, MSR AI Seminar. July 25, 2017. Redmond, Washington, United States.
- [3] Learning from Zero: Recent Advances in Bootstrapping Semantic Parsers using Crowdsourcing
USC/ISI. October 28, 2016. Los Angeles, California, United States.
- [2] Exploiting Relevance Feedback in Knowledge Graph Search
IBM Research. August 15, 2015. Yorktown Heights, New York, United States.
- [1] Search over Large-scale Knowledge Graphs
UCSB, IGERT Network Science Seminar. August 14, 2014. Santa Barbara, California, United States.