Tao Yu

Contact Information	https://taoyds.github.io/ 420 Temple Street, New Haven, CT 06511	tao.yu@yale.edu 801-493-9166

Research Interests
Natural language processing, deep learning, and machine learning. More specifically, I am currently interested in building conversational natural language interfaces for information access.

Education

Yale University, New Haven, CT

Ph.D. Computer Science - Natural Language Processing   08/2017 - 05/2021 expected
- Advisors: Dragomir R. Radev

Columbia University, SEAS, New York, NY

M.S. Data Science   GPA: 3.82/4.0   09/2015 - 05/2017
- Advisors: Owen Rambow and Kathleen McKeown

University of Utah, Salt Lake City, UT

B.S. Mathematics and Economics   GPA: 3.86/4.0   01/2012 - 05/2015
- Advisors: Jingyi Zhu and Peter Philips

Research & Industry Experience

Yale University, New York, NY

Graduate Student Research – NLP   08/2017 - Present
- Working on dialog-based robust and reliable semantic parsing/text-to-SQL systems.
- Open Text-to-SQL Challenge Series: Spider, SParC, CoSQL.

Microsoft Research, Redmond, WA

Research Intern mentored by Ahmed Hassan Awadallah – NLP   06/2020 - 08/2020

Salesforce Research, Palo Alto, CA

Research Intern mentored by Victoria Lin and Caiming Xiong – NLP   06/2019 - 08/2019

Samsung Research America, Mountain View, CA

Research Intern – NLP   06/2018 - 08/2018

Columbia CCLS & NLP Group, New York, NY

Graduate Research Assistant   05/2016 - 10/2016
- Developed systems for DARPA Text Analysis Conference (TAC) 2016 Belief and Sentiment Evaluation, and achieved the best performances on all 12 English and Spanish tasks.

AIG Science, New York, NY

App Development Intern   01/2016 - 03/2016
- Developed a Hybrid Mobile App using React Native and built backend using Flask.

Publications & Papers in Preparation

1. **T. Yu**, R. Zhang, H. Er, S. Li, E. Xue, B. Pang, V. Lin, Y. Tan, T. Shi, Z. Li, Y. Jiang, M. Yasunaga, S. Shim, T. Chen, A. Fabbri, Z. Li, L. Chen, Y. Zhang, S. Dixit, V. Zhang, C. Xiong, R. Socher, W. Lasecki, D. Radev. "CoSQL: A Conversational Text-to-SQL Challenge Towards Cross-Domain Natural Language Interfaces to Databases". *In the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019. *Long Paper*

2. R. Zhang, **T. Yu**, H. Er, S. Shim, E. Xue, V. Lin, T. Shi, C. Xiong, R. Socher, W. Lasecki, D. Radev. "Editing-based SQL Query Generation for Cross-Domain Context-Dependent Questions". *In the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019. *Long Paper*

3. **T. Yu**, R. Zhang, M. Yasunaga, Y. Tan, V. Lin, S. Li, H. Er, I. Li, B. Pang, T. Chen, E. Ji, S. Dixit, D. Proctor, S. Shim, J. Kraft, V. Zhang, C. Xiong, R. Socher, D. Radev. "SParC: Cross-Domain Semantic Parsing in Context". *In the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. *Long Paper*

4. R. Plunz, Y. Zhou, M. Vintimilla, K. Mckeown, **T. Yu**, L. Uguccioni, M. Sutto. "Twitter Sentiment in New York City Parks as Measure of Well-being". *Landscape and Urban Planning*, 2019. *Long Paper*

5. **T. Yu**, R. Zhang, K. Yang, M. Yasunaga, D. Wang, Z. Li, J. Ma, I. Li, Q. Yao, S. Roman, Z. Zhang, D. Radev. "Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task". *In the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018. *Long Paper*

6. **T. Yu**, M. Yasunaga, K. Yang, R. Zhang, D. Wang, Z. Li, D. Radev. "SyntaxSQLNet: Syntax Tree Networks for Complex and Cross Domain Text-to-SQL Task". *In the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018. *Long Paper*

7. **T. Yu**, Z. Li, Z. Zhang, R. Zhang, D. Radev. "TypeSQL: Knowledge-based Type-Aware Neural Text-to-SQL Generation". *In the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, 2018. *Short Paper*

8. M. S. Rasooli, N. Farra, A. Radeva, **T. Yu** and K. McKeown. "Cross-Lingual Sentiment Transfer with Limited Resources". *Machine Translation*, 2017. *Long Paper*

9. O. Rambow, **T. Yu**, A. Radeva, S. Hamidian, A. Fabbri, C. Hidey, K. McKeown et. al. "The Columbia-GWU System at the 2016 TAC KBP BeSt Evaluation". *NIST TAC KBP Workshop*, 2016. *Long Paper*

10. **T. Yu**, C. Hidey, O. Rambow, K. McKeown. "Leveraging Sparse and Dense Feature Combinations for Sentiment Classification". *arXiv:1708.03940 [cs.CL]*, 2017. *Short Paper*

PROFESSIONAL SERVICES

Organizing Committee Member

The Workshop on Interactive and Executable Semantic Parsing at EMNLP 2020

Program Committee Member

The Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020

The 28th International Conference on Computational Linguistics (COLING), 2020

The Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing (AACL-IJCNLP), 2020

The Conference of the Association for Computational Linguistics (ACL), 2020

The Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019

The North American Chapter of the Association for Computational Linguistics (NAACL), 2019

Awards	Yale Conference Travel Fellowship	2018
	Yale PhD Fellowship	2017
	Connie and Robert White Honor Roll Scholarship	2014
	Mathematics Department Scholarship	2014
	International Student and Scholar Services Scholarship	2014
	NSF Research Experience for Undergraduates (REU) Fellowship	2014

Teaching Experience	Natural Language Processing (CPSC 577), Yale University	2019
	Database Systems (CPSC 537), Yale University	2019
	Introduction to Natural Language Processing (CS 4705), Columbia University	2017
	Natural Language Processing (Coursera)	2016

Programming & Other Skills Proficient in Python, PyTorch, TensorFlow, R, Java, Matlab, SQL. Familiar with C, C++, JavaScript, Theano, Hadoop, Spark. FAA Student Pilot.