

## Wei (Victor) Yang

---

CONTACT INFORMATION	<a href="mailto:w85yang@uwaterloo.ca">w85yang@uwaterloo.ca</a> Homepage, Github, Linkedin, Google Scholar
INDUSTRY EXPERIENCE	<p><b>Research Engineer II, Borealis.ai</b> Toronto, Dec 2019 - Now</p> <ul style="list-style-type: none"><li>Research tasks: 1). build various model architectures for semantic parsing including RoBERTa + RAT encoder, grammar-guided LSTM/transformer decoder, meta learning and neural data augmentation; 2). lead a small research group to investigate the contributions of data from different sources in a proof-of-concept (POC) project.</li><li>Engineering tasks: 1). build and maintain a safe-guard module for out-of-domain and unanswerable question detection in a question answering system; 2). build an end-to-end inference system for natural language database interface and deploy it into the market research domain inside Royal Bank of Canada (RBC); 3). write template-matching grammar in the answer retrieval module of a FAQ system.</li></ul> <p><b>Machine Learning Engineer, RSVP.ai</b> Waterloo, May 2018 - Aug 2019</p> <ul style="list-style-type: none"><li>Build a open-domain question answering system for intelligent chatbot</li><li>Build a coreference resolution system for question understanding in chatbot</li></ul> <p><b>Software Engineer Intern, DiDi Research</b> Hangzhou, Feb 2017 - Jun 2017</p> <ul style="list-style-type: none"><li>Works on planning in a self-driving project</li></ul> <p><b>Software Engineer Intern, Netease Game</b> Hangzhou, June 2016 - Oct 2016</p> <ul style="list-style-type: none"><li>Works on an auto-testing framework for a MMORPG game (Revelation Online)</li></ul>
EDUCATION	<p><b>University of Waterloo</b>, Waterloo, ON, Canada</p> <p>M.Math., Computer Science (GPA: 86.8 / 100) Sep 2017 - Apr 2019</p> <ul style="list-style-type: none"><li>Thesis: End-to-End Neural Information Retrieval [<a href="#">Paper</a>] [<a href="#">Slide</a>]</li><li>Thesis Supervisor: Jimmy Lin</li></ul> <p><b>Singapore University of Technology and Design</b>, Singapore</p> <p>Exchange Student (GPA: 4.0 / 4.0 ) Sep 2016 - Dec 2016</p> <p><b>Zhejiang University</b>, Hangzhou, China</p> <p>B.S., Computer Science and Technology (GPA: 3.87 / 4.0) Sep 2013 - Jun 2017</p> <ul style="list-style-type: none"><li>Thesis: Cross Domain Word Embedding (advised by <a href="#">Lu Wei</a> and <a href="#">Siliang Tang</a>)</li><li>Minor: Advanced Honors Class of Engineering Education (ACEE), Chu Kochen Honors College, Zhejiang University</li></ul>
ACADEMIC EXPERIENCE	<p><b>Research Assistant</b> Sep 2017 - Apr 2019</p> <p>Data System Group, University of Waterloo, supervised by <a href="#">Jimmy Lin</a></p> <p>Research Areas: Neural Information Retrieval, Open-domain Question Answering</p> <p><b>Research Assistant</b> Oct 2016 - Feb 2017</p> <p>Singapore University of Technology and Design, supervised by <a href="#">Lu Wei</a></p> <p>Research Areas: Cross-domain Word Embedding</p> <p><b>Research Assistant</b> Jun 2016 - Sep 2016</p> <p>The Hong Kong Polytechnic University, supervised by <a href="#">Ajay Kumar</a> (IEEE Fellow)</p> <p>Research Areas: Neural Palmprint Recognition</p>
ACHIEVEMENTS	<ul style="list-style-type: none"><li>Rank 1st in the Spider Cross-domain Semntic Parsing leaderboard 2020</li><li>Rank 1st in the MS-MARCO Passage Retrieval leaderboard 2019</li><li>Student Researcher Scholarship, EMNLP/SIGIR/NAACL 2017/2019/2019</li><li>Best Paper Honourable Mention, SIGIR 2019</li><li>Graduate with Honors, Zhejiang University &amp; Zhejiang Province 2017</li></ul>

OPEN SOURCE  
CONTRIBUTION

Castor • Anserini • NCRFpp • Birch • MatchZoo • BERT-Fine\_tune • Capreolus

ACADEMIC  
SERVICE

Sub-reviewer: ACL 2020, NAACL 2019, SIGIR 2019, CIKM 2019, AAAI 2018  
Reviewer: AACL 2020, EMNLP(2021, 2020, 2019), COLING 2020, AAAI(2021, 2020),  
IJCAI 2021, NAACL 2021, ACL 2021

REFERRED  
PUBLICATIONS

1. Chenyang Huang, **Wei Yang**, Yanshuai Cao, Osmar Zaiane and Lili Mou “A Globally Normalized Neural Model for Semantic Parsing” *ACL workshop*, 2021.
2. Peng Xu, Wenjie Zi, Hamidreza Shahidi, Akos Kadar, Keyi Tang, **Wei Yang**, Jawad Ateeq, Harsh Barot, Meidan Alon and Yanshuai Cao “TURING: an Accurate and Interpretable Multi-Hypothesis Cross-Domain Natural Language Database Interface” *ACL Demo*, 2021.
3. Peng Xu, Dhruv Kumar, **Wei Yang**, Wenjie Zi, Keyi Tang, Chenyang Huang, Jackie Chi Kit Cheung, Simon J.D. Prince, Yanshuai Cao “Optimizing Deeper Transformers on Small Datasets” *ACL*, 2021. [\[Paper\]](#)
4. Zhengkai Tu\*, **Wei Yang\***, Zihang Fu, Yuqing Xie, Luchen Tan, Kun Xiong, Ming Li and Jimmy Lin “Approximate Nearest Neighbor Search and Lightweight Dense Vector Reranking in Multi-Stage Retrieval Architectures” *ICTIR*, 2020. (\* equal contribution) [\[Paper\]](#)
5. Rodrigo Nogueira, **Wei Yang**, Kyunghyun Cho, Jimmy Lin “Multi-stage document ranking with BERT” [\[Paper\]](#) [\[Code\]](#)
6. **Wei Yang\***, Yuqing Xie\*, Luchen Tan, Kun Xiong, Ming Li, and Jimmy Lin “Distant Supervision for Multi-Stage Fine-Tuning in Retrieval-Based Question Answering” (\* equal contribution) *WWW*, 2020. [\[Paper\]](#)
7. Andrew Yates, Siddhant Arora, Xinyu Zhang, **Wei Yang**, Kevin Martin Jose and Jimmy Lin “Capreolus: A Toolkit for End-to-End Neural Ad Hoc Retrieval.” *WSDM demo*, 2020. [\[Paper\]](#) [\[Code\]](#)
8. Zeynep Akkalyoncu Yilmaz, Shengjin Wang, **Wei Yang**, Haotian Zhang and Jimmy Lin “Applying BERT to Document Retrieval with Birch.” *EMNLP demo*, 2019. [\[Paper\]](#) [\[Code\]](#)
9. Jinfeng Rao, Linqing Liu, Yi Tay, **Wei Yang**, Peng Shi, Jimmy Lin “Bridging the Gap of Relevance Matching and Semantic Matching with Hierarchical Co-Attention Network.” *EMNLP*, 2019. [\[Paper\]](#) [\[Code\]](#)
10. Linqing Liu, **Wei Yang**, Jinfeng Rao, Raphael Tang, and Jimmy Lin “Incorporating Compositional and Syntactic Structures Improves Semantic Similarity Modeling.” *EMNLP*, 2019. [\[Paper\]](#) [\[Code\]](#)
11. Zeynep Akkalyoncu, **Wei Yang**, Haotian Zhang, and Jimmy Lin “Cross-Domain Modeling of Sentence-Level Evidence for Document Retrieval.” *EMNLP*, 2019. [\[Paper\]](#) [\[Code\]](#)
12. Rodrigo Nogueira, **Wei Yang**, Jimmy Lin, and Kyunghyun Cho “Document Expansion by Query Prediction.” [\[Paper\]](#) [\[Code\]](#)
13. **Wei Yang\***, Haotian Zhang\*, and Jimmy Lin “Simple Applications of BERT for Ad Hoc Document Retrieval.” (\* equal contribution) [\[Paper\]](#) [\[Code\]](#)

14. **Wei Yang**, Kuang Lu, Peilin Yang, Jimmy Lin “Critically Examining the “Neural Hype”: Weak Baselines and the Additivity of Effectiveness Gains from Neural Ranking Models.” *SIGIR*, 2019. (**Best Short Paper Honorable Mention**) [\[Paper\]](#) [\[Code\]](#)
15. **Wei Yang\***, Yuqing Xie\*, Aileen Lin, Xingyu li, Luchen Tan, Kun Xiong, Ming Li, and Jimmy Lin “End-to-End Open-Domain Question Answering with BERTserini.” (\* equal contribution) *NAACL*, 2019. [\[Paper\]](#) [\[Code\]](#)
16. **Wei Yang**, Luchen Tan, Anqi Cui, Han Li, Xi Chen, Chunwei Lu, Muzi Wang, and Jimmy Lin “Detecting Customer Complaint Escalation with Recurrent Neural Networks and Manually-Engineered Features.” *NAACL industry track*, 2019. [\[Paper\]](#)
17. **Wei Yang**, Junnan Chen, Rui Qiao, Haocheng Qin, Ming Li, and Jimmy Lin “End-to-End Neural Context Reconstruction in Chinese Dialogue.” *ACL workshop (NLP for Conversational AI)*, 2019. [\[Paper\]](#)
18. Jinfeng Rao, **Wei Yang**, Yuhao Zhang, Ferhan Ture, and Jimmy Lin “Multi-Perspective Relevance Matching with Hierarchical ConvNets for Social Media Search.” *AAAI*, 2019. [\[Paper\]](#) [\[Code\]](#)
19. **Wei Yang**, Wei Lu, Vincent W. Zheng “A Simple Regularization-based Algorithm for Learning Cross-Domain Word Embeddings.” *EMNLP*, 2017. [\[Paper\]](#) [\[Code\]](#)