

Xinya Du

Webpage: <https://www.cs.cornell.edu/~xdu/>

Email: xdu@cs.cornell.edu

Research Areas	Natural Language Processing, Machine Learning	
Education	Cornell University , Ithaca, NY, USA	
	Ph.D. in Computer Science, Advisor: Prof. Claire Cardie	2016 – 2021 (Expected)
	Shanghai Jiao Tong University , Shanghai, China	
	B.E. in Computer Science and Engineering, GPA Rank: Top 1%	2012 – 2016
Publications	GRIT: Generative Role-filler Transformers for Document-level Event Entity Extraction Xinya Du , Alexander M. Rush, and Claire Cardie <i>Proceedings of EACL, 2021</i>	
	Event Extraction by Answering (Almost) Natural Questions Xinya Du and Claire Cardie <i>Proceedings of EMNLP, 2020</i>	
	Improving Event Duration Prediction via Time-aware Pre-training Zonglin Yang, Xinya Du , Alexander M. Rush and Claire Cardie <i>Proceedings of Findings of EMNLP, 2020</i>	
	Document-Level Event Role Filler Extraction using Multi-Granularity Contextualized Encoding Xinya Du and Claire Cardie <i>Proceedings of ACL, 2020</i>	
	Leveraging Structured Metadata for Improving Question Answering on the Web Xinya Du , Ahmed Hassan Awadallah, Adam Fourney, Robert Sim, Paul Bennett and Claire Cardie <i>Proceedings of AACL, 2020</i>	
	Be Consistent! Improving Procedural Text Comprehension using Label Consistency Xinya Du , Bhavana Dalvi, Niket Tandon, Antoine Bosselut, Wen-tau Yih, Peter Clark, Claire Cardie <i>Proceedings of 17th North American Chapter of the Association for Computational Linguistics (NAACL), 2019</i>	
	Harvesting Paragraph-Level Question-Answer Pairs from Wikipedia Xinya Du and Claire Cardie <i>Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018</i>	
	Identifying Where to Focus in Reading Comprehension for Neural Question Generation Xinya Du and Claire Cardie <i>Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2017</i>	
	Learning to Ask: Neural Question Generation for Reading Comprehension Xinya Du , Junru Shao and Claire Cardie <i>Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), 2017</i> – Featured in New Scientist, TechRepublic – The first paper to propose learning-based methods for natural question generation , inspiring work in Question Answering (Alberti 2019), Summarization (Wang 2020), and Dialogue Systems (Laban 2020)	

Research Internships	Allen Institute for Artificial Intelligence Sep 2018 – Dec 2018 Topic: Improving Procedural Paragraph Understanding by Leveraging Consistency Constraints Supervisors: Bhavana Dalvi, Peter Clark
	Microsoft Research Redmond May 2018 – Aug 2018 Topic: Leveraging web-structured knowledge into plain text for machine reading/question answering. Supervisors: Paul Bennett, Ahmed Hassan Awadallah
	Cornell NLP Group, Undergrad Research Intern Aug 2015 – Dec 2015 Topic: Researched on deep learning techniques for fine-grained opinion extraction. Supervisor: Prof. Claire Cardie
Media Coverage	New Scientist: Inquisitive bot asks questions to test your understanding [Link]
	TechRepublic: How researchers trained one AI system to start asking its own questions [Link]
Talks	<i>Harvesting Paragraph-Level Question-Answer Pairs from Wikipedia</i> ACL 2018, Jul 2018
	<i>LwLL: Progress on the NLP Front</i> DARPA site visit, Cornell University (Online), Apr 2020
	<i>Event Extraction by Answering (Almost) Natural Questions</i> UIUC Class Information Extraction and Knowledge Acquisition (Online), Sep 2020 EMNLP 2020 (Online), Nov 2020
	<i>Towards More Informed Extraction of Events from Documents</i> Rising Stars in Data Science Workshop at University of Chicago, Jan 2021
Mentoring	Zonglin Yang, Cornell CS MEng student, Spring 2020 – Present with publication in Findings of EMNLP 2020
Teaching	Natural Language Processing (TA for Prof. Claire Cardie), Fall 19.
	Natural Language Processing (TA for Prof. Yoav Artzi), Spring 19.
	Software Engineering (TA for Prof. William Arms), Spring 17, Spring 18.
	Introduction to Computing Using Python (TA for Prof. Walker White), Fall 16.
Services	Program Committee Member: Year 2018: ACL, CoNLL, MRQA, W-NUT; Year 2019: NAACL, ACL, EMNLP, CoNLL, *SEM, W-NUT, IJCAI, AAAI; Year 2020: ACL, IJCAI, EMNLP, AACL.
	Journal Reviewer: IEEE Transactions on Knowledge and Data Engineering (TKDE), AI Communications
	Student Volunteer: ACL 2017, 2018; EMNLP 2017; Cornell CS PhD Visit Day 2019, 2020; Cornell CS PhD Admission Committee.
Awards & Scholarships	CDAC Rising Star 2020
	Outstanding Graduate of Shanghai Jiao Tong University 2016
	National Scholarship (Highest honor for undergraduates in China, awarded to top 1% students.) 2013
	National Olympiad in Mathematics in Provinces 2011, The First Prize. 2011
	National Olympiad in Informatics in Provinces 2010, The First Prize. 2010

Code Release

NQG [\[Link\]](#)

Open source repository for neural question generation from sentences (ACL 2017).
Including evaluation scripts and implementations for the models.

EEQA [\[Link\]](#)

Open source repository for using QA techniques for extracting events from text (EMNLP 2020).
Including different question templates and implementations for the models.

Document Event Entity Extractor [\[Link\]](#)

Open source repository for document-level entity extraction.
Including implementations for our model, evaluation scripts for document-level extraction, and analysis for predictions.

References

Claire Cardie

Professor, Department of Computer Science and Department of Information Science
Cornell University
cardie@cs.cornell.edu

Alexander M. Rush

Associate Professor, Department of Computer Science
Cornell University
arush@cornell.edu

Peter Clark

Senior Research Manager
Allen Institute for Artificial Intelligence (AI2)
peterc@allenai.org

Ahmed Hassan Awadallah

Principal Research Manager, Language and Information Technologies Team
Microsoft Research AI
hassanam@microsoft.com

[Updated on 2021-01-26]