Wei Xu

✉ weixu@cse.ohio-state.edu   ☎ +1 718 612 9347
395 Dreese Laboratories, 2015 Neil Avenue Columbus, OH, U.S.A. 43210
http://web.cse.ohio-state.edu/~weixu/

POSITION	**Assistant Professor, Ohio State University**, Columbus, OH	Aug 2016 – present
	Department of Computer Science and Engineering	
	Research Areas: Natural Language Processing, Machine Learning, Social Media	
EDUCATION	**Ph.D. in Computer Science, New York University**, New York, NY	2014
	Thesis: *Data-driven Approaches for Paraphrasing Across Language Variations* (Advisor: Ralph Grishman)	
	B.S. and M.S. in Computer Science, Tsinghua University, Beijing, CHINA	2004/2007
	Thesis: *Event-relevance Based Summarization*	
EXPERIENCE	**Postdoctoral Researcher, University of Pennsylvania**, Philadelphia, PA	Feb 2014 – Aug 2016
	Computer Information and Science Department (Advisor: Chris Callison-Burch)	
	Visiting Researcher, University of Washington, Seattle, WA	Jan 2012 – Dec 2013
	Computer Science and Engineering Department	
	Research Intern, Microsoft Research, Seattle, WA	Jun – Sep 2011
	Internet Services Research Center	
	Research Consultant, Educational Testing Service, Princeton, NJ	Mar – Dec 2010
	Natural Language and Speech Group	

GRANTS & AWARDS

CrowdFlower AI for Everyone Award (sole PI: estimated ∼$150,000 award value)   2018
DARPA Grant Computational Simulation of Online Social Behavior (co-PI 50% of $600,000) 2017 – 2021
NYU MacCracken Fellowship (total $230,000 award)   2007 – 2012

SERVICE

Workshop Chair: ACL (2017); **Publicity Chair:** NAACL (2016)
Area Chair: COLING (2018), EMNLP (2016)
NSF Review Panelist: CISE (2017); **Journal Reviewer:** TACL, JAIR
Program Committee: ACL (2018, 2017, 2015, 2014, 2013), EMNLP (2017, 2015, 2014), NAACL (2018, 2015), WWW (2017, 2016, 2015), AAAI (2016, 2015, 2012), KDD (2015), COLING (2014)
Organizer: Workshop on Noisy User-generated Text (EMNLP 2018 & 2017, COLING 2016, ACL 2015); Shared-task on Paraphrase Identification and Semantic Similarity in Twitter (SemEval 2015)

SELECTED PUBLICATIONS

A Continuously Growing Dataset of Sentential Paraphrases
Wuwei Lan, Siyu Qiu, Hua He, Wei Xu
Proceedings of EMNLP 2017, long paper (acceptance rate 25.8%)

A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter
Jeniya Tabassum, Alan Ritter, Wei Xu
Proceedings of EMNLP 2016, long paper (acceptance rate 26%), oral presentation

Optimizing Statistical Machine Translation for Simplification
Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, Chris Callison-Burch
TACL 2016; oral presentation at ACL 2016

Discovering User Attribute Stylistic Differences via Paraphrasing
Daniel Preotiuc, Wei Xu, Lyle Ungar
Proceedings of AAAI 2016, long paper (acceptance rate 26%)

Problems in Current Text Simplification Research: New Data Can Help
Wei Xu, Chris Callison-Burch, Courtney Napoles
TACL 2015; oral presentation at EMNLP 2015

Cost Optimization for Crowdsourcing Translation
Mingkun Gao, Wei Xu, Chris Callison-Burch
Proceedings of NAACL 2015, long paper (acceptance rate 29%)

SemEval-2015 Task 1: Paraphrase and Semantic Similarity in Twitter
Wei Xu, Chris Callison-Burch, William B. Dolan
Proceedings of SemEval 2015

Extracting Lexically Divergent Paraphrases from Twitter
Wei Xu, Alan Ritter, Chris Callison-Burch, William B. Dolan, Yangfeng Ji
TACL 2014; oral presentation at NAACL 2015

Infusion of Labeled Data into Distant Supervision for Relation Extraction
Maria Pershina, Bonan Min, Wei Xu, Ralph Grishman
Proceedings of ACL 2014, short paper (acceptance rate 25.2%)

Filling Knowledge Base Gaps for Distant Supervision of Relation Extraction
Wei Xu, Raphael Hoffmann, Le Zhao, Ralph Grishman
Proceedings of ACL 2013, short paper (acceptance rate 24%)

Gathering and Generating Paraphrases from Twitter with Application to Normalization
Wei Xu, Alan Ritter, Ralph Grishman
Proceedings of ACL Workshop on Building and Using Comparable Corpora 2013

Paraphrasing for Style
Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, Colin Cherry
Proceedings of COLING 2012, long paper (acceptance rate 25%)

Exploiting Syntactic and Distributional Information for Spelling Correction with Web-Scale N-grams Models
Wei Xu, Joel Tetreault, Martin Chodorow, Ralph Grishman, Le Zhao
Proceedings of EMNLP 2011, long paper (acceptance rate 23.7%)

Passage Retrieval for Information Extraction using Distant Supervision
Wei Xu, Ralph Grishman, Le Zhao
Proceedings of IJCNLP 2011, long paper

Extractive Summarization using Inter- and Intra- Event Relevance
Wenjie Li, Wei Xu, Mingli Wu, Chunfa Yuan, Qin Lu
Proceedings of COLING-ACL 2006, long paper, oral presentation (acceptance rate 23%)

SELECTED OPEN SOURCE CODE & DATA	*A Continuously Growing Dataset of Sentential Paraphrases* https://github.com/lanwuwei/paraphrase-dataset/	Sep 2017
	Text Simplification System and Newsela Corpus https://github.com/cocoxu/simplification/; https://newsela.com/data/	May 2015
	Multi-instance Learning Paraphrase Model and SemEval Twitter Paraphrase Corpus https://www.cis.upenn.edu/~xwe/multip/; http://alt.qcri.org/semeval2015/task1/	Dec 2014
	Parallel Shakespeare Corpus and Model https://github.com/cocoxu/Shakespeare/	Jul 2012

STUDENT ADVISING

Wuwei Lan (3rd year PhD student at OSU)
Mounica Maddela (1st year PhD student at OSU)
Shuaichen Chang (1st year PhD student at OSU)
Pravar Mahajan (masters student at OSU)
Piyush Ghai (masters student at OSU)

INVITED TALKS	*Learning Large-scale Paraphrases for Natural Language Understanding and Generation* IBM Thomas J. Watson Research Center, New York	Nov 2017
	Can Paraphrase be a Ultimate Solution for NLU and NLG? Google Research, New York	July 2017
	Paraphrase $\approx$ Monolingual Translation Amazon.com, Berlin, Germany	Aug 2016
	Multiple Instance Learning from Unlimited Text Microsoft Research Asia; University of Delaware; University of Edinburgh; University of North Carolina, Chapel Hill; Arizona State University; Vanderbilt University; Imperial College London; University of Waterloo; Indiana University, Bloomington; Washington University, St Louis; Simon Fraser University; University of Alberta; Yale University; University of Maryland	Oct 2015 – Dec 2016
	Learning and Generating Paraphrases from Twitter and Beyond Carnegie Mellon University; Columbia University; Johns Hopkins University	Feb 2015
	Paraphrases in Twitter Twitter.com, San Francisco	Feb 2015
	Modeling Lexically Divergent Paraphrases in Twitter (and Shakespeare!) City University of New York; IBM Research, Almaden; University of California, Berkeley; University of Texas, Austin; Yahoo!, New York; Carnegie Mellon University; Microsoft Research, Redmond	Aug 2014 – Mar 2015