

# 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/>

---

<b>RESEARCH</b>	Natural Language Processing, Machine Learning, Social Media	
<b>EMPLOYMENT</b>	<b>Assistant Professor, Ohio State University</b> , Columbus, OH <i>Department of Computer Science and Engineering</i>	Aug 2016 – present
	<b>Postdoctoral Researcher, University of Pennsylvania</b> , Philadelphia, PA <i>Computer Information and Science Department</i> (Advisor: Chris Callison-Burch)	Feb 2014 – Aug 2016
	<b>Visiting Researcher, University of Washington</b> , Seattle, WA <i>Computer Science and Engineering Department</i>	Jan 2012 – Dec 2013
	<b>Research Intern, Microsoft Research</b> , Seattle, WA <i>Internet Services Research Center</i>	Jun – Sep 2011
	<b>Research Consultant, Educational Testing Service</b> , Princeton, NJ <i>Natural Language and Speech Group</i>	Mar – Dec 2010
<b>EDUCATION</b>	<b>Ph.D. in Computer Science, New York University</b> , New York, NY Thesis: <i>Data-driven Approaches for Paraphrasing Across Language Variations</i> (Advisor: Ralph Grishman)	2014
	<b>B.S. and M.S. in Computer Science, Tsinghua University</b> , Beijing, CHINA Thesis: <i>Event-relevance Based Summarization</i>	2004/2007
<b>AWARDS &amp; GRANTS</b>	<b>NSF CRII Award RI: Learning a Timely Semantic Resource from Social Media Data</b> Sole PI, total \$175,000	2018 – 2020
	<b>CrowdFlower AI for Everyone Award</b> (\$25,000)	2018 – 2019
	<b>DARPA Research Grant</b> Computational Simulation of Online Social Behavior co-PI (50% share), total \$600,000	2017 – 2011
	<b>NYU MacCracken Fellowship</b> (total \$230,000)	2007 – 2012
<b>SELECTED PUBLICATIONS</b>	<i>Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering</i> Wuwei Lan, Wei Xu Proceedings of COLING 2018, long paper (acceptance rate 37.4%)  <i>The Importance of Subword Embeddings in Sentence Pair Modeling</i> Wuwei Lan, Wei Xu Proceedings of NAACL 2018, short paper (acceptance rate 29%)  <i>Taking the First Essential Steps in Automating the Wet Laboratory: Annotating a Corpus of Protocols for Reproducibility</i> Chaitanya Kulkarni, Wei Xu, Alan Ritter, Raghu Machiraju Proceedings of NAACL 2018, short paper (acceptance rate 29%)  <i>A Continuously Growing Dataset of Sentential Paraphrases</i> Wuwei Lan, Siyu Qiu, Hua He, Wei Xu Proceedings of EMNLP 2017, long paper (acceptance rate 25.8%)  <i>A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter</i> Jeniya Tabassum, Alan Ritter, Wei Xu Proceedings of EMNLP 2016, long paper (acceptance rate 26%), oral presentation  <i>Optimizing Statistical Machine Translation for Simplification</i> Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, Chris Callison-Burch TACL 2016; oral presentation at ACL 2016  <i>Discovering User Attribute Stylistic Differences via Paraphrasing</i> Daniel Preotiuc, Wei Xu, Lyle Ungar Proceedings of AAAI 2016, long paper (acceptance rate 26%)  <i>Problems in Current Text Simplification Research: New Data Can Help</i> 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 (acceptance rate 23%), oral presentation

## SERVICE

**Workshop Chair:** ACL (2017); **Publicity Chair:** NAACL (2018, 2016)

**Area Chair:** COLING (2018), EMNLP (2018, 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)

## INVITED TALKS

*Learning Large-scale Paraphrases for Natural Language Understanding and Generation* Nov 2017 - 2018  
Midwest Machine Learning Symposium, Chicago; Facebook, Menlo Park; Twitter, San Francisco; IBM Thomas J. Watson Research Center

*How AI Understand Language?*

Women in Analytics Conference (main stage panel), Columbus, OH

Mar 2018

*Can Paraphrase be a Ultimate Solution for NLU and NLG?*

Google Research, New York

July 2017

*Paraphrase  $\approx$  Monolingual Translation*

Amazon, Berlin, Germany

Aug 2016

*Multiple Instance Learning from Unlimited Text*

Oct 2015 – Dec 2016  
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

*Learning and Generating Paraphrases from Twitter and Beyond*

Carnegie Mellon University; Columbia University; Johns Hopkins University

Feb 2015

*Paraphrases in Twitter*  
Twitter, San Francisco

Feb 2015

*Modeling Lexically Divergent Paraphrases in Twitter (and Shakespeare!)* Aug 2014 – Mar 2015  
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

## TEACHING

*CSE 5522 Artificial Intelligence II: Advanced Techniques* Spring 2018  
[https://cocoxu.github.io/courses/5522\\_spring18.html](https://cocoxu.github.io/courses/5522_spring18.html)

**Ohio State University**, mixed undergraduate and graduate level, teaching evaluation 4.5 out of 5

*CSE 5525 Speech and Language Processing* Spring 2017  
[https://cocoxu.github.io/courses/5525\\_spring17.html](https://cocoxu.github.io/courses/5525_spring17.html)

**Ohio State University**, mixed undergraduate and graduate level, teaching evaluation 3.8 out of 5

*CSE 5539 Social Media and Text Analytics* Summer 2016, Autumn 2016, Autumn 2017  
<http://socialmedia-class.org/>

Designed a new course integrated with research, covering from basic to state-of-the-art machine learning algorithms. The course material I created has been used by other universities (e.g., University of Utah CS 1060, Georgia Tech CS 6452, Iowa State University MIS 546).

**Ohio State University**, mixed undergraduate and graduate level, teaching evaluation 4.6 out of 5 (Autumn 2016), 4.4 out of 5 (Autumn 2017)

**North American Summer School on Logic, Language, and Information (NASSLLI)**, interdisciplinary graduate level, teaching evaluation: 5.72 out of 6 (average across instructors was 5.23)