

# 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>POSITION</b>	<b>Assistant Professor, Ohio State University</b> , Columbus, OH <i>Department of Computer Science and Engineering</i> Research Areas: Natural Language Processing, Machine Learning, Social Media	Aug 2016 – present
<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>EXPERIENCE</b>	<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 – May, Sep – Dec 2010
<b>GRANTS AWARDS</b>	<b>NSF CRII Award RI: Learning a Timely Semantic Resource from Social Media Data</b> Sole PI, total \$175,000	2018 – 2020
	<b>DARPA Grant Computational Simulation of Online Social Behavior</b> co-PI (50% share), total \$600,000	2017 – 2011
	<b>CrowdFlower AI for Everyone Award</b> (estimated ~\$150,000 value)	2018 – 2019
	<b>NYU MacCracken Fellowship</b> (total \$230,000)	2007 – 2012
<b>SERVICE</b>	<b>Workshop Chair:</b> ACL (2017); <b>Publicity Chair:</b> NAACL (2018, 2016) <b>Area Chair:</b> COLING (2018), EMNLP (2016) <b>NSF Review Panelist:</b> CISE (2017); <b>Journal Reviewer:</b> TACL, JAIR <b>Program Committee:</b> ACL (2018, 2017, 2015, 2014, 2013), EMNLP (2017, 2015, 2014), NAACL (2015), WWW (2017, 2016, 2015), AAI (2016, 2015, 2012), KDD (2015), COLING (2014) <b>Organizer:</b> 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)	
<b>SELECTED PUBLICATIONS</b>	<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 Transactions of the Association for Computational Linguistics (TACL) 2016; oral presentation at ACL  <i>Discovering User Attribute Stylistic Differences via Paraphrasing</i> Daniel Preotiuc, Wei Xu, Lyle Ungar Proceedings of AAI 2016, long paper (acceptance rate 26%)	

*Problems in Current Text Simplification Research: New Data Can Help*

Wei Xu, Chris Callison-Burch, Courtney Napoles

Transactions of the Association for Computational Linguistics (TACL) 2015; oral presentation at EMNLP

*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, shared-task overview paper

*Extracting Lexically Divergent Paraphrases from Twitter*

Wei Xu, Alan Ritter, Chris Callison-Burch, William B. Dolan, Yangfeng Ji

Transactions of the Association for Computational Linguistics (TACL) 2014; oral presentation at NAACL

*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%)

*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%)

*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**

*A Continuously Growing Dataset of Sentential Paraphrases*

Sep 2017

**OPEN SOURCE  
CODE & DATA**

<https://github.com/lanwuwei/paraphrase-dataset/>

*Text Simplification System and Newsela Corpus*

May 2015

<https://github.com/cocoxu/simplification/>; <https://newsela.com/data/>

*Multi-instance Learning Paraphrase Model and SemEval Twitter Paraphrase Corpus*

Dec 2014

<https://www.cis.upenn.edu/~xwe/multi/>; <http://alt.qcri.org/semeval2015/task1/>

**INVITED  
TALKS**

How AI Understand Language?

Mar 2018

Women in Analytics Conference, Columbus, OH

*Learning Large-scale Paraphrases for Natural Language Understanding and Generation*

Nov 2017

IBM Thomas J. Watson Research Center, New York

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

July 2017

Google Research, New York

*Paraphrase  $\approx$  Monolingual Translation*

Aug 2016

Amazon.com, Berlin, Germany

*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*

Feb 2015

Carnegie Mellon University; Columbia University; Johns Hopkins University

*Paraphrases in Twitter*

Feb 2015

Twitter.com, San Francisco

*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

<b>STUDENT RESEARCH ADVISING</b>	Wuwei Lan (PhD student at OSU - phd thesis research)	Autumn 2016 - present
	Mounica Maddela (PhD student at OSU - phd thesis research)	Autumn 2017 - present
	Chaitanya Kulkarni (PhD student at OSU - primary advisor: Raghu Machiraju)	Spring 2017 - present
	Serena Davis (Masters student at OSU - research project)	Summer 2018 - present
	Shuaichen Chang (PhD student at OSU - research project)	Autumn 2017 - Spring 2018
	Pravar Mahajan (Masters student at OSU - masters project)	Winter 2016 - Autumn 2017
	Piyush Ghai (Masters student at OSU - independent study)	Autumn 2017
	Gordon Moon (PhD student at OSU - independent study)	Spring 2017
	Jeniya Tabassum (PhD student at OSU - primary advisor: Alan Ritter)	Summer 2016
	Wenchao Du (completed undergraduate student at UWaterloo, now a Masters student at CMU LTI)	2016
	Mingkun Gao (completed Masters student at UPenn, now a PhD student at UIUC)	2015
	Siyu Qiu (completed Masters student at UPenn, now at Hulu.com)	2015
	Jim Chen (completed undergraduate student at UPenn, now a PhD student at UWashington)	2014 - 2015
	Ray Lei (completed undergraduate student at UPenn, now at Microsoft Redmond)	2014
	Maria Pershina (completed PhD student at NYU, now at Goldman Sachs NYC)	2014
	<b>TEACHING</b>	<b>Instructor, The Ohio State University</b> , Columbus, OH
Course: <i>CSE 5522 Artificial Intelligence II: Advanced Techniques</i> (mixed undergraduate and graduate level) <a href="https://cocoxu.github.io/courses/5522_spring18.html">https://cocoxu.github.io/courses/5522_spring18.html</a>		
<b>Instructor, The Ohio State University</b> , Columbus, OH		Spring 2017
Course: <i>CSE 5525 Speech and Language Processing</i> (mixed undergraduate and graduate level) <a href="https://cocoxu.github.io/courses/5525_spring17.html">https://cocoxu.github.io/courses/5525_spring17.html</a>		
<b>Instructor, The Ohio State University</b> , Columbus, OH		Autumn 2016, 2017
Course: <i>CSE 5539 Social Media and Text Analytics</i> (mixed undergraduate and graduate level) <a href="http://socialmedia-class.org/">http://socialmedia-class.org/</a> 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).		
<b>Instructor</b> , New Brunswick, NJ		Summer 2016
<b>North American Summer School on Logic, Language, and Information (NASSLLI)</b> Course: <i>Social Media and Text Analytics</i> (interdisciplinary graduate level)		
<b>Instructor, University of Pennsylvania</b> , Philadelphia, PA	Summer 2015	
Course: <i>Social Media and Text Analytics</i> (interdisciplinary graduate level)		
<b>Guest Lecturer, Rutgers University</b> , New Brunswick, NJ	Spring 2015	
Course: <i>Natural Language Processing</i>		
<b>Guest Lecturer, University of Pennsylvania</b> , Philadelphia, PA	Spring 2014	
Course: <i>Advanced Topics in Natural Language Processing</i>		