

# 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>CURRENT POSITION</b>	<b>Assistant Professor, Ohio State University</b> , Columbus, OH <i>Department of Computer Science and Engineering</i>	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; Committee: Satoshi Sekine, Ernest Davis, Bill Dolan (Microsoft Research), Luke Zettlemoyer (University of Washington/Facebook AI)	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>RESEARCH</b>	Natural Language Processing, Machine Learning, Social Media	
<b>CITIZENSHIP</b>	United States	
<b>EXPERIENCE</b>	<b>Visiting Faculty, Carnegie Mellon University</b> , Pittsburgh, PA <i>Language Technologies Institute</i> (Host: Graham Neubig)	Summer 2019
	<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>	Summer 2011
	<b>Research Intern, Educational Testing Service</b> , Princeton, NJ <i>Natural Language and Speech Group</i>	Mar – Dec 2010
	<b>Engineer Intern, Amazon.com</b> , Seattle, WA <i>Catalog Quality Group</i>	Summer 2010
	<b>Visiting Researcher, Hong Kong Polytechnic University</b> <i>Department of Computing</i>	Sep 2005 – Feb 2006
<b>AWARDS</b>	<b>Best Paper Award, COLING</b> , 2018 <b>Criteo Faculty Research Award</b> (\$20,000), 2018 <b>CrowdFlower AI for Everyone Award</b> (\$25,000), 2018 <b>NSF CRII Award</b> , 2018 <b>NYU MacCracken PhD Fellowship</b> , 2007 – 2012	
<b>GRANTS</b>	<b>NSF Grant</b> , <i>Collaborative Research: Automatic Text-Simplification and Reading-Assistance to Support Self-Directed Learning by Deaf and Hard-of-Hearing Computing Workers</i> PI (100%), total \$375,732 (including \$8,000 REU supplement)	2018 – 2021
	<b>NSF CRII RI</b> : Learning a Timely Semantic Resource from Social Media Data PI (100%), total \$175,000	2018 – 2020
	<b>IARPA Research Grant</b> Better Extraction from Text Towards Enhanced Retrieval co-PI (50%), total \$850,000 (prime: Brown University)	2019 – 2023
	<b>DARPA Research Grant</b> Computational Simulation of Online Social Behavior co-PI (50%), total \$600,000 (prime: Leidos)	2017 – 2021
<b>SERVICES</b>	<b>Best Paper Award Committee</b> : EMNLP (2018) <b>Area Chair</b> : AAAI (2020), ACL (2019), NAACL (2019), COLING (2018), EMNLP (2018, 2016) <b>Program Committee</b> : ACL (2018, 2017, 2015, 2014, 2013), EMNLP (2017, 2015, 2014), NAACL (2015) WWW (2017, 2016, 2015), AAAI (2016, 2015, 2012), KDD (2015), COLING (2014) <b>Session Chair</b> : ACL (2019, 2014), NAACL (2019, 2015), EMNLP (2018, 2017, 2016, 2015), COLING (2018), AAAI (2015) <b>Publicity Chair</b> : EMNLP (2019), NAACL (2018, 2016) <b>Workshop Chair</b> : ACL (2017) <b>Steering Committee</b> : SemEval (2017) <b>NSF Review Panelist</b> : CISE (2017)	

**Journal Reviewer:** Transactions of the Association for Computational (TACL), Journal of Artificial Intelligence Research (JAIR)

**Organizer:** Workshop on Noisy User-generated Text (<http://noisy-text.github.io/>) at EMNLP (2019, 2018, 2017), COLING (2016), ACL (2015)

SemEval Task 1: Paraphrase Identification and Semantic Similarity in Twitter (2015)

Mid-Atlantic Student Colloquium on Speech, Language and Learning (2016)

## PUBLICATIONS

(Underline is used to indicate student advisees at the Ohio State University. TACL is one of the most selective venues for NLP publications, which allows authors to choose any of the ACL/NAACL/EMNLP conferences to present the paper.)

*Multi-task Pairwise Neural Ranking for Hashtag Segmentation*

Mounica Maddela, Wei Xu, Daniel Preotiuc-Pietro

Proceedings of ACL 2019, long paper (acceptance rate 25.7%)

*A Word-Complexity Lexicon and A Neural Readability Ranking Model for Lexical Simplification*

Mounica Maddela, Wei Xu

Proceedings of EMNLP 2018, long paper (acceptance rate 25.8%; oral presentation)

*Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering*

Wuwei Lan, Wei Xu

Proceedings of COLING 2018, long paper (**Best Paper Award**; selection rate  $8/888 = 0.90\%$ )

*An Annotated Corpus for Machine Reading of Instructions in Wet Lab Protocols*

Chaitanya Kulkarni, Wei Xu, Alan Ritter, Raghu Machiraju

NAACL 2018, short paper (acceptance rate 29%)

*The Importance of Subword Embeddings in Sentence Pair Modeling*

Wuwei Lan, Wei Xu

NAACL 2018, short paper (acceptance rate 29%)

*A Continuously Growing Dataset of Sentential Paraphrases*

Wuwei Lan, Siyu Qiu, Hua He, Wei Xu

EMNLP 2017, long paper (acceptance rate 25.8%)

*From Shakespeare to Twitter: What are Language Styles all about?*

Wei Xu

EMNLP 2017 Workshop on Stylistic Variation

*A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter*

Jeniya Tabassum, Alan Ritter, Wei Xu

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, long paper (oral presentation at ACL 2016)

*Discovering User Attribute Stylistic Differences via Paraphrasing*

Daniel Preotiuc-Pietro, Wei Xu, Lyle Ungar

AAAI 2016, long paper (acceptance rate 26%; oral presentation)

*Results of the WNUT16 Named Entity Recognition Shared Task*

Benjamin Strauss, Bethany Toma, Alan Ritter, Marie-Catherine de Marneffe, Wei Xu

COLING 2016 Workshop on Noisy User-generated Text (shared-task overview)

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

Wei Xu, Chris Callison-Burch, Courtney Napoles

TACL 2015, long paper (oral presentation at EMNLP 2015)

*Cost Optimization for Crowdsourcing Translation*

Mingkun Gao, Wei Xu, Chris Callison-Burch

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

SemEval 2015 (shared-task overview)

*Extracting Lexically Divergent Paraphrases from Twitter*

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

TACL 2014, long paper (oral presentation at NAACL 2015)

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

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

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

*A Preliminary Study of Tweet Summarization using Information Extraction*  
 Wei Xu, Ralph Grishman, Adam Meyers, Alan Ritter  
 NAACL Workshop on Language Analysis in Social Media 2013

*Paraphrasing for Style*  
 Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, Colin Cherry  
 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  
 EMNLP 2011, long paper (acceptance rate 23.7%)

*New York University 2011 System for KBP (Knowledge Base Population) Slot Filing*  
 Ang Sun, Ralph Grishman, Wei Xu, Bonan Min  
 TAC 2011 (best performance system in NIST KBP-2011 Evaluation)

*Passage Retrieval for Information Extraction using Distant Supervision*  
 Wei Xu, Ralph Grishman, Le Zhao  
 IJCNLP 2011, long paper (acceptance rate 36%)

*Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task*  
 Kristen Parton, Kathleen McKeown, Bob Coyne, Mona Diab, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Heng Ji, Weiyun Ma, Adam Meyers, Sara Stolbach, Ang Sun, Gokhan Tur, Wei Xu, Sibel Yaman  
 ACL 2009, long paper (acceptance rate 21%; oral presentation)

*A Parse-and-Trim Approach with Information Significance for Chinese Sentence Compression*  
 Wei Xu, Ralph Grishman  
 ACL Workshop on Language Generation and Summarisation 2009

*Transducing Logical Relations from Automatic and Manual Annotation*  
 Adam Meyers, Michiko Kosaka, Heng Ji, Nianwen Xue, Mary Harper, Ang Sun, Wei Xu, Shasha Liao  
 ACL Workshop on Linguistic Annotation 2009

*Automatic Recognition of Logical Relations for English, Chinese and Japanese in the GLARF Framework*  
 Adam Meyers, Michiko Kosaka, Nianwen Xue, Heng Ji, Ang Sun, Shasha Liao, Wei Xu  
 SemEval 2009

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

*Using Non-Local Features to Improve Named Entity Recognition Recall*  
 Xinnian Mao, Wei Xu, Yuan Dong, Haila Wang  
 PACLIC 2007

*Deriving Event Relevance from the Ontology Constructed with Formal Concept Analysis*  
 Wei Xu, Wenjie Li, Mingli Wu, Wei Li, Chunfa Yuan  
 CICLing 2006, long paper (acceptance rate 30.4%; oral presentation)

*Building Document Graph for Text Summarization: An Event-based Approach*  
 Wei Xu, Wenjie Li, Mingli Wu, Wei Li, Chunfa Yuan  
 ICCPOL 2006

*The THU/PolyU System at MSE 2006: An Event-relevance based Approach*  
 Wei Xu, Chunfa Yuan, Mingling Wu, Wenjie Li  
 MSE 2006

## STUDENTS

Wuwei Lan (PhD student at OSU)	2016 – present
Mounica Maddela (PhD student at OSU)	2017 – present
Chao Jiang (PhD student at OSU)	2018 – present
Yang Zhong (PhD student at OSU)	2019 – present

	Vishal Sunder (PhD student at OSU)	2019 – present
	Jeniya Tabassum (PhD student at OSU - co-advised with Alan Ritter)	2016 – present
	Sydney Lee (undergraduate at OSU)	2018 – present
	Sarah Flanagan (undergraduate at OSU)	2018 – present
	Nabeel Tausif (undergraduate at OSU)	2019 – present
	Raleigh Potluri (undergraduate student at OSU)	2018 – 2019
	Lillian Chow (undergraduate student at OSU)	2018 – 2019
	Rita Tong (completed undergraduate student at OSU)	2018 – 2019
	Pravar Mahajan (completed Masters student at OSU, now IBM)	2016 – 2017
	Piyush Ghai (completed Masters student at OSU, now Amazon)	2017
	Mingkun Gao (completed Masters student at UPenn, now a PhD student at UIUC)	2015
	Siyu Qiu (completed Masters student at UPenn, now at Hulu)	2015
	Jim Chen (completed undergraduate student at UPenn, now a PhD student at UW)	2014 – 2015
	Ray Lei (completed undergraduate student at UPenn, now at Microsoft Redmond)	2014
	Chaitanya Kulkarni (PhD student at OSU - advisor: Raghu Machiraju)	2017 – 2019
	Sanqiang Zhao (PhD student at UPitt - phd thesis committee)	2019
	Kai Cao (completed PhD student at NYU - phd thesis committee)	2017
	Maria Pershina (completed PhD student at NYU, now at Bloomberg - phd thesis committee)	2014
<b>TEACHING</b>	<i>CSE 5522 Artificial Intelligence II: Advanced Techniques</i>	Spring 2018, Autumn 2018
	Ohio State University, mixed undergraduate and graduate level (teaching evaluation: 4.85/5.00 Autumn 2018, 4.50/5.00 Spring 2018)	
	<i>CSE 5525 Speech and Language Processing</i>	Spring 2017
	Ohio State University, mixed undergraduate and graduate level (teaching evaluation 3.80/5.00 Spring 2017)	
	<i>CSE 5539 Social Media and Text Analytics</i> <a href="http://socialmedia-class.org/">http://socialmedia-class.org/</a> A new course integrated with research, covering from basic to state-of-the-art machine learning algorithms (teaching evaluation: 4.40/5.00 Autumn 2017, 4.60/5.00 Autumn 2016, 5.72/6.00 Summer 2015)	Summer 2016, Autumn 2016, Autumn 2017
<b>INVITED TALKS</b>	<b>Learning for Unlimited Human Language</b>	
	Peking University, Beijing, China	Dec 2018
	<b>Learning Large-scale Paraphrases for Natural Language Understanding and Generation</b>	
	Midwest Machine Learning Symposium, Chicago, IL	Jun 2018
	Facebook, Menlo Park, CA	May 2018
	Stanford Research Institute, Menlo Park, CA	May 2018
	Twitter, San Francisco, CA	May 2018
	IBM Thomas J. Watson Research Center, New York, NY	Nov 2017
	<b>How does AI Understand Language?</b>	
	Women in Analytics Conference, Columbus, OH (Main Stage Panel)	Mar 2018
	<b>Can Paraphrase be a Ultimate Solution for NLU and NLG?</b>	
	Google Research, New York, NY	Jul 2017
	<b>Paraphrase <math>\approx</math> Monolingual Translation</b>	
	Amazon, Berlin, Germany	Aug 2016
	<b>Multiple Instance Learning from Unlimited Text</b>	
	Microsoft Research Asia, Beijing, China	Dec 2016
	University of Delaware, Newark, DE	Sep 2016
	University of Edinburgh, Edinburgh, United Kingdom	May 2016
	Ohio State University, Columbus, OH	Apr 2016
	University of North Carolina, Chapel Hill, NC	Apr 2016
	Arizona State University, Tempe, AZ	Mar 2016
	Vanderbilt University, Nashville, TN	Mar 2016
	Imperial College London, London, United Kingdom	Mar 2016
	University of Waterloo, Waterloo, ON, Canada (CS Seminar)	Mar 2016
	Indiana University, Bloomington, IN (Computer Science Colloquium Series)	Feb 2016
	Washington University, St Louis, MI (Computer Science & Engineering Colloquia Series)	Feb 2016
	Simon Fraser University, Vancouver, BC, Canada	Feb 2016
	University of Alberta, Edmonton, AB, Canada	Feb 2016
	Yale University, New Haven, CT (CS Talk)	Feb 2016
	University of Maryland, College Park, MD (CLIP Colloquium)	Oct 2015
	Ohio State University, Columbus, OH (Clippers Seminar)	Oct 2015

<b>OUTREACH ACTIVITIES</b>	<b>Large-scale Paraphrase Acquisition from Twitter</b>	
	DARPA's DEFT Project Meeting, Boulder, CO	May 2015
	<b>Learning and Generating Paraphrases from Twitter and Beyond</b>	
	Carnegie Mellon University, Pittsburgh, PA	Apr 2015
	Columbia University, New York, NY (NLP Talk)	Apr 2015
	Johns Hopkins University, Baltimore, MD (CLSP Colloquium)	Feb 2015
	<b>Paraphrases in Twitter</b>	
	Twitter, San Francisco, CA	Feb 2015
	<b>Modeling Lexically Divergent Paraphrases in Twitter (and Shakespeare!)</b>	
	The City University of New York, New York, NY (NLP Seminar)	Mar 2015
	IBM Research - Almaden, San Jose, CA	Feb 2015
	University of California, Berkeley, CA	Feb 2015
	The University of Texas, Austin, TX (Forum for Artificial Intelligence)	Feb 2015
	Yahoo!, New York, NY	Dec 2014
	Carnegie Mellon University, Pittsburgh, PA (CL+NLP Lunch Seminar)	Nov 2014
	Microsoft Research, Seattle, WA (Visiting Speaker Series)	Aug 2014
	<b>Incremental Information Extraction</b>	
	Stanford Research Institute, Palo Alto, CA	Apr 2012
	IARPA's KDD Project PI Meeting, San Diego, CA	May 2011
	<b>Information Extraction Research</b>	
	University of Washington, Seattle, WA	Jan 2011
	<b>Event-based Summarization</b>	
	Thomson Reuters, Eagan, Minnesota, MN	Nov 2009
	Speaker/Judge, Ohio High School Hackathon	Mar 2019
	Speaker, Franklin Friday art and science festival in Columbus Ohio	Mar 2019
	Panelist, CogFest - Cognitive Science Festival	Apr 2018
	Mentor, Women and Underrepresented Minorities in NLP Workshop	Jun 2018
	Mentor, OSU's AI Hackathon	Apr 2018
	Speaker/Panelist, Women in Analytics Conference	Mar 2018
	Speaker, OSU's AI Club	Feb 2018
	Judge, HackOhio	Oct 2017
	Mentor, Women and Underrepresented Minorities in NLP Workshop	Jul 2017
	Judge, Ohio High School Hackathon	Mar 2017
	Volunteer, Philadelphia Science Festival	Apr 2015
<b>OPEN SOURCE CODE / DATA</b>	<i>#HashtagMaster: A Semantic Analysis Tool for Hashtags</i>	Jun 2019
	<a href="https://mounicam.github.io/hashtag_master">https://mounicam.github.io/hashtag_master</a>	
	<i>Pairwise Neural Ranking Model and SimplePPDB++</i>	Oct 2018
	<a href="https://github.com/lanwuwei/SPM_toolkit">https://github.com/lanwuwei/SPM_toolkit</a>	
	<i>SPM Toolkit for Sentence Pair Modeling</i>	Aug 2018
	<a href="https://github.com/lanwuwei/SPM_toolkit">https://github.com/lanwuwei/SPM_toolkit</a>	
	<i>LanguageNet: Large-scale Paraphrase Corpus</i>	Sep 2017
	<a href="https://github.com/lanwuwei/paraphrase-dataset">https://github.com/lanwuwei/paraphrase-dataset</a>	
	<i>Syntax MT-based Text Simplification System and SARI Evaluation Metric</i>	May 2015
	<a href="https://github.com/cocoxu/simplification/">https://github.com/cocoxu/simplification/</a> (contribution to the Joshua Machine Translation Toolkit)	
	<i>NEWSELA Text Simplification Corpus</i>	Sep 2015
	<a href="https://newseila.com/data/">https://newseila.com/data/</a> (widely adopted as the benchmark for text simplification research)	
	<i>Multiple-instance Learning Paraphrase Model</i>	Dec 2014
	<a href="https://github.com/cocoxu/multip">https://github.com/cocoxu/multip</a>	
	<i>Twitter Paraphrase Corpus</i> (shared-task at SemEval-2015)	Oct 2014
	<a href="http://alt.qcri.org/semeval2015/task1/">http://alt.qcri.org/semeval2015/task1/</a>	
	<i>Event-based Twitter Summarization System</i>	Nov 2013
	<a href="https://github.com/cocoxu/twittersummarization/">https://github.com/cocoxu/twittersummarization/</a>	
	<i>Twitter Normalization Phrase Table</i>	Oct 2014
	<a href="https://github.com/cocoxu/twitterparaphrase/">https://github.com/cocoxu/twitterparaphrase/</a>	
	<i>Parallel Shakespeare Corpus and Model</i>	Jul 2012
	<a href="https://github.com/cocoxu/Shakespeare/">https://github.com/cocoxu/Shakespeare/</a>	