

Wei Xu

✉ weixu@cse.ohio-state.edu ☎ +1 718 612 9347

The Ohio State University

Department of Computer Science and Engineering

395 Dreese Laboratories, 2015 Neil Avenue Columbus, OH, U.S.A. 43210

<http://web.cse.ohio-state.edu/~weixu/>

RESEARCH	Natural Language Processing, Machine Learning, Social Media	
CURRENT POSITION	Assistant Professor, Ohio State University , Columbus, OH <i>Department of Computer Science and Engineering</i>	Aug 2016–Present
EDUCATION	Ph.D. in Computer Science, New York University , New York, NY Thesis: Data-driven Approaches for Paraphrasing Across Language Variations	2014
	B.S. and M.S. in Computer Science, Tsinghua University , Beijing, CHINA Thesis: Event-relevance Based Summarization	2004/2007
EXPERIENCE	Postdoctoral Fellow, University of Pennsylvania , Philadelphia, PA <i>Computer Information and Science Department</i>	Feb 2014–Aug 2016
	Visiting Researcher, University of Washington , Seattle, WA <i>Computer Science and Engineering Department</i>	Jan 2012–Dec 2013
	Research Intern, Microsoft Research , Seattle, WA <i>Internet Services Research Center</i>	Jun–Sep 2011
	Consultant, Educational Testing Service , Princeton, NJ <i>Natural Language and Speech Group</i>	Mar–Dec 2010
SERVICES	Workshop Chair: ACL (2017) Area Chair: COLING (2018), EMNLP (2016) Publicity Chair: NAACL (2018, 2016) Session Chair: EMNLP (2017, 2016, 2015), NAACL (2015), AAAI (2015), ACL (2014) NSF Review Panelist: CISE (2017) Steering Committee: SemEval (2017) Program Committee: ACL (2017, 2015, 2014, 2013), EMNLP (2017, 2015, 2014), NAACL (2015) WWW (2017, 2016, 2015), AAAI (2016, 2015, 2012), KDD (2015), COLING (2014) WWW Workshop on #Microposts (2016) ACL Workshop on Social Factors in Natural Language Processing (2016) EACL Workshop on Language Analysis in Social Media (2014) Journal Reviewer: Transactions of the Association for Computational (TACL) Journal of Artificial Intelligence Research (JAIR) Workshop Organizer: Workshop on Noisy User-generated Text at ACL (2015), COLING (2016), EMNLP (2017) Shared-task Organizer: SemEval (2015): Paraphrase Identification and Semantic Similarity in Twitter (19 research teams participated; 200+ research groups worldwide have requested the data since)	
GRANTS	DARPA Grant Computational Simulation of Online Social Behavior (\$600,000 award)	2017–2021
	NYU MacCracken Fellowship (5 continuous years; \$230,000 award)	2007–2012
	Travel Grants: Google, Grace Hopper, CRA-Women	

PUBLICATIONS Refereed Journal and Conference Papers (all have acceptance rates \approx 25%)

A Continuously Growing Dataset of Sentential Paraphrases

Wuwei Lan, Siyu Qiu, Hua He, Wei Xu

Proceedings of EMNLP 2017, long paper

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

Jeniya Tabassum, Alan Ritter, Wei Xu

Proceedings of EMNLP 2016, long paper, oral presentation

Optimizing Statistical Machine Translation for Simplification

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, Chris Callison-Burch

Transactions of the Association for Computational (TACL), 4(1): pages 401-415, 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

Problems in Current Text Simplification Research: New Data Can Help

Wei Xu, Chris Callison-Burch, Courtney Napoles

Transactions of the Association for Computational (TACL), 3(1): pages 283-297, 2015
oral presentation at EMNLP 2015

Cost Optimization for Crowdsourcing Translation

Mingkun Gao, Wei Xu, Chris Callison-Burch

Proceedings of NAACL 2015, long 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 (TACL), 2(1): pages 435-448, 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

Filling Knowledge Base Gaps for Distant Supervision of Relation Extraction

Wei Xu, Raphael Hoffmann, Le Zhao, Ralph Grishman

Proceedings of ACL 2013, short paper

Paraphrasing for Style

Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, Colin Cherry

Proceedings of COLING 2012, long paper

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

Passage Retrieval for Information Extraction using Distant Supervision

Wei Xu, Ralph Grishman, Le Zhao

Proceedings of IJCNLP 2011, long paper

Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task

Kristen Parton, Kathleen R. McKeown, Bob Coyne, Mona T. Diab, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Heng Ji, Wei Yun Ma, Adam Meyers, Sara Stolbach, Ang Sun, Gokhan Tur, Wei Xu, Sibel Yaman

Proceedings of ACL-IJCNLP 2009, 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

Refereed Workshop Papers (all peer-reviewed)

From Shakespeare to Twitter: What are Language Styles all about?
Wei Xu
Proceedings of EMNLP Workshop on Stylistic Variation 2017

Results of the WNUT16 Named Entity Recognition Shared Task
Benjamin Strauss, Bethany Toma, Alan Ritter, Marie-Catherine de Marneffe, Wei Xu
Proceedings of COLING Workshop on Noisy User-generated Text 2016

Shared Tasks of the 2015 Workshop on Noisy User-generated Text: Twitter Lexical Normalization and Named Entity Recognition
Timothy Baldwin, Marie-Catherine de Marneffe, Bo Han, Young-Bum Kim, Alan Ritter, Wei Xu
Proceedings of ACL Workshop on Noisy User-generated Text 2015 (alphabetical author order)

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

Poetry of the Crowd: A Human Computation Algorithm to Convert Prose into Rhyming Verse
Quanze Chen, Chenyang Lei, Wei Xu, Ellie Pavlick, Chris Callison-Burch
Proceedings of HCOMP 2014

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

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

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

A Parse-and-Trim Approach with Information Significance for Chinese Sentence Compression
Wei Xu, Ralph Grishman
Proceedings of the ACL-IJNLP 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
Proceedings of ACL-IJNLP 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
Proceedings of NAACL-HLT Workshop on Semantic Evaluations 2009

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

Deriving Event Relevance from the Ontology Constructed with Formal Concept Analysis
Wei Xu, Wenjie Li, Mingli Wu, Wei Li, Chunfa Yuan
Proceedings of CICLing 2006

Building Document Graph for Text Summarization: An Event-based Approach

Wei Xu, Wenjie Li, Mingli Wu, Wei Li, Chunfa Yuan

Proceedings of ICCPOL 2006

The THU/PolyU System at MSE 2006: An Event-relevance based Approach

Wei Xu, Chunfa Yuan, Mingling Wu, Wenjie Li

Proceedings of MSE 2006

**STUDENT
RESEARCH
ADVISING**

Wuwei Lan (current PhD student at OSU)

Mounica Maddela (current PhD student at OSU)

Chaitanya Kulkarni (current PhD student at OSU; primary advisor: Raghu Machiraju)

Pravar Mahajan (current masters student at OSU)

Piyush Ghai (current masters student at OSU)

Mingkun Gao (completed masters student, now a PhD student at UIUC)

Wenchao Du (completed undergraduate student, now a masters student at CMU LTI)

Siyu Qiu (completed masters student, now at Hulu.com)

Jim Chen (completed undergraduate student, now a PhD student at University of Washington)

Bin Fu (completed undergraduate student, now at Google NYC)

Ray Lei (completed undergraduate student, now at Microsoft Redmond)

**INVITED
TALKS**

Can Paraphrase be a Ultimate Solution for NLU and NLG?

Google Research, New York, NY

July 2017

Paraphrase \approx Monolingual Translation

Amazon.com, Berlin, Germany

Aug 2016

Multiple Instance Learning from Unlimited Text

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

Large-scale Paraphrase Acquisition from Twitter

DARPA's DEFT Project PI Meeting, Boulder, CO

May 2015

Learning and Generating Paraphrases from Twitter and Beyond

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

Paraphrases in Twitter

Twitter.com, San Francisco, CA

Feb 2015

Modeling Lexically Divergent Paraphrases in Twitter (and Shakespeare!)

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
	Incremental Information Extraction	
	Stanford Research Institute, Palo Alto, CA	Apr 2012
	IARPA's KDD Project PI Meeting, San Diego, CA	May 2011
	Information Extraction Research	
	University of Washington, Seattle, WA	Jan 2011
	Event-based Summarization	
	Thomson Reuters, Eagan, Minnesota, MN	Nov 2009
PRESS	New Scientist - <i>One Per Cent</i>	Oct 29, 2014
	New York Times - <i>Government Aims to Build a "Data Eye in the Sky"</i>	Oct 11, 2011
OPEN SOURCE CODE / DATA	<i>A Continuously Growing Dataset of Sentential Paraphrases</i>	Sep 2017 – ongoing
	https://github.com/lanwuwei/paraphrase-dataset/	
	<i>Syntax MT-based Text Simplification System and Crowdsourced Dataset</i> (contribution to the Joshua MT Toolkit)	May 2015
	https://github.com/cocoxu/simplification/	
	<i>NEWSLA Text Simplification Corpus</i>	Sep 2015
	https://newsela.com/data/	
	<i>Multiple-instance Learning Paraphrase Model</i>	Dec 2014
	https://www.cis.upenn.edu/~xwe/multip/	
	<i>Twitter Paraphrase Corpus</i> (shared-task at SemEval-2015)	Oct 2014
	http://alt.qcri.org/semeval2015/task1/	
	<i>Event-based Twitter Summarization System</i>	Nov 2013
	https://github.com/cocoxu/twittersummarization/	
	<i>Twitter Normalization Phrase Table</i>	Oct 2014
	https://github.com/cocoxu/twitterparaphrase/	
	<i>Parallel Shakespeare Corpus and Model</i>	Jul 2012
	https://github.com/cocoxu/Shakespeare/	
TEACHING	Instructor, The Ohio State University , Columbus, OH	Spring 2017
	Course: <i>CSE 5525 Speech and Language Processing</i> (dual-listed undergraduate and graduate course)	
	https://cocoxu.github.io/courses/5525_spring17.html	
	Instructor, The Ohio State University , Columbus, OH	Autumn 2016, 2017
	Course: <i>CSE 5539 Social Media and Text Analytics</i> (dual-listed undergraduate and graduate course)	
	http://socialmedia-class.org/	
	Designed and taught a new course, covering from basic to state-of-the-art machine learning algorithms for text processing and social media analysis.	
	Instructor , New Brunswick, NJ	Summer 2016
	North American Summer School on Logic, Language, and Information (NASSLLI)	
	Course: <i>Social Media and Text Analytics</i> (interdisciplinary)	

REFERENCES

Chris Callison-Burch

Associate Professor, Computer and Information Science Department

University of Pennsylvania, Philadelphia, PA

Sloan Research Fellow (2014), Chair of the NAACL Executive Board (2012-2013)

☎ +1 267 909 2668

✉ ccb@cis.upenn.edu

<http://www.cis.upenn.edu/~ccb/>

Raymond J. Mooney

Professor, Computer Science Department, University of Texas, Austin, TX

AAAI fellow (2005), ACM fellow (2010), ACL fellow (2014)

☎ +1 512 471 9558

✉ mooney@cs.utexas.edu

<https://www.cs.utexas.edu/~mooney/>

Bill Dolan

Research Manager/Principal Researcher, Microsoft Research, Redmond, WA

Manager of the Natural Language Processing Group

☎ (425) 706-3709

✉ billdol@microsoft.com

<http://research.microsoft.com/en-us/people/billdol/>

Ralph Grishman

Professor, Computer Science Department, New York University, New York, NY

Department Chairman (1986-1988), President of ACL (1991)

☎ +1 212 998 3497

✉ grishman@cs.nyu.edu

<http://cs.nyu.edu/grishman/>

Joel Tetreault

Director of Research, Grammarly, New York, NY

Treasurer of NAACL (2013-present)

☎ +1 310 612 0315

✉ joel.tetreault@grammarly.com

<http://www.cs.rochester.edu/~tetreaul/academic.html>