

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/>

Last updated: April 4th, 2020

| | | |
|-------------------------|---|---------------------------|
| CURRENT POSITION | Assistant Professor, The 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: <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 Research) | 2014 |
| | B.S. and M.S. in Computer Science , Tsinghua University, Beijing, CHINA Thesis: <i>Event-relevance Based Summarization</i> | 2004/2007 |
| RESEARCH | Natural Language Processing, Machine Learning, Social Media | |
| CITIZENSHIP | United States | |
| EXPERIENCE | Visiting Faculty, Carnegie Mellon University , Pittsburgh, PA <i>Language Technologies Institute</i> (Host: Graham Neubig) | Summer 2019 |
| | Postdoctoral Researcher, University of Pennsylvania , Philadelphia, PA <i>Computer Information and Science Department</i> (Advisor: Chris Callison-Burch) | Feb 2014 – Aug 2016 |
| | Visiting PhD Student, 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> | Summer 2011 |
| | Research Intern, Educational Testing Service , Princeton, NJ <i>Natural Language and Speech Group</i> | Mar – May, Sep – Dec 2010 |
| | Engineer Intern, Amazon.com , Seattle, WA <i>Catalog Quality Group</i> | Summer 2010 |
| | Research Intern, Hong Kong Polytechnic University <i>Department of Computing</i> | Sep 2005 – Feb 2006 |
| AWARDS | Best Paper Award, COLING , 2018 Criteo Faculty Research Award (\$20,000), 2018 CrowdFlower AI for Everyone Award (\$25,000), 2018 NSF CRII Award , 2018 NYU MacCracken PhD Fellowship , 2007 – 2012 | |
| GRANTS | NSF Grant <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 |
| | NSF CRII RI: Learning a Timely Semantic Resource from Social Media Data PI (100%), total \$183,000 (including \$8,000 REU supplement) | 2018 – 2020 |
| | IARPA Research Grant <i>Better Extraction from Text Towards Enhanced Retrieval</i> co-PI (50%), total \$850,000 (prime: Brown University) | 2019 – 2023 |
| | DARPA Research Grant <i>Computational Simulation of Online Social Behavior</i> co-PI (50%), total \$600,000 (prime: Leidos) | 2017 – 2021 |
| SERVICES | Best Paper Award Committee: EMNLP (2018) Area Chair/Senior Area Chair: ACL (2020, 2019), AACL (2020), NAACL (2019), COLING (2018), EMNLP (2020, 2018, 2016) Program Committee: ACL (2018, 2017, 2015, 2014, 2013), EMNLP (2017, 2015, 2014), NAACL (2015) WWW (2017, 2016, 2015), AACL (2016, 2015, 2012), KDD (2015), COLING (2014) Publicity Chair: EMNLP (2019), NAACL (2018, 2016) Workshop Chair: ACL (2017) Session Chair: ACL (2019, 2014), NAACL (2019, 2015), EMNLP (2018, 2017, 2016, 2015), COLING (2018), AACL (2015) Steering Committee: SemEval (2017) | |

NSF Review Panelist: 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.)

Neural CRF Model for Sentence Alignment Model in Text Simplification

Chao Jiang, Mounica Maddela, Yang Zhong, Wuwei Lan, Wei Xu

ACL 2020, long paper

An Empirical Study of Named Entity Recognition in StackOverflow

Jeniya Tabassum, Mounica Maddela, Wei Xu, Alan Ritter

ACL 2020, long paper

Generalizing Natural Language Analysis through Span-relation Representations

Zhengbao Jiang, Wei Xu, Jun Araki, Graham Neubig

ACL 2020, long paper

Discourse Level Factors for Sentence Deletion in Text Simplification

Yang Zhong, Chao Jiang, Wei Xu, Junyi Jessy Li

AAAI 2020, long paper (acceptance rate 20.6%; oral presentation)

Multi-task Pairwise Neural Ranking for Hashtag Segmentation

Mounica Maddela, Wei Xu, Daniel Preotiuc-Pietro

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

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

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%)

Character-based Neural Networks for 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, long paper (shared-task overview)

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
ACL 2015 Workshop on Noisy User-generated Text (shared-task overview; author ordered alphabetically)

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, long paper

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, long paper

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

| | | |
|----------------------|--|----------------|
| STUDENTS | Wuwei Lan (PhD student at OSU) | 2016 – present |
| | Jeniya Tabassum (PhD student at OSU - co-advisor: Alan Ritter) | 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 |
| | Sydney Lee (undergraduate at OSU) | 2018 – present |
| | Sarah Flanagan (undergraduate at OSU) | 2018 – present |
| | Sam Stevens (undergraduate at OSU) | 2019 – present |
| | Daniel Szoke (undergraduate at OSU) | 2019 – present |
| | Ethan Lee (undergraduate at OSU) | 2019 – present |
| | Jaewook Lee (undergraduate at OSU) | 2019 – present |
| | Nicholas Motlagh (undergraduate at OSU) | 2019 |
| | 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 at IBM Research) | 2016 – 2017 |
| | Piyush Ghai (completed Masters student at OSU, now at 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) | 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 |
| TEACHING | <i>CSE 5539 Social Media and Text Analytics</i> (http://socialmedia-class.org/) | |
| | A new course integrated with research, covering from basic to state-of-the-art machine learning algorithms (teaching evaluation: 4.13/5.00 Autumn 2019, 4.40/5.00 Autumn 2017, 4.60/5.00 Autumn 2016; 5.72/6.00 at NASSLLI 2015) | |
| | <i>CSE 5522 Artificial Intelligence II: Advanced Techniques</i> | |
| | Ohio State University, mixed undergraduate and graduate level (teaching evaluation: Spring 2020, 4.85/5.00 Autumn 2018, 4.50/5.00 Spring 2018) | |
| INVITED TALKS | <i>CSE 5525 Speech and Language Processing</i> | |
| | Ohio State University, mixed undergraduate and graduate level (teaching evaluation 3.80/5.00 Spring 2017) | |
| | Understanding and Generating Human Language | |
| | University of Maryland, College Park, MD | Feb 2020 |
| | University of Massachusetts, Amherst, MA | Jan 2020 |
| | Georgia Institute of Technology, Atlanta, GA | Dec 2019 |
| | Learning for Unlimited Human Language | |
| | Peking University, Beijing, China | Dec 2018 |
| | Learning Large-scale Paraphrases for Natural Language Understanding and Generation | |
| | 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 |
| | How does AI Understand Language? | |
| | Women in Analytics Conference, Columbus, OH (Main Stage Panel) | Mar 2018 |
| | Can Paraphrase be a Ultimate Solution for NLU and NLG? | |
| | Google Research, New York, NY | Jul 2017 |
| | Paraphrase \approx Monolingual Translation | |
| | Amazon, 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 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, 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 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 |
| OUTREACH ACTIVITIES | 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 |
| | Presenter, Philadelphia Science Festival | Apr 2015 |
| OPEN SOURCE CODE / DATA | <i>#HashtagMaster: A Semantic Analysis Tool for Hashtags</i> | Jun 2019 |
| | https://mounicam.github.io/hashtag_master | |
| | <i>Pairwise Neural Ranking Model and SimplePPDB++</i> | Oct 2018 |
| | https://github.com/lanwuwei/SPM_toolkit | |
| | <i>SPM Toolkit for Sentence Pair Modeling</i> | Aug 2018 |
| | https://github.com/lanwuwei/SPM_toolkit | |
| | <i>LanguageNet: Large-scale Paraphrase Corpus</i> | Sep 2017 |
| | https://github.com/lanwuwei/paraphrase-dataset | |
| | <i>Syntax MT-based Text Simplification System and SARI Evaluation Metric</i> | May 2015 |

| | |
|--|----------|
| https://github.com/cocoxu/simplification/ (contribution to the Joshua Machine Translation Toolkit) | |
| <i>NEWSLA Text Simplification Corpus</i> | Sep 2015 |
| https://newsela.com/data/ (widely adopted as the benchmark for text simplification research) | |
| <i>Multiple-instance Learning Paraphrase Model</i> | Dec 2014 |
| https://github.com/cocoxu/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/ | |