Wei Xu

✉ xu.1265@osu.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 Aug 2016 – Present
Department of Computer Science and Engineering

Education

Ph.D. in Computer Science, New York University, New York, NY Jan 2014
Thesis: Data-driven Approaches for Paraphrasing Across Language Variations

B.S. and M.S. in Computer Science, Tsinghua University, Beijing, CHINA 2004/2007
Thesis: Event-relevance Based Summarization

Experience

Postdoctoral Fellow, University of Pennsylvania, Philadelphia, PA Feb 2014 – Aug 2016
Computer Information and Science Department

Visiting Researcher, University of Washington, Seattle, WA Jan 2012 – Dec 2013
Computer Science and Engineering Department

Intern, Microsoft Research, Seattle, WA Jun – Sep 2011
Internet Services Research Center

Consultant, Educational Testing Service, Princeton, NJ Mar – Dec 2010
Natural Language and Speech Group

Intern, Amazon.com, Seattle, WA Jun – Aug 2010
Catalog Quality Group

Services

Workshop Chair: ACL (2017)
Area Chair: EMNLP (2016)
Publicity Chair: NAACL (2016)
Session Chair: EMNLP (2016, 2015), NAACL (2015), AAAI (2015), ACL (2014)

Program Committee:
ACL (2017, 2015, 2014 – Outstanding Reviewer Award, 2013), NAACL (2015), EMNLP (2015, 2014)
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)

Steering Committee: SemEval (2017)

Thesis Committee:
Maria Pershina, Ph.D. in Computer Science, New York University (2016)
Kai Cao, Ph.D. in Computer Science, New York University (2016)

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

Workshop Organizer:
– The 1st International Workshop on Noisy User-generatedText at ACL (2015)
– The 2nd International Workshop on Noisy User-generated Text at COLING (2016)
– The 3rd International Workshop on Noisy User-generated Text at EMNLP (2017)
`http://noisy-text.github.io/`

– SemEval shared-task: Paraphrase and Semantic in Twitter (2015)
(19 research teams worldwide participated)
`http://alt.qcri.org/semeval2015/task1/`

– Mid-Atlantic Student Colloquium on Speech, Language and Learning (2016)
`http://www.mascsll.org/`

Publications

Refereed Journal and Conference Papers (all have acceptance rates ≈ 25%)

A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter
Jeniya Tabassum, Alan Ritter, **Wei Xu**
Proceedings of EMNLP 2016

Optimizing Statistical Machine Translation for Simplification
Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, Chris Callison-Burch
Transactions of the Association for Computational (TACL), 2016
oral presentation at ACL 2016

Discovering User Attribute Stylistic Differences via Paraphrasing
Daniel Preotiuc, **Wei Xu**, Lyle Ungar
Proceedings of AAAI 2016

Problems in Current Text Simplification Research: New Data Can Help
Wei Xu, Chris Callison-Burch, Courtney Napoles
Transactions of the Association for Computational (TACL), 2015
oral presentation at EMNLP 2015

Cost Optimization for Crowdsourcing Translation
Mingkun Gao, **Wei Xu**, Chris Callison-Burch
Proceedings of NAACL 2015

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), 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

Filling Knowledge Base Gaps for Distant Supervision of Relation Extraction
Wei Xu, Raphael Hoffmann, Le Zhao, Ralph Grishman
Proceedings of ACL 2013

Paraphrasing for Style
***Wei Xu**, Alan Ritter, Bill Dolan, Ralph Grishman, Colin Cherry*
Proceedings of COLING 2012

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

Passage Retrieval for Information Extraction using Distant Supervision
Wei Xu, Ralph Grishman, Le Zhao
Proceedings of IJCNLP 2011

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

Extractive Summarization using Inter- and Intra- Event Relevance
Wenjie Li, **Wei Xu**, Mingli Wu, Chunfa Yuan, Qin Lu
Proceedings of COLING-ACL 2006

Refereed Workshop Papers (all peer-reviewed)

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 2015 Workshop on Noisy User-generated Text (WNUT)

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 2013 Workshop on Building and Using Comparable Corpora (BUCC)

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

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 slot filling track)

A Parse-and-Trim Approach with Information Significance for Chinese Sentence Compression
Wei Xu, Ralph Grishman
Proceedings of the 2009 Workshop on Language Generation and Summarisation (UCNLG)

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

Domain Extension of Chinese Named Entity Recognition
Wei Xu, Bin Fu, Liu Liu, Chunfa Yuan, Wenjie Li
Frontiers of Content Computing (FCC) 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

Theses

Data-driven Approaches for Paraphrasing Across Language Variations
Wei Xu
Ph.D. Thesis, Department of Computer Science, New York University

Open Source Code / Data

Syntax MT-based Text Simplification System and Crowdsourced Dataset 2015 – Ongoing
(contribution to the Joshua MT Toolkit)
`https://github.com/cocoxu/simplification/`

NEWSELA Text Simplification Corpus Sep 2015
`https://newsela.com/data/`

Multiple-instance Learning Paraphrase Model Dec 2014
`https://www.cis.upenn.edu/~xwe/multip/`

Twitter Paraphrase Corpus (shared-task at SemEval-2015) Oct 2014
`http://alt.qcri.org/semeval2015/task1/`

Event-based Twitter Summarization System Nov 2013
`https://github.com/cocoxu/twittersummarization/`

Twitter Normalization Phrase Table Oct 2014
`https://github.com/cocoxu/twitterparaphrase/`

Parallel Shakespeare Corpus and Model Jul 2012
`https://github.com/cocoxu/Shakespeare/`

Invited Talks

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 (Special Lecture)	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
France Telecom, Beijing, CHINA	Mar 2007

Student Research Advising

Wuwei Lan (current PhD student at OSU)
Pravar Mahajan (current masters student at OSU)
Mingkun Gao (completed masters student, now a PhD student at UIUC)
Siyu Qiu (completed masters student, now at Hulu.com)
Quanze Chen (completed undergraduate student, now PhD at University of Washington)
Bin Fu (completed undergraduate student, now at Google NYC)
Chenyang Lei (completed undergraduate student, now at Microsoft Redmond)

Teaching

Instructor, The Ohio State University, Columbus, OH Spring 2017
Course: *Speech and Language Processing* (dual-listed undergraduate and graduate course)
`https://cocoxu.github.io/courses/5525_spring17.html`

Instructor, The Ohio State University, Columbus, OH Autumn 2016
Course: *Social Media and Text Analytics* (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: *Social Media and Text Analytics* (interdisciplinary)
`http://ruccs.rutgers.edu/nasslli2016/`

Instructor, University of Pennsylvania, Philadelphia, PA Summer 2015
Course: *Social Media and Text Analytics* (graduate-level)

References

Chris Callison-Burch
Assistant 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/`

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/`

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/`

Ralph Grishman
Professor, Computer Science Department, New York University, New York, NY
Department Chair (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 (2013-2016)
☎ +1 310 612 0315
🖂 joel.tetreault@grammarly.com
`http://www.cs.rochester.edu/~tetreaul/academic.html`