

# Alan Ritter

---

Assistant Professor  
Computer Science and Engineering  
The Ohio State University

## CONTACT INFORMATION

*E-mail:* ritter.1492@osu.edu  
*Office:* 595 Dreese Labs  
*Phone:* (360) 820-1090  
*WWW:* <http://aritter.github.io/>  
*Google Scholar Profile:* <https://scholar.google.com/citations?user=unXtH3IAAAAJ>

## RESEARCH INTERESTS

Natural Language Processing, Information Extraction, Machine Learning, Social Media Analysis

## EDUCATION

**University of Washington**, Seattle, WA USA

Ph.D. in Computer Science & Engineering

- Thesis Topic: Information Extraction in Social Media
- Fall 2007 - Summer 2013

**Western Washington University**, Bellingham, WA USA

B.S./M.S., Computer Science, December 2006

- Overall GPA: 4.00

## PROFESSIONAL EXPERIENCE

**The Ohio State University**, Columbus OH USA

*Assistant Professor - Computer Science and Engineering*

**August 2014 - ...**

**Microsoft Research**, Redmond WA USA

*Consultant - Natural Language Processing Group*

**August 2014 - ...**

Continuing collaboration on data-driven conversation [EMNLP 2011a] that was started during my summer internship in 2009 (<http://research.microsoft.com/en-us/projects/convo/>).

**Carnegie Mellon University**, Pittsburgh PA USA

*Postdoctoral Fellow - Machine Learning Department*

**August 2013 - August 2014**

Proposed a new method for weakly supervised event extraction based on expectation regularization. Initiated a collaboration with researchers at US-CERT to automatically extract cybersecurity events reported on social media, including denial of service attacks, data breaches and account hijacking [WWW 2015]. Co-developed a joint model of semantic and visual word senses [CVPR 2015]. Investigated extraction of structured user profiles and major life events from user-generated text on social media [ACL 2014, EMNLP 2014].

Supervisor: Tom Mitchell

**Microsoft Research**, Redmond WA USA

*Intern - Natural Language Processing Group*

**June 2009 - September 2009**

Investigated conversations on Twitter. Crawled Twitter to gather a corpus of conversations. Designed a generative model of dialogue, and conducted experiments to demonstrate the feasibility of unsupervised modeling of dialogue acts in new media. Proposed data-driven response generation, identified and addressed initial challenges, developed a system and conducted experiments. This work resulted in publication at [NAACL 2010] and [EMNLP 2011a], and led to a major new research initiative (<http://research.microsoft.com/en-us/projects/convo/>)

Supervisors: Colin Cherry, Bill Dolan

*Intern - Knowledge Tools Group (Machine Learning Department)*      **June 2008 - August 2008**  
Investigated new methods for Smart Selection in the file browser domain. Proposed the idea of learning from indirect feedback in interactive machine learning. Implemented a prototype *smart file browser*, and conducted user studies. This work resulted in an award paper at [IUI 2009].  
Supervisor: Sumit Basu

**University of Washington**, Seattle WA USA

*Research Assistant*

**April 2007 - August 2013**

Supervisors: Oren Etzioni, Mausam

**Microsoft**, Redmond WA USA

*Software Design Engineer*

**January 2007 - March 2007**

Windows Media DRM group.

**Naval Research Laboratory, Marine Meteorology Division**, Monterey, CA USA

*Intern - Data Assimilation Group*

**June 2006 - August 2006**

Used TAU (Tuning and Analysis Utilities) to instrument, profile, analyze, and improve scalability of research and operational NRL Data Assimilation Fortran/MPI code. Taught a class on automatic instrumentation of parallel programs for NRL scientists. Presented a Seminar on scalability enhancements to NAVDAS.

Supervisor: William F. Campbell

**The NetBSD Project**

*Google "Summer of Code"*

**June 2005 - August 2005**

Ported FreeBSD's NDIS Compatibility Layer to NetBSD.

<http://www.netbsd.org/foundation/press/soc-summary.html>

Mentor: Philip A. Nelson

#### AWARDS

Best Student Paper Award for *Learning to Generalize for Complex Selection Tasks* at IUI (2009)  
NDSEG Fellowship (2008)

#### GRANTS

**CRII: III: Learning to Extract Events from Knowledge Base Revisions**, U.S. NSF  
9/1/2015-9/1/2017  
Amount: \$151,299  
Role: PI

**Extracting a Realtime Cybersecurity Knowledge Graph from Text**, IARPA

Leidos Subcontract

Amount: \$710,150

Role: Ohio State PI

**Events, relationships, and script learning for situational awareness**

Software Engineering Institute, Carnegie Mellon University

Amount: \$40,000

Role: Ohio State PI

**Supporting systematic review production with article similarity network visualization**, NIH

Amount: \$100,000

Role: Co-PI

**Modeling the Spread of Information through Social and Knowledge Graphs**, DARPA

Leidos Subcontract  
Amount: \$600,000  
Role: Ohio State PI

TEACHING

**The Ohio State University**, Columbus OH USA  
*Instructor - Speech and Language Processing*  
[http://aritter.github.io/courses/5525\\_spring16.html](http://aritter.github.io/courses/5525_spring16.html)

**January 2016 - May 2016**

**The Ohio State University**, Columbus OH USA  
*Instructor - Web Information Extraction*  
[http://aritter.github.io/courses/5539\\_fall15.html](http://aritter.github.io/courses/5539_fall15.html)

**September 2014 - December 2014**

**The Ohio State University**, Columbus OH USA  
*Instructor - 5522: Survey of AI II: Advanced Techniques*  
<http://aritter.github.io/courses/5522.html>

**January 2015 - May 2015**

**The Ohio State University**, Columbus OH USA  
*Instructor - IE and NLP for the Social Web*  
<http://aritter.github.io/courses/5539.html>

**September 2014 - December 2014**

THESIS  
COMMITTEES

*Young-Bum Kim* PhD, Feb 2015  
Natural Language Technologies for Low-Resource Languages (Advisor: Benjamin Snyder)  
University of Wisconsin

PATENTS

US Patent Application US20110010669  
“Items Selection Via Automatic Generalization”

INVITED TALKS

“Data-Driven Conversation”  
Microsoft Research Faculty Summit, July 2017

“Large-Scale Learning for Information Extraction”  
Midwest Machine Learning Symposium, June 2017  
University of Colorado, March 2017  
Tsinghua University, December 2016  
University of Edinburgh, July 2016  
University of North Carolina, March 2016

“Weakly Supervised Information Extraction for the Social Web”  
Machine Learning Seminar, Purdue University, December 2015  
Statistics Department, Ohio State University, October 2015  
School of Computer Science, Carnegie Mellon University, April 2015  
CLIP, University of Maryland, March 2015

“Information Extraction from Noisy Text”  
Clunch, University of Pennsylvania, September 2014

“Extracting Knowledge from Informal Text”

University of Pittsburgh, January 2014

Machine Learning Lunch, Carnegie Mellon University, October 2013

IEEE Intelligence and Security Informatics, June 2013

Toyota Technological Institute at Chicago, April 2013

Microsoft Research, April 2013

University of Texas at Dallas, March 2013

The Ohio State University, March 2013

Machine Learning and Friends Lunch, University of Massachusetts Amherst, March 2013

Human Language Technology Center of Excellence, March 2013

“Modeling Conversations in Social Media”

Microsoft Research, September 2011

“Extracting a Calendar from Twitter”

Pacific Northwest National Laboratory, January 2013

Proteus Group, New York University, September 2012

Wavii, April 2012

Facebook, March 2012

Google Research, September 2011

Yahoo! Labs, September 2011

Twitter, May 2011

Information Sciences Institute, University of Southern California, February 2011

“Finding Contradictions in Text Using Functions”

Boeing Research

February 2008

#### PRESS

“Forget today’s news: Zapaday will break future news”

<http://thenextweb.com/media/2012/05/01/forget-todays-news-zapaday-will-break-future-news/>

<http://bit.ly/1Ndfnic>

The Next Web

May 2012

“Hello, computer!” UW prof and students search outside the box”

<http://www.geekwire.com/2011/computer-uw-prof-students-search-box/>

GeekWire

November 2011

#### ACADEMIC SERVICE Organizer

*ACL Workshop on Noisy User-Generated Text (WNUT 2015, 2016, 2017)*

*Sentiment Analysis in Twitter Shared Task (SemEval 2013, 2014, 2015, 2016)*

*Named Entity Recognition in Twitter Shared Task (at WNUT 2015, 2016)*

#### Area Chair

*ACL 2014 (NLP for the web and social media)*

*NAACL 2015 (Information Extraction and Question Answering)*

*EMNLP 2016 (Information Extraction, Information Retrieval and Question Answering)*

*NAACL 2018 (Computational Social Science)*

Steering Committee

*SemEval 2016,2018*

Journal Reviewing

*ACM Computing Surveys (2014), TACL (2013,2014,2015,2016), Computational Linguistics (2013,2015,2016), JAIR (2012), TOIS (2012), TKDE (2014,2016), JMLR MLOSS (2013), Journal of Natural Language Engineering (2014)*

Program Committee

*EMNLP (2010,2011,2012,2013,2014,2015,2017), ACL (2012,2013,2015,2016,2017), NAACL-HLT (2012,2013,2016), CoNLL (2013,2014,2015), AACL (2012,2015), EACL (2012,2014), Coling (2014,2016), ICWSM (2012), LREC (2012), IJCAI (2011), \*SEM (2013), WWW (2014,2015,2016,2017), CIKM (2014)*

Research Proposal Reviewing:

*Panelist, National Science Foundation, IIS, 2017*

*Independent Reviewer U.S. Department of Energy (2015)*

*Independent Reviewer for Hong Kong Research Grants Council (2016)*

*Independent Reviewer for Natural Sciences and Engineering Research Council of Canada (2016)*

JOURNAL PAPERS      Preslav Nakov, Sara Rosenthal, Svetlana Kiritchenko, Saif M. Mohammad, Zornitsa Kozareva, Alan Ritter, Veselin Stoyanov, Xiaodan Zhu  
“Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts”  
Language Resources and Evaluation 2016

Wei Xu, Alan Ritter, Chris Callison-Burch, William B. Dolan and Yangfeng Ji  
“Extracting Lexically Divergent Paraphrases from Twitter”  
Proceedings of TACL 2014

Alan Ritter, Luke Zettlemoyer, Mausam and Oren Etzioni  
“Modeling Missing Data in Distant Supervision for Information Extraction”  
Proceedings of TACL 2013

CONFERENCE PAPERS      Xinlei Chen, Alan Ritter, Abhinav Gupta and Tom Mitchell  
“Sense Discovery via Co-Clustering on Images and Text”  
Proceedings of CVPR 2015

Alan Ritter, Evan Wright William Casey and Tom Mitchell  
“Weakly Supervised Extraction of Computer Security Events from Twitter”  
Proceedings of WWW 2015

“Never Ending Learning”  
Tom Mitchell, William Cohen, Estevam Hruschka, Partha Talukdar, Justin Betteridge, Andrew Carlson, Bhavana Dalvi, Matt Gardner, Bryan Kisiel, Jayant Krishnmurthy, Ni Lao, Kathryn Mazaitis, Tahir Mohammad, Ndapa Nakashole, Emmanouil Antonios Platanios, Alan Ritter, Mehdi Samadi, Burr Settles, Richard Wang, Derry Wijaya, Abhinav Gupta, Xinlei Chen, Abulhair Saparov, Malcolm Greaves and Joel Welling  
Proceedings of AAAI 2015

Jiwei Li, Alan Ritter, Claire Cardie and Eduard Hovy  
“Major Life Event Extraction from Twitter based on Congratulations/Condolences Speech Acts”  
Proceedings of EMNLP 2014

Jiwei Li, Alan Ritter and Eduard Hovy  
“Weakly Supervised User Profile Extraction from Twitter”  
Proceedings of ACL 2014

Justin Betteridge, Alan Ritter and Tom Mitchell  
“Assuming Facts Are Expressed More Than Once”  
Proceedings of FLAIRS 2014

Leon Derczynski, Alan Ritter, Sam Clark and Kalina Bontcheva  
“Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data”  
Proceedings of RANLP 2013

Alan Ritter, Mausam, Oren Etzioni, Sam Clark  
“Open Domain Event Extraction from Twitter”  
Proceedings of KDD 2012

Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, Colin Cherry  
“Paraphrasing for Style”  
Proceedings of Coling 2012

Alan Ritter, Sam Clark, Mausam, Oren Etzioni  
“Named Entity Recognition in Tweets: An Experimental Study”  
Proceedings of EMNLP 2011

Alan Ritter, Colin Cherry, Bill Dolan  
“Data-Driven Response Generation in Social Media”  
Proceedings of EMNLP 2011

Alan Ritter, Mausam and Oren Etzioni  
“A Latent Dirichlet Allocation Method for Selectional Preferences”  
Proceedings of ACL 2010

Alan Ritter, Colin Cherry and Bill Dolan  
“Unsupervised Modeling of Twitter Conversations”  
Proceedings of HLT-NAACL 2010

Alan Ritter and Sumit Basu  
“Learning to Generalize for Complex Selection Tasks”  
**Best Student Paper Award**  
Proceedings of IUI 2009

Alan Ritter, Doug Downey, Stephen Soderland, and Oren Etzioni  
“It’s a Contradiction – No, It’s Not: A Case Study using Functional Relations”  
Proceedings of EMNLP 2008

WORKSHOP PAPERS Ting-Hao K. Huang, Walter S. Lasecki, Alan L. Ritter, Jeffrey P. Bigham  
“Combining Non-Expert and Expert Crowd Work to Convert Web APIs to Dialog Systems” Work-in-Progress paper in the Proceeding of Conference on Human Computation & Crowdsourcing (HCOMP 2014)

Walter S. Lasecki, Alan, Ritter and Jeffrey P. Bigham  
“Powering Spoken Language Interactions With the Crowd.”  
Proceedings of the CHI Workshop on Designing Speech and Language Interactions (DSLI 2014), 2014.

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

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

Sina Jafarpour and Chris Burges, Alan Ritter  
“Filter, Rank, and Transfer the Knowledge: Learning to Chat”  
NIPS Workshop on Advances in Ranking, Vancouver, Canada, 2009

Alan Ritter, Stephen Soderland and Oren Etzioni  
“What is this Anyway? Automatic Hypernym Discovery”  
Proceedings of the AAAI 2009 Spring Symposium on Learning by Reading and Learning to Read