

# Large-Scale Learning for Information Extraction

**Alan Ritter**

Computer Science and Engineering  
The Ohio State University



14:00-16:00, 2016/12/26  
FIT 1-312, Tsinghua University

## Abstract

Much of human knowledge is available on the internet, however natural language is notoriously difficult for computers to interpret. In this talk, I present some recent advances in extracting structured knowledge from text with an eye toward real time information in massive data streams found on social media. I argue we can not rely on traditional methods that learn from small, hand-annotated datasets if we hope to extract a broad range of relations from diverse text genres at scale. As an alternative to human labeling, I will describe an approach that reasons about latent variables to learn robust information extraction models from large, opportunistically gathered datasets. To highlight new applications this work enables, I will demonstrate a system that continuously extracts a calendar of popular events in the near future from Twitter and also a method for detecting early reports of cyber-attacks on social media. Finally, I will outline some exciting possibilities for predicting real-time updates to a knowledge graph from events mentioned in text.

## About speaker

Alan Ritter is an assistant professor in Computer Science at Ohio State University. His research interests include natural language processing, social media analysis and machine learning. Alan completed his PhD at the University of Washington and was a postdoctoral fellow in the Machine Learning Department at Carnegie Mellon University. He has received an NDSEG fellowship, a best student paper award at IUI, an NSF CRII and has served as an area chair for ACL, EMNLP and NAACL.