

# Statistical Relational Learning for Knowledge Extraction from the Web

**Hoifung Poon**

University of Washington, USA

## Abstract

Extracting knowledge from unstructured text has been a long-standing goal of NLP. The advent of the Web further increases its urgency by making available billions of online documents. To represent the acquired knowledge that is complex and heterogeneous, we need first-order logic. To handle the inherent uncertainty and ambiguity in extracting and reasoning with knowledge, we need probability. Combining the two has led to rapid progress in the emerging field of statistical relational learning. In this talk, I will show that statistical relational learning offers promising solutions for conquering the knowledge-extraction quest. I will present Markov logic, which is the leading unifying framework for representing and reasoning with complex and uncertain knowledge, and has spawned a number of successful applications for knowledge extraction from the Web. In particular, I will present OntoUSP, an end-to-end knowledge extraction system that can read text and answer questions. OntoUSP is completely unsupervised and benefits from jointly conducting ontology induction, population, and knowledge extraction. Experiments show that OntoUSP extracted five times as many correct answers compared to state-of-the-art systems, with a precision of 91%.