LTI Colloquium Series

Carnegie Mellon University
Language Technologies Institute

Heng Ji

(Rensselaer Polytechnic Institute)

Friday, March 29th

Doherty Hall 2315
2:30 pm Talk,
4:00 pm Snacks, LTI 5th Floor

UNIVERSAL INFORMATION EXTRACTION

BIOGRAPHY

Heng Ji is the Edward P. Hamilton Chair Professor in Computer Science at Rensselaer Polytechnic Institute. She received her Ph.D. in Computer Science from New York University. Her research interests focus on Natural Language Processing, especially on Information Extraction and Knowledge Base Population. She was selected as "Young Scientist" and a member of the Global Future Council on the Future of Computing by the World Economic Forum in 2016, 2017 and 2018.

She received "AI's 10 to Watch" Award by IEEE Intelligent Systems in 2013, NSF CAREER award in 2009, Google Research Awards in 2009 and 2014, IBM Watson Faculty Award in 2012 and 2014, and Bosch Research Awards in 2015, 2016 and 2017. She coordinated the NIST TAC Knowledge Base Population task since 2010, and served as the Program Committee Co-Chair of several conferences including NAACL-HLT2018.

ABSTRACT

The big data boom in recent years covers a wide spectrum of heterogeneous data types, from text to image, video, speech, and multimedia. Most of the valuable information in such "big data" is encoded in natural language, which makes it accessible to some people — for example, those who can read that particular language — but much less amenable to computer processing beyond a simple keyword search. Information Extraction (IE) and Information Retrieval (IR) on a massive scale share the same goal of creating the next generation of information access in which humans can communicate with computers in any natural language beyond keyword search, by extracting and presenting the important and relevant information embedded in big data. IE aims extract structured facts from a wide spectrum of heterogeneous unstructured data types.