



Fall 2010 Colloquium

Temple University Computer and Information Sciences

Autonomous Web-scale Information Extraction

Doug Downey
Northwestern University

Tuesday 11/16, 11am, Tech Center 111

Abstract:

Search engines are extremely useful tools for answering simple questions. However, for more complex questions -- e.g., "which nanotechnology companies are hiring on the West Coast?" -- existing search engines are less effective, because the answers are not contained on just a single page. Answering these questions requires extracting and synthesizing information across multiple documents. Currently, this is a tedious and error-prone manual process.

In this talk, Dr. Downey will describe his research aimed at automating the extraction of information from the Web. He will present a model of the redundancy inherent in the Web, and show that the model can be used to identify correct extractions autonomously, without the manually labeled examples typically assumed in previous information extraction research. Further, while the redundancy-based model alone is ineffective for the "long tail" of infrequently mentioned facts, Dr. Downey will illustrate how unsupervised language models can be leveraged to overcome this limitation.

Bio:

Doug Downey is an assistant professor in the EECS Department of Northwestern University, which he joined in the Fall of 2008. He obtained his PhD from the University of Washington, where he was advised by Oren Etzioni. His research interests are in the areas of natural language processing, machine learning, and artificial intelligence, with a particular interest in utilizing the Web to autonomously extract large knowledge bases.