



SE305

Topics in Web-Scale Knowledge Harvesting

Instructor: Partha Talukdar [ppt@serc, talukdar.net]

When: TTh 2-3:30pm [First Class: Aug 5, 2014]

Where: SERC 202

Number: SE 305 (3:1), AUG 2014

URL: <http://sites.google.com/site/se3052014>

References: Current literature

Topics

This is a research-focused course where we shall read papers to understand and extend state-of-the-art statistical machine learning techniques for various problems in web-scale knowledge base (KB) construction, e.g., entity categorization and resolution, relation extraction, distant supervision, KB inference, curriculum learning, learning from limited supervision in KB, representation learning for knowledge harvesting, etc. Particular emphasis will be given to techniques that can scale to large datasets, preferably using distributed frameworks such as Hadoop.

Evaluation

Student presentation and participation in class, and also a semester-end synthesis report or project (as preferred by the student).

Prerequisites

Interest in any of the following: machine learning, natural language processing, and large data processing