

Computer Science Seminar Series, 2009

National Capital Region

Data-Intensive Text Processing with MapReduce

Speaker: Prof. Jimmy Lin
University of Maryland

Friday, October 16, 2009
1:00PM- 2:00PM, NVC 325

Abstract

Over the past couple of decades, the field of natural language processing (and more broadly, human language technology) has seen the emergence and later dominance of empirical techniques and data-driven research. An impediment to research progress today is the need for scalable algorithms to cope with the vast quantities of available data.

The only practical solution to large-data challenges today is to distribute the computation across multiple machines. Cluster computing, however, is fraught with challenges ranging from scheduling to synchronization. Over the past few years, MapReduce has emerged as an attractive alternative to traditional programming models: it provides a simple functional abstraction that hides many system-level issues, allowing the researcher to focus on solving the problem.

In this talk, I will overview "cloud computing" projects at the University of Maryland that grapple with the issue of scalability in text processing applications.

Biography

Jimmy Lin is an associate professor in the iSchool at the University of Maryland. He is the director of the newly-formed Cloud Computing Center, and leads the Google/IBM Academic Cloud Computing Initiative at the University of Maryland. Dr. Lin's research lies at the intersection of information retrieval and natural language processing. He graduated with a Ph.D. in Electrical Engineering and Computer Science from MIT in 2004