CSE 960 Algorithms for the Web: Selected Topics of Algorithms & Complexity
Spring 2003

Instructor: Moon Jung Chung (chung@cse.msu.edu)
Office Hours: TuTh: 3:00-4:00 pm

Course Objectives

The World Wide Web has become part of our everyday life. Information retrieval, data mining, and auctions on the web is now of enormous practical interest. The aim of the course is to study the theoretical foundations of the algorithms supporting these activities and practical applications of them to real problems on the Web. In the course, we study models of web graphs, and the recent development of high-performance algorithms and applications that use the graph structure of the web.

Prerequisites: CPS830, CS835, or CSE860


References: Data Mining: Concepts and Techniques by J. Han, M. Kamber, Morgan Kaufmann Publishers, 1st edition (August 2000)

Reference Papers: Will be handout time to time.

Grading

Homework 20%
Midterm 30%
Project 30%
  Mid Progress Presentation
  Midterm Report
  Final Report
  Final Presentation
Class Participation 20% -- Presentations and Discussions

Project: A team of 2-4 students will select a topic and work on the project. Each student will present background papers and results of the project in the class.
Topics

Topics of interest include but are not limited to:

Part I: Data Mining Basics
  o Mining Association Rules
  o Apriori Algorithm
  o Classification and Prediction – Decision Tree Induction, Bayesian Classification
  o Clustering

Part II: Graph models of the Web and the Internet
  o Algorithms for analyzing the web graph
  o Graph oriented statistical sampling of the web
  o Practical discovery algorithmic techniques on the Web
  o Mining web link structure – Hub, Authority
  o Mining Web Content
  o Mining Web Usage
  o Dimension Reduction
  o Application to web-searching, information retrieval on the web, data mining.
  o Web mining and Internet security

PART III: Game theory and economics of the Internet
  o Routing on the Internet
  o Game theory
  o Nash’s equilibrium
  o Network design with selfish agents
  o Fixed point models for performance engineering in enterprise IP networks
  o Performance and measurement issues in the Web
  o Empirical studies and issues
  o Congestion pricing and capacity expansion games