CS:4440 Web Mining

Credits and contact hours: 3 semester hours
On Campus section: Two 75-minute lectures per week

Instructor:
Padmini Srinivasan
101B MacLean Hall

Textbooks: Free online versions of the following

1. Introduction to Information Retrieval, Cambridge University Press. 2008. By Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze.

2. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. By David Easley and Jon Kleinberg. Cambridge University Press.

3. Network Science By Albert-Laszlo Barabasi.

4. Selected research papers.

Evaluation:
Participation
Assignments (3 to 5)
Project
Paper

Course description: The Web as a vast, variegated and growing resource interconnects our world in many interesting ways. One area of study that has emerged somewhat recently is that of "web mining". Research and development in web mining span a spectrum of goals from methods for high quality retrieval, to monitoring information sites and streams, to distilling and summarizing web information to identifying social networks and discovering new opportunities and ideas. This course will study the web mining landscape through papers, assignments and a term project. Exposure to Perl and to database technology is preferred.

Prerequisites: CS:3330 with a minimum grade of C-

Recommendations: CS:4400

Brief list of topics covered:
Networks and the Structure of the Web (2 weeks)
Working with Text Collections (4 weeks)
Evaluating Systems (2 weeks)
TREC research problems (3 weeks)
Applications in Web Mining (4 weeks)

Disclaimer: All information is tentative and subject to change.

Spring 2020