CAP 6777 Web Mining

Credits: 3 credits

Text book, title, author, and year: Introduction to Information Retrieval, Cambridge University Press, 2008. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schützehttp://nlp.stanford.edu/IR-book/, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, 2nd Edition, Springer July 2011, Bing Liu. http://www.cs.uic.edu/~liub/WebMiningBook.html

Reference materials: Research papers which will be distributed in the class

Specific course information

Catalog description: This course teaches students basic techniques to mine the Web and information networks (including social networks and social media). Detailed topics include three aspects: (1) web Crawling, indexing, ranking and search algorithms using content and link analysis; (2) Web clustering, classification, and mining algorithms, and (3) social network analysis and online social media mining. Students will also gain experience through course project on one of the topics covered in class (such as building a search engine or an online twitter sentiment analysis tool).

Prerequisites: Prerequisites: *STA 4821 or equivalent*

Specific goals for the course: To provide certain technical skills that are important in computer science and engineering applications; to provide knowledge base for data communication, networks, and web programming. To provide knowledge on web crawling, data mining, and social network analysis. The goal of this class is for students to gain hands-on experiences on information search and web mining. Course covers techniques used to collect, analyze, and understand the data from Internet and the web (including social networks). At the end of the class, students should be able to understand the whole process of collecting information from the web, and carrying out system design for search and mining the web. We will use online web documents (such as Twitter data) as the testbed and practice web mining techniques.

Brief list of topics to be covered:

Information retrieval

- Web crawling
- Text indexing, scoring, and ranking.
- Information extraction and integration

Web Mining

- Text clustering and classification
- Online opinion mining sentiment analysis
- Recommendation systems.

Social Network Analysis

- Link analysis, RageRank, and SimRank algorithms.
- Social mining algorithms.
- Online social media analysis and monitoring