

CS 590I: Information Retrieval

Spring 2009

Instructor: Professor Luo Si

Course Description:

The explosive growth of available digital information (e.g., Web pages, emails, news, scientific literature) demands intelligent information agents that can sift through all available information and find out the most valuable and relevant information. Web search engines, such as Google, Yahoo!, and MSN, are several examples of such tools. This course studies the basic principles and practical algorithms used for information retrieval and text mining. The contents includes: statistical characteristics of text, several important retrieval models, text categorization, recommendation system, clustering, information extraction, etc. The course emphasizes both the above applications and solid modeling techniques (e.g., probabilistic modeling) that can be extended for other applications.

Students will:

1. Learn the theories and techniques behind Web search engines, E-commerce recommendation systems, etc.
2. Get hands on project experience by developing real-world applications, such as intelligent tools for improving search accuracy from user feedback, email spam detection, recommendation system, or scientific literature organization and mining.
3. Learn tools and techniques to do cutting-edge research in the area of information retrieval or text mining. This course provides a natural transition to CS690I (Advanced Topics in Information Retrieval).
4. Open the door to the amazing job opportunities in Search Technology and E-commerce companies such as Google, Microsoft, Yahoo! and Amazon.

The course is lecture-based. Students are expected to read book chapters and related research papers. The preliminary plan for evaluation is based on homework assignments: 30%, one course project: 30%, one final exam: 30%, and class participation: 10%.

Prerequisites:

C/C++ programming skills are required (e.g., CS240). Background in basic level of probability/statistics is preferred, but not required. Knowledge of Matlab is helpful.