Advances Database Systems (CS6093)
(Topics for the Web)

Computer Science and Engineering
Polytechnic Institute of NYU

Spring 2011

1 Instructors

Dr. Fernando Diaz (Yahoo! Research), {lastnamefirstinitial}@yahoo-inc.com

Dr. Cong Yu (Google Research), {fullname}@google.com

2 Course Information

Brief Description: This course covers a variety of advanced database topics. In this semester, we will focus on topics that are closely related to information retrieval and management on the Web. Specifically, we will study the following main topics: information retrieval ranking and evaluation (IRR and IRE), personalized information retrieval (IRP), structured search and information extraction (SS and IE), and web data mining (DM). We will also briefly touch upon a recent topic on human computation (HC).

Prerequisite: CS6083 or CS308 or an equivalent first course on database systems. The course will assume familiarity with basic algorithm concepts such as data structures and algorithm complexity, and basic database and information retrieval concepts such as relational data model, document model, and query processing. Some knowledge about data mining and information retrieval is a plus, but not required.

Communication Tools: Minor announcements will be posted on the web site only. Important announcements will be posted on the web site and emailed to the mailing list.

Web Site: http://www.eecs.umich.edu/~congy/cs6093/s11/ (to be moved soon)

Mailing List: spring1112.cs6093.2256@utopia.poly.edu

Time and Location: M 6:00 - 8:25pm, RH602 (6 MetroTech)

Office Hours: M 4:50 - 5:50pm, LC246, or by appointment.

Textbook and Reading Material: There is no required text book. The course will primarily rely on recent research papers. We expect students to locate most of the papers online themselves (this is a required skill for conducting research!). Certain articles that are not available online will be made available through the course web site. Classic database and information retrieval textbooks such as Database Management Systems (ISBN: 0072465638) and Introduction to Information Retrieval (ISBN: 0521865719) can be useful.

Grading: 25% class participation, 30% midterm report, 45% final report and demo. For class participation, there will be four (random) quizzes throughout the semester: each will constitute 5% of the participation and the remaining 5% will be based on your engagement in the discussion during class. The final grades will not be based on a curve.
3 Lecture Outline

Lectures 02,05,06: Structured Search and Information Extraction (by Cong)

Lectures 03,04: Information Retrieval Ranking (by Fernando)

Lectures 07,08: Web Data Mining (by Cong)

Lectures 09,10: Information Retrieval System Evaluation (by Fernando)

Lectures 11,12: Personalized Information Retrieval (by Fernando)

Lectures 13: Human Computation (by Cong)

Lectures 14: Project Demo

4 Course Projects

Course project is a required component of the course and accounts for 75% of the grade. Each group will consist of 3 to 4 students and work on a project relevant to one of the course topics. Groups of 1 or 2 are allowed only for students in the Ph.D. program and require approval from their advisors. Groups of more than 4 students are not allowed.

A list of potential project topics are available on the course web site. Project topics outside of the list can also be chosen, but will require approval from one of the instructors first. This list of available topics are on a first-come first-serve basis and you can only choose a topic after forming a group. If the project is about designing and building a prototype system, a demo of the system will be required.

5 Important Milestones

02/07: Project group formation and project selections due

03/21: Midterm report due

05/09: Project demo in class

05/13: Final report due