CS 599 – GRADUATE SPECIAL TOPICS INFORMATION RETRIEVAL

COURSE INFORMATION

Instructor   : Prof. Sampath Jayarathna, Web: http://www.cpp.edu/~ukjayarathna

Contact   : Office: 8-46, Email: ukjayarathna@cpp.edu, Phone: (909) 869-3145

Office Hours : Monday and Wednesday, 1.00 PM – 3.00 PM, or email me for an appointment

Schedule   : Monday and Wednesday, Room: 8-348, Time: 6.00 PM – 7.50 PM

Website   : http://www.cpp.edu/~ukjayarathna/courses/w17/cs599

Piazza   : www.piazza.com/csupomona/winter2017/cs599/home

Blackboard   : https://blackboard.cpp.edu/

Prerequisites : CS 331 with a grade of C or better or instructor's consent.

WHAT IS THIS COURSE ABOUT?

The CS 599, Graduate Special Topics, Information Retrieval (IR) aims to focus on the topics and underlying technologies of modern information retrieval systems, and hands-on experience by using existing information retrieval toolkits to learn algorithms, design and implementation of modern IR systems

In this course, we'll study the theory, design, and implementation of text-based and web-based information retrieval systems, including an examination of web mining algorithms and techniques at the core of modern search and data mining applications. This course will be a practice-oriented course that focuses on: IR techniques, retrieval models, indexing, text classification, document clustering, the concept of relevance, evaluation of IR systems and search engine technologies. This course will primarily focus on "text" retrieval.

WHAT WILL YOU GET FROM THIS COURSE?

After successfully completing this course, students should be able to:

• Define and explain the key concepts and models relevant to information storage and retrieval, including efficient text indexing, boolean, vector space and probabilistic retrieval models, relevance feedback, document clustering and text categorization.

• Analyze, identify and design core text based retrieval system algorithms and advanced algorithms like document clustering and text categorization/classification.

• Learn measures and techniques to evaluate IR systems and fundamental techniques to implement IR systems

• Demonstrate through involvement in a team project the central elements of team building and team management and salient features in recent research results in web search and information retrieval.

- **Required textbook**. *Introduction to Information Retrieval* C. Manning, P. Raghavan and H. Schutze Cambridge University Press, 2008. Free online version available at: http://nlp.stanford.edu/IR-book/
- We'll also read some selections from:
  - *Search Engines – Information Retrieval in Practice* W. B. Croft, D. Metzler, and T. Strohman Cambridge University Press, 2015. Free online version available at: http://ciir.cs.umass.edu/downloads/SEIRiP.pdf
  - We'll also read several papers and other resources provided in the course schedule (with links)

TENTATIVE COURSE SCHEDULE

Topics: The tentative topics are as follows. Topics and specific course activities may change as needed. PowerPoint slides will be available on the course web page before each lecture.

Boolean Retrieval
Vector Space Model
Evaluation in Information Retrieval
Relevance Feedback
Probabilistic Information Retrieval
Text Classification
Vector Space Classification
Support Vector Machines and Machine Learning on Documents
Clustering
Matrix Decomposition

WHAT YOU CAN EXPECT FROM ME:

I have an open door policy i.e., office visits. My posted office hours are times when I will make concerted effort to be available. Occasionally administrative meetings or emergencies may interfere with these posted times. The open door policy is: if my door is open, I am in and welcome walk-in visitations. In order to encourage you to find my office, the *first assignment is due at the drop box next to my office 8-46*.

I am committed to supporting students with disabilities. If you have challenges related to these issues or others I want to work with you to help you succeed. Please come and talk to me, since only you can properly communicate your situation to me.

WHAT YOU CAN GIVE TO THE CLASS:

It is extremely important for you to be engaged in the course. Otherwise, you will fall asleep and wonder what happened to your tuition dollars. So, I encourage you to ask questions during lecture and actively participate at the piazza forum. For the first few weeks, when asking a question at the class, state your name so that I know who you are.

Cell phones and Tardiness: You may have cell phones in class, but they must be on mute, or airplane mode and not answered until the end of class. You are expected to arrive on time so that you do not cause a disruption in the middle of class. I would like to start the class at the scheduled time. If you cannot make it on time or want to leave early for some reason, please let me know. Persistent tardiness will be noted.

Piazza: All questions will be fielded through Piazza. The primary benefit is that for many questions everyone can see the answer and other students can answer as well. I will endorse good student responses. Additionally, I expect you to actively participate in online discussions at Piazza. Over the course of the quarter, you should post at least one substantive, interesting post to the discussion forum. You must also respond to at least four posts made by others. You can also post private messages that can only be seen by the instructor. You will be signed up with your cpp email, but you may switch to another email.

Blackboard: Blackboard will be used primarily for grade dissemination.

Email: If you send email to me, please be sure to include your name and the course number in the body of the e-mail. You should also use an appropriate subject line that looks like "CS599-IR Project" etc. Failure to follow these guideline may result in delayed response. *Again, email should only be used in rare instances*, I will probably point you back to Piazza if you have a question related to course materials and/or relevant to other students in the class.

COURSE ACTIVITIES

The scores you receive on the various graded tasks in the class will be weighted as follows:

30%	Project: Proposal (5 pts), Progress reports(5), Presentation and Demo (10 pts), Final Report (10 pts)
20%	Final Exam
20%	Quizzes: Scheduled quizzes (2), Pop quizzes (2)
20%	Homework Assignments (5)
10%	Research Paper Summary (5)
1%	Extra Credit (1 user study participation or 1 culture report)
101%	Your Total Score for the class

Project: For the project, you will work in teams of either 3, or 4 students on a problem of your choosing that is interesting, significant, and relevant to IR. The ultimate goal of your course project is to develop a new tool to tackle some interesting real-world problem. At the end of the quarter, we will hold a competition during our regular class time for your project demonstration. All members of a group will receive the same grade on group work. Therefore, it is in your interest to choose other group members (ideally, first day of the class if possible) who have the same goal in the class as you do. It is also in your interest to work together and ensure that all tasks are completed effectively. Your scores on group work may be adjusted based on your contribution (peer-evaluation).

Final Exam: The final exam is comprehensive, closed books and will be held on **Monday, March 13 from 6.00 pm to 7.45 PM**. You may bring one standard 8.5" by 11" piece of paper with any notes you deem appropriate or significant (front and back). No iPads, iPhones, Blackberries, Android phones/tablets are allowed. Standard calculators are allowed.

In class Quizzes: Attendance in class and participation in the discussion are both important to your success in the course. As one crude measure of your participation and course preparation, you will have around 2 scheduled quizzes and 2 pop quizzes spread across the quarter. I will use them to gauge what topics we need to devote more time to and as an indicator that you were in class.

Homework: We will have 5 homework assignments, each worth 4% of your overall grade.

Research Paper Summary: There are approximately 7 papers assigned for reading (Wednesday). For each paper, you are expected to provide a summary/discussion in Piazza forum. More details on how to write a paper summary are up on piazza. For grading, I will choose the top 5 out of the 7 summaries.

Extra Credit: You can get up to one point added to your final grade through culture reports that broaden your exposure to computer science, or user study evaluation participation. Details for participating for user studies, selecting material, writing, and submitting the extra credit is on the course website. Note: Borderline grades will not be boosted if extra credit is not submitted.

GRADES

Final course grades are based on the overall average. You are guaranteed a grade based on a 10% window (e.g., 90-100% is an A). Overall class grade (not the individual grade) windows may be increased in size if the instructor finds it appropriate.

A: 90-100, B+: 85-89, B: 80-84, C+: 75-79, C: 70-74, D+: 65-69, D: 60-64, Fail (Grade F): 0-59

ATTENDANCE, MAKE-UPS AND LATE POLICIES:

All project reports, homework assignments, extra credit are due at the beginning of class in all required forms (e.g., paper and/or submit on blackboard) on the due date. Changes to a submission's due dates will be avoided because they are unfair to those students who have organized their time to complete the assigned work. Individual accommodations will be discussed if you have a valid medical excuse.

Project due dates will be set to give ample time for completion of the project and will not be extended save for the unexpected and unlikely major, long-lived catastrophe. Start projects early--last minute computer malfunctions will not be accepted as a reason for delaying a project report due date. For team project reports, only 1 report from the group required. It is you responsibility as a group to choose which team member responsible for the submission. *Unless otherwise noted, all project reports should be submitted on paper at the beginning of the class*.

An exam/quiz missed without an acceptable excuse will be recorded as a grade of zero (0). There will be no makeup examination/quiz. If you have a valid medical excuse, the missing component of your grade will be computed based on the other parts of the course content. E.g.: A missed quiz with acceptable excuse will be assigned the average grade of all quizzes, at the option of the course instructor.

Homework assignments must be turned in by the due date and time in order to contribute to your grade. Each late submission will incur a 5 points penalty per day. No submission will be accepted after 3rd day and will be recorded as a grade of zero (0). There will be no makeup for homework assignments. If you have a valid medical excuse, the missing component of your grade will be computed based on the other parts of the relevant course content. E.g.: A missed homework with acceptable excuse will be assigned the average grade of all other homework components, at the option of the course instructor. *Unless otherwise noted, all homework assignments should be submitted on paper at the beginning of the class.*

ACADEMIC OFFENSES

Scholarly dishonesty, especially plagiarism, will not be tolerated. Plagiarism is defined as "Failing to credit sources used in a work product in an attempt to pass off the work as one's own. Attempting to receive credit for work performed by another, including papers obtained in whole or in part from individuals or other sources." Students found to have engaged in plagiarism will be punished severely, typically earning an automatic F in the course and being reported to the Office of Student Conduct and Integrity. The Office of Student Conduct & Integrity investigates issues of student misconduct to determine if there has been a violation of the Student Conduct Code. If students are found responsible for a violation, students receive educational sanctions which can range anywhere from warnings to expulsion from the California State University system.

For the homework assignments, you may talk to any other class member or work in groups to discuss the problems in a general way. However, your actual detailed solution must be yours alone. If you do talk to other students, you must write on your homework assignment who it is that you discussed the problems with. Your submitted work must be written solely by you and not contain work directly copied from others.

Homework Collaboration Clarification: To clarify, your homework assignment is yours alone and you are expected to complete each homework independently. Your solution should be written by you without the direct aid or help of anyone else. However, I believe that collaboration and team work are important for facilitating learning, so I encourage you to discuss problems and general problem approaches (but not actual solutions) with your classmates. If you do have a chat with another student about a homework problem, you must inform me by writing a note on your homework submission (e.g., Bob pointed me to the relevant section for problem 3). The basic rule is that no student should explicitly share a solution with another student (and thereby circumvent the basic learning process), but it is okay to share general approaches, directions, and so on. If you feel like you have an issue that needs clarification, feel free to contact me.

DISABILITY RESOURCES

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities.

If you have a physical or a learning disability, please talk to me privately and/or contact the Disability Resource Center (DRC) at 909-869-3333. The location is at Bldg 9-103 to coordinate course accommodations. For further information, visit the DRC website at http://www.cpp.edu/~drc/index.shtml