Course title and number   Information Storage and Retrieval, CSCE470
Term   Spring 2022
Meeting times and location 8-9:15 am, Tu/Th, HRBB 124

Course Description and Prerequisites

This course will introduce the key concepts, models, algorithms and techniques at the core of modern text-based search systems. The topics covered in this course include efficient text indexing, Boolean, vector space and probabilistic retrieval models, retrieval evaluation, relevance feedback, document classification, learning to rank, document clustering and link analysis. The students will have an opportunity to implement key retrieval models on top of an open-source search engine system.

Prerequisites: CSCE 315 and junior or senior standing; or approval of the instructor. In addition, students should have had some exposure to basic probability, statistics, data structures and algorithms. You should be able to learn new software libraries on your own and design and develop functions on top.

Learning Outcomes or Course Objectives

The main objective of this course is to equip students with solid theoretical knowledge and enough practical experience to develop and diagnose their own search systems in the future. Upon successful completion of the course, a student will be able to:

- Define and explain the key concepts and models relevant to information storage and retrieval.
- Implement important algorithms, recognize and fix common problems in practice.
- Design and develop his/her own text search systems addressing customized information needs.

Instructor Information

Name   Ruihong Huang
Telephone number 979-862-2908
Email address   huangrh@cse.tamu.edu
Office hours   TBD
Office location Online

Textbook and/or Resource Material

The primary textbook: Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze, Cambridge University Press. 2008.

Instructor will provide copies of lecture slides and links to resources on the internet.

Grading Policies

The course grading policy is as follows: two programming assignments (30%), four written assignments (20%), class participation (5%), class project (25%) and final exam (20%). Class

participation will be measured through attendance in class, submission of assignments and participation in online piazza discussions.

The late policy is: for the programming/written homework assignments, you have a total of 5 late days that you can use during the semester. However, a single assignment can be submitted up to 2 days late only. For the purposes of the class, a late day is an indivisible 24-hour unit. Once you exhaust your 5 late days, we will not accept any late submissions.

Grading Scale

The grading scale is: A: 90-100, B: 80-89, C: 70-79, D: 60-69, F: 0-59. Scores are rounded up.

Attendance / Make-up Work Policies

The University views class attendance as the responsibility of an individual student. Attendance is essential to complete the course successfully. University rules related to excused/unexcused absences and make-up work are located on-line at http://student-rules.tamu.edu/rule07.

Course Topics, Calendar of Activities, Major Assignment Dates (Tentative)

Week	Topic
1	Overview and indexing
2	The Boolean Retrieval model
3	Probabilistic IR: Vector space model & BM25, #1 written assignment due
4	Probabilistic IR: Language Models
5	Retrieval Evaluation, #1 programming assignment due
6	Retrieval Feedback
7	Document Classification, #2 written assignment due
8	Learning to Rank
9	Flat Clustering: K-means, #2 Programming assignment due
10	Hierarchical Clustering: HAC
11	Link Analysis, #3 written assignment due
12	Trending Topics, #4 written assignment due
13	Trending Topics
14	Project Presentations

Other Pertinent Course Information

Americans with Disabilities Act (ADA)

Texas A&M University is committed to providing equitable access to learning opportunities for all students. If you experience barriers to your education due to a disability or think you may have a disability, please contact Disability Resources in the Student Services Building or at (979) 845-1637 or visit http://disability.tamu.edu. Disabilities may include, but are not limited to attentional, learning, mental health, sensory, physical, or chronic health conditions. All students are encouraged to discuss their disability related needs with Disability Resources and their instructors as soon as possible.

Academic Integrity
For additional information please visit: http://aggiehonor.tamu.edu