

Course Information

Course Number: CSCE 670
Course Title: Information Storage and Retrieval
Section: 600
Time: Mondays, Wednesdays, and Fridays 1:50pm – 2:40pm
Location: HRBB 113
Credit Hours: 3

Instructor Details

Instructor: Yu Zhang
Office: PETR 222
E-Mail: yuzhang@tamu.edu
Office Hours: Mondays 3:30pm – 4:30pm

Catalog Description

Information Storage and Retrieval. (3-0). Credit 3. Representation, storage, and access to very large multimedia document collections; fundamental data structures and algorithms of information storage and retrieval systems; techniques to design and evaluate complete retrieval systems, including cover of algorithms for indexing, compressing, and querying very large collections. Prerequisites: CSCE 310 or CSCE 603 or approval of instructor; graduate classification.

Course Prerequisites

Prerequisite/Corequisite(s): CSCE 310 or CSCE 603 or approval of instructor; graduate classification.

Course Learning Outcomes

Upon completion of this course, the learner will be able to:

- Define and explain the key concepts and models relevant to web search, including topics like text indexing, retrieval models, evaluation, Web crawling, link-based algorithms like PageRank, and learning to rank.
- Define and explain the key concepts and models relevant to recommender systems, including topics like collaborative filtering, matrix factorization, recommender system evaluation, and implicit recommendation.
- Design, implement, and evaluate the core algorithms underlying a fully functional web search system and recommendation system.
- Identify the salient features and apply recent research results in web search and recommender systems, including topics such as adversarial information retrieval and neural models of retrieval and recommendation.

Textbook and/or Resource Materials

This course does not mandate any textbook. The following two textbooks are optional and are available online.

- ***Introduction to Information Retrieval***, Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze, 2009.
 - Available online: <https://nlp.stanford.edu/IR-book/>
- ***Mining of Massive Datasets***, Jure Leskovec, Anand Rajaraman, and Jeffrey D. Ullman, 2014.
 - Available online: <http://www.mmds.org/>

Grading Policy

The course grading scale is:

- A : $\geq 90\%$
- B : $\geq 80\%$ and $< 90\%$
- C : $\geq 70\%$ and $< 80\%$
- D : $\geq 60\%$ and $< 70\%$
- F : $< 60\%$

Homework 30%: We will have five programming assignments (2%+7%+7%+7%+7%). All of them will be in Python. The first assignment is an introductory one designed to help you get familiar with Jupyter Notebook and the Python operations commonly used in this course.

Quizzes 20%: We will have four brief in-class quizzes (5%+5%+5%+5%).

Final 30%: We will hold a comprehensive final exam from 3:30 - 5:30 PM on May 4 in HRBB 113.

Group Project 20%: Students will work in teams of 3 or 4 to engage in a mini-research project.

Homework collaboration policy: Homeworks are individual assignments. That is, you should write your own code, by yourself. However, we do encourage you to talk amongst yourselves about approaches and methods. As an example, it is appropriate to sit in a group with no laptops to talk through the problems. It is inappropriate, however, to pair program, share code, or work through specific problems in code together. You must acknowledge all help in your homework submission.

Regrade policy: Once you receive your graded assignment (e.g., a homework assignment), you have seven days to request a regrade in writing. After seven days, there will be no regrades. You must give the instructor a written explanation of your regrade request. We reserve the right to regrade the entire assignment.

Late Work Policy

Work submitted by a student as makeup work for an excused absence is not considered late work and is exempted from the late work policy (Student Rule 7 (<https://student-rules.tamu.edu/rule07/>)).

Homework late day policy: All homeworks are due by 11:59pm on the due date. Each student gets five late days total. A late day is an indivisible 24-hour unit. For example, if the due date is 11:59pm on Monday, and you submit at 12:01am Tuesday, that is 1 late day. Once you use up the five late days, a late assignment will receive a grade of 0.

Course Schedule

(this is a rough schedule for the course; subject to change)

Week 1: Course Overview; Boolean Retrieval

Week 2: TF-IDF; Vector Space Model; BM25

Week 3: Link Analysis

Week 4: Evaluation of Search Engines

Week 5: Learning to Rank

Week 6: Recommender Systems (Content-Based; Collaborative Filtering)

Week 7: Recommender Systems (Matrix Factorization; Bayesian Personalized Ranking)

Week 8: word2vec; Neural Ranking

Week 9: Spring Break

Week 10: Transformer; BERT-Based Ranking

Week 11: Neural Collaborative Filtering; Sequential Recommendation

Week 12: Large Language Models Basics

Week 13: Large Language Models with Search Engines

Week 14: Large Language Models for Ranking and Recommendation

Week 15: Wrap-up; Project Presentations

Important Dates

Homeworks

- Homework 0: due Jan 26
- Homework 1: due Feb 9
- Homework 2: due Mar 2
- Homework 3: due Mar 30
- Homework 4: due Apr 20

Quizzes

- Quiz 1: in the Feb 6 class
- Quiz 2: in the Feb 27 class
- Quiz 3: in the Mar 25 class
- Quiz 4: in the Apr 15 class

Group Project

- Proposal: due Mar 7
- Presentation: in the Apr 20, Apr 22, Apr 24, and Apr 27 classes
- Report: due May 2

Final: May 4

AI Statement

AI policy: In principle you may submit AI-generated code, or code that is based on or derived from AI-generated code, as long as this use is properly documented in the comments: you need to include the prompt and the significant parts of the response. AI tools may help you avoid syntax errors, but there is no guarantee that the generated code is correct. It is your responsibility to identify errors in program logic through comprehensive, documented testing. Moreover, generated code, even if syntactically correct, may have significant scope for

improvement, in particular regarding separation of concerns and avoiding repetitions. The submission itself must meet our standards of attribution and validation.

Technology Support

Technology Services (IT) - Main Campus

Hours: 24/7

Phone: (979) 845-8300

Email: helpdesk@tamu.edu

Call/Chat/Email/visit: <https://it.tamu.edu/help>

Canvas LMS Technical Support

Hours: 24/7/365

Phone: (877) 354-4821

Email: support@instructure.com

Support is available by clicking the Help button at the far left in the Canvas global navigation menu.

Canvas Resources are also linked on the home page of every Canvas course.

Galveston Students:

Technology Services - Help Desk

E-mail: Helpdesk@tamug.edu

Phone: 409-740-4714

Location: Bldg 3007 (CLB), Rm 113

Hours: Monday - Friday 7:30am to 6:00pm

Canvas LMS Technical Support

Hours: 24/7/365

Phone: (877) 354-4821

Email: support@instructure.com

Support is available by clicking the Help button at the far left in the Canvas global navigation menu.

Canvas Resources are also linked on the home page of every Canvas course.

University Policies

This section outlines the university-level policies that must be included in each course syllabus. The TAMU Faculty Advisory Council established the wording of these policies.

Academic Integrity Statement and Policy

“An Aggie does not lie, cheat or steal, or tolerate those who do.”

“Texas A&M University students are responsible for authenticating all work submitted to an instructor. If asked, students must be able to produce proof that the item submitted is indeed the work of that student. Students must keep appropriate records at all times. The inability to authenticate one’s work, should the instructor request it, may be sufficient grounds to initiate an academic misconduct case” (Section 20.1.2.3, [Student Rule 20](#)).

You can learn more about the Aggie Honor System Office Rules and Procedures, academic integrity, and your rights and responsibilities at aggiehonor.tamu.edu.

University Attendance Policy

The university views class attendance and participation as an individual student responsibility. Students are expected to attend class and to complete all assignments.

Please refer to [Student Rule 7](#) in its entirety for information about excused absences, including definitions, and related documentation and timelines.

Makeup Work Policy

Students will be excused from attending class on the day of a graded activity or when attendance contributes to a student’s grade, for the reasons stated in Student Rule 7, or other reason deemed appropriate by the instructor.

Please refer to [Student Rule 7](#) in its entirety for information about makeup work, including definitions, and related documentation and timelines.

Absences related to Title IX of the Education Amendments of 1972 may necessitate a period of more than 30 days for make-up work, and the timeframe for make-up work should be agreed upon by the student and instructor” ([Student Rule 7, Section 7.4.1](#)).

“The instructor is under no obligation to provide an opportunity for the student to make up work missed because of an unexcused absence” ([Student Rule 7, Section 7.4.2](#)).

Students who request an excused absence are expected to uphold the Aggie Honor Code and Student Conduct Code. ([See Student Rule 24](#).)

Notice of Nondiscrimination

Texas A&M University is committed to providing safe and non-discriminatory learning, living, and work environments for all members of the University community. The University provides equal opportunity to all employees, students, applicants for employment or admission, and the public, regardless of race, color, sex (including pregnancy and related conditions), religion, national origin, age, disability, genetic information, or veteran status.

Texas A&M University will promptly, thoroughly, and fairly investigate and resolve all complaints of discrimination, harassment (including sexual harassment), complicity, and related retaliation based on a protected class in accordance with [System Regulation 08.01.01, University Rule 08.01.01.M1, Standard Administrative Procedure \(SAP\) 08.01.01.M1.01](#), and applicable federal and state laws. In accordance with Title IX and its implementing regulations, Texas A&M does not discriminate on the basis of sex in any educational program or activity, including admissions and employment.

The following person has been designated to handle inquiries and complaints regarding the non-discrimination policies: Jennifer M. Smith, TAMU Associate VP & Title IX Coordinator at YMCA Ste 108, College Station, TX 77843, 979-458-8407, or email civilrights@tamu.edu. For other reporting options, visit the [U.S. Department of Education Office for Civil Rights Complaint Assessment System](#) to locate the address and phone number of the office that serves your area, or call 1-800-421-3481.

Civil Rights, Free Speech, and Title IX Policies

Texas A&M University is committed to fostering a learning environment that is safe and productive for all. University policies and federal and state laws prohibit discrimination and harassment based on an individual's race, color, sex, (including pregnancy and related conditions), religion, national origin, age, disability, genetic information, veteran status, or any other legally protected characteristic. This includes forms of sex-based violence, such as sexual assault, sexual harassment, sexual exploitation, dating/domestic violence, and stalking.

Students can report discrimination/harassment, access supportive resources, or learn more about their options for resolving complaints on the [University's Civil Rights & Title IX webpage](#).

Students should be aware that all university employees (except medical or mental health providers) are mandatory reporters, which means that if they observe, experience or become aware of an incident that they reasonably believe to be discrimination/harassment alleged to have been committed by or against a person who was a student or employee at the time of the incident, the employee must report the incident to the university.

Americans with Disabilities Act (ADA) Policy

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 the Disability Resources office on your campus (resources listed below). 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.

To request academic accommodations, contact the designated ADA office based on your location:

- Texas A&M University, College of Nursing, College of Dentistry, Irma Lerma Rangel College of Pharmacy College Station, College of Medicine, School of Public Health, Institute of Biosciences and Technology, EnMed Program, Bush School in Washington DC, Mays Business School – CityCentre, TAMU Engineering Academies, Texas A&M University Higher Education Center at McAllen and Texas A&M University at Galveston should contact Disability Resources at (979) 845-1637 or disability@tamu.edu.
- Texas A&M University School of Law should contact the Office of Student Affairs at (817) 212-4111 or law-disability@law.tamu.edu to request accommodations.
- Irma Lerma Rangel College of Pharmacy in Kingsville should contact the Disability Resource Center at Texas A&M University-Kingsville at (361) 593-3024 or drc.center@tamuk.edu to request accommodations.
- Texas A&M University College of Veterinary Medicine & Biomedical Sciences in Canyon should contact the Office of Student Accessibility at West Texas A&M University – Canyon at (806) 651-2335 or osa@wtamu.edu.

If you are experiencing difficulties with your approved accommodations, contact the office responsible for approving your accommodations or the Texas A&M ADA Coordinator Julie Kuder at ADA.Coordinator@tamu.edu or (979) 458-8407.

Pregnancy Accommodations

Texas A&M provides reasonable accommodations to students due to pregnancy and/or related conditions, such as childbirth, recovery, and lactation. Students should contact the University's [Pregnancy Coordinator](#) as soon as they become aware of the need for accommodation. Depending on the circumstances, accommodations could include extended time to complete assignments or exams, changes in course sequence, or modifications to the physical classroom environment.

Texas A&M will also allow a voluntary leave of absence, ensure the availability of lactation space, and maintain grievance procedures to provide for the prompt and equitable resolution of complaints of sex discrimination. For information regarding pregnancy accommodations, email TIX.Pregnancy@tamu.edu.

Statement on Mental Health and Wellness

Texas A&M University recognizes that mental health and wellness are critical factors influencing a student's academic success and overall wellbeing. Students are encouraged to engage in healthy self-care practices by utilizing the resources and services available through [University Health Services](#). The [TELUS Health Student Support app](#) provides access to professional counseling in multiple languages anytime, anywhere by phone or chat, and the 988 Suicide & Crisis Lifeline offers 24-hour emergency support at 988 or [988lifeline.org](#).

Texas A&M College Station

Students needing a listening ear can contact University Health Services at 979.458.4584. Call 911 or visit your nearest emergency room if you are currently experiencing a life-threatening situation or if your safety is at risk. 24-hour emergency help is also available through the 988 Suicide & Crisis Lifeline (988) or at [988lifeline.org](#).

Statement on the Family Educational Rights and Privacy Act (FERPA)

FERPA is a federal law designed to protect the privacy of educational records by limiting access to these records, to establish the right of students to inspect and review their educational records, and to provide guidelines for the correction of inaccurate and misleading data through informal and formal hearings.

Currently enrolled students wishing to withhold any or all directory information items can do so within [howdy.tamu.edu](#) using the Directory Information Withholding Form. The complete [FERPA Notice to Students](#) and the student records policy is available on the Office of the Registrar webpage.

Items that can never be identified as public information are a student's social security number, citizenship, gender, grades, GPR, or class schedule. All efforts will be made in this class to protect your privacy and to ensure confidential treatment of information associated with or generated by your participation in the class.

Directory items include name, UIN, local address, permanent address, email address, local telephone number, permanent telephone number, dates of attendance, program of study (college, major, campus), classification, previous institutions attended, degrees, honors and awards received, participation in officially recognized activities and sports, medical residence location, and medical residence specialization.

Free Speech and Civil Discourse

Texas A&M recognizes that the pursuit of truth through open and robust discourse is critical to academic inquiry. However, as a community of scholars, the university has an aspirational

expectation that such discourse will be conducted in accordance with Aggie Core Values. In this “marketplace of ideas,” we encourage civil dialogue creating an environment that allows individuals to express their ideas and to have their ideas challenged in respectful and responsible ways. Students can learn more about Freedom of Expression and Free Speech on the [University's website](#) about the First Amendment.