CSCE 5200

Introduction to Information Retrieval

## Instructor Contact

**Name: Dr. Wei Jin**

**Office Location:** NTDP B190

**Phone Number:** (940) 369-7172

**Office Hours:** Monday at 5:30-7:30pm (via zoom) see Announcements for zoom meeting information.

**Email:** wei.jin@unt.edu

**Communication Expectations:** If you have questions regarding this class, please contact me via email. When you send emails, please use CSCE5200 as part of your subject. Please do not expect a response over the weekend. Feedback on assignments, exams, and grades will usually be posted within two weeks after the due date.

## Welcome to UNT!

As members of the UNT community, we have all made a commitment to be part of an institution that respects and values the identities of the students and employees with whom we interact. UNT does not tolerate identity-based discrimination, harassment, and retaliation. UNT’s full Non-Discrimination Policy can be found in the UNT Policies section of the syllabus.

## Course Description

Covers traditional material and recent advances in information retrieval, study of indexing, processing and querying textual data, basic retrieval models, algorithms and information retrieval system implementations. May covers advanced topics in intelligent information retrieval, including natural language processing techniques and link analysis.

## Course Structure

This course takes place 100% online. Your interaction with me and with your fellow students will take place in Canvas or by emails. There are 14 weeks of content that you will move through. I will open up a new module of 1 or 2 weeks each.

## Course Prerequisites

There are no required prerequisites for this course. However, in order to be successful in this course you will need to have:

* General programming experience, such as using C/C++, or Java or Python programming languages.
* Basic knowledge on data structures and algorithms, linear algebra and probability theory.

## Course Objectives

By the end of this course, students will be able to:

1. Demonstrate understanding of core concepts in text-based information retrieval techniques, such as inverted indexes, vector space model, and TF-IDF term weighting.
2. Understanding the mechanism of Boolean retrieval models. Show how Boolean queries can be processed and how to optimize query processing.
3. Demonstrate understanding of linguistic modules used in document preprocessing and show how to process queries that have spelling errors and other imprecise matches to the vocabulary used in the document collection.
4. Demonstrate understanding of ranked retrieval models and able to apply and analyze a number of algorithms for constructing the inverted index from a text collection.
5. Demonstrate understanding of frequent and basic evaluation measures for information retrieval effectiveness and able to apply them in different scenarios.
6. Design and implement a complete information retrieval system and evaluate its performance by applying the learned knowledge.

## Required Texts

Introduction to Information Retrieval, 1st Edition, by Christopher D. Manning, Hinrich Schutze, and Prabhakar Raghavan ISBN-13: 978-0521865715

### Technical Assistance

Part of working in the online environment involves dealing with the inconveniences and frustration that can arise when technology breaks down or does not perform as expected. Here at UNT we have a Student Help Desk that you can contact for help with Canvas or other technology issues.

### Rules of Engagement

Rules of engagement refer to the way students are expected to interact with each other and with their instructors. Here are some general guidelines:

* While the freedom to express yourself is a fundamental human right, any communication that utilizes cruel and derogatory language on the basis of race, color, national origin, religion, sex, sexual orientation, gender identity, gender expression, age, disability, genetic information, veteran status, or any other characteristic protected under applicable federal or state law will not be tolerated.
* Treat your instructor and classmates with respect in any communication online or face-to-face, even when their opinion differs from your own.
* Ask for and use the correct name and pronouns for your instructor and classmates.
* Speak from personal experiences. Use “I” statements to share thoughts and feelings. Try not to speak on behalf of groups or other individual’s experiences.
* Use your critical thinking skills to challenge other people’s ideas, instead of attacking individuals.
* Avoid using all caps while communicating digitally. This may be interpreted as “YELLING!”
* Be cautious when using humor or sarcasm in emails or discussion posts as tone can be difficult to interpret digitally.
* Avoid using “text-talk” unless explicitly permitted by your instructor.
* Proofread and fact-check your sources.
* Keep in mind that online posts can be permanent, so think first before you type.

See these [Engagement Guidelines](https://clear.unt.edu/online-communication-tips) (https://clear.unt.edu/online-communication-tips) for more information.

## Course Requirements

| ***Assignment*** | ***Percentage of Final Grade*** |
| --- | --- |
| ***Homeworks*** | *10%* |
| ***Quizzes*** | *15%* |
| ***Midterm Exam*** | *20%* |
| ***Final Exam*** | *20%* |
| ***Project*** | *35%* |
| ***Total Points Possible*** | *100%* |

## Grading

A: 90-100% (Outstanding, excellent work. The student performs well above the minimum criteria.)  
B: 80-89% (Good, impressive work. The student performs above the minimum criteria.)  
C: 70-79% (Solid, college-level work. The student meets the criteria of the assignment.)  
D: 60-69% (Below average work. The student fails to meet the minimum criteria.)  
F: 59% and below (Sub-par work. The student fails to complete the assignment.)

## Course Policies

### Assignment Policy and Content Responsibility

### You are expected to learn course materials and complete all readings and assignment on time. Students are responsible for all content presented in lecture slides, in the videos, and required readings from the textbook. All class materials will be posted on Canvas. For assignments (including projects), they are individual work, and no collaboration is allowed.

The University is committed to providing a reliable online course system to all users. However, in the event of any unexpected server outage or any unusual technical difficulty which prevents students from completing a time sensitive assessment activity, the instructor will extend the time windows and provide an appropriate accommodation based on the situation. Students should immediately report any problems to the instructor and contact the UNT Student Help Desk: [helpdesk@unt.edu](mailto:helpdesk@unt.edu) or 940.565.2324 and obtain a ticket number. The instructor and the UNT Student Help Desk will work with the student to resolve any issues at the earliest possible time.

### Examination Policy

For each graded exam and quiz, you will be required to use Lockdown Browser with a webcam. You are not allowed to use your cell phones for any reason during an exam or quiz. For exams and quizzes, they are closed-book and closed-notes.

Late WorkNo late submission is accepted.

Attendance Policy

While attendance will not be taken into grades, you will be expected to know and understand the requisite topics and concepts. You are responsible for checking announcements often to keep in touch with the whole class about upcoming events, assignments, or activities.

### Academic Integrity Policy

### Cheating of any sort will not be tolerated in this course.  All work turned in with your name on it must be your own work. Failure to adhere to these strict standards will be cause for disciplinary action that could be as severe as expulsion from the university. If it is determined a student cheated on any assignment in this course, they will receive an F for their final course grade and an academic integrity report will be filed with the Office of Academic Integrity. Further, UNT is now maintaining a database recording any acts of academic dishonesty that is available to employers. For more information see the UNT Student Academic Integrity Policy.