CSCE 5200: Information Retrieval and Web Search  
COURSE SYLLABUS, Spring 2022  
Department of Computer Science and Engineering  
University of North Texas

Instructor  
J. U. QUEVEDO-Torrero  
Jesusubaldo.Quevedo-torrero@unt.edu (preferred communication option)

Office: Discovery Park Building E250G  
Office Hours: Mondays and Tuesdays 11:30am to 1:30PM, Fridays 12:30: 1:30PM or by appointment.

Teaching Assistants:  
Shou Sun (shuosun2@my.unt.edu)  
Vasanth Kandagatla (vasanth.kandagatla@unt.edu)  
Alankruth Mallampeta (alankruthmallampeta@my.unt.edu)

TA Office Hours: TBD or by appointment.

Course Description:  
This course will introduce students to text-based information retrieval (IR) techniques, i.e. search engines (e.g., Google, Yahoo, etc.). Various IR models such as the Boolean model, vector space model, and probabilistic models will be studied. Efficient indexing techniques for both general document collections and specialized collections (strings, digital library) will be examined. The course will also cover web search engines techniques, such as hyperlink analysis (e.g., PageRank (used in Google search)), Web technologies and representations, and query languages, with a focus on techniques that can be used to access, retrieve, organize, and present information. Students will work on programming projects to gain hands-on experience in building an IR system.

Course Structure  
Time: Friday 9:00AM-11:50AM  
Location: NTDP B155  
Credit hours: 3  
Dates: January 18 to May 13, 2022

Communication Expectations: You will be expected to regularly check university email and attend class regularly. When you miss a class, you are expected to check the course calendar shortly after class to be aware of assignments, quizzes, and other materials. Questions not answered in class are best asked before or after class. For in-depth assistance on course content, you are expected to meet with the TA prior to meeting with the instructor. For quick questions, email is preferred, and you can expect a response within 24 hours during the work week (M-F). For involved questions or discussions not appropriate for the TA, office hours are preferred.
**Tentative topics**
Boolean retrieval
The term vocabulary & postings lists
Dictionaries and tolerant retrieval
Index construction
Scoring, term weighting & the vector space model
Computing scores in a complete search system
Evaluation in information retrieval
Relevance feedback & query expansion
Probabilistic information retrieval
Language models for information retrieval
Web search basics
Web search and retrieval
Web link analysis

**PREREQUISITE**
Programming Experience and Data Structures

**Materials**
All materials (readings, videos, tutorials, quizzes, and assignments) will be accessible online and posted on the course Canvas site on the respective class day at the latest.

**REQUIRED TEXT**
- Introduction to Information Retrieval, by C. Manning, P. Raghavan, and H. Schütze, Cambridge University Press
  Note: an online version of this book is available at [http://informationretrieval.org](http://informationretrieval.org)

**Reference Book:**
- Modern Information Retrieval, by Ricardo Baeza-Yates, Berthier Ribeiro-Neto

**Technical Requirements and Skills**
**Minimum Technology Requirements**
- Computers are **required for exams**
  - **in-class exams**: You will need to bring a laptop on exam days. These will be done individually on your computer in class. You will be expected to connect to the UNT wireless network.
  - Quizzes and exams will use the Canvas quiz system
    - [Canvas Technical Requirements](https://clear.unt.edu/supported-technologies/canvas/requirements)
- Computers are **optional during class, outside of exam times**
  - There will be occasional in-class activities along with lecture that are not required, but may help in understanding and applying the material
**Course Requirements**

**Readings, Quizzes, and Exams**

**Quizzes (in-class/take home):** These quizzes are meant to focus students on the important aspects of the readings or lectures.

**Exams:** Exam days are already posted and are considered fixed. **Missed exams:** Exams cannot be missed without prior arrangements or later documented proof of extenuating circumstances.

**Assignments.** There are some assignments. You need to solve the given problems and submit your answers by the due date.

**Project:** Students (Individual or a group of two students) are required to read a recent research paper in information retrieval and Web search. The paper list will be given later. You need to prepare a presentation and talk about the main content of the technical paper.

**GRADING**

Your course grade will be determined by a combination of the following:

- Quizzes(in-Class/take-home): 10%
- Assignments: 25%
- Paper Presentation: 20 %
- Exams: 45%
  - Exam I: 10 %
  - Exam II: 13 %
  - Final Exam: 22 %

Total 100%

**Grading Scale:** A=90, B=80-89.9, C=70-79.9, D=60-69.9, F=0-59.9 %.

**Course Evaluation**

Student Perceptions of Teaching (SPOT) is the student evaluation system for UNT and allows students the ability to confidentially provide constructive feedback to their instructor and department to improve the quality of student experiences in the course.

**Expectations**

1. **Examination Policy:** Exams will be on the computer using the Canvas quiz system. You need to bring a laptop on the appropriate exam days. Quizzes and exams must be taken in the classroom unless special accommodations have been made through the Office of Disability Accommodation (ODA). Another other accommodations must
be given by prior arrangement with the instructor, otherwise documentation proving an extenuating circumstance will have to be provided after the missed exam. Time will be limited, and all work will be individual. Exams will focus on the most recent material but are expected to be cumulative in scope.

2. **Technical errors during exams:** If during an online quiz or exam there is a technical error which affects your ability to complete the assignment, you are immediately to let the quiz or exam proctor know and the instructor will discuss ways to allow you to resume the test without giving an unfair advantage. 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.

3. Each student will be responsible for completing the assigned reading, exercises and attending classes.

4. **Attendance:** You are expected to attend lectures and to complete all readings, however, this course does not use participation points and there is no penalty for missing days without exams, or group project efforts. There is no need to let the instructor know you have missed a class; however, you are responsible for keeping up with the material covered in the class if you are not present. If a class is missed, you are expected to proactively reach out to classmates, the TA, or the instructor if there are any questions. Individual attendance is not required except on group projects and exam days. There is no direct participation grading. **There will be a sign-in sheet to record attendance for class analytics and administration.**

5. If you miss a class, you are still responsible for knowing everything that took place. Your absence does not change the due date of an assignment.

6. Generally, late submission of any class work is not allowed. However, late assignments might be submitted with reduction of points 20% per weekday.

7. **Cheating:** Cheating on tests and programs will be dealt with very severely. You must make a diligent effort to prevent other students from seeing your test answers. Keep your paper covered and do not let your eyes wander during tests. You should not receive or give help to others on any program that goes beyond help in deciphering syntax errors. **Plagiarism:** Plagiarism is a form of cheating. Copying someone else’s program, changing a few lines, and turning it in as your own is plagiarism; thus, this is cheating. Each student is to write his or her own programs.

8. Topics or discussions unrelated to class, suggestions about the logistic of the course are all welcome outside class, but are considered disruptive during class and will affect negatively your “class contribution” grade, and may impact at the discretion of the instructor your final grade.

9. If you use laptops and other electronic devices, they should be used to enhance your class engagement level.

10. **Cellular Telephones and Pagers in Class and Lab:** Along with your instructors, many students find these both distracting and rude. As a courtesy to all involved, please either turn off your cellular telephone or pager or disable the ring tone during lecture and lab. If you must use the phone, please leave the classroom or lab and go to a place that will not interrupt others.

11. **Course Schedule**
<table>
<thead>
<tr>
<th>Date</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-21-22</td>
<td>Class 1</td>
</tr>
<tr>
<td>1-28-22</td>
<td>Class 2</td>
</tr>
<tr>
<td>2-4-22</td>
<td>Class 3</td>
</tr>
<tr>
<td>2-11-22</td>
<td>Class 4</td>
</tr>
<tr>
<td>2-18-22</td>
<td>Class 5</td>
</tr>
<tr>
<td>2-25-22</td>
<td>Class 6, Exam 1</td>
</tr>
<tr>
<td>3-4-22</td>
<td>Class 7</td>
</tr>
<tr>
<td>3-11-22</td>
<td>Class 8</td>
</tr>
<tr>
<td>3-18-22</td>
<td>UNT Spring Break</td>
</tr>
<tr>
<td>3-25-22</td>
<td>Class 9</td>
</tr>
<tr>
<td>4-1-22</td>
<td>Class 10</td>
</tr>
<tr>
<td>4-8-22</td>
<td>Class 11</td>
</tr>
<tr>
<td>4-15-22</td>
<td>Class 12, Exam 2</td>
</tr>
<tr>
<td>4-22-22</td>
<td>Class 13</td>
</tr>
<tr>
<td>4-29-22</td>
<td>Class 14</td>
</tr>
<tr>
<td>5-?-22</td>
<td>Final Exam</td>
</tr>
</tbody>
</table>
UNT Policies

Academic Integrity Policy
Academic Integrity Standards and Consequences. According to UNT Policy 06.003, Student Academic Integrity, academic dishonesty occurs when students engage in behaviors including, but not limited to cheating, fabrication, facilitating academic dishonesty, forgery, plagiarism, and sabotage. A finding of academic dishonesty may result in a range of academic penalties or sanctions ranging from admonition to expulsion from the University.

ADA Policy
UNT makes reasonable academic accommodation for students with disabilities. Students seeking accommodation must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide a student with an accommodation letter to be delivered to faculty to begin a private discussion regarding one’s specific course needs. Students may request accommodation at any time; however, ODA notices of accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of accommodation for every semester and must meet with each faculty member prior to implementation in each class. For additional information see the ODA website (https://disability.unt.edu/).

Emergency Notification & Procedures
UNT uses a system called Eagle Alert to quickly notify students with critical information in the event of an emergency (i.e., severe weather, campus closing, and health and public safety emergencies like chemical spills, fires, or violence). In the event of a university closure, please refer to Blackboard for contingency plans for covering course materials.

Retention of Student Records
Student records pertaining to this course are maintained in a secure location by the instructor of record. All records such as exams, answer sheets (with keys), and written papers submitted during the duration of the course are kept for at least one calendar year after course completion. Course work completed via the Canvas online system, including grading information and comments, is also stored in a safe electronic environment for one year. Students have the right to view their individual record; however, information about student’s records will not be divulged to other individuals without proper written consent. Students are encouraged to review the Public Information Policy and the Family Educational Rights and Privacy Act (FERPA) laws and the University’s policy. See UNT Policy 10.10, Records Management and Retention for additional information.

Acceptable Student Behavior
Student behavior that interferes with an instructor’s ability to conduct a class or other students' opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Dean of Students to consider whether the student's conduct violated the Code of Student Conduct. The University's expectations for student conduct apply to all instructional forums,
including University and electronic classroom, labs, discussion groups, field trips, etc. Visit UNT’s Code of Student Conduct (https://deanofstudents.unt.edu/conduct) to learn more.

**Access to Information - Eagle Connect**
Students’ access point for business and academic services at UNT is located at: my.unt.edu. All official communication from the University will be delivered to a student’s Eagle Connect account. For more information, please visit the website that explains Eagle Connect and how to forward e-mail Eagle Connect (https://it.unt.edu/eagleconnect).

**Student Evaluation Administration Dates**
Student feedback is important and an essential part of participation in this course. The student evaluation of instruction is a requirement for all organized classes at UNT. The survey will be made available during weeks 13, 14 and 15 of the long semesters to provide students with an opportunity to evaluate how this course is taught. Students will receive an email from "UNT SPOT Course Evaluations via IASystem Notification" (noreply@iasystem.org) with the survey link. Students should look for the email in their UNT email inbox. Simply click on the link and complete the survey. Once students complete the survey they will receive a confirmation email that the survey has been submitted. For additional information, please visit the SPOT website (http://spot.unt.edu/) or email spot@unt.edu.

**Getting Help**

**Technical Assistance**
UIT Help Desk (http://www.unt.edu/helpdesk/index.htm)
Email: helpdesk@unt.edu
Phone: 940-565-2324
In Person: Sage Hall, Room 130
Walk-In Availability: 8am-9pm
**Telephone Availability:**
- Sunday: noon-midnight
- Monday-Thursday: 8am-midnight
- Friday: 8am-8pm
- Saturday: 9am-5pm
**Laptop Checkout:** 8am-7pm

**Student Support Services**
- Registrar (https://registrar.unt.edu/registration)
- Financial Aid (https://financialaid.unt.edu/)
- Student Legal Services (https://studentaffairs.unt.edu/student-legal-services)
- Career Center (https://studentaffairs.unt.edu/career-center)
- Multicultural Center (https://edo.unt.edu/multicultural-center)
- Counseling and Testing Services (https://studentaffairs.unt.edu/counseling-and-testing-services)
• Student Affairs Care Team (https://studentaffairs.unt.edu/care)
• Student Health and Wellness Center (https://studentaffairs.unt.edu/student-health-and-wellness-center)
• Pride Alliance (https://edo.unt.edu/pridealliance)
• Academic Support Services
• Academic Resource Center (https://clear.unt.edu/canvas/student-resources)
• Academic Success Center (https://success.unt.edu/asc)
• UNT Libraries (https://library.unt.edu/)
• Writing Lab (http://writingcenter.unt.edu/)
• MathLab (https://math.unt.edu/mathlab)