UTRGV

CSCI 6370.01 Information Retrieval & Web Search	**Dr. Zhixiang Chen**
Summer I 2020	Phone: 665-2857, Email: zhixiang.chen@utrgv.edu
Class time: TR 6:30 pm to 10:15 pm	Office: EIEAB 3.241
ONLINE - Asynchronous	**Office Hours: ONLINE (by Appointment)**

Textbook and/or Resource Material

We will use BlackBoard for disseminating course materials and submitting assignments.
There is no required textbook. I recommend "Modern Information Retrieval" by Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison-Wesley, 1999, or the latest edition.

Other suggested materials: will be given in class as the semester progresses.

Course Description and Prerequisites

Here are the course descriptions from UTRGV Catalog:

CSCI 6370 Topics in Computer Science
In-depth study of specific issues in computer science. Subject matter varies from semester to semester. May be repeated when subject matter changes. A total of six hours may be counted toward fulfillment of degree requirements. Prerequisite: Consent of instructor.

Specific Topics for Summer I 2020: Web Search Engines and Information Retrieval. The following will be covered: information retrieval models, evaluations, query languages, query operations, text properties and operations, indexing and searching, user interface, and web search engines, text clustering, linkage analysis. In addition, Python programming will also be covered in class.

Teaching Philosophy

I would like to help my students become self-motivated learners who not only master the content of the course but also are able to apply the learned knowledge to practice. I expect my students to gain fruitful attainment in critical thinking and technical innovation through completing integrated components of coursework.

Learning Objectives/Outcomes for the Course

After completing this course, students should be able to
- Understand how search engines work.
- Understand the limits of existing search technology.
- Learn to appreciate the sheer size of the Web.
- Learn to write code for text indexing and retrieval.
- Learn about the state of the art in IR research.
- Learn to analyze textual and semi-structured data sets.
- Learn to appreciate the diversity of texts on the Web.
- Learn to evaluate information retrieval systems.
- Learn about standardized document collections.
- Learn about text similarity measures.
- Learn about semantic dimensionality reduction.
- Learn about the idiosyncrasies of hyperlinked document collections.
- Learn about web crawling.

- Learn to use existing tools of information retrieval.
- Understand the dynamics of the Web by building appropriate mathematical models.
- Build working systems that assist users in finding useful information on the Web.

Student Learning Outcomes:

Upon successful completion of the course, students are able to:
- Demonstrate competency in basic information retrieval techniques
  - Understand the concept of index terms and their use in an inverted index.
  - Understand how to score and rank query results.
  - Understand the criteria to evaluate the results of an IR system.
  - Understand the basic ideas of text classification and clustering.
- Demonstrate competency in automated web search
  - Understand how the web is organized and its fundamental properties.
  - Understand how search engines collect and index web content.
  - Understand how web search engines present the most relevant results for a given query.
  - Understand how to design and construct software that implements significant web IR concepts.

Learning Objectives for Core Curriculum Requirements

This course is not a part of the University Core Curriculum.

Exam, Assignment and Grading:

Weekly exercises/quizzes/tests	20%
Final Exam	20%
Group Project mySearchEngine.com (for a corpus of web pages – 9,400 files, 180 folders, 37 MB) Design, Implementation and Deployment to AWS	45% Top 3 projects will receive Big Bonus Reward.
Term paper (just one)	10%
Attendance	5%
total	100%

The letter grade will be determined as follows:

A: 90-100%	B: 80-89%	C: 70-79%,	D: 60-69%	F: 0-59%

Assignment Policies:

- All assignments must be submitted through Blackboard. No other means of submission is allowed.
- Each assignment has a due date and a grace period for late submission with 20% penalty. The due dates and grace periods are enforced by Blackboard. Once the time passes, the submission option will no longer be available.
- Assignments will be graded on the basis of correctness, logic, clearness, motivation, and style.
- Unless stated otherwise, all assignments are individual assignments and are expected to be a student's own work. General discussions regarding understanding problems are encouraged but giving or

receiving major sections of solutions to problems will be considered cheating and will be dealt with on an individual basis.

Course Organization

- Week 1: Introduction, IR models, implementation
- Week 2: Evaluation, query operation, query languages
- Week 3: Text properties, Web search, spidering/crawling
- Week 4: Web interfaces, link analysis, text categorization
- Week 5: Text clustering, recommenders, topics of interests

Attendance:

You are responsible for all materials covered in class, the textbook, and homework assignments.

Changes in the course content

Changes in the course content and topics may be made by the instructor as deemed necessary.

Blackboard Orientation

To actively participate in this course, one needs to get familiar with the course environment. This course uses Blackboard tools. One can get familiar with these tools by self-enrolling and reviewing the orientation provided:
- In Blackboard, use the **Course Search** box and enter *Blackboard Orientation*
- Click on Go
- Click on the **Enroll** button to include the *orientation course* in your *Blackboard MyCourses* list

Contacting Instructor and Instructor's Response Time

Use instructor's email given above for contacting him regarding the course content. Normally, the instructor will respond within two working weekdays. Any exceptions will be announced through Blackboard announcement mechanism. Technical questions on Blackboard should be referred to Blackboard Support (see under Course Technology).

Grades are normally updated each time a grading session has been completed – typically within two working weekdays or before the next class session.

Course Delivery Method due to COVID-19 Situation

Due to the current COVID-19 pandemic, the delivery of the course, and interaction between students and instructor, will be via *online* tools. Below is a summary:
1. **Class notes, recordings, and attending classes online:** The class-notes, recordings, and other material will be available *online* at Blackboard. Students will be required to complete any assigned work before attending the class online. *Attending the online classes is mandatory*. *Blackboard Collaborate Ultra* needs to be used to attend the class online. Below are links to Blackboard Student Support:
    - Student Preparation Checklist (Technology):
        https://www.utrgv.edu/online/academic-continuity/students/index.htm
    - How-To *Guide* and *Tutorials*:
        https://www.utrgv.edu/online/academic-continuity/resources/index.htm

2. **Syllabus, grading policy, and exam dates/times:** Syllabus has the detail on grading policy and exam dates/times. There cannot be an exception. Of course, the exam dates/times may be changed due to unforeseen situation at the university level, which will be communicated to you as soon as the university announces any changes.
3. **Assignments, Quizzes and Exams:** Assignments, Quizzes and Exams are required to be submitted online through Blackboard. There will be quizzes and exams online during the class time. Note that a quiz or an exam will be conducted *synchronously* during an online class for everyone at the same time. Students may not take a quiz or an exam at different times.
4. **Online tutoring:** Teaching Assistants will be available through *online*. The ONLINE_Tutoring_Schedule (along with instructions of how to connect with a tutor online) will be available on Blackboard under **Reference Material** folder.
5. **Online office hours:** We will be meeting online at the scheduled class time, and I will be answering your questions about the recorded lectures/notes and conducting quizzes/exams. Additionally, you may request separate online appointment through email. Students may not visit the instructor or a Teaching Assistant in person at this time.
6. **Visiting on-campus computer labs:** Contact Help Desk (956-882-2020) to know which labs are available for you to use, and any restrictions associated with your visit. If any students respond YES to any of the four (4) items listed below, they should not visit campus and be present in a social-distanced space with other individuals:
   - Have you felt like you have a fever in past 48 hours?
   - Do you have a cough?
   - Have you been to Europe or Asia in the past 14 days?
   - Have you been in personal contact with someone who has known or suspected coronavirus (COVID-19) infection?

CALENDAR OF ACTIVITIES

Refer to Section "*Course Organization*" above for subject topics planned for each week.

The UTRGV academic calendar can be found at https://my.utrgv.edu/home at the bottom of the screen, *prior to login*. Some important dates for Summer I 2020 include:
- *June 1 First day of classes*
- *June 2 Last day to add a course or register for Summer I 2020*
- *June 25 Last day to drop (DR grade) a class or withdraw (grade of W)*
- *July 4 Independence Day – NO classes*
- *July 6 Study Day – NO classes*
- *July 7 Final Exams*
- *July 7 Summer I classes end; Official last day of the term*

Additional Course Information

The lab work required in the course is integrated with the course itself in the form of assignments/projects. Besides the formal allotted time, a student is expected to spend a large number of hours working on projects, assignments, reading material, recorded lectures, and videos. *The general rule is that for every 1 hour in class a student is expected to put in 3 or more hours outside the class.* You may only do work related with the class when in class. *Activities unrelated with the class (such as texting, talking, browsing, unless otherwise explicitly instructed by the instructor to do so) are prohibited during a class session.*

UTRGV Policy Statements

STUDENTS WITH DISABILITIES:
Students with a documented disability (physical, psychological, learning, or other disability which affects academic performance) who would like to receive academic accommodations should contact **Student Accessibility Services (SAS)** as soon as possible to schedule an appointment to initiate services. Accommodations can be arranged through SAS at any time, but are not retroactive. Students who experience a broken bone, severe injury, or undergo surgery during the semester are eligible for temporary services.

Pregnancy, Pregnancy-related, and Parenting Accommodations
Title IX of the Education Amendments of 1972 prohibits sex discrimination, which includes discrimination based on pregnancy, marital status, or parental status. Students seeking accommodations related to pregnancy, pregnancy-related condition, or parenting (reasonably immediate postpartum period) are encouraged to contact Student Accessibility Services for additional information and to request accommodations.

Student Accessibility Services:
Brownsville Campus: Student Accessibility Services is located in 1.107 in the Music and Learning Center building (BMSLC) and can be contacted by phone at (956) 882-7374 or via email at ability@utrgv.edu.

Edinburg Campus: Student Accessibility Services is located in 108 University Center (EUCTR) and can be contacted by phone at (956) 665-7005 or via email at ability@utrgv.edu.

MANDATORY COURSE EVALUATION PERIOD:
Students are required to complete an ONLINE evaluation of this course, accessed through your UTRGV account (http://my.utrgv.edu); you will be contacted through email with further instructions. Students who complete their evaluations will have priority access to their grades. Online evaluations will be available on or about:

Module 1	October 2nd – 8th
Module 2	November 27th – December 3rd
Full Fall Semester	November 14th – December 4th

SEXUAL MISCONDUCT and MANDATORY REPORTING:
In accordance with UT System regulations, your instructor is a "Responsible Employee" for reporting purposes under Title IX regulations and so must report to the Office of Institutional Equity & Diversity (oie@utrgv.edu) any instance, occurring during a student's time in college, of sexual misconduct, which includes sexual assault, stalking, dating violence, domestic violence, and sexual harassment, about which she/he becomes aware during this course through writing, discussion, or personal disclosure. More information can be found at www.utrgv.edu/equity, including confidential resources available on campus. The faculty and staff of UTRGV actively strive to provide a learning, working, and living environment that promotes personal integrity, civility, and mutual respect that is free from sexual misconduct, discrimination, and all forms of violence. If students, faculty, or staff would like confidential assistance, or have questions, they can contact OVAVP (Office for Victim Advocacy & Violence Prevention) at 665-8287, 882-8282, or OVAVP@utrgv.edu.

COURSE DROPS:
According to UTRGV policy, students may drop any class without penalty earning a grade of DR until the official drop date. Following that date, students must be assigned a letter grade and can no longer drop the class. Students considering dropping the class should be aware of the "3-peat rule" and the "6-drop" rule so they can recognize how dropped classes may affect their academic success. The 6-drop rule refers to Texas law that dictates that undergraduate students may not drop more than six courses during their undergraduate career. Courses dropped at other Texas public higher education institutions will count toward the six-course drop limit. The 3-peat rule refers to additional fees charged to students who take the same class for the third time.

STUDENT SERVICES:
Students who demonstrate financial need have a variety of options when it comes to paying for college costs, such as scholarships, grants, loans and work-study. Students should visit the Students Services Center (U Central) for additional information. U Central is located in BMAIN 1.100 (Brownsville) or ESSBL 1.145 (Edinburg) or can be

reached by email (ucentral@utrgv.edu) or telephone: (888) 882-4026. In addition to financial aid, U Central can assist students with registration and admissions.

Students seeking academic help in their studies can use university resources in addition to an instructor's office hours. University Resources include the Advising Center, Career Center, Counseling Center, Learning Center, and Writing Center. The centers provide services such as tutoring, writing help, critical thinking, study skills, degree planning, and student employment. Locations are:

Center Name	Brownsville Campus	Edinburg Campus
Advising Center AcademicAdvising@utrgv.edu	BMAIN 1.400 (956) 665-7120	ESWKH 101 (956) 665-7120
Career Center CareerCenter@utrgv.edu	BCRTZ 129 (956) 882-5627	ESSBL 2.101 (956) 665-2243
Counseling Center Counseling@utrgv.edu	EUCTR 109 (956) 665-2574	BSTUN 2.10 (956) 882-3897
Learning Center LearningCenter@utrgv.edu	BMSLC 2.118 (956) 882-8208	ELCTR 100 (956) 665-2585
Writing Center WC@utrgv.edu	BUBLB 3.206 (956) 882-7065	ESTAC 3.119 (956) 665-2538