Course	**CS 6322 Information Retrieval**
Professor	Sanda Harabagiu
Term	Spring 2024
Meetings	MW 4:00-5:15 PM SLC 1.102

Professor's Contact Information

Office Phone	(972) 883-4654
Office Location	ECSS 3.411
Email Address	sanda@utdallas.edu
Office Hours	MW 7:00PM-8:00 PM on eLearning Blackboard Collaborate
Other Information	Class Web Page: www.utdallas.edu/~sanda/courses/IR/cs6322.html

Instructional Mode	**F**ace-to-face. Only in exceptional cases the class will be held online.
Expectations	Students in this class are expected to **study both** the lecture material (slides and discussed handouts) as well as the textbook material. It is recommended that students first study the textbook material before it is presented in class.

Class Participation

Regular class participation is expected. Students who fail to participate in class regularly are inviting scholastic difficulty. Successful participation is defined as consistently adhering to University requirements. Questions and discussions regarding the class material shall take place ONLY during the Professor's office hours, **not through emails.** Questions and discussions regarding the homeworks shall take place ONLY during the TA's office hours, **not through emails.** Students should not ask questions through email or Teams massaging, but instead prepare to come from the beginning of office hours to the end of them – to hear all discussions that are taking place. When raising a question during office hours, students should first present what they have tried to do for solving their doubts/problems. Students should not email the TA questions about the homeworks. Failure to follow this policy will lead to a reduction of 1 point from their final grade per incidence. Students should attend the TA office hours from the beginning of the time period allotted to these office hours and participate in all the discussions regarding the homeworks. Questions about a homework grade or examination grade will be answered ONLY within 10 days AFTER the grade is issued by the Faculty/TA. Failure to attend office hours and requesting help through emails in the last minute will not be honored by the Professor or the TA, and will lead to reduction of the overall grade, without warnings.

Class Materials

The instructor may provide class materials that will be made available to all students registered for this class as they are intended to supplement the classroom experience. These materials may be downloaded during the course, however, these materials are for registered students' use

only and are copyrighted by the Professor. All materials used in this class, including slides, homeworks and exams are copyrighted by the Professor and no student is allowed to post on the Internet any of this material, even after the end of class. Copyright infringements shall be sought legally, which will affect the immigration status of the students. Using any homeworks or exams that were given in this class in previous years is considered cheating and will automatically lead to a grade of ZERO. Classroom materials may not be reproduced or shared with those not enrolled in the class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation.

General Course Information

Pre-requisites, Co-requisites, & other restrictions	CS 5343 Algorithm Analysis & Data Structures *Excellent programming skills*
Course Description	The graduate Computer Science class CS 6322 on Information Retrieval provides a strong grounding in the fundamentals of organizing on-line information, Web search/crawling and digital libraries. This course is intended to prepare you to design, develop and use information systems. The course explores the practices, issues and theoretical foundations of organizing and analyzing information and information content for the purpose of providing access to textual and non-textual information resources. Students will learn how effective information search and retrieval is interrelated with the organization and description of information to be retrieved. More information is available at: *http://www.utdallas.edu/~sanda/courses/class_description/html_ir/cs6322.html*
Learning Outcomes	CS 6322 focuses on the basic IR techniques, including indexing, clustering, crawling and satisfying information needs. The fundamental algorithms and techniques for each of these areas of information retrieval are studied. Students will learn the characteristics of Boolean and Vector Retrieval relevance models. A variety of current research topics are also covered, including modern distributed indexing along with other forms of indexing. During this class state-of-the art techniques for building search engines will be covered and each student will gain hands-on experience with programming efficient up-to-date search engines. In addition, relevance feedback principles shall be studied and knowledge about flat and hierarchical clustering shall be acquired.
Required Texts & Materials	Introduction to Information Retrieval by Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze Cambridge University Press. 2008, ISBN 978-0-521-86571-5
Suggested Texts, Readings, & Materials	As provided in class/class web page.

Assignments & Academic Calendar

January 17 2024	Introduction to Information Retrieval	Syllabus Issued
January 22 2024	Boolean Retrieval	
January 24 2024	The Term Vocabulary and Posting List	
January 29 2024	Stemming and Skip pointers	Homework 1 issued
January 31 2024	Dictionaries and Tolerant Retrieval	
February 5 2024	Index Construction	
February 7 2024	Dynamic Index Construction	
February 12 2024	Index Compression -1	Project Selection
February 14 2024	Index Compression -2	
February 19 2024	The Vector Space Model -1	Homework 1 due
February 21 2024	The Vector Space Model -2	Homework 2 issued
February 26 2024	Scoring in the Vector Space Model	
February 28 2024	Information Retrieval Evaluation	
March 4 2024	Web Search -1	
March 6 2023	Web Search -2	
March 11-17 2024	*Spring Break*	
March 18 2024	Web- Search -3	MID-TERM EXAM
March 20 2024	Web Crawling -1	
March 25 2024	Web Crawling -2	
March 27 2024	Link Analysis -1	Homework 2 due
April 1 2024	Link Analysis -2	Homework 3 issued
April 3 2024	Flat Clustering	
April 8 2024	Hierarchical Clustering -1	
April 10 2024	Hierarchical Clustering -2	
April 15 2024	Relevance Feedback -1	Homework 3 due
April 17 2024	Query Expansion-1	
April 22 2024	Query Expansion-2	
April 24 2024	Project Presentations -1	
April 29 2024	Project Presentations -2	FINAL EXAM
May 1 2024	Project Presentations -3	

Course Policies

Grading (credit) Criteria	Homeworks: 30% Project: 25% Mid-Term Exam: 20% Final Exam: 20% COMET behavior commitment: 5%
Make-up Exams	There will be **no** make-up exams
Extra Credit	There will be **no** extra-credit assignments
Late Work	If the homework is turned in after the deadline, the grade for the homework shall be reduced by 20% for the first 24 hours, 50% for the next 24 hours and shall not be accepted after 48 hours. **NO** late submissions will be allowed for the exams!!!
Special Assignments	Students will have special assignment for projects
Class Attendance	Following the CS Department attendance policy: Absence in three consecutive lectures will result in the course grade being lowered by one

	letter. Absence in four consecutive lectures will automatically result in a failing grade (F) in the course. Only for documented medical reasons presented to the instructor absences will be permitted.
Comet Creed	*This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:* *"As a Comet, I pledge honesty, integrity, and service in all that I do."* This also means that grades are **earned**, not negotiated. For example, requesting a grade the student feels they "need" instead of the earned grade is a violation of the Comet Creed, and will lead to a reduction of the grade. Email abuse will be punished with a reduction of 1point from final grade per incidence. Discussions with the instructor or TA will take place **ONLY** during office hours, on eLearning.

These descriptions and timelines are subject to change at the discretion of the Professor.