

Compiled & Shared By: ✨ Hassan Sardar Naveed

👤 "Please remember me and my family in your prayers." 🌸

📖 Bachelor of Science in Computer Science

🎓 University of the People

📞 Contact

✉ Email: hassannaveed@my.uopeople.edu

📱 WhatsApp: <https://wa.me/923022129236>

🌐 GitHub: <https://github.com/HassanSardarNaveed>

Join UoPeople Students Community

https://chat.whatsapp.com/Kv0vstTEaMUHUrlNH9SOT4?mode=ac_t

Here you will find the syllabi and primary textbooks for all UoPeople courses. The Disclaimer for Use of the Repository can be found [here](#).

Computer Science

CS 3308 Information Retrieval



CS 3308: Information Retrieval

Syllabus

Prerequisites: CS 3303: Data Structures

Course Description: This course introduces the fundamental concepts of information retrieval (IR) systems. Information Retrieval systems are systems that provide the ability to search for and find specific data or information within a collection. Although there are many implementations of IR technology, web search engines such as Google.com, Altavista.com, bing.com, and ask.com are all examples of IR technology applied to content in the world wide web.

Required Textbook and Materials: UoPeople courses use open educational resources (OER) and other materials specifically donated to the University with free permissions for educational use. Therefore, students are not required to purchase any textbooks or sign up for any websites that have a cost associated with them. The main required textbook for this course is listed below and can be readily accessed using the provided link. There may be additional required/recommended readings, supplemental materials, or other resources and websites necessary for lessons; these will be provided for you in the course's General Information and Forums area, and throughout the term via the weekly course Unit areas and the Learning Guides.

- Manning, C.D., Raghaven, P., & Schütze, H. (2009). *An Introduction to Information Retrieval* (Online ed.). Cambridge, MA: Cambridge University Press. Available at <http://nlp.stanford.edu/IR-book/information-retrieval-book.html>

Many of the optional video lectures (***please note that unit 6 does not have video lectures***) in this course take advantage of or have adapted the slides created for the Stanford University Information Retrieval course and which have been posted online at the following URL: <http://www.stanford.edu/class/cs276/> We thankfully acknowledge the work of Pandu Nayak and Prabhakar Raghaven which they made available from the textbook's web site. <http://nlp.stanford.edu/IR-book/information-retrieval-book.html>

Software Requirements/Installation: The information retrieval (IR) course provides learning experiences that address both the theory and practice of information retrieval systems. As part of this course, students will learn fundamental and critical theories of information retrieval and put those theories into practice by constructing elements of an information retrieval system. Students will be required to construct a parser, indexer, and search interface using the Python language.

For these programming assignments, you must download and install the appropriate Python interpreter for your computer and operating system. Versions of the software are available for Windows (XP, Vista, Windows7), Linux distributions, and Mac OS. Most popular distributions of Linux will either include Python or will provide an installation option for it in the software management utility.

You can find available downloads for Python v2.7.x at the following URL: <http://www.python.org/download/>

Installation is relatively straightforward. Follow the prompts when installing. Further information is available in the documentation section located here: <http://docs.python.org/>

?

Instructions to install and configure Python can be found in the Python setup and usage section of this page.

Learning Objectives and Outcomes:

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

1. Explain fundamental concepts and theories of information retrieval.
2. Differentiate between and apply index compression and search effectiveness techniques.
3. Compute weights and scores of documents within an IR system.
4. Determine the effectiveness of an information retrieval system using a known document corpus.
5. Construct a complete information retrieval system.
6. Construct a web search system by integrating indexer, search engine, and web crawler (spider) components.

Course Schedule and Topics: This course will cover the following topics in eight learning sessions, with one Unit per week. The Final Exam will take place during Week/Unit 9 (UoPeople time).

Week 1: Unit 1 - Introduction to IR, Boolean Retrieval, and Terms and Postings (Chapters 1 & 2)

Week 2: Unit 2 - Dictionaries and Index Construction (Chapters 3 & 4)

Week 3: Unit 3 - Index Compression (Chapter 5)

Week 4: Unit 4 - Scoring, Term Weighting, and the Vector Space Model (Chapter 6)

Week 5: Unit 5 - Scoring and Ranking in a Complete Search System (Chapter 7)

Week 6: Unit 6 - Evaluation in Information Retrieval (Chapter 8)

Week 7: Unit 7 - Introduction to Web Search (Chapter 19)

Week 8: Unit 8 - Web Crawling (Chapter 20 & 21)

Week 9: Unit 9 - Course Review and Final Exam

Learning Guide: The following is an outline of how this course will be conducted, with suggested best practices for students.

Unit 1: Introduction to IR, Boolean Retrieval, and Terms and Postings (Chapters 1 & 2)

- Read the Learning Guide and Reading Assignments
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Make entries to the Learning Journal
- Take the Self-Quiz

Unit 2: Dictionaries and Index Construction (Chapters 3 & 4)

- Read the Learning Guide and Reading Assignments
- Complete and submit the programming Assignment
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Make entries to the Learning Journal
- Take the Self-Quiz

Unit 3: Index Compression (Chapter 5)

- Peer assess Unit 2 Programming Assignment
- Read the Learning Guide and Reading Assignments
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Make entries to the Learning Journal
- Take the Self-Quiz

Unit 4: Scoring, Term Weighting, and the Vector Space Model (Chapter 6)

- Read the Learning Guide and Reading Assignments

- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Complete and submit the Programming Assignment
- Make entries to the Learning Journal
- Take the Self-Quiz

Unit 5: Scoring and Ranking in a Complete Search System (Chapter 7)

- Peer assess Unit 4 Programming Assignment
- Read the Learning Guide and Reading Assignments
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Complete and submit the Programming Assignment
- Make entries to the Learning Journal
- Take the Self-Quiz
- Take the Graded Quiz

Unit 6: Evaluation in Information Retrieval (Chapter 8)

- Peer assess Unit 5 Programming Assignment
- Read the Learning Guide and Reading Assignments
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Make entries to the Learning Journal
- Take the Self-Quiz

Unit 7: Introduction to Web Search (Chapter 19)

- Read the Learning Guide and Reading Assignments
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Complete and submit the Programming Assignment
- Make entries to the Learning Journal
- Take the Self-Quiz

Unit 8: Web Crawling (Chapter 20 & 21)

- Peer assess Unit 7 Programming Assignment
- Read the Learning Guide and the Reading Assignment
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Complete the Programming Assignment (non-graded)
- Make entries to the Learning Journal
- Read the Unit 9 Learning Guide carefully for instructions on the Final Exam
- Take the Review Quiz

Unit 9: Course Review and Final Exam

- Read the Learning Guide and take the Review Quiz, if you haven't already done so
- Prepare for, take, and submit the Final Exam
- The Final Exam will take place during Week/Unit 9 (UoPeople time); exact dates, times, and other details will be provided accordingly by your instructor

Course Requirements:

Programming Assignments & Assessment Forms

By the end of this course, you will have built a single cumulative Programming project. This project will be due in four parts throughout the course as Programming Assignments and Assessment Forms. You are required to submit your assignments by the indicated deadlines and, in addition, to peer assess three (3) of your classmates' assignments according to the instructions found in the Assessment Form, which is provided to you during the following week. During this peer assessment period, you are expected to provide details in the feedback section of the Assessment Form, indicating why you awarded the grade that you did to your peer. Failure to submit Programming Assignments and/or Assessment Forms may result in failure of the course.

The culminating project you will be working towards with these assignments is developing an information retrieval system. Your information retrieval system will have four main components a parser, an indexer, a search engine, and a web crawler.

Indexer Part 1 – In this assignment, you will construct a process that will generate an inverted index. You will tokenize the contents of a corpus that contains over 11,000 Reuters articles and store your index to disk in a format that will enable fast search retrieval. You will have 2 weeks to complete this assignment.

Indexer Part 2 – In this assignment, you will enhance your indexer process by incorporating a stop words functionality, term editing, a porter stemmer, and you will calculate and store tf-idf, d weighting for each unique combination of term and document in the index. You will have 1 week to complete this assignment.

Search Engine – In this assignment you will develop a basic search engine that will enable the user of the process to enter one or more search terms and the process will extract all documents from the index that contain all of the search terms, calculate the cosine similarity between each document and the query and 20 documents with the highest cosine similarity. You will have 1 week to complete this assignment.

Web Crawler – In this assignment, you will enhance a basic web crawler which you will point at a web URL and your web crawler process must tokenize each web page and populate the inverted index.

You will be required to use the functionality of the indexer that you created in the first two assignments and integrate it into a basic web crawler. Your search engine must be able to search the inverted index created by your web crawler. You will have 1 week to complete this assignment.

Discussion Assignments & Response Posts/Ratings

Some units in this course require that you complete a Discussion Assignment. You are required to develop and post a substantive response to the Discussion Assignment in the Discussion Forum. A substantive response is one that fully answers the question that has been posed by the instructor. In addition, you must extend the discussion by responding to at least three (3) of your peers' postings in the Discussion Forum and by rating their posts. Instructions for proper posting and rating are provided inside the Discussion Forum for each week. Discussion Forums are only active for each current and relevant learning week, so it is not possible to contribute to the forum once the learning week has come to an end. Failure to participate in the Discussion Assignment by posting in the Discussion Forum and responding to peers as required may result in failure of the course.

Learning Journal

Your instructor may choose to assign specific topics and/or relevant questions as a weekly Learning Journal entry for you to complete, but you are still encouraged to also use it to document your activities, record questions/problems you may have encountered, reflect on the learning process, and draft answers for other course assignments. The Learning Journal must be updated on a weekly basis because its entries will be assessed by your instructor directly as a part of your final grade. The Learning Journal will only be seen by your instructor.

Quizzes

This course will contain three types of quizzes – the Self-Quiz, the Graded Quiz, and the Review Quiz. These quizzes may contain multiple choice, true/false, or short answer questions. The results of the Self-Quiz will not count towards your final grade. However, it is highly recommended that you complete the Self-Quiz to ensure that you have adequately understood the course materials. Along with the Reading Assignments, the results of the Self-Quiz should be used as part of an iterative learning process, to thoroughly cover and test your understanding of course material. You should use the results of your Self-Quiz as a guide to go back and review relevant sections of the Reading Assignments. Likewise, the Review Quiz will not count towards your final grade, but should also be used to assist you in a comprehensive review and full understanding of all course material, in preparation for your Final Exam. Lastly, the results of the Graded Quiz will count towards your final grade.

Final Exam

The Final Exam will take place during the Thursday and Sunday of Week/Unit 9, following the completion of eight units of work. The format of the Final Exam is similar to that of the quizzes and may contain a combination of different question types. You will have one attempt to take the exam, and it will be graded electronically. Specific instructions on how to prepare for and take the Final Exam will be provided during Week 8 (located inside the Unit 9 Learning Guide). Final Exams must be taken without the use of course learning materials (both those inside and outside the course). If particular materials are allowed for use during the exam, these will be noted in the exam's instructions.

Course Forum

The Course Forum is the place to raise issues and questions relating to the course. It is regularly monitored by the instructors and is a good place to meet fellow students taking the same course. While it is not required to participate in the Course Forum, it is highly recommended.

Course Policies:

Grading Components and Weights

Each graded component of the course will contribute some percentage to the final grading scale, as indicated here:

Learning Journals	10%
Discussion Assignments	10%
Programming Assignments	30%
Graded Quiz	20%
Final Exam	30%
TOTAL	100%

Grading Scale

This course will follow the standard 100-point grading scale defined by the University of the People, as indicated here:

Letter Grade	Grade Scale	Grade Points
A+	98-100	4.00
A	93-97	4.00
A-	90-92	3.67
B+	88-89	3.33
B	83-87	3.00
B-	80-82	2.67
C+	78-79	2.33
C	73-77	2.00
C-	70-72	1.67
D+	68-69	1.33
D	63-67	1.00
D-	60-62	0.67
F	Under 60	0.00

Grade Appeal

If you believe that the final grade you received for a course is erroneous, unjust, or unfair, please contact your course instructor. This must be done within seven days of the posted final grade. For more information on this topic, please review the Grade Appeal Procedure in the University Catalog.

Participation

Non-participation is characterized by lack of any assignment submissions, inadequate contributions to the Discussion Forums, and/or lack of peer feedback to Discussion/Written Assignments. Also, please note the following important points about course participation:

- Assignments must be submitted on or before the specified deadline. A course timeline is provided in the course schedule, and the instructor will specify deadlines for each assignment.
- Any student showing non-participation for two weeks (consecutive or non-consecutive) is likely to automatically fail the course.
- Occasionally there may be a legitimate reason for submitting an assignment late. Most of the time, late assignments will not be accepted and there will be no make-up assignments.
- All students are obligated to inform their instructor in advance of any known absences which may result in their non-participation.

Academic Honesty and Integrity

When you submit any work that requires research and writing, it is essential to cite and reference all source material. Failure to properly acknowledge your sources is known as "plagiarism" – which is effectively passing off an individual's words or ideas as your own. University of the People adheres to a strict policy of academic honesty and integrity. Failure to comply with these guidelines may result in sanctions by the University, including dismissal from the University or course failure. For more information on this topic, please review the Academic Integrity Policy in the University Catalog.

Unless otherwise stated, any materials cited in this course should be referenced using the style guidelines established by the American Psychological Association (APA). The APA format is widely used in colleges and universities across the world and is one of several style and citation formats required for publication in professional and academic journals. Purdue University's Online Writing Lab (OWL) is a free

website that provides excellent information and resources for understanding and using the APA format and style. The OWL website can be accessed here: https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_style_introduction.html

Code of Conduct

University of the People expects that students conduct themselves in a respectful, collaborative, and honest manner at all times. Harassment, threatening behavior, or deliberate embarrassment of others will not be permitted. Any conduct that interferes with the quality of the educational experience is not allowed and may result in disciplinary action, such as course failure, probation, suspension, or dismissal. For more information on this topic, please review the Code of Conduct Policy in the University Catalog.