Course Syllabus

Information Retrieval Systems

INST734
CMSC828Z
Fall 2019

Learning Outcomes

Principles of organizing and providing access to information using automated information storage and retrieval systems. Retrieval system models, index language selection, data structures, user interfaces and evaluation for text and multimedia applications.

Students should be able to understand and explain the basic concepts of Information Retrieval, apply the Information Retrieval methods on real data and estimate the performance of the system.

After successfully completing this course students will be able to:

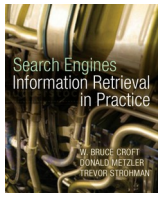
- Predict how different types of user queries perform.
- Identify a suitable method which can improve retrieval quality for particular queries.
- Analyze evaluation approaches and select the most suitable approach for a particular retrieval task/scenario.

Resources

Course website: **inst734.umiacs.io**

No textbook is required. All course material will be available freely online for the UMD students.

Recommended Resources



Donald Metzler, Trevor Strohman, and W. Bruce Croft: Search Engines: **Information Retrieval in Practice**, 2010. https://ciir.cs.umass.edu/irbook/

Dr. Petra Galuscakova
petra@umiacs.umd.edu

Class Meets
Thursdays
6:00–8:45pm
TYD 1118

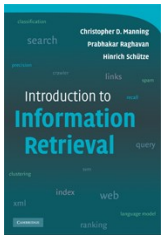
Office Hours
IRB 4138
Thursdays
3:00–4:00pm
and by appointment

Prerequisites
Students must be enrolled in CS masters or Ph.D. program or iSchool Ph.D. program. MIM students must have completed (or waived) INFM 603. MLS students must have completed (or waived) LBSC 671. Students enrolled in other programs (including iSchool Ph.D. students) should consult with the instructor to determine whether their academic preparation is appropriate for this course.

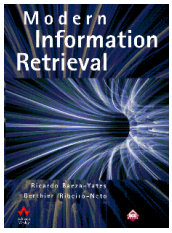
Course Syllabus

Information Retrieval Systems

INST734
CMSC828Z
Fall 2019



Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze: **Introduction to Information Retrieval**, 2008
https://nlp.stanford.edu/IR-book/information-retrieval-book.html



Ricardo Baeza-Yates and Berthier Ribeiro-Neto: **Modern Information Retrieval**, 1999.
http://people.ischool.berkeley.edu/~hearst/irbook/

Course Communication
The course material will be available on the course website. ELMS will be used for submitting homework assignments. All assessment scores will be posted on the course ELMS page.

Approach

The course is organized as a set of 15 modules. For each module, the syllabus shows the start date and the assignment which needs to be submitted on Wednesday 11:59pm EDT/EST before this module (e.g. the first assignment needs to be finished on 09/04). Readings and videos may be assigned for each module and need to be completed in advance of class. Reading assignments may vary for students with different background (CS/iSchool), but the amount of reading will be generally the same for both reading variants.

Two possible sets of assignments are available and they vary based on the student background. In the CS assignments, students will build up the ranked information retrieval system from scratch. The IS assignments are aimed at understanding the retrieval scenarios and methods and they expect the student to be able to apply the methods on real world problems.
First two assignments are common for all students and each of them should take a week to finish. Apart from this, there are 5 CS assignments, each of which should take two weeks to finish / 7 IS assignments, each of which should take a week to finish. In modules 6, 10 and 13 students are expected to review their work on the project.

Assignments

CS students are obliged to work on the CS assignments. Other students may pick up the assignment type on which they prefer to work, based on their experience. This preference needs be announced to the lecturer before the first assignment. If you are not sure which assignments set to select, please consult your preference with the lecturer.

Course Syllabus

Information Retrieval Systems

INST734
CMSC828Z
Fall 2019

The CS assignments will be more programming oriented. Preferred programming language for the assignments is Python, if you wish to use different programming language, contact the lecturer before the assignment. The assignments will be oriented on the Unix-like environment, but it should be possible to finish them without Unix. As the programming assignments are generally expected to be more time consuming than the IS assignments, students working on the IS assignments set are also obliged to send a reflection paper for each module. Reflection papers need to be submitted on Wednesday 11:59pm EDT/EST before the following module.

Projects

Student-designed projects can be either research-oriented (leading to an academic paper) or operationally oriented (leading to a complete working system). Students can either select to work on one of the projects proposed by the instructor or their own project. Students who elect to design their own project should meet with the instructor no later than September 26 to discuss their ideas. All students will be required to prepare a complete project plan with milestones no later than October 3rd.

Preliminary Outline

<table>
<tr><th></th><th>Date</th><th>Module</th><th>Assignment CS</th><th>Assignment IS</th><th>Project</th></tr>
<tr><td>1</td><td>08/29</td><td>Structure of IR systems</td><td colspan="2"></td><td></td></tr>
<tr><td>2</td><td>09/05</td><td>Boolean retrieval, inverted index</td><td colspan="2">Problem set</td><td></td></tr>
<tr><td>3</td><td>09/12</td><td>Ranked retrieval, probabilistic retrieval and language models</td><td colspan="2">Apply existing IR system on a collection.</td><td></td></tr>
<tr><td>4</td><td>09/19</td><td>Evaluation and benchmarks</td><td rowspan="2">Implement selected text processing method</td><td>Evaluate different matching methods</td><td></td></tr>
<tr><td>5</td><td>09/26</td><td>Text processing</td><td>Compare the effect of text processing approaches</td><td></td></tr>
<tr><td>6</td><td>10/03</td><td>Searching in meaning</td><td rowspan="2">Implement vector space retrieval</td><td></td><td>Project selection</td></tr>
<tr><td>7</td><td>10/10</td><td>Filtering and recommendation</td><td>Analyze the effect of query expansion methods</td><td></td></tr>
</table>

8	10/17	Learning to rank	Implement query expansion method and measure its impact	Explore query performance prediction methods	
9	10/24	Interaction and web search		Compare visualization approaches	
10	10/31	Cross-language search	Implement dictionary-based CLIR system		Project update
11	11/07	Speech and music		Examine automatic transcripts/foreign queries in search	
12	11/14	Images and video	Implement selected results visualization method	Survey multimedia retrieval systems	
13	11/21	Final exam			Project update
14	11/28	No class (Thanksgiving Day)			
15	12/05	Capstone project presentations			Final project submission

Grading

The assignments and the reflection papers submitted up to 24 hours after the deadline, will receive 50% of the grade. No credit is given after 24 hours. One lowest score from the assignment and two lowest scores from the reflection papers are dropped.

A term project needs to be completed by the end of the semester. Students may work individually or in groups. In the case of the group project, each member must clearly indicate the role and contributions.

There will be no grades for class participation, however class participation is highly encouraged. Occasionally, there might be extra assignments presented at the classes for which some extra credit may be offered.

The final examination will be the only exam given during the semester. The examination will be 2 hours, in class, at the same time that the class usually meets on the date indicated in the syllabus. Reference sheet notes are allowed to be used during the exam.

Course Syllabus

Information Retrieval Systems

INST734
CMSC828Z
Fall 2019

Component	Component CS	Component IS
Commentaries/Reflection papers		10%
Assignments	40%	30%
Term Project	30%	30%
Final Exam	30%	30%

Final Grade Cutoffs										
+	97.00%	+	87.00%	+	77.00%	+	67.00%			
A	93.00%	B	83.00%	C	73.00%	D	63.00%	F	<60.0%	
-	90.00%	-	80.00%	-	70.00%	-	60.00%			

Academic Integrity

Students are encouraged to work together to learn the materials and to learn the how to do the assignments. However, all of the material that is turned in for grading must be produced by the individual or team that is submitting the material.

The University of Maryland, College Park has a nationally recognized *Code of Academic Integrity*. This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the *Code of Academic Integrity* or the Student Honor Council, please visit http://www.shc.umd.edu.