

# CSC 395: Information Retrieval

FALL 2019

Department of Computer Science  
Grinnell College

## INSTRUCTOR

**Fahmida Hamid**

**Assistant Professor**

Office: Noyce 2611

Phone: (641) 269-3271

Office Hours: Friday, 4:00 PM ~ 6:00 PM

Email: [hamidfah@grinnell.edu](mailto:hamidfah@grinnell.edu)

## CLASS

- CSCI 395: TuTh (2:00 PM - 3:50 PM) Room #Noyce-3819

## TEXTBOOK

**Introduction to Information Retrieval**, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Cambridge University Press. 2008. ISBN: 0521865719.

## PREREQUISITE

CSC 161, STA 209, and one of MAT 218, CSC 208, or MAT 208

## COURSE OVERVIEW

Information retrieval (IR) is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information (Salton, 1968). The goal of an IR system is to retrieve relevant information based on the task (user needs) with minimum information overload and retrieval time. In this course, we plan to address the basic concepts and modeling tools in IR and study some recent research trends in the field: e.g., web search engine, link analysis, query expansion, question-answering, keyphrase extraction. Work will include programming assignments, written exercises, exams, and a group research project.

## COURSE OUTCOME

After successfully completing the course, students will have a strong understanding of the basic concepts, principles, and applications of IR. They will have a hands-on experience with several existing tools through the programming assignments and they will be solving one critical problem on their own in the project. Besides, they will advance their analytical skills through written assignments, project-report, peer reviews, and hour-exams.

## COMMUNICATION

**Course website** You will find the course website here: <http://www.cs.grinnell.edu/~hamidfah/courses/csc395fall2019/>. The website will give you the detailed schedule of the class. I may need to update it from time to time.

**Pioneerweb** We will use the pioneer-web system for submitting homework and keeping track of your grades. A course account (CSC-395) is created and you all are added to it. You will find the course materials (syllabus, lecture notes, homework problems, etc.) posted in this account.

**Email** You are encouraged to email me if you need to set up a meeting beyond the dedicated office hours. You are also welcome to ask any question through email. I will try to get back to you within next 24 hours. If your question is of interest to others in the class, we can discuss it in the class or I can send a message to all using the pioneerweb course account.

## ACTIVITIES

We plan for the following activities:

**Class-sessions:** In-class activities of the first five weeks will mostly be lectures highlighting and demonstrating important concepts from textbook readings. When appropriate, external readings may be used. Starting from week seven, when we cover different research tracks in IR through literature reviews, the class will mostly be discussions where students play an active role after my initial lecture.

**Programming assignments:** Programming exercises (group work) of varying length/difficulty will be assigned for the first five weeks of the term. Roughly two such exercises will be given. Some of variability is due to particular interests and abilities of the participants. Several topics covered in the class are appropriate to be programming assignments: for example, implementing one of the clustering algorithms for a certain dataset, building an uncompressed index and retrieval, query expansion, email filtering using some of the discussed techniques, and so on.

**Written exercises:** Students will complete one short written exercise (individual work) every alternating week to reinforce the theoretical concepts of the course. I do not plan to give you any written exercises on the seventh and the last week. The written exercises from the eighth to the thirteenth week will mostly be summarizing the assigned articles for each week or answering some analytical questions in the context of those articles.

**Exams:** Students will take one midterm and one final exam. The final exam will be comprehensive. We will use open-book policy for both of the exams.

**Research project:** We will use the class meeting times of the fourteenth week for group presentations. With a cap of 20 students, I expect there to be roughly 10 groups, leaving approximately 20 minutes for each presentation. This format is similar to that of an IR research conference. In addition to their presentations, students will submit a paper describing their work. The paper will follow a standard IR conference template and should cover related literature review, empirical studies, experimental data, significant discoveries, and discussion/future directions as core segments. Students will be submitting the paper and well-documented code by the end of week thirteen.

**Peer review:** Each student will review project papers (written by other groups) and write constructive and analytical review. We will follow a common review template. All the reviews will be sent to the corresponding groups. This activity will motivate them to writing better papers.

## EXAM

There will be one midterm exam. The two-hour mid-term exam will be on **Tuesday 08 October**.

Missing exams because of illness will require an excuse from a medical practitioner. Make-up exams for excuses other than illness will be given only in extraordinary circumstances and only at the discretion of the instructor. If you expect that you will need a make-up exam, contact your instructor at least one week in advance.

## RESEARCH PROJECT

I will expect this as a group activity. I will allow two people to form a group unless someone really wants to work alone. In case of odd class-size, I will allow only one group to have three members.

As the research project carries the highest weight, you need to carefully define the depth and scope of your work. We will consider three research tracks: **Summarization, Keyphrase extraction, and Recommender Systems**. Based on your preliminary studies and with the help and guidance of your instructor, you will define your project. It can be a **survey/analysis**; in which case, you will implement at least three existing models for solving the same problem and analyze their performances on different criteria. It can be an **implementation-oriented project**; in which case, you propose a new model, implement it, and analyze its performance. In both cases, your clarity of thinking, efficiency of implementation, depth of your analysis, structure and readability of project report, and presentations will determine your total score for the project. A detailed rubric for the research project will be provided later in class.

## ATTENDANCE

Attendance at all lectures is **mandatory**. If you miss a class for some personal reason/sickness, please send me a quick note. Also, please let me know in advance of planned absences. Note that regularly missing class is one of the potential special circumstances that could lead to a discretionary reduction in grade.

## HOMEWORK SUBMISSION POLICY

All assigned work is due on the date and at the time specified. For homework, we will use a late day policy to help you manage your work load throughout the semester. This policy works as follows:

- You have five late days to use in the semester.
- You may use one late day to turn in one homework up to 24 hours after the due date, no questions asked.
- You do not need to tell us that you are using a late day.
- Late days are automatically noted and tracked by the instructor.
- You may use up to two late days on a given homework.
- If you are working in a group, each member must use a late day in order to extend the deadline by 24 hours.

Beyond late days, homework may not be turned in after the due date. This is a strict policy in order to help us get your feedback to you in a timely manner. Only the most exceptional of circumstances discussed well in advance with the instructor (as much as the situation allows) will be entertained. In case of medical emergencies, written medical excuses from a medical practitioner (or a responsible person from student affairs) may be grounds for granting an extension.

## GRADING POLICY

My goal in the course is for everyone to be proficient in the big concepts outlined in the Overview. While this is a lot of content, I firmly believe that everyone is capable of mastering this material — earning an A in the process-with enough time, dedication, and proper study skills. My initial plan is to start with the following breakdown:

Activity	Weight
Participation/Attendance:	10%
Written Exercises:	10%
Programming Assignments:	20%
Research Project:	40%
Peer Review:	10%
Midterm Exam:	10%

**Percentages may be adjusted upwards or downwards at the discretion of the instructor.**

## ACCESS STATEMENT

If you have specific physical, psychiatric, or learning disabilities and require accommodations, please let me know early in the semester so that your learning needs may be appropriately met. Note that you will also need to provide documentation of your disability to the Dean for Student Academic Support and Advising, Autumn Wilke, located on the 3rd floor of the Rosenfield Center (x3702).

## ACADEMIC RESPONSIBILITY

Students are expected to read and abide by the principles clearly explained in the [Student Handbook](#). When in doubt, talk to your professor.

## ACADEMIC HONESTY

All academic work at Grinnell College must follow standard academic practice regarding quotation, paraphrase, and citation. Grinnell's Student Handbook provides basic guidelines.

**Homework:** Homework exercises should be entirely your own individual work, not done in collaboration with other students in the class, and not quoted or paraphrased, in whole or in part, from external sources. [Note: Although the Web can be useful for reference, much material on the Web is of poor quality. You are responsible for the quality of what you turn in, regardless of the source of the material.]

**Programming Assignments:** Programming assignments should be a group work. Each group should be of size 2. In case of odd class-size, I will allow one group of size 3. Students are allowed to use any programming language. I personally prefer Python.

**Group Presentation:** If a group of two or three people work together on a presentation, the group should turn in one written report, and all names in the group must appear at the top of the first page. A sample template will be provided later for the written report.

**Plagiarism:** If I encounter clear indications of plagiarism or academic dishonesty, the Committee on Academic Standing will deal with them. I will impose penalties for academic dishonesty only as directed by the committee.

## TENTATIVE SCHEDULE

Week	Dates	Topic	Reading	Assignments
Week 00	Aug 29	Introduction Boolean Retrieval Model	Ch 1	Wri Exer 01
Week 01	Sept 03 Sept 05	Search Engine Architecture Web Crawler & Basic Text Processing Technique	Ch 19 Ch 20	Prog Asgn 01
Week 02	Sept 10 Sept 12	Inverted Index & Query Processing Cont ...	Ch 2, 4, 5	Wri Exer 02
Week 03	Sept 17 Sept 19	Search Result Interface Boolean & Vector Space model	Ch 1, 6	Prog Asgn 02
Week 04	Sept 24 Sept 26	Probabilistic Information Retrieval Language Models	Ch 11 Ch 12	Wri Exer 03
Week 05	Oct 01 Oct 03	Evaluation in IR Discussion on possible Research Tracks	Ch 08	
Week 06	Oct 08 Oct 10	<b>Midterm Exam</b> Web-Crawling, Link Analysis, Personalization	<b>Research Proposal (Draft)</b> (Due)	
Week 07	Oct 15	Special Track: Recommender Systems	<b>Instructors Feedback</b> (research proposal)	
	Oct 17	Special Track: Keyphrase Extraction		
<b>FALL BREAK</b>				
Week 08	Oct 29	Special Track: Summarization	<b>Research Proposal (Final)</b> (Due)	
	Oct 31	<b>Proposal Presentation</b>		
Week 09	Nov 05 Nov 07	Lab Day Special Track: Learning to Rank + Lab		
Week 10	Nov 12 Nov 14	Lab Day Mid-project progress presentation + Lab		
Week 11	Nov 19 Nov 21	Lab Day Mid-project progress presentation+ Lab + Review 01		
Week 12	Nov 26 Nov 28	Lab Day THANKSGIVING RECESS (No Class)		
Week 13	Dec 03 Dec 05	Lab Day Mid-project progress presentation+ Lab + Review 02		

Week	Dates	Topic	Reading	Assignments
Week 14	Dec 10 Dec 12	Lab Day <b>Project presentation</b>		

MORE ...

**Syllabus** This syllabus may be modified as the course progresses. Notice of such changes will be announced in class or through course website.

**Have a Great Semester!**