

Information Retrieval

Fall 2018

Course Information & Syllabus

Instructor: Cornelia Caragea

Office: 851 S. Morgan (M/C 152) Room 1120 SEO Chicago, IL 60607-7053

Email: cornelia@uic.edu

WWW: <https://www.cs.uic.edu/~cornelia/>

Lectures: Tue / Thr 12:30pm - 1:45pm, Room TH 204

Office Hours: Cornelia: Thr 11:00am - 12:00pm or by appointment

Exam 1: October 9th

Exam 2: November 29th

Course Objectives: Information Retrieval (IR) refers to the processing, indexing and querying of unstructured or loosely structured information. The course objectives are to understand information retrieval algorithms and identify challenging problems on the Web. The course will cover both traditional and newly developed algorithms in information retrieval and Web search and their Web applications. Examples of topics include: indexing, processing, and querying textual data; basic retrieval models (e.g., the vector space model), “intelligent” IR systems and IR system implementations; relevance feedback and query expansion; Web crawling and search; link analysis; text classification and mining; sentiment analysis on the Web.

Course Work and Evaluation: There will be two exams for the course. Students will be evaluated based on the exams, homework assignments, and a class project. Students are encouraged to attend every lecture and to participate in class discussion.

Assignments are due by 11:59pm on the due date. Assignments may be turned in up to 3 days late, with a penalty of 10% for each day late. No credit will be given after 3 days. There will be no final exam for this class. The final is replaced by the project. The grading criterion is shown below:

Section	Weight
Homework	25%
Exams	40%
Project	30%
Class Participation	5%

Collaboration policies:

- You are encouraged to discuss the course material, concepts, and assignments, but you must write your answers independently.
- For each assignment, you are required to list students with whom you have discussed the assignment.
- Your submission should reflect your own knowledge and you should be able to reproduce the material you turn in at any time.
- Sharing answers will not be tolerated.
- Plagiarism will not be tolerated either.
- Appropriate citations for any external sources used in your work are mandatory. Never use sentences or phrases taken directly from a paper you are reviewing.

Prerequisites: Basic knowledge on probability and statistics, data structures and algorithms. Background in information retrieval is not required.

Targeted audience: *Graduate students from Computer Science and related areas.*

Attendance: Attendance is essential and thus is expected.

Required textbooks:

- Introduction to Information Retrieval by Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze. Cambridge University Press, 2008.
Online version available at: <http://nlp.stanford.edu/IR-book/>.

Other Recommended textbooks:

- Readings in Information Retrieval by K. Sparck Jones and P. Willett Morgan Kaufmann, 1997.
- Modern Information Retrieval by Ricardo Baeza-Yates and Berthier Ribeiro-Neto Addison-Wesley, 1999.

Topics: The tentative topics are as follows:

The term vocabulary and postings lists
Index construction
Scoring, term weighting and the vector space model
Computing scores in a complete search system
Evaluation in information retrieval
Relevance feedback and query expansion
Language models for information retrieval
Web search basics
Web crawling and indexes
Link analysis
Text classification and mining
Naive Bayes
Vector space classification
Sentiment analysis on the Web
Recommender Systems
Information Extraction - Keyphrase Extraction
Question Answering

Other Policies: No make-up exams and no incomplete, unless there is a very serious reason.

Students with Disabilities: Any student with a disability who needs an accommodation or other assistance in this course should make an appointment to speak with one of the instructors as soon as possible.