Winter 2018

CSI4107: Information Retrieval and the Internet

Instructor: Mohamad Hoda
E-mail: mhoda053@uottawa.ca

Office Hours: TBA or by email appointment, in SITE 5077.

Overview

Basic principles of Information Retrieval. Indexing methods. Query processing. Linguistic aspects of Information Retrieval. Agents and artificial intelligence approaches to Information Retrieval. Relation of Information Retrieval to the World Wide Web. Search engines. Servers and clients. Browser and server side programming for Information Retrieval.

Evaluation Students will be evaluated on:

- Two written and programming assignments / group project (15% each). The programming language will be Java or any other.
  Note: The assignments will be submitted electronically through Brightspace. No late assignments are accepted.
- Midterm exam (15%)
- One in-class Presentation (15%)
- Final exam (40%)
- Bonus points for class participation

Timetable (no late assignments are considered)

- Assignment 1, due Feb 17 at 10:00pm
- In-class presentation.
- Midterm (Wed, Feb 28, in class)
- Assignment 2, due Mar 24 at 10:00pm.
- Final exam (during exam period)

Recommended Textbook

Introduction to Information Retrieval, by Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze, Cambridge University Press, 2008 (online version available)

Other books:
Information Retrieval, by D. Grossman and O. Frieder, Springer, 2004 (second edition).
Another online book Information Retrieval, by C. J. van Rijsbergen (1979)
Modern Information Retrieval, by Ricardo Baeza-Yates and Berthier Ribeiro-Neto, 1999.
Companion website to this book.

Syllabus (The lecture slides will be in pdf format, you can read them with Acrobat Reader)

Credit: some of the lecture notes are initially designed by prof. Ray Mooney, University of Texas Austin

Pre-Requisites (CSI3103 or ELG3300), (CSI3125 or CSI2115 or SEG2101) or permission from the instructor.
This is a ***tentative study plan***. It is subject to minor changes during the semester.

Week	Lecture	Readings
1 Jan. 10	a- Course Introduction. b- Boolean retrieval	IIR ch. 1
2	a- Term vocabulary & postings lists b- Dictionaries & tolerant retrieval	IIR ch. 2, ch. 3
3	a- Index construction b- Index compression	IIR ch. 4, ch 5
4	Scores, weights, vector space	IIR ch. 6
5	Computing scores	IIR ch. 7
6	Evaluation & result summaries	IIR ch. 8
7	Rel. feedback, query expansion	IIR ch. 9
8	**Reading period**	
9	**Midterm Exam** Text classification, Naive Baye	IIR ch. 13
10	a- Language models for IR b- Vector space classification	IIR ch. 12, ch. 14
11	a-Support vector machines b-Learning to rank (LTR)	IIR ch. 15
12	Flat clustering	IIR ch. 16 Presentations
13	Linking Analysis	IIR ch. 21 Presentations
14	Final Exam Review	Presentations