BIM490: Introduction to Information Retrieval

Spring 2017

Ozgur Yilmazel
Computer Engineering
Anadolu University

Introduction

Information Retrieval (IR) is the area of Information Science that focuses on the electronic provision of information from textual, image, and sound databases to users in response to their information needs. IR possesses a long tradition of basic and applied research with emphasis on both the technical and human side of the provision of timely, accurate information. IR research is pursued in both academic and commercial organizations, with increasing interaction between the two groups. IR is an area of resurgent research interest resulting from the broadly recognized fact that all of the information now residing on the web or on intranets is of little use if it cannot be effectively retrieved in response to users' needs.

In this course, students will learn the full technical details of how IR is accomplished from both the theoretical and applied perspectives. While the size of the class mitigates against a real seminar format, active participation in class discussions and interchange amongst the students, professor, and guest lecturers will form a cornerstone of your learning experience. Therefore, you should be prepared to summarize readings and discuss knowledgeably the topics assigned for each class, as well as discuss what you learn from your assignments and team projects.

Assignments will include weekly readings in preparation for discussion in class, small exercises, and a semester long team project with intermediate reports.

Student Responsibilities

1. You are expected to attend every class. If you are unable to attend, it is your responsibility to inform the professor and submit your assignment for that week on time and complete the assigned reading.
2. Assignments are due in class on the date specified. If you have a medical or family emergency, you should inform the professor in advance and negotiate whether a late assignment will be accepted. Late assignments are subject to a penalty deduction of half a letter grade.
3. If you do not understand an assignment, it is your responsibility to seek clarification from the professor in advance of the due date. A grade on an assignment will not be changed due to a student's statement of lack of understanding of what was required for an assignment.
4. **Anadolu University's and the Engineering Faculty' policies on academic dishonesty will be followed. Students engaging in plagiarism, cheating, or other types of academic dishonesty will receive an F for the assignment.**

Grading

Midterm 1	20 %
Midterm 2	30 %
Final Exam	50 %

Texts

Introduction to Information Retrieval, by C.Manning, P. Raghavan, and H. Schutze
http://nlp.stanford.edu/IR-book/

You are responsible for assigned readings on the date listed in the syllabus. We are likely to start classes with a brief discussion of that week's readings and you need to be prepared to comment on the readings for the week.

Office Hours

While I am available to meet with you whenever requested, the reserving of specific office hours has not proven efficient for students or the professor. So, instead, just email me to schedule an appointment in my room. Contact me at ozgur@anadolu.edu.tr or x5352.

Class Topics, Readings & Assignments

13/02/2017 Introduction to Information Retrieval

- Modern Information Retrieval, Chapter 1.
- Bush, Vannevar. (1945). *As We May Think*. Atlantic Monthly. Pp.101-108. http://www.theatlantic.com/doc/194507/bush

20/02/2017 Boolean Retrieval

- Introduction to IR, Chapter 1.
- Managing Gigabytes Sec. 3.2
- Modern Information Retrieval 8.2

27/02/2017 The term vocabulary and postings lists

- Introduction to IR, Chapter 2
- Modern Information Retrieval, Section 7.2
- Managing Gigabytes 3.6 and 4.3

06/03/2017 Scoring, Term Weighting

- Introduction to IR Chapter 6.
- Exploting the similarity space http://goanna.cs.rmit.edu.au/%7Ejz/fulltext/sigirforum98.pdf

13/03/2017 First Midterm

20/03/2017 Vector Space Scoring

- Introduction to IR Chapter 7
- Managing Gigabytes 4.4-4.6

- Modern Information Retrieval 2.5, 2.7.1, 2.7.2

27/03/2017 Lab –Introduction to Lucene Project

- Lucene
  http://lucene.apache.org

Indexing in Lucene –

Experiments using Lucene

03/04/2017 IR Performance Evaluation

- Introduction to IR Chapter 8
- Modern Information Retrieval, Chapter 3.
- Buckley, C. & Harman, D. (2004). Reliable Information Access Final Workshop Report. ARDA-NRRC 2004 Workshops. http://nrrc.mitre.org/NRRC/Docs_Data/RIA_2003/ria_final.pdf

Automatic Scoring Routine

- **Notes on TREC Evaluation**
  http://ir.iit.edu/~dagr/cs529/files/project_files/trec_eval_desc.htm

10/04/2017 Query Expansion and Relevance Feedback

- Introduction to IR Chapter 9
- Modern Information Retrieval, Chapter 5.
- Salton, G.. & Buckley, C. (1990). *Improving Retrieval Performance by Relevance Feedback*. Journal of the American Society for Information Science. Vol. 41 (4), 288-297.
- Mitra, M., Singhal, A. & Buckley,.C. (1998). *Improving Automatic Query Expansion*. Proceedings of ACM SIGIR Conference. pp. 206-214 http://citeseer.ist.psu.edu/558131.html

17/04/2017 Web-based Retrieval Systems

- Introduction to IR Chapter 19
- Modern Information Retrieval, Chapter 13.
- Kleinberg, J. (1998). *Authoritative Sources in a Hyperlinked Environment*. http://www.cs.cornell.edu/home/kleinber/auth.pdf
- Brin, S. & Page, L. (1998). *Anatomy of a Large-Scale Hypertextual Web Search Engine*. http://www-db.stanford.edu/pub/papers/google.pdf

24/04/2017 Second MidTerm
01/05/2017 Holiday
08/05/2017 Make-up classes
14/05/2017-27/05/2017 Final Exam day