**COMP 440/540: Information Retrieval**

### Course Description

Information retrieval is the study of computational methods for organizing, analyzing, and searching large quantities of semi-structured data. IR methods are present in everything from web search engines to spam filtering software to news alerts to recommender systems. You will learn how these methods work, how they are implemented, and why they sometimes fail.

**Instructor**

* **Name:** Emine Yilmaz
* **Office: ENG Z12**
* **Phone: (212) 338 1737**
* **E-mail:** [eyilmaz@ku.edu.tr](mailto:eyilmaz@ku.edu.tr)
* **Office Hours:** Tue, Thu 16:00-17:00

**Course**

* **Lecture:** Tue, Thu 17:00-18:15pm
* **Room**: ENG B30

**Text**

* [Search Engines: Information Retrieval in Practice](http://www.search-engines-book.com/) by Croft, Metzler, and Strohman.
* [Introduction to Information Retrieval](http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html) by Manning, Raghavan, and Schütze.

### Grading

You will be graded on your project, homeworks, and exams. Since this course is about search engines in practice, most of your grade depends on the project.

* 20% - Homeworks.
* 40% - Midterm exam,
* 40% - Course project and presentations

**Lateness Policy:** Assignments must be turned in by email to the professor and TA by midnight of the indicated due date. Each day late will be penalized 20%.

**Academic Honesty Policy:** All work you turn in must be your own. This includes code. While you may look at the source code for freely-available retrieval engines, you may not copy it for your project. You may use third-party libraries to help with parsing, stemming, stopping; if you do, you must clearly indicate the names of the libraries and where to obtain them. Students caught submitting someone else's work (either another student's or answers found on the web) will be prosecuted according to University guidelines.