

# INLS 509: Information Retrieval

**Description:** The field of information retrieval (IR) is concerned with the analysis, organization, storage, and retrieval of unstructured and semi-structured data. In this course, we will focus on mostly text. While IR systems are often associated with Web search engines (e.g., Google), IR applications also include digital library search, patent search, search for local businesses, and expert search, to name a few. Likewise, IR techniques (the underlying technology behind IR systems) are used to solve a wide range of problems, such as organizing documents into an ontology, recommending news stories to users, detecting spam, and predicting reading difficulty. This course will provide an overview of the theory, implementation, and evaluation of IR systems and IR techniques. In particular, we will explore how search engines work, how they "interpret" human language, what different users expect from them, how they are evaluated, why they sometimes fail, and how they might be improved in the future.

**Prerequisites:** There are no prerequisites for this course.

**Expectations:** Information retrieval is the study of *computer*-based solutions to a *human* problem. Thus, the first half of the course will be system-focused, while the second half will be user-focused. During the first half, you should expect to see some math (e.g., basic probability and statistics and some linear algebra). However, we will focus on the concepts rather than the details.

Students will have an opportunity to explore their interests with an open-ended [literature review](#).

**Time & Location:** M,W 11:15am-12:30pm, Manning 208 (in-person)

**Instructor:** Jaime Arguello ([email](#), [web](#))

**Office Hours:** By appointment, Manning 10 (Garden Level)

**Required Textbook:** [Search Engines - Information Retrieval in Practice](#), W. B. Croft, D. Metzler, and T. Strohman. Cambridge University Press. 2009. [Available on-line](#).

**Additional Resources:** [Foundations of Statistical Natural Language Processing](#), C. Manning and H. Schutze. 1999.

[Introduction to Information Retrieval](#), C. Manning, P. Raghavan and H. Schutze. 2008.

**Other Readings:** Selected papers and chapters from other books will sometimes be assigned for reading. These will be available online.

**Course Policies:** [Laptops](#), [Attendance](#), [Participation](#), [Collaboration](#), [Plagiarism & Cheating](#), [Late Policy](#), [Use of Generative AI Tools](#)

**Grading:** 30% homework (10% each)  
15% midterm exam  
15% final exam  
30% [literature review](#) (5% proposal, 10% presentation, 15% paper)  
10% participation

**Grade Assignments:** Undergraduate grading scale: A+ 97-100%, A 94-96%, A- 90-93%, B+ 87-89%, B 84-86, B- 80-83%, C+ 77-79%, C 74-76%, C- 70-73%, D+ 67-69%, D 64-66%, D- 60-63%, F 0-59%

Graduate grading scale: H 95-100%, P 80-94%, L 60-79%, and F 0-59%.

All assignments, exams, and the literature review will be graded on a curve.

**Schedule:** Subject to change! The required textbook (Croft, Metzler, and Strohman) is denoted as CMS below.

Lecture	Date	Events	Topic	Reading Due
1	Mon. 8/21		<a href="#">Introduction to Information Retrieval: The Big Picture</a>	
2	Wed. 8/23		Course Overview: Roadmap and Expectations	CMS Ch. 1
3	Mon. 8/28		Introduction To Ad-hoc Retrieval I	CMS Ch. 2, CMS 7.0-7.1
4	Wed. 8/30		Introduction To Ad-hoc Retrieval II	

5	Mon. 9/4	Labor Day (No Class)		
6	Wed. 9/6	HW1 Out	Indexing and Query Processing	CMS Ch. 5.0-5.3
7	Mon. 9/11		Statistical Properties of Text I	CMS Ch. 4.0-4.2
8	Wed. 9/13		Statistical Properties of Text II	
9	Mon. 9/18		Text Representation	CMS Ch. 4.3-4.7, <a href="#">MRS Ch. 2</a>
10	Wed. 9/20	HW1 Due, HW2 Out	Vector Space Model I	CMS Ch. 7.0-7.1.2
11	Mon. 9/25	Well-Being Day (No Class)		
12	Wed. 9/27	Literature Review Proposal Due	Vector Space Model II	
13	Mon. 10/2		Query Likelihood Model I	CMS Ch. 7.3, CMS 4.5
14	Wed. 10/4	HW2 Due	Query Likelihood Model I	
15	Mon. 10/9		Pseudo-relevance Feedback	
16	Wed. 10/11		Evaluation Overview	CMS Ch. 8
21	Mon. 10/16	Midterm Review	Midterm Review	
22	Wed. 10/18	Midterm	Midterm	
23	Mon. 10/23		Test Collection Evaluation I	<a href="#">Sanderson '10 (pages 248-298)</a> , <a href="#">Hersh et al., '00</a> , <a href="#">Turpin &amp; Hersh '01</a>
24	Wed. 10/25		Test Collection Evaluation II	<a href="#">Sanderson '10 (pages 308-350)</a>
25	Mon. 10/30	HW3 Out	Experimentaion I	<a href="#">Smucker et al., '07</a> , <a href="#">Cross-Validation, Parameter Tunning and Overfitting</a>
26	Wed. 11/1		Experimentaion II	
27	Mon. 11/6		Interactive Information Retrieval I	
28	Wed. 11/8		Interactive Information Retrieval II	
29	Mon. 11/13	HW3 Due	Search Log Analysis	<a href="#">Joachims et al., '05</a>
30	Wed. 11/15	TREC Conference (No Class)		
31	Mon. 11/20		A/B Testinig I	<a href="#">Dmitriev et al., '17</a> , <a href="#">Video Tutorial (Kohavi et al. '17)</a>
32	Wed. 11/22	Thanksgiving (No Class)		
33	Mon. 11/27		A/B Testing II	
34	Wed. 11/29		Literature Review Presentations I	
35	Mon. 12/4		Literature Review Presentations II	
36	Wed. 12/6		Literature Review Presentations III	
37	TBD	Literature Review Due		
38	TBD	Final Exam		