

LIS 544 / IMT542 / INSC 544: Information Retrieval Systems

Information School, University of Washington

Quarter: Autumn 2010

Class Meets: Thursdays, 4:30 PM - 7:20 PM in EEB 025

Course Website: <http://courses.washington.edu/ir2010/>

Class Email List: lis544a_au10@uw.edu

Instructors: Jeff Huang (lazyjeff@uw.edu); Shawn Walker (stw3@uw.edu)

Office Hours: By Appointment

Course Description

Introduction to information systems for the storage and retrieval of unstructured information. We will examine information retrieval architectures, processes, retrieval models, archiving of web content, query languages, and methods of system evaluation. Emphasis will be given to Internet-based services for storing and accessing information to be used in integrated application development.

Day	Topic(s)	What to Have Read	What's Due/Going On
9/30	Course Introduction & Documents (J)	<ul style="list-style-type: none">Chapter 1: Boolean retrieval. In. C.D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge : Cambridge University Press, 2008. Pp. 1-17.	
10/7	Queries (J&S)	<ul style="list-style-type: none">A taxonomy of web searchAn algorithm for suffix stripping, Section 1. IntroductionRelevance feedback and query expansion	<ul style="list-style-type: none">Reading Response
10/14	Evaluation (J)	<ul style="list-style-type: none">Evaluation in information retrieval	<ul style="list-style-type: none">Queries Assignment (S)Reading Response
10/21	Ads (J)	<ul style="list-style-type: none">Computer Scientists Optimize Innovative Ad AuctionA Quality-Based Auction for Search Ad Markets with Aggregators (read to 3.3)	<ul style="list-style-type: none">Reading Response
10/28	Networks (S)	<ul style="list-style-type: none">Otte, E. and R. Rousseau, Social network analysis: a powerful strategy, also for the information sciences. Journal of Information Science, 2002. 28(6): p. 441.Garton, Laura (et al), "Studying Online Social Networks," Journal of Computer-Mediated Communication, V. 3, N. 1, June, 1997, http://jcmc.indiana.edu/vol3/issue1/garton.htmlButts, C., Social network analysis: A methodological introduction. Asian Journal of Social Psychology, 2008. 11(1): p. 13. (Skip math sections)Barzilai-Nahon Karine and Jeff Hemsley, 2011, Democracy.com: A Tale of Political Blogs and Content, 44nd Hawaiian International Conference on System Sciences, Hawaii, January [forthcoming] (Skim)	<ul style="list-style-type: none">Reading Response

Day	Topic(s)	What to Have Read	What's Due/Going On
11/4	Crawlers (S)	<ul style="list-style-type: none"> • Cho, J., & Garcia-Molina, H. (2000). The evolution of the web and implications for an incremental crawler. Proceedings of the 26th international conference on very large data bases (p. 200–209). • Mohr, G., Stack, M., Rnitovic, I., Avery, D., & Kimpton, M. (2004). Introduction to Heritrix. <i>4th International Web Archiving Workshop</i>. • http://en.wikipedia.org/wiki/Web_crawler 	<ul style="list-style-type: none"> • SNA Assignment (S) • Reading Response
11/11	Links (J&S)	<ul style="list-style-type: none"> • The PageRank citation ranking: Bringing order to the web 	<ul style="list-style-type: none"> • Crawlers Assignment (S) • Reading Response
11/18	Behavior (J)	<ul style="list-style-type: none"> • Strong Regularities in World Wide Web Surfing • Learning user interaction models for predicting web search result preferences 	<ul style="list-style-type: none"> • SEO Game (J) • Reading Response
12/2	Archiving (S)	<ul style="list-style-type: none"> • TBD 	<ul style="list-style-type: none"> • Reading Response
12/9	Social (S)	<ul style="list-style-type: none"> • TBD 	<ul style="list-style-type: none"> • Archiving Assignment (S) • Reading Response
Finals Week			<ul style="list-style-type: none"> • Pagerank Game (J)

Readings

Students are expected to complete the readings before each class. Readings will be posted on the course website. Additionally, updated reading lists and cases will be posted on the website on a regular basis. Please make sure to check the web-site on a regular basis for the most up-to-date reading lists.

Assignments & Grading

Assignments	35%
Games	35%
Group Presentation	10%
Reading Responses	20%

Assignments

Four (4) assignments will be given throughout the quarter as “labs” so you have an opportunity to use/play with some of the technologies and concepts presented in class. Handouts will be posted to the class website. Assignments are normally due the week after the topic is presented in class. See the class schedule for more specific due dates.

Games

There will be two information retrieval games (like assignments but more fun) to play. One is the SEO game which will teach you about document analysis and ranking. The other is the Pagerank Game which will teach you about Pagerank and crawling. They will be posted on the calendar. There will be intermittent deliverables throughout the quarter for these games.

Reading Responses (weekly)

Weekly reading responses provide students an opportunity to reflect upon the readings before class. Please discuss how the readings relate to information retrieval, information science, library science, or information management. Also, list any especially interesting points, questions, difficulties you in the readings. **Weekly reading responses are due the morning before class (Thursdays, 9am).**

Parameters:

- This is an individual assignment
- The length should be about 1-2 pages
- **Don't forget to proofread!**
- Important: Please submit your work via [CollectIt](#), as a Word, text, or PDF file, using the following file-naming convention: Lastname-week#.docx (e.g. Walker-week1.docx).

Group Presentation

Each student will participate in a group presentation (2-3 members) during the term.

The task is to build on the assigned readings and topics presented in class and to create a 3-5 minute video OR 10 minute presentation consisting of no more than one (1) slide critique/presenting a current issue relevant to information retrieval.

The video/presentation and ensuing discussion should take no more than 15 minutes.

Collect It Dropbox

Reading Responses and Group Presentation self-evaluations will be turned in via Collect It (a Catalyst tool) which can be found on the course website.

Grading

You will receive a decimal grade for this class.

General grading information for the University of Washington is available at http://www.washington.edu/students/genclat/front/Grading_Sys.html

Your written work will be graded based on its clarity, organization, balance, amount of pertinent detail included, depth and clarity of evaluative and analytical comments, and preparation. It will also be graded on the extent to which a good understanding of the material presented in the course is shown and on the extent to which directions are followed. If evaluative or analytical comments are required, they should be supported by factual evidence, either from readings or other documents. Other aspects of individual assignments may also be included in the grading.

Written work that shows a lack of understanding of subject matter, is unclear or poorly organized, contains few or irrelevant details, does not follow directions, contains little or unsubstantiated evaluative commentary, or is poorly written, prepared (e.g. typos, grammatical errors), or documented will receive low grades.

Late assignments: lateness will incur an automatic 25% deduction in grade *per day*. As such, any assignments more than four days late will receive a zero, and will not be accepted. Re extenuating circumstances: if at all possible please notify the instructor BEFORE the day the assignment is due.

Students with Disabilities

To request academic accommodations due to a disability, please contact Disabled Student Services: 448 Schmitz, 206-543-8924 (V/TTY). If you have a letter from DSS indicating that you have a disability which requires academic accommodations, please present the letter to me so we can discuss the accommodations you might need in the class.

Academic Conduct

The following paragraphs discussing academic integrity, copyright and privacy outline matters governing academic conduct in the iSchool and the University of Washington.

Academic Integrity

The essence of academic life revolves around respect not only for the ideas of others, but also their rights to those ideas and their promulgation. It is therefore essential that all of us engaged in the life of the mind take the utmost care that the ideas and expressions of ideas of other people always be appropriately handled, and, where necessary, cited. For writing assignments, when ideas or materials of others are used, they must be cited. The format is not that important—as long as the source material can be located

and the citation verified, it's OK. What is important is that the material be cited. In any situation, if you have a question, please feel free to ask. Such attention to ideas and acknowledgment of their sources is central not only to academic life, but life in general.

Please acquaint yourself with the University of Washington's resources on academic honesty (<http://depts.washington.edu/grading/issue1/honesty.htm>).

Students are encouraged to take drafts of their writing assignments to the Writing Center for assistance with using citations ethically and effectively. Information on scheduling an appointment can be found at: <http://www.uwtc.washington.edu/resources/eiwc/>

Copyright

All of the expressions of ideas in this class that are fixed in any tangible medium such as digital and physical documents are protected by copyright law as embodied in title 17 of the United States Code. These expressions include the work product of both: (1) your student colleagues (e.g., any assignments published here in the course environment or statements committed to text in a discussion forum); and, (2) your instructors (e.g., the syllabus, assignments, reading lists, and lectures). Within the constraints of "fair use", you may copy these copyrighted expressions for your personal intellectual use in support of your education here in the iSchool. Such fair use by you does not include further distribution by any means of copying, performance or presentation beyond the circle of your close acquaintances, student colleagues in this class and your family. If you have any questions regarding whether a use to which you wish to put one of these expressions violates the creator's copyright interests, please feel free to ask the instructor for guidance.

Privacy

To support an academic environment of rigorous discussion and open expression of personal thoughts and feelings, we, as members of the academic community, must be committed to the inviolate right of privacy of our student and instructor colleagues. As a result, we must forego sharing personally identifiable information about any member of our community including information about the ideas they express, their families, life styles and their political and social affiliations. If you have any questions regarding whether a disclosure you wish to make regarding anyone in this course or in the iSchool community violates that person's privacy interests, please feel free to ask the instructor for guidance.

Knowing violations of these principles of academic conduct, privacy or copyright may result in University disciplinary action under the Student Code of Conduct.

Student Code of Conduct

Good student conduct is important for maintaining a healthy course environment. Please familiarize yourself with the University of Washington's Student Code of Conduct at: <http://www.washington.edu/students/handbook/conduct.html>