

[SILS, U. of North Carolina, Chapel Hill](#)
INLS-509 -- Information Retrieval

[Bob Losee](#)

Manning 302

962-7150

losee at unc dot edu

Fall 2009

Brief Description:

An introductory survey of information filtering and retrieval, with an emphasis on developing the student's understanding of the relationship between the algorithms used by search engines, the query and document, and system performance. This is an information science course, not an information technology course. The course is required for students in the School's Masters in Information Science program and will emphasize basic knowledge useful for those who will be in leadership positions in a wide range of information professions.

Course WWW links:

<http://InformationRetrieval.US>

(if you forget, there is link from [my home page](#))

Course Evaluation:

Quality of class participation 30%

Critiques of readings 35%

Other homework 35%

Constructive Critiques of Readings:

For some articles listed on the course schedule, students are expected to write a *constructive critique* of the article of 5 to 10 sentences in length (maximum $\frac{3}{4}$ page single spaced, 1 page double spaced) and hand in the critique (on paper, not via email, and use serif fonts for the body of the text) by the beginning of the class on the due date listed on the [schedule](#). The critiques should be *constructive*, emphasizing ways that the research could be improved or expanded, and might include questions that arose as you read the article whose answers would be useful, possible research questions that could be turned into (and are focused enough and small enough to be) SILS Master's papers, along with methodologies for addressing these questions. Do not criticize the author's writing style or the choice of topic or how you wish the article had been written; emphasize how you personally might be able to expand on the material described in the article. Think of this as a chance to try out some two to five sentence research proposals, one of which might lead to your end-of-the-semester proposal. That, in turn, might be a possible master's paper hypothesis. The one lowest critique grade will be dropped, to cover "bad days," critiques that don't get handed in on-time, or sickness.

End of the Semester Proposal

By noon **December 11**, the Friday of the last week of class, students will hand in a printed research proposal based upon one of the critiques they wrote during the semester. The proposal should contain a clearly stated research hypothesis in the first paragraph. This research proposal should be 3 to 6 pages, single spaced, and should include enough discussion of the related literature to base the different aspects or components of your research hypothesis and how it would be answered in the context of the research literature. The proposal needs to emphasize an area dealing with system performance as opposed to user performance. Ideally, the hypothesis will be constructive, and if the proposal were carried out, possibly in a master's paper, concrete recommendations about system design would be made. Be clear about the methodology, the kind of results that might be obtained, and major problems that exist due to the methodology.

Sources of Information on Information Filtering & Retrieval

Serials:

The major serials covering IR include *Information Processing and Management* (formerly *Information Storage and Retrieval*), *Journal of the American Society for Information Science and Technology* (formerly *JASIS* and before that *American Documentation*), *Journal of Documentation*, *IEEE Trans on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Data and Knowledge Engineering*, *ACM Transactions on Information Systems*, and *Information Retrieval*.

Monographs:

(** Best works or classics marked with asterisks)

Baldi and Brunak, *Bioinformatics: The Machine Learning Approach*, MIT, 2001.

Baldi, Frasconi, and Smyth, *Modeling the Internet and the Web*, Wiley, 2003.

Baeza-Yates and Ribeiro-Neto, *Modern Information Retrieval*, Addison Wesley, 1999.

Case, Donald, *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*, Academic Press, 2002.

Chen, Li, and Wang, *Machine Learning and Statistical Modeling Approaches to Image Retrieval*, Kluwer, 2004.

Chu, Heting, *Information Representation and Retrieval in the Digital Age*, ASIST, 2003.

Croft, Metzler, and Strohman. *Search Engines: Information Retrieval in Practice*. Addison Wesley, 2009.

Dominich, Sandor. *The Modern Algebra of Information Retrieval*, Springer 2008.

- Feldman, R. and Sanger, J. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge U. Press, 2006.
- ** Foskett, A. C., *The Subject Approach to Information*, London, Lib. Assoc. Publ, 1996.
- Forsyth and Rada, *Machine Learning; Applications in Expert Systems and Information Retrieval*, Wiley, 1986.
- ** Frakes and Baeza-Yates, eds., *Information Retrieval: Data Structures & Algorithms*, Prentice Hall, 1992.
- Grossman and Frieder, *Information Retrieval: Algorithms and Heuristics*, Second edition, Springer-Verlag, 2004.
- Grefenstette, *Cross-Language Information Retrieval*, Kluwer, 1998.
- Hersh, *Information Retrieval: A Health and Biomedical Perspective*, Third edition, Springer-Verlag, 2008.
- Korfhage, *Information Storage and Retrieval*, Wiley, 1997.
- Kowalski and Maybury, *Information Storage and Retrieval Systems*, Kluwer, 2000.
- Langville and Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton, 2006.
- Losee, *Text Retrieval and Filtering*, Kluwer, 1998.
- Manning, Raghaven, and Schutze. [Introduction to Information Retrieval](#), Cambridge, 2008.
- ** Manning and Schutze, *Foundations of Statistical Natural Language Processing*, MIT Press, 1999.
- Salton, *Automatic Text Processing*, Addison-Wesley, 1989.
- ** Salton and McGill, *Introduction to Modern Information Retrieval*, McGraw Hill, 1983
- Sparck Jones and Willett, *Information Retrieval*, Morgan Kaufmann Publishers, 1997.
- Van Rijsbergen, *Geometry of Information Retrieval*, Cambridge, 2004.
- ** Van Rijsbergen, *Information Retrieval, Second Edition*, Butterworth, 1979.
- Voorhees, E. and Harman, D. *TREC: Experiment and Evaluation in Information Retrieval*, MIT, 2005.
- Wu, Xiong, and Shekhar, *Clustering and Information Retrieval*, Kluwer, 2004.

Honor Code:

Students should familiarize themselves with the University of North Carolina at Chapel Hill Honor Code that is described in University publications. It should be noted that in this course, students are expected to receive (and provide) some assistance regarding the use of hardware and software and general problem solving techniques for homework assignments. Students should NOT receive (or provide) major creative assistance or continuing minor support for projects.

Plagiarism:

Student assignments that are handed in that contain more than 5 consecutive words that the instructor feels were taken from another source without proper attribution (without the proper quote marks and citations) *definitely will be referred* to the appropriate administrative authorities who address issues of Academic Integrity (e.g. the *Honor Court*) I assume that all students are equally likely to be honest and will put an equal amount of effort into considering the possibility of plagiarism for each student's paper.

Classroom Behavior:

Separate from the Honor Code but related to respect for classmates is classroom behavior, which will be a factor in your class participation grade. Students are expected to behave in a professional manner in class. Students in class are expected to focus on classroom materials. Students are expected to avoid student-to-student conversations during class. **Use of laptop computers should be limited to taking notes for class and to using class related materials.** Similarly, materials being read should be limited to those appropriate for the classroom lecture or discussion. Students who appear to be involved in non-class related activities during class time will be graded as not participating in class. Cellular telephones and computers should have speakers or other audio devices muted before class begins so as to not disturb others.