

Course Information for CS533

Fall 2009

Prerequisites: CS333

Grades

1. Grading policy

Mid term exam	20%
Final exam	20%
4-6 Theoretical Assignments	20%
1-3 Programs	20%
Essay and presentation	10%

EngiNet students do not need to present

2. Missing grades

Grades will be posted on Blackboard. Please, check your status on blackboard periodically. A missing grade at the end of the semester will indicate that the work has NOT been done.

3. Late Homework:

Late homework will not be accepted.

4. Exams:

The exams are open book.

5. Programs

Please make an effort to make your programs easy to understand and grade. All programming assignment should have:

- A general explanation of the program
- A description of the purpose of each class and its methods, and an explanation of how it works.

- A description of the purpose, and the assumptions made about each parameter of a function.
- A comment for every variable declaration.
- The code should also be well commented.

Programs may be written in Java, C, C++, or JavaScript. The use of other languages should be discussed with instructor.

Office Hours:

I encourage you to take advantage of the office hours.

Text:

URL: <http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html>

Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.

Recommended reading

*****Bruce Croft, Donald Metzler, Trevor Strohman, "Search Engines: Information Retrieval in Practice:" Addison Wesley/Prentice Hall, 2010.

Baeza-Yates R. and Ribeiro-Neto B., "Modern Information Retrieval", Addison Wesley 1999.

Belew Richard K., "Finding Out About. A Cognitive perspective on Search Engine Technology and the WWW", Cambridge University Press, 2000.

**** Liu Bing, "Web Data Mining Exploring Hyperlinks, Contents, and Usage Data", Springer, 2007

Witten I. H., Moffat A., Bell T. C., "Managing Gigabytes, Compressing and Indexing Documents and Images 2nd Edition", Morgan Kaufmann, 1999.

****Chakrabarti Soumen "Mining the web, discovering knowledge from hypertext data", Morgan Kaufmann publishers, 2003.

Previously Used Texts:

1. Frakes W. B., and Baeza-Yates R. (editors) "Information Retrieval Data Structures and Algorithms", Prentice-Hall 1992. Source code to the programs in this book are available, from <ftp://ftp.uu.se/pub/unix/networking/wais/ir-book-sources/>
2. Kowalski G. "Information Retrieval Systems Theory and Implementation" Kluwer Academic Publishers, 1997.
3. Salton G, "Automatic Text Processing", Addison Wesley 1989.
4. Witten I. H., Moffat A., Bell T. C., "Managing Gigabytes, Compressing and Indexing Documents and Images", VNR, 1994.
5. Grossman D.A. and Frieder O, "Information Retrieval: Algorithms and Heuristics", Kluwer, August 1998

Other Books

1. "Discovering HTML 4", Bryan Pfaffenberger, Morgan Kaufman Publishers, August 1998 ISBN 0-12553-167-2.
2. "Dynamic HTML for Webmasters", Tom Dell, Morgan Kaufman publishers, September 1998, ISBN 0-12209-065-9.
3. "Principles of Multimedia Database Systems", V.S. Subrahmanian, The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor 1998, ISBN 1-55860-466-9
4. "Visual Information Retrieval" Alberto del Bimbo, June 1999, Morgan Kaufman Publishers, ISBN 1-55860-624-6.
5. Abiteboul S., Buneman P., and Suciu D., "Data on the Web: From Relations to Semistructured Data and XML", Morgan Kaufman Publishers, October 1999, ISBN 1-55860-622-X
6. Alberico, R, and Micco M., "Expert Systems for Reference and Information Retrieval", Westport, CT: Meckler, 1990.
7. Allen B. L. "Information Tasks: Toward a User Centered Approach to Information Systems" Academic Press, San Diego, CA, 1996
8. Folk M. J., Zoellick B., and Riccardi G., "File Structures, an Object-Oriented Approach with C++", Addison-Wesley, 1998
9. Frants V. J., Shapiro J., Voiskunskil V. G. "Automated Information Retrieval Theory and Methods" Academic Press, Aug 1997.
10. Gilster, Paul, "Finding it on the Internet : the Internet Navigator's guide to search tools and techniques, New York : Wiley, 1996.
11. Grefenstette G., "Cross Language Information Retrieval", Kluwer International Series on Information Retrieval, ISBN 0-7923-8122-X, March 1998
12. Harman D. K., "The Fourth Text Retrieval Conference (TREC-4), NIST Special Publications 500-236, October 1997.

13. Harman D. K., "The Third Text Retrieval Conference (TREC-3), NIST Special Publications 500-225, April 1995.
14. Hock R. "The Extreme searcher's Guide to Web Search Engines, A Handbook for the Serious Searcher", CyberAge Books, 1999
15. Lesk M., "Practical Digital Libraries Books, Bytes & Bucks", Morgan Kaufman Publishers, July 1997.
16. Mauldin M. L. "Conceptual Information Retrieval : a case study in adaptive partial parsing", Kluwer Academic Publishers, 1991
17. Maybury M. T., "Intelligent multimedia information retrieval", Menlo Park, Calif. : AAAI Press ; Cambridge, Mass, MIT Press, 1997
18. Meadow C. T., "Text Information Retrieval Systems", Academic Press 1992.
19. Miyamoto S. "Fuzzy Sets in Information Retrieval and Cluster Analysis", ISBN 0-7923-0721-6, Kluwer Academic Publishers, May 1990
20. Paul J. (editor) "Text-Based Intelligent Systems, current research and practice in information extraction and retrieval", LEA 1992.
21. Pfaffenberger B., "Web Search Strategies" MIS Press, 1996
22. Rijsbergen C. J., Crestani F., and Lalmas M., "Information Retrieval: Uncertainty and Logics Advanced Models for the Representation and Retrieval of Information", ISBN 0-7923-8302-8, Kluwer Academic Publishers, October 1998
23. Rosenfeld L. and Morville P., "Information Architecture for the World Wide Web" O'Reily & Associates 1998.
24. Salton G., and McGill M., "Introduction to Modern Information Retrieval", McGraw-Hill, New York, NY, 1983.
25. Schäuble, Peter, " Multimedia information retrieval: content-based information retrieval from large text and audio databases" Kluwer Academic, 1997.
26. Sparck Jones K. and Willett P (editors)., "Readings in Information Retrieval", Morgan Kaufman Publishers, July 1997.
27. Strzalkowski T., Editor, "Natural Language Information Retrieval" (CD-Rom), ISBN 0-7923-5685-3, Kluwer Academic Publishers, April 1999
28. Zernik U. (editor) "Lexical Acquisition, Exploiting On-Line resources to Build a Lexicon" Lawrence Erlbaum Associates, 1991.

Topics

Chapter

- | | |
|----|--|
| 01 | Boolean retrieval |
| 02 | The term vocabulary & postings lists |
| 03 | Dictionaries and tolerant retrieval |
| 04 | Index construction |
| 05 | Index compression |
| 06 | Scoring, term weighting & the vector space model |

07	Computing scores in a complete search system
08	Evaluation in information retrieval
09	Relevance feedback & query expansion
10	XML retrieval
11	Probabilistic information retrieval
12	Language models for information retrieval
13	Text classification & Naive Bayes
14	Vector space classification
15	Support vector machines & machine learning on documents
16	Flat clustering
17	Hierarchical clustering
18	Matrix decompositions & latent semantic indexing
19	Web search basics
20	Web crawling and indexes
21	Link analysis

This semester

1 -9, 10-11, 18-21