LIS618

Online Information Retrieval Techniques

last revised: 2011–09–21

The latest version of this document is available on the web at http://openlib.org/home/krichel/courses/lis618p11a/.

Course Description
This course will introduce the students to the theory of information retrieval and its application in large-scale commercial database systems and on the web.

Course objectives
On completing this course, students

- will have been introduced to information retrieval models;
- will have been introduced to several commercial database systems and be aware of their strengths and weaknesses;
- will have been introduced to expert search strategies with web search engines and databases.

The /home/krichel/liu_admin/pslo.html aimed at by this course is 3E, "Students will demonstrate appropriate techniques for identification, selection, acquisition, retrieval, evaluation and synthesis of information from a variety of information systems and services."

Prerequisites
Students should have a basic command of the Microsoft Windows operating system because the machines in the lab run on this operating system.

Instructor
Thomas Krichel
Palmer School of Library and Information Science
C.W. Post Campus of Long Island University
720 Northern Boulevard
Brookville, NY 11548–1300
krichel@openlib.org
work phone: +1–(516)299–2843

Class structure

Classes will be held on the CW Post campus of LIU, between 19:00 and 21:00. After class students can stay on for guided practice. There will be a mixture of lectures and hands-on work in the lab. Provisional class details are:

1	2011–09–12	introduction to the course
2	2011–09–19	history of information retrieval
3	2011–09–26	preprocessing of records and queries
4	2011–10–03	the Boolean model of retrieval
5	2011–10–17	introduction to search and to Dialog
6	2011–10–24	the Dialog command language
7	2011–10–31	Dialog by example
8	2011–11–07	vector model and ProQuest
9	2011–11–14	Credo and Gale
10	2011–11–21	web information retrieval
11	2011–11–28	Google and Bing
12	2011–12–05	Google Scholar, Web of Knowlegde, IR performance measures
13	2011–12–12	constructing a search interface

Slides for all classes are downloadable from the course web site.

Class mailing list
A mailing list for this class has been set up. Students who wish to stay informed are encouraged to sign up.

Readings
The powerpoint slides of the instructor are the reading prime reading material. The slides may point to other sources of reference as required.
Readings that the slides are derived from include van Rijsbergen (1979), Baeza-Yates and Ribeiro-Neto (1999), Manning, Raghavan, and Schütze (2009) and maybe even Hock (2010).
Background readings on history include Lesk (1995), Schatz (1997), Salton (1987).
Historically significant contributions Luhn (1957), Spärk-Jones (1972), and Salton, Wong, and Yang (1975). Bourne (1963) provides an overview over technology in the early 60s.
On Dialog, we use the Dialog Command Language Pocket Guide.
On search interface we will use Hearst (2009).
Some `reference questions`[1] to work on are available.

Assessment
Each student will have to prepare a search exercise and report as detailed in the first lecture. It will count for 50% of the total grade. It is due on December 12. The report must not exceed five pages. Appendices are permitted, but may not get read. The remaining 50% will come from quizzes held at the start of each lecture except the first. Quizzes will last around ten minutes. The questions aim for a precise, short answer. The worst five quiz performances will be discarded when the average is being computed.

Contact hours breakdown
This is a breakdown of additional work that has to be done outside class by week. As per New York state regulations, the course needs to contain 120 additional hours of work outside class. This is an estimated breakdown of these times by week, as ordered by the Palmer School director.

References

Baeza-Yates, Ricardo and Berthier Ribeiro-Neto (1999). *Modern Information Retrieval* (first ed.). Addison Wesley.

Bourne, Charles P. (1963). *Methods of Information Handling*. John Wiley and Sons. available at http://openlib.org/home/krichel/courses/lis618/readings/bourne63_method_infor_handl.pdf.

Hearst, Marti (2009). *Search User Interfaces*. Cambridge University Press. available at http://searchuserinterfaces.com/.

Hock, Randolph (2010). *The Exteme SearcherâĂŹs Internet Handbook : a guide for the serious searcher* (3rd ed.). Cyber Age Books.

Lesk, Michael E. (1995). The Seven Ages of Information Retrieval. UDT occasional paper, available at http://archive.ifla.org/VI/5/op/udtop5/udtop5.htm.

Luhn, Hans-Peter (1957, October). A statistical approach to mechanized encoding and searching for information. *IBM Journal*, 309–317. available at http://openlib.org/home/krichel/courses/lis618/readings/luhn57.pdf.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze (2009). *Introduction to Information Retrieval*. Cambridge University Press. available at http://nlp.stanford.edu/IR-book/.

Salton, Gerard M. (1987). Historical Note: The past Thirty Years in Information Retrieval. mimeo 87-827, Department of Computer Science, Cornell University.

Salton, Gerard M., Andrew K.C. Wong, and Chung-Shu Yang (1975). A vector space model for automatic indexing. *Communications of the ACM 18(11)*, 613–620. available at http://openlib.org/home/krichel/courses/lis618/readings/salton75.pdf.

Schatz, Bruce R. (1997). Information Retrieval in Digital Libraries: Bringing Search to the Net. *Science 275*, 327–333. available at http://openlib.org/home/krichel/courses/lis654/readings/schatz97_infor_retriev_digit_librar.pdf.

Spärk-Jones, Karen (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation 28(1)*, 11–21. available at http://openlib.org/home/krichel/courses/lis618/readings/spaerk-jones72.pdf.

[1] reference_questions.html

van Rijsbergen, Keith (1979). *Information Retrieval*. London: Butterworths. available at http://openlib.org/home/krichel/courses/lis618/readings/rijsbergen79_infor_retriev.pdf.