ITR8

Information Retrieval

2001–01–19

The latest version of this document is available on the web at http://openlib.org/home/krichel/itr8p01s/itr8p01s.html.

Course Description
Information retrieval has long been a subject that interested a small group of computer scientists, information scientists as well as some librarians. Naturally these groups do not have the same perspective of the subject. This course aims to balance the computer science and librarian perspective of information retrieval.
Information retrieval has long been confined to building application to retrieve extracts from a predefined stock of content. With the rise of the web, information retrieval issues on the web have started to attract a lot of interest. This course aims to give a balanced overview of issues that arise in tradition information retrieval and web-based systems.

Course objectives
After completing the course, students will be able to

- understand of the basic questions that arise in information retrieval
- understand issues of data encoding and compression
- organize data prior to information retrieval
- formulate effective search strategies and queries
- understand the basic concept for retrieval evaluation
- evaluate user interfaces

Prerequisites
Students should be familiar with elementary symbolic algebra.

Readings
There are two texts used in several lectures.
Korfhage, Robert R., “Information Storage and Retrieval”, published by John Wiley & Sons, 1997
This book will be referred to as RRK throughout the course documentation. Baeza-Yates, Ricardo and Berthier Ribeiro-Neto “Modern Information Retrieval”, ACM Press 1999.
This book will be referred to as BYRN throughout the course documentation. Students should not read that book unless specifically requested by the instructor.
Other papers used
Brin, Sergey and Page, Lawrence (1998) “The anatomy of a Large-Scale Hyertextual Search Engine”

Class structure
Classes will be held on Tuesdays and Thursdays between 14:00 and 15:20 in the computer lab.

	date	*topic*	*reading*	*assignment*
01	1–16	introduction of course participants		
02	1–18	introduction to the topic	RRK 1–13	
03	1–23	retrieval on the web		web retrieval exercise
04	1–25	user interfaces	BYRN ch. 10	
04	1–30	search engines	Brin and Page	user interfac
06	2–01	sofix		
07	2–6	creating well formed XML		miniexam
08	2–8	data input session I		
09	2–13	XML applications		application description
10	2–15	data input session II		
00	2–20	*no class*		
11	2–22	creating valid XML		
12	2–27	displaying XML		sofix file
13	3–01	revision class		
14	3–06	mid-term exam		
15	3–08	introduction to datasbases		
16	3–13	table operations		style sheet
17	3–20	more table operation		
18	3–22	sample queries		
19	3–27	tables in MS Access		
20	3–27	more on MS Access		

Assessment
Class discussion and activities form an important part of this course. Missing more than two classes without the approval of the instructor will result in lowering the grade for the course by one grade.
Handed-in exercises have to be handed in the Tuesday class. Each day of delay reduces the mark by one grade.
For both handed-in assignment only the three out of five that have the best marks will count. For the mini exams, only the three that have the best marks will count. Everybody is allowed to have a bad day.
The final grade will be composed as follows

item	*weight*
mini exams	25%
exercises	25%
mid-term exam	25%
final exam	25%

Instructor
Thomas Krichel
Room 330
B. Davis Schwartz Memorial Library
The Palmer School of Library and Information Science
C.W. Post Campus
Long Island University
720 Northern Boulevard, Brookville, New York 11548–1300
t.krichel@surrey.ac.uk
work phone: +1–(516)299–2843
Private contact details may be obtained from the online CV at /home/krichel/cv.html.