IV2002 Internet Search and Monitoring Techniques 7.5 credits

Teknik för internetsökning och omvärldsbevakning

This is a translation of the Swedish, legally binding, course syllabus.

Establishment

Course syllabus for IV2002 valid from Autumn 2008

Grading scale

A, B, C, D, E, FX, F

Education cycle

Second cycle

Main field of study

Specific prerequisites

For single course students:

Prerequisites (Förkunskaper): Apart from a completed upper secondaryeducation and very good knowledge of English, basic knowledge of databases and programming is recommended. It is assumed that the students use the Internet and search engines on a regular basis.

Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

Intended learning outcomes

The course gives an insight into the techniques for information searching and monitoring applied on the Internet. After the course is finished, the students should be able to:
- Understand the techniques of information retrieval.
- Understand the architecture and main algorithms used by Internet search engines as well as Business Intelligence applications.
- Set requirements to, compare and measure the quality of information retrieval tools.
- Understand why not all on-line information is covered by general-purpose Internet search engines and know the ways of finding such information
- Understand and choose between different approaches and techniques of automated question answering

Course contents

Fundamentals of Information Retrieval: Boolean, term weight- and vector-space text retrieval models; document similarity measures; quality measures - precision and recall; index of documents and its access methods; morphologic and semantic analysis in text retrieval.
Query analysis: Processing the search word and index using word stemming, query expansion, fuzzy matching, compound splitting and compound joining that increase the quality of search. Other techniques are automatic translation of search words to other languages to make cross language information retrieval.
Information clustering and presentation: Sorting of text flows using automatic clustering and semi automatic clustering. Automatic document summarization removes redundant information from a document and creates a shorter summarized document. Multi document summarization summarizes several documents to one document. Using machine translation to present results in the users native language.
Search Engines: Architecture of a search engine; crawlers and features that hinder crawling; keyword-based retrieval; link analysis and PageRank; optimization of websites for search engines (Search Engine Optimization) and search engine spamming; paid listing; meta-search engines; web directories. Furthermore, there exist authoritative information accessible over the Internet and not visible to ordinary search engines. This material resides on the "invisible web", which is largely comprised of content-rich databases from universities, libraries, associations, businesses, and government agencies.
Monitoring tools: News archives and indexing tools, news alerts and agents, and RSS based news surveillance tools.
Question-Answering Systems deliver the answer to the question the user has in mind while searching, instead of a ranked list of documents. The three main question-answering approaches are based on Natural Language Processing, Information Retrieval, and question templates.

Disposition

Half speed
Credits (hp): 7,5
Lectures: 16 lectures x 2 hours
Assignments: 3
Laborations: 2 occasions x 2 hours
Seminar task:1 lecture x 2 hours
Groups
Laborations and seminar task are all carried out in groups of maximum two students.
Assignments in groups of maximum of four students.
Laborations are carried out at university at fixed times under supervision of the course managers.
The assignments are carried at home but there are occasions of supervision where the students can ask questions and get support from the teacher.
The seminar task is prepared at home in form of 6-7 written pages, but presented at one occasion where fellow students can ask questions and criticize. We encourage that Swedish and International student to mix in groups in order to obtain higher language competence. Some of the task may contain some Swedish text.
Distance students
The distance student must only participate physical for the exam, the rest of the tasks are solved completely at distance.
The distance education students must be present at the campus for the exam, the rest of the tasks can be solved using electronic means of communication.
If a distance student has no possibility to form a group then the student is allowed to solve all tasks alone, http://www.dsv.su.se/~eriks/66BI/66BIdist.html

Course literature

- R. Baeza-Yates, B. Ribeiro-Neto: Modern Information Retrieval, Harlow Addison-Wesley, 1999

- Våge, L., Dalianis, H. och Iselid, L.: Informationssökning på Internet (Upplaga: 1:a), Studentlitteratur, 2003, 91-44-03178-5

- Mark Levene: An Introduction to Search Engines and Web Navigation, Addison Wesley, 2005

- Mike Moran, Bill Hunt: Search Engine Marketing, Inc., IBM Press, 2005

Course compendium.

Examination

- INL1 - Assignment, 4.5 credits, grading scale: P, F

- TEN1 - Examination, 3.0 credits, grading scale: A, B, C, D, E, FX, F

Based on recommendation from KTH's coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

If the course is discontinued, students may request to be examined during the following two academic years.

ssignments: 3
Laborations: 2
Seminar task:1
Written examination

The exam corresponds to 3 hp and has a grading of (A/B/C/D/E/Fx/F). The assignments, laborations and seminar task together corresponds to 4.5 hp. The assignments, laborations and seminar task have the grading (P/F).

Other requirements for final grade

To pass the course the student must pass the exam, the assignments, the laborations as well as the the seminar task. The grading of the whole course is based on the exam.

If the student is close to pass then we will give the possibility to make a complement of the examinations so the student can pass.

Ethical approach

- All members of a group are responsible for the group's work.
- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.