DD2475 Information Retrieval 9.0 credits

Informationssökning

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

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

Establishment

Course syllabus for DD2475 valid from Autumn 2010

Grading scale

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

Education cycle

Second cycle

Main field of study

Computer Science and Engineering

Specific prerequisites

Single course students: 90 university credits including 45 university credits in Mathematics or Information Technology. English B, or equivalent.

Language of instruction

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

Intended learning outcomes

After completing the course you will be able to:

- Explain the concepts of indexing, vocabulary, normalization and dictionary in Information Retrieval
- Define a boolean model and a vector space model, and explain the differences between them
- Explain the differences between classification and clustering
- Discuss the differences between different classification and clustering methods
- Choose a suitable classification or clustering method depending on the problem constraints at hand
- Implement classification in a boolean model and a vector space model
- Implement a basic clustering method
- Give account of a basic spectral method
- Evaluate information retrieval algorithms, and give an account of the difficulties of evaluation
- Explain the basics of XML and Web search.

Course contents

Basic and advanced techniques for information systems: information extraction; efficient text indexing; indexing of non-text data; Boolean and vector space retrieval models; evaluation and interface issues; XML, structure of Web search engines; clustering, classification; spectral methods, random indexing; data mining.

Course literature

C. D. Manning, P. Raghavan and H. Schütze: Introduction to Information Retrieval, Cambridge University Press, 2008.

Examination

- LAB1 - Laboratory Works, 3.0 credits, grading scale: P, F
- LAB2 - Project, 3.0 credits, grading scale: A, B, C, D, E, FX, F
- TEN1 - Exam, 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.

In this course all the regulations of the code of honor at the School of Computer science and Communication apply, see: http://www.kth.se/csc/student/hederskodex/1.17237?l=en_UK.

Other requirements for final grade

The students participating in the course are expected to take part in all activities in the course with a particular emphasis on the exercises and laboratories. In addition the course focuses on training:
   * independently acquiring knowledge
   * oral and written presentation
Examination by one written exam (TEN1; 3.0 credits), laboratory assignments (LAB1; 3.0 credits), and a project assigment assessed orally and in writing (LAB2; 3.0 credits).

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.