



DH2418 Language Engineering

6.0 credits

Språkteknologi

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

Establishment

Course syllabus for DH2418 valid from Autumn 2009

Grading scale

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

Education cycle

Second cycle

Main field of study

Computer Science and Engineering, Information Technology, Information and Communication Technology

Specific prerequisites

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

Language of instruction

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

Intended learning outcomes

The students should after the course have the knowledge to:

- Explain and use general concepts within the following levels of linguistics: morphology, syntax, semantics, discourse, and pragmatics.
- Use the knowledge about morphology, syntax, and lexical semantics in order to develop systems, and explain existing systems using these levels.
- Clarify the differences between analysis, generation, and filtering in text-based systems.
- Use general language technology tools and resources, such as part-of-speech taggers, chunkers, corpora, and lexica in order to build new applications.
- Explain and use standard methods based on rules, statistics, and machine learning.
- Apply methods based on finite automata/transducers, context-free grammars, word frequencies, n-grams, co-occurrence statistics, Markov models, and vector space models.
- Analyze and explain which problems within language technology that could be solved with usable results, and which could not be solved.
- Give details of how spelling- and grammar checkers, taggers based on machine learning, stemmers, and an algorithm for semantic content acquisition work.
- Design and carry out a simpler evaluation of a language technology system, and interpret the results.
- Independently solve a well-defined practical language technology problem, or analyze a problem theoretically.

To be able to:

- Work for a language technology company.
- Continue with studies in language technology.
- Work with a master's project in computer science or human-computer interaction with a focus on language technology.
- Be an important link between systems designers, programmers, and interaction designers in industry as well as in research projects.

Course contents

Theory:

The history and basics of language technology, morphology, syntax, and semantics, vector space models, evaluation methods, the principles and methods of terminology work, machine learning, information theory and Markov models, algorithms and data structures for efficient lexicon handling.

Methods:

Morphological analysis and generation, statistical methods in corpus linguistics, parsing and generation, part-of-speech tagging, named entity recognition and probabilistic parsing, statistical lexical semantics.

Application areas:

Spelling- and grammar checking, information retrieval, word prediction for smart text entry, text clustering and text categorization, computer assisted language learning, dialogue systems, text summarization, speech technology, localization and internationalization.

Course literature

Course literature will be announced at course home page not later than 4 weeks before the course starts. Previous year: Jurafsky & Martin, Speech and language processing and material produced at the department.

Examination

- INLA - Assignment, 1.5 credits, grading scale: A, B, C, D, E, FX, F
- TEN2 - Examination, 3.0 credits, grading scale: A, B, C, D, E, FX, F
- LAB2 - Laboratory Assignments, 1.5 credits, grading scale: P, 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.

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/heder-skodex/1.17237?l=en_UK.

Other requirements for final grade

Examination (TEN2; 3 university credits.).

Laboratory assignments (LAB2; 1,5 university credits.).

Home assignment (INLA; 1,5 university 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.