DD2418 Language Engineering 6.0 credits

Språkteknologi

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

Establishment

Course syllabus for DD2418 valid from Spring 2019

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

For non-program students, 90 credits are required, of which 45 credits have to be within mathematics or information technology. Furthermore, Swedish B or the equivalent and English A are required or the equivalent.

Language of instruction

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

Intended learning outcomes

Students should after course can:

- explain and use concepts in the following levels of linguistics: morphology, syntax, semantics, discourse and pragmatics,

- apply knowledge of morphology, syntax and lexical semantics to develop language engineering systems as well as explain structure of existing systems using these levels

- clarify the differences between analysis, generation, and filtering in text-based systems

- use basic tools in language engineering such as part-of-speech taggers, chunkers as well as various types of corpora and dictionary to be able to build own programs,

- explain and use standard methods in language technology that are based on both rules, statistics and machine learning,

- practically apply methods that are based on finite automata/transducers, context free grammars, word frequencies, n-grams, co-occurrence statistics, Markov models, and vector space models,

- analyse and explain which language engineering problems that could be solved with usable results as well as which that lie beyond the research horizon,

- explain in detail how spell checking, grammar checking, some kind of tagger using machine learning, stemmer and an algorithm for statistical extraction of related words work,

- design and carry out simple evaluations of some language engineering system as well as interpret the results,

- independently solve a well-defined practical language technology problem, or analyze a problem theoretically,

to be able to:

- work for language technology companies

- continue with language technology oriented studies,

- carry out a degree project in computer science or human computer interaction with a language engineering specialisation

- 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, generation and language statistics and corpus processing, parsing, generation, part-of-speech tagging, named entity recognition, probabilistic parsing and statistical lexical semantics.

Application areas:

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

Course literature

The reading list will be announced on the course webpage no later than 4 weeks before start of the course. The previous course offering used Jurafsky & Martin, Speech and language processing as well as material produced at the department.

Examination

- INL1 - Hand-ins, 1.5 credits, grading scale: A, B, C, D, E, FX, F
- LAB2 - Laboratory work, 4.5 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.

The course DD1418 has until now shared examination with DD2418. Students who have initiated DD2418 earlier but still not have passed the exam can take the exam in DD1418. This possibility is retained in two years.

Other requirements for final grade

For final course grade is required that student has passed the components INL1 and LAB2. The final grade will be a weighted mean of the grades on the two components.

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.