

Syllabus

for course at advanced level

Language Engineering
Språkteknologi

**6.0 Higher Education
Credits
6.0 ECTS credits**

| | |
|-------------------|---|
| Course code: | DA7032 |
| Valid from: | Autumn 2011 |
| Date of approval: | 2011-01-18 |
| Department | Department of Mathematics (incl. Math. Statistics) |
| Main field: | Computer Science |
| Specialisation: | A1N - Second cycle, has only first-cycle course/s as entry requirements |

Decision

This syllabus has been approved by the Board of the Faculty of Science at Stockholm University, October xx, 2009.

Prerequisites and special admittance requirements

For course admission knowledge equivalent to the following is required:

- the courses Object Oriented Programming, FL, 7.5 HECs (DA3002) and Probability Theory I, FL, 7.5 HECs (MT3001)
- or
- 60 HECs in Computer and Systems Sciences
- or
- 30 HECs Computer Linguistics.

Course structure

| Examination code | Name | Higher Education Credits |
|------------------|---------------------|--------------------------|
| THEO | Theory | 3 |
| ILUP | Home Assignment | 1.5 |
| LABO | Practical Exercises | 1.5 |

Course content

The course covers: Different methods for analysis, generation, and filtering of human language especially text.

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.

b. The course includes the following elements:

- Theory, 3 HECs
- Practical Exercises, 1.5 HECs
- Home Assignments, 1.5 HECs

Learning outcomes

It is expected that the student after taking the course will be able 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

Education

The education consists of lectures and practical exercises.

Participation in practical exercises and group education associated with this is compulsory. An examiner may rule that a student is not obliged to participate in certain compulsory education, if there are special grounds for this, after consultation with the relevant teacher.

Forms of examination

a. Examination for the course is in the following manner: measurement of knowledge of the element Theory takes place through written and oral examination, and of the element Home Assignment through written and oral presentation.

b. Grading is carried out according to a 7-point scale related to learning objectives:

- A = Excellent
- B = Very Good
- C = Good
- D = Satisfactory
- E = Sufficient
- Fx = Fail
- F = Fail

c. Grading criteria for the course will be distributed at the start of the course.

d. A minimum grade of E is required to pass the course, together with pass of the element Practical Exercises and participation in all compulsory education.

e. Students who fail to achieve a pass grade in an ordinary examination have the right to take at least further four examinations, as long as the course is given. The term “examination” here is used to denote also other compulsory elements of the course. Students who have achieved a pass grade on an examination may not retake this examination in order to attempt to achieve a higher grade. Students who have failed to reach a pass grade on two occasions have the right to request that a different teacher be appointed to set the grade of the course. A request for such appointment must be sent to the departmental board.

Interim

Students may request that the examination is carried out in accordance with this syllabus even after it has ceased to apply. This right is limited, however, to a maximum of three occasions during a two-year-period

after the end of giving the course. A request for such examination must be sent to the departmental board.

Limitations

The course may not be included in a degree together with the course Language Engineering, Second Cycle (DA7012), or Language Engineering for Computational Linguists, First Cycle (DA3010), or the equivalents.

Misc

The course is given as an individual course, but can also be a component of the Master's Programme in Computer Science.

Required reading

Course literature is decided by the departmental board and is described in an appendix to the syllabus.