



DEPARTMENT OF PHILOSOPHY, LINGUISTICS AND THEORY OF SCIENCE

LT2003 Natural language processing, 15 credits

Grundläggande språkteknologi, 15 högskolepoäng

Second Cycle

Confirmation

This course syllabus was confirmed by Department of Philosophy, Linguistics and Theory of Science on 2018-05-29 and was last revised on 2019-05-31 to be valid from 2019-09-02, autumn semester of 2019.

Field of education: Science 100%

Department: Department of Philosophy, Linguistics and Theory of Science

Position in the educational system

It can also be offered as a freestanding course.

The course can be part of the following programme: 1) Master in Language Technology (One year or Two years) (H2MLT)

Main field of studies

Language Technology

Specialization

A1N, Second cycle, has only first-cycle course/s as entry requirements

Entry requirements

For admission to course a Bachelor Degree in some of the following subjects:

computer science, computational linguistics or language technology, linguistics (including at least 30 credits in formal linguistics or programming), adjacent subject, for instance cognitive science, languages, philosophy or mathematics, (provided that the student has got 30 credits either in formal linguistics or programming),

or the equivalent after assessment is required for the admission to course.

Passed knowledge in English equivalent to English 5/English A (upper secondary course level) is mandatory.

Learning outcomes

Knowledge and understanding

- account for basic concepts in NLP, automata theory, probability theory, linear algebra, and calculus
- account for pros and cons with certain approaches, supervised/non supervised, different algorithms, etc.,
- account for different evaluation methods and their scope,
- account for common parsing algorithms,
- account for theories behind different data driven methods,
- know what plagiarism implies,
- have knowledge of GU rules about plagiarism.

Competence and skills

- apply current methods for basic NLP tasks such as part-of-speech tagging,
- use misc machine learning approaches to solve NLP problems together with annotated corpora,
- use basic machine learning and NLP programming tools,
- use command-line tools to manage data,
- use statistical evaluation methods,
- identify plagiarism,
- find information about plagiarism and GU rules about plagiarism.

Judgement and approach

- choose between different machine learning techniques and different data representations to solve a specific NLP task,
- choose appropriate features for evaluating a solved NLP problem.

Course content

The course gives a theoretical view of problems encountered within natural language processing, and some standard solutions.

Students will gain practical experience in programming while solving these problems. The programming language used in Introduction to programming, LT2001, will also be used in this course together with standard NLP libraries and command-line tools.

The course is divided into three main topics, one covering basic concepts and three covering NLP subfields.

1. Basic concepts:

- basic concepts in NLP
- basic automata theory and mathematical linguistics
- basic probability theory and machine learning
- basic algebra and calculus
- evaluation measurement, correctness, precision and recall.

2. Words and sentences:

- corpora and corpus annotation
- finite state methods for segmentation and morphological analysis
- statistical language modelling with n-gram Markov models
- vector space representations and operations

3. Overview of common NLP tasks, including, as time permits, a subset of:

- part-of-speech tagging
- word sense disambiguation
- machine translation
- dialog systems
- distributional semantics
- text classification

Form of teaching

Combination of lecture, demonstration, and laboratory sessions for assistance with assignments.

Language of instruction: English

Assessment

The examination consists of a combination of take-home programming exercises and projects, written assignments, written and/or oral tests. Obligatory attendance may be required for some course components.

A student who has failed a test twice has the right to change examiner, unless weighty argument can be adduced. The application shall be sent to the board of the department and has to be in writing. The total number of exam sessions is five, when feasible.

Completion of examined student achievement is admitted

Grades

The grading scale comprises: Pass with Distinction (VG), Pass (G) and Fail (U).

For the grade Pass is required:

- completed take-home assignments
- passed written/online tests, if any are assigned

To pass with distinction is required:

- exceptional performance on take-home assignments
- passed written/online tests, if any are assigned

Course evaluation

Students who are currently taking the course or have completed it will be given the opportunity to express their views and share their experiences in an anonymous course evaluation. A compilation of the course evaluation and the course coordinator's reflections on it will be made available to the students within reasonable time after the end of the course. The next time the course is taught the compilation and any measures based on it will be presented to the students.