LT2124   Themes in NLP and language technology, 7.5 credits

Teman i datalingvistik och språkteknologi, 7,5 högskolepoäng

Second Cycle

Confirmation

This course syllabus was confirmed by Department of Philosophy, Linguistics and Theory of Science on 2020-11-23 and was last revised on 2023-05-29 to be valid from 2023-08-28, autumn semester of 2023.

Field of education: Science 100%
Department: Department of Philosophy, Linguistics and Theory of Science

Position in the educational system

Can 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	*Specialization*
Language Technology	A1N, Second cycle, has only first-cycle course/s as entry requirements

Entry requirements

Students with an undergraduate degree (at least three year full-time study) in

- language technology, computational linguistics or computer science;
- linguistics (with at least 30 hec, corresponding to half a year full-time study, in formal linguistics);
- adjacent subjects, eg. cognitive science, languages, philosophy or mathematics can also be considered, provided that the student can show a background in either programming or formal linguistics corresponding to 30 hec, half a year full-time study;
- or a certificate from the department that your qualifications are equivalent.

English 6 or equivalent is also required.

Learning outcomes

Knowledge and understanding
- account for pros and cons of different NLP approaches,
- describe common parsing algorithms,
- account for theories behind different data driven methods,
- explain differences between methods and applications in NLP,

Competence and skills
- apply current methods for basic NLP tasks,
- use miscellaneous software tools to solve NLP problems together with annotated corpora,
- write simple programs that apply NLP tools and software libraries,
- perform work according to a predetermined schedule,

Judgement and approach
- choose between different techniques and data representations to solve a specific NLP task,
- select existing applications and techniques in order to apply them to a new problem,
- choose appropriate features for evaluating a possibly solved NLP problem.

Course content

The course gives a high-level of view of applications and techniques 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, another covering the processing of textual data, and another giving an overview of common NLP tasks.

1. Basic concepts:
   - supervision
   - evaluation
   - symbolic vs. statistical processing
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 contemporary NLP tasks, including, as time permits:

- part-of-speech tagging
- word sense disambiguation
- machine translation
- distributional semantics
- text classification
- image captioning

Form of teaching

The teaching is given in the form of lectures, laboratories, assignments, seminars, exercises, individual work, or group work.

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.

The grading teacher may request completion of examined student achievements.

A student who has taken two exams in a course or part of a course without obtaining a pass grade is entitled to the nomination of another examiner. The student needs to contact the department for a new examiner, preferably in writing, and this should be approved by the department unless there are special reasons to the contrary (Chapter 6 Section 22 of the Higher Education Ordinance).

If a student has received a recommendation from the University of Gothenburg for special educational support, where it is compatible with the learning outcomes of the course and provided that no unreasonable resources are required, the examiner may decide to allow the student to sit an adjusted exam or alternative form of assessment.

In the event that a course has ceased or undergone major changes, students are to be guaranteed at least three examination sessions (including the ordinary examination session) over a period of at least one year, but no more than two years, after the course has ceased/been changed. The same applies to placements and professional placements (VFU), although this is restricted to just one additional examination session.

Grades

4/4

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.

Additional information

This 7.5 credit course is ideally intended to be taken alongside the 7.5 credit LT2123 in the H2LTG program and together supplant the 15-credit course LT2003. It is also intended to supplant LT2114.

The course requires access to a computer (or similar) with internet access.