# Text Mining

## Single subject and programme course

## 6 credits

## Text Mining

## 732A92

## Valid from: 2017 Autumn semester

**Determined by**

The Quality Board at the Faculty of Arts
and Sciences

**Date determined**
2016-04-13

**Revision date**
2016-09-08

**Offered for the last time**
Autumn semester 2022

**Replaced by**
732A81

## Main field of study

Statistics

## Course level

Second cycle

## Advancement level

A1N

## Course offered for

- Masters Programme in Statistics and Machine Learning

## Entry requirements

- Bachelor's degree equivalent to a Swedish Kandidatexamen of 180 ECTS credits including an in-depth academic paper 15 ECTS credits in
    - statistics
    - mathematics
    - applied mathematics
    - computer science
    - engineering
- Passed courses in
    - calculus
    - linear algebra
    - statistics
    - programming
- English corresponding to the level of English in Swedish upper secondary education (Engelska 6/B)
  (Exemption from Swedish)

## Intended learning outcomes

After completion of the course the student should on an advanced level be able to:

- use basic methods for information extraction and retrieval of textual data,
- apply text processing techniques to prepare documents for statistical modelling,
- apply relevant statistical models for analyzing textual data and correctly interpret the results,
- use statistical models for prediction of textual information,
- evaluate the performance of statistical models for textual data.

## Course content

The course presents how textual data can be retrieved, linguistically pre-processed and subsequently analyzed quantitatively using formal statistical methods and models. The course brings together expertise from the areas of database methodology, computational linguistics and statistics.

The following topics are covered:
Introduction and overview of quantitative text analysis and its applications; Information extraction; Web crawling; Information retrieval; Tf-idf; Vector space models; Text preprocessing; Bag of words; N-grams; Sparsity and smoothing for text; Document classification; Sentiment analysis; Model evaluation; Topic models.

## Teaching and working methods

The teaching comprises lectures, lab exercises and a text mining project. The lectures are devoted to presentations of concepts, and methods. The computer lab exercises are devoted to practical application of text mining tools. In the project work, the student will get hands-on experience in solving a text mining problem. Homework and independent study are a necessary complement to the course. Language of instruction: English.

## Examination

Written report on the Text mining project. Written reports on lab assignments. Detailed information about the examination can be found in the course's study guide.

If the LiU coordinator for students with disabilities has granted a student the right to an adapted examination for a written examination in an examination hall, the student has the right to it. If the coordinator has instead recommended for the student an adapted examination or alternative form of examination, the examiner may grant this if the examiner assesses that it is possible, based on consideration of the course objectives.

Students failing an exam covering either the entire course or part of the course twice are entitled to have a new examiner appointed for the reexamination.

Students who have passed an examination may not retake it in order to improve their grades.

## Grades

ECTS, EC

## Other information

Planning and implementation of a course must take its starting point in the wording of the syllabus. The course evaluation included in each course must therefore take up the question how well the course agrees with the syllabus.

The course is carried out in such a way that both men´s and women´s experience and knowledge is made visible and developed.

# Department

Institutionen för datavetenskap