

SYLLABUS

Text Mining 7.5 credits

D7058E

Text Mining

Course syllabus admitted: Autumn 2023 Sp 1 - Present

DECISION DATE
2022-08-22

Text Mining 7.5 credits D7058E

Text Mining

Second cycle, D7058E

Education level	Grade scale	Subject	Subject group (SCB)
Second cycle	U G VG	Informationsteknik	Computer Technology

Main field of study

Information Systems Sciences, Computer Science and Engineering

Entry requirements

In order to meet the general entry requirements for the course, you must have accomplished a minimum of 180 ECTS of university studies, out of which 60 ECTS in the areas of computer or system science. The studies shall have included Introductory Programming (for example D0009E or D0007N) and Fundamentals of Databases (for example D0004N or D0018E). The Text Mining course also requires a basic data mining course (for example D0025E Data Mining).

Good knowledge in English equivalent to English 6.

More information about the English language requirements [<http://www.ltu.se/edu/bli-student/Application-process/English-language-requirements-1.109316?l=en>].

Selection

The selection is based on 30-285 credits

Course Aim

The objective of the course is for the student to develop their knowledge and skills in Text Mining. After passing the course, the student should be able to:

- [1]. Explain and use text preprocessing techniques
- [2]. Describe a text analytics system together with its components, optional and mandatory ones
- [3]. Explain how text could be analyzed
- [4]. Evaluate results of text analytics
- [5]. Analyze and reflect on the various techniques used in text analytics and the parameters needed as well as the problem solved
- [6]. Plan & execute a text analytics experiment

Contents

The Text Mining course is focusing on the importance and the difficulty of analyzing text. The Text Mining course is designed to provide students with knowledge relevant to both preprocessing of text as well as analytics of text. The Text Mining course, however, focuses on wide range of algorithms, techniques, and tools. These include standard methods, such as: tokenization, TF-IDF, n-grams, Named Entity Extraction, Sentiment Analysis, and Topic Modeling. Furthermore, recent trends in machine learning and deep learning are also covered, including: Word2Vec, Semantic Hashing, and Recurrent Neural Networks for Natural Language Processing. Various examples and use cases are used across the course.

Realization

Each course occasion's language and form is stated and appear on the course page on Luleå University of Technology's website.

Lectures, labs, assignments, case studies and/or project work. During the course, the students work with individual tasks and/or group work. Some assignments or case studies in the course might contain work in contact with or about the industry. The student uses different methods and techniques, and it is important to choose the right method, technique or computer support for each task. Before and after the tasks are solved, there are lectures to present and discuss different solutions.

Teaching is in English and on the Internet for distance students or on campus for students living here. IT support: Learning management system, e-mail and phone. The learning management system is used for delivering course material, information and submissions. Knowledge is shared and created within the course through virtual meetings with teachers and other students for discussions, supervision, teamwork and seminars. For students on campus, there will be meetings on campus.

Examination

If there is a decision on special educational support, in accordance with the Guideline Student's rights and obligations at Luleå University of Technology, an adapted or alternative form of examination can be provided.

Through written tests, individual and group/project assignment, different student abilities are examined. Those are: the ability to explain and use text mining techniques and the ability to address business problems using text mining individually and in groups.

Unauthorized aids during exams and assessments

If a student, by using unauthorized aids, tries to mislead during an exam or when a study performance is to be assessed, disciplinary measures may be taken. The term "unauthorized aids" refers to aids that the teacher has not previously specified as permissible aids and that may assist in solving the examination task. This means that all aids not specified as permissible are prohibited. The Swedish version has interpretative precedence in the event of a conflict.

Remarks

Technical Requirements: access to computer with administrative rights, web camera, microphone and Internet connection.

Course offered by

Department of Computer Science, Electrical and Space Engineering

Modules

Code	Description	Grade scale	Cr	Status	From period	Title
0002	Individual task	U G#	1.5	Mandatory	A20	
0003	Group-/Project work	U G#	2	Mandatory	A20	
0004	Written exam/Individual exam	U G VG	4	Mandatory	A21	

Study guidance

Study guidance for the course is to be found in our learning platform Canvas before the course starts. Students applying for single subject courses get more information in the Welcome letter. You will find the learning platform via My LTU.

Last revised

by Jonny Johansson, HUL SRT 2022-08-22

Syllabus established

by Jonny Johansson, HUL SRT 2020-02-21