Linnæus University

Jnr: 2020/1241-3.1.2.2

Course syllabus

Faculty of Technology
Department of Computer Science and Media Technology

4DV808 Datalogisk och Visuell Textanalys, 5 högskolepoäng
Computational and Visual Text Analysis, 5 credits

Main field of study
Computer Science

Subject Group
Informatics/Computer and Systems Sciences

Level of classification
Second Level

Progression
A1N

Date of Ratification
Approved by Faculty of Technology 2020-03-30
The course syllabus is valid from spring semester 2021

Prerequisites

- 90 credits in Computer Science (including a degree project at Bachelor level).
- English B/English 6 or the equivalent.

Objectives

After completing the course the student shall be able to:
Knowledge and understanding

- A.1 Define and explain the standard approaches for preprocessing and modeling text data,
- A.2 describe the main classes of existing natural language processing and text mining approaches,
- A.3 discuss the design space for text visualization techniques, and
- A.4 establish the relation between computational and visual methods for text data analysis.

Skills and abilities

- B.1 Implement software for automatic preprocessing and computational analysis of text data, and
- B.2 design and develop custom solutions for visual text analysis.

Judgement and approach

- C.1 Compare and criticize the discussed computational and visual analysis

approaches in the light of current theories and research, and
- C.2 make well-grounded design choices in the context of various tasks and data constraints.

Content

This course provides an introduction to the variety of analytical methods for text data. Texts surround us in our professional and daily lives in form of written communication, document collections, social media streams, etc. Computer science methods for text analytics can thus be useful for scientific and engineering tasks, including domain applications for literature, social media, or medical text data, for instance. This course combines two perspectives: computational (i.e., natural language processing) and visual (i.e., information visualization for raw and derived text data) to support various analytical tasks, e.g., topic analysis, opinion mining, and named entity recognition.

The following topics are covered:

- Definitions and standard models for text data
- Standard text preprocessing methods (tokenization, stop-word filtering)
- Vector representations and transformations for text data (BoW, document-term matrix, TF-IDF, word2vec)
- Overview of traditional and modern computational text analysis tasks and methods
- Overview of tasks and design options for text visualization techniques
- Interaction between computational and visual methods for applied analytical tasks (sentiment analysis, topic modeling, named entity recognition)
- Domain applications of text visualization and visual text analysis
- Open challenges in visual text analysis
- Overview of software tools and libraries for computational and visual text analysis

Type of Instruction

The instruction consists of lectures, seminars, and teacher-supervised laboratory sessions.

Examination

The examination of the course is divided as following:

Code	Appellation	Grade	Credits
2101	Programming project	AF	2.00
2102	Oral presentation	AF	1.00
2103	Oral exam	AF	2.00

The course is assessed with the grades A, B, C, D, E, Fx or F.
The grade A constitutes the highest grade on the scale and the remaining grades follow in descending order where the grade E is the lowest grade on the scale that will result in a pass. The grade F means that the student's performance is assessed as fail (i.e. received the grade F).

Assessment of student performance is made through a programming project, an oral presentation, and an oral exam. Repeat examination is offered in accordance with Local regulations for courses and examination at the first- and second-cycle level at Linnaeus University.

To pass the course, grade E or higher is required for all parts. The final grade is decided from: Programming project (40%) and Oral presentation (20%), and Oral exam (40%).

If the university has decided that a student is entitled to special pedagogical support due to a disability, the examiner has the right to give a customized exam or to have the student conduct the exam in an alternative way.

Objectives achievement

The examination parts are linked to the learning outcomes as follows:

Goal	2101	2102	2103
A.1		✓	✓
A.2	✓		✓
A.3			✓
A.4	✓		✓
B.1	✓		
B.2	✓		
C.1	✓	✓	✓
C.2	✓		✓

Course Evaluation

During the implementation of the course or in close conjunction with the course, a course evaluation is to be carried out. Results and analysis of the course evaluation are to be promptly presented as feedback to the students who have completed the course. Students who participate during the next course instance receive feedback at the start of the course. The course evaluation is to be carried out anonymously.

Other

Grade criteria for the A–F scale are communicated to the student through a special document. The student is to be informed about the grade criteria for the course by the start of the course at the latest.

The course is conducted in such a way that the course participants' experiences and knowledge are made visible and developed. This means, for example, that we have an inclusive approach and strive for no one to feel excluded. This can be expressed in different ways in a course, for example by using the gender neutral example.

Required Reading and Additional Study Material

Aggarwal, Charu C., Machine Learning for Text, Springer, latest edition. Pages: 200 of 452.

Cao, Nan, and Cui, Weiwei, Introduction to Text Visualization, Atlantis Press, latest edition. Pages: 114 of 114.

Compendium of scientific articles.