ID2211 Data Mining, Basic Course 7.5 credits

Datautvinning, grundkurs

This is a translation of the Swedish, legally binding, course syllabus.

If the course is discontinued, students may request to be examined during the following two academic years

Establishment

Course syllabus for ID2211 valid from Spring 2019

Grading scale

A, B, C, D, E, FX, F

Education cycle

Second cycle

Main field of study

Computer Science and Engineering

Specific prerequisites

Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

Intended learning outcomes

The course studies fundamentals of data mining including Information Network Analysis and mining as well as the basic techniques for mining and analyzing text data in natural languages.

In particular the course will cover the basics of graph theory, network structure and link analysis and as well as basics of Mining and Analytics of texts in natural language.

After this course, students will be able to mine and analyse information networks and natural language texts. In particular, the student will be able to

- summarize and describe the fundamental concepts of graph theory and apply them in practice for graph analysis
- summarize and describe the fundamental principles of natural language analysis and apply them in practice for mining texts
- elaborate on and apply algorithms for massive linked data problems (e.g. graph clustering, community detection etc.).

Course contents

- Basic Definitions of Graph Theory, Strong and weak Ties, Degree Distributions and Clustering Measures.
- Erdos-Renyi, Wats-Strogatz, Configuration Model, The Small-World Effect.
- Random walks on Graphs, Page Rank.
- Cascading Behaviour, Epidemics.
- Label Propagation, Link Prediction.
- Distributional Semantics, Word Embeddings, Sentiment Analysis.
- Topic Modelling, Document summarization, Text Segmentation Learning.

Course literature

The contents of the course are derived from the following textbooks as well as from number of reserach papers:

- John Hopcroft and Ravindran Kannan ” Foundations of Data Science” (2013).
- David Easley and Jon Kleinberg “Networks, Crowds, and Markets: Reasoning About a Highly Connected World” (2010).
- A. Rajaraman and J. D. Ullman, Mining of massive datasets. Cambridge University Press, 2012 (alternative: J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, 3-rd Ed., Morgan Kaufmann, 2012).

Examination

- PRO1 - Project, 3.0 credits, grading scale: P, F

- TEN1 - Examination, 4.5 credits, grading scale: A, B, C, D, E, FX, F

Based on recommendation from KTH's coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

Ethical approach

- All members of a group are responsible for the group's work.
- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.