

## Course memo

*Marcel Bollmann*

### Introduction

This document contains general information about the course *Text Mining*. The course is co-taught as **732A81** and **TDDE16**; unless a course code is specifically mentioned in the text, the information here applies to both course codes equally.

### Course content

Text Mining develops methods for accessing information in and extracting knowledge from large volumes of text. The overall aim of this course is to provide you with practical experience of the main steps of text mining: information retrieval, processing of text data, modelling and analysis of experimental results. The course ends with an individual project where you work on a self-defined problem.

### Intended learning outcomes

On completion of the course, you should be able to:

1. implement text mining methods and apply them to practical problems
2. analyse and summarise results from text mining experiments
3. identify, formulate and solve problems within the area of text mining
4. clearly present and discuss the conclusions of a project work

For each intended learning outcome, there is a set of more specific knowledge requirements that express what you need to demonstrate in order to attain a particular grade. These knowledge requirements are listed under *Examination* below.

## Teaching and working methods

The course is taught in the form of lectures, lab sessions, and supervision in connection with an individual project. You are also expected to study independently, both individually and in groups. When you plan your time for the course, you should calculate approximately:

- **40 hours** to prepare for, attend, and follow up on the **lectures**.
- **40 hours** to prepare for, carry out, and follow up on the **labs**.
- **80 hours** to plan, carry out, and document the **project**.

You are entitled to individual project supervision during the study period you are registered for.

## Course literature

The reading for this course consists of excerpts from the following books, which are available to read for free online, as well as selected research articles.

- Jacob Eisenstein. *Introduction to Natural Language Processing*. Manuscript, November 2018.
- Dan Jurafsky and James H. Martin. *Speech and Language Processing*. Draft chapters of 3rd edition, January 2023.
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press. 2009. (Online version)

## Examination

The course has the following examination modules:

**LAB1** Laboratory work, 3 credits (*U, G*)

**PRA1** Project, 3 credits ( **732A81** *Fx, E, D, C, B, A*; **TDDE16** *U, 3, 4, 5*)

To pass the course, you must have a passing grade on both modules (*G* on the lab module; on the project module: **732A81** *E* or higher / **TDDE16** *3* or higher). Your final grade is your grade on the project module (PRA1).

## LAB1: Laboratory work

The lab module tests to what extent you are able to (a) implement text mining methods and apply them to practical problems (learning outcome 1), and (b) analyse and summarise the results of text mining experiments (learning outcome 2).

*Knowledge requirements* You implement the text mining methods that are covered in the course and apply them to pre-defined, small-scale problems according to given instructions. You analyse experimental results with given evaluation methods and summarise them with simple judgements.

*Form of examination* This module is examined by a series of **lab assignments**, or ‘labs’, which are done in pairs; as well as **written reflections** on these assignments, which are done individually. Each lab assignment is examined based on a notebook with your code (one submission per lab group), while the reflection consists of brief, written answers to one or more questions related to the lab assignments (one submission per student).

*Grade requirements* For a passing grade (G), you need a *Pass* on all labs (i.e., notebooks and reflection questions).

*Examination date* The formal examination date for this module is the last ordinary examination date for the course, **2025-01-18**.

**Each lab assignment has two due dates:** The first due date is in the week after the assignment was released; the second due date is the formal examination date for the course session, i.e. 2025-01-18.



Why should I try to meet the first due date?

If you meet the first due date, you will get formative feedback on your assignment and the chance to revise it before its final assessment after the examination deadline. If you do *not* submit by the first due date, we will only assess your submission after the second due date has passed!

*Feedback* To get feedback about how well you meet the knowledge requirements for this module, you can attend the tutored lab sessions and submit your assignments in time for the first due date. You can also get feedback from the examiner.

## PRA1: Project

The project module primarily tests to what extent you are able to (a) identify, formulate and solve problems within the area of text mining (learning outcome 3), and (b) clearly present and discuss the conclusions of a project work (learning outcome 4).

*Knowledge requirements* You identify and formulate a *substantial* text mining problem with some help from a teacher. In working on your problem, you implement and apply suitable text mining methods, analyse experimental results with appropriate evaluation methods and summarise them with well-developed judgements. You clearly present and discuss the conclusions of your work.

*Form of examination* This module is examined by an **individual written project report**. The report is assessed based on the following criteria: the overall substance of the work, the suitability and correctness of the technical approach, the soundness of the evaluation methods and analysis, as well as the overall clarity and presentation of the report.

*Grade requirements* Detailed grading criteria are made available in a separate document on the course website.

*Examination date* The formal examination date for this module is the last ordinary examination date for the course, **2025-01-18**.

*Feedback* To get feedback about how well you meet the knowledge requirements for this module, you can book individual supervision sessions with the examiner.

## Cheating and plagiarism

Each piece of work you present for examination must be entirely your own. You are not allowed to give or receive aid on an assignment unless such collaboration is explicitly permitted. The use of prohibited aids is cheating.

The following, in particular, is not allowed:

- copying code from other lab groups, or letting other lab groups copy your code
- making lab solutions available via public channels, such as GitHub

When using external sources (such as text or code) in work that you present for examination, you must appropriately acknowledge these sources. This rule also applies to materials obtained from the internet. Failure to acknowledge your sources is plagiarism.

*Use of generative AI* For using generative AI tools including (but not limited to) Copilot or ChatGPT, the following rules apply:

- You take responsibility for each piece of work you present for examination. This means that *you* are responsible for ensuring that your work does not contain plagiarised code or text, and acknowledges any sources appropriately and correctly. We therefore strongly discourage using AI to fully automate the generation of code or text for your report – besides hurting your own learning, it may also result in plagiarism.
- Any use of generative AI must be declared when handing in a lab assignment or the project report.

If you are unsure about what you should or shouldn't use generative AI for, you are always welcome to contact the examiner.



Consequences of cheating and plagiarism

We **must** report suspected cheating and plagiarism cases to the University Disciplinary Board. No exceptions exist to this rule, no matter how slight the potential offence.

## Additional examination

For each module, there are two additional examination opportunities during the year following the course session:

- during the re-exam period for HT2 (last day: **2025-03-20**)
- during the re-exam period *before* the summer break (last day: **2025-06-12**)

After this, you can still be examined in connection with the next course session. Note however that the next session may feature different content, different assignments and different examination requirements.

To take an additional examination in the LAB1 module, you must submit (or re-submit) all assignments (i.e., labs and/or reflections) that you do not already have a *Pass* on. To

take an additional examination in the PRA1 module, you must submit (or re-submit) your project report. The same instructions apply as for the ordinary examination. The submission deadline is always the last day of the relevant exam period.

If you are not registered for the most recent course session, you need to contact the examiner before submitting your labs. The last day to contact the examiner is ten (10) working days before the relevant examination date.

## Other course policies

### Feedback policy

*What you can expect from us* We try our best to give you prompt, constructive, and meaningful feedback on how well you meet the knowledge requirements set out for the course. Our focus is on *formative feedback*, which you can use to improve your learning (and we can use to improve our teaching!) while the course is ongoing.

*What we expect from you* We expect you to familiarise yourself with the knowledge requirements set out for the course and to actively seek our feedback on how well you meet these requirements. We also expect you to reflect on the feedback we provide and grasp opportunities to put it to good use.

### Communication policy

*What you can expect from us* When you contact us via email or Teams chat, you can expect an answer during standard working hours, 9–17. (Do not expect a response to email/chat in the evening or weekend.) For more personal contact, you can talk to the examiner in class or book an appointment. The link for appointment booking can be found on the course website.

*What we expect from you* The course website and the documents linked there are the primary sources of information about the course, and we expect you to keep yourself up-to-date with what we publish there. We also send out information via the University's email list for the course, and we expect you to check for and read this information while the course is ongoing.

## Special needs

*Accessibility* If there is any portion of the course that is not accessible to you due to challenges with technology or the course format, please let the examiner know so we can make appropriate accommodations.

*Students with disabilities* If you have a documented disability, you should contact the examiner as soon as possible regarding accommodations.