Information about the project report

Marco Kuhlmann

Introduction

The project module of the course 732A92 / TDDE16 Text Mining is examined based on a written report. The knowledge requirements for the module are as follows:

You apply the text mining techniques covered in the course to self-defined problems. You evaluate the performance of methods and systems with appropriate validation methods and interpret your results with well-developed judgements.

A *problem* can either be a specific task ('build a system for sentiment classification of movie reviews') or the answering of a limited-scale research question ('investigate which text segmentation strategy yields the most coherent topic models'). *Well-developed judgements* are judgements that, among others, are based on a substantial amount of relevant material (such as experiments or theoretical analysis), contain nuanced arguments (rather than unsubstantiated opinion), and are clearly written.

Suggested structure

The suggested structure for the report is the following:

1. Introduction. Introduce the task or research question that you have addressed in your project. What were you trying to do? Why did you choose this project?

2. Theory. Present relevant theoretical background, and in particular the models that you have used. Where appropriate, use mathematical formulas.

3. Data. Present your data. What information does it contain? Where did you get it from? What preprocessing did you do, if any?

4. Method. Explain how you carried out your study. Aim to be detailed enough for others to reproduce your results.

5. Results. Present your results in an objective way. Use tables and charts, but do not forget to also include a summary in text form. Do not interpret your results.

6. Discussion. Analyse your results and discuss the possibilities and limitations of your technical approach. Compare your study to related work.

7. Conclusion. Based on your results and their analysis, what new knowledge do you take away from your project?

Formal requirements

Your report must meet the following formal requirements. If you fail to comply with these, I will ask you to revise your report and submit it again in connection with the next examination date.

Title page The first page of your report must contain the title of your project as well as your full name, LiU-ID, and course code (732A92 or TDDE16). Do not use a generic title such as 'Text Mining Project Report'; instead, find a title that expresses what you did in your project, such as 'Predicting Personality Types from Written Text'.

Abstract Every report must include an abstract. This abstract must be one paragraph, and may not exceed 250 words. The abstract should provide a concise summary of your project's purpose, method, and results. Note that the abstracts of all projects will be published on the course website.

Length Your report should contain between 2,000 and 4,000 words, corresponding to approximately 4–8 pages of single-spaced text. This guidance excludes the title and abstract, as well as any non-textual elements such as figures, tables, mathematical formulas. Even references are excluded from this guidance.

References When using ideas, code, or text from others, you must appropriately cite your sources. This also applies to materials obtained from the internet. When formatting references, choose whatever style you are comfortable with, but be consistent.[1]

Proofreading Before submitting your report, you should carefully proofread your text and check it for errors in grammar and spelling. If you need help with your writing, consider contacting Academic English Support (*Språkverkstaden*).

[1] For an example style, see Section V of the IEEE Editorial Style Manual.

Assessment criteria

When grading your report, I will assess it with respect to the criteria defined below. For each criterion, I will assign one of the scores 0 (below expectation), 3 (meets expectation), or 5 (exceeds expectation). The descriptors for each score hopefully give you a good idea of what I am looking for in a report. To get a passing grade, your score *for each criterion* must be at least 3. Your grade is determined by the total score, as specified in the following table:

Total score	15	17	19	21	23
Grade 732A92	E	D	C	B	A
Grade TDDE16	3	3	4	5	5

Clarity Is it clear what was done in this project, why it was done, and how it was done? Is the report well-written and well-structured?

0 There are some important questions about the method, results, or analysis that even expert readers are not able to resolve.

3 Any student who has successfully completed the course should understand what was done in this project, why it was done, and how it was done.

5 The report is well-polished. In regard to clarity of presentation, it would be acceptable for an academic journal or conference proceedings.

Soundness and correctness Is the technical approach sound and well-chosen? Are the claims made in the report supported by proper experiments, and are the results of these experiments correctly interpreted?

0 Troublesome. There may be some ideas worth salvaging here, but the work should really have been done or evaluated differently. The claims made in the report have no support in the experimental results.

3 Fairly reasonable work. The approach is not bad, the methods are appropriate, and at least the main claims are probably correct. The report contains a discussion of the possibilities and limitations of the technical approach.

5 The approach is very apt, and the claims are convincingly supported. The report contains a well-developed discussion of the possibilities and limitations of the technical approach, including the reliability and validity of the results.

Related work Does the report show awareness and understanding of related work documented in scientific sources? Is it clear where the work done in the project sits with respect to that related work?

0 The report shows little awareness and understanding of related work. References to scientific sources are missing or incomplete. There is no account of how the work done in the project compares to the related work.

3 The report shows some awareness and understanding of related work. Scientific sources are adequately referenced. The relation between the work done in the project and the work documented in the scientific sources is clear.

5 The report features a precise and enlightening comparison with related work. References are complete and consistently formatted. The majority of scientific sources are peer-reviewed research articles.

Creativeness How creative is the project? For example: Does the project target a new problem? Does it contribute a new data set? Does it use any machine learning models that were not covered in the course?

0 There are no or very few creative elements in this project. The project is essentially a repetition of one of the lab assignments.

3 The project contains at least one creative element.

5 There are many creative elements in this project. The project goes significantly beyond what has been covered in the course.

Substance Based on the report, does this project have enough substance, or would there have been room for more ideas, results, or analysis? (The expected amount of work for the project module is 88 hours.)

0 Seems thin. I (the examiner) would have expected significantly more ideas, results, or analysis for a project with this timeframe.

3 Represents an appropriate amount of work for a project in this course.

5 Contains significantly more ideas, experiments, and analysis than what I (the examiner) would have expected for a project with this timeframe.

Together with your grade, you will receive a short written motivation based on the descriptors for each criterion.