# PWM-CS-HS3: Computational Social Science III: Computational Text Analysis for Social Science: Term Paper Requirements

Sergei Pashakhin

February 5, 2024

## Deadline

- In Winter Semester 23/34, the hard deadline is March 31.
- Please be mindful that I need time to evaluate your work. This can impact the timing of getting your grade registered.
- My advice is to submit your work no later than March 24.

# Grading Criteria

Your paper will be evaluated in the following form.

Criterion	Notes	Evaluation
Basics/Literature		
Structure & Presentation		
Correctness / relevance		
Argumentation		
Reflection of results, limitations		
Innovative / Creative work		

Conclusion : Grade:

For each criterion, there can be one of the following marks in "Evaluation" column:

- '++' full match to the criterion with evidence of a significant effort to exceed the standard: +2 points.
- '+' full match to the criterion with weak evidence of an effort to exceed the standard: +1 point.
- '0' complete match to the criterion with no evidence of effort to exceed the standard.
- '-' the work does not fully satisfy the criterion with one mistake: -1 point.
- '--' the work does not fully satisfy the criterion with more than one mistake: -2 points

This is added up and results in the following grading scale:

Grade	1,0	1,3	1,7	2	2, 3	2,7	3,0	3,3	3,7	4,0
Points	12 - 11	10 - 9	8 - 7	6 - 5	4 - 3	2 - 1	0 - 1	-23	-45	-67

#### Criteria and standard

- *Literature*. How do you solve the problem of finding your literature. How well your literature covers the issue. How connected and/or independent are your sources. Your awareness of how the literature fits together. The standard: the work uses current literature; the work highlights uncertainties and seeks to collaborate them with independent sources; the literature is presented in your own words - only necessary paraphrasing and no plagiarism.
- Structure & Presentation. How well do sections of your paper support each other. Coherence of presentation. The standard: methods, their discussion and implementation are the most important sections of the paper; tables and figures are used in the main body only when they help to communicate your idea.
- Correctness and relevance. The writing stays close to your problem, literature, and data. The standard: writing does not deviate into loosely related topics; all the used concepts are defined in the main body; when necessary you explicitly say what perspective you adopt on an issue and explain why.

- Argumentation. How are your claims supported. Logical issues. How the text anticipates potential criticism and preemptively responds to it. The standard: you show that you are aware of all the limitations we covered during the seminar for each method.
- *Reflection of results, limitations.* Your ability to critically assess yourself and your work. The standard: you have a section titled "limitations."
- Innovative / Creative work. This reflects the degree of 'extra work.' How far you went beyond the ideas and materials we engaged in during the seminar. The standard: you do not do any extra work beyond looking for the relevant literature. An example of extra work: reproducing findings from other papers; validating results one package gives you with an alternative implementation or package; adding techniques we did not cover in the seminar.

## Choosing a topic

We have two types of term paper: project paper and review paper.

#### Project paper

A project paper documents your work on a research question with a dataset.

- 1. Pick a dataset.
- 2. Come up with a research question that is possible to answer with the data and methods discussed during the seminar.
- 3. Apply methods to data.
- 4. Summarize findings.
- 5. Discuss limitations and possibilities to improve the analysis.

#### Where to look for data

- https://datasetsearch.research.google.com/
- Data Is Plural repository by https://www.data-is-plural.com/
- https://osf.io/

- https://dataverse.harvard.edu/ and other instance see https://en.wikipedia.org/wiki/Dataverse?
- https://www.kaggle.com/datasets/
- https://github.com/ search for lists of datasets.

#### **Review** paper

A review paper considers a problem in computational text analysis and available solutions. The goal of a review paper is to compare possible solutions and discuss options for a potential researcher.

For example, a review paper can consider the task of hate speech detection. The paper, then, examines definitions and operationalization of hate-speech. Then, the paper describes the scope of possible solutions. For instance, some languages may have available dictionary resources and pretrained classifiers. Other researchers may propose language-agnostic solutions - i.e. a full-cycle pipeline for hate-speech detection applicable to any language and dependent only on the information available in the data at hand. The review paper describes available solutions, looks into their application in published research, and compares the benefits and limitations of each solution. Finally, it is possible to focus a review paper on a specific language: i.e. "hate-speech detection in German language."

#### Example topics

- "Validity of sentiment analysis: do sentiment analysis methods measure what they intend?"
- "BERTopic, LDA and STM: evaluating the fit of topic modeling algorithms to the most common research designs"
- "Can we do hypothesis testing with text data? The issue of sample representativeness"
- "A review of approaches to automatic frame extraction from news items"

### Format

• The paper length must not exceed 10 pages (excluding references and appendices). R code should be part of the appendices.

- Use APA citation style.
- You can consult these resources for templating and styling: https://vc.uni-bamberg.de/moodle/course/view.php?id=26902
- Submit a PDF file with your writing by email: sergei.pashakhin@unibamberg.de
  - Submit code in plain text format: i.e. \*.R, \*.Rmd, \*.qmd. If you would like to share a compiled notebook mixing together code and output, then submit it in \*.HTML.