UNIVERSITÄT BONN

CAISA Lab
https://caisa-lab.github.io/

b-it Bonn-Aachen International Center for Information Technology

Introduction to Natural Language Processing
Lecturer: Prof. Dr. Lucie Flek
Course Coordinator: Vahid Sadiri Javadi
Exercise Instructor: David Kaczér

Problem Formulation (PF) – Guideline

When formulating a research problem for a project, it is important to consider the following points:

1. Topic (Title):

The general area of interest or field of study that the research project addresses.

2. Motivation:

A clear and concise statement of the problem or issue that the research project aims to address. This involves describing the background of the research topic and the context within which the research problem exists. In this section, you list a set of specific and measurable objectives that the research project aims to achieve. This involves formulating a clear and specific research question that the research aims to answer. The research question should be feasible and realistic and should align with the research objectives.

3. Dataset:

It's important to get a sense of your dataset's basic structure and quality. The data exploration step is crucial for understanding the structure, quality, and nuances of your dataset before building any models. In this section, you explore the characteristics of the text itself. Calculating basic statistics such as average sentence length, word count, and vocabulary size helps you understand the complexity and variability of your text data. Visualization plays a big role in exploring the data. Creating word clouds or frequency distribution plots is a great way to identify the most common words and their relative importance. You can also perform parts-of-speech (POS) tagging to look at the linguistic structure of your text.

4. Methodology (Possible Solutions):

This involves outlining the research methodology that will be used to answer the research question. This includes describing the research design and data modeling methods that will be used. It also helps us if you provide reasons why you use certain techniques as well as the hypothesis of what you expect from those techniques.

5. Expected Results:

It is important to note that expected results are not definitive or certain, and may change as the research progresses. Therefore, it is important to remain open to unexpected results and to be prepared to adapt the research methodology and the research question as needed.

6. Evaluation Metrics:

Defining appropriate evaluation metrics is essential for evaluating the performance of NLP models and demonstrating their effectiveness. It is important to choose metrics that are appropriate for the specific task being investigated such as comparing the proposed model's performance to a set of baseline models or using statistical tests to ensure that the observed improvements are significant.

7. Challenges and Limitations:

This involves identifying the potential limitations of the research, and explaining how they will be addressed. This is important for demonstrating the credibility and validity of the research.

8. Task Assignment:

To ensure effective task assignments, it is important to develop a detailed project plan and to communicate the tasks and expectations clearly to the research team. (Who does what?)