Creating a Synthetic Evaluation Dataset for Serbian SentiWordNet Using Large Language Models

**Abstract. Brief overview of the problem and research objective**

**Summary of the methodology used**

**Key findings**

**Implications of the research**

# **Introduction**

* Background on sentiment analysis and its importance
* Overview of SentiWordNet and its role in sentiment analysis
* Challenges in creating SentiWordNet for under-represented languages like Serbian
* Introduction to the concept of synthetic evaluation datasets
* Objectives of the paper

# Literature Review

* Studies related to SentiWordNet and its applications
* Research on synthetic dataset generation for sentiment analysis
* Few-shot learning and its relevance to sentiment analysis
* Overview of Large Language Models (LLMs) in sentiment analysis
* Gap in research for Serbian SentiWordNet and the need for evaluation datasets

# Methodology

* Overview of the Serbian WordNet and selection of synsets for evaluation
* Detailed description of the few-shot learning approach using LLMs
  + Selection criteria for training examples
  + Configuration of the LLM
* Process of labeling the selected synsets to create the synthetic evaluation dataset
  + Criteria for label assignment
  + Validation process for the assigned labels
* Development of evaluation metrics for assessing the synthetic dataset

# Implementation

* Preparation of the few-shot learning environment
* Selection and preparation of the LLM
* Step-by-step process of creating the synthetic dataset
  + Data processing and cleaning
  + Few-shot learning execution
  + Labeling and validation

# Results

* Performance evaluation of the few-shot learning approach
* Discussion on the reliability and validity of the synthetic evaluation dataset

# Discussion

* Implications of the findings for the development of Serbian SentiWordNet
* Challenges encountered during the research and how they were addressed
* Potential for applying the methodology to other under-represented languages
* Suggestions for improving the dataset generation process

# Conclusion

* Summary of the research findings
* Contribution of the study to the field of sentiment analysis and language resources
* Future research directions

**References**

* List of cited works