

MultiLingual: Multilingual Contextual Summarization for Global Enterprises

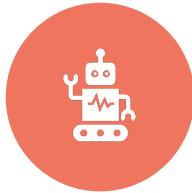
COMP8420 - Adv NLP

Team Members: Muhammad Haris Rizwan & Syed Rafay Ali | Student IDs: 47565284 & 47833920

Date: May 31st, 2024



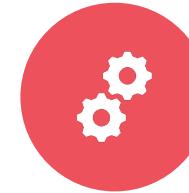
Company Overview and Role



Company: The IT company, MultiLinguaAI, specializes in providing advanced NLP solutions to global enterprises.



NLP Services Offered: These include sentiment analysis, text summarization, named entity recognition, and chatbots etc.



Our Role: As Engineers, we are responsible for the technical development and implementation of a multilingual summarization tool. This includes data collection, preprocessing, model training, evaluation, and integration into the company's systems.

Problem Statement & Objective



PROBLEM STATEMENT: GLOBAL ENTERPRISES NEED ACCURATE, CONTEXT-PRESERVING SUMMARIES FOR DOCUMENTS IN MULTIPLE LANGUAGES.

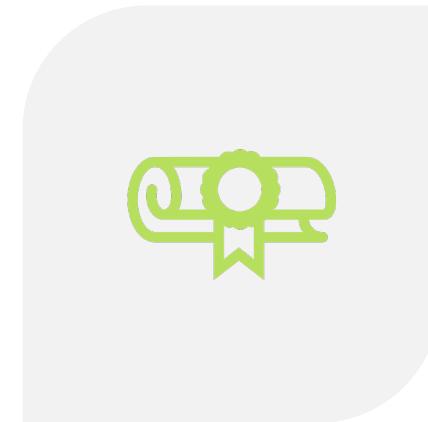


OBJECTIVE: DEVELOP A MULTILINGUAL SUMMARIZATION TOOL THAT MAINTAINS THE CONTEXT AND KEY INFORMATION ACROSS DIFFERENT LANGUAGES.

Project Scope



REAL-WORLD CHALLENGE: HANDLING
AND SUMMARIZING LARGE VOLUMES
OF MULTILINGUAL DOCUMENTS.



TARGET USERS: GLOBAL ENTERPRISES
WITH DIVERSE LINGUISTIC
DOCUMENTATION NEEDS.

Methodology



Data Collection: Multilingual dataset with documents and summaries.



Sources: Public corpora, translated documents.



Pre-processing: Tokenization, normalization, language detection.



Model Training: Using models like mBERT, XLM-R, or multilingual T5.



Evaluation: ROUGE, BLEU, METEOR, human evaluation for context accuracy.

Experimental Results

1

Comparison: Baseline models vs. our model.

2

Key Metrics: ROUGE/BLEU scores, human evaluation feedback.

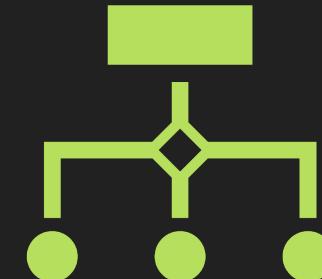
3

Ablation Study: Impact of different components on performance.

Summary



Summary: Developing a tool for summarizing multilingual documents while preserving context.



Future Work: Enhancements and potential real-world integration.

References

1. MLSUM: The Multilingual Summarization Corpus (<https://aclanthology.org/2020.emnlp-main.647/>)
2. XL-Sum: Large-Scale Multilingual Abstractive Summarization (chrome-extension://efaidnbmnnibpcajpcglclefindmkaj/<https://aclanthology.org/2020.emnlp-main.647.pdf>)
3. Contrastive Aligned Joint Learning for Multilingual Summarization (chrome-extension://efaidnbmnnibpcajpcglclefindmkaj/<https://aclanthology.org/2021.findings-acl.242.pdf>)

Q&A

Engagement: Open the floor for any questions from the audience to discuss the project in detail and gather feedback.

