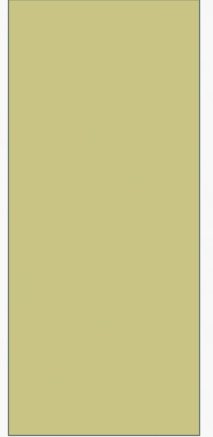


TURKISH TEXT SUMMARIZATION

GÜNDÜZ VEHBI DEMIRCI
EMIR GÜLÜMSER



OUTLINE

- Problem Description
- Methodology
- Motivation / Importance
- Expected Results

PROBLEM DESCRIPTION

- We plan to do summarization method that generates summaries for the texts.
- We plan to rank the sentences according to their importance.
- Finally, display the results to the users.
- We only focus on Turkish texts.

METHODOLOGY

- We consider to use Java programming language.
- First, we would clean the text (stop word list, stemming).
- We plan to evaluate the words for candidate sentences using feature calculation (TF X IDF calculation, discretization).

METHODOLOGY

- We may use training data (not decided).
- In the extraction stage, using training data and feature calculation, candidate words will be extracted.
- Combining the candidate words, we would rank the sentences according to threshold.
- We would evaluate our sentences using ROUGE evaluation technique which is used to compare summarization outputs with human generated summaries.

MOTIVATION / IMPORTANCE

- Due to huge amount of data on the Internet IR becomes more important.
- Text search engines return too much documents for a person to identify which one is relevant.
- Therefore there is need for technologies that help people on this purpose.
- Presenting document summaries will help people to find their desired documents.

EXPECTED RESULTS

- Firstly we expect to see our summaries match with test data.
- But there is possibility to see some of important sentences are not included or vise versa.