

Topic Extraction From Turkish News Articles

Anıl Armağan, Fuat Basık,
Fatih Çalışır, Arif Usta

Agenda

- Introduction
- Project Description
- Motivation
- Methodology
- Expected Results
- Conclusion

Introduction

- Huge volume, variety and velocity of data.
- Hard to find, interesting information from that data.
- Even reading news and finding out what is happening in the world is hard.

Project Description

- Using text mining and information retrieval techniques
- From a given dataset of Turkish news articles
- Extract what happened yesterday, with key terms
- Sort the different events with respect to importance.

Motivation

- Working with Turkish Language has its' own challenges.
 - Agglutinating structure.
- Working with text documents is important to learn information retrieval systems.
- As long as we use text based search engines, text processing will always be important.

Methodology

- Preprocessing
- Stop word removal
- Term Weighting
 - Different approaches will be applied.
 - TF-IDF, TF-PDF with different n-grams
- Find out most important terms, and documents containing these terms.

Expected Results

- We expect to find out
 - A good term weighting method for Turkish
 - Effect of the n-grams for Turkish texts
 - A good tool that gives most important news of yesterday.

Conclusion

- Huge information needs to be summarized.
- Turkish language has its' own challenges to process.
- Summarizing what happened yesterday from Turkish news articles:
 - Beneficial
 - Useful
 - Important

Questions

- Thank you for listening.