CHRONOLOGICAL NEWS SUMMARIZATION

TOLGA ÇEKİÇ

- MOTIVATION
- METHODOLOGY
- TEST ENVIRONMENT

MOTIVATION

- A summary of multiple news about a topic
- Covering a time frame for news
- Preserving the chronological ordering

METHODOLOGY

- Using Keyphrase Extraction algorithm for finding key phrases on multiple documents
- Use keyphrases for summarization
- Apply this to a sliding window approach for preserving chronology

KEYPHRASE EXTRACTION ALGORITHM (KEA)

- A machine learning algorithm
- Identify candidate phrases and determine which are key phrases
- Based on two features

1. TFxIDF
2. First Occurence

KEYPHRASE SUMMARIZATION

- Rank keyphrases by score
- Choose ‘n’ most important keyphrases (n=5)
- Extract sentences that include keyphrases
- Eliminate similar sentences

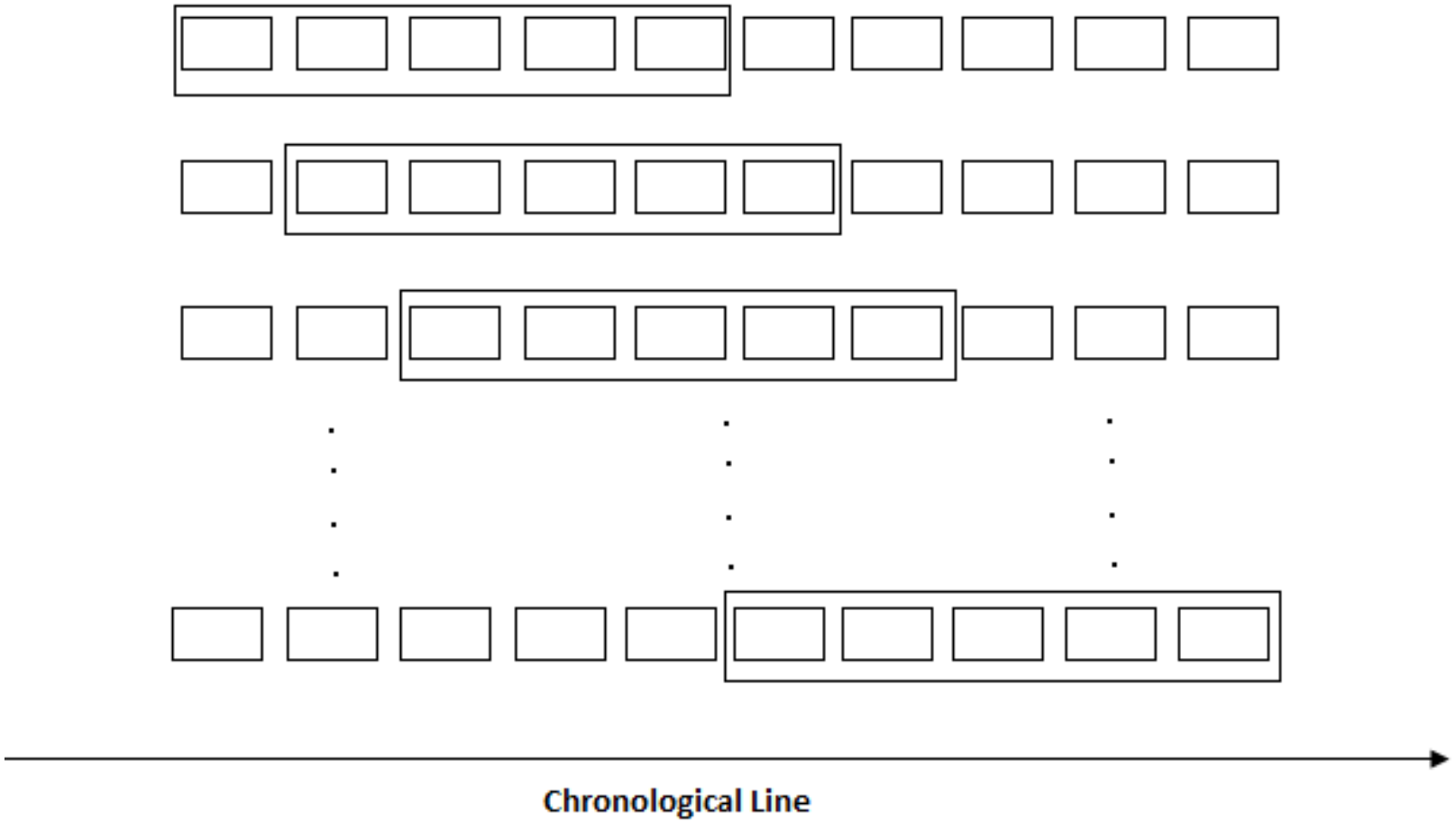
CHRONOLOGICAL SUMMARY

- Some differences from a usual summary
- Need time features on documents
- Possible problem: omission of different but same events

2008: Barack Obama has won the election

2012: Barack Obama has won the election

SLIDING WINDOW APPROACH



TEST ENVIRONMENT

- BilSum 2005 data set

XML format, chronologically listed set of news about a topic

Provides ground truth

TEST ENVIRONMENT

- Zemberek: used for stemming processes
- ROUGE for evaluation

Compares user generated summaries with auto generated summaries, calculates a score

THANK YOU FOR LISTENING