Text Summarization

Text Summarization

Introduction

- based on transformers
- condenses longer texts into shorter, coherent summaries
- preserves its key information and main ideas
- used in
  - news article summarization
  - document summarization

Text Summarization

Types of Text Summarization

Abstractive Summarization

- model creates summary in own words
- rephrases and restructures content for more concise and coherent summary
- requires deeper understanding of input text

Extractive Summarization

- model selects and extracts sentences from input text
- does not generate entirely new sentences
- identifies most important passages

Text Summarization

How does it work?

Preprocessing

- irrelevant information removed from text, e.g. formatting tags, special characters
- tokenization performed

Transformer

- typically variants of BERT, GPT, T5
- Encoder-Decoder Architecture
  - encoder creates fixed-length vector representation of text
  - decoder uses encoding to generate summary
- training data has
  - input texts
  - human-generated summaries
- model learns minimize difference between its summary and reference (true) summary