AASD 4004
Machine Learning - II

Applied AI Solutions Developer Program

Module 07
Text Summarization

Vejey Gandyer

Agenda

Text Summarization

Why

Extractive Summarization

Abstractive Summarization

Text Summarization

What is it?

Text Summarization

Refers to the task of creating a summary of a large piece of text.

Idea: To create a **coherent** summary that captures the key ideas in the text
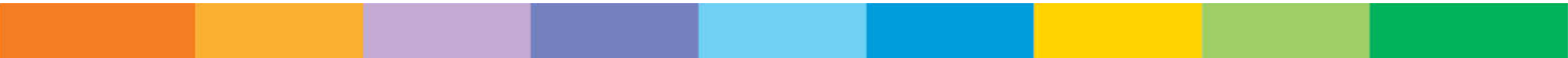
Types

Extractive Vs Abstractive

Query-focused Vs Query-independent

Single-document Vs Multi-document

Why

Why Text Summarization

Create short summaries of longer documents for human readers

Search Engine crawlers can index on short summaries rather than on long text

Extractive Summarization

Extractive Summarization

Selecting **important** sentences from a long piece of text and showing them together as a summary

Abstractive Summarization

Abstractive Summarization

Generating an **abstract** of the text instead of picking sentences from within the text

Summarization using Sumy

```python
from sumy.parsers.html import HtmlParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.text_rank import TextRankSummarizer

url = "https://en.wikipedia.org/wiki/Automatic_summarization"
parser = HtmlParser.from_url(url, Tokenizer("english"))
summarizer = TextRankSummarizer()
for sentence in summarizer(parser.document, 5):
    print(sentence)
```

Summarization using Gensim

```python
from gensim.summarization import summarize
text = "some text you want to summarize"
print(summarize(text))
```

Exercise: Build a Text Summarizer

Building a Text Summarizer

Create a Text Summarizer using

- Sumy

- Gensim

- Summa

Further Reading

Summarization libraries
- Sumy https://github.com/miso-belica/sumy
- Gensim https://radimrehurek.com/gensim/
- Summa https://github.com/summanlp/textrank