

Lab 10: Text Summarization

Generate a summary for story or news article

```
In [1]: import re

In [2]: from collections import Counter

In [3]: def preprocess_text(text):
    text=text.lower()
    text=re.sub(r'^a-zA-Z0-9\s|', '',text)
    text=re.sub(r'\s+', ' ', text).strip()
    return text

In [4]: def tokenize_text(text):
    return text.split()

In [5]: def calculate_tf(tokens):
    tf_counter=Counter(tokens)
    total_tokens=len(tokens)
    tf={token: count/total_tokens for token,count in tf_counter.items()}
    return tf

In [6]: def calculate_idf(documents):
    idf={}
    total_documents=len(documents)
    all_tokens=[token for doc in documents for token in set(doc)]
    token_counter=Counter(all_tokens)
    for token, count in token_counter.items():
        idf[token]=total_documents/count
    return idf

In [7]: def calculate_tfidf(tf,idf):
    tfidf={token:tf_val*idf[token] for token,tf_val in tf.items()}
    return tfidf

In [8]: def generate_summary(text,ratio=0.2):
    text=preprocess_text(text)
    tokens=tokenize_text(text)
    tf=calculate_tf(tokens)
    idf=calculate_idf([tokens])
    tfidf=calculate_tfidf(tf,idf)
    sorted_tokens=sorted(tfidf.items(),key=lambda x: x[1], reverse=True)
    num_tokens=int(len(tokens)*ratio)
    summary_tokens=[token for token,_ in sorted_tokens[:num_tokens]]
    summary=' '.join(summary_tokens)
    return summary

In [9]: def read_document_from_file(file_path):
    with open(file_path,'r',encoding='utf-8') as file:
        document=file.read()
    return document

In [13]: file_path='textfile.txt'
document=read_document_from_file(file_path)

summary=generate_summary(document)
print("Original Document:")
print(document)

print("\nSummary:")
print(summary)
```

Original Document:

Once upon a time, in a land far, far away, there lived a young girl named Alice. Alice was a curious and adventurous girl, and she loved to explore the woods near her home. One day, while Alice was exploring, she came across a strange rabbit hole. She peered down into the hole, but it was too dark to see anything. Alice was curious, so she climbed down the hole.

Alice found herself in a strange and wonderful world. She met all sorts of interesting creatures, including a talking caterpillar, a Cheshire cat, and a Mad Hatter. Alice had many adventures in this new world, but she eventually grew homesick and wanted to return home. Alice found her way back to the rabbit hole and climbed out. She was back in the woods near her home, but she knew that she would never forget her adventures in Wonderland.

Alice ran home and told her parents all about her adventures. They didn't believe her at first, but then Alice showed them the white rabbit's pocket watch that she had taken with her from Wonderland. Her parents were amazed, and they were so happy that Alice had had such a wonderful adventure.

Alice never forgot her adventures in Wonderland. She often told stories about her adventures to her children and grandchildren. And whenever she saw a white rabbit, she would smile and remember her magical journey.

Summary:

she her a alice and in the was to adventures home hole but had rabbit that wonderland far girl curious woods near strange down so climbed found wonderful world all back would never told parents about they white were once upon time land away there lived

In []: