AI Mini Project

-made by Kushagra Gupta

Introduction

With the recent advancements in the tech AI community the 'NEWS ANALYZER project' thoroughly got my attention. On exploring, I found various bits and in-depth knowledge of Python inbuilt libraries and their various functions. Further, I got to know the real-life implementation of these project.

My project performs various natural language processing (NLP) tasks on an article from the CNN website (here taken for example).

TechStack/ Resources Used

The '**newspaper'** library is used to scrape the article and the '**nltk'** and '**rake_nltk'** libraries are used for various NLP tasks such as tokenization, stopword removal, keyword extraction, and named entity recognition. The '**spacy'** library is also used for named entity recognition and visualization, and the '**matplotlib'** and '**seaborn'** libraries are used to plot the results of the named entity analysis.

Firstly, the required '**nltk'** resources are downloaded and the article is scraped from the specified URL using the '**Article'** class from the '**newspaper'** library. The article's metadata (title, authors, and publication date) is then printed to the console.

Next, the article's summary is printed and the sentiment of the article text is analyzed using the '**TextBlob'** class from the '**textblob'** library.

The article text is then tokenized and stopwords are removed using the '**word_tokenize'** and '**stopwords'** functions from '**nltk'**. Keywords are extracted from the article text using the '**Rake'** class from the '**rake_nltk'** library. Named entities are extracted from the article text using the '**nlp'** function from '**spacy'** and visualized using the '**displacy'** function. The frequency of each named entity is counted using the '**Counter'** class and plotted using '**matplotlib'** and '**seaborn'**.

Summary

Overall, My project demonstrates how various NLP libraries and techniques can be used to deeply analyze and visualize the content of a news article.

Some Glimpses of working of Project:

Print the article Summary

```
In [5]: print(f'Summary: {article.summary}')
```

```
Summary: New York CNN Business —TikTok is scrambling to keep its American presence alive, and Microsoft has emerged as a potent
ial, albeit unlikely, savior.
It's clear why TikTok would be interested in the deal: Trump has said he will ban the app if it doesn't find an American buyer
by September 15.
So the TikTok deal would mark a significant push into the consumer space.
As part of the deal, Microsoft committed to ensuring American users' private data is stored in the United States.
"I don't think there will be significant antitrust scrutiny on this deal, because Microsoft doesn't have any footprint in consu
mer social media," Jaluria said.
```

Analyze the sentiment of the article

```
In [6]: analysis = TextBlob(article.text)
        print("The URL text: ",f'Sentiment: {"positive" if analysis.polarity > 0 else "negative" if analysis.polarity < 0 else "neutral"}')
```

```
The URL text:  Sentiment: positive
```

Visualize the named entities

```
In [11]: displacy.render(doc, style='ent', jupyter=True)
```

But there are risks, too.

As part of the deal, Microsoft ORG committed to ensuring American NORP users' private data is stored in the United States GPE . Microsoft ORG and TikTok ORG parent ByteDance ORG will need US GPE officials' sign-off on the deal and may have to make some kind of payment to the US Treasury ORG . And acquiring an app created in China GPE could bring new scrutiny on Microsoft ORG at a time when US GPE - China GPE relations, especially over technology, are at a low point.

Microsoft ORG has largely avoided the recent wave of regulatory scrutiny directed at its Big Tech peers, but that could change if it acquires a controversial social media app created in China GPE with a massive user base. Already, White House ORG trade adviser Peter Navarro PERSON has floated a suggestion that Microsoft ORG be required to "divest its Chinese NORP holdings" in order to be allowed to buy TikTok ORG , he

---

Count the frequency of each named entity

```
In [12]: entity_counts = Counter([entity[1] for entity in entities])
         print(f'Entity Counts: {entity_counts}')
```

```
Entity Counts: Counter({'ORG': 58, 'GPE': 18, 'PERSON': 15, 'DATE': 10, 'NORP': 4, 'MONEY': 4, 'CARDINAL': 1, 'PRODUCT': 1, 'WO
RK_OF_ART': 1})
```

Plot the frequency of named entities

```
In [13]: sns.barplot(x=list(entity_counts.keys()), y=list(entity_counts.values()))
         plt.show()
```