

9913_python_exp_6

March 15, 2024

```
[4]: import nltk
from nltk.tokenize import word_tokenize
from nltk.tag import pos_tag
import re
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')

def count_articles(text):
    tokens = word_tokenize(text)
    tags = pos_tag(tokens)
    article_count = {'a': 0, 'an': 0, 'the': 0}

    for i in range(len(tags)):
        if tags[i][0].lower() in ['a', 'an', 'the'] and tags[i][1] == 'DT':
            article_count[tags[i][0].lower()] += 1

    return article_count

def extract_dates(text):
    # Define regex patterns for different date formats
    date_patterns = [
        r'\d{1,2}/\d{1,2}/(?:\d{4}|\d{2})', # 15/11/2012 or 15/11/12
        r'\d{1,2}(?:st|nd|rd|th)?(?:\n| )?(?:January|February|March|April|May|June|July|August|September|October|November|December)[, ]?', # 15th March 1999 or 15th of March, 1999
        r'(?:(January|February|March|April|May|June|July|August|September|October|November|December)[, ]?)\d{1,2}(?:st|nd|rd|th)?(?:\n| )?(?:\d{4}|\d{2})' # March 15th, 1999
    ]

    dates = []
    for pattern in date_patterns:
        dates.extend(re.findall(pattern, text))

    return dates

text = ""
```

Delhi, is a metropolitan and the capital region of India which includes the
↳national capital city, New Delhi.

It is the second most populous metropolis in India after Mumbai and the largest
↳city in terms of area.

Mumbai, also known as Bombay, is the capital city of the Indian state of
↳Maharashtra.

It is the most populous city in India, and the fourth most populous city in the
↳world, with a total metropolitan area population of approximately 20.5
↳million.

New York is a state in the Northeastern region of the United States. New York
↳is the 27th-most extensive, the 3rd-most populous, and the 7th-most densely
↳populated of the 50 United States.

The Indian Rebellion of 1857 began as a mutiny of sepoys of the East India
↳Company's army on 10 May 1857, in the town of Meerut, and soon escalated
↳into other mutinies and civilian rebellions largely in the upper Gangetic
↳plain and central India, with the major hostilities confined to present-day
↳Uttar Pradesh, Bihar, northern Madhya Pradesh, and the Delhi region.

The Boston Tea Party (referred to in its time simply as "the destruction of the
↳tea" or by other informal names and so named until half a century later,[2])
↳was a political protest by the Sons of Liberty in Boston, a city in the
↳British colony of Massachusetts, against the tax policy of the British
↳government and the East India Company that controlled all the tea imported
↳into the colonies. On December 16, 1773, after officials in Boston refused
↳to return three shiploads of taxed tea to Britain, a group of colonists
↳boarded the ships and destroyed the tea by throwing it into Boston Harbor.
↳The incident remains an iconic event of American history, and other
↳political protests often refer to it

""

```
articles_count = count_articles(text)
dates = extract_dates(text)

print(f"Frequency of Article (a) is {articles_count['a']}, Frequency of Article
↳(an) is {articles_count['an']}, Frequency of Article (the) is
↳{articles_count['the']}")
print("\nDates found in Paragraphs:", ", ".join(dates))
```

Frequency of Article (a) is 8, Frequency of Article (an) is 1, Frequency of
Article (the) is 34

Dates found in Paragraphs: 10 May 1857

```
[nltk_data] Downloading package punkt to C:\Users\Mark
[nltk_data]      Lopes\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]      C:\Users\Mark Lopes\AppData\Roaming\nltk_data...
[nltk_data] Package averaged_perceptron_tagger is already up-to-
```

[nltk_data] date!