
Umang kumar

Captions and Hashtags Generation for any Blog URLs using Python

7th August 2023

OVERVIEW

I have to write the code to generate the Captions and Hashtags of any blog URLs to upload or share on the facebook, LinkedIn, Twitter, or any other social media platform using python.

GOALS

Generate the dynamically Captions and Hashtags for a blog URL.

METHODS

1. First we have to fetch the Blogs URLs when it will requests to us.
2. Now, we have to extract the all contents from the Blog URLs.
3. Split those all extracted contents or sentences into the words.
4. And now generate the Captions from those words.
5. Here we will install nltk package and import them for filter the words and recognize the tokens (Noun, Verb, Obj, etc..)
6. After recognizing the Tokens System will be asked for blog_url and then It will generate the both captions and hashtags for the Blog URLs.

Code

Input

Blog_url = "<https://nec.edu.in/blog/>"

Coding

```
pip install nltk

import requests

from bs4 import BeautifulSoup

import re

def fetch_blog_content(url):

    try:

        response = requests.get(url)

        response.raise_for_status()

        soup = BeautifulSoup(response.content, 'html.parser')

        paragraphs = soup.find_all('p')

        blog_text = ' '.join([re.sub(r'<.+?>', '', str(para)) for para in paragraphs])

        return blog_text

    except requests.exceptions.RequestException as e:

        print(f"Error fetching the blog content: {e}")

        return None

def generate_captions(blog_text, max_chars_per_caption=100):

    words = blog_text.split()

    captions = []

    current_caption = ""

    for word in words:

        if len(current_caption) + len(word) + 1 <= max_chars_per_caption:

            current_caption += " " + word
```

```
        else:

            captions.append(current_caption.strip())

            current_caption = word

    if current_caption:

        captions.append(current_caption.strip())

    return captions

def generate_hashtags(blog_text, max_hashtags=20):

    import nltk

    from nltk.tokenize import word_tokenize

    from nltk.corpus import stopwords

    from nltk.probability import FreqDist

    nltk.download("punkt")

    nltk.download("stopwords")

    words = word_tokenize(blog_text)

    filtered_words = [word.lower() for word in words if word.isalpha() and
word.lower() not in stopwords.words("english")]

    hashtags = [word.strip("#.,!?") for word in words if word.startswith("#")]

    additional_hashtags = [word for word in words if word not in hashtags and
not word.startswith("#")]

    hashtags.extend(additional_hashtags)

    fdist = FreqDist(filtered_words)

    common_words = fdist.most_common(max_hashtags)

    hashtags = [word for word, _ in common_words]

    return hashtags
```

```

blog_url = "https://nec.edu.in/blog/"

blog_content = fetch_blog_content(blog_url)

if blog_content:

    captions = generate_captions(blog_content)

    print("Captions:")

    for i, caption in enumerate(captions):

        print(f"Caption {i+1}: {caption}")

    hashtags = generate_hashtags(blog_content)

    print("\nHashtags:")

    for tag in hashtags:

        print(f"#{tag}")

else:

    print("Failed to fetch blog content.")

```

Output

Captions:

Caption 1: National Engineering College / Blog The award was given by Dr.K.K.Surappa, Vice chancellor, Anna

Caption 2: University and was received by from left NEC NSS Programme officer Mr.J.Thamba, NEC NSS Secretary

Caption 3: Mr.S.Thanga... From Left : Mr.M.Veera Shanthi Ram receiving winning memento from Heike Mock, Director

Caption 4: (DAAD) on 07th April 2018 at Jadavpur Univeristy, Dr.Michael Feiner (Consul General... 2017-2018

Caption 5: Engineering Cut-off List Lakshmi Ammal Memorial 10th All India Hockey Tournament - 2018 was held

Caption 6: during May 17 - 27, 2018 National Engineering College, Kovilpatti NCC units 2/9(TN) SIG. COY. NCC,

Caption 7: Tirunelveli and 3(TN) GIRLS BN NCC, Tirunelveli conducted Nominal roll from 07.00 am to 08.00... NEC

Caption 8: is preparing to provide new innovative platform, for its student community including alumni, to

Caption 9: initiate new venture "Startup" in the campus. Ir this context,... The college offers various

Caption 10: engineering and technology courses, many of which are accredited by the National Board of

Caption 11: Accreditation. These courses provide a syllabus that... © National Engineering College 2023 | All

Caption 12: rights reserved.

Hashtags:

#engineering

#national

#college

#nec

#ncc

#left

#nss

#tn

#tirunelveli

#provide

#new

#courses

#blog

#award

#given

#vice

#chancellor

#anna

#university

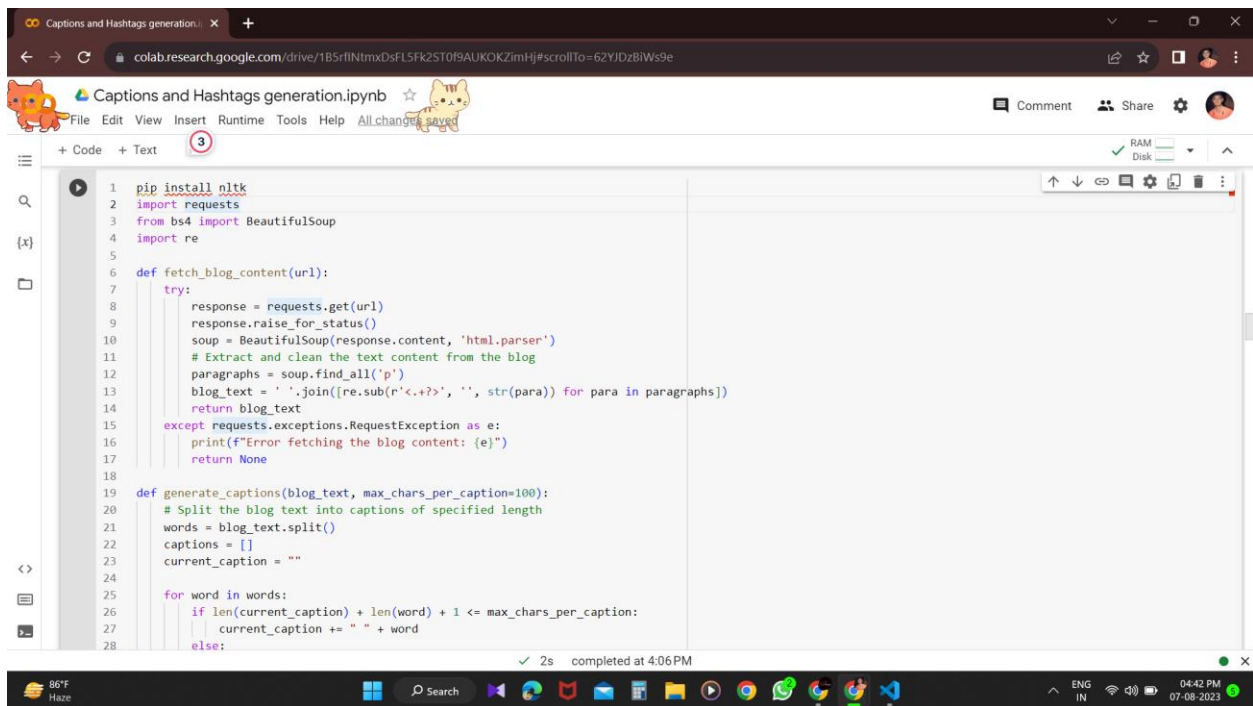
#received

[nltk_data] Downloading package punkt to /root/nltk_data...

[nltk_data] Package punkt is already up-to-date!

[nltk_data] Downloading package stopwords to /root/nltk_data...

[nltk_data] Package stopwords is already up-to-date!



```
1 pip install nltk
2 import requests
3 from bs4 import BeautifulSoup
4 import re
5
6 def fetch_blog_content(url):
7     try:
8         response = requests.get(url)
9         response.raise_for_status()
10        soup = BeautifulSoup(response.content, 'html.parser')
11        # Extract and clean the text content from the blog
12        paragraphs = soup.find_all('p')
13        blog_text = ' '.join([re.sub(r'<.+?>', '', str(para)) for para in paragraphs])
14        return blog_text
15    except requests.exceptions.RequestException as e:
16        print(f"Error fetching the blog content: {e}")
17        return None
18
19 def generate_captions(blog_text, max_chars_per_caption=100):
20     # Split the blog text into captions of specified length
21     words = blog_text.split()
22     captions = []
23     current_caption = ""
24
25     for word in words:
26         if len(current_caption) + len(word) + 1 <= max_chars_per_caption:
27             current_caption += " " + word
28         else:
```

Captions and Hashtags generation.ipynb

```
25 for word in words:
26     if len(current_caption) + len(word) + 1 <= max_chars_per_caption:
27         current_caption += " " + word
28     else:
29         captions.append(current_caption.strip())
30         current_caption = word
31
32 if current_caption:
33     captions.append(current_caption.strip())
34
35 return captions
36
37 def generate_hashtags(blog_text, max_hashtags=20):
38     # Process the blog text to extract relevant keywords for hashtags
39     # Here, we'll use the Python Natural Language Toolkit (NLTK) for keyword extraction
40     import nltk
41     from nltk.tokenize import word_tokenize
42     from nltk.corpus import stopwords
43     from nltk.probability import FreqDist
44
45     nltk.download("punkt")
46     nltk.download("stopwords")
47
48     # Tokenize the blog text into words
49     words = word_tokenize(blog_text)
50
51     # Remove stopwords and non-alphabetic words
52     filtered_words = [word.lower() for word in words if word.isalpha() and word.lower() not in stopwords.words("english")]
53
54     # Calculate the frequency distribution of words
55     fdist = FreqDist(filtered_words)
56
57     # Get the most common words as potential hashtags
58     common_words = fdist.most_common(max_hashtags)
59
60     # Extract the words for hashtags
61     hashtags = [word for word, _ in common_words]
62
63     return hashtags
64
65 # Example usage
66 blog_url = "https://nec.edu.in/blog/"
67 blog_content = fetch_blog_content(blog_url)
68
69 if blog_content:
70     captions = generate_captions(blog_content)
71     print("Captions:")
72
73     hashtags = generate_hashtags(blog_content)
74     print("Hashtags:")
75
76     # Print the generated captions and hashtags
77     print(captions)
78     print(hashtags)
```

2s completed at 4:06 PM

Captions and Hashtags generation.ipynb

```
49 words = word_tokenize(blog_text)
50
51 # Remove stopwords and non-alphabetic words
52 filtered_words = [word.lower() for word in words if word.isalpha() and word.lower() not in stopwords.words("english")]
53
54 hashtags = [word.strip("#,!") for word in words if word.startswith("#")]
55 additional_hashtags = [word for word in words if word not in hashtags and not word.startswith("#")]
56 hashtags.extend(additional_hashtags)
57
58 # Calculate the frequency distribution of words
59 fdist = FreqDist(filtered_words)
60
61 # Get the most common words as potential hashtags
62 common_words = fdist.most_common(max_hashtags)
63
64 # Extract the words for hashtags
65 hashtags = [word for word, _ in common_words]
66
67 return hashtags
68
69 # Example usage
70 blog_url = "https://nec.edu.in/blog/"
71 blog_content = fetch_blog_content(blog_url)
72
73 if blog_content:
74     captions = generate_captions(blog_content)
75     print("Captions:")
76
77     hashtags = generate_hashtags(blog_content)
78     print("Hashtags:")
79
80     # Print the generated captions and hashtags
81     print(captions)
82     print(hashtags)
```

2s completed at 4:06 PM

Captions and Hashtags generation.ipynb

```
73
74 if blog_content:
75     captions = generate_captions(blog_content)
76     print("Captions:")
77     for i, caption in enumerate(captions):
78         print(f"Caption {i+1}: {caption}")
79
80     hashtags = generate_hashtags(blog_content)
81     print("\nHashtags:")
82     for tag in hashtags:
83         print(f"# {tag}")
84 else:
85     print("Failed to fetch blog content.")
86
```

Captions:

Caption 1: National Engineering College / Blog The award was given by Dr.K.K.Surappa, Vice chancellor, Anna
Caption 2: University and was received by from left NEC NSS Programme officer Mr.J.Thamba, NEC NSS Secretary
Caption 3: Mr.S.Thanga... From Left : Mr.M.Veera Shanthi Ram receiving winning memento from Heike Mock, Director
Caption 4: (DAAD) on 07th April 2018 at Jadavpur Univeristy, Dr.Michael Feiner (Consul General... 2017-2018
Caption 5: Engineering Cut-off List Lakshmi Ammal Memorial 10th All India Hockey Tournament - 2018 was held
Caption 6: during May 17 - 27, 2018 National Engineering College, Kovilpatti NCC units 2/9(TN) SIG. COV. NCC,
Caption 7: Tirunelveli and 3(TN) GIRLS BN NCC, Tirunelveli conducted Nominal roll from 07.00 am to 08.00... NEC
Caption 8: is preparing to provide new innovative platform, for its student community including alumni, to
Caption 9: initiate new venture "Startup" in the campus. Ir this context,... The college offers various
Caption 10: engineering and technology courses, many of which are accredited by the National Board of
Caption 11: Accreditation. These courses provide a syllabus that... © National Engineering College 2023 | All
Caption 12: rights reserved.

Hashtags:

2s completed at 4:06 PM

Captions and Hashtags generation.ipynb

```
Captions:
Caption 1: National Engineering College / Blog The award was given by Dr.K.K.Surappa, Vice chancellor, Anna
Caption 2: University and was received by from left NEC NSS Programme officer Mr.J.Thamba, NEC NSS Secretary
Caption 3: Mr.S.Thanga... From Left : Mr.M.Veera Shanthi Ram receiving winning memento from Heike Mock, Director
Caption 4: (DAAD) on 07th April 2018 at Jadavpur Univeristy, Dr.Michael Feiner (Consul General... 2017-2018
Caption 5: Engineering Cut-off List Lakshmi Ammal Memorial 10th All India Hockey Tournament - 2018 was held
Caption 6: during May 17 - 27, 2018 National Engineering College, Kovilpatti NCC units 2/9(TN) SIG. COV. NCC,
Caption 7: Tirunelveli and 3(TN) GIRLS BN NCC, Tirunelveli conducted Nominal roll from 07.00 am to 08.00... NEC
Caption 8: is preparing to provide new innovative platform, for its student community including alumni, to
Caption 9: initiate new venture "Startup" in the campus. Ir this context,... The college offers various
Caption 10: engineering and technology courses, many of which are accredited by the National Board of
Caption 11: Accreditation. These courses provide a syllabus that... © National Engineering College 2023 | All
Caption 12: rights reserved.

Hashtags:
#engineering
#national
#college
#nec
#ncc
#left
#nss
#tn
#tirunelveli
#provide
#new
#courses
#blog
#award
#given
```

2s completed at 4:06 PM

The screenshot shows a Jupyter Notebook interface with the following content:

```

Caption 9: initiate new venture "Startup" in the campus. In this context, The college offers various
Caption 10: engineering and technology courses, many of which are accredited by the National Board of
Caption 11: Accreditation. These courses provide a syllabus that... @ National Engineering College 2023 | All
Caption 12: rights reserved.

Hashtags:
#engineering
#national
#college
#nec
#ncc
#left
#nss
#tn
#tirunelveli
#provide
#new
#courses
#blog
#award
#given
#vice
#chancellor
#anna
#university
#received

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
  
```

The interface includes a browser window at the top, a file explorer on the left, and a terminal at the bottom showing the execution of NLTK commands.