

climate_change

Dongyoung Kim

2024-01-02

Contents

1	Possible Variables to Extract	2
2	Parsing Ideas	2
2.1	User name	2
2.1.1	Corresponding HTML codes	2
2.1.2	Idea	2
2.2	Dates and Time	2
2.2.1	Corresponding HTML codes	2
2.2.2	Idea	2
2.2.3	Python Code	2
2.3	Word count and Bytes	3
2.3.1	Corresponding HTML codes	3
2.3.2	Idea	3
2.4	Chapter	3
2.4.1	Corresponding HTML codes	3
2.4.2	Idea	4
2.5	Tags	4
2.5.1	Corresponding HTML codes	4
2.5.2	Idea	4
2.6	Comments	4
2.6.1	Corresponding HTML codes	4
2.6.2	Idea	4
2.7	Actual Text	4
2.7.1	Corresponding HTML codes	4

1 Possible Variables to Extract

Names
user__name
dates
time
word_count
bytes
chpater
tag
comments
actual__text

2 Parsing Ideas

2.1 User name

2.1.1 Corresponding HTML codes

```
<a href="/wiki/User:Clayoquot" class="mw-userlink" title="User:Clayoquot"
↪ data-mw-revid="1190970117"><bdi>Clayoquot</bdi></a>
```

2.1.2 Idea

- Find “User:” and extract the user name
- Find

2.2 Dates and Time

2.2.1 Corresponding HTML codes

```
<span class="mw-changeslist-time">10:40</span>
<a href="/w/index.php?title=Climate_change&oldid=1190008379"
↪ class="mw-changeslist-date" title="Climate change">10:40, 15 December 2023</a>
```

2.2.2 Idea

- Find “mw-changeslist-time” and extract the time
- Find “class=”mw-changeslist-date” title=“Climate change”>

2.2.3 Python Code

```

from bs4 import BeautifulSoup

# Your HTML content
html_content = '''
<span class="mw-changeslist-time">10:40</span>
<a href="/w/index.php?title=Climate_change&oldid=1190008379"
  ↪ class="mw-changeslist-date" title="Climate change">10:40, 15 December 2023</a>
'''

# Parse the HTML content
soup = BeautifulSoup(html_content, 'lxml')

# Extract time using the 'mw-changeslist-time' class
time = soup.find(class_="mw-changeslist-time").text

# Extract date using the 'mw-changeslist-date' class
date = soup.find(class_="mw-changeslist-date").text

print(f"Time: {time}")
print(f>Date: {date}")

```

2.3 Word count and Bytes

2.3.1 Corresponding HTML codes

```

<span class="history-size mw-diff-bytes" data-mw-bytes="297168">297,168 bytes</span>
<span dir="ltr" class="mw-plusminus-neg mw-diff-bytes" title="299,444 bytes after change
  ↪ of this size">-81</span>

```

2.3.2 Idea

- Find “data-mw-bytes” and extract the bytes
- Find class=“mw-plusminus-neg mw-diff-bytes” and extract bytes and word count

2.4 Chapter

2.4.1 Corresponding HTML codes

```

<a href="/wiki/Climate_change#Modelling" title="Climate change">→Modelling</a>
<span class="autocomment"><a href="/wiki/Climate_change#Modelling" title="Climate
  ↪ change">→Modelling</a>: </span>

```

2.4.2 Idea

2.5 Tags

2.5.1 Corresponding HTML codes

```
<a href="/wiki/Wikipedia:VisualEditor" title="Wikipedia:VisualEditor">Visual edit</a>
```

2.5.2 Idea

2.6 Comments

2.6.1 Corresponding HTML codes

```
<span dir="auto"><span class="autocomment"><a href="/wiki/Climate_change#Modelling"
→ title="Climate change">→Modelling</a>: </span> added wikilinks to GST, although I am
→ struggling to figure out what exactly should be in the GST article (as to not
→ overlap with all the other articles)</span>
```

2.6.2 Idea

2.7 Actual Text

2.7.1 Corresponding HTML codes