```
Help on module Co_Occurrence:
NAME
  Co_Occurrence
CLASSES
  builtins.object
    CoOccur
  class CoOccur(builtins.object)
  | CoOccur(dataframe)
   | This class contains methods for updating the document frequency and
   | tokenizing the article.
   | Methods defined here:
   | __init__(self, dataframe)
      This is the constructor for this class which initialises the instance
       variable.
       :param dataframe: This is a pandas dataframe storing every news article.
   | find(self)
       This method updates the document frequency by iterating through each
       document in the dataframe.
       :return: It returns a list which contains all unique words across all
```

documents(document corpus).

| I |  |
|---|--|
| ı | spell_check(self, query)   |
| I | This method checks for spelling errors in the query, and updates any     |
| I | misspelt word.It uses the python library 'pyspellchecker'                |
| I | :param query: query is a list of words which may contain misspelt words. |
| I | :return: It returns a list which contains correct words in query         |
| I | as well as the incorrect words which have now been modified.             |
| I |  |
| I | tokenize(self, article)  |
| 1 | This method tokenises a string into important words. It first case folds |
| 1 | the string, creates tokens, removes stop words and then stems word using |
| 1 | Porter Stemmer.  |
| I | :param article: It is a news article in form of string.                  |
| I | :return: It returns a list of tokenized words.                           |
| I |  |
| I |  |
| 1 | Data descriptors defined here:   |
| 1 |  |
| 1 | dict   |
| 1 | dictionary for instance variables (if defined)                           |
| I |  |
| 1 | weakref  |
| 1 | list of weak references to the object (if defined)                       |
| I |  |
| 1 |  |
| I | Data and other attributes defined here:                                  |
| 1 |  |

```
| df = Empty DataFrame
| Columns: []
| Index: []

DATA

docFreq = {}

ps = <PorterStemmer>

stop_words = {'!', '"', "''', '', ...}

stopwords = <WordListCorpusReader in 'C:\\Users\\Smit\\AppData\\Roamin...

FILE

c:\users\smit\pycharmprojects\searchengine\co_occurrence.py</pre>
```