

Install newspaper3k

```
Text_Summary.py 2
Text_Summary.py > ...
1 #pip install newspaper3k
2
3 from newspaper import Article
4
5 url = 'https://www.sciencedaily.com/releases/2021/08/210811162816.htm'
```

PROBLEMS (2) OUTPUT DEBUG CONSOLE TERMINAL

PS C:\ISSR\Master2022_2023\Sim2_New_20230412\Sim2\Information System\AssignmentList\Lect7_8_LDA>
Requirement already satisfied: lxml>=3.6.0 in c:\python311\lib\site-packages (from newspaper3k) (4.9.2)
Collecting nltk>=3.2.1 (from newspaper3k)
 Downloading nltk-3.8.1-py3-none-any.whl (1.5 MB)
 1.5/1.5 MB 1.6 MB/s eta 0:00:00
Requirement already satisfied: requests>=2.10.0 in c:\python311\lib\site-packages (from newspaper3k) (2.28.2)
Collecting feedparser>=5.2.1 (from newspaper3k)
 Downloading feedparser-6.0.10-py3-none-any.whl (81 kB)
 81.1/81.1 kB 753.9 kB/s eta 0:00:00
Collecting tldextract>=2.0.1 (from newspaper3k)
 Downloading tldextract-3.4.4-py3-none-any.whl (93 kB)
 93.3/93.3 kB 5.5 MB/s eta 0:00:00
Collecting feedfinder2>=0.0.4 (from newspaper3k)
 Downloading feedfinder2-0.0.4.tar.gz (3.3 kB)
 Installing build dependencies ... done
 Getting requirements to build wheel ... done
 Preparing metadata (pyproject.toml) ... done
Collecting jieba3k>=0.35.1 (from newspaper3k)
 Downloading jieba3k-0.35.1.zip (7.4 MB)
 7.4/7.4 MB 2.7 MB/s eta 0:00:00
 Installing build dependencies ... done
 Getting requirements to build wheel ... done
 Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: python-dateutil>=2.5.3 in c:\python311\lib\site-packages (from newspaper3k) (2.8.2)
Collecting tinysegmenter==0.3 (from newspaper3k)
 Downloading tinysegmenter-0.3.tar.gz (16 kB)
 Go to Settings to activate Windows.

```
Text_Summary.py 2
Text_Summary.py > ...
1 #pip install newspaper3k
2
3 from newspaper import Article
4
5 url = 'https://www.sciencedaily.com/releases/2021/08/210811162816.htm'
```

PROBLEMS (2) OUTPUT DEBUG CONSOLE TERMINAL

Requirement already satisfied: python-dateutil>=2.5.3 in c:\python311\lib\site-packages (from newspaper3k) (2.8.2)
Collecting tinysegmenter==0.3 (from newspaper3k)
 Downloading tinysegmenter-0.3.tar.gz (16 kB)
 Installing build dependencies ... done
 Getting requirements to build wheel ... done
 Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: soupsieve>1.2 in c:\python311\lib\site-packages (from beautifulsoup4>=4.4.1->newspaper3k) (2.4.1)
Requirement already satisfied: six in c:\python311\lib\site-packages (from feedfinder2>=0.0.4->newspaper3k) (1.16.0)
Collecting sgmlib3k-1.0.0 (from feedparser>=5.2.1->newspaper3k)
 Downloading sgmlib3k-1.0.0.tar.gz (5.8 kB)
 Installing build dependencies ... done
 Getting requirements to build wheel ... done
 Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: click in c:\python311\lib\site-packages (from nltk>=3.2.1->newspaper3k) (8.1.3)
Requirement already satisfied: joblib in c:\python311\lib\site-packages (from nltk>=3.2.1->newspaper3k) (1.2.0)
Collecting regex>=2021.8.3 (from nltk>=3.2.1->newspaper3k)
 Downloading regex-2023.5.5-cp311-cp311-win_amd64.whl (267 kB)
 267.9/267.9 kB 785.6 kB/s eta 0:00:00
Requirement already satisfied: tqdm in c:\python311\lib\site-packages (from nltk>=3.2.1->newspaper3k) (4.65.0)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\python311\lib\site-packages (from requests>=2.10.0->newspaper3k) (3.1.0)
Requirement already satisfied: idna<4,>=2.5 in c:\python311\lib\site-packages (from requests>=2.10.0->newspaper3k) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\python311\lib\site-packages (from requests>=2.10.0->newspaper3k) (1.26.15)
Requirement already satisfied: certifi>=2017.4.17 in c:\python311\lib\site-packages (from requests>=2.10.0->newspaper3k) (2023.7.22)
 Go to Settings to activate Windows.

```
Text_Summary.py 2
Text_Summary.py > ...
1 #pip install newspaper3k
2
3 from newspaper import Article
4
5 url = 'https://www.sciencedaily.com/releases/2021/08/210811162816.htm'

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL
1.26.15)
Requirement already satisfied: certifi>=2017.4.17 in c:\python311\lib\site-packages (from requests>=2.10.0->newspaper3k) (202
2.12.7)
Collecting requests-file>=1.4 (from tldextract>=2.0.1->newspaper3k)
  Downloading requests_file-1.5.1-py3-none-any.whl (3.7 kB)
Collecting filelock>=3.0.8 (from tldextract>=2.0.1->newspaper3k)
  Downloading filelock-3.12.0-py3-none-any.whl (10 kB)
Requirement already satisfied: colorama in c:\python311\lib\site-packages (from click->nlTK>=3.2.1->newspaper3k) (0.4.6)
Building wheels for collected packages: tinysegmenter, feedfinder2, jieba3k, sgmlib3k
  Building wheel for tinysegmenter (pyproject.toml) ... done
  Created wheel for tinysegmenter: filename=tinysegmenter-0.3-py3-none-any.whl size=13568 sha256=82518d6d5c59a6b15c362a5b8364
ed6ce851165d3c6ea704e73517b511b43059
  Stored in directory: c:\users\mahmoudattaali\appdata\local\pip\cache\wheels\fc\ab\8\c3a9ae6d828bd346be695f7ff54612cd22b7
285b9eab6d7d6a7f55bf83c689ef2
  Stored in directory: c:\users\mahmoudattaali\appdata\local\pip\cache\wheels\3a\al\46\8e68055c1713f9c4598774c15ad0541f26d542
5ee7423b6493
  Building wheel for sgmlib3k (pyproject.toml) ... done
  Created wheel for sgmlib3k: filename=sgmlib3k-1.0.0-py3-none-any.whl size=6060 sha256=95d319da758209386dc1d5ddf69830a9eab
32ee03c4669dc3f8b39737343955d
  Stored in directory: c:\users\mahmoudattaali\appdata\local\pip\cache\wheels\3b\25\2a\105d6a15df6914f4d15047691c6c28f9052cc1
173e40285d03
Successfully built tinysegmenter feedfinder2 jieba3k sgmlib3k
Installing collected packages: tinysegmenter, sgmlib3k, jieba3k, regex, PyYAML, filelock, feedparser, cssselect, requests-fi
le, nltk, feedfinder2, tldextract, newspaper3k
Successfully installed PyYAML-6.0 cssselect-1.2.0 feedfinder2-0.0.4 feedparser-6.0.10 filelock-3.12.0 jieba3k-0.35.1 newspape
r3k-0.2.8 nltk-3.8.1 regex-2023.5.5 requests-file-1.5.1 sgmlib3k-1.0.0 tinysegmenter-0.3 tldextract-3.4.4

Screen Reader Optimized Ln 1, Col 2 Spaces: 4 UTF-8 CRLF Python 3.11.1 64-bit
```

Install Spacy

```
Text_Summary.py X
Text_Summary.py > ...
12
13 import spacy
14 from spacy.lang.en.stop_words import STOP_WORDS
15 from string import punctuation

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
ModuleNotFoundError: No module named 'spacy'
PS C:\ISSR\Master2022_2023\Sim2_New_20230412\Sim2\Information System\AssignmentList\Lect7_8_LDA> pip install spacy
Collecting spacy
  Downloading spacy-3.5.3-cp311-cp311-win_amd64.whl (12.2 MB)
    12.2/12.2 MB 3.2 MB/s eta 0:00:00
Collecting spacy-legacy<3.1.0,>=3.0.11 (from spacy)
  Downloading spacy_legacy-3.0.12-py2.py3-none-any.whl (29 kB)
Collecting spacy-loggers<2.0.0,>=1.0.0 (from spacy)
  Downloading spacy_loggers-1.0.4-py3-none-any.whl (11 kB)
Collecting murmurhash<1.1.0,>=0.28.0 (from spacy)
  Downloading murmurhash-1.0.9-cp311-cp311-win_amd64.whl (18 kB)
Collecting cymem<2.1.0,>=2.0.2 (from spacy)
  Downloading cymem-2.0.7-cp311-cp311-win_amd64.whl (28 kB)
Collecting preshed<3.1.0,>=3.0.2 (from spacy)
  Downloading preshed-3.0.8-cp311-cp311-win_amd64.whl (91 kB)
    91.9/91.9 kB 1.0 MB/s eta 0:00:00
Collecting thinc<8.2.0,>=8.1.8 (from spacy)
  Downloading thinc-8.1.10-cp311-cp311-win_amd64.whl (1.5 MB)
    1.5/1.5 MB 3.2 MB/s eta 0:00:00
Collecting wasabi<1.2.0,>=0.9.1 (from spacy)
  Downloading wasabi-1.1.1-py3-none-any.whl (27 kB)
Collecting srsly<3.0.0,>=2.4.3 (from spacy)
  Downloading srsly-2.4.6-cp311-cp311-win_amd64.whl (478 kB)
    478.8/478.8 kB 1.6 MB/s eta 0:00:00
Collecting catalogue<2.1.0,>=2.0.6 (from spacy)
  Downloading catalogue-2.0.8-py3-none-any.whl (17 kB)
Collecting typer<0.8.0,>=0.3.0 (from spacy)
```

Text_Summary.py X

Text_Summary.py > ...

12
13 import spacy
14 from spacy.lang.en.stop_words import STOP_WORDS
15 from string import punctuation

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

Code + - - - ^ X

Downloading catalogue-2.0.8-py3-none-any.whl (17 kB)
Collecting typer<0.8.0,>=0.3.0 (from spacy)
Downloading typer-0.7.0-py3-none-any.whl (38 kB)
Collecting pathy>=0.10.0 (from spacy)
Downloading pathy-0.10.1-py3-none-any.whl (48 kB)
48.9/48.9 kB 1.2 MB/s eta 0:00:00
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in c:\python311\lib\site-packages (from spacy) (6.3.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\python311\lib\site-packages (from spacy) (4.65.0)
Requirement already satisfied: numpy>=1.15.0 in c:\python311\lib\site-packages (from spacy) (1.24.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\python311\lib\site-packages (from spacy) (2.28.2)
Collecting pydantic!=1.8,!1.8.1,<1.11.0,>=1.7.4 (from spacy)
Downloading pydantic-1.10.8-cp311-cp311-win_amd64.whl (2.1 MB)
2.1/2.1 MB 610.1 kB/s eta 0:00:00
Requirement already satisfied: Jinja2 in c:\python311\lib\site-packages (from spacy) (3.1.2)
Requirement already satisfied: setuptools in c:\python311\lib\site-packages (from spacy) (65.5.0)
Requirement already satisfied: packaging>=20.0 in c:\python311\lib\site-packages (from spacy) (23.0)
Collecting langcodes<4.0.0,>=3.2.0 (from spacy)
Downloading langcodes-3.3.0-py3-none-any.whl (181 kB)
181.6/181.6 kB 1.8 MB/s eta 0:00:00
Requirement already satisfied: typing-extensions>=4.2.0 in c:\python311\lib\site-packages (from pydantic!=1.8,!1.8.1,<1.11.0,>=1.7.4->spacy) (4.5.0)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\python311\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.1.0)
Requirement already satisfied: idna<4,>=2.5 in c:\python311\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\python311\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (1.26.15)
Requirement already satisfied: certifi>=2017.4.17 in c:\python311\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2022.12.7)

Screen Reader Optimized Ln 12, Col 1 Spaces: 4 UTF-8 CRLF Python 3.11.1 64-bit

Text_Summary.py X

Text_Summary.py > ...

12
13 import spacy
14 from spacy.lang.en.stop_words import STOP_WORDS
15 from string import punctuation
16 from heapq import nlargest
17
18 def summarize(text, per):
19 nlp = spacy.load('en_core_web_sm')
20 doc = nlp(text)
21 tokens=[token.text for token in doc]
22 word_frequencies={}
23 for word in doc:

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

Code + - - - ^ X

22.12.7)
Collecting blis<0.8.0,>=0.7.8 (from thinc<8.2.0,>=8.1.8->spacy)
Downloading blis-0.7.9-cp311-cp311-win_amd64.whl (7.0 MB)
7.0/7.0 MB 4.0 MB/s eta 0:00:00
Collecting confection<1.0.0,>=0.0.1 (from thinc<8.2.0,>=8.1.8->spacy)
Downloading confection-0.0.4-py3-none-any.whl (32 kB)
Requirement already satisfied: colorama in c:\python311\lib\site-packages (from tqdm<5.0.0,>=4.38.0->spacy) (0.4.6)
Requirement already satisfied: click<9.0.0,>=7.1.1 in c:\python311\lib\site-packages (from typer<0.8.0,>=0.3.0->spacy) (8.1.3)
Requirement already satisfied: MarkupSafe>=2.0 in c:\python311\lib\site-packages (from Jinja2->spacy) (2.1.2)
Installing collected packages: cymem, wasabi, spacy-loggers, spacy-legacy, pydantic, murmurhash, langcodes, catalogue, blis, typer, srsly, preshed, pathy, confection, thinc, spacy
Successfully installed blis-0.7.9 catalogue-2.0.8 confection-0.0.4 cymem-2.0.7 langcodes-3.3.0 murmurhash-1.0.9 pathy-0.10.1 preshed-3.0.8 pydantic-1.10.8 spacy-3.5.3 spacy-legacy-3.0.12 spacy-loggers-1.0.4 srsly-2.4.6 thinc-8.1.10 typer-0.7.0 wasabi-1.1.1
PS C:\ISSR\Master2022_2023\Sim2_New_20230412\Sim2\Information System\AssignmentList\Lect7_8_LDA>

Go to Settings to activate Windows.