

```
pip install nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.3.2)
Requirement already satisfied: regex<=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.2)
```

```
import nltk
nltk.__version__
```

```
'3.8.1'
```

```
pip install spacy
```

```
Requirement already satisfied: spacy in /usr/local/lib/python3.10/dist-packages (3.7.4)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.0.9)
Requirement already satisfied: thinc<8.3.0,>=8.2.2 in /usr/local/lib/python3.10/dist-packages (from spacy) (8.2.3)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.1.2)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.0.10)
Requirement already satisfied: weasel<0.4.0,>=0.1.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.3.4)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (0.9.0)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.10/dist-packages (from spacy) (6.4.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (4.66.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.31.0)
Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.10/dist-packages (from spacy) (2.6.1)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.1.3)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from spacy) (67.7.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (23.2)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (3.3.0)
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.10/dist-packages (from spacy) (1.25.2)
Requirement already satisfied: annotated-types>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!1.8.1,<3.0.0,')
Requirement already satisfied: pydantic-core==2.16.2 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!1.8.1,<3.0.0,')
Requirement already satisfied: typing-extensions>=4.6.1 in /usr/local/lib/python3.10/dist-packages (from pydantic!=1.8,!1.8.1,<3.0.0,')
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.6)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3.6)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (2023.11.17)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.10/dist-packages (from thinc<8.3.0,>=8.2.2->spacy) (0.7)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.10/dist-packages (from thinc<8.3.0,>=8.2.2->spacy) (0.8)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.10/dist-packages (from typer<0.10.0,>=0.3.0->spacy) (8)
Requirement already satisfied: cloudpathlib<0.17.0,>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from weasel<0.4.0,>=0.1.0->spacy) (0.16.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2->spacy) (2.1.5)
```

```
import spacy
spacy.__version__
```

```
'3.7.4'
```

```
nlp = spacy.load("en_core_web_sm")
```

```
import re
```

```
type(nlp)
```

```
spacy.lang.en.English
def __call__(text: Union[str, Doc], *, disable: Iterable[str]=SimpleFrozenList(),
component_cfg: Optional[Dict[str, Dict[str, Any]]]=None) -> Doc
```

A text-processing pipeline. Usually you'll load this once per process, and pass the instance around your application.

Defaults (class): Settings, data and factory methods for creating the `nlp` object and processing pipeline.
lang (str): IETF language code, such as 'en'.

```
%%writefile text.txt
test text
```

```
Writing text.txt
```

```
pip install PyPDF2
```

```
Collecting PyPDF2
  Downloading pypdf2-3.0.1-py3-none-any.whl (232 kB)
  232.6/232.6 kB 4.2 MB/s eta 0:00:00
Installing collected packages: PyPDF2
Successfully installed PyPDF2-3.0.1
```

```
import PyPDF2 as pdf
```

```
read pages
```

```
Declaration = open('/content/US Declaration.pdf', 'rb')
reader = pdf.PdfReader(Declaration)
reader
```

```
<PyPDF2._reader.PdfReader at 0x7b3efc5fbfd0>
```

```
len(reader.pages)
```

```
7
```

```
page1 = reader.pages[0]
page1_text = page1.extract_text()
print(page1_text)
```

```
AP American Government
Declaration of Independence , 1776
```

Required Foundation Documents

Declaration of Independence, July 4, 1776

When in the course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the laws of nature and of nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation.

We hold these truths to be self-evident:

That all men are created equal; that they are endowed by their Creator with certain unalienable rights; that among these are life, liberty, and the pursuit of happiness; that, to secure these rights, governments are instituted among men, deriving their just powers from the consent of the governed; that whenever any form of government becomes destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying its foundation on such principles, and organizing its powers in such form, as to them shall seem most likely to effect their safety and happiness. Prudence, indeed, will dictate that governments long established should not be changed for light and transient causes; and accordingly all experience hath shown that mankind are more disposed to suffer, while evils are sufferable than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, pursuing invariably the same object, evinces a design to reduce them under absolute despotism, it is their right, it is their duty, to throw off such government, and to provide new guards for their future security. Such has been the patient sufferance of these colonies; and such is now the necessity which constrains them to alter their former systems of government. The history of the present King of Great Britain is a history of repeated injuries and usurpations, all having in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world.

He has refused his assent to laws, the most wholesome and necessary for the public good.

He has forbidden his governors to pass laws of immediate and pressing importance, unless suspended in their operation till his assent should be obtained; and, when so suspended, he has utterly neglected to attend to them.

He has refused to pass other laws for the accommodation of large districts of people, unless those people would relinquish the right of representation in the legislature, a right inestimable to them, and formidable to tyrants only.

He has called together legislative bodies at places unusual, uncomfortable, and distant from the depository of their public records, for the sole purpose of fatiguing them into compliance with his measures.

He has dissolved representative houses repeatedly, for opposing, with manly firmness, his invasions on the rights of the people.

He has refused for a long time, after such dissolutions, to cause others to be elected; whereby the legislative powers, incapable of annihilation, have returned to the people at large for their exercise; the state remaining, in the mean time, exposed to all the dangers of invasions from without and convulsions within.

He has endeavored to prevent the population of these states; for that purpose obstructing the laws for naturalization of foreigners; refusing to pass others to encourage their migration hither, and raising the conditions of new appropriations of lands.

He has obstructed the administration of justice, by refusing his assent to laws for establishing judiciary powers.

He has made judges dependent on his will alone, for the tenure of their offices, and the amount and payment of their salaries.


```
0,  
0,  
0,  
0,  
0,  
722,  
667,  
667.
```

```
out = open("new pdf file.pdf", "wb")  
writer.write(out)
```

```
(False, <_io.BufferedWriter name='new pdf file.pdf'>)
```

```
out.close()
```

merge pages toghother

```
from PyPDF2 import PdfReader, PdfWriter, PdfMerger
```

```
merge = PdfMerger()
```

```
file1 = "/content/F1.pdf"  
file2 = "/content/F2.pdf"
```

```
merge.append(PdfReader(open(file1, 'rb')))  
merge.append(PdfReader(open(file2, 'rb')))
```

```
merge.write('merged.pdf')
```

add new pages

```
add1 = open(file1, 'rb')  
add2 = open(file2, 'rb')
```

```
reader1 = pdf.PdfReader(add1)  
reader2 = pdf.PdfReader(add2)
```

```
output = PdfWriter()
```

```
output.add_page(reader1.pages[2])  
output.add_page(reader2.pages[5])
```

```
    '/Subtype': '/Image',
    '/Type': '/XObject',
    '/Width': 43},
  '/Image6': {'/BitsPerComponent': 8,
    '/ColorSpace': '/DeviceRGB',
    '/Filter': '/FlateDecode',
    '/Height': 43,
    '/SMask': {'/BitsPerComponent': 8,
      '/ColorSpace': '/DeviceGray',
      '/Filter': '/FlateDecode',
      '/Height': 43,
      '/Subtype': '/Image',
      '/Type': '/XObject',
      '/Width': 43},
    '/Subtype': '/Image',
    '/Type': '/XObject',
    '/Width': 43}}},
  '/Rotate': 0,
  '/Type': '/Page',
  '/Parent': {'/Type': '/Pages',
    '/Count': 2,
    '/Kids': [IndirectObject(4, 0, 135510449644560),
      IndirectObject(49, 0, 135510449644560)]}}
```

```
outputStream = open('thelastcopy.pdf', 'wb')
output.write(outputStream)
outputStream.close()
```

Start coding or [generate](#) with AI.