Encoding

July 15, 2025

```python
[1]: import nltk
```

```python
[2]: pip install jieba
```

```
Defaulting to user installation because normal site-packages is not writeable
Looking in links: /usr/share/pip-wheels
Collecting jieba
  Downloading jieba-0.42.1.tar.gz (19.2 MB)
                              19.2/19.2

MB 49.5 MB/s eta 0:00:000:00:01[36m0:00:01
  Preparing metadata (setup.py) … done
Building wheels for collected packages: jieba
  Building wheel for jieba (setup.py) … done
  Created wheel for jieba: filename=jieba-0.42.1-py3-none-any.whl
size=19314458
sha256=ecac6a7dcea0c96b01cfc40d8a42a2a3ea04cece475c757e85b31c9331d48306
  Stored in directory: /home/e9f2e6f2-2728-40a7-9938-
d80ebfe53d70/.cache/pip/wheels/c9/69/31/d56d90b22a1777b0b231e234b00302a55be25593
0f8bd92dcd
Successfully built jieba
Installing collected packages: jieba
Successfully installed jieba-0.42.1
Note: you may need to restart the kernel to use updated packages.
```

```python
[1]: import jieba
```

```python
[6]: with open('Python_lab.txt','r',encoding='utf8') as text:
         data=tetx.read()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[6], line 2
      1 with open('Python_lab.txt','r',encoding='utf8') as text:
----> 2     data=tetx.read()

NameError: name 'tetx' is not defined
```

```python
[3]: data
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[3], line 1
----> 1 data

NameError: name 'data' is not defined
```

```python
[5]: import os
     print(os.getcwd())
```

```
/home/e9f2e6f2-2728-40a7-9938-d80ebfe53d70
```

```python
[8]: with open('Python_lab.txt','r',encoding='utf8') as text:
         data=text.read()
```

```python
[9]: data
```

```
[9]: '          ",        \n     "    Python ",            \n     "        ",
     \n     "            ",          \n     "        "'
```

```python
[10]: seg_list= jieba.cut(data, cut_all=True)
```

```python
[11]: seg_list
```

```
[11]: <generator object Tokenizer.cut at 0x7d0f091fa500>
```

```python
[12]: new=(" ".join(seg_list))
```

```
Building prefix dict from the default dictionary …
Dumping model to file cache /tmp/jieba.cache
Loading model cost 0.775 seconds.
Prefix dict has been built successfully.
```

```python
[13]: new
```

```
[13]: '                        ",                        \n                "
        Python     ",                        \n                "
       ",                                                        \n
       "                        ",                        \n                "
           "'
```

```python
[14]: f= open('Python_lab.txt', 'wb')
      f.write(new.encode('utf8'))
      f.close()
```

[ ]: