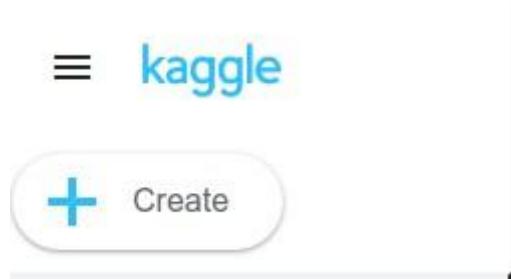


Kaggle 基本使用说明

1. 使用邮箱注册 Kaggle (<https://www.kaggle.com/>)
2. 登录 Kaggle 后，点击左上角新建 Jupyter Notebook，如下图：



3. 新建示例 jupyter-notebook 文件 demo_code，如下：

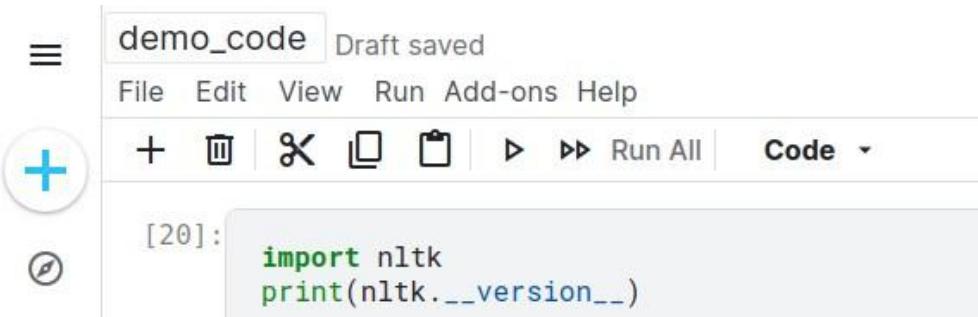
```
import nltk
print(nltk.__version__)

#先分句再分词
sents = nltk.sent_tokenize("And now for something completely different. I love you.")
word = []
for sent in sents:
    word.append(nltk.word_tokenize(sent))
print(word)

#分词
text = nltk.word_tokenize("And now for something completely different.")
print(text)
#词性标注
tagged = nltk.pos_tag(text)
print (tagged[0:6])
#命名实体识别
entities = nltk.chunk.ne_chunk(tagged)
print (entities)
```

```
import jieba
sent='中文分词是文本处理不可或缺的一步！'
seg_list=jieba.cut(sent,cut_all=False)
print('精确模式:','.'.join(seg_list))
seg_list=jieba.cut(sent)
print('默认模式:','.'.join(seg_list))
seg_list=jieba.cut_for_search(sent)
print('搜索引擎模式:','.'.join(seg_list))
seg_list=jieba.cut(sent,cut_all=True)
print('全模式:','.'.join(seg_list))
```

4. 敲入上面两段代码后, run all, 如下图:



5. 每个 cell 下会有相应的输出, 并且页面底部的控制台也会显示每个 cell 的输出, 如下图:

The screenshot shows a Kaggle Jupyter Notebook session. The top part of the image shows a code cell [20]:

```
import nltk
print(nltk.__version__)

#先分句再分词
sents = nltk.sent_tokenize("And now for something completely different. I love you.")
word = []
for sent in sents:
    word.append(nltk.word_tokenize(sent))
    print(word)

#分词
text = nltk.word_tokenize("And now for something completely different")
print(text)
#词性标注
tagged = nltk.pos_tag(text)
print(tagged[0:6])
#命名实体识别
entities = nltk.chunk.ne_chunk(tagged)
print(entities)
```

The bottom part of the image shows the 'Console' output:

```
[[{'And': 'CC', ('now', 'RB'), ('for', 'IN'), ('something', 'NN'), ('completely', 'RB'), ('different', 'JJ'): .}]]
```

Annotations at the bottom of the console output:

- 精确定义：中文/分词/是/文本处理/不可或缺/的/一步/！
- 默认模式：中文/分词/是/文本处理/不可或缺/的/一步/！
- 搜索引擎模式：中文/分词/是/文本/本处/处理/文本处理/不可或缺/不可或缺/的/一步/！
- 全模式：中文/分词/是/文本/文本处理/本处/处理/不可/不可或缺/或缺/的/一步/！

Enter console command here

6. Kaggle 上已经安装了绝大部分的 Python 第三方包, 如需安装 Kaggle 上没有的第三方包, 可在上图中控制台底部的命令行, 使用 pip 命令自行安装, 或者在 jupyter-notebook 的 cell 中使用 pip 命令安装, pip 前要加英文感叹号, 如! pip install jieba。

7. 除上述基本功能外, Kaggle 还有许多便利的功能, 且都是免费的, 请自行探索。