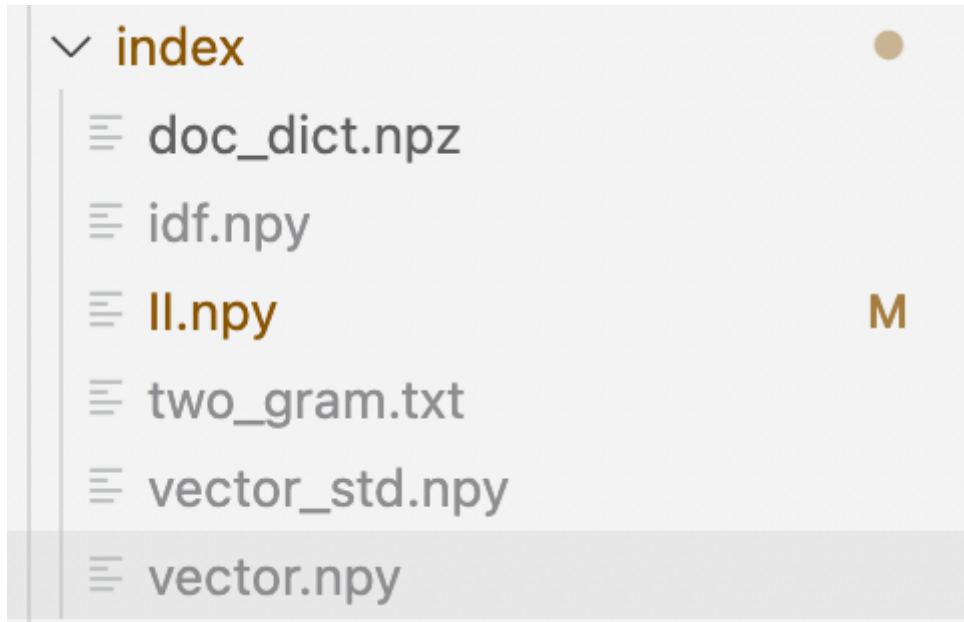# Web Search Engine 使用手册

Group 5

一、下载与运行

使用web search engine的命令如下

```
git clone ...
cd $WEBHOME/src
python console.py
```

- 该应用目前支持Linux和macOS下运行
- 请事先安装使用的库（如numpy）以及模型（如wordnet）
- 在运行 `console.py` 之前，请确认 `index` 目录下有如下文件:

如果文件不全，请将index目录删除，重新运行 `console.py`，应用将生成上述文件。

二、操作介绍

运行 `console.py` 后一段时间，程序读取创建/读取index完毕，进入开始界面。

此处展示了目前的search mode，支持的指令等信息。如果想要重新展示导航栏，请输入

```
:info
```

会重新展示导航栏。

应用接收的输入有两种，第一种为command，以 `:` 开头，`:` 后为执行的命令。另外一种为search，不以任何字符开头，用户直接输入即为搜索内容

下面介绍各种命令

:switch

切换搜索模式，具体使用在下面展示

:open

打开特定功能，如 `open extend` 为开启同义词扩展功能

:close

关闭特定功能，与open相同

:top

修改topk搜索中展示的个数

:clr

清屏，与bash中clear命令相同。

:exit

退出程序

三、搜索使用方法与示例

布尔查询

切换模式为 bool search，输入搜索的单词，将bool运算符（NOT，AND，OR）大写。

```
:switch bool
NOT rise AND fall
```


搜索返回满足结果的文档名，如果文档较多，则会询问是否要全部展示，输入y全部展示，输入n只展示前5个。

通配符查询

切换模式为wildcard search，输入通配表达式

```
:switch wildcard
*formatio*
```

```
>>> :switch wildcard
Switch search mode to wildcard
>>> *formatio*
['reformation', 'information', 'transformation', 'formation']
Search for reformation total: 1
Doc names: 544.html
 544.html: ...  reflects contract reformation agreemats with producers.
...
Search for information total: 150
The answers may be too long to show all of them,
 do u wanna show them all?[y/n]n
Doc names: 496.html 853.html 856.html 1616.html 1666.html
 1856.html 1878.html 1963.html 1964.html 2390.html
 2578.html 2609.html 2783.html 2803.html 2951.html
 3013.html 3098.html 3109.html 3332.html 3394.html
 3409.html 3539.html 3775.html 3980.html 3982.html
 3995.html 4237.html 4291.html 4303.html 4340.html
 4372.html 4630.html 4908.html 4933.html 4970.html
 4976.html 5034.html 5109.html 5145.html 5153.html
 5250.html 5331.html 5391.html 5469.html 5756.html
 6062.html 6085.html 6375.html 6402.html 6450.html
 6471.html 6473.html 6505.html 6537.html 6774.html
 6797.html 6893.html 6961.html 7012.html 7030.html
 7552.html 7583.html 7605.html 7733.html 7747.html
 7970.html 7977.html 8115.html 8298.html 8497.html
 8569.html 8778.html 8799.html 8906.html 8932.html
 8977.html 9047.html 9075.html 9539.html 9582.html
 9606.html 9634.html 9733.html 9734.html 9926.html
 9977.html 10340.html 10452.html 10473.html 10494.html
 10537.html 10548.html 10703.html 10784.html 10931.html
 11040.html 11348.html 11365.html 11504.html 11925.html
 11997.html 12468.html 12917.html 13039.html 13072.html
 14504.html 14918.html 14941.html 14949.html 15091.html
 15110.html 15344.html 15351.html 15372.html 15545.html
 15894.html 15970.html 16611.html 16656.html 16723.html
 16740.html 16875.html 17863.html 17873.html 17906.html
 18024.html 18071.html 18080.html 18146.html 18575.html
 18638.html 19136.html 19683.html 19763.html 19968.html
 20186.html 20262.html 20459.html 20485.html 20618.html
 20767.html 20812.html 20911.html 21006.html 21013.html
 21030.html 21216.html 21367.html 21501.html 21517.html

 496.html: ... provided
 they keep the information secret.
     In his...
 853.html: ... SEC had asked for
 information a few months ago as...
 856.html: ... told Reuters he had no information to
 substantiate the...
 1616.html: ... Bolster of Munger Oil
 Information Services, which tracks...
 1666.html: ... provide underwriting information
```

首先返回所有满足通配符的单词，然后一一对这些单词进行搜索返回搜索结果，并且高亮单词所在位置

短语查询

切换模式为phrase search，输入搜索的单词

```
:switch phrase
commodity company
```

```
>>> :switch phrase
Switch search mode to phrase
>>> commodity company
Search for commodity company total: 1
Doc names: 1960.html
 1960.html: ... find an
  overseas commodity company to make an offer for...
```

返回含有该短语的文档，高亮单词所在位置

Topk查询

切换模式为topk search，输入搜索的单词或短语

```
:switch topk
government
```

```
>>> :switch topk
Switch search mode to topk
>>> government
--------------------top 10--------------------
|----------------------------------------------|
|           Doc name        |        score      |
|----------------------------------------------|
|           6650.html       |       0.2561      |
|----------------------------------------------|
|           3146.html       |       0.2334      |
|----------------------------------------------|
|          20055.html       |       0.2143      |
|----------------------------------------------|
|           7574.html       |       0.2059      |
|----------------------------------------------|
|          15193.html       |       0.1992      |
|----------------------------------------------|
|           4084.html       |       0.1842      |
|----------------------------------------------|
|            930.html       |       0.1729      |
|----------------------------------------------|
|          11341.html       |       0.1702      |
|----------------------------------------------|
|          11269.html       |       0.1667      |
|----------------------------------------------|
|          15951.html       |       0.1649      |
|----------------------------------------------|
```

返回前k个得分最高的文档以及他们的得分

拼写矫正

打开拼写矫正，输入单词后会自动矫正

```
:open correct
kindom
```

```
>>> :open correct
Open correct successfully!
>>> kindom
kindom ---> kingdom
Search for kingdom total: 27
The answers may be too long to show all of them,
 do u wanna show them all?[y/n]n
Doc names: 18.html 304.html 730.html 1217.html 2966.html
 3283.html 3589.html 3747.html 4067.html 4592.html
 5167.html 5171.html 5273.html 5554.html 6215.html
 6996.html 7337.html 8039.html 8748.html 9170.html
 9443.html 15238.html 15471.html 15481.html 16096.html
 16180.html 19262.html
 18.html: ... Foods
  Ltd, a United Kingdom blender, and has licensing...
 304.html: ... GAS
  lt SHV (United Kingdom) Holding Co Ltd>
  said...
 730.html: ... company in the United Kingdom, will expand its
  specialty...
 1217.html: ... unit in the United
  Kingdom, Pye TVT Limited.
  ...
 2966.html: ... paper to banks in the Kingdom.

  ...
```

返回矫正后正确的单词，和正确单词的搜索结果

同义词扩展

打开同义词扩展，输入单词后会将单词扩展并展示

```
:open extend
well
```

```
>>> :open extend
Open extend successfully!
>>> well
Extended command:

well
good

Search for well total: 488
The answers may be too long to show all of them,
 do u wanna show them all?[y/n]n
Doc names: 12.html 110.html 247.html 345.html 372.html
 390.html 435.html 505.html 550.html 696.html
 925.html 926.html 942.html 960.html 977.html
 1193.html 1212.html 1252.html 1335.html 1361.html
 1377.html 1478.html 1520.html 1611.html 1616.html
 1619.html 1661.html 1682.html 1692.html 1696.html
 1709.html 1711.html 1735.html 1809.html 1839.html
 1919.html 1926.html 1959.html 1963.html 1967.html
 1990.html 2061.html 2080.html 2121.html 2197.html
```

首先返回扩展之后的单词，然后返回这些单词的搜索结果。