

分词 API 使用说明

/api/wordseg

请求说明

- 请求方式: GET
- 请求参数: text, 待分词的文本, 类型为字符串
- 返回结果: CSV 文件

示例

假设我们要对文本 "这是一个测试" 进行分词, 可以使用以下 curl 命令进行测试:

```
1 curl -X GET "http://112.4.97.79:6080/api/wordseg?
  text=%E8%BF%99%E6%98%AF%E4%B8%80%E4%B8%AA%E6%B5%8B%E8%AF%95" -H "accept:
  application/csv" -o result.csv
```

其中, `http://112.4.97.79:6080` 是 API 的地址, `%E8%BF%99%E6%98%AF%E4%B8%80%E4%B8%AA%E6%B5%8B%E8%AF%95` 是 URL 编码后的文本 "这是一个测试", `-H "accept: application/csv"` 表示要求返回 CSV 格式的文件, `-o result.csv` 表示将结果保存到本地文件 `result.csv` 中。

返回结果

返回的结果是一个 CSV 文件, 包含了分词结果和词性。例如, 对于文本 "这是一个测试", 分词结果为:

```
1 这   r
2 是   v
3 一个 m
4 测试 vn
```

则返回的 CSV 文件内容为:

```
1 这,r
2 是,v
3 一个,m
4 测试,vn
```

/api/userdict

请求说明

- 请求方式: POST
- 请求参数: userdict, 上传的用户自定义词典文件, 类型为文件
- 返回结果: JSON 格式的成功信息

示例

数据格式:

```
1 | 云计算 5
2 | 李小福 2 nr
3 | 创新办 3 i
4 | easy_install 3 eng
5 | 好用 300
6 | 韩玉赏鉴 3 nz
7 | 八一双鹿 3 nz
8 | 台中
9 | 凯特琳 nz
10 | Edu Trust认证 2000
```

词性说明参考 <https://blog.pulipuli.info/2017/11/fasttag-identify-part-of-speech-in.html>

假设我们要上传一个名为 `userdict.txt` 的自定义词典文件，可以使用以下 `curl` 命令进行测试:

```
1 | curl -X POST "http://112.4.97.79:6080/api/userdict" -H "accept:
   | application/json" -H "Content-Type: multipart/form-data" -F
   | "userdict=@userdict.txt"
```

其中, `http://112.4.97.79:6080` 是 API 的地址, `-H "accept: application/json"` 表示要求返回 JSON 格式的结果, `-H "Content-Type: multipart/form-data"` 表示上传的文件是表单数据, `-F "userdict=@userdict.txt"` 表示上传名为 `userdict.txt` 的文件。

返回结果

如果上传成功, 则返回 JSON 格式的成功信息:

```
1 | {
2 |   "success": true
3 | }
```

如果上传失败, 则返回错误信息。