

抽取式文本摘要算法研究与应用

🕒 Created @Feb 7, 2021 9:04 PM

开发环境

| Python + Django + MySql

1. Python 3.8, 需要安装下列库, 一般直接 `pip install` + 包名即可
 - Django 3.1.6
 - requests 2.25.1
 - snownlp 0.12.3
 - beautifulsoup4 4.9.3
 - mysql-connector-python 8.0.23
2. MySQL 8.x 版本, 可以直接用安装包安装好, 自己配置好用户名和密码
3. PyCharm Professional 版本, 可以直接导入 Django 项目, 方便启动和调试

项目结构

爬虫部分

utils/spider.py 文件, 向新浪滚动新闻发起请求, 获取新闻的相关内容, 同时生成新闻相关的摘要和关键词

数据库设置

通过运行 TEXT_SUMMARY.sql 脚本, 创建数据库 schema 和表, 还可以选择性插入一些数据

utils/database.py 是数据库 CRUD 方法的封装

Django 网站展示部分

- app 文件夹下文件
 - static: 静态文件, css 和 js 代码
 - templates: HTML 网页模版文件, 可以通过 Django 注入数据到模版中

- templatetags: 自定义的一个 tag, 用于分页功能
- views.py: 最主要的网页后台逻辑处理部分, 处理 HTTP 请求并返回数据
- 其他文件大多数是配置文件, 且基本在创建 Django 项目时生成, 不需要做改动

运行方式

- 安装好数据库, 并且到 `database.py`、`spider.py`、`app/views.py` 中改成对应的用户名和密码,

```
DatabaseConnector(user='root', password='+sknLv5T')
```

- 如果使用 PyCharm 专业版, 直接导入 Django 项目, 安装相关依赖, 可以像视频中一样直接在 PyCharm 中启动
- 如果是社区版 PyCharm, 安装好相关依赖库, 通过以下命令启动项目

```
python manage.py runserver 8000
```

根据提示, 浏览器打开 <http://localhost:8000> 即可使用