

NewsFeeds

序言

项目概述

本项目实现了一个新闻资讯收集、整理，进行个性化展示的网站。

网站服务器端使用 Node.js 作为服务器框架，使用 MongoDB 提供数据存储服务，并使用 Redis 作为数据缓存服务。在用户管理方面实现了邮箱认证注册登录，用户定制关键词等功能。

在新闻数据爬取上使用 Python BeautifulSoup 框架爬取并解析各大主流新闻网站的新闻页面，获取所需要的新闻数据；在新闻数据处理上通过维基百科提供的语料库，使用 Word2Vec 等自然语言处理相关的工具进行中文关键词建模，实现了按新闻关键词进行聚类，并根据用户感兴趣关键词进行新闻个性推荐；同时，使用了微软认知服务提供的推荐 API，用于针对用户的特定行为进行建模推荐。

主要源代码结构

NewsFeeds

📁 Graphs 实际运行效果截图, 系统架构图等

📄 [README.md](#) 开发文档

📁 Scripts 新闻数据相关脚本

📄 config.ini 配置文件

📄 [modelUpdater.py](#) 推荐系统相关实现脚本

📄 [newsCrawler.py](#) 新闻爬取脚本

📄 [service.py](#) 服务器后台常驻进程脚本

📁 Server 新闻网站服务器

📄 app.js Node.js 入口脚本

📁 bin 服务器启动脚本

📄 config.json 网站配置文件（服务器地址，数据库源等）

📄 package.json Node.js 依赖模块声明

📁 public 网站静态文件

📄 routes Node.js 路由

📄 scripts 网站后端逻辑实现

📄 views 网站前端实现

192 directories, 65 files

运行环境

- 操作系统：Windows/Linux/OSX
- 服务器软件：Node.js 6.9.0 及以上
- 逻辑数据库：MongoDB 2.4.1 及以上
- 缓存数据库：Redis 3.2.9 及以上

使用说明

1. 进入服务器所在目录 Server
2. 根据各个服务器配置，按如下说明更改配置文件 config.json:

```
{
  "host": " 服务器主机地址",
  "port": " 服务器主机端口",
  "connect": "MongoDB 连接字符串",
  "newscol": "MongoDB 新闻数据库名",
  "usercol": "MongoDB 用户数据库名",
  "mailoptions": {
    "service": " 邮件服务",
    "email": " 邮箱账号",
    "password": " 邮箱密码"
  },
  "redis": {
    "port": "Redis 端口",
    "host": "Redis 服务器地址"
  },
  "database": {
    "host": "MongoDB 数据库服务器地址",
    "port": "MongoDB 服务器端口",
    "user": "Mongo 用户名",
    "password": " 密码"
  },
  "api": {
    "modelid": " 微软认知服务-推荐系统模型 ID",
    "endpoint": "https://westus.api.cognitive.microsoft.com/recommendations/v4.0",
    "token": " 微软认知服务认证 Token"
  },
  "genres": {
    "personal": " 推荐",
    "hot": " 热门",
    "society": " 社会",
    "domestic": " 国内",
    "global": " 国际",
    "technology": " 科技",
    "finance": " 经济",
    "war": " 军事",
    "education": " 教育",
    "car": " 汽车",
  }
}
```

```
    "game": " 游戏",  
    "discover": " 探索",  
    "entertain": " 娱乐",  
    "fashion": " 时尚",  
    "health": " 健康",  
    "history": " 历史",  
    "mobile": " 数码",  
    "sport": " 体育",  
    "travel": " 旅游"  
  }  
}
```

3. 安装 Node.js 依赖包：

```
npm install
```

4. 启动服务器

```
npm start
```

前端实现

本网站前端主要使用 Bootstrap+jQuery 框架实现。

新闻爬取

新闻推荐

成果展示