

使用方法

项目构建环境系统默认为Windows 10，其他平台未经过测试

环境配置

```
Neo4j Desktop 1.2.7+  
python3.7+
```

安装

```
$ pip3 install -r requirements.txt
```

项目使用ChromeDriver为ChromeDriver Canary版本，需根据自身系统Chrome版本进行调整

使用备份数据库(database_backups)

恢复

```
$ ./neo4j-admin dump --database=neo4j --to=$PATH\KGRM-master\database_backups\backup.dump
```

备份

```
$ ./neo4j-admin load --from=$PATH\KGRM-master\database_backups\backup.dump --  
database=neo4j --force
```

数据处理流程

- 爬取获得JSON文件.

```
$ cd /kg_spider/kg_spider/spiders  
$ scrapy runspider xxx.py -o xxx.json
```

- data_clean.py中delete_whitespace去除冗余数据
- json2csv.py中cut_data剥离出name、price、content、entirety四个文件
- 对于name文件在txt_clean.py中addentity添加前置实体和关系、append添加后置实体和关系
- 移动node和relation文件到import目录，执行cypher语句

```
USING PERIODIC COMMIT 2000  
LOAD CSV WITH HEADERS FROM "file:///node.csv" AS line  
CREATE (:product { title: line.title })
```

```
USING PERIODIC COMMIT 2000  
LOAD CSV WITH HEADERS FROM "file:///relation.csv" AS line  
MATCH (entity1:newNode{title:line.newNode})  
MATCH (entity2:product{title:line.product})  
CREATE (entity1)-[:RELATION {type : line.relation }]->(entity2)
```

运行

flask 1.2.1+

建议使用virtual environment(venv)

```
$ export FLASK_APP=app.py  
$ flask run
```