

# MSQA使用说明

## 0. 数据集更新

模型默认数据集问题集等都放在./dataset文件夹下，如果需要替换知识库与问题集，将数据集放入dataset就好，如果涉及到改名等操作需要在MSQA.py文件的前几行将文件路径修改一下，修改成新的路径即可：

```
Zhfile = 'dataset/zh_triples.txt'
Enfile = 'dataset/en_triples.txt'
AlgnmentfileZE = 'dataset/zh_en_alignment.txt'
AlgnmentfileEZ = 'dataset/en_zh_alignment.txt'
EnRDF = "En_triple.rdf"
ZhRDF = "Zh_triple.rdf"
```

## 1. 构建RDF文件

数据集更新完成后需要先将知识库文档转为RDF文件才能进行SPARQL查询，命令如下：

```
python RDF_Generator.py --zh '中文知识库文件路径'
python RDF_Generator.py --en '英文知识库文件路径'
```

```
E:\dxy\lab\MSQA>python RDF_Generator.py --zh ./dataset/zh_triples.txt
building zh_triple.rdf...
done.
```

构建好的rdf文件在当前目录下为Zh\_triple.rdf和En\_triple.rdf

## 2. 问题集测试

建立Result目录存放结果，MSQA模型结果会输出到该文件夹中

```
python MSQA.py --q ./dataset/en_en_zh.txt --p ./dataset/question_predicates.txt
```

--q输入问题集路径，--p输入谓词划分文件路径

```
E:\dxy\lab\MSQA>python MSQA.py --q ./dataset/en_en_zh.txt --p ./dataset/question_predicates.txt
```

输出的结果记录在./Result文件夹中

## 3. 用到的lib

```
import rdflib
import argparse
import Levenshtein
from gurobipy
```