**考试说明**

**一：考试内容**

内容：给定10个query（见query.txt），从大规模文本(见raw\_data.txt)中查找每个query最相关的10个文档。

返回格式：结果保存为result.json文件。注意：不要额外添加任何符号。

要求:

1. 使用json进行存储成dict的格式。
2. 编码方式:UTF\_8
3. 去掉句末的空格
4. 提示：在进行文件写入时，注意将json.dump的ensure\_ascii设置为False，保证文件中文字的正常显示。
5. 每个query对应一个字典（dict），字典的key为该query的编号(从0开始到9)，value为一个list，list里包含该query对应的10个检索结果（顺序不做要求）。

**eg：**{"0":[Answer1,Answer2,…,Answer10], "1":[Answer1,Answer2,…,Answer10],

…,

"9":[Answer1,Answer2,…,Answer10]}

1. 若返回的检索结果中有重复，进行去重，最终需要返回10个互不相同的检索结果。

**二：考试提交内容**

* 代码文件：整个代码整合到一个**code.py**文件中（若有多个文件，整合为code.zip），代码将进行查重检测。
* 结果文件：**result.json**
* 说明文件：word文件，文件名：说明文档。内容如下：
  + 整个检索流程图
  + 基于流程图进行额外阐述说明
  + 基于自己的代码，对主要功能函数进行说明（注意：不需要粘贴长篇代码，只需对函数功能以及参数等进行说明）
* 提交方法：将上述三个文件放入一个文件夹并压缩，**文件夹命名格式：学号\_姓名。**将压缩文件发送到监考人员指定邮箱。邮件主题：**学号\_姓名。**

**三：注意事项**

1.注意所提交的格式，必须按照指定要求的格式提交，不按要求者，将会扣分。

2.必须在考试时间内发送到指定邮箱，逾期无效（6.27 下午6.前）。

3.考试所提交所有内容将进行重复性检查，一旦发现抄袭，抄袭者和被抄袭者都为0分。