

信息检索实验三：IR Evaluation

0.个人信息:

姓名：李心怡 学号：201800122025 班级：数据18.1

1.实验目的:

实现以下指标评价，并对Experiment2的检索结果进行评价:

- Mean Average Precision (MAP)
- Mean Reciprocal Rank (MRR)
- Normalized Discounted Cumulative Gain (NDCG)

2.软件环境:

python3.6, Anaconda3, Spyder,pycharm

3.实验内容与设计:

(1) 实验内容:

在实验三已经给出的架构上，根据给出的result和qrels数据集，进行三个检索结果评估函数的填充。这三个函数是用的方法分别为：MAP、MRR、NDCG。最后每个函数返回一个实数值，作为评价结果的度量。

(2) 算法描述:

1.MAP:

MAP评估计算公式如下所示:

$$MAP(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} AP(q_j)$$

其中，Q代表总的查询集，|Q|代表查询的个数。AP是另外一种对查询结果的评价指标，计算公式如下所示:

$$AP(q_j) = \frac{1}{m_j} \sum_{k=1}^{m_j} Precision(R_{jk})$$

其中， m_j 是对于本查询结果 q_j 的相关集的文档数量， R_{jk} 为查询 q_j 从上到下的排序检索结果的集合，直到获得相关结果 dk 。Precision是到 R_{jk} 为止目前真实相关个数与目前截止到 R_{jk} 的总个数的比值。

因此，MAP的主要计算大致要经历两个for循环，外层对每个查询的AP进行叠加，内层对单个查询的AP进行计算。最终除以|Q|，得到最终MAP值答案。

2.MRR:

MRR评价参数是由RR评价参数演变过来的。RR参数的计算如下公式所示：

$$\text{Reciprocal Rank score} = \frac{1}{K}$$

上式中，K指的是作为查询结果的result集中每一个查询对应真实相关文档集合qrels中的首次出现的位置。K相当于是一个索引值。

MRR指的是对所有查询取RR的平均值。

3.NDCG:

NDCG评价参数公式如下所示：

$$\text{NDCG}_n = \frac{\text{DCG}_n}{\text{IDCG}_n}$$

上式中，分子DCG是Discounted Cumulative Gain，表示对查询结果的评估。之所以要除以IDCG（Ideal Discounted Cumulative Gain）（理想值），是因为对DCG的归一化，以便于对不同的DCG值进行比较。

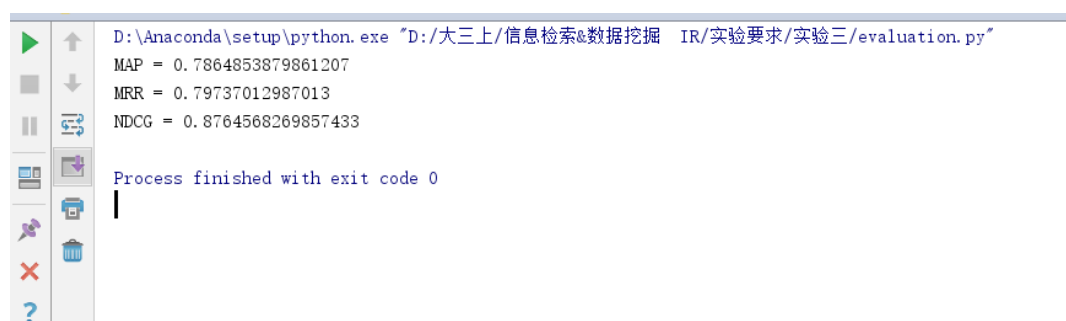
DCG 的计算公式如下所示：

$$\text{DCG}_n = \text{rel}_1 + \sum_{i=2}^n \frac{\text{rel}_i}{\log_2 i}$$

上式中，rel指的是每个文档的相关度gain值，这个值由题目给出。i指的是关于此查询，相关度排在第i个文档的位置。

IDCG与NDCG的计算方法几乎一致，不同之处在于将查询结果按照gain（rel）值进行了降序排序，重新进行DCG的计算。在此不再加以赘述。

(3) 测试结果:



```
D:\Anaconda\setup\python.exe "D:/大三上/信息检索&数据挖掘 IR/实验要求/实验三/evaluation.py"
MAP = 0.7864853879861207
MRR = 0.79737012987013
NDCG = 0.8764568269857433

Process finished with exit code 0
```

4.分析与探讨:

对于这三种查询结果评价，对于同一个查询集，大致结果是差不多的，因为在选取的方式上略有差异，答案可能会有少许差异，但不影响我们对相关性进行判断。

每种方法都恰到好处的处理了归一化的操作，使得不同长度查询之间可以进行比较。有利于用户进行相关性的比对。

