

实验 1 搜索引擎性能的评价

Report

计 22 滕爽 2012011270

同组： 董欣 计 25 2012011361

一、查询样例的集合

样例集合满足以下要求：

- 1、数据来源：百度、搜狗和 360 三个搜索引擎
 - 2、数据规模：10 个关键词并对每个关键词获取了其前 10 位的搜索结果。
 - 3、数据比例：导航类/信息类/事物类，比例 2:5:3 冷门：普通：热门 = 2:4:4
 - 4、数据具体关键词：清华大学、12306、Python2.7 下载、红高粱大结局、水煮肉做法、汪涵救场、减肥、房贷新政、失语症、lovable
- 具体情况如下表所示：
- 5、数据需求：数据选取反应了不同用户的需求，如 python 面向程序员，失语症为医学词汇等。

具体情况如下表所示：

分类	搜索关键字	搜索目标
导航类	清华大学	清华大学的官网主页
	12306	12306火车票预订官网
信息类	汪涵救场	我是歌手主持人汪涵救场词
	减肥	减肥方法
	房贷新政	房贷新政出台
	失语症	失语症的介绍
	lovable	歌曲
事务类	python2.7 下载	下载Python2.7
	红高粱大结局	希望能了解到红高粱的大结局是什么
	水煮肉做法	找到水煮肉的详细做法

二、Pooling 集合的构建

根据构建的样例集合在常用的三个搜索引擎（百度、360 搜、搜狗）对这部分查询词进行了查询。每个查询词保留了前十位的结果。将重复的搜索结果删除后得到的 pooling 表见附件 pooling1.xlsx, pooling2.xlsx。

三、构建相关标注集合

针对已有的信息需求标注 Pooling 集合中的结果，标注为答案与非答案两类，其中答案表示为 1，非答案表示为 0。

经过 URL 的判重最终每个查询词得到的总返回结果、以及相关的结果的 pooling

集合规模如下表。

关键词	相关结果数	总结果数
清华大学	1	23
12306	10	26
Python2.7 下载	21	22
红高粱大结局	15	25
水煮肉做法	21	24
汪涵救场	19	29
减肥	8	29
房贷新政	15	29
失语症	20	26
lovable	7	23

四、性能评价

1) MAP:

MAP 方法是 Mean Average Precision，即平均准确率法的简称。其定义是求每个相关文档检索出后的准确率的平均值（即 Average Precision）的算术平均值（Mean）。这里对准确率求了两次平均，因此称为 Mean Average Precision。

2) P@10:

P@N 本身是 Precision@N 的简称，指的是对特定的查询，考虑位置因素，检测前 N 条结果的准确率。在这个问题中，我们取 N=10。

3) MRR:

MRR 是平均排序倒数（Mean Reciprocal Rank）的简称，MRR 方法主要用于寻址类检索（Navigational Search）或问答类检索（Question Answering），这些检索方法只需要一个相关文档，对召回率不敏感，而是更关注搜索引擎检索到的相关文档是否排在结果列表的前面。

相关结果如下表所示：

搜索引擎	关键词	MAP	MRR	P@10
360	清华大学	1	1	0.1
	12306	0.789285714	1	0.5
	Python2.7 下载	1	1	1
	红高粱大结局	0.891723356	1	0.7
	水煮肉做法	0.931795635	1	0.8
	汪涵救场	0.703174603	0.5	0.7
	减肥	0.155555556	0.111111111	0.2
	房贷新政	0.215740741	0.142857143	0.3
	失语症	0.869047619	1	0.7
	lovable	0.5	0.5	0.2
百度	清华大学	1	1	0.1
	12306	0.666666667	1	0.2
	Python2.7 下载	0.962654321	1	0.9
	红高粱大结局	0.737755102	1	0.7
	水煮肉做法	0.97654321	1	0.9
	汪涵救场	1	1	0.8
	减肥	0.155555556	0.111111111	0.2
	房贷新政	0.768707483	1	0.7
	失语症	1	1	1
	lovable	0.444444444	1	0.3
搜狗	清华大学	1	1	0.1
	12306	0.916666667	1	0.3
	Python2.7 下载	0.928262787	1	0.9
	红高粱大结局	0.452380952	0.333333333	0.5
	水煮肉做法	0.946781305	1	0.9
	汪涵救场	0.572222222	0.5	0.5
	减肥	0.3875	0.25	0.4
	房贷新政	0.436243386	0.2	0.6
	失语症	1	1	0.6
	lovable	0.5	0.5	0.2
搜索引擎	MAP		MRR	P@10
360	0.705632322		0.725396825	0.52
百度	0.771232678		0.911111111	0.58
搜狗	0.714005732		0.678333333	0.5

4) 准确召回率曲线:

以召回率为横轴，准确率为纵轴的曲线。准确率和召回率是互相影响的，理想情况下肯定是做到两者都高，但是一般情况下准确率高、召回率低，召回率低、准确率高。如果是做搜索，那就是保证召回的情况下提升准确率；如果做疾病监测、反垃圾，则是保准确率的条件下，提升召回。

所有的曲线见附件[统计结果.xlsx](#)。

五、分析与总结

从准确性上来看百度搜索引擎 > 搜狗搜索引擎 > 360 搜索引擎。

从 P@10 评价标准来看，百度优于 360 与 sougou，且 360 能力略优于搜狗。

从 MRR 标准来看，百度远远领先于搜狗与 360，说明百度能在页面的开始满足用户的需求。搜狗比 360 拥有更好的排版。

这 3 个搜索引擎的性能有一定的差距，总体上，百度的性能最佳，这符合客观用户选择的结果。其次是搜狗。360 搜索引擎的性能相对较差。

还有各项数据相关度较低的原因可能是植入数据太多，且各个搜索引擎的广告一致程度不相同。

关于中英文搜索, 由 lovable 的结果可以看出，大部分都是翻译词典，而非歌曲的检索。说明各个搜索引擎在英文搜索方面有很大的不足，有长足的发展空间。