Experiment1: Inverted index and Boolean Retrieval Model

Microblog Track

TREC home   Data home   NIST HOME

TREC 2015 Microblog Track
TREC 2014 Microblog Track
TREC 2013 Microblog Track
TREC 2012 Microblog Track
TREC 2011 Microblog Track

https://trec.nist.gov/data/microblog.html

Experiment1: Inverted index and Boolean Retrieval Model

1 {"userName": "Mariah Peoples", "clusterNo": 82, "text": "House may kill Arizona-style immigration bill, Rep. Rick Rand says: The House is unlikely to pass the \"Ari... http://tinyurl.com/4jrjcdz", "timeStr": "Sun Jan 23 00:02:37 +0000 2011", "tweetId": "28965792812892160", "errorCode": "200", "textCleaned": " ", "relevance": 2}
2 {"userName": "servando", "clusterNo": 35, "text": "Mourners recall Sarge Shriver's charity, idealism \n   (AP): AP – R. Sargent Shriver was always an optimist, pio... http://bit.ly/gqMcdG", "timeStr": "Sun Jan 23 00:07:48 +0000 2011", "tweetId": "28967095878287360", "errorCode": "200", "textCleaned": " ", "relevance": 1}
3 {"userName": "Heide Eversoll", "clusterNo": 60, "text": "Bass Fishing Techniques: 2 Fantastic Tips To Improve Your Casting Skills", "timeStr": "Sun Jan 23 00:10:05 +0000 2011", "tweetId": "28967672074993664", "errorCode": "200", "textCleaned": " ", "relevance": 2}
4 {"userName": "Ailsa Hung", "clusterNo": 97, "text": "#Financial Aid | Proper Method Of Getting Financial Aid For Education http://ping.fm/BK0R3 #applying-for-financial-aid financial-aid-essay #", "timeStr": "Sun Jan 23 00:11:03 +0000 2011", "tweetId": "28967914417688576", "errorCode": "200", "textCleaned": " ", "relevance": 2}
5 {"userName": "Brothy", "clusterNo": 89, "text": "Supreme Court: NASA's intrusive background checks OK http://bit.ly/h2jgy9", "timeStr": "Sun Jan 23 00:13:18 +0000 2011", "tweetId": "28968479176531969", "errorCode": "200",

30548条tweets，用于创建inverted index

Experiment1: Inverted index and Boolean Retrieval Model

2014 Microblog Track
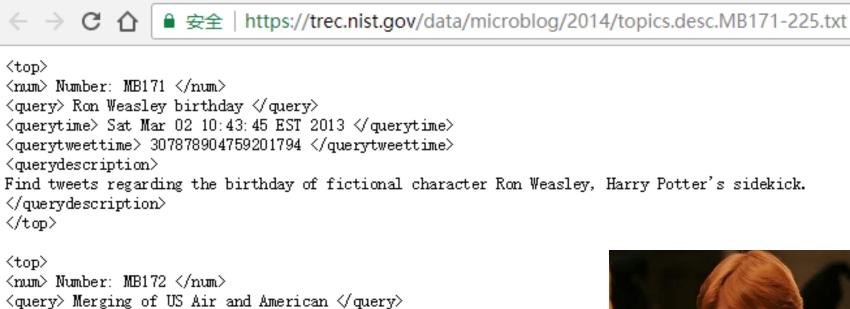
TREC home   Data home   NIST HOME

TREC 2014
Track Web Page: https://github.com/lintool/twitter-tools/wiki

TREC 2014 test topics
TREC 2014 topics with added "Description" field

TREC 2014 relevance judgments
Cluster annotations for the TTG task
Evaluation script for the TTG task

https://trec.nist.gov/data/microblog2014.html

Experiment1: Inverted index and Boolean Retrieval Model

🔒 安全 | https://trec.nist.gov/data/microblog/2014/topics.desc.MB171-225.txt

```
<top>
<num> Number: MB171 </num>
<query> Ron Weasley birthday </query>
<querytime> Sat Mar 02 10:43:45 EST 2013 </querytime>
<querytweettime> 307878904759201794 </querytweettime>
<querydescription>
Find tweets regarding the birthday of fictional character Ron Weasley, Harry Potter's sidekick.
</querydescription>
</top>


<top>
<num> Number: MB172 </num>
<query> Merging of US Air and American </query>
<querytime> Sun Feb 17 16:14:40 EST 2013 </querytime>
<querytweettime> 303251140382973952 </querytweettime>
<querydescription>
Find information on the merger of US Airways and American airlines.
</querydescription>
</top>
```


Query: Ron Weasley birthday 55个queries，用于测试Boolean Retrieval Model

Experiment1: Inverted index and Boolean Retrieval Model

任务：

- 使用我们介绍的方法，在tweets数据集上构建inverted index;
- 实现Boolean Retrieval Model，使用TREC 2014 test topics 进行测试；
- Boolean Retrieval Model：
  - Input：a query (like Ron and Weasley)
  - Output: print the qualified tweets.
  - 支持and, or ,not；查询优化可以选做；

注意：

- 对于tweets与queries使用相同的预处理；

Deadline: 2021.10.13