



# 搜索排序演进101

周金阳

twitter@ailurus1991



我们好像需要一个搜索功能了！小王，就你吧！







Elasticsearch 拯救人类！





不要给我说什么  
底层原理、框架内核！  
老夫敲代码就是  
一把 **梭**！





**复制！**  
**粘贴！**  
拿起键盘就是  
**干！**













 Search or jump to...  Pull requests Issues Marketplace Explore  + 

 medcl / [elasticsearch-analysis-ik](#)  Watch 460  Star 5,033  Fork 1,329

 Search or jump to...  Pull requests Issues Marketplace Explore  + 

 medcl / [elasticsearch-analysis-pinyin](#)  Watch 86  Star 895  Fork 225

 Search or jump to...  Pull requests Issues Marketplace Explore  + 

This repository has been archived by the owner. It is now read-only.

```
4 query = {
5     "multi_match": {
6         "query": query_content,
7         "fields": ["username^3", "title^3", "subtitle^2", "content"],
8         "type": "best_fields"
9     }
10 }
```





Query: 引力波

```
{  
  "title": "引力波，一个世纪的求索",  
  "content": "400多年前，当伽利略第一次将望远镜指向星空时，  
  一个新的时代开启了。从那时起，天文学家用一个又一个震撼人心的  
  观测不断拓展着人类的视野。经过漫长的发展和技术进步，今天的天文  
  观测早已今非昔比，然而本质上，观测星空的天文学家手里的工具  
  基本上万变不离其宗：用越来越大的口径接收来自天体的光子，并用  
  越来越灵敏的探测器记录它们。而今年，在爱因斯坦提出广义相对论  
  整整一个世纪的历史性时刻，我们站在了一个新时代的起点：通过位  
  于美国列文斯顿和汉福德的高新引力波探测器，人类有望在不远的将  
  来捕捉到时空的涟漪，用一种前所未有的方式看待这个世界。那么，  
  什么是引力波呢？要理解引力波，其实也不难。"  
}
```

```
{  
  "title": "引力波",  
  "content": "引力波引力波引力波"  
}
```



排序问题真让人头疼...

老板

头条号 / 好医行



```
{  
  "title": "引力波，一个世纪的求索",  
  "content": "400多年前，当伽利略第一次将望远镜指向星空时，  
一个新的时代开启了。从那时起，天文学家用一个又一个震撼人心的  
观测不断拓展着人类的视野。经过漫长的发展和技术进步，今天的天文  
观测早已今非昔比，然而本质上，观测星空的天文学家手里的工具  
基本上万变不离其宗：用越来越大的口径接收来自天体的光子，并用  
越来越灵敏的探测器记录它们。而今年，在爱因斯坦提出广义相对论  
整整一个世纪的历史性时刻，我们站在了一个新时代的起点：通过位  
于美国列文斯顿和汉福德的高新引力波探测器，人类有望在不远的将  
来捕捉到时空的涟漪，用一种前所未有的方式看待这个世界。那么，  
什么是引力波呢？要理解引力波，其实也不难。"  
}
```

```
{  
  "title": "引力波",  
  "content": "引力波引力波引力波"  
}
```

$f(x)$

87.93

15.43

$g(q, x) \rightarrow f(x) * g(q, x)$

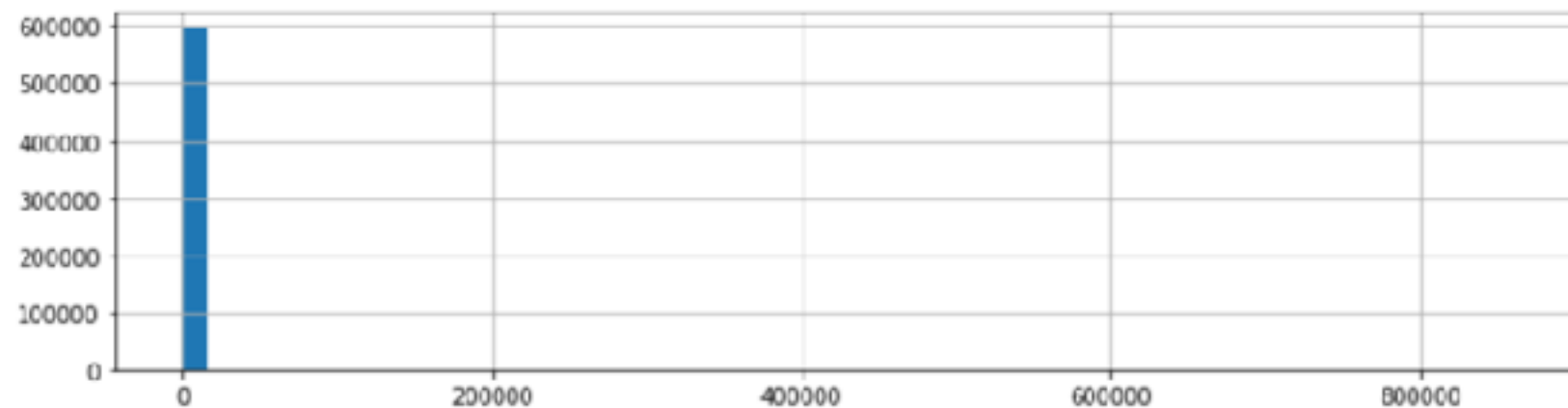


- 作者是否认证?
- 作者过去发布文章数?
- 阅读量?
- 发布日距今天数?
- 。 。 。



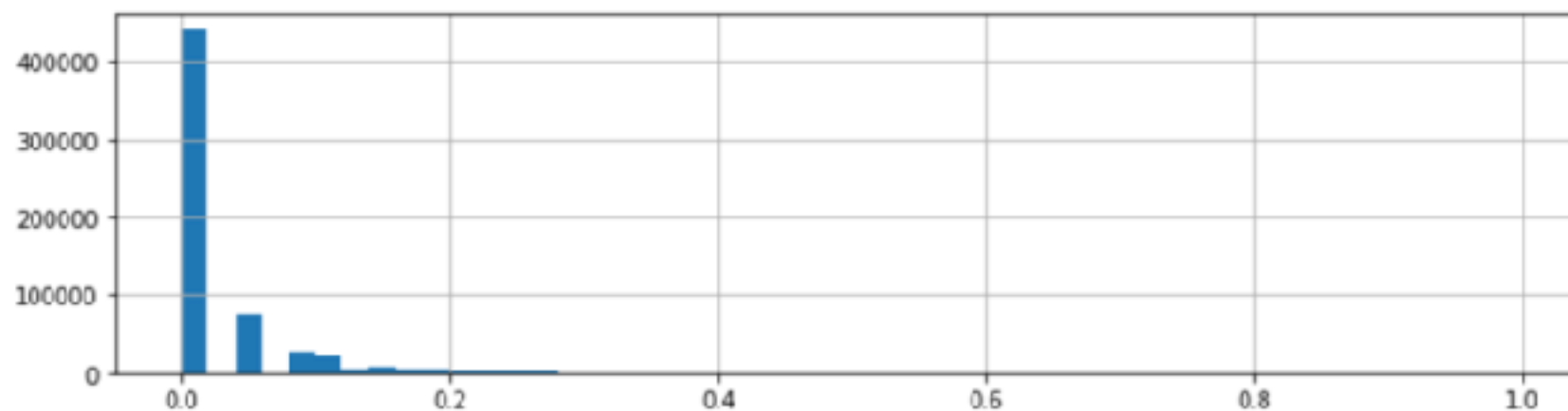
```
data["questions_count"].hist(figsize=(12, 3), bins=50)
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f74f79d0518>



```
data["questions_count_log10"] = np.log10(data["questions_count"]+1)/np.log10(data["questions_count"].max())  
data["questions_count_log10"].hist(figsize=(12, 3), bins=50)
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f74f77289b0>



$$F(x, y, z) = a_1 * x + a_2 * y + a_3 * z + \dots$$

```
"functions": [  
  {  
    "script_score": {  
      "script": {  
        "params": {  
          "bias": bias,  
          "doc_boost": doc_boost},  
        "inline": "doc['document_available'].value "  
          " * ( "  
          " Math.log10(params.bias + _score)"  
          " + (params.doc_boost * doc['document_score'].value)"  
          ")"  
        }  
      }  
    }  
  ]
```

有了



elasticsearch

是否

就算是

拥有  
春天



头发渐渐消失

# Learning To Rank

grade	keywords	article	titleScore	contentScore	upvotes	timestamp	feat_1
4	读书	a1	1.1	1.5	4	788714	10
2	读书	a2	2.5	3.1	2	238797	4
1	读书	a3	0.7	0.1	10	107413	5

LightGBM *dmlc* **XGBoost**  Yandex CatBoost

Query: 读书



**elasticsearch**

top N re-score

o19s/elasticsearch-learning-to-rank



```
"query": {
  "balabala"
},
"rescore": {
  "window_size": 1000,
  "query": {
    "rescore_query": {
      "sltr": {
        "params": {
          "keywords": query_content
        },
        "model": "model_XGBOOST_AUC_87",
      }
    }
  }
}
```

ensemble/blend  
online A/B test

```
"query": {
  "balabala"
},
"rescore": {
  "window_size": 1000,
  "query": {
    "rescore_query": {
      "sltr": {
        "params": {
          "keywords": query_content
        },
        "model": "model_RankLib_AUC_84",
      }
    }
  }
}
```

# 个性化?



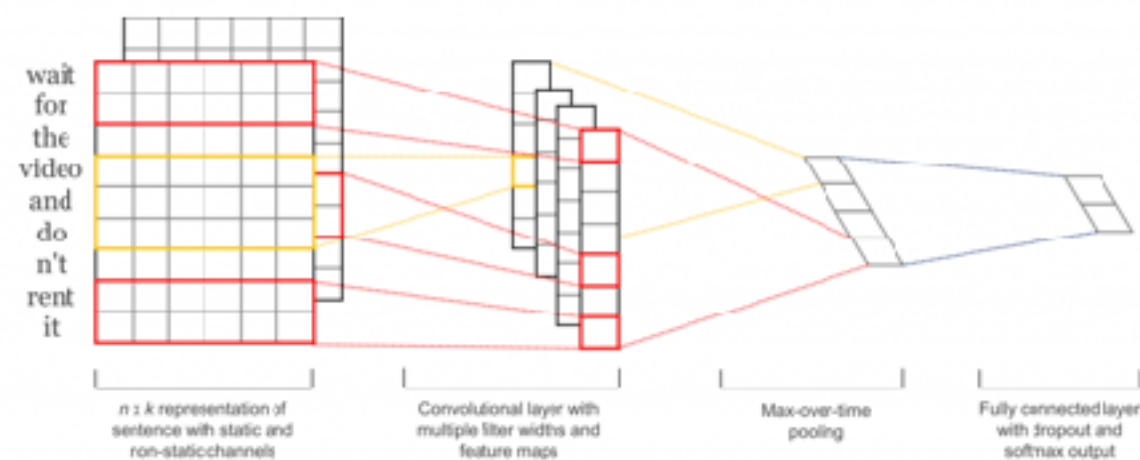
```
FILTER1 = {  
  "terms": {  
    "id": ["item1", "item3", "item10"]  
  }  
}
```

```
"functions": [  
  {  
    "filter": { FILTER1 },  
    "weight": 1.1  
  }  
]
```

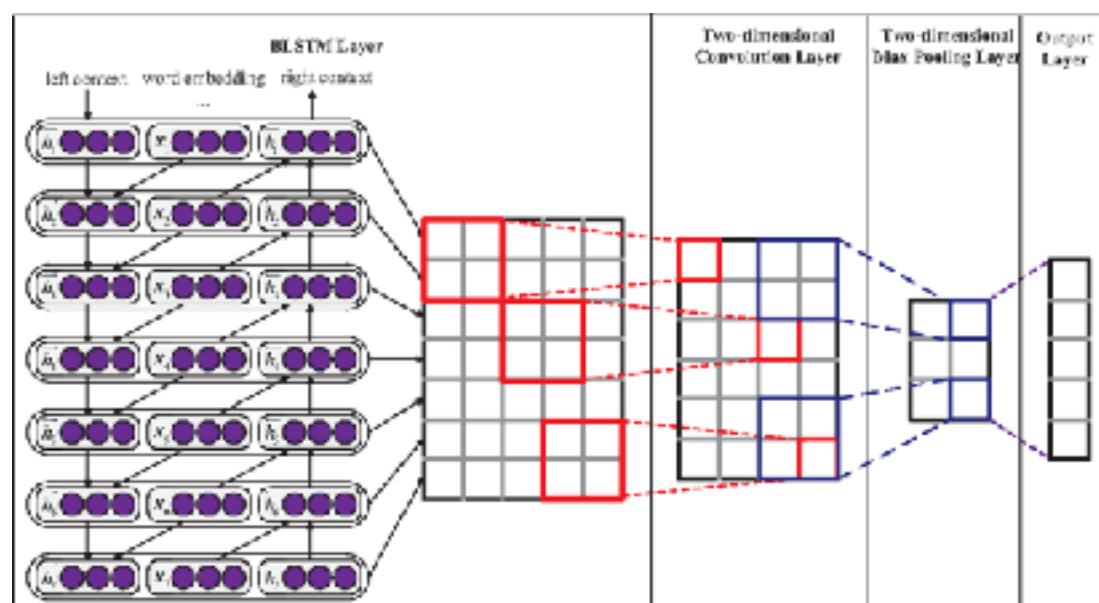
```
FILTER2 = {  
  "term": {  
    "users_who_may_like": "user01"  
  }  
}
```

```
FILTER3 = {  
  "terms": {  
    "related_items": ["item1", "item3"]  
  }  
}
```

这tm也太糙了...



## Text CNN

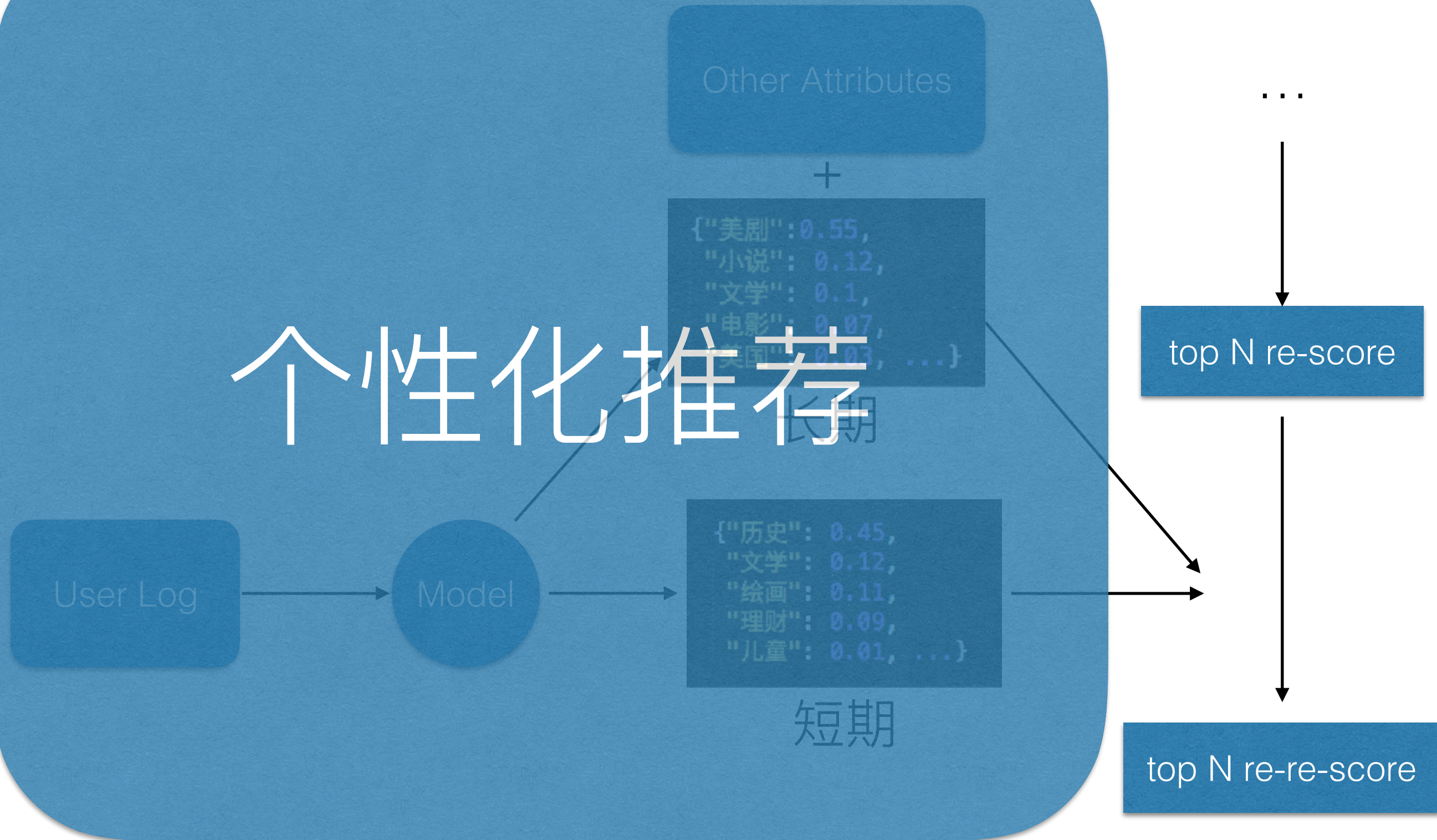


## BiLSTM+CNN

HAN, BiLSTM, Fasttext...



# 个性化推荐

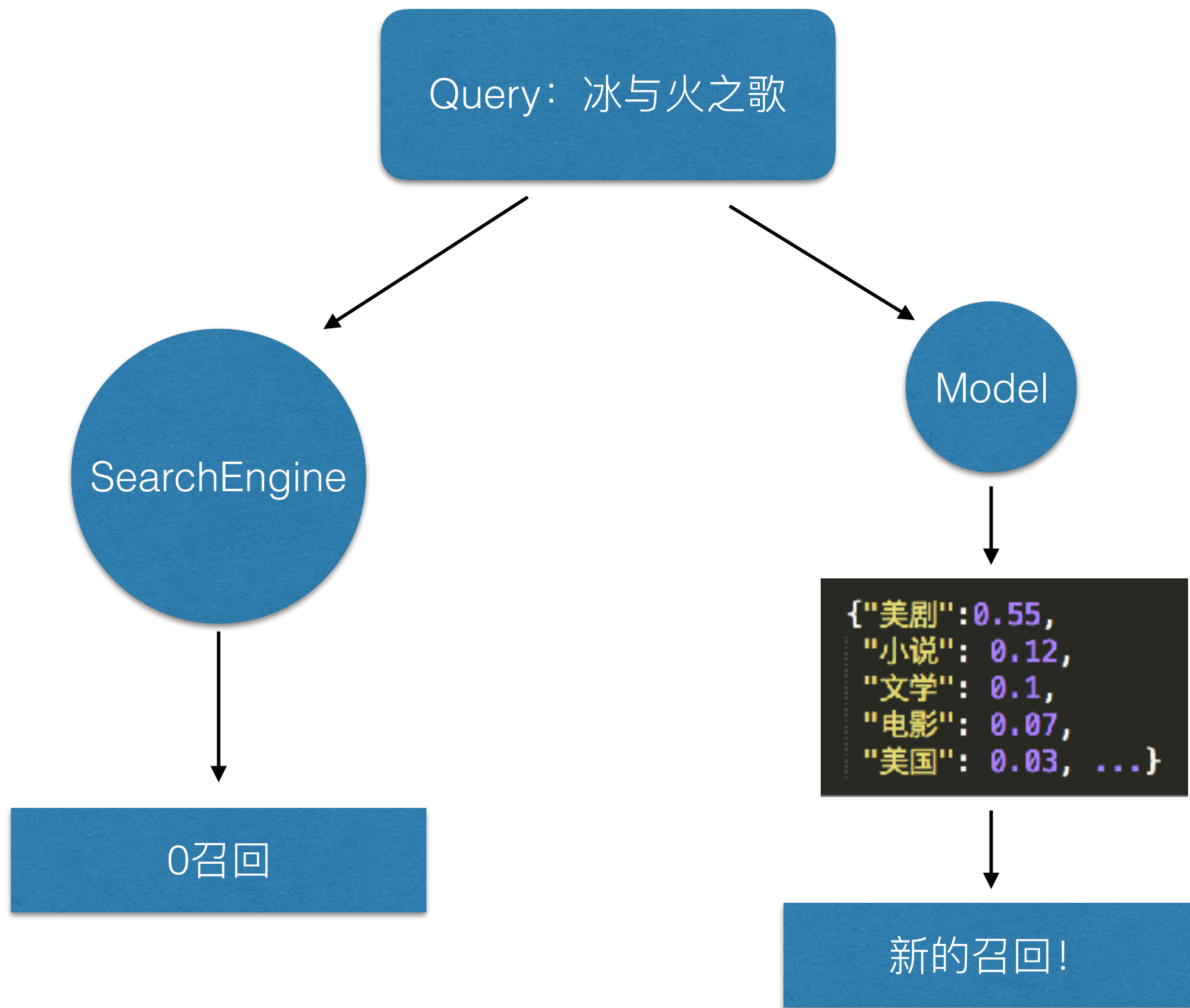




# 推荐和搜索



甚至还可以...



# End

detailed tutorial coming soon~  
twitter@ailurus1991



elastic  
中文社区

专业、垂直、纯粹的 Elastic 开源技术交流社区  
<https://elasticsearch.cn/>