

# 头条易重构——爬虫接口文档

---

头条易正式使用token: EZ6JOcuvrvdeKZmaBv48M7VVRQl4Wj

## 获取账号详情接口

- 用途：获取头条号的详细信息
- 参数：
  - tth\_id: URL 参数, tth 账号的 id
  - fields: 所需要返回的字段名, 不传则返回全部字段
- 举例：GET [https://ocean.tarsocial.com/api/toutiao/accounts/6036925076?access\\_token=EZ6JOcuvrvdeKZmaBv48M7VVRQl4Wj&fields=nickname&fields=classification&fields=rank&fields=score&fields=word\\_cloud&fields=similarity&fields=classification\\_prob&fields=description](https://ocean.tarsocial.com/api/toutiao/accounts/6036925076?access_token=EZ6JOcuvrvdeKZmaBv48M7VVRQl4Wj&fields=nickname&fields=classification&fields=rank&fields=score&fields=word_cloud&fields=similarity&fields=classification_prob&fields=description)
- 返回：{  
    "classification\_prob": {},  
    "classification": null,  
    "description": "星雅假日--特别邀请您和您的家人前往一个个被众神宠爱的地方。",  
    "nickname": "星雅假日",  
    "rank": 0,  
    "score": 385,  
    "similarity": null,  
    "word\_cloud": {  
        "事发时": 2,  
        "假日": 4,  
        "出发": 3,  
        "单身": 4,  
    }  
}

## 相似账号接口

- 举例：GET [http://ocean.tarsocial.com/api/toutiao/accounts/6036925076/similarity?access\\_token=EZ6JOcuvrvdeKZmaBv48M7VVRQl4Wj](http://ocean.tarsocial.com/api/toutiao/accounts/6036925076/similarity?access_token=EZ6JOcuvrvdeKZmaBv48M7VVRQl4Wj)
- 参数：
  - tth\_id: URL 参数, tth 账号的 id
- 返回：[  
    {

```
"avatar": "http://p2.pstatp.com/large/4535/6240930786",
"nickname": "乐活南京",
"tth_id": "4281149695"
},
{
"avatar": "http://p2.pstatp.com/large/2b6001fdd436e07a466",
"nickname": "南京助手",
"tth_id": "3349126634"
},
{
"avatar": "http://p1.pstatp.com/large/2182/4912626123",
"nickname": "龙虎网",
"tth_id": "3949749842"
},
{
"avatar": "http://p1.pstatp.com/large/8590/2594556838",
"nickname": "硬腿子",
"tth_id": "5545714477"
},
{
"avatar": "http://p3.pstatp.com/large/10062/7868356858",
"nickname": "门派",
"tth_id": "5791988176"
}
]
```

## 文章列表

- 支持任意账号，任意起止时间，任意分页
- 参数：
  - start\_time: 文章开始时间
  - end\_time: 文章结束时间
  - sort: 排序字段
  - fields: 逗号分隔的字段
  - page\_num: 页
  - page\_size: 每页包含记录条数

- 举例：http -p hbHB GET  
[http://ocean.tarsocial.com/api/toutiao/accounts/6036925076/articles?access\\_token=EZ6JOcuvrvdeKZmaBv48M7VVRQl4Wj&fields=post\\_time,read\\_num,comment\\_num,url,tth\\_id,title&start\\_time=2016-05-01T00:00:00&end\\_time=2016-06-01T00:00:00&sort=post\\_time:desc&page\\_size=1&page\\_num=1](http://ocean.tarsocial.com/api/toutiao/accounts/6036925076/articles?access_token=EZ6JOcuvrvdeKZmaBv48M7VVRQl4Wj&fields=post_time,read_num,comment_num,url,tth_id,title&start_time=2016-05-01T00:00:00&end_time=2016-06-01T00:00:00&sort=post_time:desc&page_size=1&page_num=1)

- 返回：[  
 {  
   "comment\_num": 3,  
   "post\_time": "2016-05-31T14:56:00",  
   "read\_num": 7856,  
   "title": "端午小长假来了，从南京出发一起去徽行古道吧，最全攻略奉上！",  
   "tth\_id": "6036925076",  
   "url": "<http://toutiao.com/item/6290743203490431489/>"  
 }  
 ]

返回tth\_id在start\_time与 end\_time之间按post\_time倒序的第1页，每页page\_size=1个文章的fields字段。

## 爆文接口

- 头条文章字段

```

"article_id": "6279168810503111170",
  "comment_num": 163,
  "content_type": 0,
  "copyright": 0,
  "description": "掐指一算，嫁过来苏家已经有五年多的时间了。苏家在我们这
里算是大户人家，当初我嫁过来的时候，很多亲戚朋友都羡慕得不得了，直说我这是上辈子修来的福分，
但是，这其中的苦楚，也只有我自己知道。在没结婚之前，我在一家陶瓷厂当文员，不夸张地说，我长得
很漂亮，当时厂里有很多小伙子都在追求我。",
  "insert_time": "2016-07-06T13:31:22",
  "post_time": "2016-04-30T16:55:00",
  "read_num": 389305,
  "title": "老公心疼我不让我出去工作，听到他跟小姑子的悄悄话我想哭",
  "tth_id": "5504388330",
  "update_time": null,
  "url": "http://toutiao.com/item/6279168810503111170/"

```

- 头条文章搜索接口接受字段

```
{
    'q': fields.Str(missing=None),
    'query': fields.Str(missing=None),
    'sort': fields.Str(missing=None),
    'fields': fields.Str(missing=None),
    'page_num': fields.Int(missing=1),
    'page_size': fields.Int(missing=10)
}
```

q 是一个字符串，表示搜索的字段和内容，es 风格

sort 是字符串，表示排序字段，es 风格

fields 是字符串，表示需要返回的字段，逗号分隔

page\_num 是整型，表示页数

page\_size 是整型，表示每页的条数

query 类型是一个 json 序列化的字符串，是一个综合查询字段，可以包括以上所有的内容，遇到复杂查询条件，q 无法满足时，必须使用 query，es DSL 风格。

详情参见 <https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl.html>

本次查询需要以下字段：query，page\_num，page\_size，以及 fields

**query** 如下，将三个变量替换，序号化赋值给 **query** 即可

```

{
  "query": { // 这个 query 要包含在内一起序列化
    "bool": {
      "must": [
        {
          "range": {
            "post_time": {
              "gte": "2017-05-30T00:00:00", //或者 "2017-05-30",为开始日期
              "lte": "2017-06-30T00:00:00" //或者 "2017-06-30",为结束日期
            }
          }
        },
        {
          "range" : {
            "read_num": {
              "gte": 1125000 // 为阅读数下限
            }
          }
        }
      ]
    }
  },
  "sort": {
    "read_num": "desc"
  }
}

```

page\_num, page\_size 按需传递, 默认为第一页, 每页10条。

fields为所需字段, 尽量传过来, 提高网络性能。

最后得到请求, url 为 <http://ocean.tarsocial.com/api/search/toutiao/articles>, json 为

```

{
  "query": query,
  "page_num": page_num,
  "page_size": page_size,
  "fields": fields
}

```

- 返回值字段

data: 列表, 列出所有命中的记录, 每条记录的格式如头条文章字段

first: 第一页页码, 一般为1

last: 最后一页页码

next: 下一页页码

page\_num: 当前页面

page\_size: 当前每页条数

prev: 上一页页码

total: 总记录数

## 添加文章监测及查看监测结果

- 举例： POST [http://ocean.tarsocial.com/api/toutiao/monitor/article?access\\_token=EZ6jOcuvrndeKZmaBv48M7VVRQl4Wj&monitor\\_days\\_from\\_post=3&url='http://www.toutiao.com/i6434052525963346434/'](http://ocean.tarsocial.com/api/toutiao/monitor/article?access_token=EZ6jOcuvrndeKZmaBv48M7VVRQl4Wj&monitor_days_from_post=3&url='http://www.toutiao.com/i6434052525963346434/')

返回 {"project\_id": project\_id}

- 举例： GET [http://ocean.tarsocial.com/api/toutiao/monitor/article/htujTn7zB/statistic?access\\_token=EZ6jOcuvrndeKZmaBv48M7VVRQl4Wj](http://ocean.tarsocial.com/api/toutiao/monitor/article/htujTn7zB/statistic?access_token=EZ6jOcuvrndeKZmaBv48M7VVRQl4Wj)

返回

```
{
  "abstract":
    "\u4fc4\u7f57\u65af\u4e0e\u7f8e\u56fd\u7684\u5173\u7cfb\u5251\u62e8\u5f29\u5f20\u4e4b\u9645\u5ff0c\u5317\u7ea6\u4e5f\u6ca1\u95f2\u7740\u5ff0c\u7adf\u4e5f\u64b8\u8d77\u8896\u5b50\u5e2e\u7f8e\u56fd\u603c\u4fc4\u7f57\u65af\u30026\u670821\u65e5\u5ff0c\u4fc4\u7f57\u65af\u536b\u661f\u901a\u8baf\u793e\u62a5\u9053\u62ab\u9732\u5ff0c\u4fc4\u7f57\u65af\u9632\u957f\u7ecd\u4f0a\u53e4\u5927\u5c06\u4e58\u673a\u98de\u5f80\u52a0\u91cc\u5b81\u683c\u52d2\u9014\u4e2d\u5ff0c\u5317\u7ea6\u4e00\u67b6F-16\u6218\u6597\u673a\u5728\u6ce2\u7f57\u7684\u6d77\u516c\u6d77\u4e0a\u7a7a\u66fe\u8bd5\u56fe\u63a5\u8fd1\u7ecd\u4f0a\u53e4\u7684\u5ea7\u673a\u5ff0c\u4f46\u88ab\u62a4\u822a\u7684\u4fc4\u7f57\u65af\u82cf-27\u6218\u6597\u673a\u53ca\u65f6\u9a71\u8d76\u5f00\u3002",
  "article_id": "6434052525963346434",
  "capture_num": 0,
  "comment_num": 349,
  "content": "",
  "create_time": "2017-06-28T18:55:52",
  "data_time": "2017-07-01T18:05:23",
  "end_time": "2017-06-29T18:55:52",
  "id": 117330,
  "interval": 3600,
  "like_num": 0,
  "media_id": "6579376245",
  "media_name": "\u5927\u56fd\u4e4b\u7ffc",
  "media_type": 3,
  "mp_idx": null,
  "post_time": "2017-06-21T19:29:00",
  "project_id": "htujTn7zB",
  "read_num": 556798,
  "real_comment_num": 349,
  "real_like_num": 0,
```

```
"real_read_num": 556798,
"real_repost_num": 0,
"repost_num": 0,
"score": 0.0,
"start_time": "2017-06-21T19:29:00",
"title":
"\u7a81\u53d1\u5f0c\u5317\u7ea6F16\u6218\u673a\u76ef\u4e0a\u4fc4\u7f57\u65af\u9632\u9
57\u4e13\u673a\u5f01\u5e78\u597d\u4fc4\u82cf27\u53ca\u65f6\u51fa\u73b0"
}
```

## 新增头条号（用于平台同步新流量主）

POST [http://ocean.tarsocial.com/api/toutiao/accounts?access\\_token=EZ6JOcuvrvdeKZmaBv48M7VVRQl4Wj&tth\\_id={tth\\_id}](http://ocean.tarsocial.com/api/toutiao/accounts?access_token=EZ6JOcuvrvdeKZmaBv48M7VVRQl4Wj&tth_id={tth_id})

成功则返回 200 内容为 json 格式，为基本数据。

失败返回非200，内容为 json 格式，为异常信息。