

1、算法介绍

relevance score 算法，简单来说就是计算出，一个索引中的文本，与索引中的文本，与搜索文本，他们之间的关联匹配程度

ElasticSearch 使用的是 term frequency/inverse document frequency 算法，简称 TF/IDF 算法

Term frequency: 搜索文本中的**各个词条**在 field 文本中出现了多少次，**出现次数越多，就越相关**

例如：搜索请求 “hello world”

doc1: hello, today is very good

doc2: hi world, how are you

Inverse document frequency: 搜索文本中的各个词条在整个索引的**所有文档**中出现了多少次，**出现的次数越多，就越不相关**

以上的搜索示例，假如在 index 中有1万条 document，hello 这个单词在所有的 document 中，一共出现了 1000次，world 这个单词在所有的 document 中，一共出现了 100次

因此根据 TF/IDF 算法 doc2 更加相关

Field-length norm: 和 field 长度有关，如果搜索关键字在 document 都有出现，field 越长相关性越差，field 越短相关性越高

例如：搜索请求 “hello world”

doc1: {"title": "hello article", "content": "...一万的字母..."}

doc2: {"title": "my article", "content": "...一万个字母..."}

如果 hello 和 world 在整个 index 的出现次数一样，也就是排除 inverse document frequency 的影响，doc1 的相关度更高

2、如何查看 es 是怎么计算 score 的

1 使用以下命令 `explain = true`

```
2 GET /website/article/_search?explain=true
3 {
4   "query": {
5     "match": {
6       "content": " this is my first"
7     }
8   }
9 }
```

返回结果（比较长。。。。，就是上面那三个计算原则的综合结果）

```
1 {
2   "took": 12,
3   "timed_out": false,
4   "_shards": {
5     "total": 5,
6     "successful": 5,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": 3,
12    "max_score": 1.1507283,
13    "hits": [
14      {
15        "_shard": "[website][0]",
16        "_node": "fzY8ok7NSumyQK66Foz6Tw",
17        "_index": "website",
18        "_type": "article",
19        "_id": "DXS1s2EBdXiUQYE3zPcm",
20        "_score": 1.1507283,
21        "_source": {
22          "title": "first article",
23          "content": "this is my first article in elasticsearch",
24          "post_date": "2018-02-20",
25          "author": "daiwei"
26        },
27        "_explanation": {
28          "value": 1.1507283,
29          "description": "sum of:",
30          "details": [
31            {
32              "value": 0.2876821,
33              "description": "weight(content:this in 0) [PerFieldSimilarity],
result of:",
```

```

"details": [
  {
    "value": 0.2876821,
    "description": "score(doc=0,freq=1.0 = termFreq=1.0\n),
product of:",
    "details": [
      {
        "value": 0.2876821,
        "description": "idf, computed as log(1 + (docCount -
docFreq + 0.5) / (docFreq + 0.5)) from:",
        "details": [
          {
            "value": 1,
            "description": "docFreq",
            "details": []
          },
          {
            "value": 1,
            "description": "docCount",
            "details": []
          }
        ]
      },
      {
        "value": 1,
        "description": "tfNorm, computed as (freq * (k1 + 1)) /
(freq + k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
        "details": [
          {
            "value": 1,
            "description": "termFreq=1.0",
            "details": []
          },
          {
            "value": 1.2,
            "description": "parameter k1",
            "details": []
          },
          {
            "value": 0.75,
            "description": "parameter b",
            "details": []
          },
          {
            "value": 7,
            "description": "avgFieldLength",

```

```

77         "details": []
78     },
79     {
80         "value": 7,
81         "description": "fieldLength",
82         "details": []
83     }
84 ]
85 }
86 ]
87 }
88 ]
89 },
90 {
91     "value": 0.2876821,
92     "description": "weight(content:is in 0) [PerFieldSimilarity],
result of:",
93     "details": [
94         {
95             "value": 0.2876821,
96             "description": "score(doc=0,freq=1.0 = termFreq=1.0\n),
product of:",
97             "details": [
98                 {
99                     "value": 0.2876821,
100                     "description": "idf, computed as log(1 + (docCount -
docFreq + 0.5) / (docFreq + 0.5)) from:",
101                     "details": [
102                         {
103                             "value": 1,
104                             "description": "docFreq",
105                             "details": []
106                         },
107                         {
108                             "value": 1,
109                             "description": "docCount",
110                             "details": []
111                         }
112                     ]
113                 },
114                 {
115                     "value": 1,
116                     "description": "tfNorm, computed as (freq * (k1 + 1)) /
(freq + k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
117                     "details": [
118                         {

```

```

119         "value": 1,
120         "description": "termFreq=1.0",
121         "details": []
122     },
123     {
124         "value": 1.2,
125         "description": "parameter k1",
126         "details": []
127     },
128     {
129         "value": 0.75,
130         "description": "parameter b",
131         "details": []
132     },
133     {
134         "value": 7,
135         "description": "avgFieldLength",
136         "details": []
137     },
138     {
139         "value": 7,
140         "description": "fieldLength",
141         "details": []
142     }
143 ]
144 }
145 ]
146 }
147 ]
148 },
149 {
150     "value": 0.2876821,
151     "description": "weight(content:my in 0) [PerFieldSimilarity],
result of:",
152     "details": [
153         {
154             "value": 0.2876821,
155             "description": "score(doc=0,freq=1.0 = termFreq=1.0\n),
product of:",
156             "details": [
157                 {
158                     "value": 0.2876821,
159                     "description": "idf, computed as log(1 + (docCount -
docFreq + 0.5) / (docFreq + 0.5)) from:",
160                     "details": [
161                         {

```

```

162         "value": 1,
163         "description": "docFreq",
164         "details": []
165     },
166     {
167         "value": 1,
168         "description": "docCount",
169         "details": []
170     }
171 ]
172 },
173 {
174     "value": 1,
175     "description": "tfNorm, computed as (freq * (k1 + 1)) /
(freq + k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
176     "details": [
177         {
178             "value": 1,
179             "description": "termFreq=1.0",
180             "details": []
181         },
182         {
183             "value": 1.2,
184             "description": "parameter k1",
185             "details": []
186         },
187         {
188             "value": 0.75,
189             "description": "parameter b",
190             "details": []
191         },
192         {
193             "value": 7,
194             "description": "avgFieldLength",
195             "details": []
196         },
197         {
198             "value": 7,
199             "description": "fieldLength",
200             "details": []
201         }
202     ]
203 }
204 ]
205 }
206 ]

```

```

    },
    {
      "value": 0.2876821,
      "description": "weight(content:first in 0) [PerFieldSimilarity],
result of:",
      "details": [
        {
          "value": 0.2876821,
          "description": "score(doc=0,freq=1.0 = termFreq=1.0\n),
product of:",
          "details": [
            {
              "value": 0.2876821,
              "description": "idf, computed as log(1 + (docCount -
docFreq + 0.5) / (docFreq + 0.5)) from:",
              "details": [
                {
                  "value": 1,
                  "description": "docFreq",
                  "details": []
                },
                {
                  "value": 1,
                  "description": "docCount",
                  "details": []
                }
              ]
            },
            {
              "value": 1,
              "description": "tfNorm, computed as (freq * (k1 + 1)) /
(freq + k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
              "details": [
                {
                  "value": 1,
                  "description": "termFreq=1.0",
                  "details": []
                },
                {
                  "value": 1.2,
                  "description": "parameter k1",
                  "details": []
                },
                {
                  "value": 0.75,
                  "description": "parameter b",

```

```

249         "details": []
250     },
251     {
252         "value": 7,
253         "description": "avgFieldLength",
254         "details": []
255     },
256     {
257         "value": 7,
258         "description": "fieldLength",
259         "details": []
260     }
261 ]
262 }
263 ]
264 }
265 ]
266 }
267 ]
268 }
269 },
270 {
271     "_shard": "[website][3]",
272     "_node": "fzY8ok7NSumyQK66Foz6Tw",
273     "_index": "website",
274     "_type": "article",
275     "_id": "DnS1s2EBdXiUQYE31PdG",
276     "_score": 0.8630463,
277     "_source": {
278         "title": "second article",
279         "content": "this is my second article in elasticsearch",
280         "post_date": "2018-03-20",
281         "author": "zzq"
282     },
283     "_explanation": {
284         "value": 0.8630463,
285         "description": "sum of:",
286         "details": [
287             {
288                 "value": 0.2876821,
289                 "description": "weight(content:this in 0) [PerFieldSimilarity],
result of:",
290                 "details": [
291                     {
292                         "value": 0.2876821,
293                         "description": "score(doc=0,freq=1.0 = termFreq=1.0\n),

```



```

product of:",
294         "details": [
295             {
296                 "value": 0.2876821,
297                 "description": "idf, computed as  $\log(1 + (\text{docCount} - \text{docFreq} + 0.5) / (\text{docFreq} + 0.5))$  from:",
298                 "details": [
299                     {
300                         "value": 1,
301                         "description": "docFreq",
302                         "details": []
303                     },
304                     {
305                         "value": 1,
306                         "description": "docCount",
307                         "details": []
308                     }
309                 ]
310             },
311             {
312                 "value": 1,
313                 "description": "tfNorm, computed as  $(\text{freq} * (\text{k1} + 1)) / (\text{freq} + \text{k1} * (1 - \text{b} + \text{b} * \text{fieldLength} / \text{avgFieldLength}))$  from:",
314                 "details": [
315                     {
316                         "value": 1,
317                         "description": "termFreq=1.0",
318                         "details": []
319                     },
320                     {
321                         "value": 1.2,
322                         "description": "parameter k1",
323                         "details": []
324                     },
325                     {
326                         "value": 0.75,
327                         "description": "parameter b",
328                         "details": []
329                     },
330                     {
331                         "value": 7,
332                         "description": "avgFieldLength",
333                         "details": []
334                     },
335                     {
336                         "value": 7,

```

```

337         "description": "fieldLength",
338         "details": []
339     }
340 ]
341 }
342 ]
343 }
344 ]
345 },
346 {
347     "value": 0.2876821,
348     "description": "weight(content:is in 0) [PerFieldSimilarity],
result of:",
349     "details": [
350     {
351         "value": 0.2876821,
352         "description": "score(doc=0,freq=1.0 = termFreq=1.0\n),
product of:",
353         "details": [
354         {
355             "value": 0.2876821,
356             "description": "idf, computed as log(1 + (docCount -
docFreq + 0.5) / (docFreq + 0.5)) from:",
357             "details": [
358             {
359                 "value": 1,
360                 "description": "docFreq",
361                 "details": []
362             },
363             {
364                 "value": 1,
365                 "description": "docCount",
366                 "details": []
367             }
368         ]
369     },
370     {
371         "value": 1,
372         "description": "tfNorm, computed as (freq * (k1 + 1)) /
(freq + k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
373         "details": [
374         {
375             "value": 1,
376             "description": "termFreq=1.0",
377             "details": []
378         },

```

```

379         {
380             "value": 1.2,
381             "description": "parameter k1",
382             "details": []
383         },
384         {
385             "value": 0.75,
386             "description": "parameter b",
387             "details": []
388         },
389         {
390             "value": 7,
391             "description": "avgFieldLength",
392             "details": []
393         },
394         {
395             "value": 7,
396             "description": "fieldLength",
397             "details": []
398         }
399     ]
400 }
401 ]
402 }
403 ]
404 },
405 {
406     "value": 0.2876821,
407     "description": "weight(content:my in 0) [PerFieldSimilarity],
result of:",
408     "details": [
409         {
410             "value": 0.2876821,
411             "description": "score(doc=0,freq=1.0 = termFreq=1.0\n),
product of:",
412             "details": [
413                 {
414                     "value": 0.2876821,
415                     "description": "idf, computed as log(1 + (docCount -
docFreq + 0.5) / (docFreq + 0.5)) from:",
416                     "details": [
417                         {
418                             "value": 1,
419                             "description": "docFreq",
420                             "details": []
421                         },

```

```

422         {
423             "value": 1,
424             "description": "docCount",
425             "details": []
426         }
427     ],
428 },
429 {
430     "value": 1,
431     "description": "tfNorm, computed as (freq * (k1 + 1)) /
(freq + k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
432     "details": [
433         {
434             "value": 1,
435             "description": "termFreq=1.0",
436             "details": []
437         },
438         {
439             "value": 1.2,
440             "description": "parameter k1",
441             "details": []
442         },
443         {
444             "value": 0.75,
445             "description": "parameter b",
446             "details": []
447         },
448         {
449             "value": 7,
450             "description": "avgFieldLength",
451             "details": []
452         },
453         {
454             "value": 7,
455             "description": "fieldLength",
456             "details": []
457         }
458     ]
459 }
460 ]
461 }
462 ]
463 }
464 ]
465 }
466 },

```

```

467 {
468   "_shard": "[website][4]",
469   "_node": "fzY8ok7NSumyQK66Foz6Tw",
470   "_index": "website",
471   "_type": "article",
472   "_id": "D3S1s2EBdXiUQYE33PfR",
473   "_score": 0.8630463,
474   "_source": {
475     "title": "third article",
476     "content": "this is my third article in elasticsearch",
477     "post_date": "2018-04-20",
478     "author": "zzw"
479   },
480   "_explanation": {
481     "value": 0.8630463,
482     "description": "sum of:",
483     "details": [
484       {
485         "value": 0.2876821,
486         "description": "weight(content:this in 0) [PerFieldSimilarity],
result of:",
487         "details": [
488           {
489             "value": 0.2876821,
490             "description": "score(doc=0,freq=1.0 = termFreq=1.0\n),
product of:",
491             "details": [
492               {
493                 "value": 0.2876821,
494                 "description": "idf, computed as log(1 + (docCount -
docFreq + 0.5) / (docFreq + 0.5)) from:",
495                 "details": [
496                   {
497                     "value": 1,
498                     "description": "docFreq",
499                     "details": []
500                   },
501                   {
502                     "value": 1,
503                     "description": "docCount",
504                     "details": []
505                   }
506                 ]
507               },
508               {
509                 "value": 1,

```

```

510         "description": "tfNorm, computed as (freq * (k1 + 1)) /
(freq + k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
511         "details": [
512             {
513                 "value": 1,
514                 "description": "termFreq=1.0",
515                 "details": []
516             },
517             {
518                 "value": 1.2,
519                 "description": "parameter k1",
520                 "details": []
521             },
522             {
523                 "value": 0.75,
524                 "description": "parameter b",
525                 "details": []
526             },
527             {
528                 "value": 7,
529                 "description": "avgFieldLength",
530                 "details": []
531             },
532             {
533                 "value": 7,
534                 "description": "fieldLength",
535                 "details": []
536             }
537         ]
538     }
539 ]
540 }
541 ]
542 },
543 {
544     "value": 0.2876821,
545     "description": "weight(content:is in 0) [PerFieldSimilarity],
result of:",
546     "details": [
547         {
548             "value": 0.2876821,
549             "description": "score(doc=0,freq=1.0 = termFreq=1.0\n),
product of:",
550             "details": [
551                 {
552                     "value": 0.2876821,

```

```

553         "description": "idf, computed as  $\log(1 + (\text{docCount} - \text{docFreq} + 0.5) / (\text{docFreq} + 0.5))$  from:",
554         "details": [
555             {
556                 "value": 1,
557                 "description": "docFreq",
558                 "details": []
559             },
560             {
561                 "value": 1,
562                 "description": "docCount",
563                 "details": []
564             }
565         ],
566     },
567     {
568         "value": 1,
569         "description": "tfNorm, computed as  $(\text{freq} * (\text{k1} + 1)) / (\text{freq} + \text{k1} * (1 - \text{b} + \text{b} * \text{fieldLength} / \text{avgFieldLength}))$  from:",
570         "details": [
571             {
572                 "value": 1,
573                 "description": "termFreq=1.0",
574                 "details": []
575             },
576             {
577                 "value": 1.2,
578                 "description": "parameter k1",
579                 "details": []
580             },
581             {
582                 "value": 0.75,
583                 "description": "parameter b",
584                 "details": []
585             },
586             {
587                 "value": 7,
588                 "description": "avgFieldLength",
589                 "details": []
590             },
591             {
592                 "value": 7,
593                 "description": "fieldLength",
594                 "details": []
595             }
596         ]

```

```

597         }
598     ]
599 }
600 ]
601 },
602 {
603     "value": 0.2876821,
604     "description": "weight(content:my in 0) [PerFieldSimilarity],
result of:",
605     "details": [
606         {
607             "value": 0.2876821,
608             "description": "score(doc=0,freq=1.0 = termFreq=1.0\n),
product of:",
609             "details": [
610                 {
611                     "value": 0.2876821,
612                     "description": "idf, computed as log(1 + (docCount -
docFreq + 0.5) / (docFreq + 0.5)) from:",
613                     "details": [
614                         {
615                             "value": 1,
616                             "description": "docFreq",
617                             "details": []
618                         },
619                         {
620                             "value": 1,
621                             "description": "docCount",
622                             "details": []
623                         }
624                     ]
625                 },
626                 {
627                     "value": 1,
628                     "description": "tfNorm, computed as (freq * (k1 + 1)) /
(freq + k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
629                     "details": [
630                         {
631                             "value": 1,
632                             "description": "termFreq=1.0",
633                             "details": []
634                         },
635                         {
636                             "value": 1.2,
637                             "description": "parameter k1",
638                             "details": []

```



```

639         },
640         {
641             "value": 0.75,
642             "description": "parameter b",
643             "details": []
644         },
645         {
646             "value": 7,
647             "description": "avgFieldLength",
648             "details": []
649         },
650         {
651             "value": 7,
652             "description": "fieldLength",
653             "details": []
654         }
655     ]
656 }
657 ]
658 }
659 ]
660 }
661 ]
662 }
663 }
664 ]
665 }
666 }

```

3、分析一个document 是如何被匹配上的

```

1 GET website/article/DXS1s2EBdXiUQYE3zPcm/_explain
2 {
3     "query": {
4         "match": {
5             "content": " this is my first"
6         }
7     }
8 }

```

结果如下:

```
1 {
2   "_index": "website",
3   "_type": "article",
4   "_id": "DXS1s2EBdXiUQYE3zPcm",
5   "matched": true,
6   "explanation": {
7     "value": 1.1507283,
8     "description": "sum of:",
9     "details": [
10      {
11        "value": 0.2876821,
12        "description": "weight(content:this in 0) [PerFieldSimilarity], result
of:",
13        "details": [
14          {
15            "value": 0.2876821,
16            "description": "score(doc=0,freq=1.0 = termFreq=1.0\n), product
of:",
17            "details": [
18              {
19                "value": 0.2876821,
20                "description": "idf, computed as log(1 + (docCount - docFreq +
0.5) / (docFreq + 0.5)) from:",
21                "details": [
22                  {
23                    "value": 1,
24                    "description": "docFreq",
25                    "details": []
26                  },
27                  {
28                    "value": 1,
29                    "description": "docCount",
30                    "details": []
31                  }
32                ]
33              },
34              {
35                "value": 1,
36                "description": "tfNorm, computed as (freq * (k1 + 1)) / (freq
+ k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
37                "details": [
38                  {
39                    "value": 1,
```

```

40         "description": "termFreq=1.0",
41         "details": []
42     },
43     {
44         "value": 1.2,
45         "description": "parameter k1",
46         "details": []
47     },
48     {
49         "value": 0.75,
50         "description": "parameter b",
51         "details": []
52     },
53     {
54         "value": 7,
55         "description": "avgFieldLength",
56         "details": []
57     },
58     {
59         "value": 7,
60         "description": "fieldLength",
61         "details": []
62     }
63 ]
64 }
65 ]
66 }
67 ]
68 },
69 {
70     "value": 0.2876821,
71     "description": "weight(content:is in 0) [PerFieldSimilarity], result
of:",
72     "details": [
73         {
74             "value": 0.2876821,
75             "description": "score(doc=0,freq=1.0 = termFreq=1.0\n), product
of:",
76             "details": [
77                 {
78                     "value": 0.2876821,
79                     "description": "idf, computed as log(1 + (docCount - docFreq +
0.5) / (docFreq + 0.5)) from:",
80                     "details": [
81                         {
82                             "value": 1,

```

```

83         "description": "docFreq",
84         "details": []
85     },
86     {
87         "value": 1,
88         "description": "docCount",
89         "details": []
90     }
91 ]
92 },
93 {
94     "value": 1,
95     "description": "tfNorm, computed as (freq * (k1 + 1)) / (freq
+ k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
96     "details": [
97         {
98             "value": 1,
99             "description": "termFreq=1.0",
100             "details": []
101         },
102         {
103             "value": 1.2,
104             "description": "parameter k1",
105             "details": []
106         },
107         {
108             "value": 0.75,
109             "description": "parameter b",
110             "details": []
111         },
112         {
113             "value": 7,
114             "description": "avgFieldLength",
115             "details": []
116         },
117         {
118             "value": 7,
119             "description": "fieldLength",
120             "details": []
121         }
122     ]
123 }
124 ]
125 }
126 ]
127 },

```

```

128     {
129         "value": 0.2876821,
130         "description": "weight(content:my in 0) [PerFieldSimilarity], result
of:",
131         "details": [
132             {
133                 "value": 0.2876821,
134                 "description": "score(doc=0,freq=1.0 = termFreq=1.0\n), product
of:",
135                 "details": [
136                     {
137                         "value": 0.2876821,
138                         "description": "idf, computed as log(1 + (docCount - docFreq +
0.5) / (docFreq + 0.5)) from:",
139                         "details": [
140                             {
141                                 "value": 1,
142                                 "description": "docFreq",
143                                 "details": []
144                             },
145                             {
146                                 "value": 1,
147                                 "description": "docCount",
148                                 "details": []
149                             }
150                         ]
151                     },
152                     {
153                         "value": 1,
154                         "description": "tfNorm, computed as (freq * (k1 + 1)) / (freq
+ k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
155                         "details": [
156                             {
157                                 "value": 1,
158                                 "description": "termFreq=1.0",
159                                 "details": []
160                             },
161                             {
162                                 "value": 1.2,
163                                 "description": "parameter k1",
164                                 "details": []
165                             },
166                             {
167                                 "value": 0.75,
168                                 "description": "parameter b",
169                                 "details": []

```

```

170         },
171         {
172             "value": 7,
173             "description": "avgFieldLength",
174             "details": []
175         },
176         {
177             "value": 7,
178             "description": "fieldLength",
179             "details": []
180         }
181     ]
182 }
183 ]
184 }
185 ]
186 },
187 {
188     "value": 0.2876821,
189     "description": "weight(content:first in 0) [PerFieldSimilarity],
result of:",
190     "details": [
191         {
192             "value": 0.2876821,
193             "description": "score(doc=0,freq=1.0 = termFreq=1.0\n), product
of:",
194             "details": [
195                 {
196                     "value": 0.2876821,
197                     "description": "idf, computed as log(1 + (docCount - docFreq +
0.5) / (docFreq + 0.5)) from:",
198                     "details": [
199                         {
200                             "value": 1,
201                             "description": "docFreq",
202                             "details": []
203                         },
204                         {
205                             "value": 1,
206                             "description": "docCount",
207                             "details": []
208                         }
209                     ]
210                 },
211                 {
212                     "value": 1,

```

```

213         "description": "tfNorm, computed as (freq * (k1 + 1)) / (freq
+ k1 * (1 - b + b * fieldLength / avgFieldLength)) from:",
214         "details": [
215             {
216                 "value": 1,
217                 "description": "termFreq=1.0",
218                 "details": []
219             },
220             {
221                 "value": 1.2,
222                 "description": "parameter k1",
223                 "details": []
224             },
225             {
226                 "value": 0.75,
227                 "description": "parameter b",
228                 "details": []
229             },
230             {
231                 "value": 7,
232                 "description": "avgFieldLength",
233                 "details": []
234             },
235             {
236                 "value": 7,
237                 "description": "fieldLength",
238                 "details": []
239             }
240         ]
241     }
242 ]
243 }
244 ]
245 }
246 ]
247 }
248 }

```