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

```
使用以下命令 explain = true
```

```
GET /website/article/_search?explain=true
{
  "query": {
    "match": {
      "content": " this is my first"
    }
  }
}
```

返回结果（比较长。。。。，就是上面那三个计算原则的综合结果）

```
{
  "took": 12,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": 3,
    "max_score": 1.1507283,
    "hits": [
      {
        "_shard": "[website][0]",
        "_node": "fzY8ok7NSumyQK66Foz6Tw",
        "_index": "website",
        "_type": "article",
        "_id": "DXS1s2EBdXiUQYE3zPcm",
        "_score": 1.1507283,
        "_source": {
          "title": "first article",
          "content": "this is my first article in elasticsearch",
          "post_date": "2018-02-20",
          "author": "daiwei"
        },
        "_explanation": {
          "value": 1.1507283,
          "description": "sum of:",
          "details": [
            {
              "value": 0.2876821,
              "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",
```

```
                        "details": []
                      },
                      {
                        "value": 7,
                        "description": "fieldLength",
                        "details": []
                      }
                    ]
                  }
                ]
              }
            ]
          },
          {
            "value": 0.2876821,
            "description": "weight(content:is 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",
                        "details": []
                      },
                      {
                        "value": 7,
                        "description": "fieldLength",
                        "details": []
                      }
                    ]
                  }
                ]
              }
            ]
          },
          {
            "value": 0.2876821,
            "description": "weight(content:my 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",
                        "details": []
                      },
                      {
                        "value": 7,
                        "description": "fieldLength",
                        "details": []
                      }
                    ]
                  }
                ]
              }
            ]
```

```
            },
            {
              "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",
```

```
                        "details": []
                      },
                      {
                        "value": 7,
                        "description": "avgFieldLength",
                        "details": []
                      },
                      {
                        "value": 7,
                        "description": "fieldLength",
                        "details": []
                      }
                    ]
                  }
                ]
              }
            ]
          }
        },
        {
          "_shard": "[website][3]",
          "_node": "fzY8ok7NSumyQK66Foz6Tw",
          "_index": "website",
          "_type": "article",
          "_id": "DnS1s2EBdXiUQYE31Pdg",
          "_score": 0.8630463,
          "_source": {
            "title": "second article",
            "content": "this is my second article in elasticsearch",
            "post_date": "2018-03-20",
            "author": "zzq"
          },
          "_explanation": {
            "value": 0.8630463,
            "description": "sum of:",
            "details": [
              {
                "value": 0.2876821,
                "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",
                          "details": []
                        },
                        {
                          "value": 7,
```

```
                        "description": "fieldLength",
                        "details": []
                      }
                    ]
                  }
                ]
              }
            ]
          },
          {
            "value": 0.2876821,
            "description": "weight(content:is 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",
                          "details": []
                        },
                        {
                          "value": 7,
                          "description": "fieldLength",
                          "details": []
                        }
                      ]
                    }
                  ]
                }
              ]
            },
            {
              "value": 0.2876821,
              "description": "weight(content:my 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",
                          "details": []
                        },
                        {
                          "value": 7,
                          "description": "fieldLength",
                          "details": []
                        }
                      ]
                    }
                  ]
                }
              ]
            }
          ]
        }
      },
```

```
      {
        "_shard": "[website][4]",
        "_node": "fzY8ok7NSumyQK66Foz6Tw",
        "_index": "website",
        "_type": "article",
        "_id": "D3S1s2EBdXiUQYE33PfR",
        "_score": 0.8630463,
        "_source": {
          "title": "third article",
          "content": "this is my third article in elasticsearch",
          "post_date": "2018-04-20",
          "author": "zzw"
        },
        "_explanation": {
          "value": 0.8630463,
          "description": "sum of:",
          "details": [
            {
              "value": 0.2876821,
              "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",
                        "details": []
                      },
                      {
                        "value": 7,
                        "description": "fieldLength",
                        "details": []
                      }
                    ]
                  }
                ]
              }
            ]
          },
          {
            "value": 0.2876821,
            "description": "weight(content:is 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",
                        "details": []
                      },
                      {
                        "value": 7,
                        "description": "fieldLength",
                        "details": []
                      }
                    ]
```

```
                    }
                  ]
                }
              ]
            },
            {
              "value": 0.2876821,
              "description": "weight(content:my 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",
                          "details": []
                        },
                        {
                          "value": 7,
                          "description": "fieldLength",
                          "details": []
                        }
                      ]
                    }
                  ]
                }
              ]
            }
          ]
        }
      }
    ]
  }
}
```

3、分析一个document 是如何被匹配上的

```
GET website/article/DXS1s2EBdXiUQYE3zPcm/_explain
{
  "query": {
    "match": {
      "content": " this is my first"
    }
  }
}
```

结果如下：


```
{
  "_index": "website",
  "_type": "article",
  "_id": "DXS1s2EBdXiUQYE3zPcm",
  "matched": true,
  "explanation": {
    "value": 1.1507283,
    "description": "sum of:",
    "details": [
      {
        "value": 0.2876821,
        "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",
                    "details": []
                  },
                  {
                    "value": 7,
                    "description": "fieldLength",
                    "details": []
                  }
                ]
              }
            ]
          }
        ]
      },
      {
        "value": 0.2876821,
        "description": "weight(content:is 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",
                    "details": []
                  },
                  {
                    "value": 7,
                    "description": "fieldLength",
                    "details": []
                  }
                ]
              }
            ]
          }
        ]
      },
```

```
      {
        "value": 0.2876821,
        "description": "weight(content:my 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",
                    "details": []
                  },
                  {
                    "value": 7,
                    "description": "fieldLength",
                    "details": []
                  }
                ]
              }
            ]
          }
        ]
      },
      {
        "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",
                    "details": []
                  },
                  {
                    "value": 7,
                    "description": "avgFieldLength",
                    "details": []
                  },
                  {
                    "value": 7,
                    "description": "fieldLength",
                    "details": []
                  }
                ]
              }
            ]
          }
        ]
      }
    ]
  }
}
```