

Full-Text Search

Internals

Philipp Krenn

@xeraa



Developer 🥑

**Who uses a
Database?**

**Who uses
Search?**



Question <https://sli.do/xeraa>

Answer <https://twitter.com/xeraa>

Store

Apache Lucene Elasticsearch



Example

These are **not** the droids you
are looking for.

html_strip **Char Filter**

These are not the droids you are looking for.

standard **Tokenizer**

**These are not the droids you are
looking for**

lowercase **Token Filter**

**these are not the droids you are
looking for**

stop **Token Filter**

droids you looking

snowball **Token Filter**

droid you look

Analyze

```
GET /_analyze
{
  "analyzer": "english",
  "text": "These are not the droids you are looking for."
}
```

```
{  
  "tokens": [  
    {  
      "token": "droid",  
      "start_offset": 18,  
      "end_offset": 24,  
      "type": "<ALPHANUM>",  
      "position": 4  
    },  
    {  
      "token": "you",  
      "start_offset": 25,  
      "end_offset": 28,  
      "type": "<ALPHANUM>",  
      "position": 5  
    },  
    ...  
  ]  
}
```



```
GET /_analyze
{
  "char_filter": [
    "html_strip"
  ],
  "tokenizer": "standard",
  "filter": [
    "lowercase",
    "stop",
    "snowball"
  ],
  "text": "These are <em>not</em> the droids you are looking for."
}
```



```
{
  "tokens": [
    {
      "token": "droid",
      "start_offset": 27,
      "end_offset": 33,
      "type": "<ALPHANUM>",
      "position": 4
    },
    {
      "token": "you",
      "start_offset": 34,
      "end_offset": 37,
      "type": "<ALPHANUM>",
      "position": 5
    },
    ...
  ]
}
```

Stop Words

**a an and are as at be but by for if in into is
it no not of on or such that the their then
there these they this to was will with**

**[https://github.com/apache/lucene-solr/blob/master/lucene/
core/src/java/org/apache/lucene/analysis/standard/
StandardAnalyzer.java#L44-L50](https://github.com/apache/lucene-solr/blob/master/lucene/core/src/java/org/apache/lucene/analysis/standard/StandardAnalyzer.java#L44-L50)**

Always Use Stop Words?

To be, or not to be.

Languages

Arabic, Armenian, Basque, Brazilian, Bulgarian, Catalan, CJK, Czech, Danish, Dutch, English, Finnish, French, Galician, German, Greek, Hindi, Hungarian, Indonesian, Irish, Italian, Latvian, Lithuanian, Norwegian, Persian, Portuguese, Romanian, Russian, Sorani, Spanish, Swedish, Turkish, Thai

More Language Plugins

Core: ICU (Asian languages), Kuromoji (advanced Japanese), Phonetic, SmartCN, Stempel (Polish), Ukrainian

Community: Hebrew, Vietnamese, Network Address Analysis, String2Integer,...

Polish

To nie są droidy, których szukasz.

```
GET /_analyze
{
  "analyzer": "polish",
  "text": "To nie są droidy, których szukasz."
}
```


Polish

droid szukać

Polish with the English Analyzer

nie są droidi których szukasz

Polish Stop Words

<https://github.com/apache/lucene-solr/blob/master/lucene/analysis/stempel/src/resources/org/apache/lucene/analysis/pl/stopwords.txt>

Detect Languages

[https://github.com/spinscale/
elasticsearch-ingest-langdetect](https://github.com/spinscale/elasticsearch-ingest-langdetect)

Language Rules

English: Philipp's → philipp

French: l'église → eglis

German: äußerst → ausserst

Phonetic

```
GET /_analyze
{
  "tokenizer": "standard",
  "filter": [
    {
      "type": "phonetic",
      "encoder": "beider_morse",
      "languageset": "any"
    }
  ],
  "text": "These are not the droids you are looking for."
}
```


Phonetic

... **drDts** **drits** **drots** **iou** **ari** **ori**
loknk...

Another Example

**Obi-Wan never told you what happened to
your father.**

Another Example

**obi wan never told you what
happen your father**

Another Example

No. I am your father.

Another Example

i am your father

Inverted Index

	ID 1	ID 2	ID 3
am	0	0	1[2]
droid	1[4]	0	0
father	0	1[9]	1[4]
happen	0	1[6]	0
i	0	0	1[1]
look	1[7]	0	0
never	0	1[2]	0
obi	0	1[0]	0
told	0	1[3]	0
wan	0	1[1]	0
what	0	1[5]	0
you	1[5]	1[4]	0
your	0	1[8]	1[3]

To / The Index

```
PUT /starwars
{
  "settings": {
    "number_of_shards": 1,
    "analysis": {
      "filter": {
        "my_synonym_filter": {
          "type": "synonym",
          "synonyms": [
            "father,dad",
            "droid => droid,machine"
          ]
        }
      }
    }
  },
}
```



```
"analyzer": {  
  "my_analyzer": {  
    "char_filter": [  
      "html_strip"  
    ],  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "stop",  
      "snowball",  
      "my_synonym_filter"  
    ]  
  }  
}  
},  
,
```

```
"mappings": {  
  "_doc": {  
    "properties": {  
      "quote": {  
        "type": "text",  
        "analyzer": "my_analyzer"  
      }  
    }  
  }  
}
```


Synonyms

Index synonym **or query time** synonym_graph

GET /starwars/_mapping

GET /starwars/_settings

```
PUT /starwars/_doc/1
{
  "quote": "These are <em>not</em> the droids you are looking for."
}
PUT /starwars/_doc/2
{
  "quote": "Obi-Wan never told you what happened to your father."
}
PUT /starwars/_doc/3
{
  "quote": "<b>No</b>. I am your father."
}
```

GET /starwars/_doc/1

GET /starwars/_doc/1/_source

Multi Lingual

Index

PUT /starwars_en/_doc/1

Type

Field

```
{ "quote_en": "...", "quote_de": "..."
```

PS: Single Type per Index

Search

```
POST /starwars/_search
{
  "query": {
    "match_all": { }
  }
}
```


GET **VS** POST

```
{
  "took": 1,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "failed": 0
  },
  "hits": {
    "total": 3,
    "max_score": 1,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "2",
        "_score": 1,
        "_source": {
          "quote": "Obi-Wan never told you what happened to your father."
        }
      },
      ...
    ]
  }
}
```

```
POST /starwars/_search
```

```
{  
  "query": {  
    "match": {  
      "quote": "droid"  
    }  
  }  
}
```

```
{
  "took": 2,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "failed": 0
  },
  "hits": {
    "total": 1,
    "max_score": 0.39556286,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "1",
        "_score": 0.39556286,
        "_source": {
          "quote": "These are <em>not</em> the droids you are looking for."
        }
      }
    ]
  }
}
```

POST /starwars/_search

```
{  
  "query": {  
    "match": {  
      "quote": "dad"  
    }  
  }  
}
```

```

...
"hits": {
  "total": 2,
  "max_score": 0.41913947,
  "hits": [
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "3",
      "_score": 0.41913947,
      "_source": {
        "quote": "<b>No</b>. I am your father."
      }
    },
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "2",
      "_score": 0.39291072,
      "_source": {
        "quote": "Obi-Wan never told you what happened to your father."
      }
    }
  ]
}

```



```
POST /starwars/_doc/0/_explain
```

```
{  
  "query": {  
    "match": {  
      "quote": "dad"  
    }  
  }  
}
```

```
{  
  "_index": "starwars",  
  "_type": "_doc",  
  "_id": "0",  
  "matched": false  
}
```


POST /starwars/_doc/**1**/_explain

```
{  
  "query": {  
    "match": {  
      "quote": "dad"  
    }  
  }  
}
```

```
{
  "_index": "starwars",
  "_type": "_doc",
  "_id": "1",
  "matched": false,
  "explanation": {
    "value": 0,
    "description": "no matching term",
    "details": []
  }
}
```

POST /starwars/_doc/2/_explain

```
{  
  "query": {  
    "match": {  
      "quote": "dad"  
    }  
  }  
}
```

```
{  
  "_index": "starwars",  
  "_type": "_doc",  
  "_id": "2",  
  "matched": true,  
  "explanation": {  
    ...  
  }  
}
```

```
POST /starwars/_search
{
  "query": {
    "match": {
      "quote": "machine"
    }
  }
}
```

```
{
  "took": 2,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": 1,
    "max_score": 1.2499592,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "1",
        "_score": 1.2499592,
        "_source": {
          "quote": "These are <em>not</em> the droids you are looking for."
        }
      }
    ]
  }
}
```



```
POST /starwars/_search
```

```
{  
  "query": {  
    "match_phrase": {  
      "quote": "I am your father"  
    }  
  }  
}
```

```
{
  "took": 3,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "failed": 0
  },
  "hits": {
    "total": 1,
    "max_score": 1.5665855,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "3",
        "_score": 1.5665855,
        "_source": {
          "quote": "<b>No</b>. I am your father."
        }
      }
    ]
  }
}
```



```
POST /starwars/_search
{
  "query": {
    "match_phrase": {
      "quote": {
        "query": "I am father",
        "slop": 1
      }
    }
  }
}
```

```
{
  "took": 16,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "failed": 0
  },
  "hits": {
    "total": 1,
    "max_score": 0.8327639,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "3",
        "_score": 0.8327639,
        "_source": {
          "quote": "<b>No</b>. I am your father."
        }
      }
    ]
  }
}
```

```
POST /starwars/_search
{
  "query": {
    "match_phrase": {
      "quote": {
        "query": "I am not your father",
        "slop": 1
      }
    }
  }
}
```

```
{
  "took": 5,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "failed": 0
  },
  "hits": {
    "total": 1,
    "max_score": 1.0409548,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "3",
        "_score": 1.0409548,
        "_source": {
          "quote": "<b>No</b>. I am your father."
        }
      }
    ]
  }
}
```

```
POST /starwars/_search
```

```
{
```

```
  "query": {
```

```
    "match": {
```

```
      "quote": {
```

```
        "query": "van",
```

```
        "fuzziness": "AUTO"
```

```
      }
```

```
    }
```

```
  }
```

```
}
```

```
{
  "took": 14,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "failed": 0
  },
  "hits": {
    "total": 1,
    "max_score": 0.18155496,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "2",
        "_score": 0.18155496,
        "_source": {
          "quote": "Obi-Wan never told you what happened to your father."
        }
      }
    ]
  }
}
```



```
POST /starwars/_search
{
  "query": {
    "match": {
      "quote": {
        "query": "ovi-van",
        "fuzziness": 1
      }
    }
  }
}
```

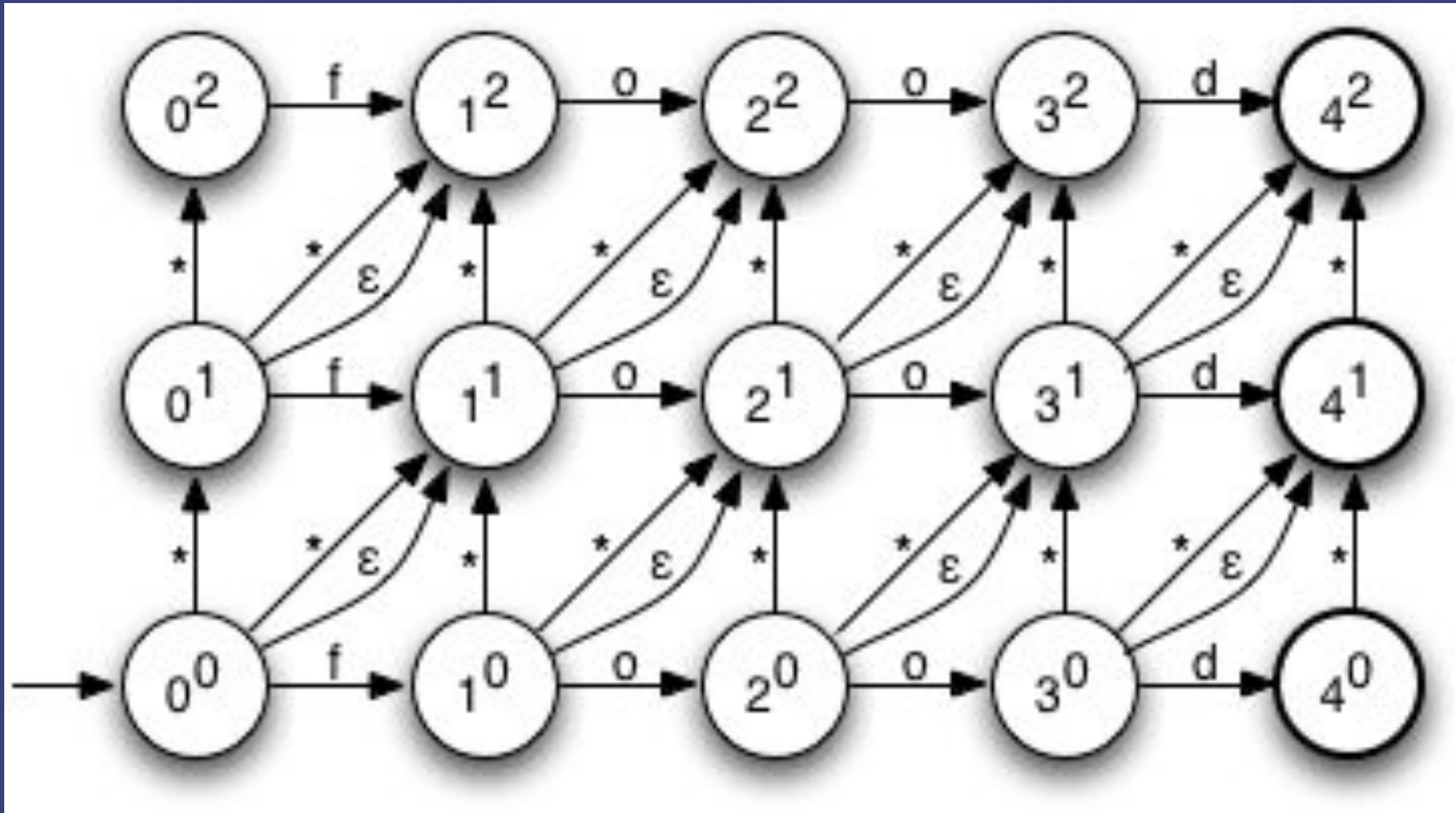
```
{
  "took": 109,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "failed": 0
  },
  "hits": {
    "total": 1,
    "max_score": 0.3798467,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "2",
        "_score": 0.3798467,
        "_source": {
          "quote": "Obi-Wan never told you what happened to your father."
        }
      }
    ]
  }
}
```


FuzzyQuery History

<http://blog.mikemccandless.com/2011/03/lucenes-fuzzyquery-is-100-times-faster.html>

Before: Brute force

Now: Levenshtein Automaton



<http://blog.notdot.net/2010/07/Damn-Cool-Algorithms-Levenshtein-Automata>

```
SELECT *  
  FROM starwars  
 WHERE quote LIKE "?an" OR  
        quote LIKE "V?n" OR  
        quote LIKE "Va?"
```

Scoring

Term Frequency / Inverse Document Frequency (TF/IDF)

Search one term

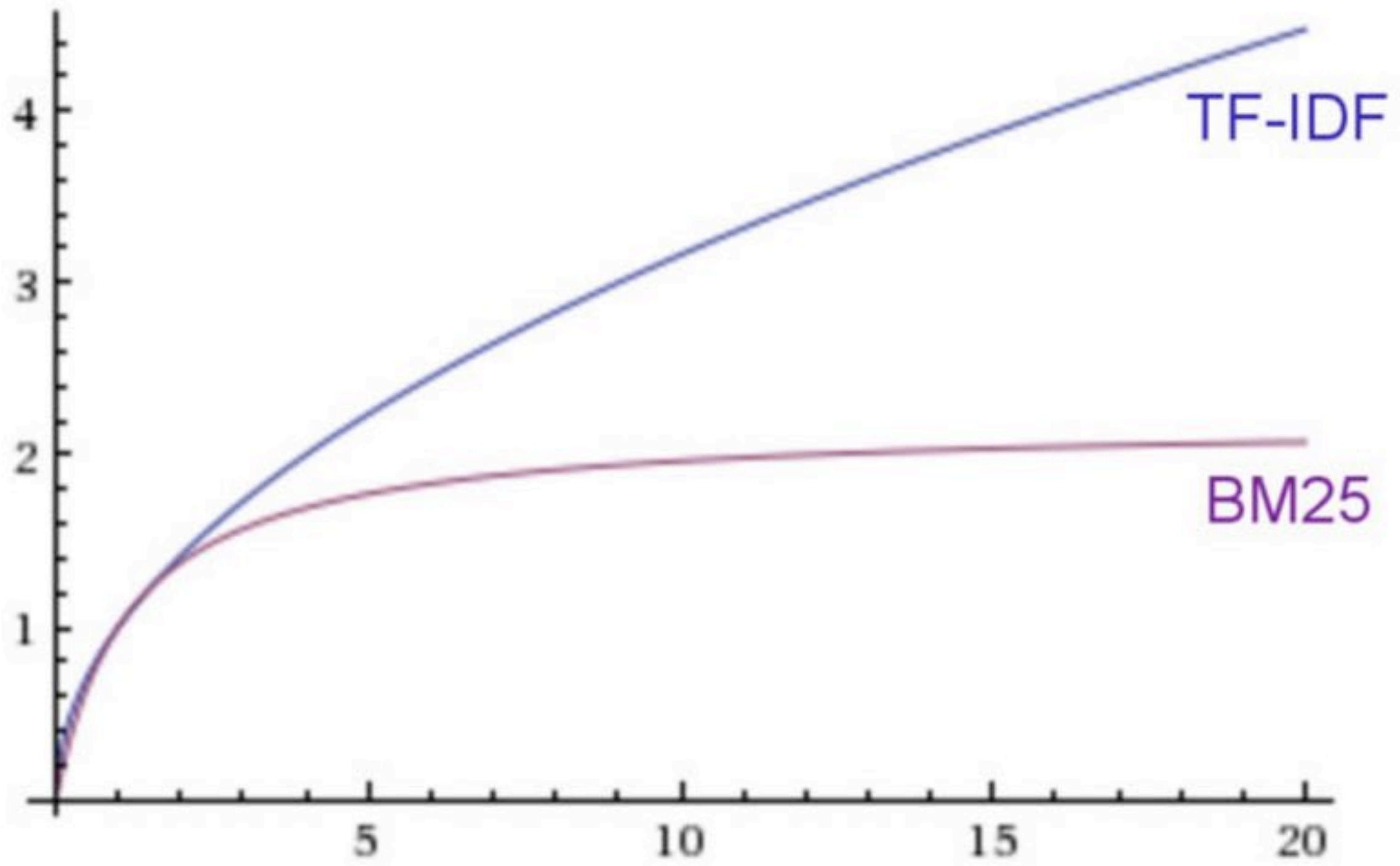
BM25

Default in Elasticsearch 5.0

<https://speakerdeck.com/elastic/improved-text-scoring-with-bm25>

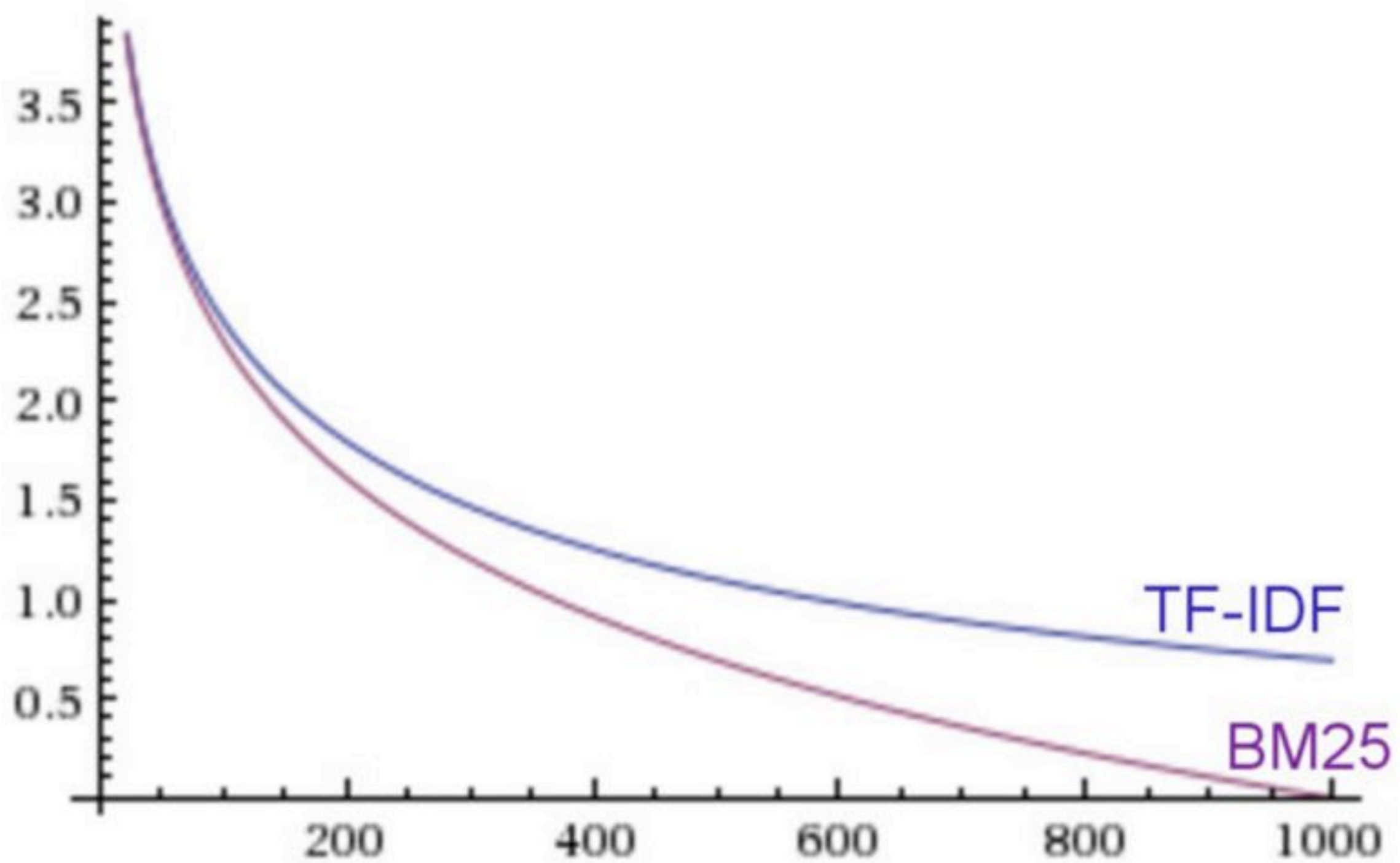
Term Frequency

$$tf(t \text{ in } d) = \sqrt{\text{frequency}}$$



Inverse Document Frequency

$$idf(t) = 1 + \log\left(\frac{numDocs}{docFreq + 1}\right)$$



Field-Length Norm

$$\textit{norm}(d) = \frac{1}{\sqrt{\textit{numTerms}}}$$

```
POST /starwars/_search?explain=true
```

```
{  
  "query": {  
    "match": {  
      "quote": "father"  
    }  
  }  
}
```

```
...
"_explanation": {
  "value": 0.41913947,
  "description": "weight(Synonym(quote:dad quote:father) in 0) [PerFieldSimilarity], result of:",
  "details": [
    {
      "value": 0.41913947,
      "description": "score(doc=0,freq=2.0 = termFreq=2.0\n), product of:",
      "details": [
        {
          "value": 0.2876821,
          "description": "idf(docFreq=1, docCount=1)",
          "details": []
        },
        {
          "value": 1.4569536,
          "description": "tfNorm, computed from:",
          "details": [
            {
              "value": 2,
              "description": "termFreq=2.0",
              "details": []
            }
          ],
          ...
        }
      ]
    }
  ]
}
```

Score

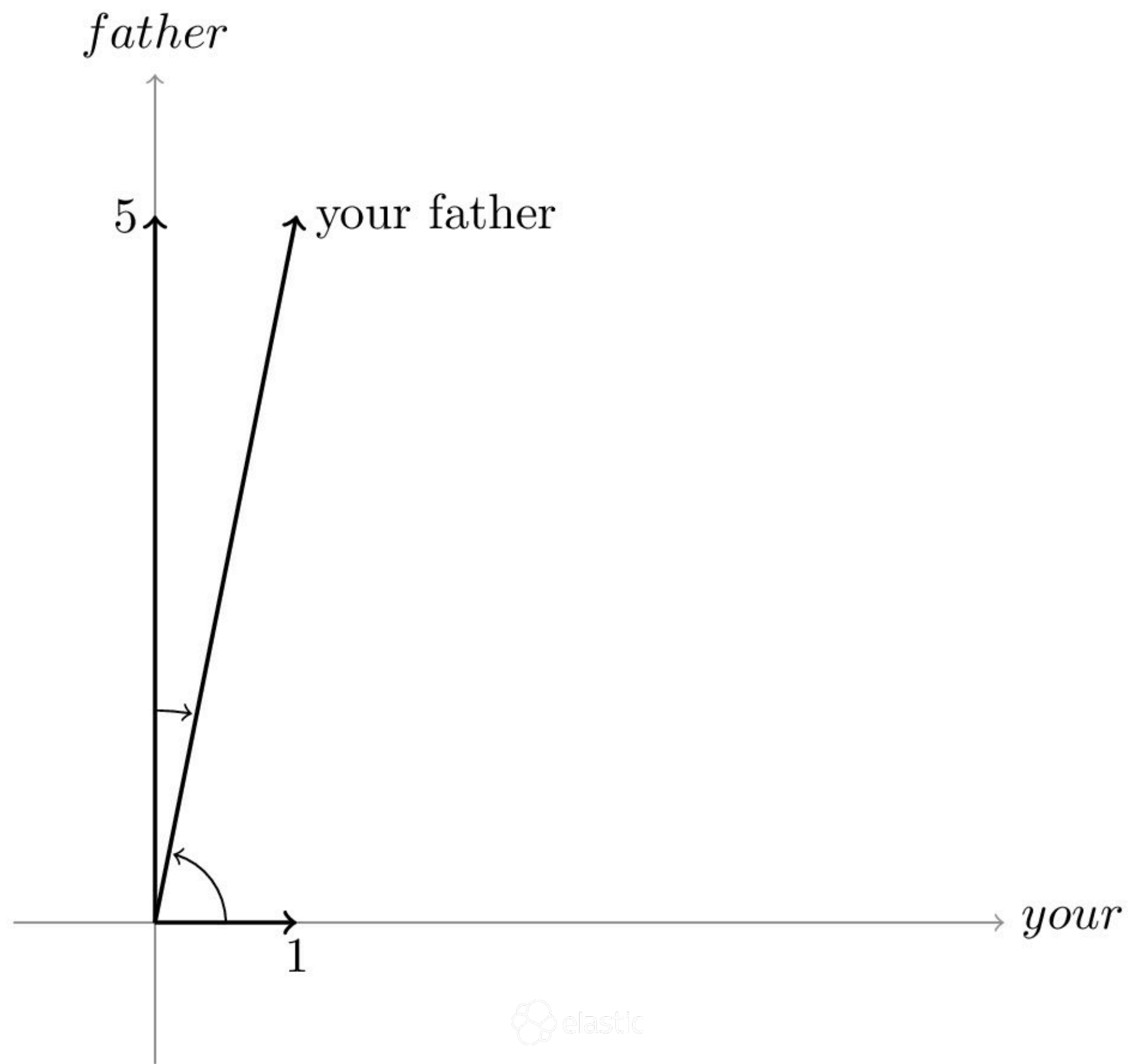
0.41913947: i am your father

**0.39291072: obi wan never told you
what happen your father**

Vector Space Model

Search multiple terms

Search your father



Coordination Factor

Reward multiple terms

Search for 3 terms

1 term: $X \cdot \frac{1}{3}$

2 terms: $(X + Y) \cdot \frac{2}{3}$

3 terms: $(X + Y + Z) \cdot \frac{3}{3}$

Practical Scoring Function

Putting it all together

```
score(q,d) =  
  queryNorm(q)  
  • coord(q,d)  
  •  $\Sigma$  (  
    tf(t in d)  
    • idf(t)2  
    • t.getBoost()  
    • norm(t,d)  
  ) (t in q)
```

Function Score

**Script, weight, random, field value, decay
(geo or date)**


```
POST /starwars/_search
{
  "query": {
    "function_score": {
      "query": {
        "match": {
          "quote": "father"
        }
      },
      "random_score": {}
    }
  }
}
```

Compare Scores

"100% perfect" vs a "50%" match

***Don't do this. Seriously.
Stop trying to think about
your problem this way,
it's not going to end well.***

— [https://wiki.apache.org/lucene-java/
ScoresAsPercentages](https://wiki.apache.org/lucene-java/ScoresAsPercentages)

```
GET /starwars/_analyze
{
  "analyzer" : "my_analyzer",
  "text": "These are my father's machines."
}
```

```
{ "tokens": [  
  {  
    "token": "my",  
    "start_offset": 10,  
    "end_offset": 12,  
    "type": "<ALPHANUM>",  
    "position": 2  
  },  
  {  
    "token": "father",  
    "start_offset": 13,  
    "end_offset": 21,  
    "type": "<ALPHANUM>",  
    "position": 3  
  },  
  {  
    "token": "dad",  
    "start_offset": 13,  
    "end_offset": 21,  
    "type": "SYNONYM",  
    "position": 3  
  },  
  {  
    "token": "machin",  
    "start_offset": 22,  
    "end_offset": 30,  
    "type": "<ALPHANUM>",  
    "position": 4  
  }  
]
```

```
PUT /starwars/_doc/4
{
  "quote": "These are my father's machines."
}
```

```
POST /starwars/_search
{
  "query": {
    "match": {
      "quote": "my father machine"
    }
  }
}
```

```
"hits": {
  "total": 4,
  "max_score": 2.92523,
  "hits": [
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "4",
      "_score": 2.92523,
      "_source": {
        "quote": "These are my father's machines."
      }
    },
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "1",
      "_score": 0.8617505,
      "_source": {
        "quote": "These are <em>not</em> the droids you are looking for."
      }
    }
  ],
  ...
}
```


2.92523 == 100%

```
DELETE /starwars/_doc/4
```

```
POST /starwars/_search
```

```
{  
  "query": {  
    "match": {  
      "quote": "my father machine"  
    }  
  }  
}
```



```
"hits": {
  "total": 3,
  "max_score": 1.2499592,
  "hits": [
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "1",
      "_score": 1.2499592,
      "_source": {
        "quote": "These are <em>not</em> the droids you are looking for."
      }
    },
    ...
  ]
}
```

1.2499592 == 43%
or 100%?

```
PUT /starwars/_doc/4
{
  "quote": "These droids are my father's father's machines."
}
```

```
POST /starwars/_search
{
  "query": {
    "match": {
      "quote": "my father machine"
    }
  }
}
```

```
"hits": {
  "total": 4,
  "max_score": 3.0068164,
  "hits": [
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "4",
      "_score": 3.0068164,
      "_source": {
        "quote": "These droids are my father's father's machines."
      }
    },
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "1",
      "_score": 0.89701396,
      "_source": {
        "quote": "These are <em>not</em> the droids you are looking for."
      }
    }
  ],
  ...
}
```

3.0068164 == 103%?



PS: Shards

Default?

Effect on IDF?

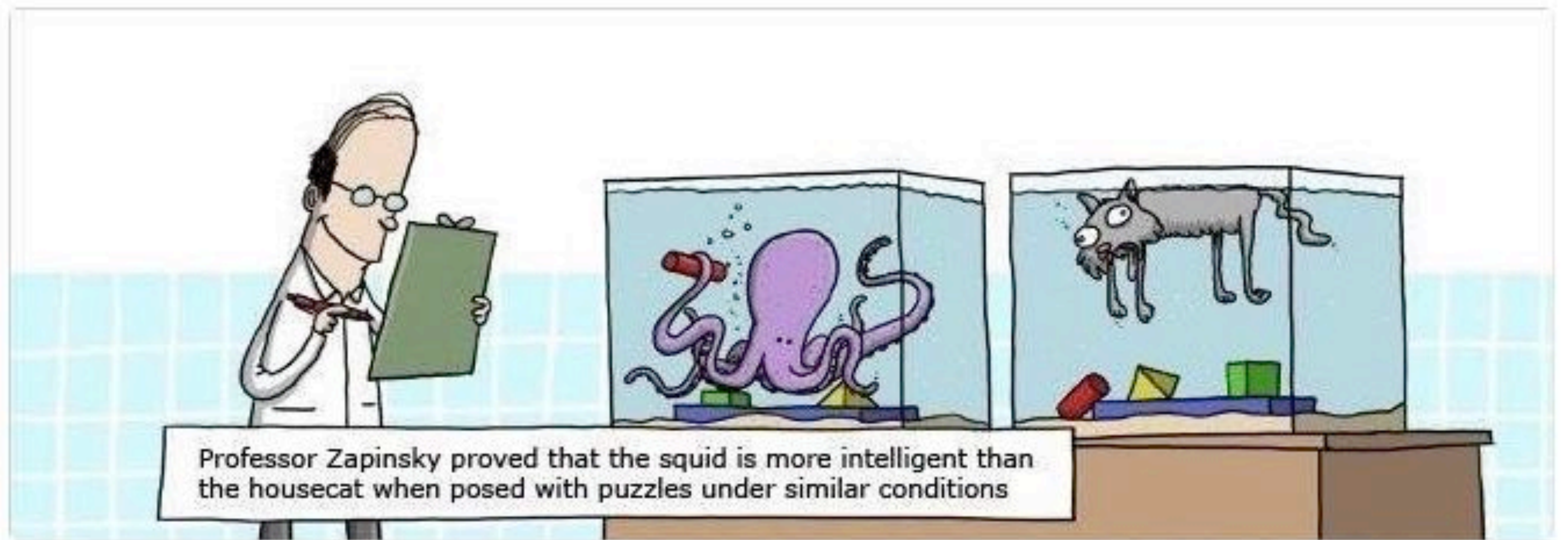
Distributed Frequency Search

```
GET starwars/_search?search_type=dfs_query_then_fetch
{
  ...
}
```

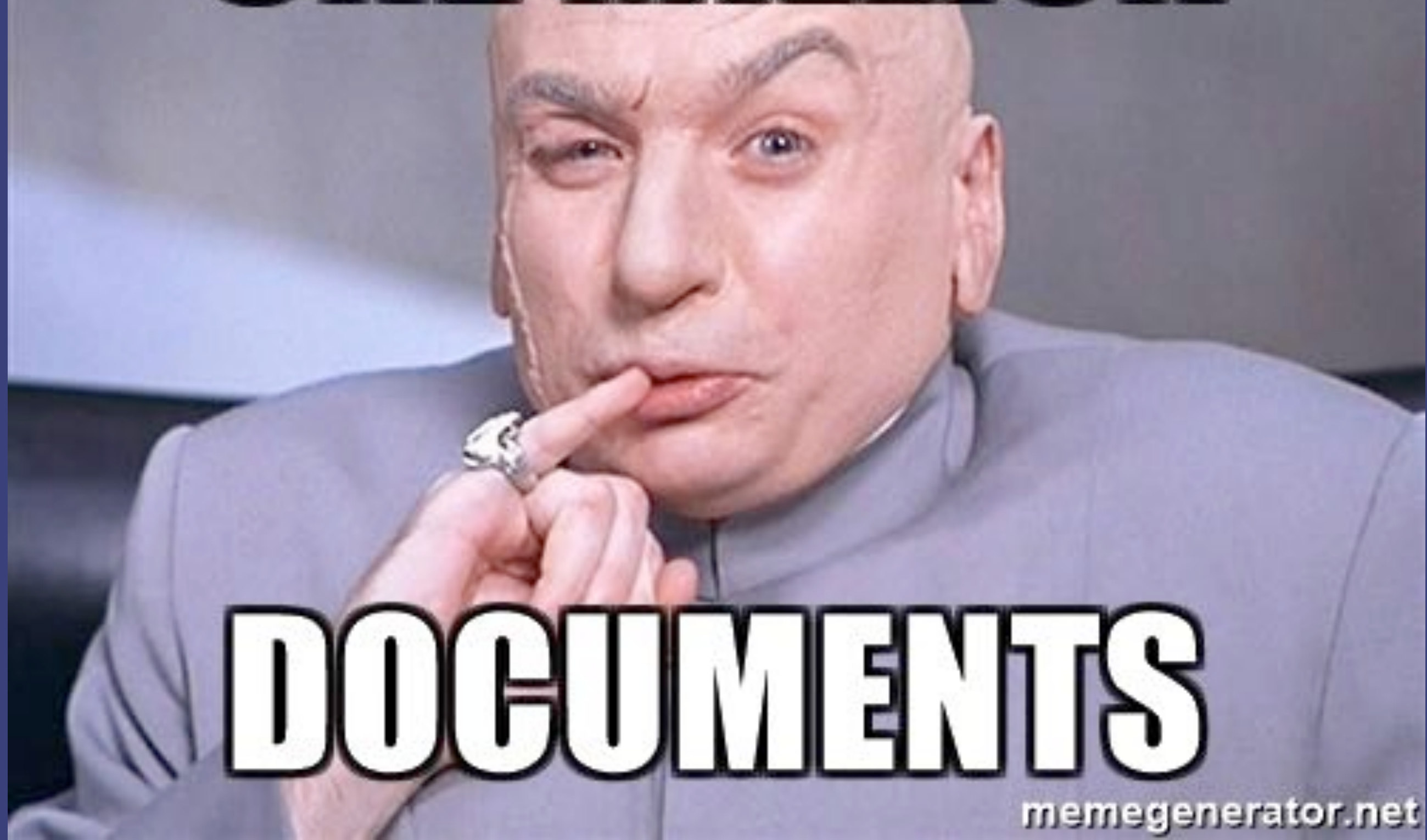

***Don't use
dfs_query_then_fetch **in**
production. It really isn't
required.***

— <https://www.elastic.co/guide/en/elasticsearch/guide/current/relevance-is-broken.html>

Performance



ONE MILLION



DOCUMENTS

Conclusion

Indexing

Formatting

Tokenize

Lowercase, Stop Words, Stemming

Synonyms

Scoring

Term Frequency

Inverse Document Frequency

Field-Length Norm

Vector Space Model

Advanced Queries

Highlighting

NGrams & Edge Grams

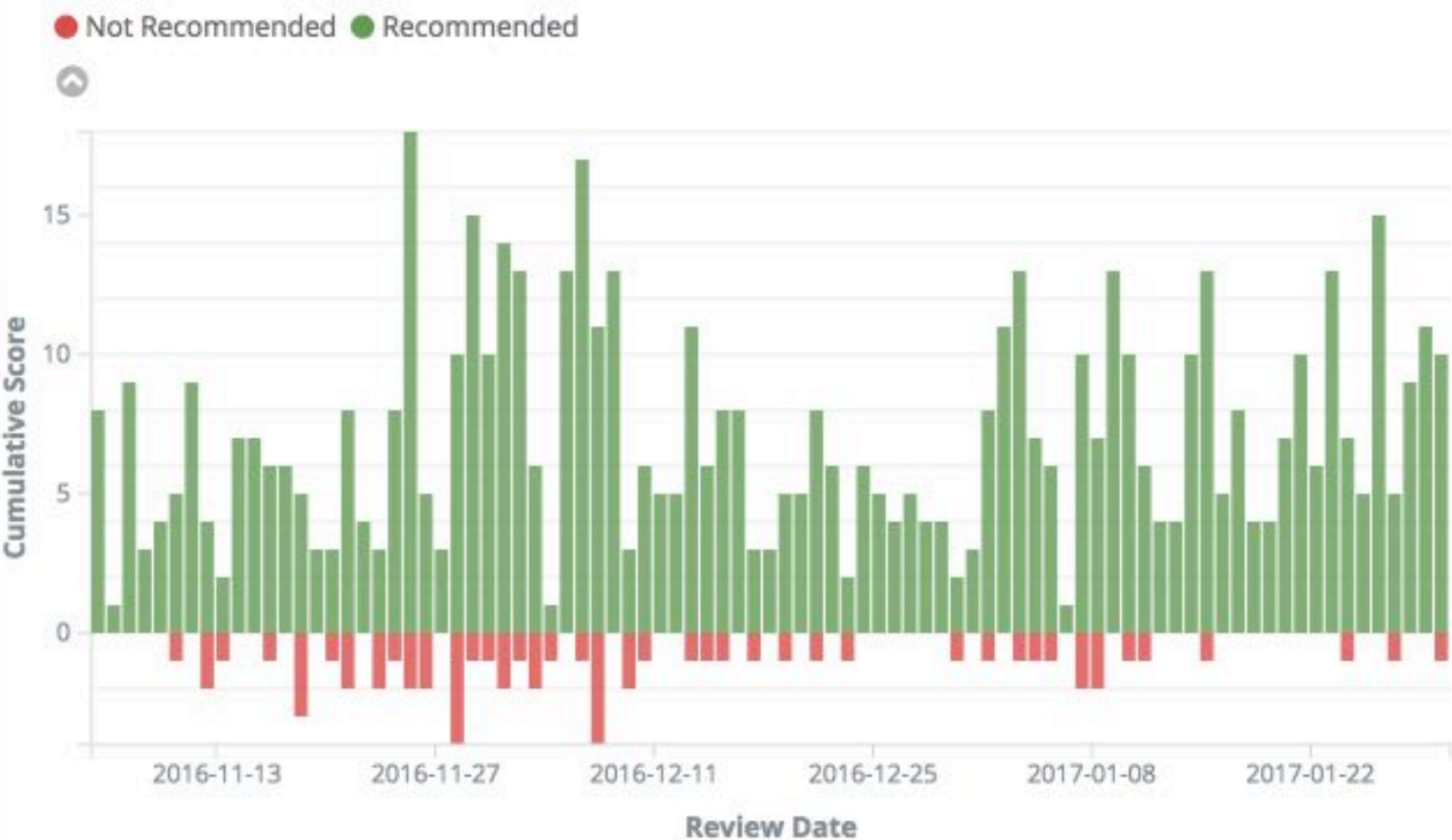
Multiple Analyzers

Reindex & Alias

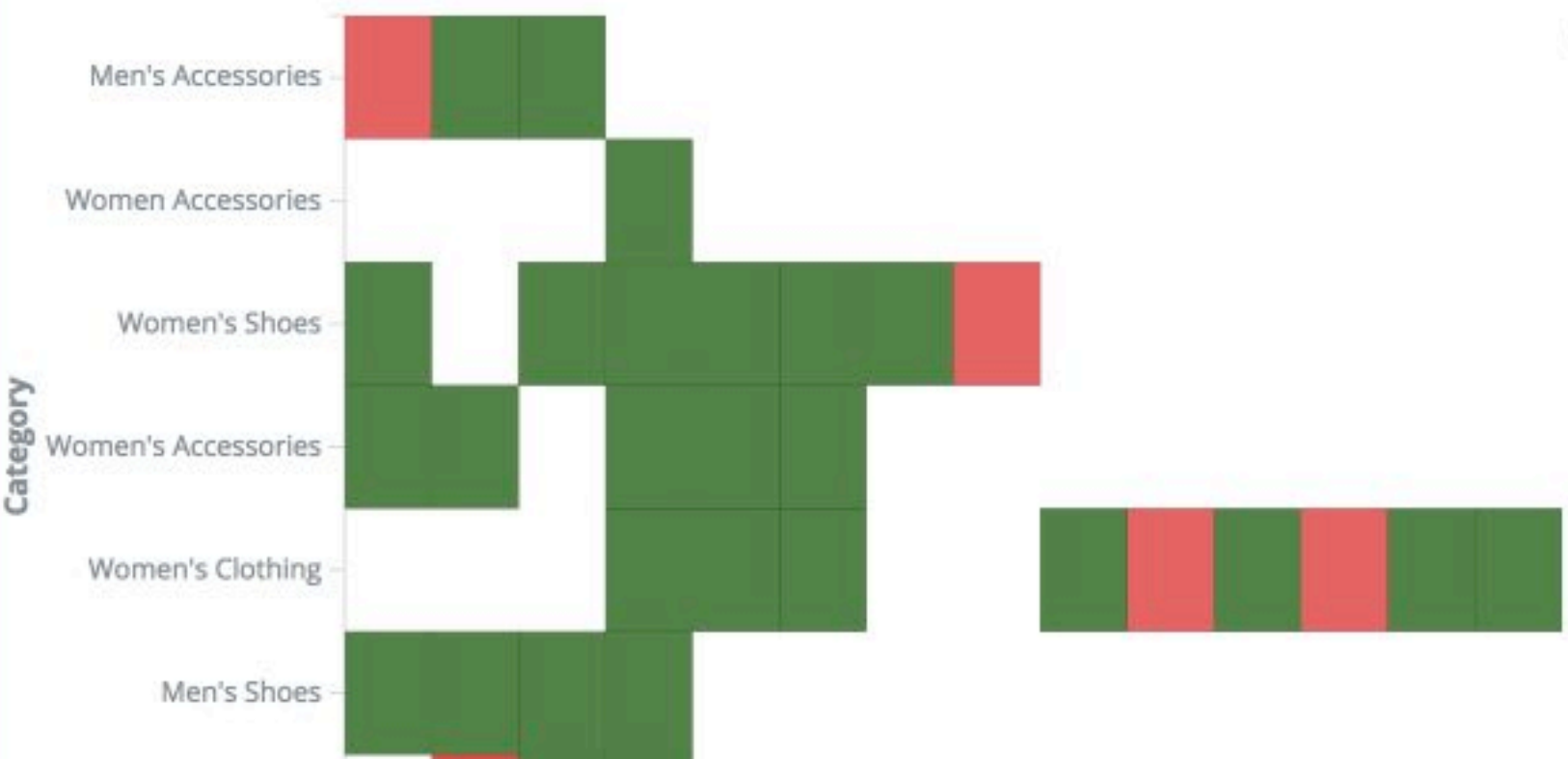
There is more

Elastic Stack

Recommended Over Time



Reviews by Manufacturer and Category



Top Reviewed SKUs with Ratings



Product: Review Keywords

The container is too small to display the entire cloud. Tags might be cropped or omitted.





cloud

<https://cloud.elastic.co>

Email

Password

Log in

[Forgot password?](#)

Don't have an account? [Sign up now.](#)



Have questions? Email us at support@elastic.co

[Trademarks](#) · [Terms of Use](#) · [Privacy](#) · [Brand](#)

© 2018. All Rights Reserved – Elasticsearch

Elasticsearch is a trademark of Elasticsearch BV, registered in the U.S. and in other countries.

Apache, Apache Lucene, Apache Hadoop, Hadoop, HDFS and the yellow elephant logo are trademarks of the Apache Software Foundation in the United States and/or other countries.





swifttype
an Elastic company

Thank You!

Questions?

Philipp Krenn

@xeraa

PS: Stickers

More

```
POST /starwars/_search
{
  "query": {
    "match": {
      "quote": "father"
    }
  },
  "highlight": {
    "type": "unified",
    "pre_tags": [
      "<tag>"
    ],
    "post_tags": [
      "</tag>"
    ],
    "fields": {
      "quote": {}
    }
  }
}
```



```
...
"hits": {
  "total": 3,
  "max_score": 0.631961,
  "hits": [
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "4",
      "_score": 0.631961,
      "_source": {
        "quote": "These droids are my father's father's machines."
      },
      "highlight": {
        "quote": [
          "These droids are my <tag>father's</tag> <tag>father's</tag> machines."
        ]
      }
    },
    ...
  ]
}
```

Boolean Queries

must must_not should filter

POST /starwars/_search

```
{
  "query": {
    "bool": {
      "must": {
        "match": {
          "quote": "father"
        }
      },
      "should": [
        {
          "match": {
            "quote": "your"
          }
        },
        {
          "match": {
            "quote": "obi"
          }
        }
      ]
    }
  }
}
```

```
...
"hits": {
  "total": 3,
  "max_score": 2.117857,
  "hits": [
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "2",
      "_score": 2.117857,
      "_source": {
        "quote": "Obi-Wan never told you what happened to your father."
      }
    },
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "3",
      "_score": 1.3856719,
      "_source": {
        "quote": "<b>No</b>. I am your father."
      }
    }
  ],
  ...
}
```

```
POST /starwars/_search
```

```
{  
  "query": {  
    "bool": {  
      "filter": {  
        "match": {  
          "quote": "father"  
        }  
      },  
      "should": [  
        {  
          "match": {  
            "quote": "your"  
          }  
        },  
        {  
          "match": {  
            "quote": "obi"  
          }  
        }  
      ]  
    }  
  }  
}
```

```
...
"hits": {
  "total": 3,
  "max_score": 1.6694657,
  "hits": [
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "2",
      "_score": 1.6694657,
      "_source": {
        "quote": "Obi-Wan never told you what happened to your father."
      }
    },
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "3",
      "_score": 0.8317767,
      "_source": {
        "quote": "<b>No</b>. I am your father."
      }
    }
  ],
}
```


Named Queries & `minimum_should_match`

```
POST /starwars/_search
{
  "query": {
    "bool": {
      "must": {
        "match": { "quote": "father" }
      },
      "should": [
        {
          "match": {
            "quote": { "query": "your", "_name": "quote-your" }
          }
        },
        {
          "match": {
            "quote": { "query": "obi", "_name": "quote-obi" }
          }
        },
        {
          "match": {
            "quote": { "query": "droid", "_name": "quote-droid" }
          }
        }
      ],
      "minimum_should_match": 2
    }
  }
}
```



```
...
  "hits": {
    "total": 1,
    "max_score": 2.117857,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "2",
        "_score": 2.117857,
        "_source": {
          "quote": "Obi-Wan never told you what happened to your father."
        },
        "matched_queries": [
          "quote-obi",
          "quote-your"
        ]
      }
    ]
  }
}
```

Boosting

>1 increase, <1 decrease, <0 punish

```
POST /starwars/_search
{
  "query": {
    "bool": {
      "must": {
        "match": {
          "quote": "father"
        }
      },
      "should": [
        {
          "match": {
            "quote": "your"
          }
        },
        {
          "match": {
            "quote": {
              "query": "obi",
              "boost": 3
            }
          }
        }
      ]
    }
  }
}
```

```
...
"hits": {
  "total": 3,
  "max_score": 4.2368493,
  "hits": [
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "2",
      "_score": 4.2368493,
      "_source": {
        "quote": "Obi-Wan never told you what happened to your father."
      }
    },
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "3",
      "_score": 1.3856719,
      "_source": {
        "quote": "<b>No</b>. I am your father."
      }
    }
  ],
  ...
}
```

Search for father, **but**
prefer father father

POST /starwars/_search

```
{  
  "query": {  
    "bool": {  
      "must": {  
        "match": {  
          "quote": "father father"  
        }  
      }  
    }  
  }  
}
```



```
...
  "hits": [
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "4",
      "_score": 1.263922,
      "_source": {
        "quote": "These droids are my father's father's machines."
      }
    },
    {
      "_index": "starwars",
      "_type": "_doc",
      "_id": "3",
      "_score": 1.1077905,
      "_source": {
        "quote": "<b>No</b>. I am your father."
      }
    }
  ],
```

POST /starwars/_search

```
{
  "query": {
    "bool": {
      "must": {
        "match": {
          "quote": "father father"
        }
      },
      "should": {
        "match_phrase": {
          "quote": "father father"
        }
      }
    }
  }
}
```



```
...
  "hits": {
    "total": 3,
    "max_score": 3.3799262,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "4",
        "_score": 3.3799262,
        "_source": {
          "quote": "These droids are my father's father's machines."
        }
      },
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "3",
        "_score": 1.1077905,
        "_source": {
          "quote": "<b>No</b>. I am your father."
        }
      }
    ]
  },
  ...
}
```

Suggestion

Suggest a similar text

`_search` **end point**

`_suggest` **deprecated since 5.0**

POST /starwars/_search

```
{
  "query": {
    "match": {
      "quote": "druir"
    }
  },
  "suggest": {
    "my_suggestion" : {
      "text" : "druir",
      "term" : {
        "field" : "quote"
      }
    }
  }
}
```

```
...
  "hits": {
    "total": 0,
    "max_score": null,
    "hits": []
  },
  "suggest": {
    "my_suggestion": [
      {
        "text": "dru",
        "offset": 0,
        "length": 4,
        "options": [
          {
            "text": "droid",
            "score": 0.5,
            "freq": 1
          }
        ]
      }
    ]
  }
}
```

NGram

Partial matches

Edge Gram

```
GET /_analyze
{
  "char_filter": [
    "html_strip"
  ],
  "tokenizer": {
    "type": "ngram",
    "min_gram": "3",
    "max_gram": "3",
    "token_chars": [
      "letter"
    ]
  },
  "filter": [
    "lowercase"
  ],
  "text": "These are <em>not</em> the droids you are looking for."
}
```



```
{
  "tokens": [
    {
      "token": "the",
      "start_offset": 0,
      "end_offset": 3,
      "type": "word",
      "position": 0
    },
    {
      "token": "hes",
      "start_offset": 1,
      "end_offset": 4,
      "type": "word",
      "position": 1
    },
    {
      "token": "ese",
      "start_offset": 2,
      "end_offset": 5,
      "type": "word",
      "position": 2
    },
    {
      "token": "are",
      "start_offset": 6,
      "end_offset": 9,
      "type": "word",
      "position": 3
    },
    ...
  ]
}
```

```
GET /_analyze
{
  "char_filter": [
    "html_strip"
  ],
  "tokenizer": {
    "type": "edge_ngram",
    "min_gram": "1",
    "max_gram": "3",
    "token_chars": [
      "letter"
    ]
  },
  "filter": [
    "lowercase"
  ],
  "text": "These are <em>not</em> the droids you are looking for."
}
```



```
{
  "tokens": [
    {
      "token": "t",
      "start_offset": 0,
      "end_offset": 1,
      "type": "word",
      "position": 0
    },
    {
      "token": "th",
      "start_offset": 0,
      "end_offset": 2,
      "type": "word",
      "position": 1
    },
    {
      "token": "the",
      "start_offset": 0,
      "end_offset": 3,
      "type": "word",
      "position": 2
    },
    {
      "token": "a",
      "start_offset": 6,
      "end_offset": 7,
      "type": "word",
      "position": 3
    },
    {
      "token": "ar",
      "start_offset": 6,
      "end_offset": 8,
      "type": "word",
      "position": 4
    },
    ...
  ]
}
```

Combining Analyzers

Reindex

Store multiple times

Tune BM25

Combine scores

BM25 Revisited

$$\sum_i^n IDF(q_i) \frac{f(q_i, D) * (k1 + 1)}{f(q_i, D) + k1 * (1 - b + b * \frac{fieldLen}{avgFieldLen})}$$

<https://www.elastic.co/blog/practical-bm25-part-2-the-bm25-algorithm-and-its-variables>

b **field length amplification**

Default 0.75

k1 **term frequency saturation**

Default 1.2

```
PUT /starwars_v42
{
  "settings": {
    "number_of_shards": 1,
    "index": {
      "similarity": {
        "default": {
          "type": "BM25",
          "b": 0,
          "k1": 0
        }
      }
    }
  },

```



```
"analysis": {
  "filter": {
    "my_synonym_filter": {
      "type": "synonym",
      "synonyms": [
        "droid,machine",
        "father,dad"
      ]
    },
    "my_ngram_filter": {
      "type": "ngram",
      "min_gram": "3",
      "max_gram": "3",
      "token_chars": [
        "letter"
      ]
    }
  },
}
```

```
"analyzer": {
  "my_lowercase_analyzer": {
    "char_filter": [
      "html_strip"
    ],
    "tokenizer": "whitespace",
    "filter": [
      "lowercase"
    ]
  },
  "my_full_analyzer": {
    "char_filter": [
      "html_strip"
    ],
    "tokenizer": "standard",
    "filter": [
      "lowercase",
      "stop",
      "snowball",
      "my_synonym_filter"
    ]
  },
}
```



```
"my_ngram_analyzer": {  
  "char_filter": [  
    "html_strip"  
  ],  
  "tokenizer": "whitespace",  
  "filter": [  
    "lowercase",  
    "stop",  
    "my_ngram_filter"  
  ]  
}  
}  
},
```

```
"mappings": {
  "_doc": {
    "properties": {
      "quote": {
        "type": "text",
        "fields": {
          "lowercase": {
            "type": "text",
            "analyzer": "my_lowercase_analyzer"
          },
          "full": {
            "type": "text",
            "analyzer": "my_full_analyzer"
          },
          "ngram": {
            "type": "text",
            "analyzer": "my_ngram_analyzer"
          }
        }
      }
    }
  }
}
```

```
POST /_reindex
{
  "source": {
    "index": "starwars"
  },
  "dest": {
    "index": "starwars_v42"
  }
}
```

Aliases

Atomic remove **and** add

Point to multiple indices (read-only)

```
PUT _alias
{
  "actions": [
    {
      "add": {
        "index": "starwars_v42",
        "alias": "starwars_extended"
      }
    }
  ]
}
```

```
POST /starwars/_search
```

```
{  
  "query": {  
    "match": {  
      "quote": "droid"  
    }  
  }  
}
```



```
"hits": [  
  {  
    "_index": "starwars",  
    "_type": "_doc",  
    "_id": "4",  
    "_score": 1.1533037,  
    "_source": {  
      "quote": "These droids are my father's father's machines."  
    }  
  },  
  {  
    "_index": "starwars",  
    "_type": "_doc",  
    "_id": "1",  
    "_score": 1.1295731,  
    "_source": {  
      "quote": "These are <em>not</em> the droids you are looking for."  
    }  
  }  
]
```

```
POST /starwars_extended/_search
{
  "query": {
    "match": {
      "quote.full": "droid"
    }
  }
}
```



```
"hits": [  
  {  
    "_index": "starwars_v42",  
    "_type": "_doc",  
    "_id": "1",  
    "_score": 0.6931472,  
    "_source": {  
      "quote": "These are <em>not</em> the droids you are looking for."  
    }  
  },  
  {  
    "_index": "starwars_v42",  
    "_type": "_doc",  
    "_id": "4",  
    "_score": 0.6931472,  
    "_source": {  
      "quote": "These droids are my father's father's machines."  
    }  
  }  
]
```

There are no "best" **b** and
k1 values

```
POST /starwars_extended/_search?explain=true
{
  "query": {
    "multi_match": {
      "query": "obiwan",
      "fields": [
        "quote",
        "quote.lowercase",
        "quote.full",
        "quote.ngram"
      ],
      "type": "most_fields"
    }
  }
}
```

```
...
"hits": {
  "total": 1,
  "max_score": 0.4912064,
  "hits": [
    {
      "_shard": "[starwars_v42][2]",
      "_node": "BCDwzJ4WSw2dyoGLTzw1qw",
      "_index": "starwars_v42",
      "_type": "_doc",
      "_id": "2",
      "_score": 0.4912064,
      "_source": {
        "quote": "Obi-Wan never told you what happened to your father."
      },
    },
  ],
  ...
}
```

Whitespace Tokenizer

```
"weight(  
  Synonym(quote.ngram:biw quote.ngram:iwa quote.ngram:obi quote.ngram:wan)  
in 0) [PerFieldSimilarity], result of:"
```

```
POST /starwars_extended/_search
{
  "query": {
    "multi_match": {
      "query": "you",
      "fields": [
        "quote",
        "quote.lowercase^5",
        "quote.full",
        "quote.ngram"
      ],
      "type": "best_fields"
    }
  }
}
```



```
"hits": [  
  {  
    "_index": "starwars_v42",  
    "_type": "_doc",  
    "_id": "1",  
    "_score": 2.1939285,  
    "_source": {  
      "quote": "These are <em>not</em> the droids you are looking for."  
    }  
  },  
  {  
    "_index": "starwars_v42",  
    "_type": "_doc",  
    "_id": "2",  
    "_score": 2.1939285,  
    "_source": {  
      "quote": "Obi-Wan never told you what happened to your father."  
    }  
  },  
  {  
    "_index": "starwars_v42",  
    "_type": "_doc",  
    "_id": "3",  
    "_score": 0.1990188,  
    "_source": {  
      "quote": "<b>No</b>. I am your father."  
    }  
  }  
]
```

Multi Match Type

best_fields **Score of the best field (default)**

cross_fields **All terms in at least one field**

most_fields **Score sum of all fields**

phrase

Different Analyzers for Indexing and Searching

Per query

In the mapping

```
POST /starwars_extended/_search
{
  "query": {
    "match": {
      "quote.ngram": {
        "query": "the",
        "analyzer": "standard"
      }
    }
  }
}
```

```
...
"hits": [
  {
    "_index": "starwars_extended",
    "_type": "_doc",
    "_id": "2",
    "_score": 0.38254172,
    "_source": {
      "quote": "Obi-Wan never told you what happened to your father."
    }
  },
  {
    "_index": "starwars_extended",
    "_type": "_doc",
    "_id": "3",
    "_score": 0.36165747,
    "_source": {
      "quote": "<b>No</b>. I am your father."
    }
  }
]
...
```

Edge Gram vs Trigram

Extending a mapping

Testing a custom mapping

```
POST /starwars_extended/_close
PUT /starwars_extended/_settings
{
  "analysis": {
    "filter": {
      "my_edgegram_filter": {
        "type": "edge_ngram",
        "min_gram": 3,
        "max_gram": 10
      }
    },
    "analyzer": {
      "my_edgegram_analyzer": {
        "char_filter": [
          "html_strip"
        ],
        "tokenizer": "standard",
        "filter": [
          "lowercase",
          "my_edgegram_filter"
        ]
      }
    }
  }
}
POST /starwars_extended/_open
```

```
GET starwars_extended/_analyze
{
  "text": "Father",
  "analyzer": "my_edgegram_analyzer"
}
```



```
{
  "tokens": [
    {
      "token": "fat",
      "start_offset": 0,
      "end_offset": 6,
      "type": "<ALPHANUM>",
      "position": 0
    },
    {
      "token": "fath",
      "start_offset": 0,
      "end_offset": 6,
      "type": "<ALPHANUM>",
      "position": 0
    },
    {
      "token": "fathe",
      "start_offset": 0,
      "end_offset": 6,
      "type": "<ALPHANUM>",
      "position": 0
    },
    {
      "token": "father",
      "start_offset": 0,
      "end_offset": 6,
      "type": "<ALPHANUM>",
      "position": 0
    }
  ]
}
```

```
PUT /starwars_extended/_doc/_mapping
{
  "properties": {
    "quote": {
      "type": "text",
      "fields": {
        "edgegram": {
          "type": "text",
          "analyzer": "my_edgegram_analyzer",
          "search_analyzer": "standard"
        }
      }
    }
  }
}
```



```
PUT /starwars_extended/_doc/4
{
  "quote": "I find your lack of faith disturbing."
}
PUT /starwars_extended/_doc/5
{
  "quote": "That... is your failure."
}
```

```
GET /starwars_extended/_doc/4/_termvectors
{
  "fields": [
    "quote.edagram"
  ],
  "offsets": true,
  "payloads": true,
  "positions": true,
  "term_statistics": true,
  "field_statistics": true
}
```

```
{
  "_index": "starwars_v42",
  "_type": "_doc",
  "_id": "4",
  "_version": 1,
  "found": true,
  "took": 3,
  "term_vectors": {
    "quote.edgegram": {
      "field_statistics": {
        "sum_doc_freq": 26,
        "doc_count": 2,
        "sum_ttf": 26
      },
      "terms": {
        "dis": {
          "doc_freq": 1,
          "ttf": 1,
          "term_freq": 1,
          "tokens": [
            {
              "position": 6,
              "start_offset": 26,
              "end_offset": 36
            }
          ]
        }
      },
      "dist": {
        "doc_freq": 1,
        "ttf": 1,
        ...
      }
    }
  }
}
```

POST /starwars_extended/_search

```
{  
  "query": {  
    "match": {  
      "quote": "fail"  
    }  
  }  
}
```

```
POST /starwars_extended/_search
{
  "query": {
    "match": {
      "quote.lowercase": "fail"
    }
  }
}
```

```
POST /starwars_extended/_search
{
  "query": {
    "match": {
      "quote.full": "fail"
    }
  }
}
```



```
POST /starwars_extended/_search
{
  "query": {
    "match": {
      "quote.ngram": "fail"
    }
  }
}
```

```
...
"hits": {
  "total": 2,
  "max_score": 1.0135446,
  "hits": [
    {
      "_index": "starwars_v42",
      "_type": "_doc",
      "_id": "4",
      "_score": 1.0135446,
      "_source": {
        "quote": "I find your lack of faith disturbing."
      }
    },
    {
      "_index": "starwars_v42",
      "_type": "_doc",
      "_id": "5",
      "_score": 0.50476736,
      "_source": {
        "quote": "That... is your failure."
      }
    }
  ]
}
...
```



```
POST /starwars_extended/_search
{
  "query": {
    "match": {
      "quote.edgogram": "fail"
    }
  }
}
```

```
...
"hits": {
  "total": 1,
  "max_score": 0.39556286,
  "hits": [
    {
      "_index": "starwars_v42",
      "_type": "_doc",
      "_id": "5",
      "_score": 0.39556286,
      "_source": {
        "quote": "That... is your failure."
      }
    }
  ]
}
...
```

Trainings

<https://training.elastic.co>

The End