binlogic DataOps

" A journey between **infrastructure** and **data** "

JUNE 20-21 2019 WTCBCN

Philipp Krenn

@xeraa

Full-Text Search Internals



Who is using databases?

Who is using search?


elastic



elasticsearch



elastic

Developer 🥑

Store

Apache Lucene
Elasticsearch



elastic cloud

https://cloud.elastic.co



docker

elastic

```yaml
---
version: '2'
services:

  elasticsearch:
    image: docker.elastic.co/elasticsearch/elasticsearch:$ELASTIC_VERSION
    environment:
      - bootstrap.memory_lock=true
      - "ES_JAVA_OPTS=-Xms512m -Xmx512m"
      - discovery.type=single-node
    ulimits:
      memlock:
        soft: -1
        hard: -1
    mem_limit: 1g
    volumes:
      - esdata1:/usr/share/elasticsearch/data
    ports:
      - 9200:9200

  kibana:
    image: docker.elastic.co/kibana/kibana:$ELASTIC_VERSION
    links:
      - elasticsearch
    ports:
      - 5601:5601

volumes:
  esdata1:
    driver: local
```



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

elastic

```
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

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 (better Polish stemming), Ukrainian (stemming)

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

Language Rules

English: `Philipp's` → `philipp`

French: `l'église` → `eglis`

German: `äußerst` → `ausserst`

Spanish

Éstos no son los androides que estáis buscando.

elastic

Spanish

est android buscand

Spanish with the English Analyzer

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

elastic

```
PUT /starwars
{
  "settings": {
    "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": {
  "properties": {
    "quote": {
      "type": "text",
      "analyzer": "my_analyzer"
    }
  }
}
```

```
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/_source/1
```

Search

elastic

```
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/_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
      }
    }
  }
}
```

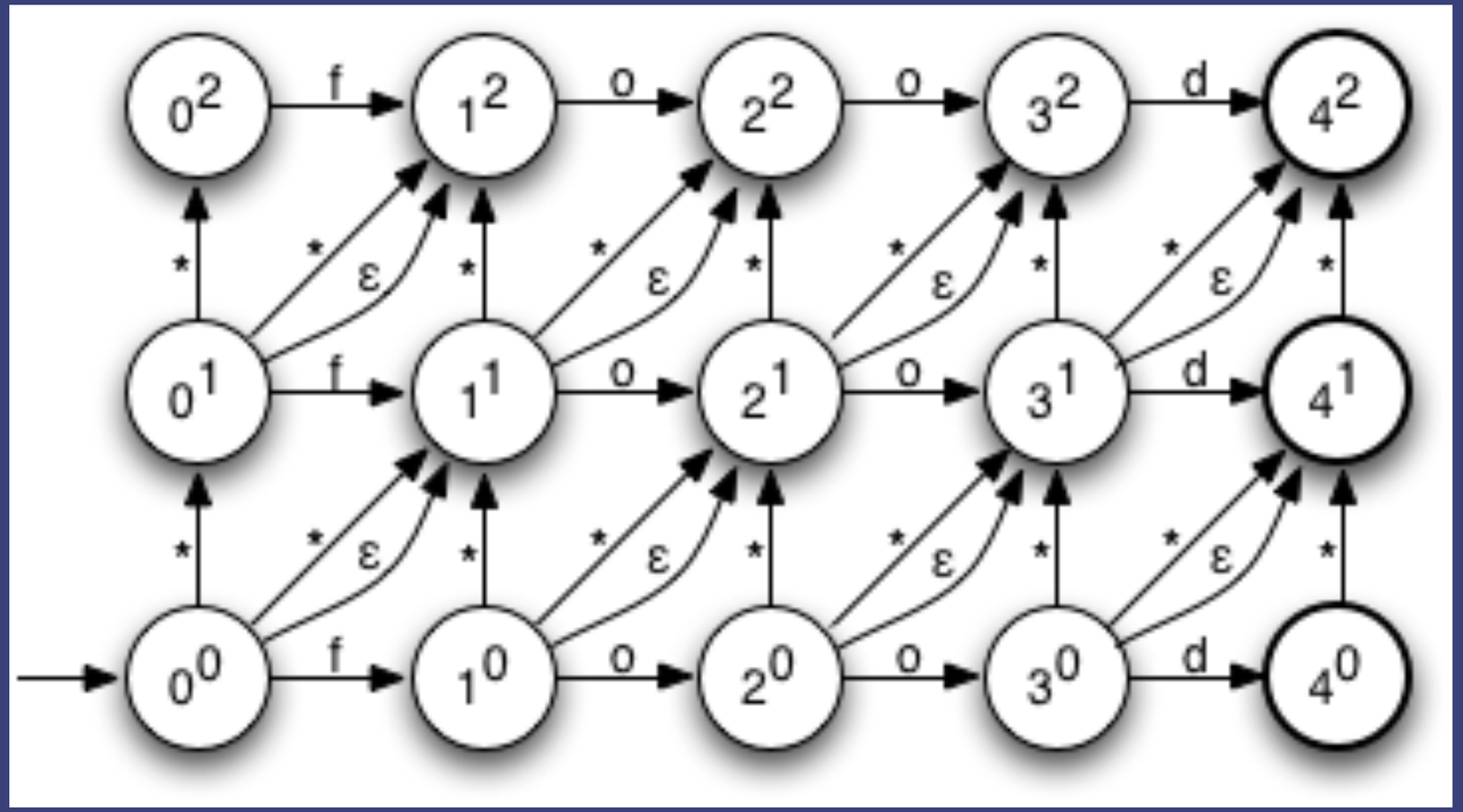
```
{
  "took": 109,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "failed": 0
  },
  "hits": {
    "total": 1,
    "max_score": 0.3798467,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "2",
        "_score": 0.3798467,
        "_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

elastic

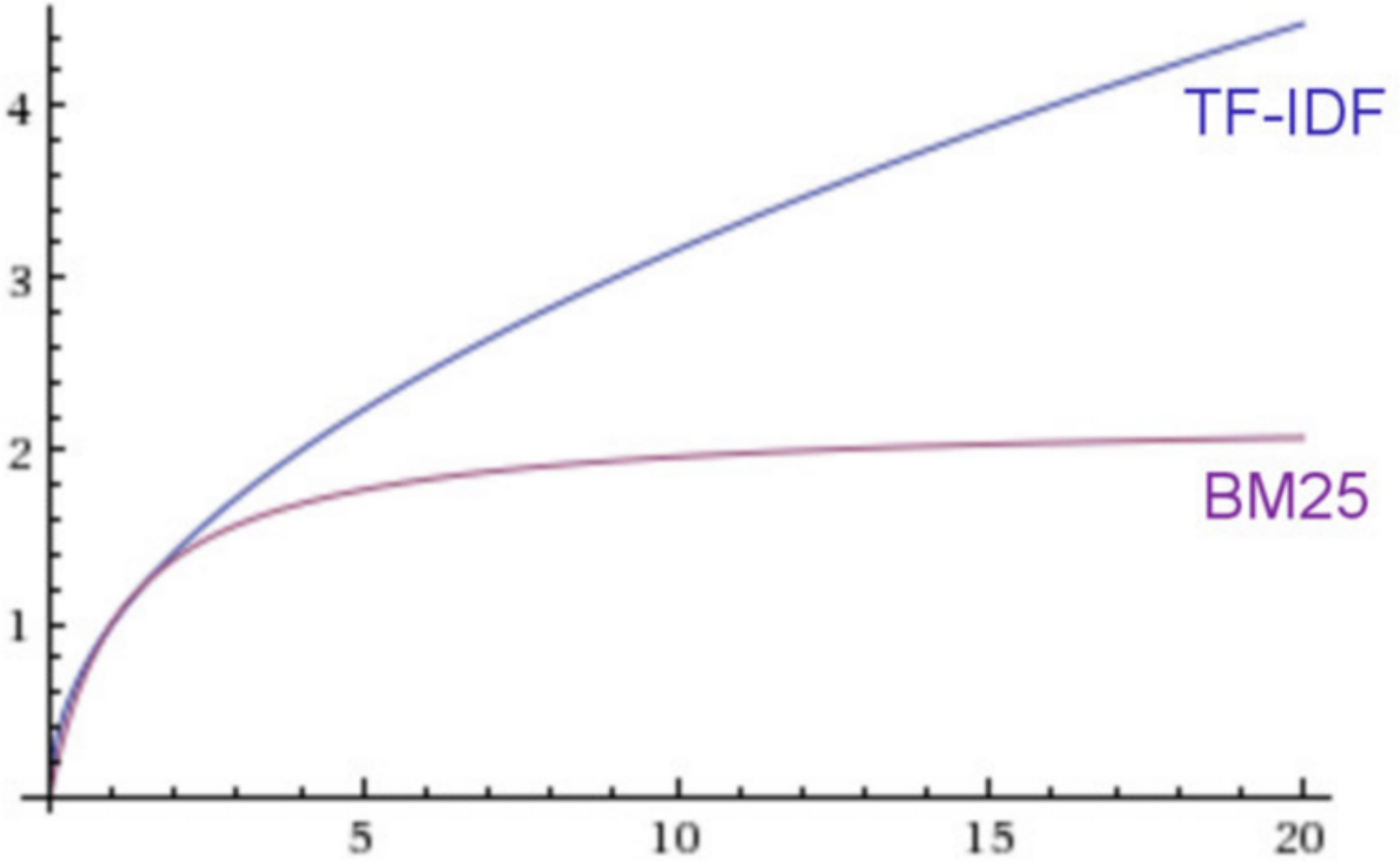
BM25

Default in Elasticsearch 5.0

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

Term Frequency

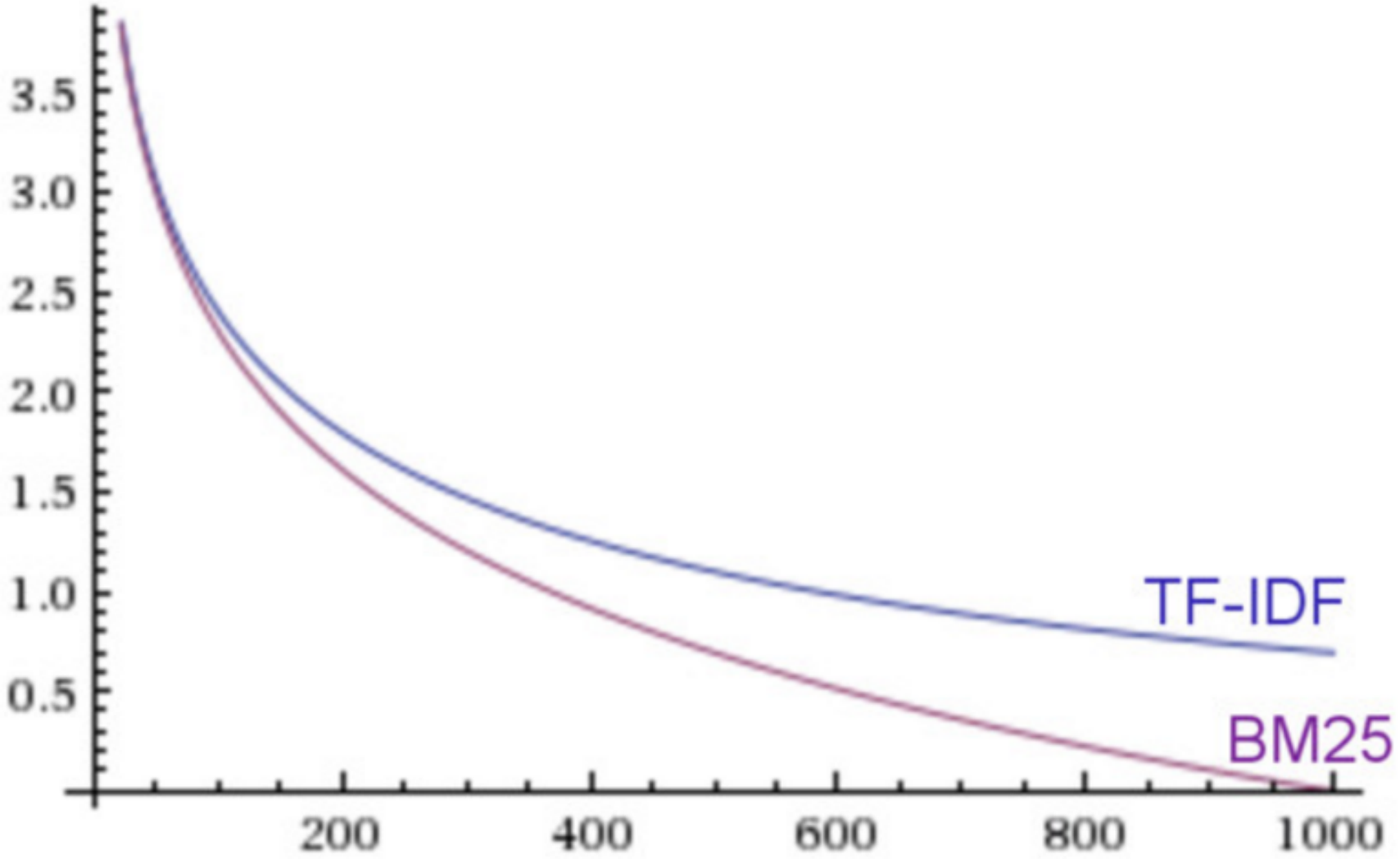
$$tf(t\ in\ d) = \sqrt{frequency}$$



elastic

Inverse Document Frequency

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



elastic

Field-Length Norm

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

```
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

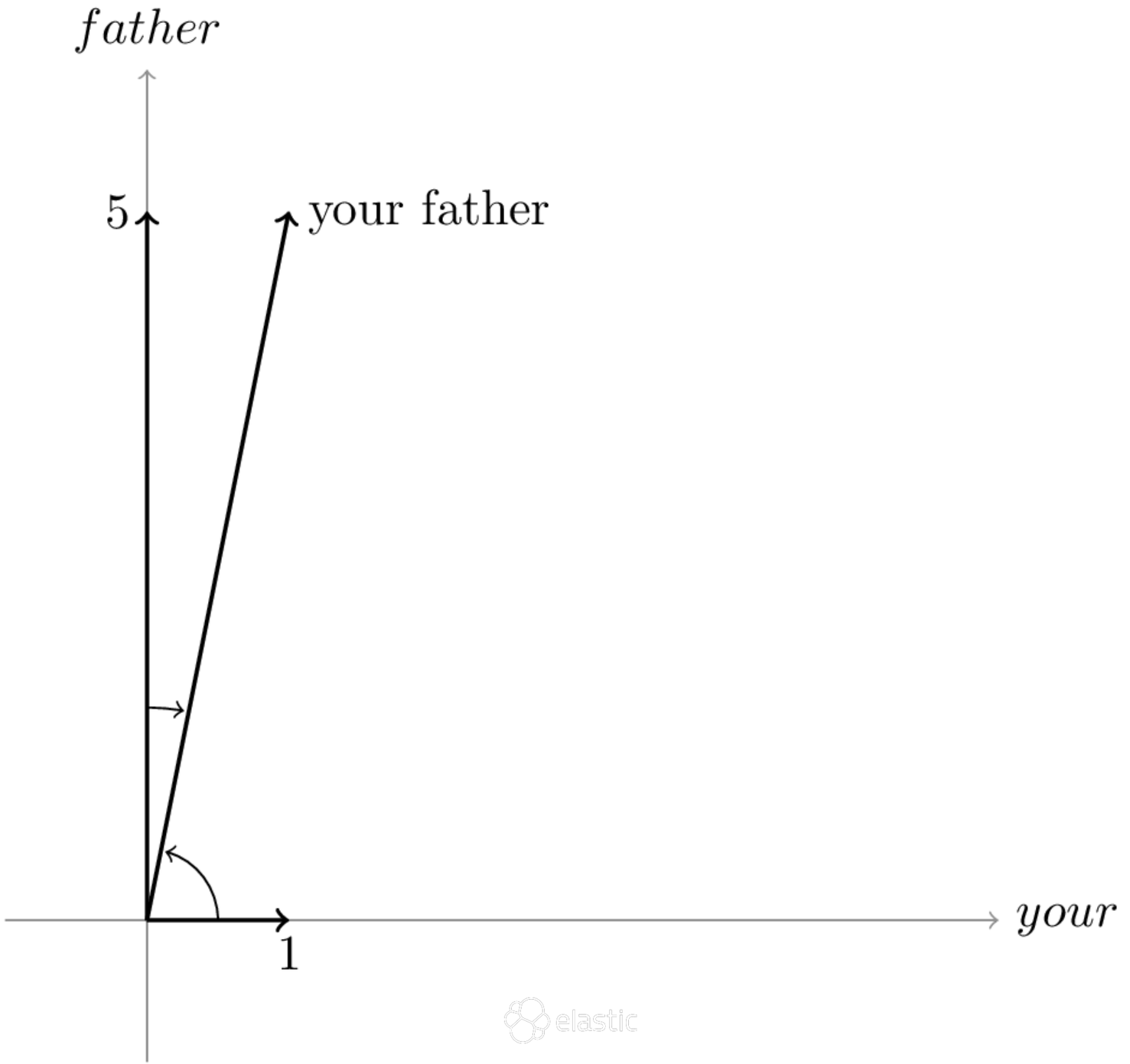
Vector Space Model

Search multiple terms

elastic

Search your father

elastic


father
5
your father
your
1

Coordination Factor

Reward multiple terms

elastic

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

elastic

```
score(q,d)  =
              queryNorm(q)
            · coord(q,d)
            · Σ (
                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

```
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."
}

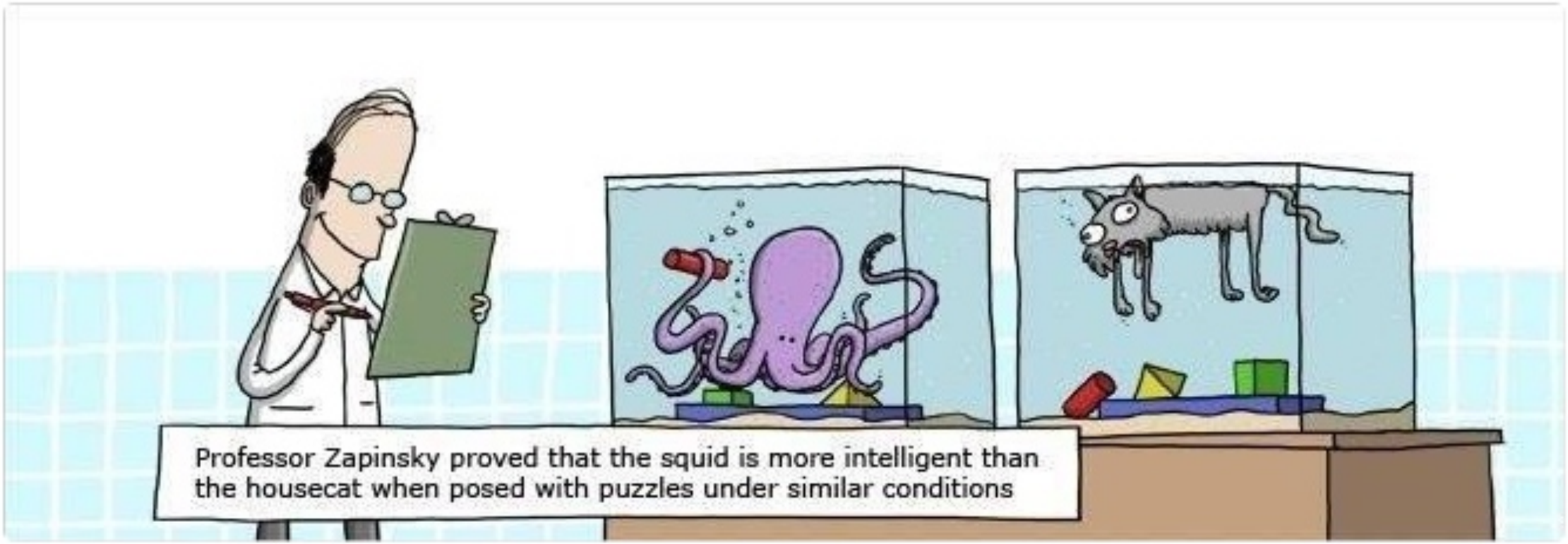
POST /starwars/_search
{
  "query": {
    "match": {
      "quote": "my father machine"
    }
  }
}
```

```
"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%?



Performance


Professor Zapinsky proved that the squid is more intelligent than the housecat when posed with puzzles under similar conditions

elastic


ONE MILLION
DOCUMENTS
memegenerator.net

Conclusion

Indexing

Formatting

Tokenize

Lowercase, Stop Words, Stemming

Synonyms

Scoring

Term Frequency

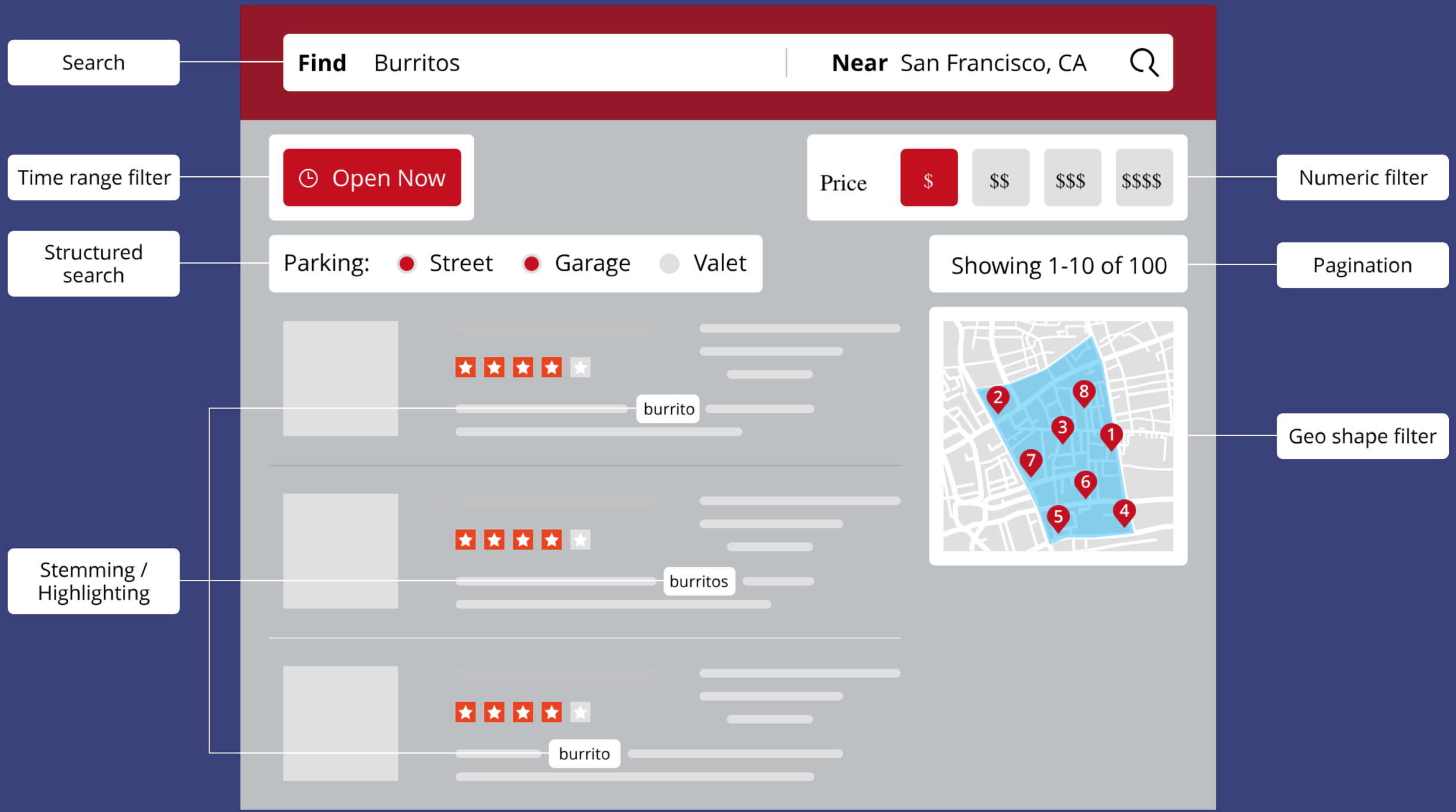
Inverse Document Frequency

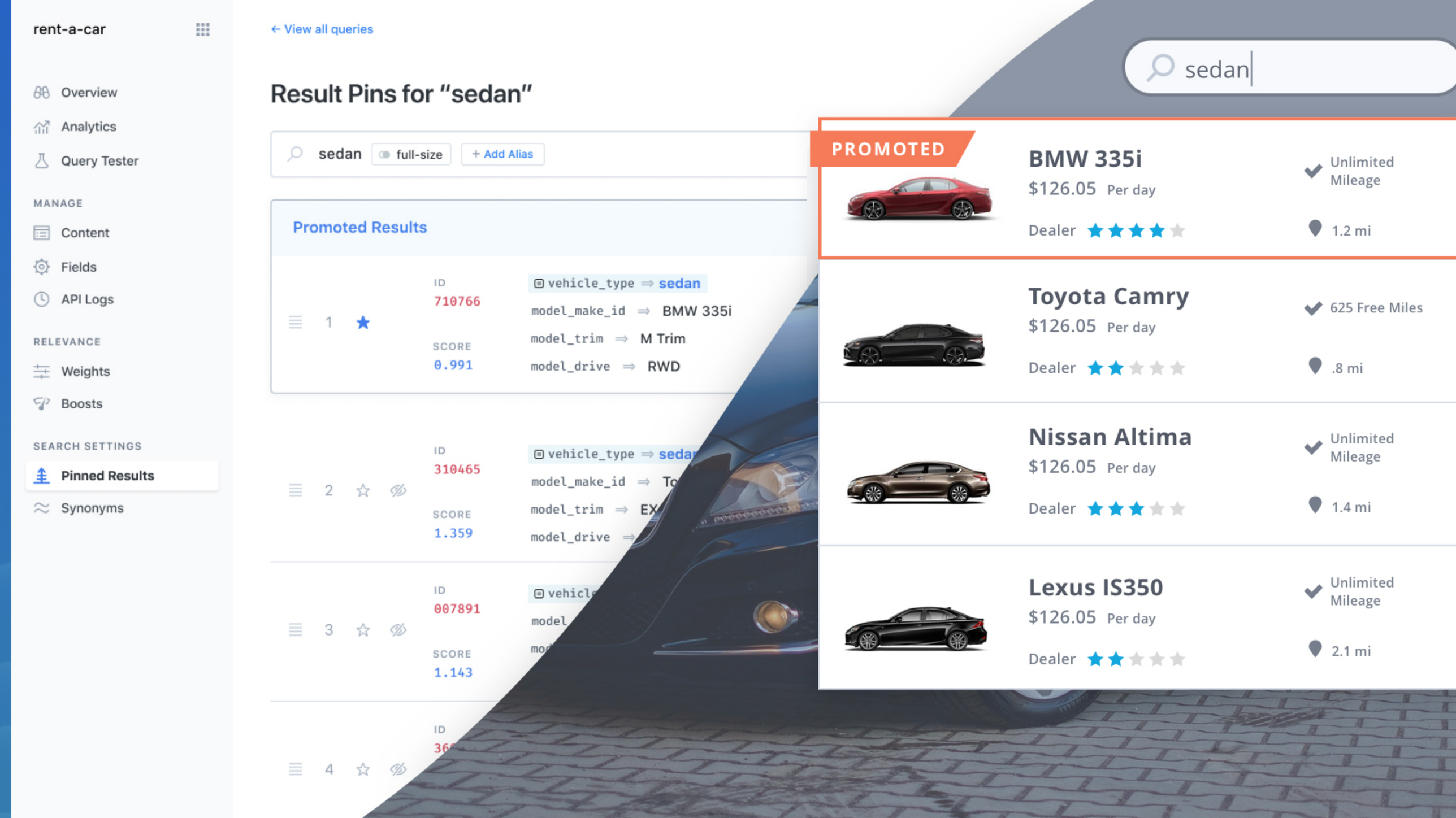
Field-Length Norm

Vector Space Model



app search


Search
Find Burritos
Near San Francisco, CA
Time range filter
Open Now
Price
$
$$
$$$
$$$$
Numeric filter
Structured search
Parking:
Street
Garage
Valet
Showing 1-10 of 100
Pagination
burrito
burritos
burrito
Stemming / Highlighting
Geo shape filter


rent-a-car
Overview
Analytics
Query Tester
MANAGE
Content
Fields
API Logs
RELEVANCE
Weights
Boosts
SEARCH SETTINGS
Pinned Results
Synonyms
← View all queries
Result Pins for "sedan"
sedan
full-size
+ Add Alias
Promoted Results
1
ID
710766
SCORE
0.991
vehicle_type ⇒ sedan
model_make_id ⇒ BMW 335i
model_trim ⇒ M Trim
model_drive ⇒ RWD
2
ID
310465
SCORE
1.359
3
ID
007891
SCORE
1.143
4
sedan
PROMOTED
BMW 335i
$126.05 Per day
Dealer
Unlimited Mileage
1.2 mi
Toyota Camry
$126.05 Per day
Dealer
625 Free Miles
.8 mi
Nissan Altima
$126.05 Per day
Dealer
Unlimited Mileage
1.4 mi
Lexus IS350
$126.05 Per day
Dealer
Unlimited Mileage
2.1 mi

Thank You!

Questions?

Philipp Krenn @xeraa

PS: Stickers

elastic

The End

elastic

More

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

```
...
"hits": [
  {
    "_index": "starwars",
    "_type": "_doc",
    "_id": "3",
    "_score": 0.41913947,
    "_source": {
      "quote": "<b>No</b>. I am your father."
    },
    "highlight": {
      "quote": [
        "<b>No</b>. I am your <tag>father</tag>."
      ]
    }
  },
  ...
```

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": 2,
    "max_score": 0.96268076,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "2",
        "_score": 0.96268076,
        "_source": {
          "quote": "Obi-Wan never told you what happened to your father."
        }
      },
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "3",
        "_score": 0.73245656,
        "_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": 2,
    "max_score": 0.56977004,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "2",
        "_score": 0.56977004,
        "_source": {
          "quote": "Obi-Wan never told you what happened to your father."
        }
      },
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "3",
        "_score": 0.31331712,
        "_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": 1.8154771,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "2",
        "_score": 1.8154771,
        "_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": 2,
    "max_score": 1.5324509,
    "hits": [
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "2",
        "_score": 1.5324509,
        "_source": {
          "quote": "Obi-Wan never told you what happened to your father."
        }
      },
      {
        "_index": "starwars",
        "_type": "_doc",
        "_id": "3",
        "_score": 0.73245656,
        "_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": "drui"
    }
  },
  "suggest": {
    "my_suggestion" : {
      "text" : "drui",
      "term" : {
        "field" : "quote"
      }
    }
  }
}
```

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

NGram

Partial matches

Trigram

Edge Gram

elastic

```
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

Combine scores

```
PUT /starwars_v42
{
  "settings": {
    "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": {
    "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"
  }
}
```

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

Aliases

Atomic `remove` **and** `add`

Point to multiple indices (read-only)

```
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": "BCDwzJ4WSw2dyoGLTzwlqw",
      "_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",
        "quote.full^5",
        "quote.ngram"
      ],
      "type": "best_fields"
    }
  }
}
```

```
"hits": [
  {
    "_index": "starwars_v42",
    "_type": "_doc",
    "_id": "1",
    "_score": 1.6022799,
    "_source": {
      "quote": "These are <em>not</em> the droids you are looking for."
    }
  },
  {
    "_index": "starwars_v42",
    "_type": "_doc",
    "_id": "2",
    "_score": 1.4997643,
    "_source": {
      "quote": "Obi-Wan never told you what happened to your father."
    }
  },
  {
    "_index": "starwars_v42",
    "_type": "_doc",
    "_id": "3",
    "_score": 0.38650417,
    "_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

elastic

```
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/_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/_termvectors/4
{
  "fields": [
    "quote.edgegram"
  ],
  "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.edgegram": "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."
      }
    }
  ]
  ...
```