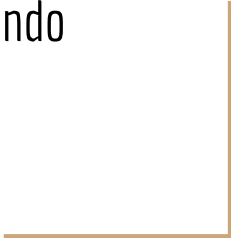


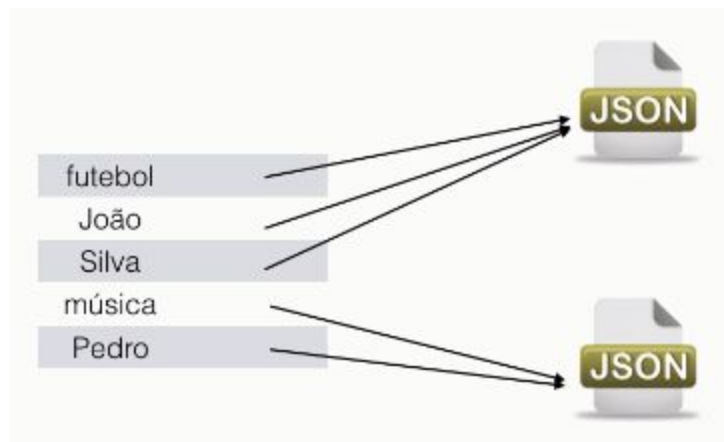


# Elasticsearch

pesquisando e analisando  
seus dados



# Sinônimos em português



# Analyzers Comuns

- Espaço em Branco (whitespace)
- Simples (simple)
- Padrão (standard)
- Idioma (e.g: portuguese, english)

# Índice invertido

T  
o  
k  
e  
n  
s

futebol

joao

silva

music

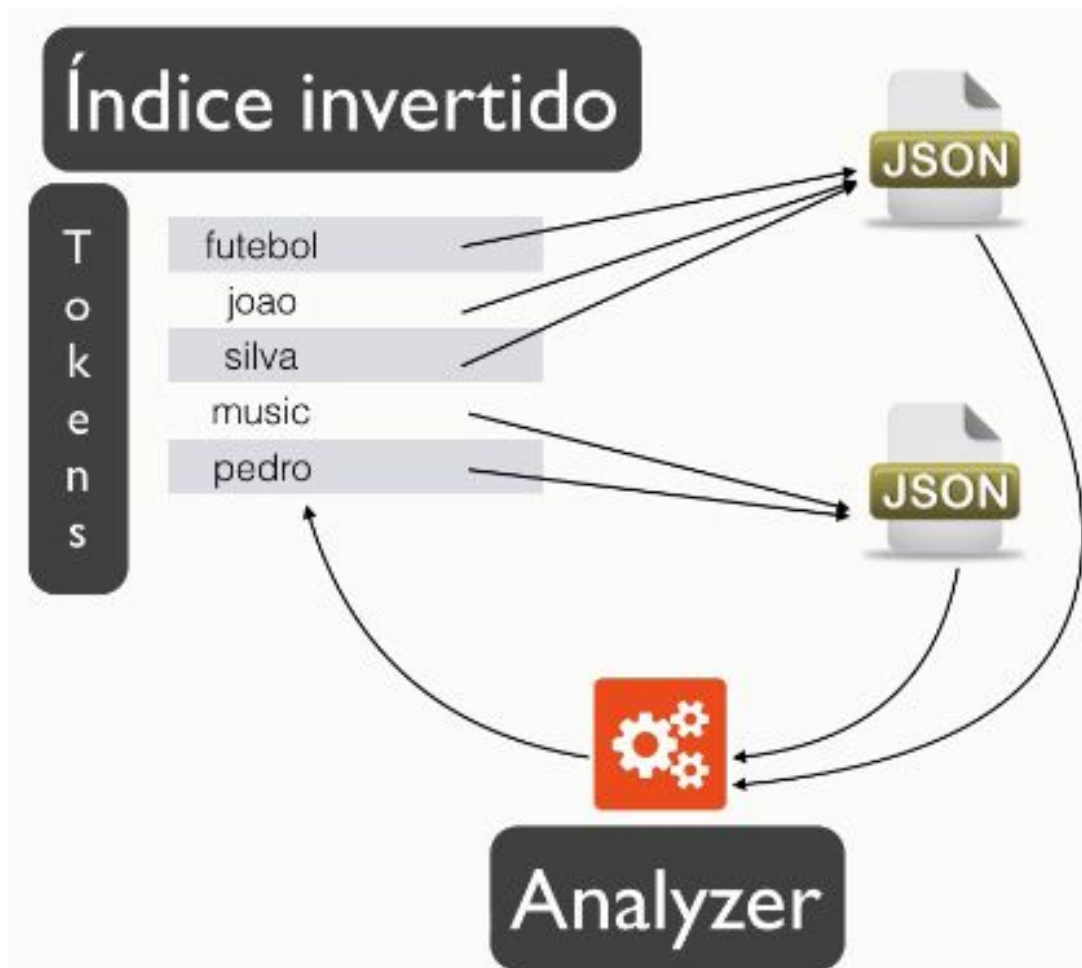
pedro

JSON

JSON



Analyzer



# Índice invertido

T  
o  
k  
e  
n  
s

futebol

joao

silva

music

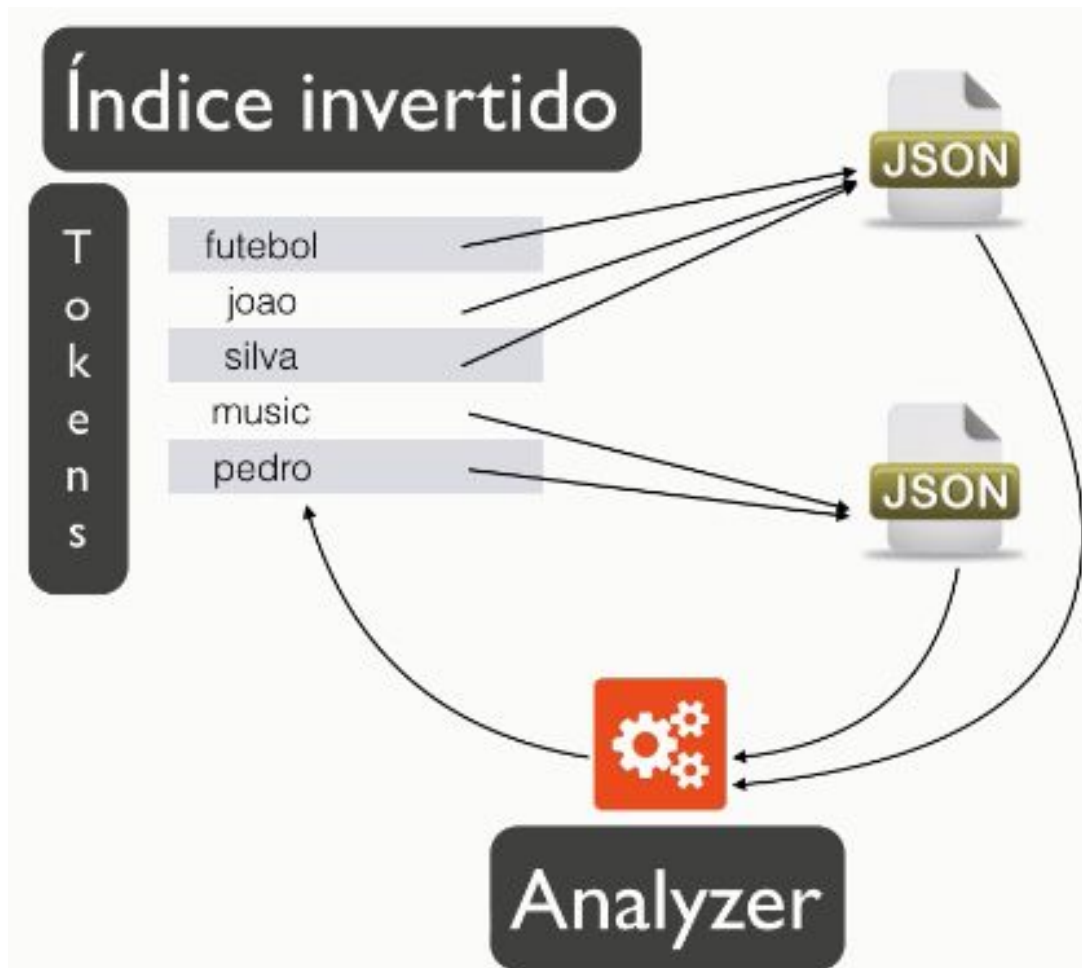
pedro

JSON

JSON



Analyzer





NOVO  
DOCUMENTO



NOVO  
DOCUMENTO



ANÁLISE



NOVO  
DOCUMENTO



ANÁLISE



DOCUMENTO  
ARMAZENADO

```
1 PUT /portuguese_example
2 {
3   "settings": {
4     "analysis": {
5       "filter": {
6         "portuguese_stop": {
7           "type": "stop",
8           "stopwords": "_portuguese_"
9         },
10        "portuguese_keywords": {
11          "type": "keyword_marker",
12          "keywords": ["exemplo"]
13        },
14        "portuguese_stemmer": {
15          "type": "stemmer",
16          "language": "light_portuguese"
17        }
18      },
19      "analyzer": {
20        "rebuilt_portuguese": {
21          "tokenizer": "standard",
22          "filter": [
23            "lowercase",
24            "portuguese_stop",
25            "portuguese_keywords",
26            "portuguese_stemmer"
27          ]
28        }
29      }
30    }
31  }
32 }
```

```
"analyzer": {  
  "portuguese": {  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "portuguese_stop",  
      "portuguese_keywords",  
      "portuguese_stemmer"  
    ]  
  }  
}
```

```
"analyzer": {  
  "portuguese": {  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "portuguese_stop",  
      "portuguese_keywords",  
      "portuguese_stemmer"  
    ]  
  }  
}
```

```
"analyzer": {  
  "portuguese": {  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "portuguese_stop",  
      "portuguese_keywords",  
      "portuguese_stemmer"  
    ]  
  }  
}
```

```
"analyzer": {  
  "portuguese": {  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "portuguese_stop",  
      "portuguese_keywords",  
      "portuguese_stemmer"  
    ]  
  }  
}
```

"O RATO roeu a ROUPA do rei de Roma"

=> token filter: lowercase =>

["o", "rato", "roeu", "a",  
"roupa", "do", "rei", "de",  
"roma";

```
"analyzer": {  
  "portuguese": {  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "portuguese_stop",  
      "portuguese_keywords",  
      "portuguese_stemmer"  
    ]  
  }  
}
```

```
"analyzer": {  
  "portuguese": {  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "portuguese_stop",  
      "portuguese_keywords",  
      "portuguese_stemmer"  
    ]  
  }  
}
```

**"o rato roeu a roupa do rei de Roma"**

**=> token filter: stop =>**

**["rato", "roeu", "roupa", "rei",  
"Roma"]**

```
"analyzer": {  
  "portuguese": {  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "portuguese_stop",  
      "portuguese_keywords",  
      "portuguese_stemmer"  
    ]  
  }  
}
```

```
"analyzer": {  
  "portuguese": {  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "portuguese_stop",  
      "portuguese_keywords",  
      "portuguese_stemmer"  
    ]  
  }  
}
```

**"Quem com ferro fere, com ferro será ferido"**

```
"analyzer": {  
  "portuguese": {  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "portuguese_stop",  
      "portuguese_keywords",  
      "portuguese_stemmer"  
    ]  
  }  
}
```

**=> token filter: stemmer =>**

**["ferro", "fere", "ferr", "ferid"]**

**"Quem com ferro fere, com ferro será ferido"**

```
"analyzer": {  
  "portuguese": {  
    "tokenizer": "standard",  
    "filter": [  
      "lowercase",  
      "portuguese_stop",  
      "portuguese_keywords",  
      "portuguese_stemmer"  
    ]  
  }  
}
```

**=> token filter: stemmer =>**

**["ferro", "fere", "ferr", "ferid"]**

 capitulo1.txt	10/12/2019 19:03
 capitulo2.txt	10/12/2019 20:45
 capitulo4.txt	14/12/2019 11:22
 capitulo5.txt	14/12/2019 13:14
 capitulo6.txt	10/12/2019 19:03
 capitulo7.txt	10/12/2019 19:03
 capitulo8.txt	10/12/2019 19:03

```
252 PUT /indice_com_sinonimo
253 {
254   "settings": {
255     "index": {
256       "number_of_shards": 3,
257       "number_of_replicas": 0
258     },
259     "analysis": {
260       "filter": {
261         "filtro_de_sinonimos": {
262           "type": "synonym",
263           "synonyms": [
264             "esporte,futebol,society,fut,pelada"
265           ]
266         }
267       },
268       "analyzer": {
269         "sinonimos": {
270           "tokenizer": "standard",
271           "filter": [
272             "lowercase",
273             "filtro_de_sinonimos"
274           ]
275         }
276       }
277     }
278   }
279 }
```

```
"filter": {  
  "filtro_de_sinonimos": {  
    "type": "synonym",  
    "synonyms": [  
      "esporte,futebol,society,fut,pelada"  
    ]  
  }  
}
```

```
GET /indice_com_sinonimo/_analyze
{
  "analyzer": "sinonimos",
  "text": "eu gosto de jogar society"
}
```

```
{
  "tokens" : [
    {
      "token" : "eu",
      "start_offset" : 0,
      "end_offset" : 2,
      "type" : "<ALPHANUM>",
      "position" : 0
    },
    {
      "token" : "gosto",
      "start_offset" : 3,
      "end_offset" : 8,
      "type" : "<ALPHANUM>",
      "position" : 1
    }
  ]
}
```

Caso #1

"text": "eu gosto de jogar society"

Caso #2

"text": "eu gosto de praticar esporte"

Caso #3

"text": "encontraram ouro na serra pelada"

# Sinônimos em português - 2

**“esporte,futebol,basquete,society,volei”**

```
GET /indice_com_sinonimo/_analyze
{
  "analyzer": "sinonimos",
  "text": "eu gosto de praticar esporte"
}
```

**“esporte,futebol,basquete,society,volei”**

**“eu gosto de jogar society”**

**“eu gosto de jogar futebol”**

**“eu gosto de praticar esporte”**

**“eu gosto de jogar basquete”**

**“eu gosto de jogar volei”**

**“Praticantes de esporte”**

**“Jogadores de basquete”**

```
"synonyms": [  
  "futebol => futebol,society",  
  "society => society,futebol",  
  "esporte => esporte,futebol,society,volei,basquete"  
]
```

**[TERMO A SER INDEXADO] => [LISTA DE TOKENS  
GERADOS NO ÍNDICE]**

**[TERMO MAIS GENÉRICO] => [LISTA DE  
TERMOS MAIS ESPECÍFICOS]**

**[COLEÇÃO] => [LISTA DE ELEMENTOS]**

**"nordeste" => "alagoas,  
bahia, ceara, maranhão,  
paraíba, pernambuco,  
piauí, rio grande do norte,  
sergipe"**

 capitulo1.txt	10/12/2019 19:03
 capitulo2.txt	10/12/2019 20:45
 capitulo4.txt	14/12/2019 11:22
 capitulo5.txt	14/12/2019 13:14
 capitulo6.txt	10/12/2019 19:03
 capitulo7.txt	10/12/2019 19:03
 capitulo8.txt	10/12/2019 19:03

```
GET /indice_com_sinonimo_2/_analyze
{
  "analyzer":"sinonimos",
  "text":"pelada"
}
```

```
GET /indice_com_sinonimo_2/_analyze
{
  "analyzer":"sinonimos",
  "text":"futebol"
}
```

**"futebol => futebol,society",**

```
GET /indice_com_sinonimo_2/_analyze
{
  "analyzer":"sinonimos",
  "text":"esporte"
}
```

```
GET /indice_com_sinonimo_2/_analyze
{
  "analyzer":"sinonimos",
  "text":"esportes"
}
```

**"esporte => esporte,futebol,society,volei,basquete",  
"esportes => esporte,futebol,society,volei,basquete"**