

12 - Advanced Search

INTRODUCING LOCATION SEARCH

- Busca de localização
- geoquery
 - Largamente suportado no ES
 - Data types:
 - geo_point
 - geo_shape
 - Queries suportadas
 - bounding_box
 - geo_distance
 - geo_shape
- bounding_box query
 - Busca em uma dada área (e.g. um retângulo)
 - definido com dois pontos de latitude e longitude
- geo_distance query
 - busca pontos em torno de um círculo dado por um ponto e um raio.
- geo_shape query

Busca pontos dentro de um polígono dado por um conjunto de pontos
Cada aresta é definida um par de pontos

GEOSPATIAL DATA TYPES

- geo_point
 - representa um ponto
 - Tem uma latitude e uma longitude

Listing 12.2 Indexing a bus stop with a location defined as a string

```
POST bus_stops/_doc
{
  "name": "London Bridge Station",
  "location": "51.07, 0.08"
}
```

← Inputs the location as a string with latitude and longitude values

-
- Existem várias formas de definir um ponto para o Elastic
 - "location" : "POINT (51.49 0.14)"
 - "location" : {"lon":-0.12, "lat":51.50}
 - "location" : [51.54, 0.23]
 - "location" : "gcpvh2bg7sff" (geohash)
- Suporta formato WKT
- Dependendo do formato, longitude ou latitude podem ter que vir antes no array.
- geo_shape
 - Representa uma forma geométrica
 - pontos, multipontos, linhas, polígonos, multipolígono.

GEOSPATIAL QUERIES

- e.g. buscar o restaurante mais próximo de casa
- geo_bounding_box query:
 - Busca documentos contidos em uma dado retângulo definido em um georetângulo
- geo_distance query:
 - Busca endereços dentro de uma certa distância de um dado ponto
- geo_shape query:
 - Acha formas geométricas dentro de outra forma geométrica (e.g. fazendas dentro de um dado território)

THE GEO_BOUIDING_BOX QUERY

- Busca endereços dentro de uma forma (círculo, retângulo, polígono, etc)

Listing 12.6 Matching restaurant locations in a georectangle

```
GET restaurants/_search
{
  "query": {
    "geo_bounding_box": {
      "location": {
        "top_left": {
          "lat": 52,
          "lon": 0.2
        },
        "bottom_right": {
          "lat": 49,
          "lon": 0.1
        }
      }
    }
  }
}
```

← Constructs a georectangle

← Sets the document's geo_point field

← Defines the top_left point, formed by a latitude/longitude pair

← Defines the bottom_right point, formed by a latitude/longitude pair

- Também podemos buscar por data types geoshape, não apenas pontos.
 - Basta buscar num index que contém geoshape e não geopoints, o resto é igual.

Listing 12.8 Geoquery with a location represented as WKT

```
GET restaurants/_search
{
  "query": {
    "bool": {
      "must": [
        {
          "match_all": {}
        }
      ],
      "filter": [
        {
          "geo_bounding_box": {
            "location": {
              "wkt": "BBOX(0.08, 0.04, 52.00, 49.00)"
            }
          }
        }
      ]
    }
  }
}
```

← Sets the coordinates in WKT format

- Não há diferença exceto de estilo em buscar por array ou WKT ou JSON.

THE GEO_DISTANCE QUERY

- Usado para buscar lista de endereços em torno de um ponto



Figure 12.5 Returning schools with a `geo_distance` query

Listing 12.9 Searching for restaurants within a given radius

```
GET restaurants/_search
{
  "query": {
    "geo_distance": {
      "distance": "175 km",
      "location": {
        "lat": 50.00,
        "lon": 0.10
      }
    }
  }
}
```

Declares the `geo_distance` query

Sets the vicinity of an area to search (distance from a central point)

Sets the central location, defined as a point on the map

- Distância pode ser em km ou milhas.

THE GEO_SHAPE QUERY

- Usado para buscar formas dentro de formas, não apenas pontos.

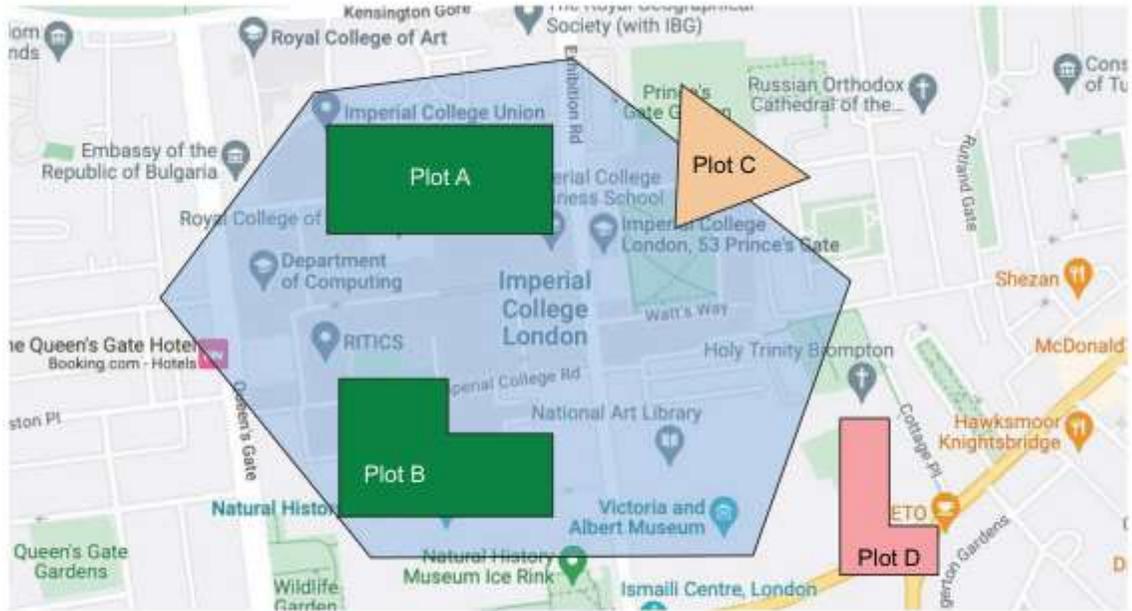


Figure 12.6 Plots of farmland in an area of London

- Podemos retornar os documentos que, em relação à forma envelope da query:
 - Têm **qualquer ponto** em comum (intersects) -- default
 - Estão inteiramente dentro da forma da query (within)
 - Contêm a forma da query (contains)
 - Estão fora (disjoint)

THE SHAPE QUERY

- Criar formas 2D
 - Desenhos técnicos
- Podemos buscá-las com queries:
 - Por exemplo, todas as formas dentro de um dado envelope:

Listing 12.13 Searching for all the shapes in a given envelope

```
GET myshapes/_search
{
  "query": {
    "shape": {
      "myshape": {
        "shape": {
          "type": "envelope",
          "coordinates": [[10, 16], [14, 10]]
        }
      }
    }
  }
}
```

Annotations for Listing 12.13:

- Specifies a shape query (points to "shape" in the outer object)
- Field on which the query is run (points to "myshape" in the outer object)
- Shape we want to construct (points to "shape" in the inner object)
- The envelope constructed by the given coordinates (points to "coordinates" in the inner object)

THE SPAN QUERY

- Busca de conjuntos de palavras em uma dada ordem com um dado intervalo de palavras entre elas.
 - **Regular query:** "Does this document contain 'apple' and 'pie'?" (Doesn't care if it's "apple pie", "pie made of apple", or "apple juice and blueberry pie").
 - **Span query:** "Does this document contain 'apple' immediately followed by 'pie'?" or "Does 'apple' appear within 3 words of 'pie'?"

Plato is my friend. Aristotle is my friend. But my greatest friend is truth
Isaac Newton



Figure 12.8
Finding a quote using
a positional query

- o
- Bom para buscar jargões técnicos específicos em textos complexos.
- Algumas variações:
 - o span_first query:
 - Busca se uma palavra aparece nas n primeiras palavras de um documento
 - o span_near query:
 - Busca se uma palavra aparece nas n primeiras palavras após (ou antes) uma outra palavra

Listing 12.16 Searching for documents with terms near each other

```
GET quotes/_search
{
  "query": {
    "span_near": {
      "clauses": [
        {
          "span_term": {
            "quote": "plato"
          }
        },
        {
          "span_term": {
            "quote": "aristotle"
          }
        }
      ],
      "slop": 3,
      "in_order": true
    }
  }
}
```

The span_near query definition with a couple of clauses

The clauses consisting of two independent span_terms to search for individual words

The slop attribute gives the acceptable number of positions between the words.

The in_order attribute sets the order of the attributes.

- o span_within query:
 - Busca termos cercados por outros dois termos
- o span_or query:
 - Combina vários span_queries com o operador OR, retorna caso dê match em qqr um deles
- o Há outros span queries ainda.

SPECIALIZED QUERIES

- ES tem várias outras queries bem nichadas
- distance_feature query:
 - o Dá um boost no score para um certo critério
 - e.g. se está mais próximo de uma localização ou uma data

Listing 12.21 Boosting scores of universities closer to London Bridge

```
GET universities/_search
{
  "query": {
    "distance_feature": {
      "field": "location",
      "origin": [-0.0860, 51.5048],
      "pivot": "10 km"
    }
  }
}
```

The location to search for

The distance_feature query declaration

Focal point where the distance is measured from an origin

Distance from the focal point

■ - pinned query:

- Fixa documentos no topo da lista independente do resultado da query

Listing 12.24 Modifying search results by adding sponsored results

```
GET iphones/_search
{
  "query": {
    "pinned": {
      "ids": ["1", "3"],
      "organic": {
        "match": {
          "name": "iPhone 12"
        }
      }
    }
  }
}
```

Specifies the pinned query

List of document IDs scored higher than the rest of the results

Carries out the query search

A match query that searches for the iPhone 12

○

- more_like_this query:

- Busca itens similares a um dado critério

Listing 12.26 Searching for More Like This documents

```
GET profiles/_search
{
  "query": {
    "more_like_this": {
      "fields": ["name", "profile"],
      "like": "Sotherby",
      "min_term_freq": 1,
      "max_query_terms": 12,
      "min_doc_freq": 1
    }
  }
}
```

Defines the more_like_this query

Searches the given input in fields

Defines the query criterion

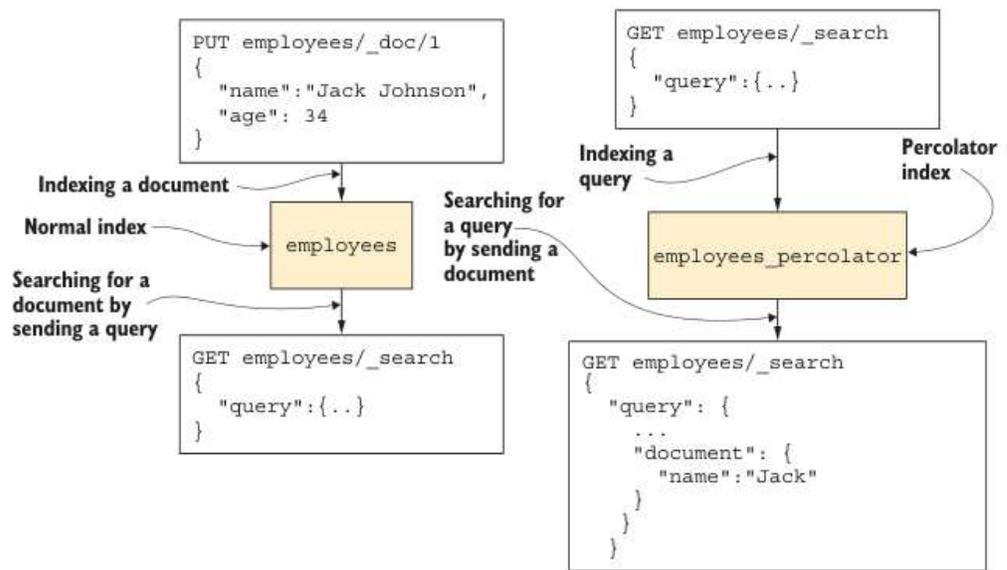
Sets the term frequency (defaults to 2)

Sets the number of terms to be selected

○

- percolate query:

- Procura por queries feitas para um dado documento
- Útil para buscar por queries que no passado retornaram nulo mas hoje contêm algum resultado (e.g. um produto que apareceu em estoque)
 - Inverso do usual



■ **Figure 12.14** Normal vs. percolate queries

- indexamos queries em um dado index percolator e depois buscamos as mesmas queries

Listing 12.32 Searching for queries in the percolator index

```
GET tech_books_percolator/_search
{
  "query": {
    "percolate": {
      "field": "query",
      "document": {
        "name": "Python in Action",
        "tags": ["Python", "Software Programming"]
      }
    }
  }
}
```

Annotations:

- Specifies the percolate query (points to "percolate")
- Sets the field's name to query (points to "field")
- Specifies the document consisting of the original book indexed into tech_books (points to "document")

Summary

- Elasticsearch supports the `geo_point` and `geo_shape` data types to work with geodata.
- Geospatial queries fetch locations and addresses using coordinates formed from longitude and latitude values.
- A `geo_bounding_box` query fetches the addresses in a georectangle, which is constructed using a pair of longitude and latitude values as the top-left and bottom-right coordinates.
- A `geo_distance` query finds locations inside a circular area with a center provided as a pivot and the radius as the distance from the pivot.
- A `geo_shape` query fetches all suitable locations inside a given envelope formed by coordinates.
- A `geo_shape` query searches 2D shapes in a rectangular coordinate system (Cartesian plane).
- Span queries are advanced queries that work with lower-level positions of individual tokens or words. Elasticsearch supports several span queries: `span_first`, `span_within`, `span_near`, and more.
- A `distance_feature` query is a specialized query where the proximity of documents to a given focal point increases the relevance score and, thus, gives the documents higher priority.
- A `pinned` query lets us bundle additional (even unmatched) documents with the original result set, potentially creating sponsored search results.
- A `more_like_this` query finds related or similar-looking results.
- A `percolate` query lets us notify users about their negative searches at a future date.