

Atividade 3 Marco Carujo

July 20, 2020

```
[25]: %load_ext lab_black
```

0.0.1 Aluno Marco Antonio Moreira Carujo

1 ATIVIDADE 3

Utilizando os arquivos baixados no Lab 1 e a estrutura do Lab 2 para busca-los no Elasticsearch

E tambem utilizar a biblioteca python **ElasticSearch** (baixo nivel), cuja a documentação pode ser acessada no link [aqui](#) para fazer uma busca genérica.

```
[26]: from elasticsearch import Elasticsearch

es = Elasticsearch()
if es.ping():
    print("ElasticSearch ONLINE !")
```

ElasticSearch ONLINE !

2 Busca

Search(body=None, index=None, params=None, headers=None) - Returns results matching a query. <https://www.elastic.co/guide/en/elasticsearch/reference/master/search-search.html>

Parameters: - **body** – The search definition using the Query DSL - **index** – A comma-separated list of index names to search; use `_all` or empty string to perform the operation on all indices - **__source** – True or false to return the `__source` field or not, or a list of fields to return - **__source_excludes** – A list of fields to exclude from the returned `__source` field - **__source_includes** – A list of fields to extract and return from the `__source` field - **allow_no_indices** – Whether to ignore if a wildcard indices expression resolves into no concrete indices. (This includes *all string* or *when no indices have been specified*) - **allow_partial_search_results** – Indicate if an error should be returned if there is a partial search failure or timeout Default: True - **analyze_wildcard** – Specify whether wildcard and prefix queries should be analyzed (default: false) - **analyzer** – The analyzer to use for the query string - **batched_reduce_size** – The number of shard results that should be reduced at once on the coordinating node. This value should be used as a protection mechanism to reduce the memory overhead per search request if the potential number of shards in the request can be large. Default: 512 - **ccs_minimize_roundtrips** – Indicates whether network round-trips should be minimized as part of cross-cluster search requests execution Default: true - **default_operator** – The

default operator for query string query (AND or OR) Valid choices: AND, OR Default: OR - df - The field to use as default where no field prefix is given in the query string - docvalue_fields - A comma-separated list of fields to return as the docvalue representation of a field for each hit - expand_wildcards - Whether to expand wildcard expression to concrete indices that are open, closed or both. Valid choices: open, closed, hidden, none, all Default: open - explain - Specify whether to return detailed information about score computation as part of a hit - from - Starting offset (default: 0) - ignore_throttled - Whether specified concrete, expanded or aliased indices should be ignored when throttled - ignore_unavailable - Whether specified concrete indices should be ignored when unavailable (missing or closed) - lenient - Specify whether format-based query failures (such as providing text to a numeric field) should be ignored - max_concurrent_shard_requests - The number of concurrent shard requests per node this search executes concurrently. This value should be used to limit the impact of the search on the cluster in order to limit the number of concurrent shard requests Default: 5 - pre_filter_shard_size - A threshold that enforces a pre-filter roundtrip to prefilter search shards based on query rewriting if the number of shards the search request expands to exceeds the threshold. This filter roundtrip can limit the number of shards significantly if for instance a shard can not match any documents based on its rewrite method ie. if date filters are mandatory to match but the shard bounds and the query are disjoint. - preference - Specify the node or shard the operation should be performed on (default: random) - q - Query in the Lucene query string syntax - request_cache - Specify if request cache should be used for this request or not, defaults to index level setting - rest_total_hits_as_int - Indicates whether hits.total should be rendered as an integer or an object in the rest search response - routing - A comma-separated list of specific routing values - scroll - Specify how long a consistent view of the index should be maintained for scrolled search - search_type - Search operation type Valid choices: query_then_fetch, dfs_query_then_fetch - seq_no_primary_term - Specify whether to return sequence number and primary term of the last modification of each hit - size - Number of hits to return (default: 10) - sort - A comma-separated list of : pairs - stats - Specific 'tag' of the request for logging and statistical purposes - stored_fields - A comma-separated list of stored fields to return as part of a hit - suggest_field - Specify which field to use for suggestions - suggest_mode - Specify suggest mode Valid choices: missing, popular, always Default: missing - suggest_size - How many suggestions to return in response - suggest_text - The source text for which the suggestions should be returned - terminate_after - The maximum number of documents to collect for each shard, upon reaching which the query execution will terminate early. - timeout - Explicit operation timeout - track_scores - Whether to calculate and return scores even if they are not used for sorting - track_total_hits - Indicate if the number of documents that match the query should be tracked. A number can also be specified, to accurately track the total hit count up to the number. - typed_keys - Specify whether aggregation and suggester names should be prefixed by their respective types in the response - version - Specify whether to return document version as part of a hit

```
[27]: import pandas as pd
      from IPython.display import display

      # função para auxiliar o print
      def eprint(response):
          results = []
          for doc in response["hits"]["hits"]:
              line = {}
```

```

line["id"] = doc["_id"]
line["score"] = doc["_score"]
line["titulo"] = doc["_source"]["titulo"]
line["autor"] = doc["_source"]["autor"]
line["texto"] = doc["_source"]["texto"]
results.append(line)
display(pd.DataFrame.from_dict(results, orient="columns"))

```

2.0.1 Procurando em qualquer campo

Passando como busca “realidade aumentada” em qualquer campo do documento e limitando o tamanho da busca como 10 documentos.

```
[28]: res = es.search(q="realidade aumentada", index="artigos-index", size=10)
      eprint(res)
```

	autor	id \
0	Alan Ferreira Alves, Cícero Francisco Bezerra...	AemBZHMB9RWPpyMKDEM_y
1	Eduarda Queiroz, Rafaela Moura, Ellen Souza	AOmBZHMB9RWPpyMKDEM_y
2	Deyse Mara Romualdo Soares, Gabriela Teles, Ro...	KemBZHMB9RWPpyMKDEM_z
3	Elvis Medeiros de Melo, Dennys Leite Maia	DemBZHMB9RWPpyMKDEM_y
4	Aladir Ferreira da Silva Júnior, Leizer Fernan...	MOmBZHMB9RWPpyMKDEM_z
5	Amanda Maria D. de Oliveira, Adriana Virgínia ...	RemBZHMB9RWPpyMKDEM_z
6	Lucas O. Lopes, Paula R. P. Oliveira, Karoline...	JumBZHMB9RWPpyMKDEM_z
7	Elvis Medeiros de Melo, Dennys Leite Maia	DOmBZHMB9RWPpyMKDEM_y
8	Danielle Moura Santos, Raphaela Delmondes do N...	JemBZHMB9RWPpyMKDEM_z
9	Daniel de Sant'anna Martins, Anderson Paulo da...	B-mBZHMB9RWPpyMKDEM_y

	score	texto \
0	7.623015	IV Congresso sobre Tecnologias na Educação (Ct...
1	4.881375	IV Congresso sobre Tecnologias na Educação (Ct...
2	2.422138	Tecnologias Digitais nos Processos de Ensin...
3	2.113137	O Uso de Dispositivos Móveis para o Tratamen...
4	1.927623	Programando a Várias Mãos: um Relato de Ex...
5	1.565196	Software Educativo Encantando Cordel Amanda Ma...
6	1.290093	O "Maker" na Escola: uma Reflexão sobre Tecnol...
7	1.156423	O Uso do Smartphone por Estudantes de Pós-Gra...
8	1.100505	Tecnologia Digital na Aprendizagem Baseada em ...
9	1.094950	LudoEscola: a Gamificação como Ferramenta...

	titulo
0	Investigação de Novas Estratégias para o Ensin...
1	Como a Realidade Aumentada tem Auxiliado no Pr...
2	Tecnologias Digitais nos Processos de Ensino e...
3	O Uso de Dispositivos Móveis para o Tratamento...
4	Programando a Várias Mãos: um Relato de Experi...
5	Software Educativo Encantando Cordel
6	O "Maker" na Escola: uma Reflexão sobre Tecnol...

```

7 0 Uso do Smartphone por Estudantes de Pós-Grad...
8 Tecnologia Digital na Aprendizagem Baseada em ...
9 LudoEscola: a Gamificação como Ferramenta de T...

```

2.1 Procurando no campo ‘autor’

```
[29]: res = es.search(q="autor:queiroz", index="artigos-index", size=10)
      eprint(res)
```

```

                                autor                                id \
0  Eduarda Queiroz, Rafaela Moura, Ellen Souza  AOmBZHMB9RWPYMKDEM_y

                                score                                texto \
0  3.527312  IV Congresso sobre Tecnologias na Educação (Ct...

                                titulo
0  Como a Realidade Aumentada tem Auxiliado no Pr...

```

2.2 Procurando no campo ‘titulo’

```
[30]: res = es.search(q="titulo:software", index="artigos-index", size=10)
      eprint(res)
```

```

                                autor                                id \
0  Amanda Maria D. de Oliveira, Adriana Virgínia ...  RemBZHMB9RWPYMKDEM_z
1  Sonia Azevedo de Medeiros, Taciana Pontual Fal...  H-mBZHMB9RWPYMKDEM_y

                                score                                texto \
0  3.244864  Software Educativo Encantando Cordel Amanda Ma...
1  2.498678  0 Uso do Software SEBRAN ABC como Recurso Pote...

                                titulo
0                                     Software Educativo Encantando Cordel
1  0 Uso do Software SEBRAN ABC como Recurso Pote...

```

2.3 Procurando no campo ‘texto’

```
[31]: res = es.search(q="texto:software sebran", index="artigos-index", size=10)
      eprint(res)
```

```

                                autor                                id \
0  Sonia Azevedo de Medeiros, Taciana Pontual Fal...  H-mBZHMB9RWPYMKDEM_y
1  Paul Symon Ribeiro Rocha, Carlos Vieira Ramos,...  BOmBZHMB9RWPYMKDEM_y
2      Eduarda Queiroz, Rafaela Moura, Ellen Souza  AOmBZHMB9RWPYMKDEM_y
3  Amanda Maria D. de Oliveira, Adriana Virgínia ...  RemBZHMB9RWPYMKDEM_z

```

4	Ariane N. Rodrigues, Arianne Sarmento Torcate,...	H0mBZHMB9RWPYMKDEM_y
5	Raquel S. Freire, Mércia V. Campos Figueiredo...	S0mBZHMB9RWPYMKDEM_z
6	Ana Carolina Candido de Melo, Ellen Polliana R...	HemBZHMB9RWPYMKDEM_y
7	Elvis Medeiros de Melo, Dennys Leite Maia	DemBZHMB9RWPYMKDEM_y
8	Kátia Carla S. Silva, Mariana Boulitreau S. C....	NemBZHMB9RWPYMKDEM_z
9	Elvis Medeiros de Melo, Dennys Leite Maia	D0mBZHMB9RWPYMKDEM_y

	score	texto \
0	7.665312	0 Uso do Software SEBRAN ABC como Recurso Pote...
1	3.777315	A Utilização de Softwares no Ensino de Matem...
2	1.484964	IV Congresso sobre Tecnologias na Educação (Ct...
3	1.479431	Software Educativo Encantando Cordel Amanda Ma...
4	1.433400	A Relação entre a Ludicidade e a Cognição e...
5	1.414263	Recurso Educacional Digital (RED) para Trabal...
6	1.380582	A Tecnologia Assistiva e a Inclusão de Educan...
7	1.295870	0 Uso de Dispositivos Móveis para o Tratamen...
8	1.252132	Gamificação como Tecnologia Educacional Auxili...
9	1.215152	0 Uso do Smartphone por Estudantes de Pós-Gra...

	titulo
0	0 Uso do Software SEBRAN ABC como Recurso Pote...
1	A Utilização de Softwares no Ensino de Matemát...
2	Como a Realidade Aumentada tem Auxiliado no Pr...
3	Software Educativo Encantando Cordel
4	A Relação entre a Ludicidade e a Cognição em u...
5	Recurso Educacional Digital (RED) para Trabalh...
6	A Tecnologia Assistiva e a Inclusão de Educand...
7	0 Uso de Dispositivos Móveis para o Tratamento...
8	Gamificação como Tecnologia Educacional Auxili...
9	0 Uso do Smartphone por Estudantes de Pós-Grad...