



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА КОМПЬЮТЕРНЫЕ СИСТЕМЫ И СЕТИ (ИУ6)

НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.04.01 Информатика и вычислительная техника

МАГИСТЕРСКАЯ ПРОГРАММА 09.04.01/07 Интеллектуальные системы анализа,
обработки и интерпретации больших данных

О Т Ч Е Т

по лабораторной работе № 5

Название: Работа с документацией NoSQL БД Elasticsearch и реализация
процесса data science

Дисциплина: Технология параллельных систем баз данных

Студент

ИУ6-12М

(Группа)

(Подпись, дата)

Д.С. Каткова

(И.О. Фамилия)

Преподаватель

(Подпись, дата)

А.Д. Пономарев

(И.О. Фамилия)

Москва, 2023

Цель лабораторной работы – изучение модели представления данных и способы работы с документацией NoSQL базой данных Elasticsearch. Получение навыков инсталляции, индексации и поиска в Elasticsearch, реализация этапов процесса Data Science, тестирование работы с большим объёмом данных.

Ход работы

Запрос 1. С тремя обязательными ключевыми словами

Текст первого запроса 1 представлен на рисунке 1.

```
1 from elasticsearch import Elasticsearch
2 import pprint
3
4 client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
5 indexName = "medical"
6 docType = "diseases"
7 searchFrom = 0
8 searchSize = 3
9
10 # 1
11 searchBody = {
12     "_source": ["name"],
13     "query": {
14         "simple_query_string": {
15             "query": '+fatigue+fever+joint pain',
16             "fields": ["fulltext", "title^5", "name^10"]
17         }
18     }
19 }
20 result = client.search(index=indexName, body=searchBody, from_=searchFrom, size=searchSize)
21
22 pprint.pprint(result)
```

Рисунок 1 – Текст 1 запроса

Результат выполнения показан на рисунке 2.

```
{'_shards': {'failed': 0, 'skipped': 0, 'successful': 1, 'total': 1},
 'hits': {'hits': [{'_id': 'Chronic fatigue syndrome',
                    '_index': 'medical',
                    '_score': 117.35565,
                    '_source': {'name': 'Chronic fatigue syndrome'},
                    '_type': 'diseases'},
                  {'_id': 'Chronic fatigue immune dysfunction syndrome',
                    '_index': 'medical',
                    '_score': 100.70878,
                    '_source': {'name': 'Chronic fatigue immune dysfunction '
                                     'syndrome'},
                    '_type': 'diseases'},
                  {'_id': 'Encephalomyelitis, myalgic',
                    '_index': 'medical',
                    '_score': 39.607258,
                    '_source': {'name': 'Encephalomyelitis, myalgic'},
                    '_type': 'diseases'}]},
 'max_score': 117.35565,
 'total': {'relation': 'eq', 'value': 48}},
 'timed_out': False,
 'took': 1531}
```

Рисунок 2 – Результат выполнения 1 запроса

Первым результатом поиска стала болезнь - Синдром хронической усталости (Chronic fatigue syndrome). Волчанка (Lupus erythematosus) не входит в число первых 3 из 48 подходящих болезней. Результаты сортируются по рейтингу соответствия (переменная `_score`), в нем учитывается, насколько хорошо болезнь соответствует запросу, сколько раз было найдено ключевое слово, назначенные веса и т.д.

Запрос 2. Как изменятся результаты поиска болезней, если в "query" добавить новый симптом rash (сыпь)?

Текст запроса 2 представлен на рисунке 3.

```
1 from elasticsearch import Elasticsearch
2 import pprint
3
4 client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
5 indexName = "medical"
6 docType = "diseases"
7 searchFrom = 0
8 searchSize = 3
9
10 # 1
11 searchBody = {
12     "_source": ["name"],
13     "query": {
14         "simple_query_string": {
15             "query": '+fatigue+fever+rash+joint pain',
16             "fields": ["fulltext", "title^5", "name^10"]
17         }
18     }
19 }
20 result = client.search(index=indexName, body=searchBody, from_=searchFrom, size=searchSize)
21
22 pprint.pprint(result)
```

Рисунок 3 – Текст 2 запроса

Результат выполнения показан на рисунке 4.

```
{'_shards': {'failed': 0, 'skipped': 0, 'successful': 1, 'total': 1},
 'hits': {'hits': [{'_id': 'Chikungunya',
                    '_index': 'medical',
                    '_score': 13.879888,
                    '_source': {'name': 'Chikungunya'},
                    '_type': 'diseases'},
                  {'_id': 'Cutaneous lupus erythematosus',
                    '_index': 'medical',
                    '_score': 13.509659,
                    '_source': {'name': 'Cutaneous lupus erythematosus'},
                    '_type': 'diseases'},
                  {'_id': 'Lupus erythematosus',
                    '_index': 'medical',
                    '_score': 13.509659,
                    '_source': {'name': 'Lupus erythematosus'},
                    '_type': 'diseases'}],
         'max_score': 13.879888,
         'total': {'relation': 'eq', 'value': 16}},
 'timed_out': False,
 'took': 99}
```

Рисунок 4 – Результат выполнения 2 запроса

Первым результатом поиска стала болезнь – Лихорадка Чикунгунья (Chikungunya). Два вида волчанки (Lupus erythematosus и Cutaneous lupus

erythematosus) теперь входят в число первых 3 из 48 подходящих болезней. Результаты сортируются по рейтингу соответствия (переменная `_score`), в нем учитывается, насколько хорошо болезнь соответствует запросу, сколько раз было найдено ключевое слово, назначенные веса и т.д.

Запрос 3. Как изменятся результаты поиска болезней, если в "query" добавить ещё один симптом `sensitivity to sunlight` (чувствительность к солнечному свету)?

Текст запроса 3 представлен на рисунке 5.

```
1 from elasticsearch import Elasticsearch
2 import pprint
3
4 client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
5 indexName = "medical"
6 docType = "diseases"
7 searchFrom = 0
8 searchSize = 3
9
10 # 1
11 searchBody = {
12     "_source": ["name"],
13     "query": {
14         "simple_query_string": {
15             "query": '+fatigue+fever+rash+"sensitivity to sunlight"+"joint pain"',
16             "fields": ["fulltext", "title^5", "name^10"]
17         }
18     }
19 }
20 result = client.search(index=indexName, body=searchBody, from_=searchFrom, size=searchSize)
21
22 pprint.pprint(result)
23
```

Рисунок 5 – Текст 3 запроса

Результат выполнения показан на рисунке 6.

```
{'_shards': {'failed': 0, 'skipped': 0, 'successful': 1, 'total': 1},
 'hits': {'hits': [{'_id': 'Cutaneous lupus erythematosus',
                    '_index': 'medical',
                    '_score': 20.492167,
                    '_source': {'name': 'Cutaneous lupus erythematosus'},
                    '_type': 'diseases'},
                  {'_id': 'Lupus erythematosus',
                    '_index': 'medical',
                    '_score': 20.492167,
                    '_source': {'name': 'Lupus erythematosus'},
                    '_type': 'diseases'}],
 'max_score': 20.492167,
 'total': {'relation': 'eq', 'value': 2}},
 'timed_out': False,
 'took': 108}
```

Рисунок 6 – Результат выполнения 3 запроса

Два вида волчанки (`Lupus erythematosus` и `Cutaneous lupus erythematosus`) теперь являются единственными болезнями, подходящими по заданному условию поиска.

Запрос 4. Выявить симптомы болезней, в имени которых встречается слово diabetes (диабет).

Текст запроса 4 представлен на рисунке 7.

```
1 from elasticsearch import Elasticsearch
2 import pprint
3
4 client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
5 indexName = "medical"
6 docType = "diseases"
7 searchFrom = 0
8 searchSize = 3
9
10 searchBody={
11     "_source": ["name"],
12     "query": {
13         "query_string" : {
14             "query": "name: diabetes"
15         }
16     },
17     "aggregations" : {
18         "DiseaseKeywords" : {
19             "significant_terms" : { "field" : "fulltext", "size" : 10 }
20         }
21     }
22 }
23 result=client.search(index=indexName, body=searchBody, from_ = searchFrom, size=searchSize)
24
25 pprint.pprint(result)
26
```

Рисунок 7 – Текст 4 запроса

Результат выполнения показан на рисунке 8.

```
{ '_shards': {'failed': 0, 'skipped': 0, 'successful': 1, 'total': 1},
  'aggregations': {'DiseaseKeywords': {'bg_count': 5916,
    'buckets': [{'bg_count': 11,
      'doc_count': 10,
      'key': 'siphon',
      'score': 55.64185034528427},
      {'bg_count': 9,
      'doc_count': 9,
      'key': 'diabainein',
      'score': 55.11446409989595},
      {'bg_count': 9,
      'doc_count': 9,
      'key': 'bainein',
      'score': 55.11446409989595},
      {'bg_count': 8,
      'doc_count': 8,
      'key': 'ndi',
      'score': 48.99063475546306},
      {'bg_count': 11,
      'doc_count': 9,
      'key': 'passer',
      'score': 45.04086652161575},
      {'bg_count': 7,
      'doc_count': 7,
      'key': 'sapere',
      'score': 42.86680541103018},
      {'bg_count': 7,
      'doc_count': 7,
      'key': 'tasty',
      'score': 42.86680541103018},
      {'bg_count': 7,
      'doc_count': 7,
      'key': 'sapidus',
      'score': 42.86680541103018},
      {'bg_count': 10,
      'doc_count': 8,
      'key': 'avp',
      'score': 39.140894901144634},
      {'bg_count': 6,
      'doc_count': 6,
      'key': 'defect,the',
      'score': 36.74297606659729}],
    'doc_count': 31}},
```

```

{'hits': {'hits': [{ '_id': 'Diabetes insipidus',
                     '_index': 'medical',
                     '_score': 6.0656104,
                     '_source': {'name': 'Diabetes insipidus'},
                     '_type': 'diseases'},
                   { '_id': 'Diabetes mellitus',
                     '_index': 'medical',
                     '_score': 6.0656104,
                     '_source': {'name': 'Diabetes mellitus'},
                     '_type': 'diseases'},
                   { '_id': 'Phosphate diabetes',
                     '_index': 'medical',
                     '_score': 6.0656104,
                     '_source': {'name': 'Phosphate diabetes'},
                     '_type': 'diseases'}]},
  'max_score': 6.0656104,
  'total': {'relation': 'eq', 'value': 31}},
'timed_out': False,
'took': 4707}

```

Рисунок 8 – Результат выполнения 4 запроса

Были выделены 10 симптомов, которые чаще всего присутствуют при заболеваниях, имеющих в названии слово «diabetes». Также выделены 3 подвида диабета из подходящих 31 болезней, обнаруженных по указанному запросу.

Пятью наиболее важными симптомами являются siphon, diabainein, bainein, ndi, passer. Был проанализирован 31 документ.

Запрос выполнялся 4707 миллисекунд.

Запрос 5. Выявить симптомы болезней, в имени которых встречаются слова diabetes (диабет) и insipidus (несахарный).

Текст запроса 5 представлен на рисунке 9.

```

1 from elasticsearch import Elasticsearch
2 import pprint
3
4 client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
5 indexName = "medical"
6 docType = "diseases"
7 searchFrom = 0
8 searchSize = 3
9
10 searchBody={
11     "_source": ["name"],
12     "query": {
13         "query_string" : {
14             "query": "name: diabetes+insipidus"
15         }
16     },
17     "aggregations" : {
18         "DiseaseKeywords" : {
19             "significant_terms" : { "field" : "fulltext", "size" : 10 }
20         }
21     }
22 }
23 result=client.search(index=indexName, body=searchBody, from_ = searchFrom, size=searchSize)
24
25 pprint.pprint(result)

```

Рисунок 9 – Текст 5 запроса

Результат выполнения показан на рисунке 10.

```
{ '_shards': {'failed': 0, 'skipped': 0, 'successful': 1, 'total': 1},
  'aggregations': {'DiseaseKeywords': {'bg_count': 5916,
    'buckets': [{'bg_count': 11,
      'doc_count': 10,
      'key': 'siphon',
      'score': 55.64185034528427},
      {'bg_count': 9,
      'doc_count': 9,
      'key': 'diabainein',
      'score': 55.11446409989595},
      {'bg_count': 9,
      'doc_count': 9,
      'key': 'bainein',
      'score': 55.11446409989595},
      {'bg_count': 8,
      'doc_count': 8,
      'key': 'ndi',
      'score': 48.99063475546306},
      {'bg_count': 11,
      'doc_count': 9,
      'key': 'passer',
      'score': 45.04086652161575},
      {'bg_count': 7,
      'doc_count': 7,
      'key': 'sapere',
      'score': 42.86680541103018},
      {'bg_count': 7,
      'doc_count': 7,
      'key': 'tasty',
      'score': 42.86680541103018},
      {'bg_count': 7,
      'doc_count': 7,
      'key': 'sapidus',
      'score': 42.86680541103018},
      {'bg_count': 10,
      'doc_count': 8,
      'key': 'avp',
      'score': 39.140894901144634},
      {'bg_count': 6,
      'doc_count': 6,
      'key': 'defect.the',
      'score': 36.74297606659729}],
    'doc_count': 31}},
  'hits': {'hits': [{'_id': 'Diabetes insipidus',
    '_index': 'medical',
    '_score': 13.403998,
    '_source': {'name': 'Diabetes insipidus'},
    '_type': 'diseases'},
    {'_id': 'Central diabetes insipidus',
    '_index': 'medical',
    '_score': 11.578066,
    '_source': {'name': 'Central diabetes insipidus'},
    '_type': 'diseases'},
    {'_id': 'Nephrogenic diabetes insipidus',
    '_index': 'medical',
    '_score': 11.578066,
    '_source': {'name': 'Nephrogenic diabetes insipidus'},
    '_type': 'diseases'}],
    'max_score': 13.403998,
    'total': {'relation': 'eq', 'value': 31}},
  'timed_out': False,
  'took': 171}
```

Рисунок 10 – Результат выполнения 5 запроса

Был выполнен поиск симптомов для болезней, в имени которых встречаются слова diabetes (диабет) и insipidus (несахарный). Симптомы аналогичны, однако можно заметить, что в качестве трех первых наиболее подходящих болезней указаны только те подвиды диабета, в название которых входят слова diabetes и insipidus.

Запрос 6. Выполните нечёткий поиск (fuzzy) для имени болезни diabetse (в названии болезни случайно поменяли местами две последние буквы).

Текст запроса 6 представлен на рисунке 11.

```
1 from elasticsearch import Elasticsearch
2 import pprint
3
4 client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
5 indexName = "medical"
6 docType = "diseases"
7 searchFrom = 0
8 searchSize = 3
9
10 searchBody={
11     "_source": ["name"],
12     "query": {
13         "fuzzy" : {
14             "name": "diabetse"
15         }
16     },
17     "aggregations" : {
18         "DiseaseKeywords" : {
19             "significant_terms" : { "field" : "fulltext", "size" : 10 }
20         }
21     }
22 }
23
24 result=client.search(index=indexName, body=searchBody, from_ = searchFrom, size=searchSize)
25
26 pprint.pprint(result)
```

Рисунок 11 – Текст 6 запроса

Результат выполнения показан на рисунке 12.

```
{'_shards': {'failed': 0, 'skipped': 0, 'successful': 1, 'total': 1},
 'aggregations': {'DiseaseKeywords': {'bg_count': 5916,
                                       'buckets': [{'bg_count': 11,
                                                     'doc_count': 10,
                                                     'key': 'siphon',
                                                     'score': 43.6178107606679},
                                                     {'bg_count': 9,
                                                     'doc_count': 9,
                                                     'key': 'diabainein',
                                                     'score': 43.2073469387755},
                                                     {'bg_count': 9,
                                                     'doc_count': 9,
                                                     'key': 'bainein',
                                                     'score': 43.2073469387755},
                                                     {'bg_count': 8,
                                                     'doc_count': 8,
                                                     'key': 'ndi',
                                                     'score': 38.4065306122449},
                                                     {'bg_count': 11,
                                                     'doc_count': 9,
                                                     'key': 'passer',
                                                     'score': 35.30471243042671},
                                                     {'bg_count': 7,
                                                     'doc_count': 7,
                                                     'key': 'sapere',
                                                     'score': 33.60571428571429},
                                                     {'bg_count': 7,
                                                     'doc_count': 7,
                                                     'key': 'tasty',
                                                     'score': 33.60571428571429},
                                                     {'bg_count': 7,
                                                     'doc_count': 7,
                                                     'key': 'sapidus',
                                                     'score': 33.60571428571429},
                                                     {'bg_count': 10,
                                                     'doc_count': 8,
                                                     'key': 'avp',
                                                     'score': 30.679510204081634},
                                                     {'bg_count': 6,
                                                     'doc_count': 6,
                                                     'key': 'vaptanes',
                                                     'score': 28.804897959183673}],
                                       'doc_count': 35}}}
```



```

'hits': {'hits': [{'_id': 'Diabetes insipidus',
                    '_index': 'medical',
                    '_score': 5.307409,
                    '_source': {'name': 'Diabetes insipidus'},
                    '_type': 'diseases'},
                  {'_id': 'Diabetes mellitus',
                    '_index': 'medical',
                    '_score': 5.307409,
                    '_source': {'name': 'Diabetes mellitus'},
                    '_type': 'diseases'},
                  {'_id': 'Phosphate diabetes',
                    '_index': 'medical',
                    '_score': 5.307409,
                    '_source': {'name': 'Phosphate diabetes'},
                    '_type': 'diseases'}]},
'max_score': 5.307409,
'total': {'relation': 'eq', 'value': 35}},
'timed_out': False,
'took': 278}

```

Рисунок 12 – Результат выполнения 6 запроса

Нечёткий поиск (fuzzy) для имени болезни diabetse успешно выполнен.

Запрос 7. С фильтрацией.

Текст запроса 7 представлен на рисунке 13.

```

1 from elasticsearch import Elasticsearch
2 import pprint
3
4 client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
5 indexName = "medical"
6 docType = "diseases"
7
8 filterQuery={
9     "_source": ["name", "title", "fulltext"],
10    "query": {
11        "constant_score" : {
12            "filter": {
13                "term": {
14                    "title": "headache"
15                }
16            }
17        }
18    }
19 }
20
21 result=client.search(index=indexName, body=filterQuery)
22
23 pprint.pprint(result)

```

Рисунок 13 – Текст 7 запроса

Результат выполнения показан на рисунке 14.

```
{'_shards': {'failed': 0, 'skipped': 0, 'successful': 1, 'total': 1},
'aggregations': {'DiseaseKeywords': {'bg_count': 5916,
                                      'buckets': [],
                                      'doc_count': 2}},
'hits': {'hits': [{'_id': 'Cluster headache',
                    '_index': 'medical',
                    '_score': 1.0,
                    '_source': {'fulltext': 'Cluster headache (CH) is a '
                                           'neurological disorder '
                                           'characterized by recurrent severe '
                                           'headaches on one side of the '
                                           'head, typically around the eye. '
                                           'There is often accompanying eye '
                                           'watering, nasal congestion, or '
                                           'swelling around the eye on the '
                                           'affected side. These symptoms '
                                           'typically last 15 minutes to 3 '

```

Рисунок 14 – Результат выполнения 7 запроса

Запрос 8. С сортировкой.

Текст запроса 8 представлен на рисунке 15.

```
1 from elasticsearch import Elasticsearch
2 import pprint
3
4 client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
5 indexName = "medical"
6 docType = "diseases"
7
8 sortQuery={
9     "_source": ["name", "title"],
10     "sort": [{"title": "desc"}]
11 }
12 result=client.search(index=indexName, body=sortQuery)
13
14 pprint.pprint(result)
15
```

Рисунок 15 – Текст 8 запроса

Результат выполнения показан на рисунке 16.

```

warnings.warn(message, category=ElasticsearchWarning)
{'_shards': {'failed': 0, 'skipped': 0, 'successful': 1, 'total': 1},
 'hits': {'hits': [{'_id': 'Shared psychotic disorder',
                    '_index': 'medical',
                    '_score': None,
                    '_source': {'name': 'Shared psychotic disorder',
                                'title': 'Folie à deux'},
                    '_type': 'diseases',
                    'sort': ['à']},
                  {'_id': 'Zygomycosis',
                    '_index': 'medical',
                    '_score': None,
                    '_source': {'name': 'Zygomycosis', 'title': 'Zygomycosis'},
                    '_type': 'diseases',
                    'sort': ['zygomycosis']},
                  {'_id': 'Weissenbacher-Zweymüller syndrome',
                    '_index': 'medical',
                    '_score': None,
                    '_source': {'name': 'Weissenbacher-Zweymüller syndrome',
                                'title': 'Weissenbacher-Zweymüller syndrome'},
                    '_type': 'diseases',
                    'sort': ['zweymüller']},
                  {'_id': 'Zwisch-Kaye syndrome',
                    '_index': 'medical',
                    '_score': None,
                    '_source': {'name': 'Zwisch-Kaye syndrome',
                                'title': 'Zwisch-Kaye syndrome'},
                    '_type': 'diseases',
                    'sort': ['zwisch-kaye']}]}]}

```

Рисунок 16 – Результат выполнения 8 запроса

Результаты запроса отсортированы по текстовому полю "title" в обратном порядке.

Запрос 9. Изучение документа по идентификатору

Текст запроса 9 представлен на рисунке 17.

```

1 from elasticsearch import Elasticsearch
2 import pprint
3
4 client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
5 indexName = "medical"
6 docType = "diseases"
7
8 idQuery={
9     "query": {
10         "terms": {
11             "_id": ["Stomatitis ", "Alkaptonuria"]
12         }
13     }
14 }
15 result=client.search(index=indexName, body=idQuery)
16
17 pprint.pprint(result)
18

```

Рисунок 17 – Текст 9 запроса

Результат выполнения показан на рисунке 18.

```
{'_shards': {'failed': 0, 'skipped': 0, 'successful': 1, 'total': 1},
  'hits': {'hits': [{'_id': 'Alkaptonuria',
                      '_index': 'medical',
                      '_score': 1.0,
                      '_source': {'fulltext': 'Alkaptonuria is a rare inherited '
                                             'genetic disorder in which the '
                                             'body cannot process the amino '
                                             'acids phenylalanine and tyrosine, '
                                             'which occur in protein. It is '
                                             'caused by a mutation in the HGD '
                                             'gene for the enzyme homogentisate '
                                             '1,2-dioxygenase (EC 1.13.11.5); '
                                             'if a person inherits abnormal '
                                             'copies from each parent (it is a '
                                             'recessive condition), the body '
                                             'accumulates an intermediate '
                                             'substance called homogentisic '

```

Рисунок 18 – Результат выполнения 9 запроса

Фрагмент текста из выполненного запроса:

Алкаптонурия - это редкое наследственное генетическое заболевание, при котором организм не может перерабатывать аминокислоты фенилаланин и тирозин, содержащиеся в белке. Это вызвано мутацией в гене HGD фермента гомогентизат 1,2-диоксигеназы (EC 1.13.11.5); если человек наследует аномальные копии от каждого из родителей (это рецессивное состояние), в организме накапливается промежуточное вещество, называемое гомогентизиновой кислотой, в крови и тканях. Гомогентизиновая кислота и ее окисленная форма алкаптон выделяются с мочой, придавая ей необычно темный цвет. Накапливающаяся гомогентизиновая кислота вызывает повреждение хряща (охроноз, приводящее к остеоартриту) и сердечных клапанов, а также осаждение в виде камней в почках и других органах. Симптомы обычно развиваются у людей старше 30 лет, хотя темное изменение цвета мочи присутствует с рождения. Помимо лечения осложнений (таких как обезболивание и замена суставов при повреждении хряща), препарат нититинон оказывает полезное действие. подавляют выработку гомогентизиновой кислоты, и продолжаются исследования, может ли это улучшить симптомы. Алькаптонурия - заболевание редкое; он встречается у одного из 250 000 человек, но чаще встречается в Словакии и Доминиканской Республике. == Признаки и симптомы == У пациентов с болезнью черной кости в детском или молодом возрасте протекает бессимптомно, но у них моча может стать коричневой или даже чернильно-черной, если ее собрать и оставить на открытом воздухе. Пигментация может быть отмечена в хряще уха и других хрящах, а также в склере и лимбе роговицы глаза. После 30 лет у людей начинают появляться боли в несущих нагрузку суставах позвоночника, бедер и колен. Боль может быть настолько сильной, что мешает повседневной деятельности и может повлиять на трудоспособность. Операция по замене суставов (бедра и плеча) часто необходима в относительно молодом возрасте. В долгосрочной перспективе поражение позвоночных суставов приводит к уменьшению подвижности грудной клетки и может повлиять на дыхание...

Запрос 10

Текст запроса 10 представлен на рисунке 19.

```

1 from elasticsearch import Elasticsearch
2 import pprint
3
4 client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
5 indexName = "medical"
6 docType = "diseases"
7 searchQuery={
8     "query": {
9         "bool": {
10             "must": {
11                 "match" : {
12                     "title": "Chronic"
13                 }
14             },
15             "filter": {
16                 "match" : {
17                     "fulltext": "Americans"
18                 }
19             },
20             "must_not": {
21                 "match" : {
22                     "name": "Epidemic encephalomyelitis"
23                 }
24             },
25             "should": [
26                 { "match" : {
27                     "title": "leukemia"
28                 } },
29                 { "match" : {
30                     "title": "syndrome"
31                 } }
32             ],
33             "minimum_should_match": 1,
34             "boost" : 1.0
35         }
36     }
37 }
38 result=client.search(index=indexName, body=searchQuery)
39
40 pprint.pprint(result)

```

Рисунок 19 – Текст 10 запроса

Запрос соответствия (match query) – это стандартный запрос для выполнения полнотекстового поиска, включая параметры нечеткого сопоставления.

Результат выполнения показан на рисунке 20. Всего было найдено 4 описания.

```

{'_shards': {'failed': 0, 'skipped': 0, 'successful': 1, 'total': 1},
 'hits': {'hits': [{'_id': 'Chronic myelogenous leukemia',
                     '_index': 'medical',
                     '_score': 10.107349,
                     '_source': {'fulltext': 'Chronic myelogenous leukemia '
                                             '(CML), also known as chronic '
                                             'myeloid leukemia, is a cancer of '
                                             'the white blood cells. It is a '
                                             'form of leukemia characterized by '
                                             'the increased and unregulated '
                                             'growth of myeloid cells in the '
                                             'bone marrow and the accumulation '
                                             'of these cells in the blood. CML '
                                             'is a clonal bone marrow stem cell '
                                             'disorder in which a proliferation '
                                             'of mature granulocytes '
                                             '(neutrophils, eosinophils and '
                                             'basophils) and their precursors '
                                             'is found. It is a type of '

```

Рисунок 20 – Результат выполнения 10 запроса

Фрагмент текста 1 описания:

Хронический миелолейкоз (ХМЛ), также известный как хронический миелоидный лейкоз, представляет собой рак лейкоцитов. Это форма лейкемии, характеризующаяся усиленным и нерегулируемым ростом миелоидных клеток в костном мозге и накоплением этих клеток в крови. ХМЛ - это заболевание клональных стволовых клеток костного мозга, при котором обнаруживается пролиферация зрелых гранулоцитов (нейтрофилов, эозинофилов и базофилов) и их предшественников. Это тип миелопролиферативного новообразования, связанного с характерной хромосомной транслокацией, называемой филадельфийской хромосомой. СМЛ в основном лечат с помощью таргетных препаратов, называемых ингибиторами тирозинкиназы (ТКИ), которые с 2001 года привели к значительному увеличению показателей долгосрочной выживаемости. Эти препараты революционизировали лечение этого заболевания и позволили большинству пациентов иметь хорошее качество жизни по сравнению с прежними химиотерапевтическими препаратами. В западных странах на ХМЛ приходится 15 \ u2013 25% всех лейкозов у взрослых и 14% лейкозов в целом (включая педиатрическое население, где ХМЛ менее распространен).

== Признаки и симптомы == Наличие ХМЛ зависит от стадии заболевания на момент постановки диагноза, поскольку известно, что в некоторых случаях он пропускает стадии. У большинства пациентов (~ 90%) диагноз ставится на хронической стадии, которая чаще всего протекает бессимптомно. В этих случаях он может быть диагностирован случайно по повышенному количеству лейкоцитов при обычном лабораторном тесте. Он также может проявляться симптомами, указывающими на гепатоспленомегалию и возникающую в результате боль в верхнем квадранте. Увеличенная селезенка может оказывать давление на желудок, вызывая потерю аппетита и, как следствие, потерю веса. Он также может проявляться легкой лихорадкой и ночным потоотделением из-за повышенного базального уровня метаболизма. Некоторые (10 \ u2013 109 / л), не реагирует на терапию Постоянная или усиливающаяся спленомегалия, не реагирует на терапию Постоянный тромбоцитоз (> 1000 \ u2013 109 / л), не реагирует на терапию Постоянная тромбоцитопения (20% бластов в крови или костном мозге Наличие экстрамедуллярного разрастания бластов == Лечение == Единственное лечебное лечение ХМЛ - это трансплантация костного мозга или трансплантация аллогенных стволовых клеток. Помимо этого, существует четыре основных принципа лечения. При ХМЛ: лечение ингибиторами тирозинкиназы, миелосупрессивная терапия или лейкоферез (для противодействия лейкоцитозу во время раннего лечения), спленэктомия и лечение интерфероном альфа-2b. Из-за высокого среднего возраста пациенток с ХМЛ ХМЛ относительно редко встречается у беременных женщин, несмотря на это, однако, хронический миелолейкоз можно лечить с относительной безопасностью в любое время во время беременности с помощью гормонов интерферона-альфа.

=== Хроническая фаза === В прошлом антиметаболиты (например, цитарабин, гидроксимочевина), алкилирующие агенты, интерферон альфа 2b и стероиды использовались для лечения ХМЛ в хронической фазе, но с 2000-х годов были заменены препаратами ингибиторов тирозинкиназы Bcr-Abl, которые специфически нацелены на BCR-ABL, конститутивно активированный слитый белок тирозинкиназы, вызванный транслокацией филадельфийской хромосомы. Несмотря на переход к замене цитотоксических противоопухолевых препаратов (стандартные противоопухолевые препараты) ингибиторами тирозинкиназы, иногда гидроксимочевина все еще используется для противодействия повышенному количеству лейкоцитов, возникающему при лечении ингибиторами тирозинкиназы, такими как иматиниб; в этих ситуациях он 27 может быть предпочтительным миелосупрессивным средством из-за его относительного отсутствия лейкемогенных эффектов и, следовательно, относительного отсутствия возможности возникновения вторичных гематологических злокачественных новообразований в результате лечения. IRIS, международное исследование, в котором сравнивали комбинацию интерферона / цитарабина и первого из этих новых препаратов иматиниба, с долгосрочным наблюдением, продемонстрировало явное превосходство целевого ингибирования тирозинкиназы над существующими методами лечения. == ==

Иматиниб ===== Первым из этого нового класса препаратов был мезилат иматиниба (продаваемый как Gleevec или Glivec), одобренный Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов США (FDA) в 2001 году. Было обнаружено, что иматиниб подавляет прогрессирование ХМЛ у большинства пациентов (65–75%) достаточно для восстановления нормальной популяции стволовых клеток костного мозга (цитогенетический ответ) со стабильными пропорциями созревающих лейкоцитов. Поскольку некоторые лейкозные клетки (по оценке с помощью ОТ-ПЦР) сохраняются почти у всех пациентов, лечение необходимо продолжать бесконечно. С момента появления иматиниба ХМЛ стал первым раком, при котором стандартное лечение может обеспечить пациенту нормальную продолжительность жизни. ===== Дазатиниб, nilотиниб, радотиниб и босутиниб ===== Чтобы преодолеть резистентность к иматинибу и повысить чувствительность к ингибиторам ТК, позже были разработаны четыре новых агента. Первый, дазатиниб, блокирует еще несколько онкогенных белков в дополнение к более сильному ингибированию белка BCR-ABL и был первоначально одобрен в 2007 г. FDA США для лечения ХМЛ у пациентов, которые были либо устойчивы к иматинибу, либо не переносили его. Второй новый ингибитор ТК, nilотиниб, также был одобрен FDA по тому же показанию. В 2010 г. nilотиниб и дазатиниб были также одобрены для терапии первой линии, что сделало три препарата этого класса доступными для лечения впервые выявленного ХМЛ. В 2012 году...

Вывод: при выполнении лабораторной работы была изучена модель представления данных и способы работы с документной NoSQL базой данных Elasticsearch. Были получены навыки инсталляции, индексации и поиска в Elasticsearch, реализованы этапы процесса Data Science, протестирована работа с большим объёмом данных.