Cryptocurrencies Information Retrieval System

Milestone 2

João Matos, up201703884
Rui Pinto, up201806441
Tiago Gomes, up201806658

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Tools used

Elasticsearch:

- More popular tool
- More resources available to help us

Kibana:

- Attractive UI
- Frontend for Elasticsearch

Documents

- 9575 documents
- Data about cryptocurrencies


```
{
    "id": "terra-luna",
    "block_time_in_minutes": "0",
    "hashing_algorithm": "",
    "categories": [
      "Cosmos Ecosystem",
      "Terra Ecosystem",
      "Decentralized Finance (DeFi)"
    ],
    "genesis_date": "",
    "developer_score": "0.0",
    "community_score": "42.143",
    "liquidity_score": "71.152",
    "description": "Terra is a decentralized financial payment network that..."  ,
    "homepage_link": [ "https://terra.money"  ],
    "blockchain_site": [ "https://finder.terra.money/"  ],
    "subreddit_url": "https://www.reddit.com/r/terraluna/"  ,
    "github": [ "https://github.com/terra-project/core"  ],
    "image_url": "https://assets.coingecko.com/coins/images/8284.png?15  67147",
    "all_time_high(usd)": "49.7",
    "all_time_high_date": "2021-10-04T16:35:04.475Z" ,
    "market_cap": "17353003992.0",
    "current_price": "43.24",
    "price_change_percentage_1y": 13885.02113,
    "price_change_percentage_30d": 17.64327,
    "price_change_percentage_7d": 1.0319,
    "news": [
      {
        "title": "XDEFI Wallet launches liquidity program as it integrates Terra"  ,
        "article": "XDEFI Wallet, a browser-based service for Decentralized Finance..."
        "url": "https://coinmarketcap.com/headlines/news/xdefi-erra/"
      }
    ]
}
```

Schema

Data indexing:

- Define the mapping for the data types
- Tell Elasticsearch the types that should be used for each of the attributes in our dataset.

```
"mappings": {
      "properties": {
          "id":  { "type": "text" },
          "categories":  { "type": "text" },
          "block_time_in_minutes": { "type": "integer" },
          "hashing_algorithm": { "type": "text", "analyzer": "my_analyzer",
              "fields": {
                  "keyword": {
                      "type": "keyword"
                  }
              }
          },
          "genesis_date": { "type": "date", "ignore_malformed": true },
          "developer_score": { "type": "float" },
          "community_score": { "type": "float" },
          "liquidity_score": { "type": "float" },
          "description":  { "type": "text", "analyzer": "my_analyzer" },
          "homepage_link": { "type": "keyword", "index": false },
          "blockchain_site": { "type": "keyword", "index": false },
          "subreddit_url": { "type": "keyword", "index": false },
          "github": { "type": "keyword", "index": false },
          "image_url": { "type": "keyword", "index": false },
          "all_time_high(usd)": { "type": "double" },
          "all_time_high_date": { "type": "date" },
          "market_cap": { "type": "double" },
          "current_price": { "type": "double" },
          "price_change_percentage_1y": { "type": "double" },
          "price_change_percentage_30d": { "type": "double" },
          "price_change_percentage_7d": { "type": "double" },
          "news": {
              "type": "nested",
              "properties": {
                  "title": { "type": "text", "analyzer": "my_analyzer" },
                  "article": { "type": "text", "analyzer": "my_analyzer" },
                  "url": { "type": "keyword", "index": false }
              }
          }
      }
  }
```

Importing the data

POST /cryptos/_bulk


```
{"index": {"_id":381}}
{"id":"apyswap","block_time_in_minutes" :"0","hashing_algorithm" :"","categories" :"","genesis_date" :"", … }
{"index": {"_id":393}}
{"id":"aragon","block_time_in_minutes" :"0","hashing_algorithm" :"","categories" :["Cosmos Ecosystem" ,"Software"], … }
{"index": {"_id":438}}
{"id":"armor","block_time_in_minutes" :"0","hashing_algorithm" :"","categories" :["Insurance","Decentralized Finance (DeFi)" ], …
}
{"index": {"_id":446}}
{"id":"arsenal-fan-token" ,"block_time_in_minutes" :"0","hashing_algorithm" :"","categories" :["Fan Token"], … }

.
.
.
```

Character Filters, Token Filters, Tokenizers

- Custom analyzer
- Custom character and token filters
- Improved quality of results

```
"settings" : {
        "analysis" : {
            "char_filter" : {
                "remove_comma_number" : {
                    "type" : "pattern_replace" ,
                    "pattern" : "(?<=\\d),(?=\\d)" ,
                    "replacement" : ""
                },
                "swap_dollar_symbol" : {
                    "type" : "pattern_replace" ,
                    "pattern" : "(\\$)(\\d+)" ,
                    "replacement" : "$2$1"
                },
                "add_usd_word" : {
                    "type" : "pattern_replace" ,
                    "pattern" : "(?:(\\d+)\\$)" ,
                    "replacement" : "$1 usd"
                },
                "remove_dollar_symbol" : {
                    "type" : "pattern_replace" ,
                    "pattern" : "\\$([A-Za-z]+)" ,
                    "replacement" : "$1"
                },
                "add_percentage_word" : {
                    "type" : "pattern_replace" ,
                    "pattern" : "(\\d+)\\%" ,
                    "replacement" : "$1 percent"
                },
                "add_times_word" : {
                    "type" : "pattern_replace" ,
                    "pattern" : "(\\d+)[xX]" ,
                    "replacement" : "$1 times"
                },
                "join_dot_separated_word" : {
                    "type" : "pattern_replace" ,
                    "pattern" : "(?<=[A-Za-z])\\.(?=[A-Za-z])" ,
                    "replacement" : ""
                },
                "remove_special_chars" : {
                    "type" : "pattern_replace" ,
                    "pattern" : "[©®™]" ,
                    "replacement" : ""
                }
```

```
"filter" : {
                "synonym" : {
                    "type" : "synonym" ,
                    "synonyms" : [
                        "dollar, usd" ,
                        "proof work => pow" ,
                        "proof stake => pos" ,
                        "football => soccer"
                    ]
                },
                "remove_urls" : {
                    "type" : "keep_types" ,
                    "types" : ["<URL>"] ,
                    "mode" : "exclude"
                },
                "no_stem" : {
                    "type" : "keyword_marker" ,
                    "keywords" : ["pow", "pos"]
                }
            },
            "analyzer" : {
                "my_analyzer" : {
                    "char_filter" : [
                        "html_strip" ,
                        "remove_comma_number" ,
                        "remove_special_chars" ,
                        "join_dot_separated_word" ,
                        "swap_dollar_symbol" ,
                        "add_usd_word" ,
                        "remove_dollar_symbol" ,
                        "add_percentage_word" ,
                        "add_times_word"
                    ],
                    "tokenizer" : "uax_url_email" ,
                    "filter" : [
                        "lowercase" ,
                        "remove_urls" ,
                        "stop" ,
                        "synonym" ,
                        "no_stem" ,
                        "stemmer" ]
                }
            }
        }
```

Queries

- Coins that can be mined, with a low block time, were created a few years ago, and have a price lower than 1$.
- Coins whose price went down and were completely abandoned by its developers.
- Coins with a price spike in the last month that have high liquidity.
- News related to China's strong restrictions regarding bitcoin mining.
- News related to blockchain-based games based on NFTs.
- Crypto fan tokens related to football clubs.

Example Retrieval

Retrieve coins with the following characteristics:

- Can be mined
- Have a short block time
- Were created a few years ago
- Have price lower than 1$

```
GET /cryptos/_search
{
 "_source": ["genesis_date", "current_price", "id", "hashing_algorithm",
"block_time_in_minutes", "description"],
 "size": 250,
 "query": {
   "function_score": {
     "query": {
       "bool": {
         "must": [
           {
            "range": {
              "genesis_date": {
                "gte": "now/y-5y"
              }
            }
           },
           {
            "range": {
                "current_price": {
                  "lt": 1
                }
            }
           },
           {
            "script": {
              "script": {
                "source": "doc['block_time_in_minutes'].value == 0"
              }
            }
           }
         ],
         "must_not": {
           "multi_match": {
             "query": "pos",
             "fields": [ "hashing_algorithm", "description" ]
           }
         }
       }
     },
     "functions": [
       {
         "filter": {
           "match": {
             "description": "pow"
           }
         },
         "weight": 10
       },
       {
         "filter": {
           "term": {
             "hashing_algorithm.keyword": ""
           }
         },
         "weight": 0.5
       }
     ]
   }
 }
}
```

Search Results


```
{
 "took": 2,
 "timed_out": false,
 "_shards": {
   "total": 1,
   "successful": 1,
   "skipped": 0,
   "failed": 0
 },
 "hits": {
   "total": {
     "value": 7,
     "relation": "eq"
   },
   "max_score": 30,
   "hits": [
     {
       "_index": "cryptos",
       "_type": "_doc",
       "_id": "3503",
       "_score": 30,
       "_source": {
         "block_time_in_minutes": "0",
         "description": "Garlicoin is a new, freshly
baked cryptocurrency, born from ...",
         "id": "garlicoin",
         "genesis_date": "2018-01-16",
         "current_price": "0.050413",
         "hashing_algorithm": "Scrypt-N"
       }
     },
```

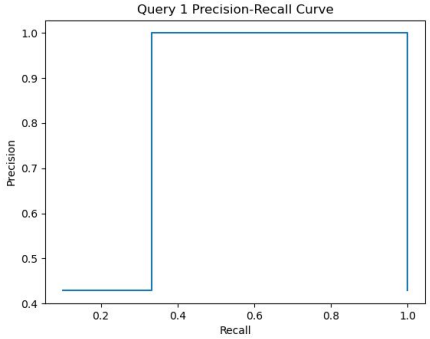
```
{
       "_index": "cryptos",
       "_type": "_doc",
       "_id": "5658",
       "_score": 30,
       "_source": {
         "block_time_in_minutes": "0",
         "description": "The bitcoin network is a peer-to-peer...",
         "id": "nework",
         "genesis_date": "2016-01-24",
         "current_price": "0.00626908",
         "hashing_algorithm": "Scrypt"
       }
     },
     {
       "_index": "cryptos",
       "_type": "_doc",
       "_id": "6678",
       "_score": 30,
       "_source": {
         "block_time_in_minutes": "0",
         "description": "Quadrant is a blockchain-based...",
         "id": "quadrant-protocol",
         "genesis_date": "2018-06-25",
         "current_price": "0.0069677",
         "hashing_algorithm": "Ethash"
       }
     },
```

Evaluation

- Based on tutorial code
- Added F1 calculation
- Added final precision and recall
- Metrics used:
  - Average precision
  - Precision at 3
  - Precision
  - Recall
  - F1 score

Query 1 Metrics

Metric	Value
Average Precision	0.808333
Precision at 3 (P@3)	1
Precision	0.428571
Recall	1
F1 Measure	0.6



Questions?