POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Systems and Methods for Big and Unstructured Data Project

Author(s): **Balestrieri Niccolò - 10936955**

Bertogalli Andrea - 10702303

Tombini Nicolò - 10912627

Academic Year: 2023-2024

Contents

1 | Introduction

In this report we will present a comprehensive analysis of the ArXiv dataset, a collaboratively funded, community-supported resource founded by Paul Ginsparg in 1991 and currently maintained and operated by Cornell University. Our project delves into the intricate web of academic knowledge encapsulated within ArXiv, utilizing advanced technologies to extract meaningful insights.

Our group strategically leveraged the power of two cutting-edge technologies to unravel the complexities inherent in the ArXiv dataset. The first technology used is Neo4j, a graph database management system. This choice was driven by the dataset's inherent graph-like structure, where nodes represent authors and articles, and relationships encapsulate the act of writing or not papers. Neo4j in handling complex relationships and its native support for graph structures made it the ideal choice for modeling and querying the interconnected web of academic contributions within the ArXiv repository.

Complementing our graph-oriented approach, we embraced Elasticsearch as our second technological choice. Elasticsearch, renowned for its robust full-text search capabilities, are well integrated with our project objectives. This technology addresses the need for powerful and efficient textual searches on the vast array of academic papers within the ArXiv dataset. Its flexibility and scalability make it a valuable asset in uncovering relevant information through textual queries, enhancing the overall usability and accessibility of the dataset.

In the following sections of this report, we will delve into our web application built on top of these two technologies, whose purpose is to show 20 queries (10 for each technology) in a user-friendly way.

In this chapter additional useful information are reported.

2 | Data wrangling / Data generation

The chosen dataset, in JSON format, presents several issues to address. Firstly, for computational and resource reasons, it was decided to limit the dataset to 30,000 randomly selected data points from the entire dataset.
Subsequently, cleaning was performed on the authors' names and dates. Moreover, upon examining a random abstract text, it was observed that the text is available in Latex format. Therefore, both the title and the abstract text were transformed to UTF-8 format using the pylatexenc library.

Before:

```
This result confirms that dominant contributions to the electric
and magnetic polarizabilities may be represented in terms of two-photon
couplings to the $\sigma$-meson having the predicted mass $m_\sigma=666$ MeV
and two-photon width $\Gamma_{\gamma\gamma}=2.6$ keV.
```

After:

This result confirms that dominant contributions to the electric and magnetic polarizabilities may be represented in terms of two-photon couplings to the σ-meson having the predicted mass $m_\sigma = 666$ MeV and two-photon width $\Gamma_{\gamma\gamma} = 2.6$ keV.

Later on, we created the clean and well-formatted dataset in .csv format, which was then imported and utilized on Neo4j and ElasticSearch.

3 | Dataset

In general the dataset is a file JSON following this structure:

```
{
    "id":"0704.0001",
    "submitter":"Pavel Nadolsky",
    "authors":"E. L. Berger, P. M. Nadolsky, C.-P. Yuan",
    "title":"Paper title",
    "comments":"37 pages, 15 figures; published version",
    "journal-ref":"Phys.Rev.D76:013009,2007",
    "doi":"10.1103/PhysRevD.76.013009",
    "report-no":"ANL-HEP-PR-07-12",
    "categories":"hep-ph",
    "license":"",
    "abstract":"Paper abstract",
    "versions":[
        0:{
            "version":"v1"
            "created":"Mon, 2 Apr 2007 19:18:42 GMT"
        }
    ],
    "update_date":"2008-11-26"
}
```

After cleaning and parsing dataset from JSON into CSV format, we chose to use the same dataset for both technologies with slight modifications in terms of attributes. The details of these modifications will be explained in the following sections.

3.1. Neo4j

Regarding Neo4j, the dataset used considers authors and articles.

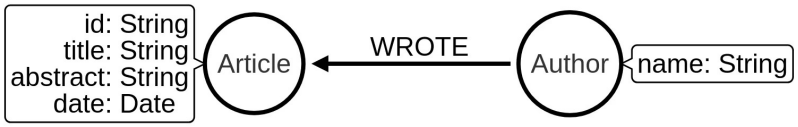

Figure 3.1: Neo4j DB structure.

As seen in the picture above, we have 2 types of nodes: **Article** and **Author**.

For **Article**, we have:

- **id**: article ID, String
- **title**: article title, String
- **abstract**: summary of the article content, String
- **date**: article publication date, Date

For **Author**, we have:

- **name**: author's name, String

Clearly, the relation **WROTE** is present if the author wrote the relative article.

3.2. ElasticSearch

Instead, for what concerns ElasticSearch, we started again from CSV dataset but we had to create a manual mapping, and for this purpose it was necessary to recover a sort of JSON file.

```
mapping = {
    "settings":{
        "number_of_shards": 3,
        "number_of_replicas": 3
    },
    "mappings": {
        "properties": {
            "submitter": {"type": "keyword"},
            "authors": {"type": "text"},
            "title": {"type": "text"},
            "abstract": {"type": "text"},
            "comments": {"type": "text"},
            "ref": {"type": "keyword"},
            "categories": {"type": "keyword"},
        }
    }
}
```

Here we can spot these attributes:

- **submitter**: the author who submitted the journal, keyword
- **authors**: list of paper authors, text
- **title**: article title, text
- **abstract**: summary of the article content, text
- **comments**: comments related to an article, text
- **ref**: journal acronym, keyword
- **categories**: kind of categories related to an articole, keyword

4 | Queries

4.1. Neo4j

Query 1. *Retrieves the articles from 2007 which have been written by at least 25 authors.*

```
1 MATCH (ar:Article)<-[:WROTE]-(au:Author)
2 WITH count(au) as number_of_authors, ar
3 WHERE date(ar.date).year = 2007 AND number_of_authors >= 25
4 MATCH p=(ar)<-[:WROTE]-(au:Author)
5 RETURN p;
```

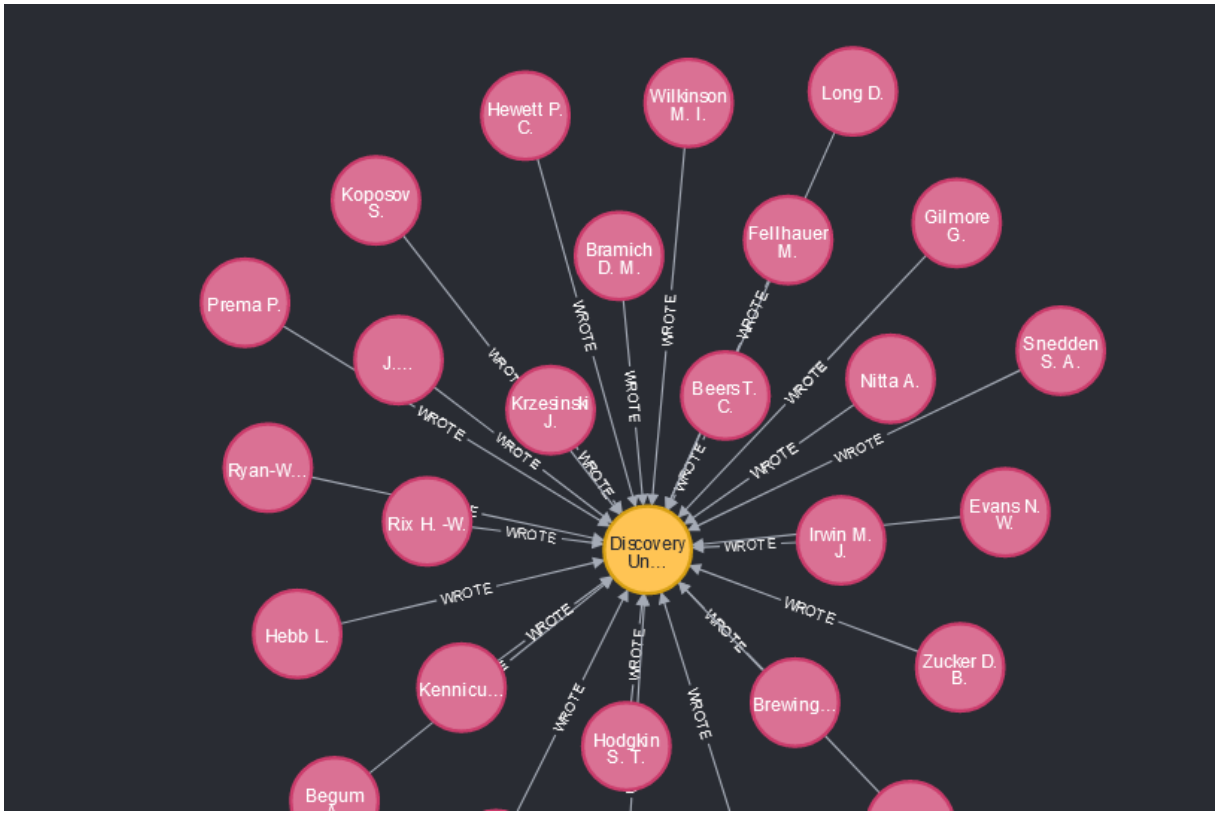

Figure 4.1: Partial outcome query 1

Query 2. *Retrieves the article with the highest number of authors.*

```
MATCH (ar:Article)<-[:WROTE]-(au:Author)
WITH ar, count(au) AS number_of_authors
WITH ar, number_of_authors
ORDER BY number_of_authors DESC
LIMIT 1
MATCH p=(ar)<-[:WROTE]-(Author)
RETURN p;
```

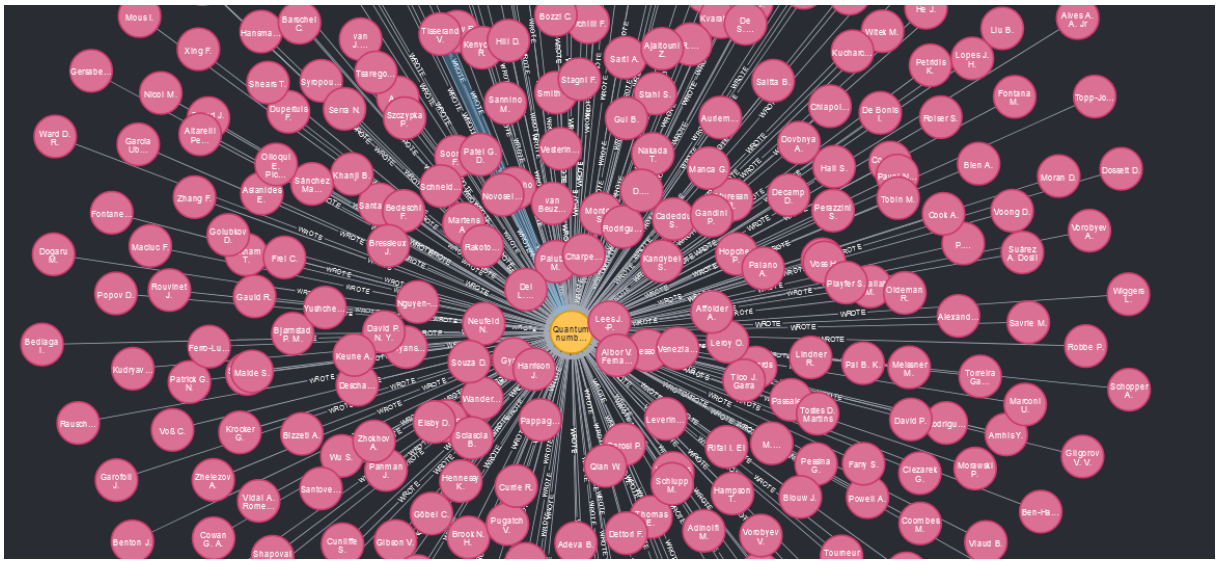

Figure 4.2: Partial outcome query 2

Query 3. *Returns all the papers that have Matteucci Matteo among the authors.*

```
MATCH (ar:Article)<-[:WROTE]-(au:Author)
WHERE au.name CONTAINS "Matteucci Matteo"
MATCH p=(ar)<-[:WROTE]-(Author)
RETURN p;
```

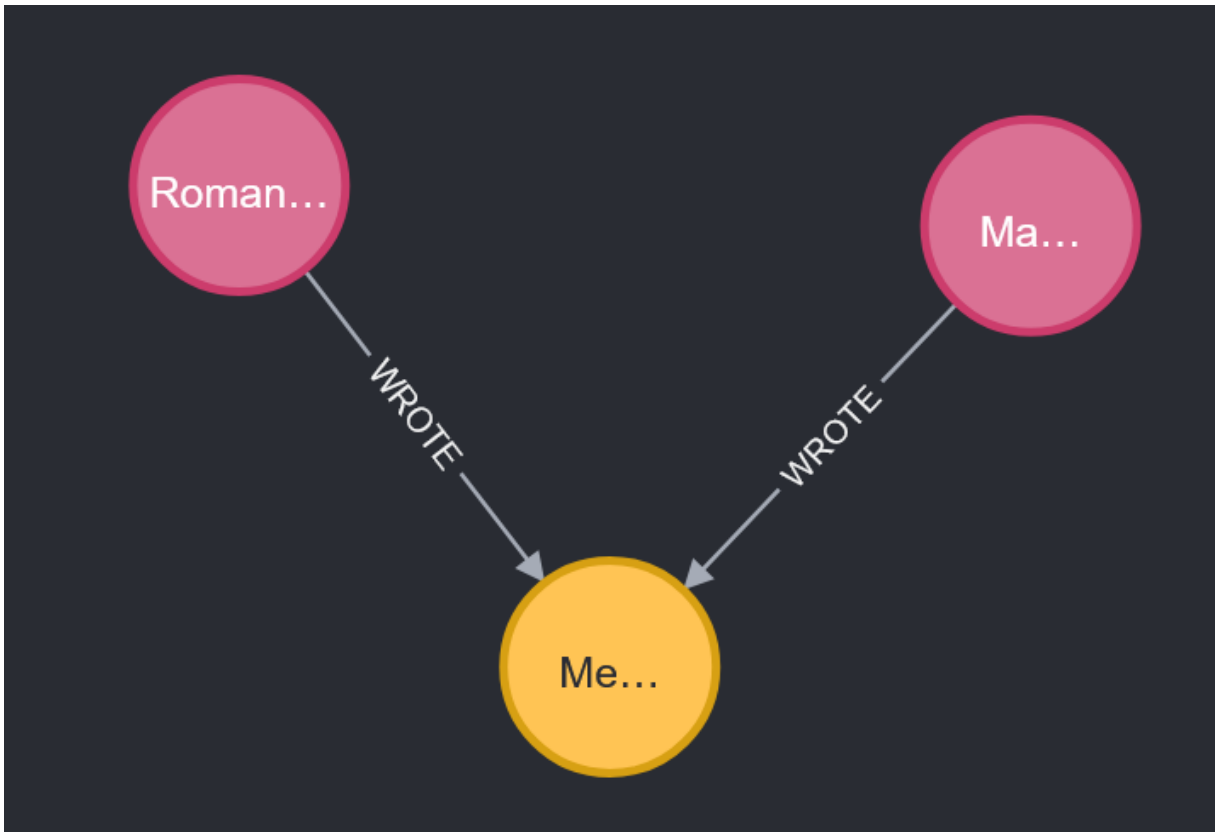

Figure 4.3: Query 3

Query 4. *Returns all the collaborators of Yann LeCun and their articles*

```
MATCH (ar:Article)<-[:WROTE]-(y:Author {name: "LeCun Yann"}),
(au:Author)-[:WROTE]->(ar)
MATCH p=(:Article)<-[:WROTE]-(au)
RETURN p;
```

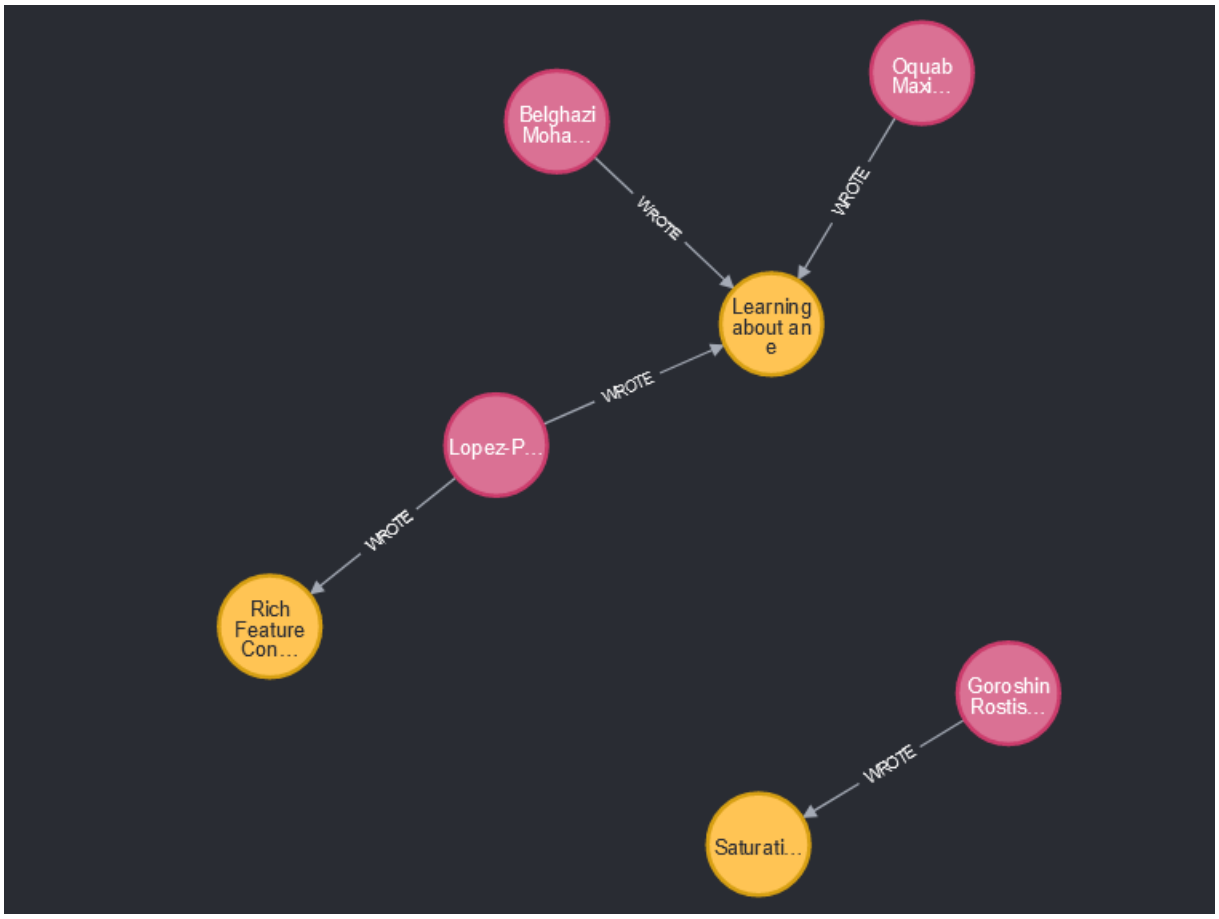

Figure 4.4: Query 4

Query 5. *Returns all the authors who wrote an article about Quantum physics in feb 2019.*

```
MATCH (a:Author)-[:WROTE]->(ar:Article)
WHERE date(ar.date).month = 2 AND date(ar.date).year = 2019
   AND ar.abstract CONTAINS 'Quantum'
MATCH p=(ar)<-[:WROTE]-(a)
RETURN p;
```

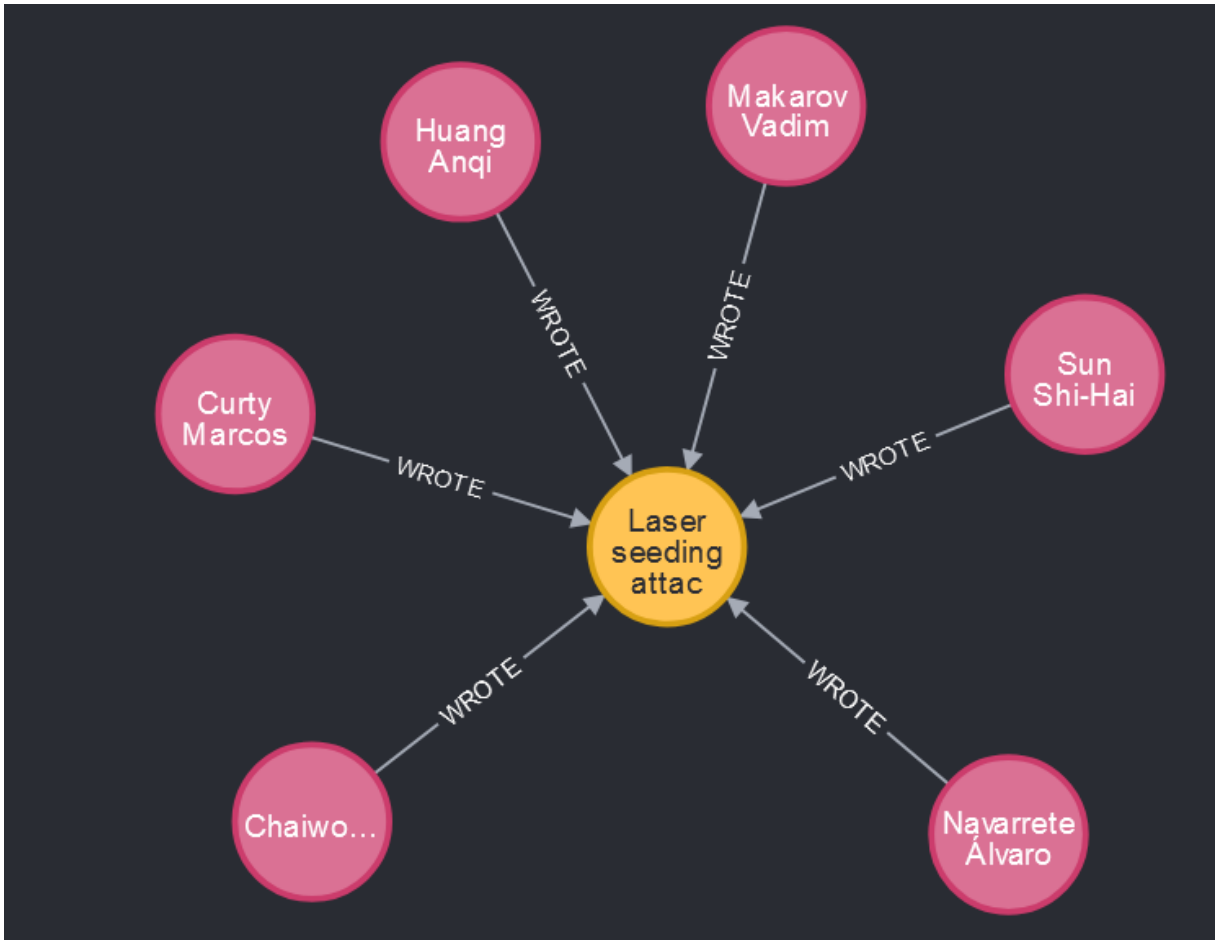

Figure 4.5: Query 5

Query 6. *Gets all articles by authors who have written articles in at least three different years.*

```
MATCH (ar:Article)<-[:WROTE]-(au:Author)-[:WROTE]->(other_ar:
   Article)
WITH au, count(DISTINCT date(other_ar.date).year) AS
   unique_years
WHERE unique_years >= 3
MATCH p=(ar)<-[:WROTE]-(au)
RETURN p
limit 250;
```

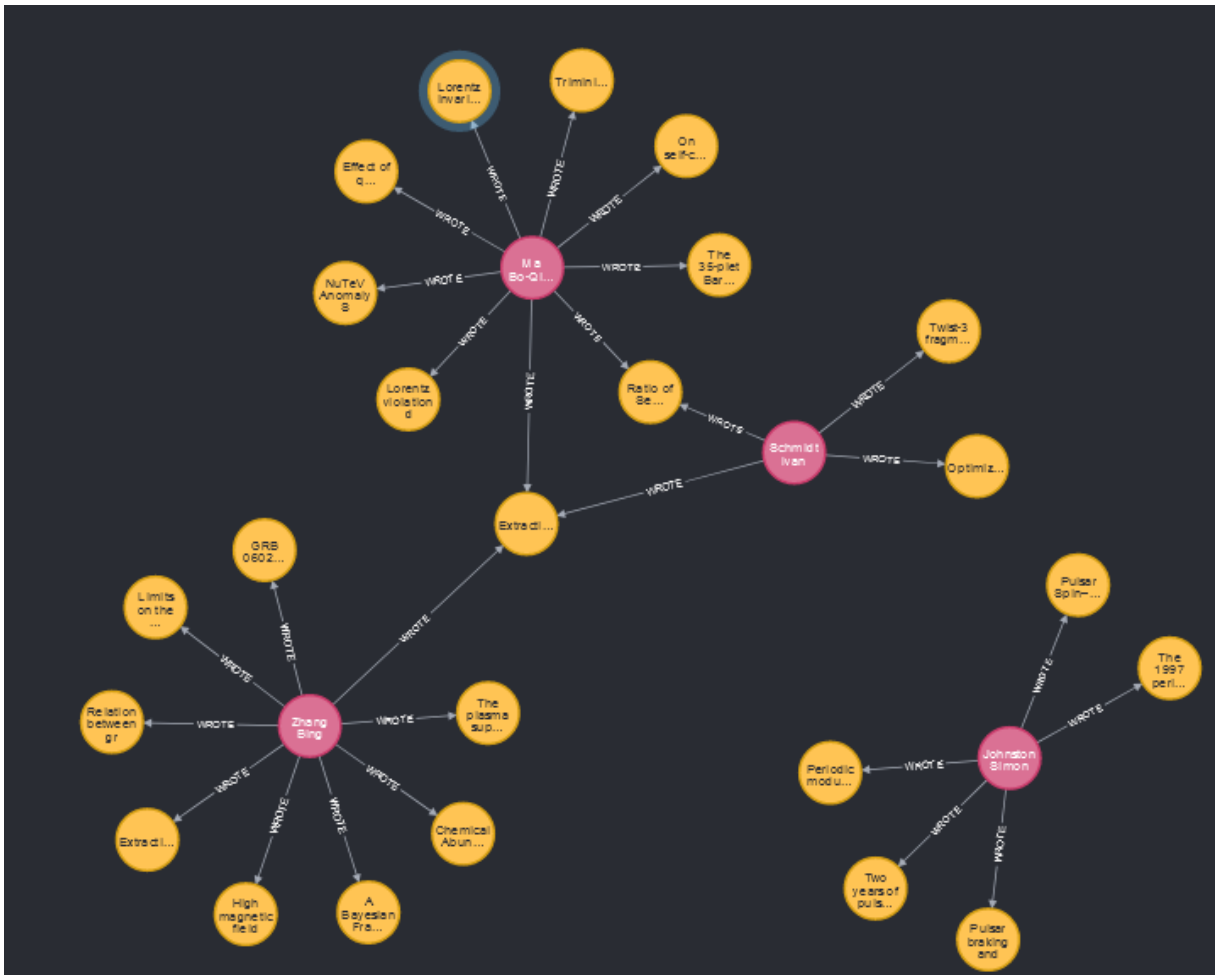

Figure 4.6: Partial outcome query 6

Query 7. *Returns articles written by authors who have also published articles with 'Machine Learning' in the abstract.*

```
MATCH (ar:Article)<-[:WROTE]-(au:Author)-[:WROTE]->(ml:
   Article)
WHERE ml.abstract CONTAINS 'Machine Learning'
MATCH p=(ar)<-[:WROTE]-(au)
RETURN p;
```

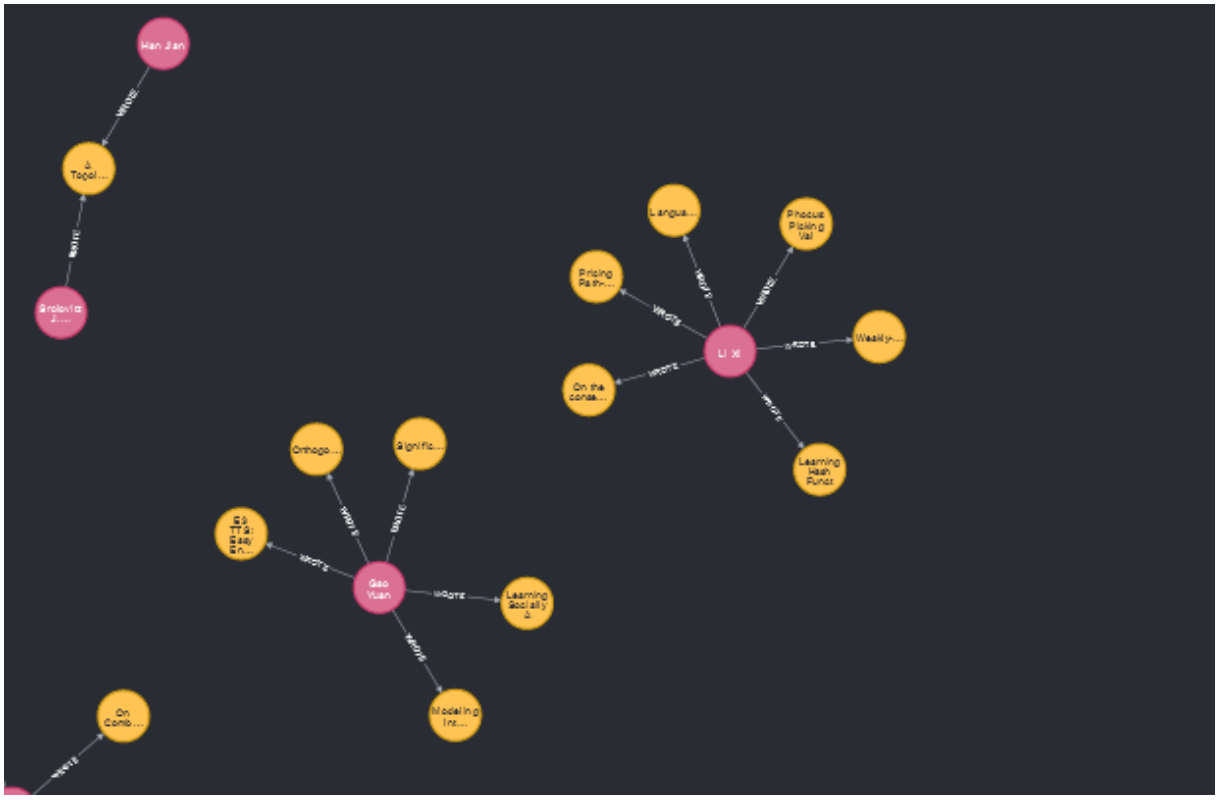

Figure 4.7: Partial outcome query 7

Query 8. *Retrieves articles with titles containing the word 'Graph' and lists their authors.*

```
MATCH (ar:Article)<-[:WROTE]-(au:Author)
WHERE ar.title CONTAINS 'Graph'
MATCH p=(ar)<-[:WROTE]-(au)
RETURN p
LIMIT 250;
```

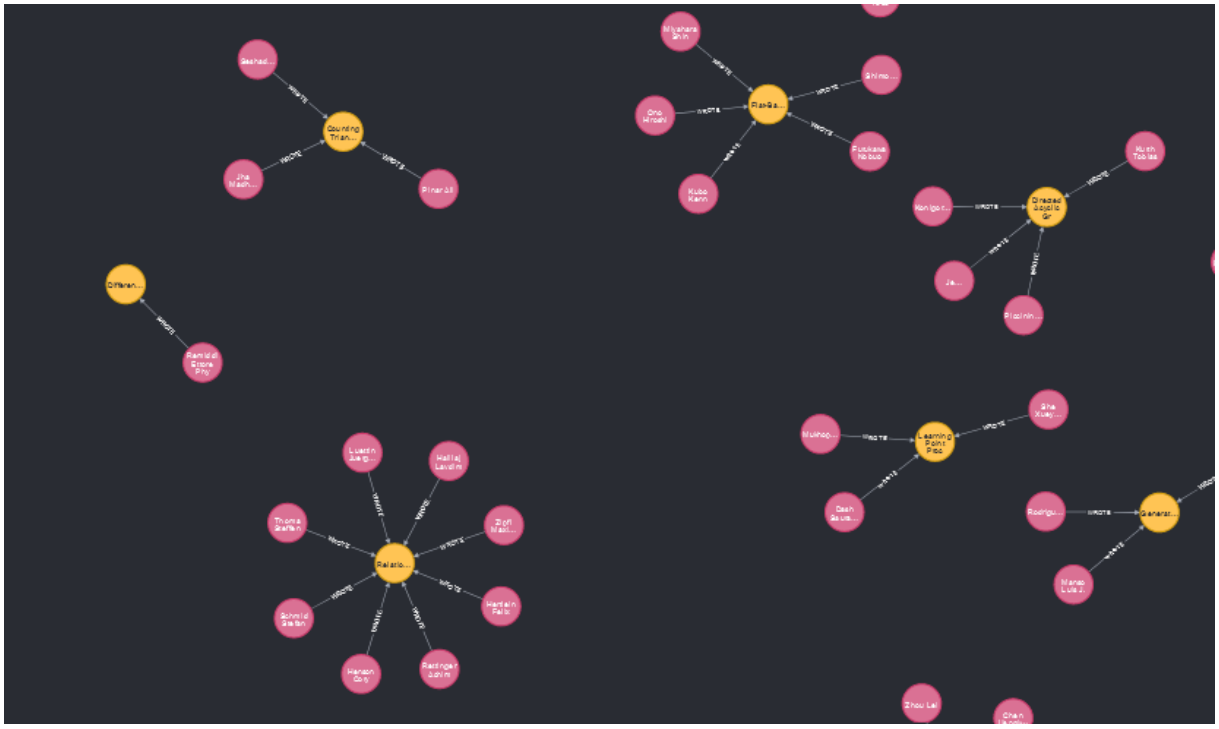

Figure 4.8: Partial outcome query 8

Query 9. *Retrivies articles written by authors who have not contributed to any articles in the year 2023 and returns the articles along with their author relationships.*

```
MATCH (ar:Article)<-[:WROTE]-(au:Author)
WITH ar, au
WHERE NOT (au)-[:WROTE]->(:Article {date: '2023'})
MATCH p=(ar)<-[:WROTE]-(au:Author)
RETURN p
LIMIT 50;
```

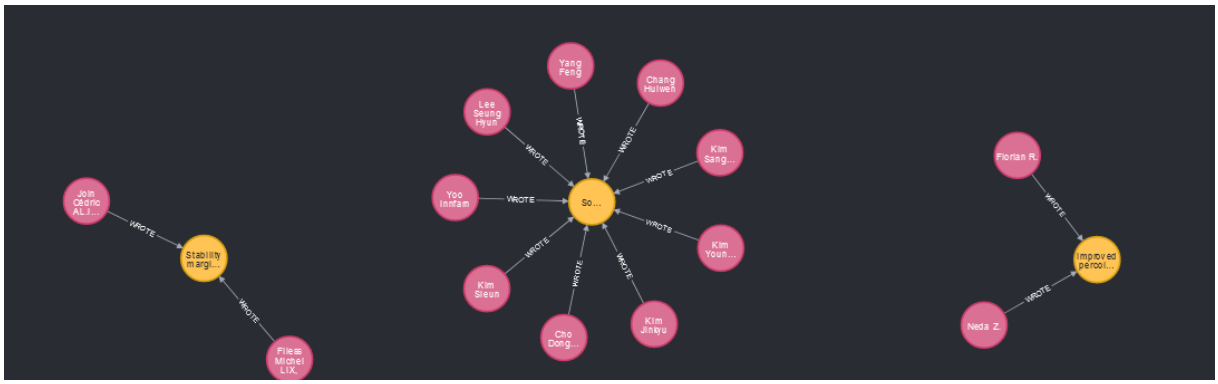

Figure 4.9: Query 9

Query 10. *Retrieves articles with the most recent publication date and their authors.*

```
MATCH (ar:Article)<-[:WROTE]-(au:Author)
WITH ar, date(ar.date) AS publication_date
ORDER BY publication_date DESC
LIMIT 1
MATCH p=(ar)<-[:WROTE]-(au)
RETURN p;
```

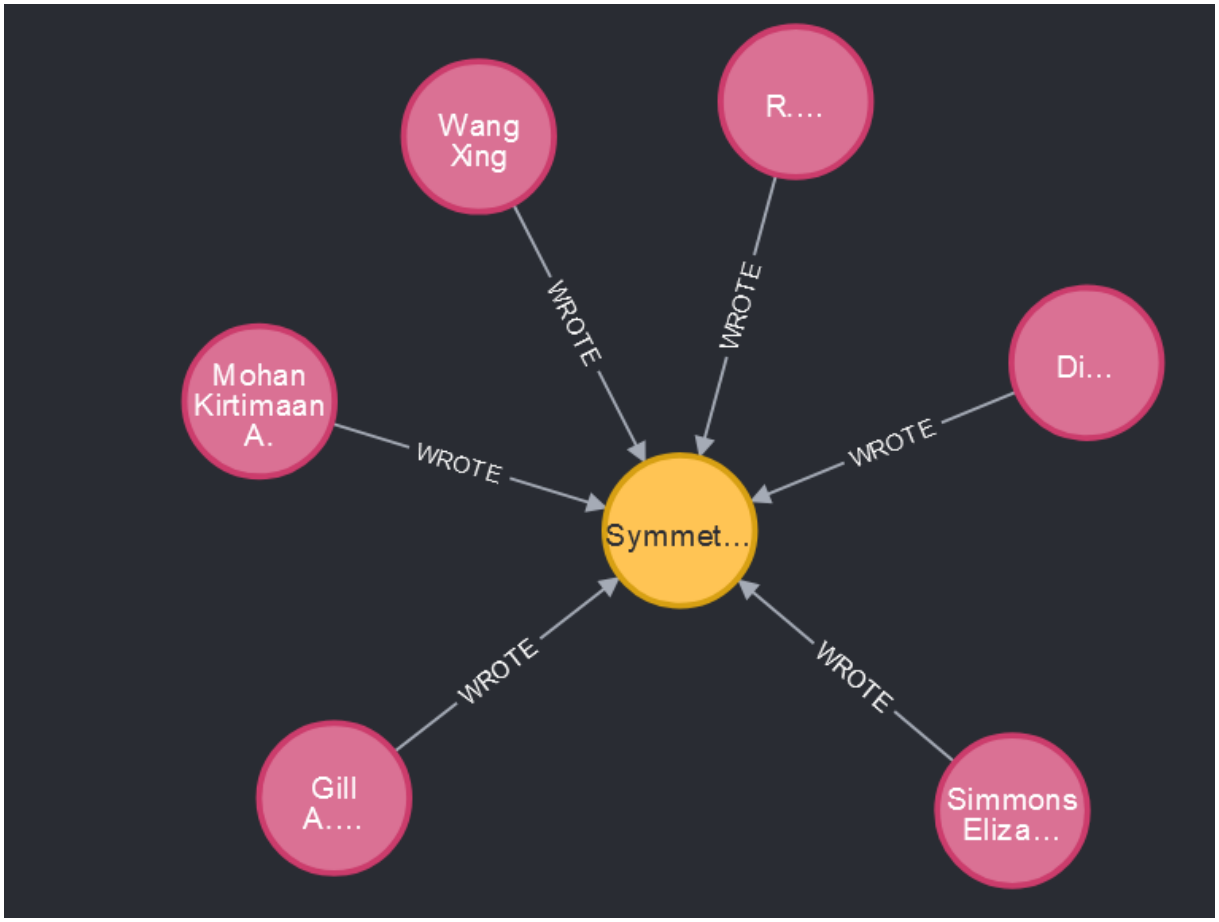

Figure 4.10: Query 10

4.2. ElasticSearch

Query 11. *Retrieves all the articles where the title contains the word "AI" but do not contains the word "ethic", prioritizing articles where the abstract contains the word "Neural".*

```
GET /arxiv/_search
{
    "query": {
        "bool": {
            "must": [
                { "match": { "title": "AI" } }
            ],
            "must_not": [
                { "match": { "title": "Ethic" } }
            ],
            "should": [
                { "match": { "abstract": "Neural" } }
            ]
        }
    }
}
```

With this partial result:

```
"hits": [
      {
        "_index": "arxiv",
        "_id": "13383",
        "_score": 11.909199,
        "_source": {
          "submitter": "Maxime Gariel",
          "authors": "Maxime Gariel, Brian Shimanuki, Rob
             Timpe, Evan Wilson",
          "title": "Framework for Certification of AI-Based
             Systems",
          "comments": "9 pages, 10 figures",
          "ref": "",
          "categories": "cs.LG cs.CV cs.SE",
```

```
13          "abstract": """The current certification process
               for aerospace software is not adapted to "AI-
               based" algorithms such as deep neural networks.
               Unlike traditional aerospace software, the
               precise parameters optimized during neural
               network training are as important as (or more
               than) the code processing the network and they
               are not directly mathematically understandable.
               Despite their lack of explainability such
               algorithms are appealing because for some
               applications they can exhibit high performance
               unattainable with any traditional explicit line-
               by-line software methods.  This paper proposes
               a framework and principles that could be used to
                establish certification methods for neural
               network models for which the current
               certification processes such as DO-178 cannot be
                applied. While it is not a magic recipe, it is
               a set of common sense steps that will allow the
               applicant and the regulator increase their
               confidence in the developed software, by
               demonstrating the capabilities to bring together
               , trace, and track the requirements, data,
               software, training process, and test results."""
14        }
15      },
16      ...
17    ]
```

Query 12. *Counts document by submitter.*

```
GET /arxiv/_search
{
    {
      "size": 0,
      "aggs": {
        "doc_by_submitter": {
          "terms": {
            "field": "submitter"
          }
        }
      }
    }
}
```

With this partial result:

```
 "aggregations": {
    "doc_by_submitter": {
      "doc_count_error_upper_bound": 9,
      "sum_other_doc_count": 29713,
      "buckets": [
        {
          "key": "",
          "doc_count": 186
        },
        {
          "key": "EPTCS",
          "doc_count": 46
        },
        ...
      ]
    }
}
```

Query 13. *Retrieves all the articles with 18 pages and 8 figures.*

```
GET /arxiv/_search
{
 "query": {
        "match": {
            "comments":{
                "query": "18 pages 8 Figures",
                "operator": "and"
            }
        }
    }
}
```

With this partial result:

```
"hits": [
      {
        "_index": "arxiv",
        "_id": "1289",
        "_score": 9.578911,
        "_source": {
          "submitter": "Evgeny Grines",
          "authors": "Aleksei M. Arefev, Evgeny A. Grines,
             Grigory V. Osipov",
          "title": "Heteroclinic cycles and chaos in a system
              of four identical phase   oscillators with
             global biharmonic coupling",
          "comments": "18 pages, 8 figures",
          "ref": "",
          "categories": "nlin.CD math.DS",
          "abstract": "abstract text"
        }
      },
      {
        "_index": "arxiv",
        "_id": "6045",
        "_score": 9.578911,
        "_source": {
```

```
            "submitter": "Fei Wu",
            "authors": "Fei Wu, Chen Wu, Zhongzhou Ren",
            "title": "Neutron stars including the effects of
               chaotic magnetic fields and the   anomalous
               magnetic moments",
            "comments": "18 pages, 8 figures",
            "ref": "Chinese Physics C 41 (2017) 045102",
            "categories": "astro-ph.HE nucl-th",
            "abstract": "abstract text"
          }
        },
        ...
    ]
```

Query 14. *Counts all "hep-ph" articles for each submitter.*

```
GET /arxiv/_search
{
"size":0,
   "query":{
      "match":{
         "categories":"hep-ph"
      }
   },
   "aggs":{
      "hep_articles_per_submitter":{
         "terms":{
            "field":"submitter"
         }
      }
   }
}
```

With this partial result:

```
"aggregations":{
      "hep_articles_per_submitter":{
         "doc_count_error_upper_bound":3,
         "sum_other_doc_count":948,
         "buckets":[
            {
               "key":"",
               "doc_count":43
            },
            {
               "key":"Adrian Signer",
               "doc_count":4
            },
            ...
        ]
    }
}
```

Query 15. *Counts for each categories how many articles with the word "Quantum" in the title.*

```
GET /arxiv/_search
{
   "size":0,
    "query":{
        "match":{
            "title": "quantum"
        }
    },
    "aggs":{
        "docs_per_category":{
            "terms":{
                "field": "categories"
            }
        }
    }
}
```

With this partial result:

```
"aggregations":{
      "docs_per_category":{
         "doc_count_error_upper_bound":9,
         "sum_other_doc_count":788,
         "buckets":[
            {
               "key":"quant-ph",
               "doc_count":388
            },
            {
               "key":"cond-mat.mes-hall",
               "doc_count":112
            },
            {
               "key":"hep-th",
               "doc_count":46
            },
```

```
18                  ...
19              ]
20      }
21  }
```

Query 16. *Finds all articles with with "Neural Network" in the title, favouring artlicles containing "deep learning" in the abstract and favouring articles containing "cnn" in the abstract with double importance.*

```
GET /arxiv/_search
{
"query":{
      "bool":{
         "must":[
            {
               "match":{
                  "title":"neural network"
               }
            }
         ],
         "should":[
            {
               "match":{
                  "abstract":"deep learning"
               }
            },
            {
               "match":{
                  "abstract":{
                     "query":"CNN",
                     "boost":2.0
                  }
               }
            }
         ]
      }
   }
}
```

With this partial result:

```
"hits":[
         {
            "_index":"arxiv",
```

```
            "_id":"2833",
            "_score":31.632092,
            "_source":{
               "submitter":"Xingquan Liu",
               "authors":"X. Zhang, Y. Huang, W. Lin, X. Liu,
                   H. Zheng, R. Wada, A. Bonasera, Z.\n  Chen
                  , L. Chen, J. Han, R. Han, M. Huang, Q. Hu,
                   Q. Leng, C. W. Ma, G. Qu, P.\n  Ren, G.
                  Tian, Z. Xu, Z. Yang, and L. Zhang",
               "title":"Determining impact parameters of
                  heavy-ion collisions at  low-intermediate
                  incident energies using deep learning with
                  convolutional  neural network",
               "comments":"21 pages, 16 figures, 1 table",
               "ref":"",
               "categories":"nucl-th",
               "abstract":"abstract text."
            }
         },
         {
            "_index":"arxiv",
            "_id":"15136",
            "_score":30.780548,
            "_source":{
               "submitter":"Tingle Li",
               "authors":"Tingle Li, Qingjian Lin, Yuanyuan
                  Bao, Ming Li",
               "title":"Atss-Net: Target Speaker Separation
                  via Attention-based Neural Network",
               "comments":"Submitted to Interspeech 2020",
               "ref":"",
               "categories":"eess.AS cs.SD",
               "abstract":"abstract text."
            }
         },
         ...
   ]
```

Query 17. *Counts articles about AI but not about ethic by category.*

```
GET /arxiv/_search
{
"size":0,
   "query":{
      "bool":{
         "must":[
            {
               "match":{
                  "title":"AI"
               }
            }
         ],
         "must_not":[
            {
               "match":{
                  "title":"ethic"
               }
            },
            {
               "match":{
                  "abstract":"ethic"
               }
            }
         ]
      }
   },
   "aggs":{
      "docs_per_category":{
         "terms":{
            "field":"categories"
         }
      }
   }
}
```

With this partial result:

```
"aggregations":{
      "docs_per_category":{
         "doc_count_error_upper_bound":0,
         "sum_other_doc_count":38,
         "buckets":[
            {
               "key":"cs.AI",
               "doc_count":6
            },
            {
               "key":"cs.CL",
               "doc_count":3
            },
            {
               "key":"cs.CV",
               "doc_count":3
            },
            ...
         ]
    }
}
```

Query 18. *Counts for each category how many articles have been submitted by each submitter.*

```
GET /arxiv/_search
{
"size":0,
   "aggs":{
      "by_categories":{
         "terms":{
            "field":"categories"
         },
         "aggs":{
            "by_submitter":{
               "terms":{
                  "field":"submitter"
               }
            }
         }
      }
   }
}
```

With this partial result:

```
"aggregations":{
      "by_categories":{
         "doc_count_error_upper_bound":204,
         "sum_other_doc_count":23623,
         "buckets":[
            {
               "key":"astro-ph",
               "doc_count":1064,
               "by_submitter":{
                  "doc_count_error_upper_bound":3,
                  "sum_other_doc_count":1034,
                  "buckets":[
                     {
                        "key":"",
                        "doc_count":11
```

```
                        },
                        {
                            "key":"Charles Dermer",
                            "doc_count":3
                        },
                        {
                            "key":"Alain Jorissen",
                            "doc_count":2
                        },
                        ...
                    ]
                }
            }
        ]
    }
}
```

Query 19. *Counts for each submitter, for each categories the articles about AI favouring articles about transformers.*

```
GET /arxiv/_search
{
"size":0,
   "query":{
      "bool":{
         "must":[
            {
               "match":{
                  "title":"AI"
               }
            },
            {
               "match":{
                  "abstract":"AI"
               }
            }
         ],
         "should":[
            {
               "match":{
                  "abstract":"Transformers"
               }
            }
         ]
      }
   },
   "aggs":{
      "by_submitter":{
         "terms":{
            "field":"submitter"
         },
         "aggs":{
            "by_category":{
               "terms":{
                  "field":"submitter"
```

```
                }
            }
        }
    }
}
```

With this partial result:

```
"aggregations":{
   "by_submitter":{
      "doc_count_error_upper_bound":0,
      "sum_other_doc_count":43,
      "buckets":[
         "0":{
            "key":"Abeba Birhane",
            "doc_count":1,
            "by_category":{
               "doc_count_error_upper_bound":0,
               "sum_other_doc_count":0,
               "buckets":[
                  "0":{
                     "key":"Abeba Birhane",
                     "doc_count":1
                  }
               ]
            }
         },
         "1":{
            "key":"Admela Jukan",
            "doc_count":1,
            "by_category":{
               "doc_count_error_upper_bound":0,
               "sum_other_doc_count":0,
               "buckets":[
                  "0":{
                     "key":"Admela Jukan",
                     "doc_count":1
```

```
                    }
                ]
            }
        },
        "2":{
            "key":"Aleksander Slominski",
            "doc_count":1,
            "by_category":{
                "doc_count_error_upper_bound":0,
                "sum_other_doc_count":0,
                "buckets":[
                    "0":{
                        "key":"Aleksander Slominski",
                        "doc_count":1
                    }
                ]
            }
        },
        "3":{
            "key":"Alexander Preuhs",
            "doc_count":1,
            "by_category":{
                "doc_count_error_upper_bound":0,
                "sum_other_doc_count":0,
                "buckets":[
                    "0":{
                        "key":"Alexander Preuhs",
                        "doc_count":1
                    }
                ]
            }
        },
        "4":{
            "key":"Ameet Deshpande",
            "doc_count":1,
            "by_category":{
                "doc_count_error_upper_bound":0,
```

```
                "sum_other_doc_count":0,
                "buckets":[
                   "0":{
                      "key":"Ameet Deshpande",
                      "doc_count":1
                   }
                ]
             }
          },
          ...
    ]
  }
}
```

Query 20. *Retrieves Articles with Attention in Title, Excluding CNN in Abstract, and Emphasizing Bidirectional LSTM and Transformer.*

```
GET /arxiv/_search
{
   "query":{
      "bool":{
          "must_not":[
            {
               "match":{
                  "abstract":"CNN"
               }
            }
         ],
         "must":[
            {
               "match":{
                  "title":"Attention"
               }
            }
         ],
         "should":[
            {
               "match":{
                  "abstract": {
                      "query": "Bidirectional LSTM",
                      "operator": "and",
                      "boost": 4.0
                  }
               }
            },
            {
               "match":{
                  "abstract":{
                     "query": "Transformer",
                     "boost": 2.0
                  }
               }
```

```
            }
        ]
    }
  }
}
```

With this partial result:

```
"hits":[
        {
            "_index":"arxiv",
            "_id":"7600",
            "_score":23.900904,
            "_source":{
               "submitter":"Jaeyoung Kim",
               "authors":"Jaeyoung Kim, Mostafa El-Khamy,
                  Jungwon Lee",
               "title":"T-GSA: Transformer with Gaussian-
                  weighted self-attention for speech
                  enhancement",
               "comments":"5 pages, Submitted to ICASSP 2020"
                  ,
               "ref":"",
               "categories":"eess.AS cs.SD",
               "abstract":"abstract text"
            }
        },
        {
            "_index":"arxiv",
            "_id":"1253",
            "_score":21.699568,
            "_source":{
               "submitter":"Xiangyu Chen",
               "authors":"Xiangyu Chen, Xintao Wang, Wenlong
                  Zhang, Xiangtao Kong, Yu Qiao,\n  Jiantao
                  Zhou, and Chao Dong",
               "title":"HAT: Hybrid Attention Transformer for
                   Image Restoration",
```

```
                "comments":"Extended version of HAT",
                "ref":"",
                "categories":"cs.CV",
                "abstract":"abstract text"
            }
        },
        ...
    ]
```

5 | Extra

To enhance our project, we considered creating a web app that would simulate a user-friendly graphical interface for performing the desired queries.

5.0.1. DISCLAIMER

Due to temporal limit we had, we didn't manage error handling perfectly: so you have to write syntactically correct queries.

5.1. WebApp structure

As shown in the photo below, our application is built on a client-server architecture that leverages the HTTP protocol. The user initiates an HTTP request to the server (where a NodeJS instance is running), and the server responds with the correct HTML page along with the corresponding scripts and style sheets.

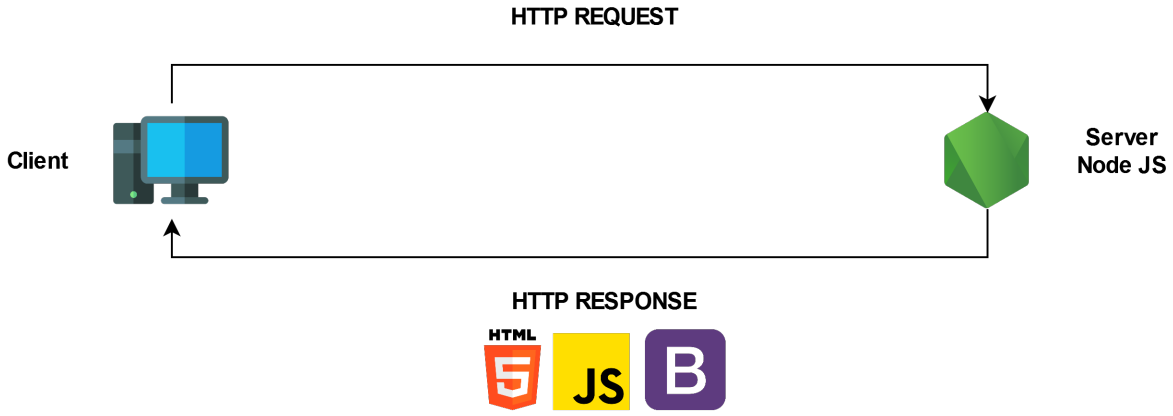

Figure 5.1: WebApp structure

Later, as we will see in the explanatory screenshots, the user selects the desired technology (Neo4j, ElasticSearch) and performs the query.
Subsequently, our server communicates with Neo4j or ElasticSearch, depending on the user's choice, to execute the requested query and return the result in the form of a JSON file, which will then be parsed on the client side.

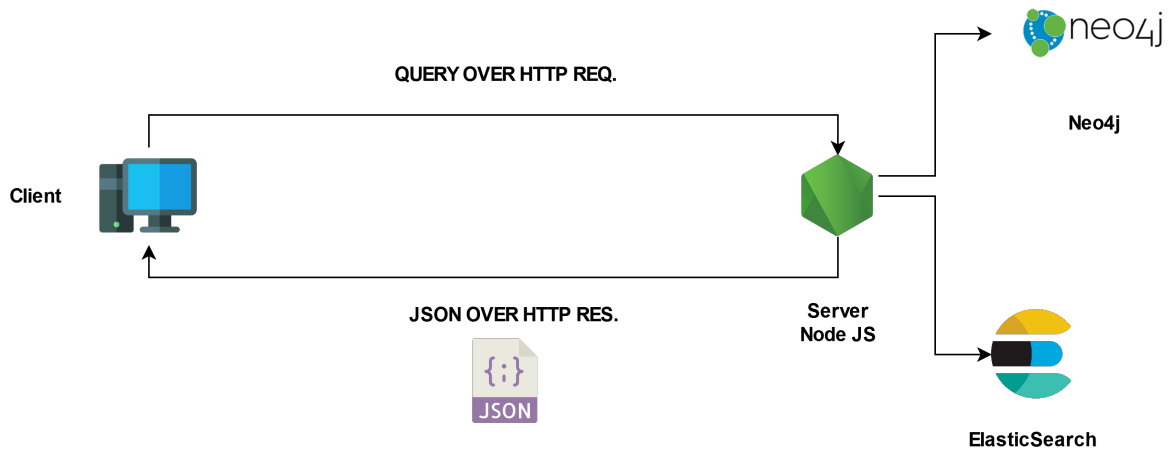

Figure 5.2: WebApp user interaction

5.2. WebApp screens

5.2.1. Neo4j screens

Initially, the website appears as follows: at the center, there is a canvas where the graph representing the result of the performed query will be drawn. In the toolbar, it's possible to select the desired technology, choose a query from the provided options, clear the input, and initiate the query. Finally, if the user wishes, they can write their own custom query in the lower bar.

Figure 5.3: Main screen for Neo4j

Clicking the yellow star reveals a screen with 10 queries that can be clicked to automatically insert them into the text field below.

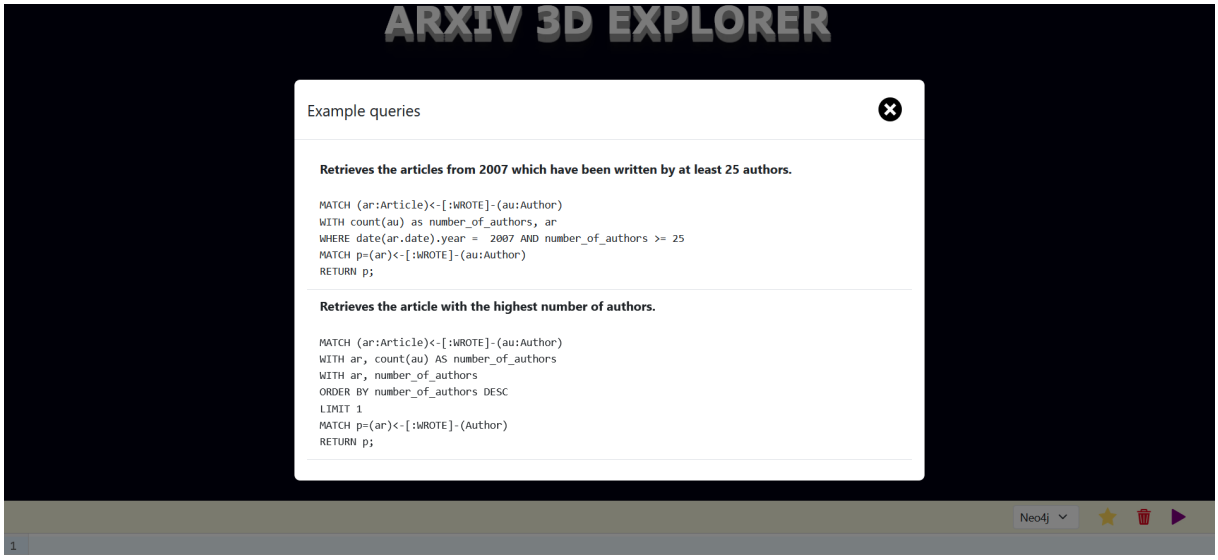

Figure 5.4: Queries screen for Neo4j

When the client receives the JSON file containing the query result, parsing is performed, and the corresponding graph is displayed. It's important to note that nodes represent authors and articles, while edges represent relationships between authors and articles.

Figure 5.5: Result screen for Neo4j

5.2.2. ElasticSearch screens

The main screen is almost identical to the one before, with the only difference being the presence of a container in the center where the query result will be displayed.

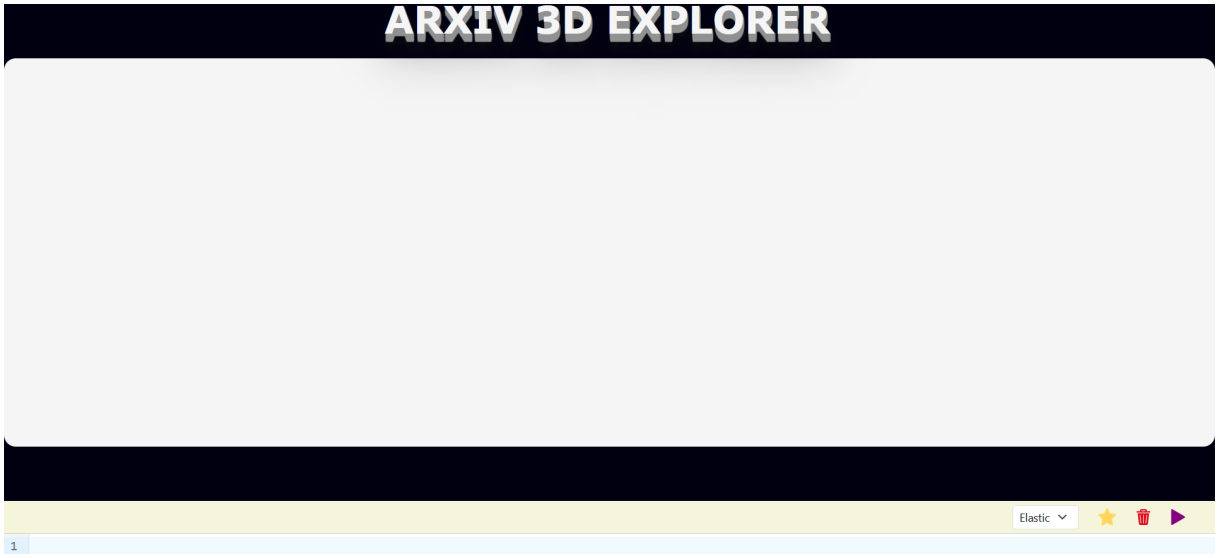

Figure 5.6: Main screen for ElasticSearch

Clicking the yellow star reveals a screen with 10 queries that can be clicked to automatically insert them into the text field below.

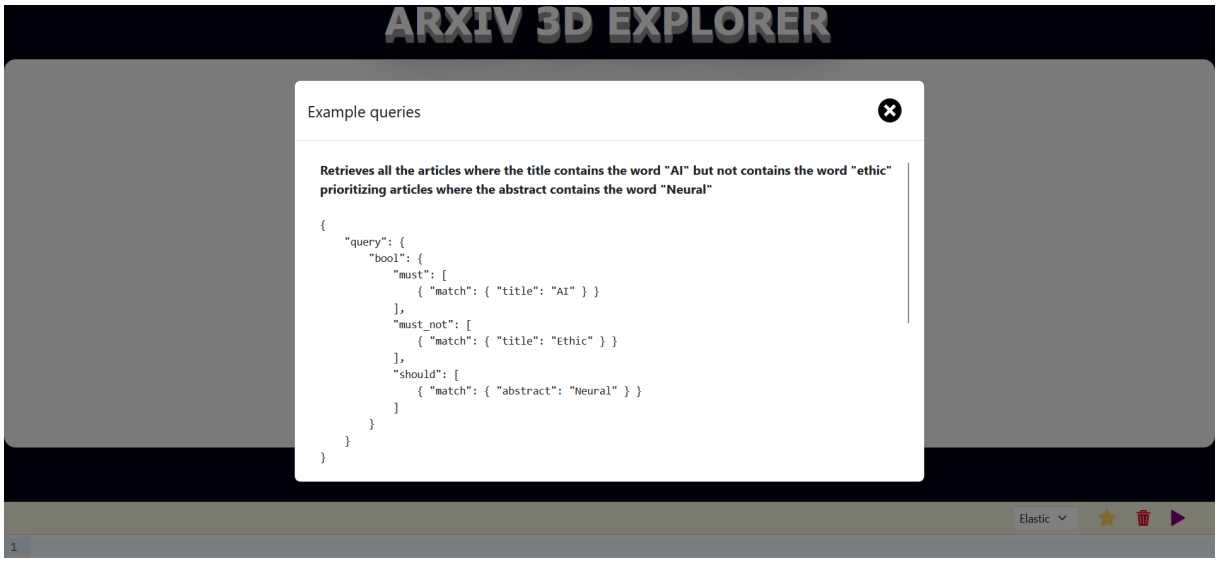

Figure 5.7: Queries screen for ElasticSearch

When the client receives the JSON file containing the query result, parsing is performed, and the corresponding parsed text is shown.

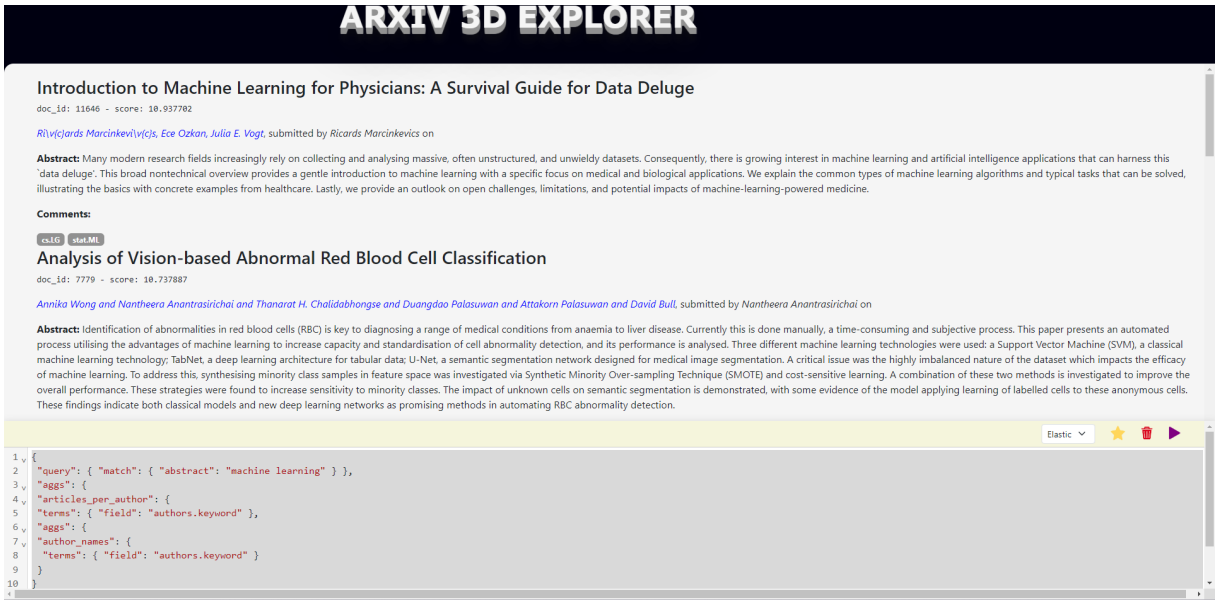

Figure 5.8: Result screen for ElasticSearch

List of Figures