Assignment 1

Student

- Lixiao Yang, ly364@drexel.edu

1. Academic Honesty Statement

Please copy and paste **the academic honesty statement** here.

Student Names: **Lixiao Yang**

Date: **April 14, 2023**

You only need to submit this once for the entire course.
You will not receive any grade without the statement.

2. Assignment Data

1. breakthrough drug for schizophrenia
2. new schizophrenia drug
3. new approach for treatment of schizophrenia
4. new hopes for schizophrenia patients

3. Assignment Tasks

3.1. Term-Document Matrix Representation (1 point)

Term	Doc 1	Doc 2	Doc 3	Doc 4
breakthrough	1	0	0	0
drug	1	1	0	0
for	1	0	1	1
schizophrenia	1	1	1	1
new	0	1	1	1
approach	0	0	1	0
treatment	0	0	1	0
of	0	0	1	0

Term	Doc 1	Doc 2	Doc 3	Doc 4
hopes	0	0	0	1
patients	0	0	0	1

3.2. Inverted Index Representation (1 point)

Dictionary	Postings
breakthrough	1
drug	1, 2
for	1, 3, 4
schizophrenia	1, 2, 3, 4
new	2, 3, 4
approach	3
treatment	3
of	3
hopes	4
patients	4

3.3. Indexing with ElasticSearch (1 point)

```
PUT ly364_info300_schizophrenia/_doc/1
{
  "content": "breakthrough drug for schizophrenia"
}

PUT ly364_info300_schizophrenia/_doc/2
{
  "content": "new schizophrenia drug"
}

PUT ly364_info300_schizophrenia/_doc/3
{
  "content": "new approach for treatment of schizophrenia"
}

PUT ly364_info300_schizophrenia/_doc/4
{
```

```
    "content": "new hopes for schizophrenia patients"
  }
```

3.4. Retrieval with ElasticSearch (0.5 points)

Request:

```
GET ly364_info300_schizophrenia/_search
```

Response:

```
{
  "took" : 444,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 4,
      "relation" : "eq"
    },
    "max_score" : 1.0,
    "hits" : [
      {
        "_index" : "ly364_info300_schizophrenia",
        "_type" : "_doc",
        "_id" : "1",
        "_score" : 1.0,
        "_source" : {
          "content" : "breakthrough drug for schizophrenia"
        }
      },
      {
        "_index" : "ly364_info300_schizophrenia",
        "_type" : "_doc",
        "_id" : "2",
        "_score" : 1.0,
        "_source" : {
          "content" : "new schizophrenia drug"
        }
      },
```

```json
    {
      "_index" : "ly364_info300_schizophrenia",
      "_type" : "_doc",
      "_id" : "3",
      "_score" : 1.0,
      "_source" : {
        "content" : "new approach for treatment of schizophrenia"
      }
    },
    {
      "_index" : "ly364_info300_schizophrenia",
      "_type" : "_doc",
      "_id" : "4",
      "_score" : 1.0,
      "_source" : {
        "content" : "new hopes for schizophrenia patients"
      }
    }
  ]
 }
}
```

3.5. Boolean Query

```
"schizophrenia" AND "drug"
```

3.5.1. Manual Analysis (0.75 point)

1. Documents containing `schizophrenia`: 1, 2, 3, 4
2. Documents containing `drug`: 1, 2
3. Merging results of 1 and 2 with `AND`: 1, 2

3.5.2. ElasticSearch Query (1 point)

Request:

```
GET ly364_info300_schizophrenia/_search
{
  "query": {
    "bool": {
      "must": [
        {"term": {"content": "schizophrenia"}},
        {"term": {"content": "drug"}}
```

```
        ]
      }
    }
  }
```

Response:

```
{
  "took" : 4,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 2,
      "relation" : "eq"
    },
    "max_score" : 0.9245879,
    "hits" : [
      {
        "_index" : "ly364_info300_schizophrenia",
        "_type" : "_doc",
        "_id" : "2",
        "_score" : 0.9245879,
        "_source" : {
          "content" : "new schizophrenia drug"
        }
      },
      {
        "_index" : "ly364_info300_schizophrenia",
        "_type" : "_doc",
        "_id" : "1",
        "_score" : 0.8365319,
        "_source" : {
          "content" : "breakthrough drug for schizophrenia"
        }
      }
    ]
  }
}
```

3.6. Compound Query

```
"for" AND ("drug" OR "approach")
```

3.6.1. Manual Analysis (1.25 point)

1. Documents containing `for`: 1, 3, 4
2. Documents containing `drug`:1, 2
3. Documents containing `approach`: 3
4. Merging results of 2 and 3 with `OR`: 1, 2, 3
5. Merging results of 1 and 4 with `AND`: 1, 3

3.6.2. ElasticSearch Query (1.5 points)

Request:

```
GET ly364_info300_schizophrenia/_search
{
  "query": {
    "bool": {
      "must": [
        {"term": {"content": "for"}},
        {
          "bool": {
            "should": [
              {"term": {"content": "drug"}},
              {"term": {"content": "approach"}}
            ]
          }
        }
      ]
    }
  }
}
```

Response:

```
{
  "took" : 6,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
```

```
      "skipped" : 0,
      "failed" : 0
    },
    "hits" : {
      "total" : {
        "value" : 2,
        "relation" : "eq"
      },
      "max_score" : 1.3733702,
      "hits" : [
        {
          "_index" : "ly364_info300_schizophrenia",
          "_type" : "_doc",
          "_id" : "3",
          "_score" : 1.3733702,
          "_source" : {
            "content" : "new approach for treatment of schizophrenia"
          }
        },
        {
          "_index" : "ly364_info300_schizophrenia",
          "_type" : "_doc",
          "_id" : "1",
          "_score" : 1.0998137,
          "_source" : {
            "content" : "breakthrough drug for schizophrenia"
          }
        }
      ]
    }
}
```