

Assignment 1

Student(s)

- Weimao Ke, wk77@drexel.edu
- John Smith, jps2020@drexel.edu

1. Academic Honesty Statement

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

Student Names: **Weimao Ke and John Smith**

Date: **July 1, 2020**

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)

You can use a table like this for the matrix representation:

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
hopes	0	0	0	1
patients	0	0	0	1

Replace table content with actual terms and values.

3.2. Inverted Index Representation (1 point)

Inverted index (table/figure) here. Note that this is NOT the same as the above matrix.

Term	Doc 1
approach	doc3
breakthrough	doc1
drug	doc1,doc2
for	doc1,doc3,doc4
hopes	doc4
new	doc2,doc3,doc4
of	doc3
patients	doc4
schizophrenia	doc1,doc2,doc3,doc4
treatment	doc3

3.3. Indexing with ElasticSearch (1 point)

```
PUT /zhangqirui_info300_schizophrenia/_doc/1
{
  "content": "Schizophrenia breakthrough drug."
}

PUT /zhangqirui_info300_schizophrenia/_doc/2
{
  "content": "New schizophrenia drug."
}

PUT /zhangqirui_info300_schizophrenia/_doc/3
{
  "content": "New approach for schizophrenia treatment"
}

PUT /zhangqirui_info300_schizophrenia/_doc/4
{
  "content": "New hopes for schizophrenia patients"
}
```

Resopnse:

```
{
  "_index": "zhangqirui_info300_schizophrenia",
```

```
"_id": "1",
"_version": 3,
"result": "updated",
"_shards": {
  "total": 2,
  "successful": 1,
  "failed": 0
},
"_seq_no": 8,
"_primary_term": 1
}

{
  "_index": "zhangqirui_info300_schizophrenia",
  "_id": "2",
  "_version": 3,
  "result": "updated",
  "_shards": {
    "total": 2,
    "successful": 1,
    "failed": 0
  },
  "_seq_no": 9,
  "_primary_term": 1
}

{
  "_index": "zhangqirui_info300_schizophrenia",
  "_id": "3",
  "_version": 3,
  "result": "updated",
  "_shards": {
    "total": 2,
    "successful": 1,
    "failed": 0
  },
  "_seq_no": 10,
  "_primary_term": 1
}

{
  "_index": "zhangqirui_info300_schizophrenia",
  "_id": "4",
  "_version": 3,
  "result": "updated",
  "_shards": {
    "total": 2,
    "successful": 1,
    "failed": 0
  },
  "_seq_no": 11,
  "_primary_term": 1
}
```

3.4. Retrieval with ElasticSearch (0.5 points)

Request:

```
GET /zhangqirui_info300_schizophrenia/_search
{
  "query": {
    "match_all": {}
  }
}
```

Response:

```
{
  "took": 17,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 4,
      "relation": "eq"
    },
    "max_score": 1,
    "hits": [
      {
        "_index": "zhangqirui_info300_schizophrenia",
        "_id": "1",
        "_score": 1,
        "_source": {
          "content": "Schizophrenia breakthrough drug."
        }
      },
      {
        "_index": "zhangqirui_info300_schizophrenia",
        "_id": "2",
        "_score": 1,
        "_source": {
          "content": "New schizophrenia drug."
        }
      },
      {
        "_index": "zhangqirui_info300_schizophrenia",
        "_id": "3",
        "_score": 1,
```

```
    "_source": {
      "content": "New approach for schizophrenia treatment"
    }
  },
  {
    "_index": "zhangqirui_info300_schizophrenia",
    "_id": "4",
    "_score": 1,
    "_source": {
      "content": "New hopes for schizophrenia patients"
    }
  }
]
}
```

3.5. Boolean Query

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

3.5.1. Manual Analysis (0.75 point)

Documents 1 and 2 both contain the word "drug" along with "schizophrenia". Thus, the returned documents for this query would be:

1. Documents containing **schizophrenia**:doc1,doc2,doc3,doc4
2. Documents containing **drug**:doc1,doc2
3. Merging results of 1 and 2 with **AND**:doc1,doc2

3.5.2. ElasticSearch Query (1 point)

Request:

```
GET /zhangqirui_info300_schizophrenia/_search
{
  "query": {
    "bool": {
      "must": [
        { "match": { "content": "schizophrenia" } },
        { "match": { "content": "drug" } }
      ]
    }
  }
}
```

Response:

```

{
  "took": 2,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 2,
      "relation": "eq"
    },
    "max_score": 1.0898671,
    "hits": [
      {
        "_index": "zhangqirui_info300_schizophrenia",
        "_id": "1",
        "_score": 1.0898671,
        "_source": {
          "content": "Schizophrenia breakthrough drug."
        }
      },
      {
        "_index": "zhangqirui_info300_schizophrenia",
        "_id": "2",
        "_score": 1.0898671,
        "_source": {
          "content": "New schizophrenia drug."
        }
      }
    ]
  }
}

```

3.6. Compound Query

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

3.6.1. Manual Analysis (1.25 point)

1. Documents containing **for**:doc1,doc3,doc4
2. Documents containing **drug**:doc1,doc2
3. Documents containing **approach**:doc3
4. Merging results of 2 and 3 with **OR**:doc1,doc2,doc3
5. Merging results of 1 and 4 with **AND**:doc3

3.6.2. Elasticsearch Query (1.5 points)

Request:

```
GET /zhangqirui_info300_schizophrenia/_search
{
  "query": {
    "bool": {
      "must": [
        { "match": { "content": "for" } },
        {
          "bool": {
            "should": [
              { "match": { "content": "drug" } },
              { "match": { "content": "approach" } }
            ]
          }
        }
      ]
    }
  }
}
```

Response:

```
{
  "took": 1,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 1,
      "relation": "eq"
    },
    "max_score": 1.786113,
    "hits": [
      {
        "_index": "zhangqirui_info300_schizophrenia",
        "_id": "3",
        "_score": 1.786113,
        "_source": {
          "content": "New approach for schizophrenia treatment"
        }
      }
    ]
  }
}
```

```
}  
}
```