

COMP9313-2019T2 Assignment #3 (ElasticSearch)

Student Number: z5195349 Student Name: Wenxun Peng

1. Spark-submit command

```
spark-submit --class "CaseIndex" --packages org.scalaj:scalaj-http_2.11:2.4.2 --master local[2] target/scala-2.11/caseindex_2.11-1.0.jar cases_test
```

2. Index design

```
1 {
2   "cases": {
3     "properties": {
4       "name": {
5         "type": "text"
6       },
7       "url": {
8         "type": "text"
9       },
10      "catchphrases": {
11        "type": "text"
12      },
13      "sentences": {
14        "type": "text"
15      },
16      "person": {
17        "type": "text"
18      },
19      "location": {
20        "type": "text"
21      },
22      "organization": {
23        "type": "text"
24      }
25    }
26  }
27 }
```

My index has 7 fields:

- 1.Name: The title of this case
- 2.URL: The original source of the legal cases
- 3.Catchphrases: The summary of the cases
- 4.Sentences: The list of sentences contained in the legal cases
- 5.Person: The list of people mentioned in the legal cases
- 6.Location: The list of locations mentioned in the legal cases
- 7.Organization: The list of organizations mentioned in the legal cases

3. Solution implementation

Step1: Create an index named legal_idx and map it in the mentioned fields above by using HTTP request to ElasticSearch server.

Step2: Load all the local XML files and extract the above fields respectively, and then send HTTP requests which include the catchphrases and sentences to the Stanford CoreNLP server to determine the name entities (People, Locations and Organizations are needed in my solutions, in fact, there are more entities, such as Date, Number and so on.). Get the JSON format file as responses and then handle it.

Step3: Extract what we want, such as get the PERSON, LOCATION and ORGANIZATION attribute and put them in a set to prepare to use

Step4: Create a new document for each XML file in JSON format and then send it to ElasticSearch server by HTTP request.

4. Queries example

Due to the limitation of the length of this report, only part of the output is displayed.

(1) Queries based on entity type

Command:

```
curl -X GET http://localhost:9200/legal\_idx/cases/\_search?pretty&q=location:Melbourne
```

```

"took": 1,
"timed_out": false,
"_shards": {
  "total": 5,
  "successful": 5,
  "skipped": 0,
  "failed": 0
},
"hits": {
  "total": 5,
  "max_score": 1.0,
  "hits": [
    {
      "_index": "legal_idx",
      "_type": "cases",
      "_id": "Beyazkilinc v Manager Baxter Immigration Reception & Processing Centre [2006] FCA 16 (18 January 2006)",
      "_score": 1.0,
      "_source": {
        "name": "Beyazkilinc v Manager Baxter Immigration Reception & Processing Centre [2006] FCA 16 (18 January 2006)",
        "url": "http://www.austlii.edu.au/au/cases/cth/FCA/2006/16.html",
        "catchphrases": "removal from australia of illegal non-citizen under s 198(6) interlocutory application whether s
plicant's mental condition migration",
        "sentences": "1 This application came on for hearing yesterday morning when the applicant sought an interlocutory
for psychiatric care and treatment. It was adjourned for further hearing this afternoon because I was informed that the seco
8 (Cth) (the Act) early tomorrow morning. The application to transfer him for treatment from the Baxter Centre in those circ
pendent from removing the applicant from Australia until he is deemed medically fit for removal from Australia, and an inter

```

Command:

`curl -X GET http://localhost:9200/legal_idx/cases/_search?pretty&q=person:John`

```

{
  "took": 1,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": 2,
    "max_score": 0.20410335,
    "hits": [
      {
        "_index": "legal_idx",
        "_type": "cases",
        "_id": "Tower Software Engineering Pty Limited; Pendant Software Pty Limited v Harwood [2006] FCA 717 (6 June 2006)",
        "_score": 0.20410335,
        "_source": {
          "name": "Tower Software Engineering Pty Limited; Pendant Software Pty Limited v Harwood [2006] FCA 717 (6 June 2006)",
          "url": "http://www.austlii.edu.au/au/cases/cth/FCA/2006/717.html",
          "catchphrases": "interlocutory application to restrain defendant from taking any further step in a proceeding/applicati

```

Command:

`curl -X GET`

`"http://localhost:9200/legal_idx/cases/_search?pretty&q=organization:State%20Bank%20New%20South%20Wales"`

```

{
  "took": 4,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": 4,
    "max_score": 4.3116612,
    "hits": [
      {
        "_index": "legal_idx",
        "_type": "cases",
        "_id": "Australian Liquor, Hospitality and Miscellaneous Workers Union v Prestige Property Services Pty Ltd [2006] FCA 11 (23 January 2006)",
        "_score": 4.3116612,
        "_source": {
          "name": "Australian Liquor, Hospitality and Miscellaneous Workers Union v Prestige Property Services Pty Ltd [2006] FCA 11 (23 January 2006)",
          "url": "http://www.austlii.edu.au/au/cases/cth/FCA/2006/11.html",
          "catchphrases": "award breach failure to pay rates of pay specified rates of pay for employees of contractors cleaning premises of state govern
mentality ordinary meaning of instrumentality whether meaning affected by history of award and context in which it was made transition from state to feder

```

(1) Query based on general terms

Command:

`curl -X GET`

`http://localhost:9200/legal_idx/cases/_search?pretty&q=(criminal%20AND%20law)`

```

"took": 1,
"timed_out": false,
"_shards": {
  "total": 5,
  "successful": 5,
  "skipped": 0,
  "failed": 0
},
"hits": {
  "total": 5,
  "max_score": 1.0,
  "hits": [
    {
      "_index": "legal_idx",
      "_type": "cases",
      "_id": "Beyazkilinc v Manager Baxter Immigration Reception & Processing Centre [2006] FCA 16 (18 January 2006)",
      "_score": 1.0,
      "_source": {
        "name": "Beyazkilinc v Manager Baxter Immigration Reception & Processing Centre [2006] FCA 16 (18 January 2006)",
        "url": "http://www.austlii.edu.au/au/cases/cth/FCA/2006/16.html",
        "catchphrases": "removal from australia of illegal non-citizen under s 198(6) interlocutory application whether s
plicant's mental condition migration",

```