

Assignment3

How to run?

- change to the directory where build.sbt file is located, then run the following commands:
9313, sbt package, elasticsearch, corenlp-server.sh
- **spark-submit --class "CaseIndex" --packages org.scalaj:scalaj-http_2.11:2.4.2,org.scalatestplus.play:scalatestplus-play_2.11:4.0.3 --master local[2] JAR_FILE_PATH FULL_PATH_OF_DIRECTORY_WITH_CASE_FILES**

Design & Implementation:

1. First, get all input files from input directory(if .DS_Store included, please delete it)
2. Create a new index to store **legal_idx** data
3. Create a new mapping, with a single type (named **cases**)

```
"mappings" : {  
  "cases" : {  
    "properties" : {  
      "catchphrase" : {  
        "type" : "text"  
      },  
      "id" : {  
        "type" : "text"  
      },  
      "location" : {  
        "type" : "text"  
      },  
      "name" : {  
        "type" : "text"  
      },  
      "organization" : {  
        "type" : "text"  
      },  
      "person" : {  
        "type" : "text"  
      },  
      "sentence" : {  
        "type" : "text"  
      },  
      "url" : {  
        "type" : "text"  
      }  
    }  
  }  
}
```

Index Design:

- **id**: filename of case
- **name**: The title of the case
- **url**: The original source of the legal case
- **catchphrase**: The summary of the case
- **sentence**: The list of sentences contained in the legal case
- **person**: The list of people found in the legal case
- **location**: The list of locations found in the legal case
- **organization**: The list of organizations found in the legal case

4. Use a loop to deal with each file, and do as following:
 - Load file content, and get file name
 - Get all information from file content, such as title, url, catchphrases, sentences
5. Pass all sentences to Named Entity Recognition API then parse response to JSON object
6. Get people, locations, organisations entities form JSON object, then convert them to String
7. Create a new document with above data and and sending HTTP request to Elasticsearch server.

```
// create a new document  
val Data = s""""{"id":"${filename}","name":"${name}","url":"${url}",  
  "catchphrase":"${catchphrase_all.toString.filter(_ >= ' ')}",  
  "sentence":"${new_sentence_list}",  
  "person":"${people_list}","location":"${locations_list}","organization":"${organizations_list}"}"""  
  
val update_document = Http("http://localhost:9200/legal_idx/cases/"+filename+"?pretty")  
  .postData(Data).method("PUT").header("Content-Type", "application/json")  
  .option(HttpOptions.readTimeout(10000)).asString
```

Example Queries

(use only search based on query strings):

- **person:** curl -X GET "http://localhost:9200/legal_idx/cases/_search?pretty&q=person:John"
- **Location:** 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" : 2,
  "max_score" : 0.6682933,
  "hits" : [
    {
      "_index" : "legal_idxxx",
      "_type" : "cases",
      "_id" : "06_717",
      "_score" : 0.6682933,
      "_source" : {
        "id" : "06_717",
        "name" : "Tower Software I",
        "url" : "http://www.austlii.edu.au/au/other/dfat/special/interloc/interloc.htm",
        "catchphrase" : "interloc"
      }
    }
  ]
}
```

```
"took" : 1,  
"timed_out" : false,  
"_shards" : {  
  "total" : 5,  
  "successful" : 5,  
  "skipped" : 0,  
  "failed" : 0  
},  
"hits" : {  
  "total" : 1,  
  "max_score" : 0.2876821,  
  "hits" : [  
    {  
      "_index" : "legal_idxxx",  
      "_type" : "cases",  
      "_id" : "06_11",  
      "_score" : 0.2876821,  
      "_source" : {  
        "id" : "06_11",  
        "name" : "Australian Liquor  
url" : "http://www.austlii.edu.au/au/other/dfat/special/  
catchphrase" : "award b
```

- **Organization:** curl -X GET "http://localhost:9200/legal_idx/cases/_search?pretty&q=organization:Commonwealth%20Savings%20Bank%20of%20Australia"
- **General:** curl -X GET "http://localhost:9200/legal_idx/cases/_search?pretty&q=(criminal%20AND%20law)"

```
{
  "took" : 8,
  "timed_out" : false,
  "_shards" : {
    "total" : 5,
    "successful" : 5,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : 5,
    "max_score" : 2.2520907,
    "hits" : [
      {
        "_index" : "legal_idxxx",
        "_type" : "cases",
        "_id" : "06_11",
        "_score" : 2.2520907,
        "_source" : {
          "id" : "06_11",
          "name" : "Australian Li
          "url" : "http://www.aus
          "catchphrase" : "award
```

```
"took" : 1,
"timed_out" : false,
"_shards" : {
  "total" : 5,
  "successful" : 5,
  "skipped" : 0,
  "failed" : 0
},
"hits" : {
  "total" : 2,
  "max_score" : 1.0326371,
  "hits" : [
    {
      "_index" : "legal_idxxx",
      "_type" : "cases",
      "_id" : "06_717",
      "_score" : 1.0326371,
      "_source" : {
        "id" : "06_717",
        "name" : "Tower Software",
        "url" : "http://www.austlii.edu.au/au/other/dfat/special/trade/catchphrase" : "interloper"
```