**Assignment 3**

**Spark-submit command**



**Index Design**

****

My mapping has 8 fields:

1. Id: filename of case report
2. Name: The title of the case
3. URL: the original source of the legal report
4. Catchphrase: The summary of the case
5. Sentence: The list of sentences contained in the legal case report
6. Person: The list of people found in the legal case report
7. Location: The list of locations found in the legal case report
8. Organization: The list of organizations found in the legal case report

**Solution Implementation**

1. First, create an index and mapping as mentioned previously by HTTP request.
2. Then, get directory path from argument. To process each file, starting with getting information from each file.



1. Next, sending each sentence to CoreNLP server to find name entities. Parsing the response to JSON object and extracting “person”, “location” and “organization” entities from response.



1. Finally, create a new document and sending HTTP request to Elasticsearch server.



**Queries example**

1. General term search

**Command:** curl -X GET "http://localhost:9200/legal\_idx/cases/\_search?pretty&q=(criminal%20AND%20law)"

**Result:**



1. Entity search

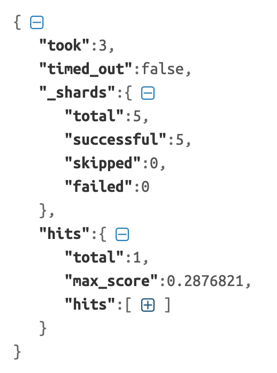
**Command:** curl -X GET "http://localhost:9200/legal\_idx/cases/\_search?pretty&q=person:John"

**Result:**



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

**Result:**

****

**Command:** curl -X GET "http://localhost:9200/legal\_idx/cases/\_search?pretty&q=organization:State%20Bank%20of%20New%20South%20Wales"

**Result:**

