

Command to run after “CaseIndex.jar” created

```
spark-submit --class "CaseIndex" --packages org.scalaj:scalaj-http_2.11:2.3.0,org.scala-lang.modules:scala-xml_2.11:1.2.0,com.typesafe.play:play-json_2.11:2.7.4 --master local[2] CaseIndex.jar cases_test
```

External Package Dependency

Following are what external package dependency and its functionality within this assignment (Also shown in sbt file).

1. **scalaj-http** For HTTP requests and response retrieval in Scala
2. **scala-xml** For XML file extraction in Scala
3. **play-json** For JSON-format information parsing in Scala

Overview

Generally, this assignment is based on elastic-search and core-NLP services. After parsing the files, sentences are sent to core-NLP service to get the prescribed labels (i.e. person, organization and location). Then, these labels consist of indexes, which are, therefore, sent to elastic-search service as payload with all other labelled information extracted from XML file. Finally, users can do queries on entity types (i.e. the labels) or on general terms through elastic search service.

The whole workflow is described as below:

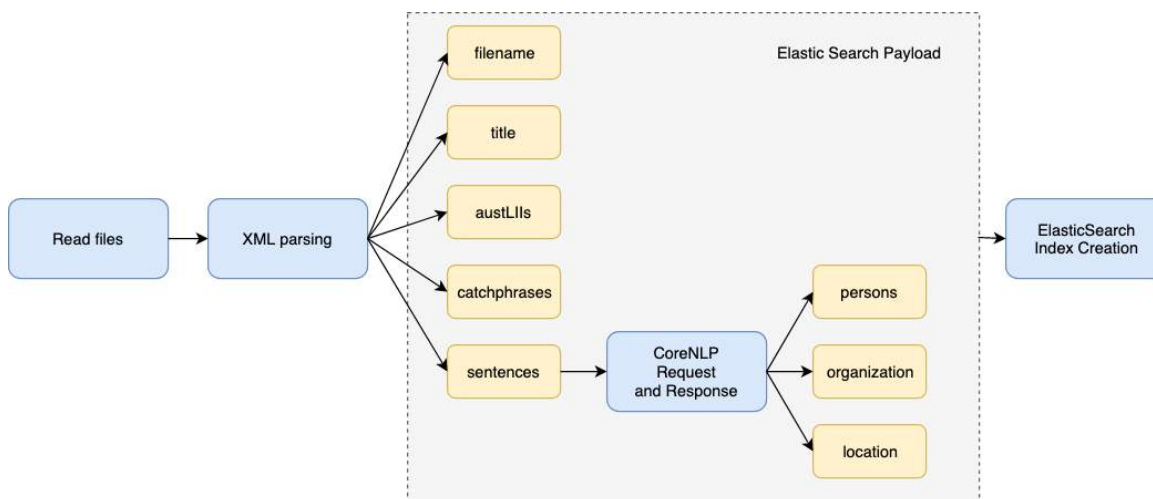


Fig 1. Workflow for this assignment

The Implementation of Data Curation

At first, Java IO methods are used for file reading inside the directory which is input as a command-line argument. Then Scala XML methods are used to extract information from XML files by label matching, storing texts from XML texts to a JSON-like formatted string.

The Implementation of CoreNLP Parsing

CoreNLP is used for parsing sentences to extract labels (displayed in “ner” labels). When standardized payload sent to the services, the labels are returned. Therefore, it is convenient to storing only the prescribed labels to three lists, respectively.

The Implementation of ElasticSearch Index Creation

Initially, ScalaJ HTTP methods are used to create entity format in elastic search service, such that the format includes file names, titles of files, austLIIs, catchphrases, sentences, persons, organizations and locations. When the three labels are extracted by CoreNLP service, they consist of payload and are sent to ElasticSearch, which returns 201 created.

Example Queries

1. curl -X GET "http://localhost:9200/legal_idx/cases/_search?pretty&q=location:Melbourne"

```
curl -X GET "http://localhost:9200/legal_idx/cases/_search?pretty&q=location:Melbourne"

{"took": 4,
 "timed_out": false,
 "errors": 0,
 "total": 5,
 "successful": 5,
 "skipped": 0,
 "failed": 0
},
{
  "hits": 1,
  "total": 1,
  "max_score": 0.8658662,
  "hits": [
    {
      "_id": "190.11.101",
      "_type": "legal_idx",
      "_score": 0.8658662,
      "_source": {
        "title": "190.11.101",
        "text": "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/other/aufrca/2006/11.html",
        "abstract": "Failure to pay rates of pay specified; rates of pay for employees of contractors cleaning premises of state government departments and instrumentalities; whether Victorian Arts Centre falls within the category of state government departments and instrumentalities; whether Victorian Arts Centre falls within the category of state government instrumentalities; whether meaning affected by history of award and context in which it was made; transition from state to federal regulation of contract cleaning industry; whether common understanding; purpose of provision in award; whether term in award; relevance of list of government bodies in another award in an unrelated industry; relevance of capital listings; relevance of possible anomalies not affecting case itself; quantum of appropriate remedy; state government departments and instrumentalities; workplace relations; awards and awards."
      }
    }
  ]
}
```

2. curl -X GET "http://localhost:9200/legal_idx/cases/_search?pretty&q=organization:Victorian"

```
curl -X GET "http://localhost:9200/legal_idx/cases/_search?pretty&q=organization:Victorian"

{"took": 4,
 "timed_out": false,
 "errors": 0,
 "total": 5,
 "successful": 5,
 "skipped": 0,
 "failed": 0
},
{
  "hits": 1,
  "total": 1,
  "max_score": 0.5354405,
  "hits": [
    {
      "_id": "190.11.101",
      "_type": "legal_idx",
      "_score": 0.5354405,
      "_source": {
        "title": "190.11.101",
        "text": "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/other/aufrca/2006/11.html",
        "abstract": "Failure to pay rates of pay specified; rates of pay for employees of contractors cleaning premises of state government departments and instrumentalities; whether Victorian Arts Centre falls within the category of state government departments and instrumentalities; whether meaning affected by history of award and context in which it was made; transition from state to federal regulation of contract cleaning industry; whether common understanding; purpose of provision in award; whether term in award; relevance of list of government bodies in another award in an unrelated industry; relevance of capital listings; relevance of possible anomalies not affecting case itself; quantum of appropriate remedy; state government departments and instrumentalities; workplace relations; awards and awards."
      }
    }
  ]
}
```

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

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

{"took": 5,
 "timed_out": false,
 "errors": 0,
 "total": 5,
 "successful": 5,
 "skipped": 0,
 "failed": 0
},
{
  "hits": 2,
  "total": 2,
  "max_score": 0.5362922,
  "hits": [
    {
      "_id": "190.11.101",
      "_type": "legal_idx",
      "_score": 0.5362922,
      "_source": {
        "title": "190.11.101",
        "text": "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/other/aufrca/2006/11.html",
        "abstract": "Failure to pay rates of pay specified; rates of pay for employees of contractors cleaning premises of state government departments and instrumentalities; whether Victorian Arts Centre falls within the category of state government departments and instrumentalities; whether meaning affected by history of award and context in which it was made; transition from state to federal regulation of contract cleaning industry; whether common understanding; purpose of provision in award; whether term in award; relevance of list of government bodies in another award in an unrelated industry; relevance of capital listings; relevance of possible anomalies not affecting case itself; quantum of appropriate remedy; state government departments and instrumentalities; workplace relations; awards and awards."
      }
    },
    {
      "_id": "190.11.102",
      "_type": "legal_idx",
      "_score": 0.5362922,
      "_source": {
        "title": "190.11.102",
        "text": "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/other/aufrca/2006/11.html",
        "abstract": "Failure to pay rates of pay specified; rates of pay for employees of contractors cleaning premises of state government departments and instrumentalities; whether Victorian Arts Centre falls within the category of state government departments and instrumentalities; whether meaning affected by history of award and context in which it was made; transition from state to federal regulation of contract cleaning industry; whether common understanding; purpose of provision in award; whether term in award; relevance of list of government bodies in another award in an unrelated industry; relevance of capital listings; relevance of possible anomalies not affecting case itself; quantum of appropriate remedy; state government departments and instrumentalities; workplace relations; awards and awards."
      }
    }
  ]
}
```

4. curl -X GET "http://localhost:9200/legal_idx/cases/_search?pretty&q=(criminal%20AND%20law)"

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

{"took": 5,
 "timed_out": false,
 "errors": 0,
 "total": 5,
 "successful": 5,
 "skipped": 0,
 "failed": 0
},
{
  "hits": 2,
  "total": 2,
  "max_score": 0.5362922,
  "hits": [
    {
      "_id": "190.11.101",
      "_type": "legal_idx",
      "_score": 0.5362922,
      "_source": {
        "title": "190.11.101",
        "text": "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/other/aufrca/2006/11.html",
        "abstract": "Failure to pay rates of pay specified; rates of pay for employees of contractors cleaning premises of state government departments and instrumentalities; whether Victorian Arts Centre falls within the category of state government departments and instrumentalities; whether meaning affected by history of award and context in which it was made; transition from state to federal regulation of contract cleaning industry; whether common understanding; purpose of provision in award; whether term in award; relevance of list of government bodies in another award in an unrelated industry; relevance of capital listings; relevance of possible anomalies not affecting case itself; quantum of appropriate remedy; state government departments and instrumentalities; workplace relations; awards and awards."
      }
    },
    {
      "_id": "190.11.102",
      "_type": "legal_idx",
      "_score": 0.5362922,
      "_source": {
        "title": "190.11.102",
        "text": "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/other/aufrca/2006/11.html",
        "abstract": "Failure to pay rates of pay specified; rates of pay for employees of contractors cleaning premises of state government departments and instrumentalities; whether Victorian Arts Centre falls within the category of state government departments and instrumentalities; whether meaning affected by history of award and context in which it was made; transition from state to federal regulation of contract cleaning industry; whether common understanding; purpose of provision in award; whether term in award; relevance of list of government bodies in another award in an unrelated industry; relevance of capital listings; relevance of possible anomalies not affecting case itself; quantum of appropriate remedy; state government departments and instrumentalities; workplace relations; awards and awards."
      }
    }
  ]
}
```