Preethi Sureshkumar (psureshk)

HOME WORK 3 – Execution and Deployment Architecture with CPE

Aim :

To use CPE instead of UIMA Document Analyzer, to integrate a remote UIMA-AS service into the CPE pipeline and to deploy the aggregate analysis engine developed in homework 2 as a UIMA-AS service.

Creating and Running CPE:

Generally, a CPE is comprised of three components – Collection Reader, Analysis Engines and CASConsumer. FileSystemCollectionReader.java and FileSystemsCOllectionReader.xml files were already available in the Examples project. The input folder path has been modified to suit hw3-psureshk project's needs. The InputData folder is located under /src/main/resources/ folder and it consists of two text files – q001.txt and q002.txt. As mentioned in the handout, the CASConsumer is created based on the Evaluator (in my case AnswerScoreAnnotator) of homework 2. The Aggregate Analysis Engine of the previous homework has been used as the AAE descriptor in the CPE configurator UI. The file is then saved as hw3-psureshk-CPE.xml.

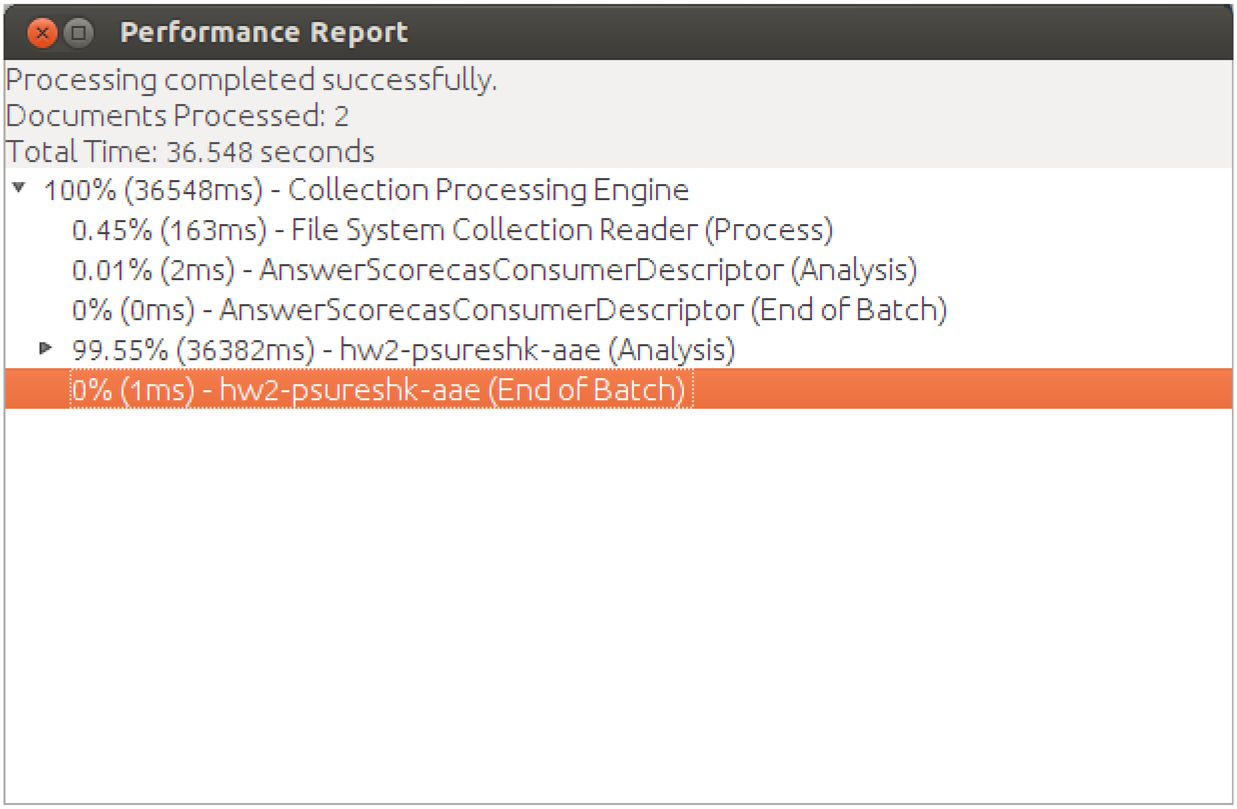
Output is :

```
+ 1.0 Booth shot Lincoln.
+ 0.9 Lincoln was shot by Booth.
- 0.9 Lincoln shot Booth.
+ 0.85 Booth assassinated Lincoln.
- 0.8 Booth was shot by Lincoln.
+ 0.75 Lincoln was assassinated by Booth.
- 0.75 Lincoln assassinated Booth.
- 0.65 Booth was assassinated by Lincoln.
Accuracy at : 1

+ 1.0 John loves Mary.
+ 0.8 John loves Mary with all his heart.
- 0.8 John doesn't love Mary.
- 0.7 Mary doesn't love John.
+ 0.7 Mary is dearly loved by John.
Accuracy at : 1
```

Preethi Sureshkumar (psureshk)

The Aggregate Analysis Engine **implements Stanford Corenlp** annotations to easily get the lemma and dependencies from the text files and t. The Stanford Corenlp has been implemented locally and the speed of the AAE is **36.548** seconds. Semantic Graphs, Lemma and POS annotations are used.



```
Processing completed successfully.
Documents Processed: 2
Total Time: 36.548 seconds
▾ 100% (36548ms) - Collection Processing Engine
    0.45% (163ms) - File System Collection Reader (Process)
    0.01% (2ms) - AnswerScorecasConsumerDescriptor (Analysis)
    0% (0ms) - AnswerScorecasConsumerDescriptor (End of Batch)
  ▸ 99.55% (36382ms) - hw2-psureshk-aae (Analysis)
    0% (1ms) - hw2-psureshk-aae (End of Batch)
```

Creating a UIMA-AS Client:

A client descriptor file has been created with the name scnlp-psureshk-client.xml which points to the remote UIMA-AS service (Stanford CoreNLP) using the brokerURL = tcp://mu.lti.cs.cmu.edu:61616 and endpoint ScnlpQueue. The timeout has been configured to be 500 secs. Once the stanford core nlp is called the control then comes to the new aggregated analysis engine similar to homework 2 but modified to meet the NamedEntityMention annotations of ClearTK. The dependencies are updated in the pom.xml of hw3-psureshk. A separate CPE descriptor has been created for this part of the homework and is executed. The accuracy in homework 2 is **1.** The accuracy by calling the Stanford NLP remotely is **0.5**. The speed of the remote execution is secs.

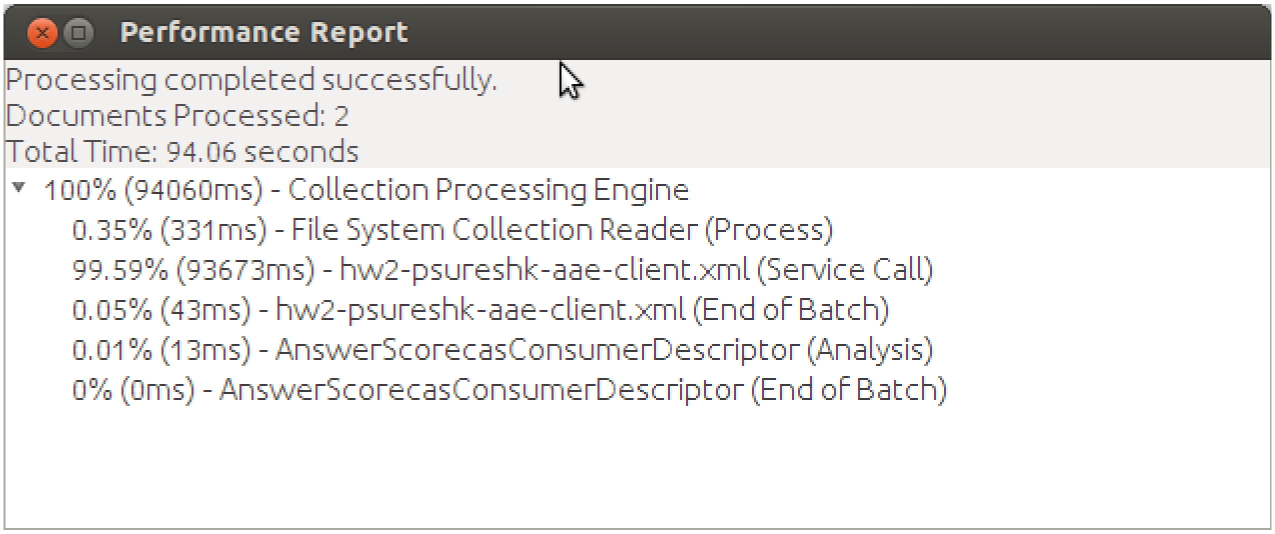
Deploying homework 2 as UIMA-AS service:

Deployment descriptor file was created by name hw2-psureshk-aae-deploy.xml which specified the brokerURL and QueueName relevant to the homework2 analysis engine. The current directory in the terminal is set to a writable directory called ActiveMQ. The local broker is started using the following command - /home/psureshk/apache-uima-as-2.4.0/bin/startBroker.sh. Then the homework2 has to be

Preethi Sureshkumar (psureshk)

deployed as a remote UIMA_AS service by using the command - /home/psureshk/apache-uima-as-2.4.0/bin/deployAsyncService.sh /home/psureshk/apache-uima-as-2.4.0/hw3-psureshk/src/main/resources/descriptors/hw2-psureshk-aae-deploy.xml.

A client descriptor has to be created which points to the above exposed remote service. A new CPE descriptor is created which has the client descriptor placed instead of the analysis engine descriptor. Once everything is setup, we can run the CPE and the result is obtained. It ran at a speed of 94.06 seconds.



```
Processing completed successfully.
Documents Processed: 2
Total Time: 94.06 seconds
▾ 100% (94060ms) - Collection Processing Engine
    0.35% (331ms) - File System Collection Reader (Process)
    99.59% (93673ms) - hw2-psureshk-aae-client.xml (Service Call)
    0.05% (43ms) - hw2-psureshk-aae-client.xml (End of Batch)
    0.01% (13ms) - AnswerScorecasConsumerDescriptor (Analysis)
    0% (0ms) - AnswerScorecasConsumerDescriptor (End of Batch)
```