11-791 Design and Engineering of Intelligent Information System
Assignment 3

Laleh Roosta Pour, AndrewID: lroostap

1. Task 1

For this task I created 3 main components of CPE: collection reader, analysis engine, and cas consumer . I created the collection reader descriptor "FileSystemCollectionReaderDescriptor.xml" and put the "FileSystemCollectionReader" as the corresponding java class. I imported uimaj-examples to my eclipse and used it as a help to do my homework. I created a casConsumer descriptor and in the corresponding java class I putt the printing parts of the project. I moved the printing part from the AnswerScoringAnnotator component of my hw2 to the consumer. I used the hw2 analysis engine. Using the CPE GUI, I ran my CPE process and saved the CPE, xml file as hw3-lroostap-CPE.xml.

2. Task 2

2.1 Task 2.1

I downloaded UIMA-AS, read the README file, installed it and set the proper variables and path.

2.2 Task 2.2

In this task, I created the UIMA-AS client descriptor, scnlp-lroostap-client.xml which connects to the remote Stanford Core-NLP service. I added the dependencies the required dependencies. Then I imported the Named Entity type to my AnswerScoringAnnotator, so that I would be able to integrate the Name Entity annotations from the Stanford CoreNLP service into my answer scoring component. I calculated a score from the Name Entity annotations and at the end put the average of ngram score and the name entity score as the final score. The precision for with the new scores didn't change, however, when I changed the ratio between the ngramScore and neScore it was shown that increasing the impact of neScore (giving more weight to neScore, not just calculating the average of neScore and ngramScore) results in lower precision. In hw2 I get the following results for the second input:

```
Question: Booth shot Lincoln?
Precision at 4 : 0.50000
```

After adding the new scores from NameEntity component and put the average of two scores and the final score I get this result for the second input:

```
Question: Booth shot Lincoln?
Precision at 4 : 0.50000
```

For calculating the neScore, first I checked the number of name entities and their type for the question, then for each answer I count the number of similar name entities with the question. At the end I compute the score by dividing the number of similar entities for each answer by the max number of name entities overall (I used this as some sort of normalization for the scores).

To show the effect of neScore I calculated the final score as: 80% (neScore) + 20% (ngramScore) and the result for the second input is:

```
Question: Booth shot Lincoln?
Precision at 4 : 0.25000
```

Th runtime for the scnlp is 1.351 seconds and hw2 processing time is less that 1 second (900 ms). The numbers are similar and it is important to mention that I used the Stanfordnlp component for tokenization in hw2.

2.3 Task 2.3

For this task, I collected all the dependencies with the dependency:copy-dependencies plug-in in "target/dependency" folder and pointed the UIMA_CLASSPATH to this folder and also ""target/-classes". I created "hw2-lroostap-aae-deploy.xml" deploy client file. Then I ran the broker and deploy my file. I started a UIMA-AS broker locally, and deployed my service to my local broker. I also created the client descriptor hw2-lroostap-aae-client.xml and sent my service through the CPE descriptor hw3-lroostap-aae-as-CPE.xml by calling the created client.