NAME: Ying Sheng
Andrew ID: yingshen
We expect you to highlight the features of your design and your system, evaluation

results, and comparison. Finally

Report for HW3:

Task 1.2:
Analysis engine (AE) implemented in HW2 has the property of processing the document one by one. CPE in this task can apply AE to a collection of documents.

The design here has 3 component:
> 1. FileSystemCollectionReader: preparing raw data to data format usable for AE
> 2. Analysis Engine: the AAE created in HW2, the evaluator component is deleted
> 3. CasConsumer: collect the Cas data generated by AE and output in a certain format, (for here, stdout is used.)

Task 2.2:
UIMA AS allows us to call service not stored locally. A client containing information of *Broker, endpoint* can call service through Broker and receive the replies. It is equivalent as an AE and can be incorporated into CPE.

The design here add a scnlp-client before hw2-aae in 2nd part of CPE:
> 1. FileSystemCollectionReader: preparing raw data to data format usable for AE
> 2. Analysis Engine: scnlp-client which calls remote service & local hw2-aae(incorporate the result of scnlp (NameEntityMention used))
> 3. CasConsumer: collect the Cas data generated by AE and output in a certain format, (for here, stdout is used.)

The incorporation of NameEntityMention in AnswerScoreAnnotator:
1. Collect all the string annotated by NameEntityMention
2. For each answer, count the number of NameEntityMention also found in question and total number of NameEntityMention
3. Integrate the information above with NGram analysis, calculate score.

Compare Result:
Speed: hw2 pipeline is much faster (113ms vs 928ms)
Accuracy: both are the same (0.5 for input 1, 0.67 for input 2).
> The reason might be only NameEntityMention is used. In our example, names appear similarly in each answer. Thus, the incorporation of this information won't change the result a lot.

```
100% (928ms) – Collection Processing Engine
    11.64% (108ms) – File System Collection Reader (Process)
    77.8% (722ms) – scnlp-yingshen-client.xml (Service Call)
    4.96% (46ms) – scnlp-yingshen-client.xml (End of Batch)
▶   3.99% (37ms) – hw2-yingshen-aae (Analysis)
    0% (0ms) – hw2-yingshen-aae (End of Batch)
    1.62% (15ms) – evaluatorCasConsumerDescriptor (Analysis)
    0% (0ms) – evaluatorCasConsumerDescriptor (End of Batch)
```

```
100% (113ms) – Collection Processing Engine
    58.41% (66ms) – File System Collection Reader (Process)
▶   30.97% (35ms) – hw2-yingshen-aae (Analysis)
    0% (0ms) – hw2-yingshen-aae (End of Batch)
    10.62% (12ms) – evaluatorCasConsumerDescriptor (Analysis)
    0% (0ms) – evaluatorCasConsumerDescriptor (End of Batch)
```

Task 2.3:
A broker is started and hw2-aae is deployed as service in the broker.

The design here used hw2-aae-client to call the hw2-aae service as remote service:
> 1. FileSystemCollectionReader: preparing raw data to data format usable for AE
> 2. Analysis Engine: hw2-aae-client
> 3. CasConsumer: collect the Cas data generated by AE and output in a certain format, (for here, stdout is used.)