

11791 Design and Engineering of Intelligent Information System

Jung In Lee

Andrew ID: junginl

HW3. Execution Architecture with CPE and Deployment Architecture with UIMA-AS

Task 1. Execution Architecture with CPE

As covered in HW2, the UIMA Analysis Engines operate on one document, but we may want to deal with collections of data instead of just one. This is where Collection Processing Engine (CPE) comes into play. The Collection Processing Architecture allows the system to run on collections of unstructured raw data.

The components of CPE are:

- Collection Reader:
Read in a collection of data and process it to be analyzed by Analysis Engines
- Analysis Engines:
Take a CAS (initialized by Cas Initializer) and analyze its contents. Produces an enriched CAS.
- Cas Consumer:
Consume the enriched CAS produced by Analysis Engines.

Outline of the pipeline:

Collections of data → Collection Reader → Cas Initializer → Analysis Engine → Cas Consumer

For this particular task, we have:

FileSystemCollectionReader → Aggregate Analysis Engine → Cas Consumer
 QuestionAnnotator
 AnswerAnnotator
 TokenAnnotator
 NgramAnnotator
 AnswerScoreAnnotator
 EvaluationAnnotator

I copied and pasted the FileSystemCollectionReader and XmiWriterCasConsumer from the uimaj-examples to create the CPE. Then I tried to modify the EvaluationAnnotator to make it into a Cas Consumer.

Task 2. Deployment Architecture with UIMA-AS

This task involves using UIMA Asynchronous Scaleout (UIMA-AS) to incorporate remote NLP into our pipeline. The second part of this task is deploying the pipeline created in HW2 as a remote service. The only difference is the insertion of the UIMA-As Remote Client part in the previous pipeline.

The new pipeline looks like this:

FileSystemCollectionReader → UIMA-AS Remote Client → AAE → CAS Consumer

Where AAE involves:

QuestionAnnotator

AnswerAnnotator

TokenAnnotator

NgramAnnotator

AnswerScoreAnnotator

EvaluationAnnotator