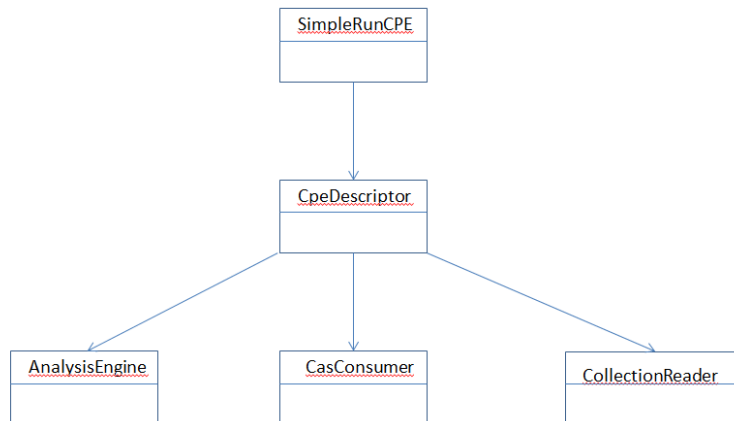


# Software Engineering for Intelligent Information Systems HW1 Report

Tan Li (tanl)

## 1. UML



Here is a rough UML class diagram for the project. `SimpleRunCPE` serves as the entry and it uses `CpeDescriptor` to map to three other descriptors, analysis engine, cas consumer and collection reader. Then we can use these descriptors to find the java classes.

## 2. Type System

I created the gene type and it has a parameter called Gene. I also used the provided `StanfordCoreNLP` package and the `LingPipe` library.

## 3. General Data Flow

The `SimpleRunCPE.java` serves as the entry of the application. We already set the `CpeDescriptor.xml` as the argument. Then it will look into the descriptor to locate information including collection reader, analysis engine and cas consumer. Then it will

go to these descriptors to find the input and output information and annotator information. Then it will find the java class to process the input and return the output.

#### 4. Other thoughts

I have never studied in the NLP area before. So at first, when I was exposed to UIMA, I'm totally confused what it is about. But after learning about UIMA and try to go through the tutorial, I became clear about the framework. Even though this might not be my field of research in the future, it is also very interesting to know how to find a target text in the document and make annotations. I think the next project will be more challenging and I wish to learn more in the project.