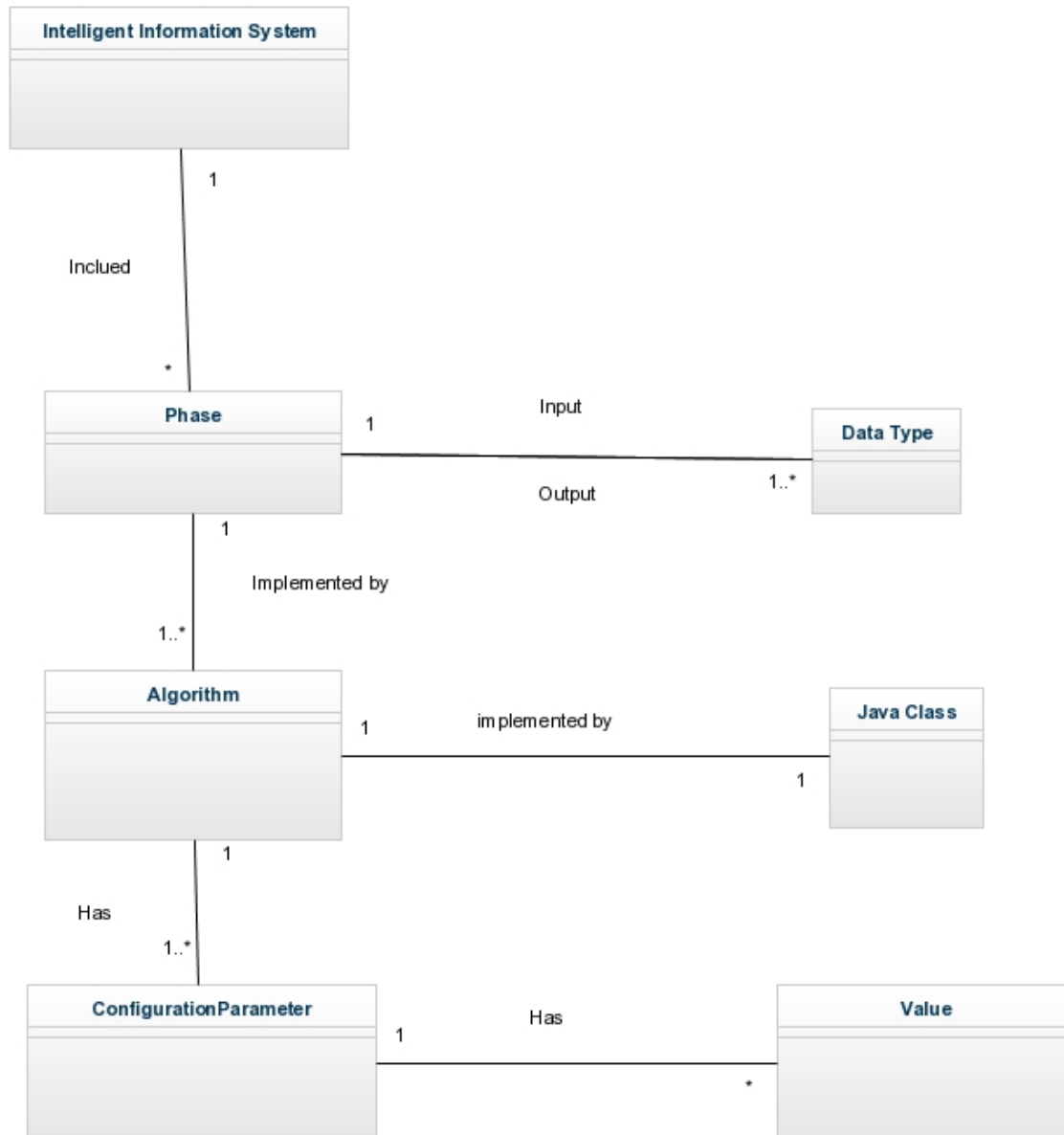


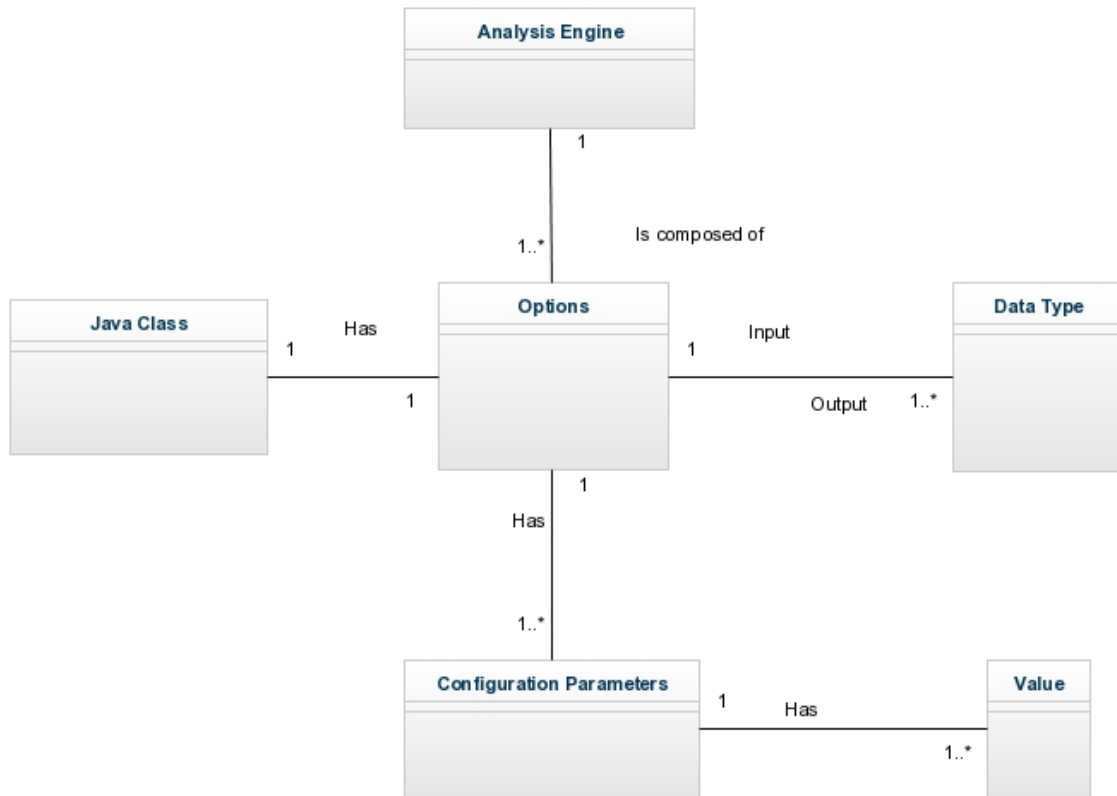
# Report

Hw1-Chongshm

## Task 1.1 Domain diagram for Intelligent Information System



## Task 1.2 Domain diagram for Analysis Engine



In this project, I implemented the **Singleton design pattern**.

1. Please identify/describe any machine learning techniques used?

**Sorry, I did not use any machine learning techniques.**

2. Please identify/describe any NLP techniques/components used?

**I used the Stanford-nlp.jar. And also I used the Lingpipe.jar.**

3. Please identify/describe any external (marked up text) training data used?

**I took the ne-en-bio-genetag.HmmChunker file as the training data.**

4. Please identify/describe any external lexical resources(terminology lists)used?

**The Stanford-nlp provided the terminology list to distinguish the none. But, the result is not able to reach the satisfactory level.**

5. Please describe any rule sets used?

**I just referred to the Lingpipe 4.1.0, and follow the dictionary that it provided.**

6. If your system interacts with or uses data from any biological database(s), please describe?

**The ne-en-bio-genetag.HmmChunker file could be seen as the biological database. In this database, we could use chunk function from lingpipe API to extract the Gene name.**

7. Please identify/describe any other relevant resources used to train/develop your system?

**When I find the Gene name tag extracted from Stanford-nlp cannot meet the demand. Then, I find the Lingpipe 4.1.0 from the internet, by using its API, I get more accurate results.**

8. Please describe the general data flow in your system?

**By using the filereader class, the system could read the data from the "hw1.in". then save the sentence in the "hw1.in" line by line. Then using the annotator to process each gene tag and save it with**

**the consumer class. Then, the annotator will process another potential gene data. Until finished, I could get a hw1-chongshm.out.**

9. Other information of interest?

**I finished the FileCollectionReader line by line. And compared the performance of Lingpipe NER and Standford-nlp NER, the precision of using Stanford-nlp API was just nearly 10%. However, the precision of Lingpipe could reach the 80% precision level. If there exists a more comprehensive biological dictionary, we may get more precise result.**