

# 11-791 Design and Engineering of Intelligent Information System Fall 2012

## Homework 1

### *UML Design and Name Entity Recognition Implementation with UIMA SDK*

#### Important dates

- **Hand out: October 1.**<sup>a</sup>
- **Turn in: October 15.** Besides all your Java source codes, and UIMA type system, collection reader and analysis engine descriptors, which are essential to your UIMA collection processing engine, you are also required to answer a set of UML design related problems, write a report for your implementation of the name entity recognizer, and include the Javadocs generated from your source codes.

If you have ever looked into your `target/` directory or Maven course-release repository for Homework 0, then you will find that Maven will package the binary files, source files, and Javadocs into different jars, and deploy them on the server, which means all you need to do is to put all the source codes and descriptors, as well as the documents, in the right place of your hw1-ID project. You should organize your project in the same hierarchy as shown below<sup>b</sup>:

```
hw1-ID
|- pom.xml
\-- src
    \-- main
        \-- java
            | \-- **/*.java    // your Java codes go into this folder
            |                  // as you did before
        \-- resources
            | \-- **/*.xmi     // all your descriptors go into the
            |                  // resources folder
        \-- docs
            | \-- hw1-ID-uml.pdf    // your UML design writeup
            \-- hw1-ID-report.pdf // your report for the pipeline
```

<sup>a</sup>This version was built on September 28, 2012

<sup>b</sup>To simplify your submission process and our evaluation process, we ask you to create `src/main/resources/docs` for your documents. But you should keep in mind it is NOT a good practice to have documentation within a `src` folder.

Several notes about organizing your Maven project and other additional information:

1. **Submission:** The same way as you did for Homework 0 (set up GitHub repo, create Maven project, write your code, submit to Maven repo!), except that the name has changed to hw1-ID from hw0-ID.
2. **Your source files and descriptors:** `**/*.java` and `**/*.xmi` tell you that you don't need to flatten your folder hierarchy, instead we encourage you to place your Java codes in the right package, and similarly, you can create subfolders for different types of descriptors, e.g., `src/main/resources/descriptors/ner` for all the analysis engine descriptors name related entity recognition task. We will look into your jar package.
3. **UML design answers and report:** We will pull out your documents from your jar files, and Prof. Eric Nyberg will look into your report and answers for the questions, so that remember to include your ID as part of file names, and we highly encourage you to submit only PDF documents.
4. **Javadocs:** Please refer to any best practise (yes, there might be more than one) to write your Javadoc, e.g., <http://www.oracle.com/technetwork/java/javase/documentation/index-137868.html>. We expect you to describe your major components at the class level of your Javadoc, and put additional comments on the most important methods if any. It might be a good idea to include your UML design in your Javadoc if you think the developers can have a general idea by just looking at your diagram.
5. **Performance evaluation:** We believe you may have different ideas about how to design the type system, and how to implement a name entity recognizer, which means we won't have a shared type system to test your components (but we will have one in Homework 2). Therefore, you are required to write a collection reader to load the input text and a CAS consumer to generate your output (both are independent of your specific type system), and our evaluation will be based on the output from your CAS consumer, similar as we did for Homework 0.

### Useful information

1. We encourage you to start Homework 1 as early as possible, especially if you haven't done the reading assignments for UIMA Specification, SDK, or even haven't had a chance to browse the UIMA portal.
2. For any question regarding the UML design part of Homework 1, please send mails directly to Prof. Eric Nyberg ([ehn@cs.cmu.edu](mailto:ehn@cs.cmu.edu)), and for other questions, please send us mails: Zi Yang ([ziy@cs.cmu.edu](mailto:ziy@cs.cmu.edu)) or Rui Liu ([ruil@cs.cmu.edu](mailto:ruil@cs.cmu.edu)).
3. Again, both source files and derived pdf file of this assignment are publicly available on my GitHub

<http://github.com/ziy/software-engineering-preliminary>

Please feel free to fork the project and send a pull request back to me as some of you did for Homework 0 for any error. Or you can just report an issue at

<http://github.com/ziy/software-engineering-preliminary/issues>