

ISMLA 17/18: UIMA Exercises

1 Project Setup

Set up a Maven/Java Eclipse project with the following dependencies:

group id	artifact id	version
org.apache.uima	uimaj-core	2.10.1
org.apache.uima	uimaj-tools	2.10.1
org.apache.uima	uimaj-document-annotation	2.10.1
org.apache.opennlp	opennlp-tools	1.8.2

Add a `src/main/resources` directory and add it to the project build path.

2 Type System Setup

1. Add a type system descriptor in `src/main/resources`.
2. Add type `Sentence` and `Token`. For the `Token` type add a lemma and part of speech feature. The name space should be the group id plus “types”.
3. Generate Java code from the XML type system.

3 Analysis Engine Setup and Implementation

1. Write an `AnalysisEngine` that annotates `Sentence` annotations (see exercise 2) and wraps around the OpenNLP sentence detection tool. Load the model in an overridden `initialize` method and run the `sentPosDetect` method to obtain the spans serving as a indices for your annotations. Use <https://opennlp.apache.org/docs/1.8.2/manual/opennlp.html#tools.sentdetect> as reference. The path to the model should be parametrized via a `ConfigurationParameter`.
2. Test your component with two texts by running your analysis engine with the `DocumentAnalyzer` on the following data:
`https://www.gutenberg.org/files/1342/1342-0.txt`
`https://www.gutenberg.org/files/11/11-0.txt`