

ISMLA 17/18: DKPro Exercises

1. Download the DKPro example projects from <https://github.com/dkpro/dkpro-core-examples>, unpack the archive, and import the projects via *File* → *Import* → *Maven* → *Existing Maven Project*.
2. Explore the *tokenizedwriter-asl* project. Read up on the uimaFIT components referenced by the project. Run the main class `TokenizedWriterPipeline` and inspect the output.
3. Based on the existing pipeline add a new class in this project that provides the following functionality by assembling DKPro components:
 - reading all files in a directory
 - sentence detection
 - tokenization
 - lemmatization
 - POS Tagging
 - Named Entity Recognition
 - writing the output as XMI files

You will need to add further DKPro core dependencies from the central Maven repository (<https://mvnrepository.com/artifact/de.tudarmstadt.ukp.dkpro.core>).

4. Inspect the XMI output of your pipeline with the AnnotationViewer (like for DocumentAnalyzer).
5. If time permits:
 - add more DKPro NLP components to your pipeline
 - port the simple whitespace tokenizer from the beginning of the class to be DKPro compatible