

Programming Lab III, WS 2016/2017

Handout 6

Dr. Marc Zimmermann, Jens Dörpinghaus
2016-11-22

Sixth task – Extending your existing application building on the previous reader, annotator and writer.

Has to be completed finally on 2016-12-13, 23:59pm.

The next task will be to add new functionality, which makes use of the already developed reader, annotator and writer. This time we want to create alignments of all genes/proteins mentioned in PubMed abstracts. Here's the list of topics you have to complete:

- Download the XMIs generated by SCAIView and open them in the CAS Editor (<https://tor-2.scai.fraunhofer.de/gf/project/pl3/docman/tasks/task06/>).
 - Which Version of the TypeSystem and Annotator has been used? Is it compatible with your version? If it is not compatible – modify the *pom* accordingly.
 - In the *DocumentView* you find annotations made by *ProMiner*. Which entities have been tagged?
 - In the *header* you find *meta information* on the article – which information is stored? In which view is the header stored?
 - Prepare a **screenshot** of the XML and highlight and comment the corresponding items in it.
- Use your *already developed* CollectionReader, Annotator and Consumer to extract all genes and proteins from the article stored in the XMI and create pairwise alignments for all occurrences of these in each article. The alignments should be written out as HTML (i.e. produce one html document per input XMI).
 - Extract all annotated genes and fetch the corresponding sequences.
 - Update and extend your existing modules. Comment and **commit all changes**.
 - Each HTML document should contain and print the meta information from input XMI. Commit the **resulting HTML files**. Hint: use `getAnnotationIndex(Header.type`
`)`
- Should be unnecessary to mention: implement the *needed* JUnit tests.

Good luck and have fun!