



cTAKES 4.0 - Clinical Documents Pipeline

 [cTAKES 4.0 - Chunker](#)

 [cTAKES 4.0 Component Use Guide](#)

[cTAKES 4.0 - Constituency Parser](#) 

Contents of this Page

- [Overview of Clinical Documents Pipeline](#)
- [Analysis engines \(annotators\)](#)
 - [AggregateCdaProcessor.xml](#) for CDA documents conforming to the provided DTD
 - [AggregatePlaintextProcessor.xml](#) for plain text documents

cTAKES 4.0 Links

[Apache cTAKES download site](#)

Documentation:

- [cTAKES 4.0](#)
- [cTAKES 4.0.0.1](#)
- [cTAKES 4.0 User Install Guide](#)
- [cTAKES 4.0 Developer Install Guide](#)
- [cTAKES 4.0 Component Use Guide](#)
- [cTAKES 4.0 Dictionaries and Models](#)
- [Documentation Conventions](#)

Overview of Clinical Documents Pipeline

This project is the top-level, main project for processing a clinical document through the entire cTAKES pipeline, including sentence detection, part of speech tagging POS, chunking, named entity recognition, context detection, and negation detection.

The pipeline can process two types of documents

- plain text files
- Clinical Document Architecture (CDA) XML files that conform to the DTD provided

Analysis engines (annotators)

AggregateCdaProcessor.xml for CDA documents conforming to the provided DTD

The file `desc/analysis_engine/AggregateCdaProcessor.xml` is the aggregate analysis engine to use to run the entire pipeline, including the `CdaCasInitializer` analysis engine, which reads CDA documents that conform to the DTD provided, and create Segment annotations based on the sections within the CDA document.

Parameters

`ChunkerCreatorClass` - the full class name of an implementation of the interface `edu.mayo.bmi.uima.chunker.ChunkerCreator`

AggregatePlaintextProcessor.xml for plain text documents

The file `desc/analysis_engine/AggregatePlaintextProcessor.xml` is the aggregate analysis engine to use to run the entire pipeline, including the `SimpleSegmentAnnotator` analysis engine, which creates a Segment annotation that wraps the entire plain text document. Other annotators in the pipeline require at least one Segment annotation.

Parameters

`SegmentID` - the identifier or name to assign to the Segment annotation

`ChunkerCreatorClass` - the full class name of an implementation of the interface `edu.mayo.bmi.uima.chunker.ChunkerCreator`



The `ChunkCreatorClass` parameter of both annotators is set to `edu.mayo.bmi.uima.chunker.PhraseTypeChunkCreator` so that each phrase type gets its own type of annotation, rather than having all chunks be of type `Chunk`.