cTAKES 3.0 Component Use Guide

Contents of this Page

cTAKES 3.0 Links

Apache cTAKES web site
Documentation:

- cTAKES 3.0
- cTAKES 3.0 User Install Guide
- cTAKES 3.0 Developer Install Guide
- cTAKES 3.0 Component Use Guide
- Documentation Conventions

Component Overview

cTAKES consists of a number of components. Each one has unique qualities and capabilities. Each component includes at least one analysis engine (annotator), some include more. You will want to assess each component's usefulness to you. UIMA provides the tooling for selecting which annotators are used together and the order in which annotators are run. Each section in this Guide covers one component.

cTAKES provides two variants of the original cTAKES pipeline which discovers Named Entities and assigns attributes to them:

- for processing plain text notes: cTAKESdesc/cdpdesc/analysis_engine/AggregatePlaintextProcessor.xml
- for processing Clinical Document Architecture (CDA) formatted notes: cTAKESdesc/cdpdesc/analysis_engine/AggregateCdaProcessor.xml
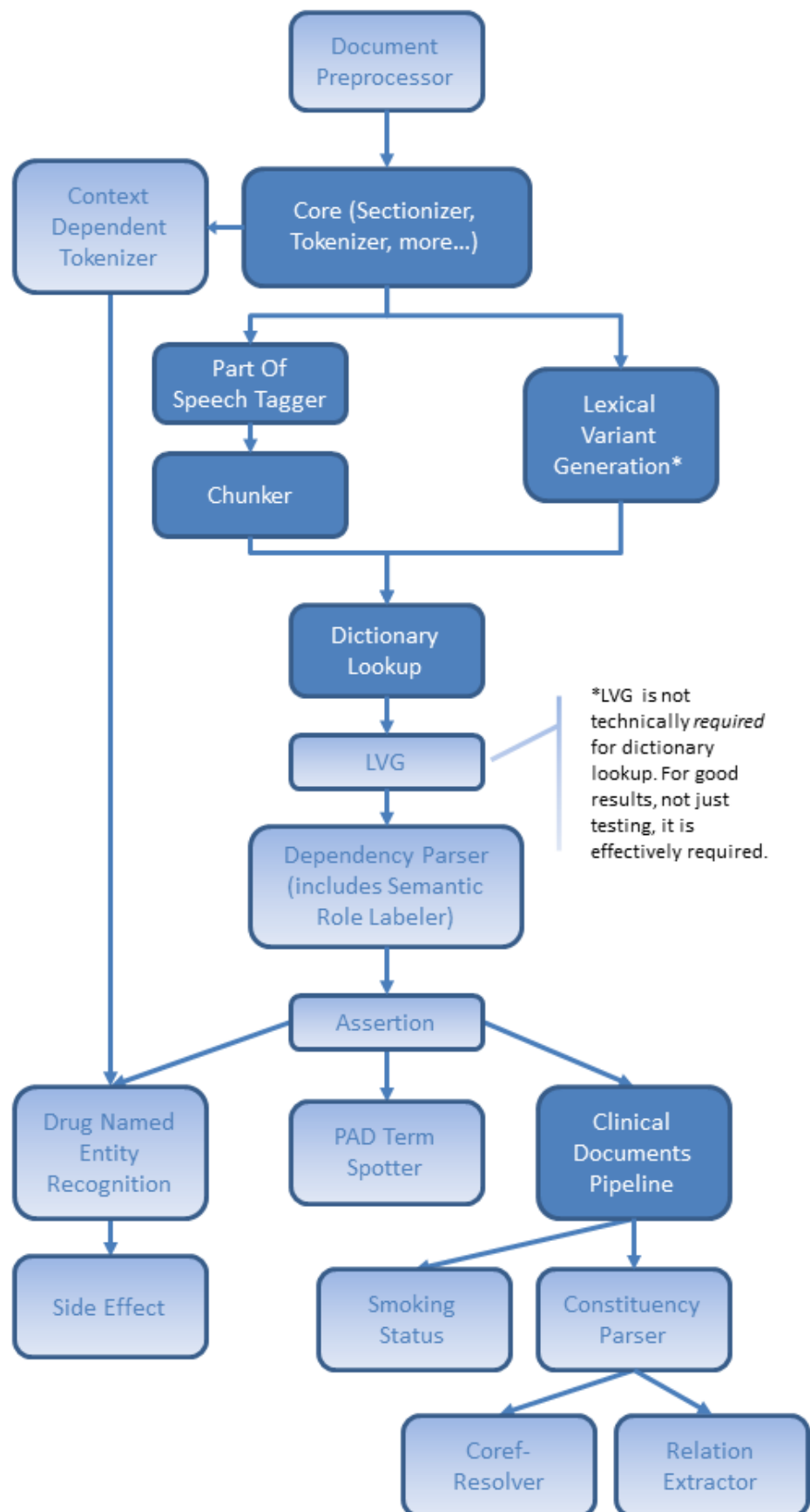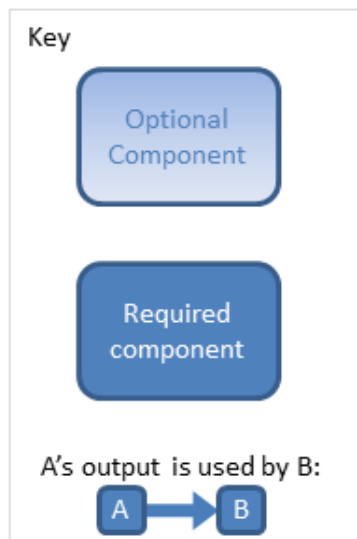
Both variants use the same set of components except that the Document Preprocessor is not used for plain text.

⚠ cTAKES is not designed to be thread safe and has not been tested for thread safety.

Component Dependencies

This diagram shows which components rely on the output of another component. Following the diagram is a textual description.

cTAKES
Component
Dependencies

Document Preprocessor

Context Dependent Tokenizer

Core (Sectionizer, Tokenizer, more...)

Part Of Speech Tagger

Lexical Variant Generation*

Chunker

Dictionary Lookup

Key

Optional Component

Required component

A's output is used by B:

A → B

LVG

*LVG is not technically *required* for dictionary lookup. For good results, not just testing, it is effectively required.

Dependency Parser (includes Semantic Role Labeler)

Assertion

Drug Named Entity Recognition

PAD Term Spotter

Clinical Documents Pipeline

Side Effect

Smoking Status

Constituency Parser

Coref-Resolver

Relation Extractor

- If the input is a CDA document, the Document Preprocessor is needed at the start of the pipeline, and its output is used by Core.
- The output of Core is used by several components, including
  - Context Dependent Tokenizer
  - Part of Speech Tagger
  - LVG
- The output of the Part of Speech Tagger is used by the Chunker
- The outputs of the Chunker and of LVG are used by Dictionary Lookup
  - LVG is not strictly required by the Dictionary Lookup but better results are achieved if LVG is used.
- The output of Dictionary Lookup can be used without using LVG, the Semantic Role Labeler (which is part of the Dependency Parser) or the Assertion component, depending on which attributes are of interest.
- The output of Dictionary Lookup is typically used by LVG, the Semantic Role Labeler (which is part of the Dependency Parser) and the Assertion component.
  Note that prior to cTAKES 2.5, the output of Dictionary Lookup was used by NE Contexts instead.
- Depending upon which pipeline was used, the output of the Assertion annotator (or the Dictionary Lookup directly) is then used by one of the following
  - PAD Term Spotter
  - Clinical Documents Pipeline
  - Drug NER
- If the Drug NER pipeline was used, the output of the Context Dependent Tokenizer is used by the Drug NER component.
- If the Side Effect pipeline was used, the output of Drug NER is used by the Side Effect component
- If the Constituency Parser or Coref-resolver pipeline was used, the output of Clinical Documents Pipeline is used by the Constituency Parser
- If the Coref-resolver pipeline was used, the output of the Constituency Parser is used by the Co-ref resolver.
- If the Smoking Status pipeline was used, the output of the Clinical Documents Pipeline is used by the Smoking Status component.

Components

- cTAKES 3.0 - Assertion
- cTAKES 3.0 - Chunk Adjuster
- cTAKES 3.0 - Chunker
- cTAKES 3.0 - Clinical Documents Pipeline
- cTAKES 3.0 - Constituency Parser
- cTAKES 3.0 - Context Dependent Tokenizer
- cTAKES 3.0 - Core
- cTAKES 3.0 - Dependency Parser and Semantic Role Labeler
- cTAKES 3.0 - Dictionary Lookup
- cTAKES 3.0 - Document Preprocessor
- cTAKES 3.0 - Drug Named Entity Recognition
- cTAKES 3.0 - LVG
- cTAKES 3.0 - NE Contexts
- cTAKES 3.0 - PAD Term Spotter
- cTAKES 3.0 - POS Tagger
- cTAKES 3.0 - Relation Extractor
- cTAKES 3.0 - Side Effect
- cTAKES 3.0 - Smoking status