

# cTAKES 4.0 Component Use Guide

## Contents of this Page

- [Component Overview](#)
- [Component Documentation](#)
- [Component Dependencies](#)

## 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](#)

## 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 was not originally designed to be thread safe.

If you would like to experiment with making it thread safe, see class `ThreadSafeLvg` in `ctakes-lvg`.

## Component Documentation

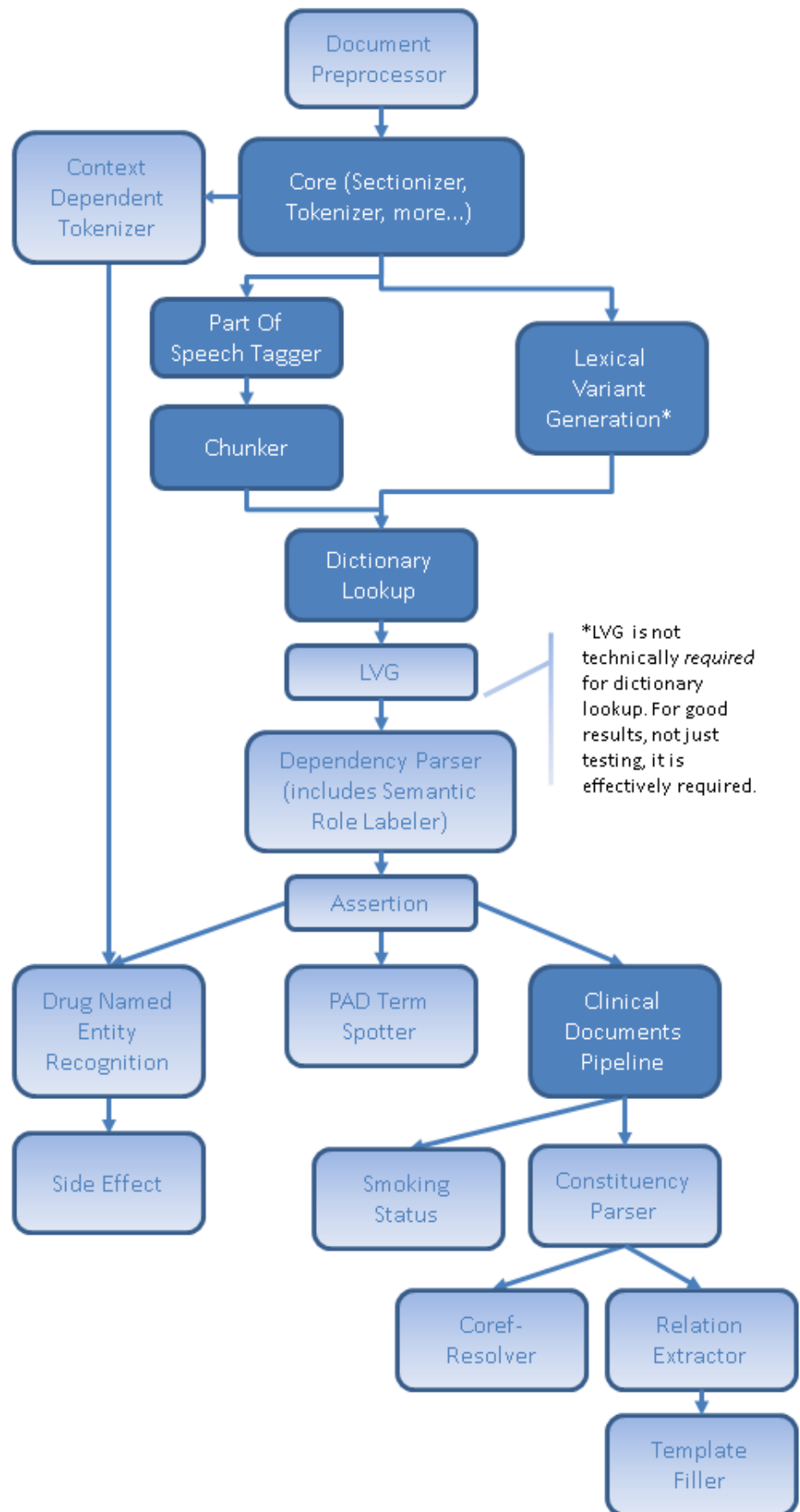
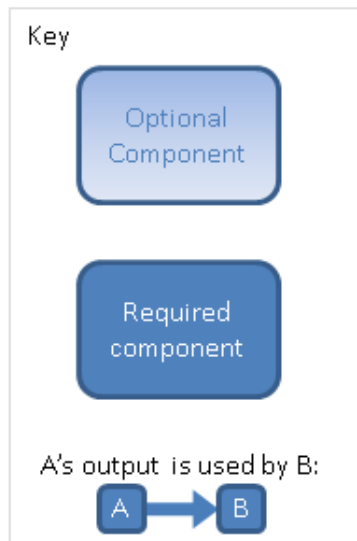
- [cTAKES 4.0 - Assertion](#) (named entity attributes: negation, aka polarity; history; subject; and more)
- [cTAKES 4.0 - Chunk Adjuster](#)
- [cTAKES 4.0 - Chunker](#)
- [cTAKES 4.0 - Clinical Documents Pipeline](#)
- [cTAKES 4.0 - Constituency Parser](#)
- [cTAKES 4.0 - Context Dependent Tokenizer](#)
- [cTAKES 4.0 - Core](#) (sentence detectors, tokenizers, and more)
- [cTAKES 4.0 - Coreference](#)
- [cTAKES 4.0 - Dependency Parser and Semantic Role Labeler](#)
- [cTAKES 4.0 - Document Preprocessor](#) (for CDA documents, not needed for plaintext input)
- [cTAKES 4.0 - Drug Named Entity Recognition](#)
- [cTAKES 4.0 - Fast Dictionary Lookup](#)
- [cTAKES 4.0 - GUI](#) (dictionary creator; pipeline fabricator/runner)
- [cTAKES 4.0 - LVG](#)
- [cTAKES 4.0 - NE Contexts](#) (named entity attributes including negation)
- [cTAKES 4.0 - Negation Annotators](#)
- [cTAKES 4.0 - Original Dictionary Lookup](#) (consider [Fast Dictionary Lookup](#) instead)
- [cTAKES 4.0 - POS Tagger](#)
- [cTAKES 4.0 - Relation Extractor](#)
- [cTAKES 4.0 - Semantic Similarity](#)
- [cTAKES 4.0 - Sense Disambiguator Annotator](#)
- [cTAKES 4.0 - Side Effect](#)
- [cTAKES 4.0 - Smoking status](#)
- [cTAKES 4.0 - Template Filler](#)
- [cTAKES 4.0 - Temporal Module](#)

- [cTAKES 4.0 - YTEX DBCollectionReader](#)
- [cTAKES 4.0 - YTEX DBConsumer](#)
- [cTAKES 4.0 - YTEX SentenceAnnotator](#)

## Component Dependencies

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

# Apache cTAKES Component Dependencies



- 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
    - 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.
  - If the Template Filler pipeline was used, the output of the Relation Extractor is used by the Template Filler.