

CLAMP (Clinical Language Annotation, Modeling and Processing toolkit) System, Version 0.10.17, 09/25/2015

CONTENTS OF THIS FILE

-
- * Introduction
- * System Requirements
- * Installation
- * Package description
- * Maintainers

Introduction

The CLAMP System is a comprehensive clinical Natural Language Processing software that enables automatic encoding of clinical information in narrative patient reports. In addition to running a clinical concept extraction pipeline as well as an annotation pipeline, the individual components of the system can also be used as independent modules. The system lends itself for diverse applications in a broad range of clinical domains.

Requirements

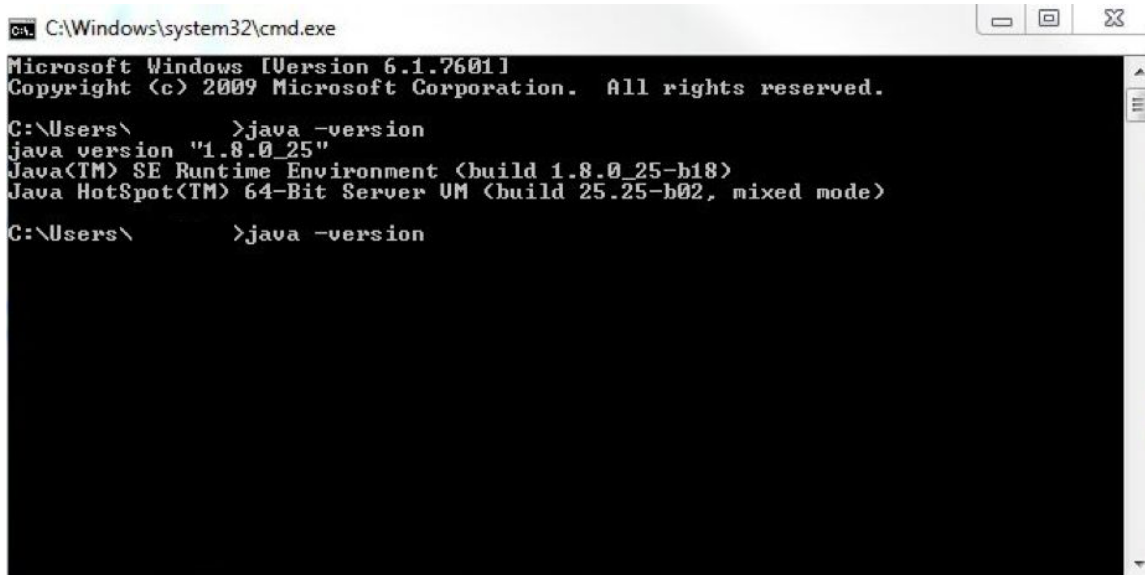
CLAMP is a stand-alone Java application based on Eclipse platform technologies. CLAMP uses the Apache UIMA (Unstructured Information Management Architecture) framework. The annotation module of CLAMP incorporates the brat rapid annotation tool . For the other individual constituents, Apache OpenNLP toolkit, Liblinear and CRF Suite are utilized in addition to in-house rule-based components. CLAMP also use the UIMA Ruta (Rule based Text Annotation) as a rule engine to help users specify rules.

CLAMP is distributed as a ready-to-use binary package that can either be executed at the command line or carries the associated Graphic User Interface (GUI). Our distribution package includes components for jar files, CRFSuite, and a Lucene index of all the restriction level 0 & level 1 UMLS concepts.

The only prerequisite necessary to compile CLAMP is JRE 1.8 (Java Runtime Environment). Please ensure that you have Java 8 or higher installed in your system. Run the following command in both Mac and Windows to check your version:

```
java -version
```

Here is an example of what you will see when running the command in Windows:

A screenshot of a Windows command prompt window. The title bar shows the path 'C:\Windows\system32\cmd.exe'. The window content displays the following text:

```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\ >java -version
java version "1.8.0_25"
Java(TM) SE Runtime Environment (build 1.8.0_25-b18)
Java HotSpot(TM) 64-Bit Server VM (build 25.25-b02, mixed mode)

C:\Users\ >java -version
```

If your java version is not 1.8, it is available for download from the Oracle website at <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>.

An UMLS account is required in order to use Level 2 and higher data in the UMLS encoding component of the system. The account can be created at <https://uts.nlm.nih.gov/home.html>. The CLAMP system comes with the Level 0 and Level 1 data by default.

CLAMP also uses the computer's default browsers to visualize the clinical documents. Since all browsers do not completely support all the aspects of the technologies used to implement the visualization, limitations exist in term of running the CLAMP annotation module in the browsers. On the Windows OS, the Internet Explorer should be higher than IE9; On Macintosh computers, Safari (all versions) works well with CLAMP.

Installation

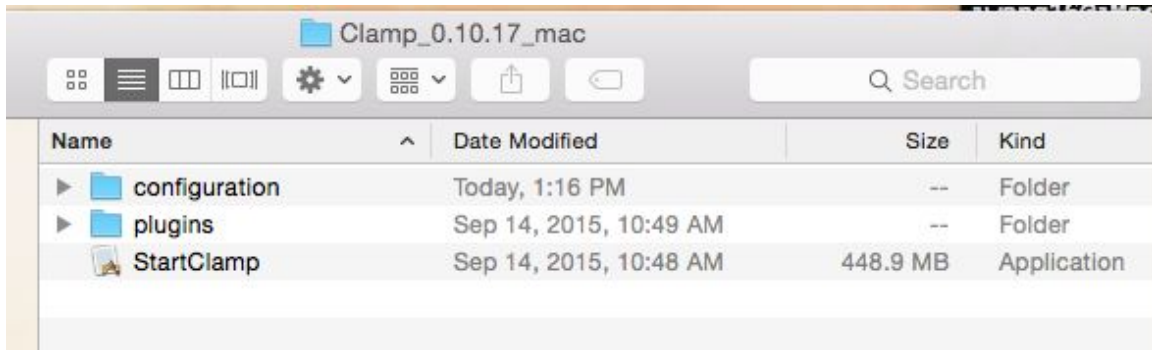
The CLAMP System is provided as a .zip file. After downloading the compressed file, unzip the package in the directory of choice and voila!! the system is ready for use. You can run the system by clicking on the startCLAMP icon to launch the GUI of CLAMP. Installation instructions are the same for both Windows and Macintosh computers. For the CLAMP command line version please refer to the readme file.

Package description

Since CLAMP is a stand-alone eclipse plugin, its folder structure is similar to other eclipse plugins.

Configuration Folder: This folder contains CLAMP configuration files.

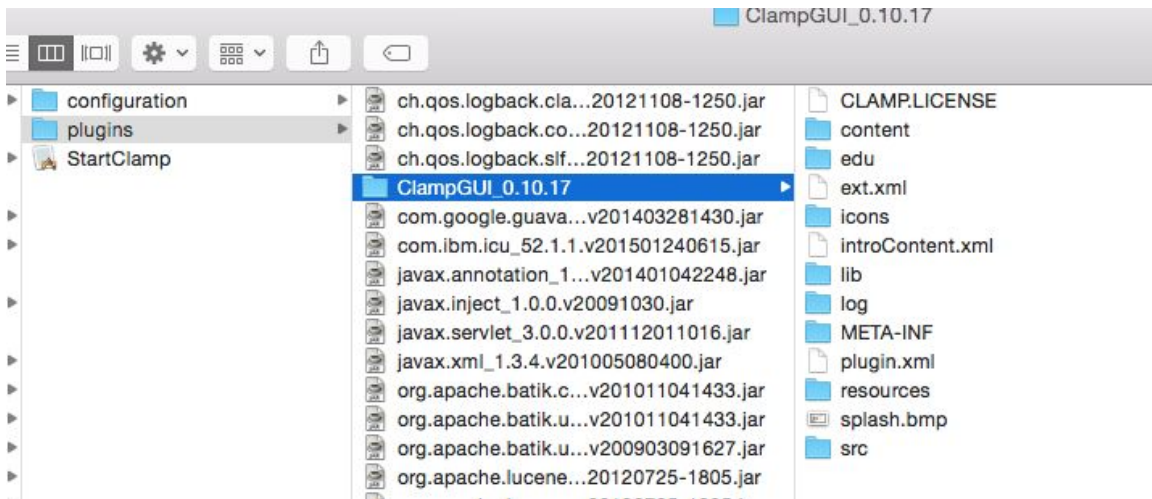
StartClamp: This is the launching point for the CLAMP GUI. In Windows, this is an executable file while in Mac, this is an application.



Plugins Folder: This folder contains the following items:

1. All the plugins that are used by CLAMP GUI such as org.eclipse.wst, org.eclipse.ui, etc.
2. The CLAMP dependencies files
3. .jar files that are Eclipse generated files
4. A folder that called “ClampGUI_0.10.17”

Now, we will describe the content of ClampGUI_0.10.17 in detail.



ClampGUI_0.10.17 folder:

CLAMP.LICENSE: is the CLAMP license file. The contents of the license file are tailored to reflect the kind of license that you have executed for CLAMP.

ext.xml, Content: Contains the help documents for CLAMP;

edu: Contains the compiled java class file, and the resources that are used internally such as dictionaries and the machine learning model files.

icons: Contains the icons used in CLAMP;

log: Includes CLAMP run-time log files;

resources: This folder includes 3rd-party libraries. Currently it has two items:

1. CRFSuite: the CRF implementation for Name Entity Recognition task;
2. Umls_index: the Lucene index built for CLAMP based on the UMLS thesaurus. If you want to use UMLS terminologies, then you will need to create an UMLS account. Please follow the following link to create an UMLS account if you do not have any.

<https://uts.nlm.nih.gov/license.html>

META-INF, and plugin.xml: The content of META-INF along with the plugin.xml files represent the config files for CLAMP.

Lib: Contains the 3rd-party libraries used in CLAMP. The following table enlists libraries included in this folder:

groupid	artifactId	version
org.ini4j	ini4j	0.5.2
com.google.code.gson	gson	2.3
org.apache.uima	uimafit-core	2.1.0
org.apache.uima	uimaj-core	2.6.0
org.apache.uima	uimaj-cpe	2.6.0
org.apache.uima	uimaj-document-annotation	2.6.0
de.bwaldvogel	liblinear	1.94
org.apache.lucene	lucene-core	5.2.1
org.apache.lucene	lucene-analyzers-common	5.2.1
org.apache.lucene	lucene-queryparser	5.2.1
org.apache.opennlp	opennlp-tools	1.5.1-incubating
org.apache.ctakes	ctakes-type-system	3.2.0
org.cleartk	cleartk-named-entity	0.6.6
org.cleartk	cleartk-ml	2.0.0
com.googlecode.clearnlp	clearnlp	1.3.1
org.apache.ctakes	ctakes-assertion	3.2.1
org.apache.ctakes	ctakes-assertion-res	3.2.1
commons-codec	commons-codec	20041127.09
dom4j	dom4j	1.6.1

org.cleartk	cleartk-ml-liblinear	2.0.0
org.apache.ctakes	ctakes-relation-extractor	3.2.2
org.apache.ctakes	ctakes-dependency-parser	3.2.2
commons-cli	commons-cli	1.3.1
org.apache.uima	ruta-ep-engine	2.3.0

Contact Us

The CLAMP System was developed by the Xu group from the School of Biomedical Informatics at the University of Texas Health Science Center in Houston.

For Technical Issues:

Please contact: Jingqi.Wang@uth.tmc.edu

For any other issues:

Please contact: Anupama.E.Gururaj@uth.tmc.edu