CLAMP(Clinical Language Annotation, Modeling, and Processing Toolkit)
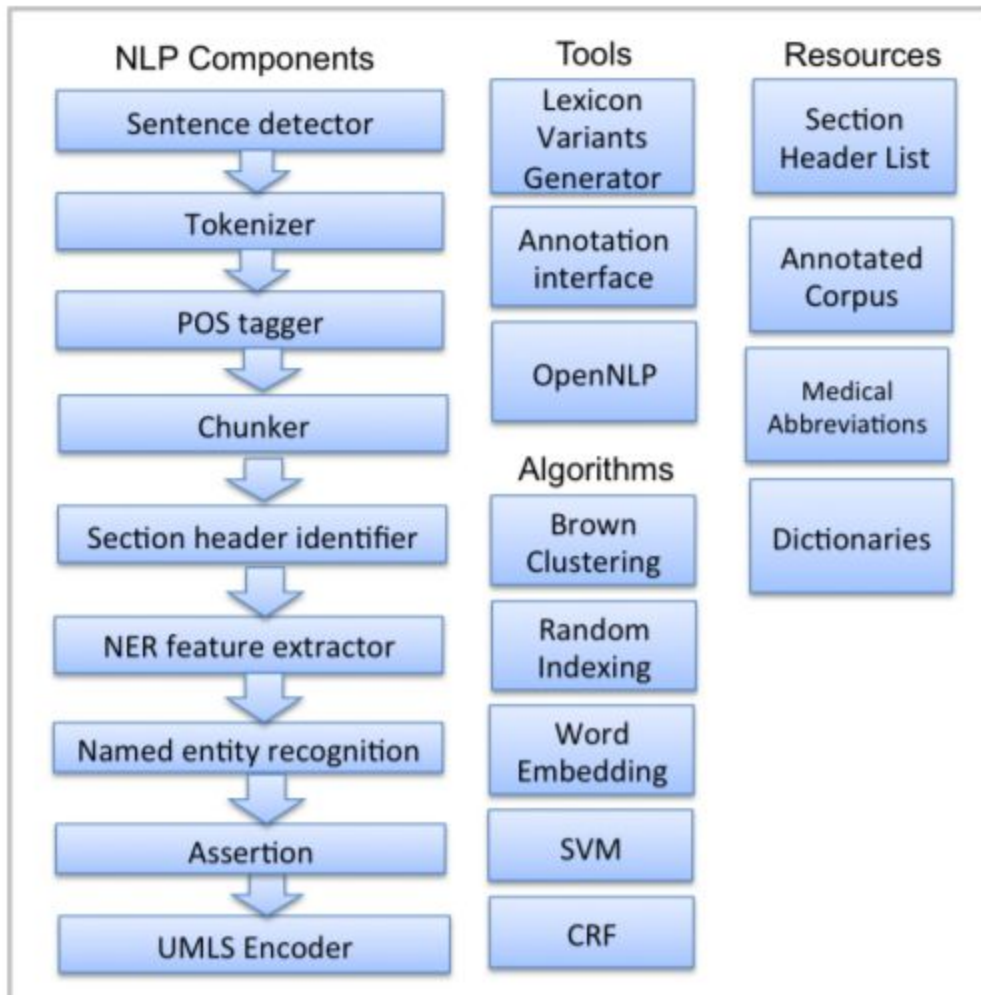
Introduction:

In the healthcare industry, natural language processing has numerous potential applications. Natural language processing can improve the precision of electronic health records(EHRs) by making an interpretation of narrative content into normalized information. This is incredibly important and can be used for clinical decision making.

What is CLAMP?

CLAMP is a comprehensive clinical Natural Language Processing (NLP) software that enables recognition and automatic encoding of clinical information in narrative patient reports. It is based on several award-winning methods and applications developed at the University of Texas Health Science Center at Houston, School of Biomedical Informatics. CLAMP is a stand-alone Java application based on the Eclipse platform technologies. The only prerequisite necessary to compile CLAMP is JRE 1.8.
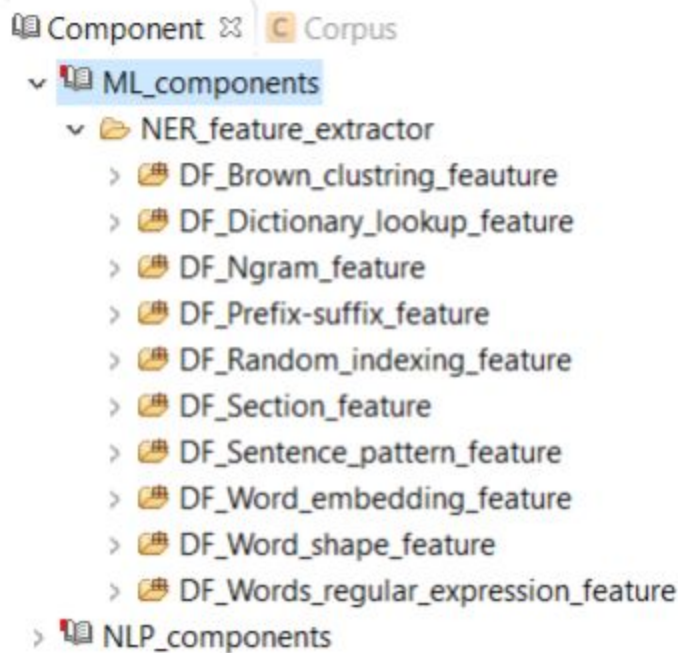
Natural Language Processing Components of CLAMP

NLP components are used for processing text. CLAMP offers multiple NLP components that are the building blocks for performing any of the NLP tasks. The individual components are provided as pre-built modules that are to be used in building pipelines for automatic text processing, as well as training customized machine learning models. The following picture displays the CLAMP NLP components, as well as its associated tools, algorithms, and resources.

NLP Components	Tools	Resources
Sentence detector	Lexicon Variants Generator	Section Header List
Tokenizer	Annotation interface	Annotated Corpus
POS tagger	OpenNLP	Medical Abbreviations
Chunker	**Algorithms**	Dictionaries
Section header identifier	Brown Clustering	
NER feature extractor	Random Indexing	
Named entity recognition	Word Embedding	
Assertion	SVM	
UMLS Encoder	CRF	

ML component of CLAMP

This component consists of different feature extractors. Here is the list of NER feature extractors:

Component   C Corpus

- ∨ ML_components
  - ∨ NER_feature_extractor
    - > DF_Brown_clustring_feauture
    - > DF_Dictionary_lookup_feature
    - > DF_Ngram_feature
    - > DF_Prefix-suffix_feature
    - > DF_Random_indexing_feature
    - > DF_Section_feature
    - > DF_Sentence_pattern_feature
    - > DF_Word_embedding_feature
    - > DF_Word_shape_feature
    - > DF_Words_regular_expression_feature
- > NLP_components

Output Visualization

One can view original content annotated with recognized clinical concepts. Different types of clinical concepts will be highlighted in different colors. Here is a view of text annotated with recognized clinical concepts.

7 Description: Discharge summary of a patient with *mood swings* `problem` and *oppositional and defiant behavior* `problem`.

9 (Medical Transcription Sample Report)

11 DISCHARGE SUMMARY

13 SUMMARY OF TREATMENT PLANNING:

15 Two *major problems* `problem` were identified at the admission of this adolescent:

17 1. *Mood swings* `problem`.

19 2. *Oppositional and defiant behavior* `problem`.

21 A developmentally appropriate *group oriented therapy program* `treatment` was *the primary treatment modality* `treatment` for this adolescent. He participated in at least eight *psychoeducational and activity groups* `treatment`. The attending psychiatrist provided *evaluation* `test` for and management of *psychotropic medications* `treatment` and collaborated with the treatment team. The clinical therapist facilitated *individual, group, and family therapy* `treatment` at least twice per week.

Why CLAMP?

1. CLAMP components are built on proven methods in many clinical NLP challenges including the I2B2 clinical NER (2009/2010-#2), SHARE/CLEF (2013-#1), SemEval2014 UMLS encoding (#1).
2. Depending on the task, users can train their own model for the machine learning-based components of CLAMP and evaluate custom models using a custom corpus.
3. Users can import clinical text corpora into the CLAMP workspace and annotate files using the built-in annotation tool that can be utilized in CLAMP projects, both as training and test datasets.
4. CLAMP allows building the NLP pipelines by offering all the requisite components such as named entity recognition, assertion, and component customizations.
5. All the knowledge resources required for CLAMP components like dictionaries, section header list, or medical abbreviation list are provided along with it.
6. CLAMP is built on the UIMA(Unstructured Information Management Architecture) framework and is therefore compatible with other systems such as cTAKES(clinical Text Analysis and Knowledge Extraction System). Further, CLAMP also utilizes the cTAKES' type system for lower linguistic level annotations.

Conclusion

CLAMP is an extremely helpful tool for recognizing and annotating clinical information in health records. CLAMP also enables users to build their own machine learning models based on a pre-annotated corpus. These models can be utilized for predictions on new files.

References:

1. Soysal, E., Wang, J., Jiang, M., Wu, Y., Pakhomov, S., Liu, H., & Xu, H. (2017). CLAMP–a toolkit for efficiently building customized clinical natural language processing pipelines. Journal of the American Medical Informatics Association, 25(3), 331-336.

   Link to website: (https://clamp.uth.edu/index.php)

2. Savova, Guergana K et al. "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications." *Journal of the American Medical Informatics Association: JAMIA* vol. 17,5 (2010): 507-13. doi:10.1136/jamia.2009.001560

3. Apache UIMA Framework: http://uima.apache.org/index.html