**HNER - Information pack**

The library has been developed in Java as an extension over the General Architecture for Text Engineering[[1]](#footnote-1) (GATE) framework. This extension is designed to load big collections of documents, with new features to create, evaluate and store gold standards and improvements to work with other libraries such as CoreNLP[[2]](#footnote-2), OpenNLP[[3]](#footnote-3) and Freeling[[4]](#footnote-4).

The NER process is mainly guided by linguistic models: gazetteers and rules. The framework provides a robust grammar engine called JAPE for recognizing patterns over documents. JAPE is a version of Common Pattern Specification Language (CPSL) and is based on finite state transducers. HNER library provides plugins and methods to ease the creation of gazetteers from external dictionaries and to support executions and evaluations of the implemented rules. The first probabilistic model implemented in the library has integrated the Conditional Random Fields (CRF) model from Mallet[[5]](#footnote-5).

Repo: <https://github.com/oeg-upm/hner>

**Implementation for Panama Papers.**

The implementation for Panama Papers is focused on the recognition of the following multilingual entities over highly unstructured texts:

* Companies
* Person Names
* Countries
* Addresses
* Emails
* Telephone/fax

Repo: <https://github.com/oeg-upm/ICIJ>

**Implementation for the Spanish drug regulatory agency (AEMPS)**

This implementation is oriented for Spanish disease entity recognition and linking with SNOMED-CT Terminology or MedDRA dictionary, both in their Spanish version.

Repo: <https://github.com/oeg-upm/AEMPS>

1. https://gate.ac.uk/ [↑](#footnote-ref-1)
2. https://stanfordnlp.github.io/CoreNLP/ [↑](#footnote-ref-2)
3. https://opennlp.apache.org/ [↑](#footnote-ref-3)
4. http://nlp.lsi.upc.edu/freeling/node/1 [↑](#footnote-ref-4)
5. http://mallet.cs.umass.edu/ [↑](#footnote-ref-5)