

# Joint Entity and Relation Extraction for clinical decision making

IXA Group

Alicia Pérez, Olatz Pérez de Viñaspre, Maite Oronoz, Arantza Casillas

Contact: [alicia.perez@ehu.eus](mailto:alicia.perez@ehu.eus)

## 1 Description

This research project is devoted to Medical Relation Extraction. Medical documents contain a variety of information, and as in the case of relations, not always all the information is explicit. To extract this type of information is very important when the experts have to make a decision. Relations can be of different nature and can appear in many types of documents such as: medical, legal....

Challenges and open gaps in the literature: There are two main approaches to deal with relation extraction:

1. In a sequential or cascade approach, entities are extracted first, and in a second step, a classifier is in charge of labelling relations between entities. The weakness of this approach rests on the fact that the errors from the entity recogniser are propagated to the relation extraction system.
2. Joint approaches exploit contextual features that serve for both entity and relation annotation at the same time. Deep neural networks offer the context to facilitate joint approaches.

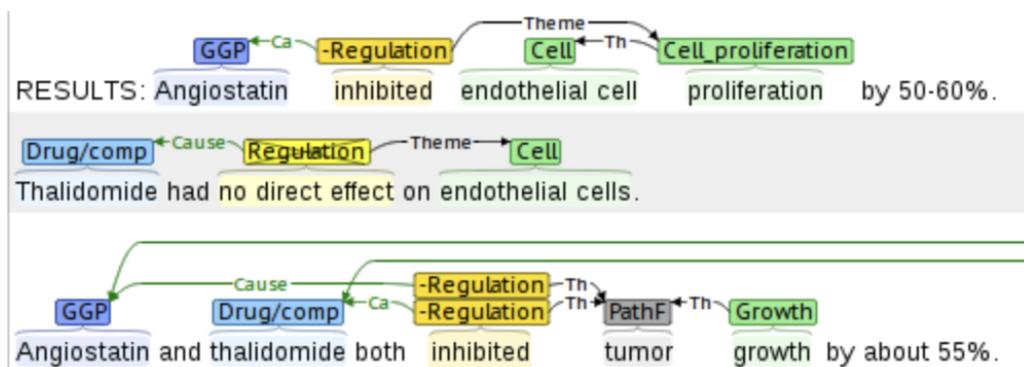


Figure 1: Entities and relations. Source of the figure: <https://brat.nlp.lab.org>

This project is co-located to support a research project: Dot-Health.

## **2 Goals**

The student will apply different techniques and relatedness metrics in order to build a prototype able to identify relations not explicitly expressed in documents. The key objectives are:

- Apply different relation extraction techniques, the most popular of the state of the art.
- Evaluate and compare the selected techniques.

## **3 Requirements**

- Good programming skills.
- The master dissertation can be written in English, Spanish or Basque.

## **4 Learning outcomes:**

the student will acquire background in information extraction reinforcing the following areas:

- deep learning applied to information extraction
- relatedness and confidence metrics for information extraction

## **5 Framework**

Python

## **6 Task and plan**

4 weeks: study literature, python specific libraries

12 weeks: experimentation and evaluation

6 weeks: write down the thesis