

This document aims to describe the basic annotation guidelines followed during the training of a NER tagger using greek legislative texts, in the scope of GSoC2019.

We decided not to train extra named entities labels since it could require immense amounts of data and used the original annotation scheme of greek spaCy. This scheme contains 6 labels: ORG for organisation, PERSON for person , GPE for geopolitical entities, LOC for locations, PRODUCT for products and EVENT for events.

Since we couldn't redetermine entity labels we tried to apply this scheme to the scope of our problem. More specifically we used:

- ORG for organisations, government authorities, courts, churches and anything that falls into this category in the sense that it is an organised structure and people are part of it.
- GPE for countries and federations. To distinguish from the above "Ελλάδα" (meaning Greece) is a GPE but "Ελληνικό Δημόσιο" (meaning Greek state or the conglomerate of Greek state mechanisms) is tagged as ORG.
- LOC for locations and geographic indications but not for countries.
- PERSON for either a specific individual (e.g. 'Νίκος Κούκος') or the office that an individual may hold, his administrative role("ο Υπουργός Υγείας")
- PRODUCTS for actual products but also for any product of the praxis of an organisation or a GPE, meaning also pacts, agreements, programs, guidelines or systems.
- EVENT for events such as "Αργία".