

SX2 Proiektua

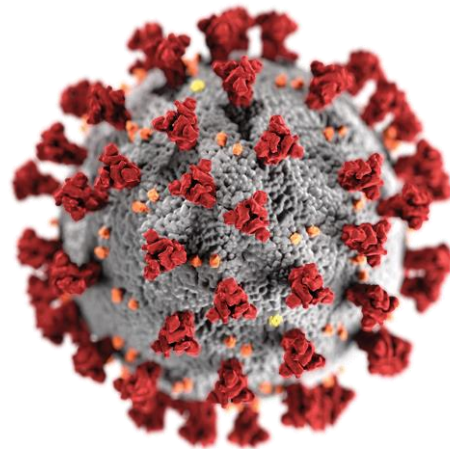
Iker Ossa

Aurkibidea

- Sarrera
- Erlazionatutako lanak
- Sistema
- Datuak
- Emaitzak
- Analisia
- Ondorioak

Sarrera

- COVID-19-ren propietateak atera
- Spacy analizatzaile sintaktikoa biomedikuntzara egokitu



Erlazionatutako lanak

- Kode lagungarria
- Craft-en artikulua akademikoa
- Ebaluaziorako programak

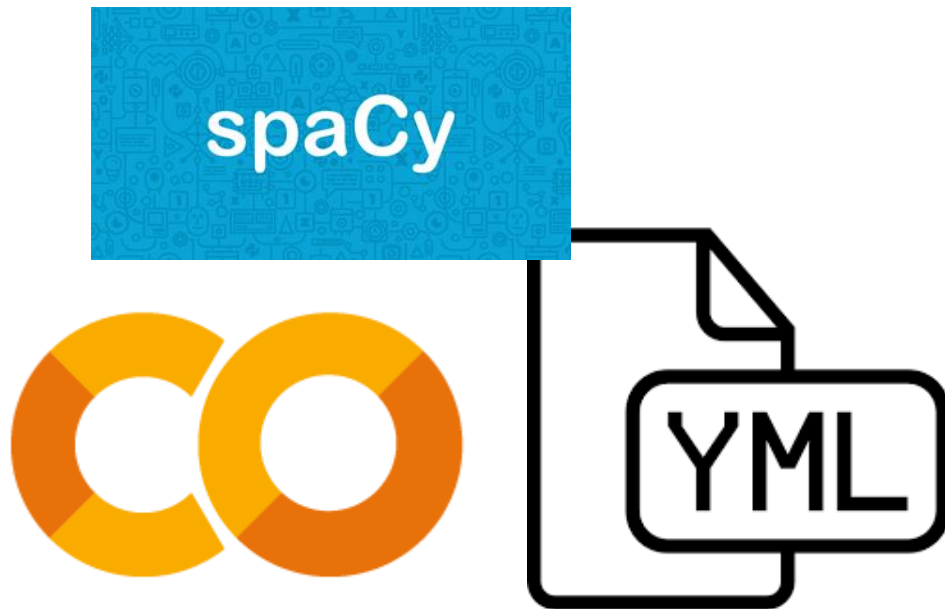
CRAFT Structural Annotation Task

Developing dependency parsers for the full text biomedical literature



Sistema

- Google colab
- Spacy analizatzaille sintaktikoa
- Tagger parser ud



Datuak

- ConLL-U fitxategiak

```
en_esl-ud-dev.conllu
# sent_id = 0100_2000_12-doc2638.xml-6
# text =
# error_annotation =
1      -      -      PRON    EX      -      2      expl    -      -
2      -      -      VERB    VBZ     -      0      root    -      -
3      -      -      DET     DT      -      5      det     -      -
4      -      -      ADJ     JJ      -      5      amod    -      -
5      -      -      NOUN    NN      -      2      nsubj   -      -
6      -      -      PART    TO      -      7      mark    -      -
7      -      -      VERB    VB      -      5      acl     -      -
8      -      -      ADV     RB      -      7      advmod  -      -
9      -      -      SCONJ   IN      -      12     mark    -      -
10     -      -      DET     DT      -      11     det     -      -
11     -      -      NOUN    NNS     -      12     nsubj   -      -
12     -      -      VERB    VBP     -      2      advcl   -      -
13     -      -      PART    RB      Polarity=Neg 12     advmod  -      -
14     -      -      VERB    VB      -      12     xcomp   -      -
15     -      -      PRON    NN      -      14     obj     -      -
16     -      -      PUNCT   .       -      2      punct   -      -
```

Emaitzak

Z1

```
{'LAS': 66.77320685751666  
  
'MLAS': 35.19622158840356,  
  
'BLEX': 44.040828919334444}
```

Z2

===== Results ==

TOK	99.45
TAG	95.02
POS	95.51
MORPH	95.80
UAS	85.88
LAS	83.78
SENT P	94.32
SENT R	90.65
SENT F	92.45
SPEED	14100

Analisia

- Ez da artikuluko emaitzen berdina
- Z1-etik Z2 hobekuntza handia

Parser	Word vectors	Extra data	LAS(dev)	LAS(test)
Turku	Bio, word2vec/CBOW (window 2)	4k sentences, pmcoa articles	89.92	89.695
Turku	Bio, word2vec/CBOW (window 2)	No	89.86	89.650
Turku	Bio (CRAFT tokens), word2vec/CBOW (defaults)	No	89.78	89.536

Ondorioak

- Artikuloko emaitzak lortzeko hurbil
- Hobekuntzak egin daitezke



GALDERARIK?

