(1) PROCESAR TEXTOS CON TIKA

QUEDARNOS CON:

- TÍTULO (TITLE)
- DESCRIPTION
- X-TIKA:content : Nos podemos quedar con las frases que tengan longitud ≥ 20 (para quitar todo lo que no nos interesa)

~~TIKA~~

Tika_reader.py
Tika_procesor.py
~~Process_text.py~~

(2) SACAR VECTORES

(3) APLICAR ALGORITMO CONGLOMERADOS

---

(4) ENTIDADES NOMBRADAS

- SPACY → APLICAR A
- STANFORD
- CoNLL2002 CHUNKER (BIGRAMAS)

ner_stanford.py
spacy_exercise2.py

ner_train_chunker.py

TRATAMIENTO DEL TEXTO

(A) Stopwords
Puntuación
Lematizar

Process_Text.py

(B) Traducir de texto de inglés a español y español a inglés (funciona mejor un diccionario que otro en un idioma que en otro?)

TextBlob