

Downloading https://raw.githubusercontent.com/stanfordnlp/stanza-resources/master/resources_2021-03-13_00:56:30.zip INFO: Downloading default packages for language: es (Spanish)...

Downloading <http://nlp.stanford.edu/software/stanza/1.2.0/es/default.zip>: 100%|██████████| 2021-03-13 00:58:24 INFO: Finished downloading models and saved to /root/stanza_resources



▼ Extracción de texto

```
# tema que se buscará
tema = 'Felicidad'

# enlace para acceder al contenido de la Wikipedia
response = requests.get(f'http://es.wikipedia.org/wiki/{tema}')

# transformar el sitio web a texto plano
soup = BeautifulSoup(response.text)

# obtener el texto por párrafos
parrafos = soup.find_all('p')

# obtener solo el primer parrafo
texto = parrafos[0].text

# mostrar el texto con el que trabajaremos
print(texto)
```

La felicidad es una emoción que se produce en un ser consciente cuando logra o alguien c



▼ Usando librería SpaCy

```
# importando línea de comandos de spacy
import spacy.cli

# descarga de paquete del idioma español
spacy.cli.download('es')

✓ Download and installation successful
You can now load the model via spacy.load('es_core_news_sm')
✓ Linking successful
/usr/local/lib/python3.7/dist-packages/es_core_news_sm -->
/usr/local/lib/python3.7/dist-packages/spacy/data/es
You can now load the model via spacy.load('es')
```

```
# cargando paquete español a una variable
```

```

nlp = spacy.load('es_core_news_sm')

# convirtiendo texto a un formato reconocido por spacy
texto_doc = nlp(texto)

# tokenización, lematización y POS tagging
print('palabra -- lematización -- pos tagging\n')
for token in texto_doc:
    print(f'{token.text} -- {token.lemma_} -- {token.pos_}')

    palabra -- lematización -- pos tagging

    La -- La -- DET
    felicidad -- felicidad -- NOUN
    es -- ser -- AUX
    una -- uno -- DET
    emoción -- emoción -- NOUN
    que -- que -- PRON
    se -- se -- PRON
    produce -- producir -- VERB
    en -- en -- ADP
    un -- uno -- DET
    ser -- ser -- NOUN
    consciente -- consciente -- ADJ
    cuando -- cuando -- SCONJ
    logra -- lograr -- VERB
    o -- o -- CONJ
    alguien -- alguien -- PRON
    da -- dar -- VERB
    cuyo -- cuyo -- PRON
    sentimiento -- sentimiento -- NOUN
    como -- como -- SCONJ
    cosquillas -- cosquillar -- NOUN
    o -- o -- CONJ
    llegar -- llegar -- VERB
    a -- a -- ADP
    un -- uno -- DET
    momento -- momento -- NOUN
    de -- de -- ADP
    conformación -- conformación -- NOUN

    --
    -- SPACE

```

▼ Utilizando la suite Stanza (Stanford)

¿Qué es estanza?

Una suite de librerías que ejecutaban varias tareas de NLP para el grupo de investigación en NLP de Stanford, esta suite se unificó en un solo servicio que le llamaron **CoreNLP** con base en código java: <https://stanfordnlp.github.io/CoreNLP/index.html>

Para python existía StanfordNLP: <https://stanfordnlp.github.io/stanfordnlp/index.html>

Sin embargo, StanfordNLP ha sido deprecada y las nuevas versiones de la suite de NLP reciben mantenimiento bajo el nombre de Stanza: <https://stanfordnlp.github.io/stanza>

```
# definiendo el idioma y las funciones que se aplicarán
stanza_nlp = stanza.Pipeline('es', processors='tokenize, lemma, pos')

# convirtiendo el texto a un formato reconocido por stanza
texto_stanza = stanza_nlp(texto)

2021-03-13 00:58:35 WARNING: Language es package default expects mwt, which has been added
2021-03-13 00:58:35 INFO: Loading these models for language: es (Spanish):
=====
| Processor | Package |
-----
| tokenize  | ancora  |
| mwt       | ancora  |
| pos       | ancora  |
| lemma     | ancora  |
=====

2021-03-13 00:58:35 INFO: Use device: cpu
2021-03-13 00:58:35 INFO: Loading: tokenize
2021-03-13 00:58:35 INFO: Loading: mwt
2021-03-13 00:58:35 INFO: Loading: pos
2021-03-13 00:58:35 INFO: Loading: lemma
2021-03-13 00:58:35 INFO: Done loading processors!
```

```
# tokenización, lematización y POS tagging
print('palabra -- lematización -- pos tagging\n')
for oracion in texto_stanza.sentences:
    for palabra in oracion.words:
        print(f'{palabra.text} -- {palabra.lemma} -- {palabra.pos}')
```

📄 palabra -- lematización -- pos tagging

```
La -- el -- DET
felicidad -- felicidad -- NOUN
es -- ser -- AUX
una -- uno -- DET
emoción -- emoción -- NOUN
que -- que -- PRON
se -- él -- PRON
produce -- producir -- VERB
en -- en -- ADP
un -- uno -- DET
ser -- ser -- NOUN
consciente -- consciente -- ADJ
cuando -- cuando -- CONJ
logra -- lograr -- VERB
o -- o -- CONJ
```

alguien -- alguien -- PRON
da -- dar -- VERB
cuyo -- cuyo -- PRON
sentimiento -- sentimiento -- NOUN
como -- como -- SCONJ
cosquillas -- cosquilla -- NOUN
o -- o -- CCONJ
llegar -- llegar -- VERB
a -- a -- ADP
un -- uno -- DET
momento -- momento -- NOUN
de -- de -- ADP
conformación -- conformación -- NOUN