ACTIVIDAD 1:
CATEGORÍAS GRAMATICALES Y EXTRACCIÓN DE ENTIDADES.

Última actualización: **02-Abril-2021**
CUNEF – c/ Leonardo Prieto Castro, 2. 28040 Madrid. Tfno. (++34) 91 448 08 92. www.cunef.edu

Índice.

1. Objetivo

El objetivo de esta práctica es realizar un notebook en Python que realice un análisis morfológico y extraiga las entidades que se presenten en un texto. Las entidades para extraer serán de tipo organización, localización y persona.

Para realizar esta actividad se utilizará un sistema ya entrenado, bien sea el entrenamiento con que se cuenta en **NLTK** o en uno modelo de lenguaje de **spaCy**.

2. Guion de la actividad

1. En esta actividad se le facilitan dos ficheros de texto, uno en español y otro en inglés. Son los ficheros **esp.txt** e **ing.txt.**
2. Lea estos ficheros desde un programa Python, y realice las labores de pre-procesamiento habituales: división en frases, división en palabras y conversión a las formas normales. Utilice los frameworks que prefiera, se sugiere **NLTK** o **spaCy**
3. Realice un análisis morfológico (POS, Part of Speech) de los términos incluidos.
4. Por último, extraiga entidades de los textos anteriores
5. Una vez realizados todos los análisis, liste en pantalla las anotaciones obtenidas: frases, palabras, categoría gramatical del análisis morfológico, y entidades extraídas: localizaciones, organizaciones y personas
6. Comente las diferencias entre los resultados obtenidos en español y en inglés. ¿Hay diferencias entre utilizar **NLTK** y **spaCy**?

.

3. Formato y fecha de entrega

El entregable correspondiente a esta actividad será el notebook Python que ejecute las tareas anteriores.

La fecha límite de entrega para esta Actividad es el **25 de Abril de 2021**