

Lingüística Computacional Ejercicio 5
--

Los puzzles de palabras suelen (o solían o suelen en otros países) ser entretenimientos frecuentes. Incluso hay aplicaciones para algunos de ellos. Uno bastante popular es el *target puzzle*. Si no lo conocen, pueden ver sus características en la web. En el capítulo 2 del Book de NLTK viene incluso una implementación.

Para construir un puzzle interesante, o para obtener todas las palabras permitidas por uno concreto, necesitamos (i) una gran base de datos léxica, y (ii) un programa.

1. (1 point) La base de datos léxica vamos a obtenerla del Corpus de Referencia del Español Actual. Descargar la lista total de frecuencias de la web de la RAE. En Python, abrir el fichero, obtener las palabras y guardarlas en una variable cuyo contenido pueden guardar en un fichero para usos posteriores.
2. (2 points) Construir una función para que, dados (i) un puzzle cuyas 9 letras son introducidas por el usuario desde el teclado, y (ii) la base de palabras del CREA, se obtengan todas las palabras posibles de acuerdo a las reglas del puzzle.

No copiar la implementación del NLTK, ni ninguna de las miles que existen en la red.

3. (2 points) *Soundex* es un algoritmo para representar nombres. Si no lo conoce, lea sus reglas en la web. Escribir en la forma de script una implementación de Soundex adaptado al español. El usuario escribe por teclado un nombre y le devuelve el resultado.