|  |
| --- |
| Procesamiento de lenguaje natural |
| Práctica 2: Auto-evaluación (Código) |
|  |
| Gabriel Vázquez Torres |
|  |

Procesamiento de lenguaje natural

Práctica 2: Auto-evaluación (Código)

# Introducción

A continuación, una descripción acerca de cómo se ha conseguido obtener generalidad y cobertura para conseguir la efectividad y generalidad de la práctica. Además, se añadirán algunos comentarios acerca de la metodología seguida y el diseño de ciertas reglas.

# Comentarios acerca del código

El código se ha transformado para que pueda funcionar correctamente en una versión de Python mayor a 3.

Además, he creado una librería complementaria para que pueda funcionar. Todo el código creado por mí se puede encontrar en el GitHub: “https://github.com/Gabvaztor/LSI\_IA\_UNED/tree/master/Procesamiento\_de\_lenguaje\_natural”

# Descripción de la auto-evaluación

Para la realización de la práctica he seguido, principalmente, cuatro pasos:

1. **Información a partir de doctores filólogos:** Gracias a mi trabajo como investigador en la Universidad de Sevilla y a conocer doctores filólogos, he tenido la posibilidad de preguntar acerca de ciertas reglas que se cumplen en el castellano entre lexemas y morfemas. Ello me ha proporcionado una serie de conocimientos para crear ciertas expresiones regulares y entender por qué, para algunas palabras del texto, era mejor no utilizar expresiones para encontrarlas (ya que no estaba comprobado que, en el castellano, esas expresiones tuvieran más del 50% de éxito).
2. **Información a partir del libro oficial de la asignatura:** Gracias a este libro he podido esclarecer y entender diferentes puntos de vista de una forma más sintetizada. Además, he podido entender qué son los autómatas y el funcionamiento de las expresiones regulares. También he podido entender algunas diferencias que existen entre el inglés y el castellano intentando captar algunas de las normas que cumple el idioma castellano para realizar la práctica.
3. **Estudio de documentos de filología encontrados por la WEB:** Esta ha sido una de las partes más extensas de todo el proceso de aprendizaje, a parte del libro. He encontrado mucha información acerca de ciertas terminaciones de verbos, adverbios y sustantivos, por ejemplo. Todo esto ha estado centrado en la eficiencia y generalidad de la práctica para nuevos textos. Existen numerosas WEBs y foros donde se puede encontrar mucha información relevante.
4. **Estudio de los lexemas y morfemas por mi cuenta:** A partir de libros de texto de primaria y secundaria, además de las conjugaciones de los verbos regulares en internet, he encontrado diferentes patrones que se cumplen en numerosos casos. Esto me ha servido a la hora de encontrar, sobre todo, patrones en las conjugaciones de los verbos regulares.

Así mismo, es importante comentar que el etiquetado morfológico dado para el castellano tiene una serie limitaciones. Por ejemplo, está limitado a establecer una única forma gramatical en una palabra que, en castellano, contiene dos. Un buen ejemplo sería la palabra “amarlo”. Contiene el infinitivo del verbo “amar” y el determinante “lo”. ¿Cómo clasificar esto según las limitaciones encontradas? En mi caso, en el código tengo comentado la expresión regular: “*(r'.\*arl(a|e|o|as|os|es)$', 'VMN0000 y DET')”* que representaría tanto al verbo acabado en infinitivo “ar” seguido de un determinante.

A modo de ejemplo, he añadido al diccionario la forma “dict['del'] = 'PREP y DET'” (entre otros), que representaría la conjugación de las palabras “de” y “el”, para que podáis valorar que lo he tenido en cuenta.

Una vez que he creado todas las expresiones regulares que he encontrado útiles, he analizado el texto y, para los que cometían errores ya que eran palabras no tomadas en cuenta en dichas expresiones, he ido añadiendo al diccionario cada una de ellas para potenciar la efectividad del texto.

En casos generales, la mayoría de las palabras serán tomadas como sustantivos masculinos o femeninos, ya que son las palabras que más existen en el castellano (conocimiento adquirido por los cuatro pasos comentados previamente.

A forma de resumen, he intentado crear las expresiones regulares teniendo en cuenta el texto dado (tratando de obtener el 100% de acierto, aunque fuese añadiendo palabra a palabra al diccionario) y la generalidad dentro de ciertos porcentajes estadísticos de palabras que podrían errar y acertar. Casi todas las expresiones regulares producen imprecisiones por lo que se ha tenido en cuenta esto con énfasis.