**TAREA 41: DICCIONARIO**

* **EXPRESIONES REGULARES:** En computo teórico y teoría de lenguajes formales una expresión regular, o expresión racional, también son conocidas como regex o regexp, por su contracción de las palabras inglesas regular expresión, es una secuencia de caracteres que conforma un patrón de búsqueda. Se utilizan principalmente para la búsqueda de patrones de cadenas de caracteres u operaciones de sustituciones.

Las expresiones regulares son patrones utilizados para encontrar una determinada combinación de caracteres dentro de una cadena de texto. Las expresiones regulares proporcionan una manera muy flexible de buscar o reconocer cadenas de texto.

La mayoría de las formalizaciones proporcionan los siguientes constructores: una expresión regular es una forma de representar los lenguajes regulares (finitos o infinitos) y se construye utilizando caracteres del alfabeto sobre el cual se define el lenguaje.

* **PROCESAMIENTO DE LENGUAJE NATURAL:** El procesamiento de lenguaje natural, abreviado PLN (NPL, natural language processing) es un campo de las ciencias de la computación, de la inteligencia artificial y de la lingüística que estudia las interacciones entre las computadoras y el lenguaje humano. Se ocupa de la formulación e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas y máquinas por medio del lenguaje natural, es decir, de las lenguas del mundo. No trata de la comunicación por medio de lenguas naturales de una forma abstracta, sino de diseñar mecanismos para comunicarse que sean eficaces computacionalmente (que se puedan realizar por medio de programas que ejecuten o simulen la comunicación). Los modelos aplicados se enfocan no solo a la compresión del lenguaje de por sí, sino a aspectos generales cognitivos humanos y a la organización de la memoria. El lenguaje natural sirve solo de medio para estudiar estos fenómenos. Hasta la década de 1980, la mayoría de los sistemas de PLN se basaban en un complejo conjunto de reglas diseñadas a mano. A partir de finales de 1980, sin embargo, hubo una revolución de PLN con la introducción de algoritmos de aprendizaje automático para el procesamiento del lenguaje.
* **SCRAPPING:** Consiste en extraer datos de cualquier documento, pagina web… de forma automatizada. Es decir, convertimos en una base de datos estructurada la información que podemos encontrar publicada. La traducción literal de data scrapping en español sería raspado de datos.

Copiar datos de una pagina web y convertirlos a Excel de forma manual se consideraría extracción de datos. Si automatizamos este trabajo recurriendo a bots o robots estaremos hablando de web scrapping.

El scrapping de datos es el mas común, pero podemos hacer scrapping de imágenes, de videos, y de cualquier tipo de archivos.

* **CRAWLING:** Un crawler, o araña, se arrastra por las diferentes paginas web imitando el comportamiento humano. Por lo que se puede decir que el ejercicio de búsqueda y rastreo, se le atribuye al crawling, y la parte de descargar información y almacenarla se denomina scrapping.