



Entender que son las expresiones regulares y el manejo de estas con cierta solvencia

Las **expresiones regulares** o **regex** (*regular expression*) son patrones utilizados en programación para encontrar una determinada combinación de caracteres dentro de una cadena de texto. Se utilizan mucho en la gestión de la información de cualquier aplicativo. Todos los principales lenguajes de programación (C++, java, ruby, python, php, ...) utilizan expresiones regulares. En ocasiones, el desarrollador de aplicaciones de Inteligencia Artificial, debe trabajar con cadenas de texto para:

- 1.- La **búsqueda de elementos** particulares dentro de un largo texto. Por ejemplo, es posible que desee identificar todas las direcciones de correo electrónico existentes en un contenido muy extenso.
- 2.- Para **reemplazar**, por ejemplo, un fichero de *lenguaje de marcas de hipertexto* (HTML) con unas etiquetas deficientes por otras correctas.
- 3.- Para **validar una contraseña o correo electrónico** de un formulario. Por ejemplo, se puede querer verificar que una contraseña cumpla con ciertos criterios, tales como: una combinación de mayúsculas y minúsculas, dígitos, signos de puntuación. O para validar si un usuario que se registra a una página web a metido un correo electrónico correcto mediante la definición de unos patrones predeterminados: la existencia de un @, ...
- 4.- Para **automatizar ciertas acciones**. Por ejemplo, es posible que se desee procesar ciertos archivos en un directorio, pero solo si cumplen unas condiciones predeterminadas.
- 5.- Para **reformatear texto**. Por ejemplo, se pueden importar datos de un programa como archivo de texto, modificar su contenido y exportarlo a un editor de texto una vez corregido.
- 6.- Para técnicas de **scrapping**. ATENTO! Esta técnica de análisis, captura e ingesta de datos se utilizan mucho para construir posteriormente soluciones de Inteligencia Artificial. Es una técnica que se utiliza en múltiples lenguajes de programación y existen diferentes librerías que nos ayudan en la construcción de este tipo de arañas (Para php la librería CURL, para python las librerías beautifulsoup o scrapy, para C++ ...). Ojo, no son lo mismo el crawling y el scraping; en este [link](#) se explican las diferencias. En muchas ocasiones esta técnica de ingesta y análisis de datos se hace desde páginas web de terceros por lo que entenderás que tener ciertas nociones de HTML para el raspado o scrapping de contenido web será necesario para el desarrollo de soluciones de IA.
- 7.- Para **buscar coincidencias** entre diferentes textos.

He aquí **unas guías** sobre la sintaxis utilizada para las expresiones regulares en diferentes lenguajes de programación:

- 1.- Para python: <https://platzi.com/blog/expresiones-regulares-python/>
- 2.- Para javascript: <https://www.arkaitzgarro.com/javascript/capitulo-11.html>
- 3.- Para php: <https://diego.com.es/expresiones-regulares-en-php>
- 4.- Para ...

Los alumnos podéis disponer en Internet de **diferentes herramientas** para probar vuestra sintaxis para expresiones regulares:

<https://regex101.com/>

<https://www.online-toolz.com/langs/es/tool-es-regexp-editor.html>

<https://www.metriplica.com/informes-y-estudios-de-analitica-digital/recursos-herramientas-de-analitica/expresiones-regulares/>

Para esta tarea proponemos el siguiente ejercicio: el @egger mediante técnicas de **Regex** debe **calcular el número de caracteres, el número palabras y ranking de palabras por frecuencia de uso** del siguiente texto. La aplicación debe servir para cualquier otro texto:

"En Strandhill, Irlanda, se cruzó en mi camino Chris, un señor de los que inspiran y se posicionan como referente. Fue una pieza fundamental en un momento de pura congelación. Te cuento? Strandhill es una playa donde el mar ruge muy bien, siempre está lleno de surfistas en busca de buenas olas. Y allí estaba yo también. Después de unos meses viviendo en una ciudad sin costa, mis ganas por hacer un poco de surfing estaban por las nubes. Aunque tenía un «pequeño» problema: no tenía equipo, ni tabla, ni neopreno, y tampoco había ninguna tienda para alquilarlo. Todo se puso a rodar enseguida. Escribí a un famoso surfista de la zona y recibí una respuesta increíble. «Mi casa está en la calle x, la puerta está abierta y tienes la tabla en la esquina. Ven cuando quieras», me dijo. Y eso hice, fui para allá y la cogí. Aunque el neopreno no me lo pudo prestar, y no porque se negara? Lamentablemente le sacaba unos 15 cm de altura. Luego, en la playa, un alemán me solucionó el tema del neopreno. Me prestó uno que había por su maletero, uno muy fino, de los que uso yo en el Mediterráneo en otoño o primavera. Y claro, era invierno y estábamos en Irlanda. El caso es que salí del agua más pronto de lo previsto y con las manos, la cabeza y los labios congelados. Me empecé a cambiar en el mismo paseo que contorneaba la costa y, estando semidesnudo, se me acercó Chris. «Está fría el agua, eh», apuntó este fenómeno. Chris superaba los 65 años y todos los días hacía un recorrido de decenas de kilómetros para llegar hasta allí. Sus gracietas y su buena conversación me hicieron apartar el frío de la parte de mi cabeza que se encarga de pensar, y hasta cantamos juntos la canción de Annie. Sé que esto último puede sonar raro, ¿quién canta Annie semidesnudo y congelado en un paseo de Irlanda con un señor que acaba de conocer? Pero? seguro que a ti también te han pasado cosas así."

Ref: <https://pegameunviaje.com/3-anecdotas-divertidas-de-mis-viajes/>

NOTA:

1.- Para una de las ramas de la Inteligencia Artificial, la del Procesamiento del Lenguaje Natural, conocer estas técnicas va a ser útil ya que nos proveerá de una herramienta de analítica de texto muy avanzada. El PLN se utiliza para filtros de correo electrónico, asistentes inteligentes (Siri, Ok Google), traductores, llamadas telefónicas digitales, ...

2.- Para la ingestión de diferentes fuentes de datos durante el desarrollo soluciones de Machine Learning o Deep Learning el raspado de información también será necesario. Imagina que estás desarrollando una solución de predicción de venta de refrescos de una tienda de comestibles. Seguramente la meteorología es una variable que afecta directamente a este consumo. Pues bien, podrás obtener las previsiones meteorológicas a partir de páginas web de meteorología mediante técnicas de scrapping y utilizar este dato en tu modelo predictivo.

#HASHTAGS (etiquetas de ayuda para búsqueda de información relevante)

#expresiones-regulares #Regex #procesamiento-del-lenguaje-natural #web-scrapping

LINKS DE INTERÉS

<https://robologs.net/2019/05/05/como-utilizar-expresiones-regulares-regex-en-python/>
<https://geekflare.com/es/regular-expression-tester/>
<https://www.metriplica.com/informes-y-estudios-de-analitica-digital/recursos-herramientas-de-analitica/expresiones-regulares/>
<https://www.youtube.com/watch?v=M72lwALYRJU>

DICCIONARIO

expresiones-regulares | procesamiento-del-lenguaje-natural | scrapping | crawling

PUNTUACIÓN

Programación: 6
Redes: 1
Seguridad: 1
Algoritmia: 5