

## PRÁCTICA 1

## LOS ELEMENTOS DE LA TABLA PERIÓDICA

MARIO MARTÍNEZ GARCÍA

## CONTEXTO

A lo largo de esta práctica se implementarán técnicas de web scraping mediante el lenguaje de programación Python con el que se extraerá información sobre los elementos de la tabla periódica. Esta práctica se desarrolla en el *máster de ciencia de datos de la UOC*, más concretamente sobre la asignatura de *Tipología y ciclo de vida de los datos*.

## IMAGEN DESCRIPTIVA

La imagen muestra una tabla periódica de los elementos. Cada elemento está representado por un cuadrado que contiene su símbolo químico y su número atómico. La tabla está organizada en periodos (filas) y grupos (columnas). Los grupos están etiquetados como IA, IIA, IIIA, IVA, VA, VIA, VIIA, y GRUPO CERO. Los periodos están etiquetados como PERIODO 1, 2, 3, 4, 5, 6, y 7. Hay una leyenda que indica 'TAMBIÉN ATÓMICO' con un punto verde y 'radio 100 pm'. La tabla incluye elementos como H, He, Li, Be, B, C, N, O, F, Ne, Na, Mg, Al, Si, P, S, Cl, Ar, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Br, Kr, Rb, Sr, Y, Zr, Nb, Mo, Tc, Ru, Rh, Pd, Ag, Cd, In, Sn, Sb, Te, I, Xe, Cs, Ba, La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Fr, Ra, Ac, Th, Pa, U, Np, Pu, Am, Cm, Bk, Cf, Es, Fm, Md, No, y Lr.

Ilustración 1: Tabla periódica de los elementos

## DESCRIPCIÓN

Sobre el dataset aparecerán los 118 elementos de la tabla periódica: los metales, los no metales y los metaloides. Estos se disponen ordenados de menor a mayor número atómico para una mejor comprensión. Sobre el dataset se han intentado recopilar algunos de los atributos más importantes que nos permitan interactuar con el archivo.

## CONTENIDO

Todos los atributos han sido extraídos de la [Tabla periódica](#) localizada en *Rincón Educativo* excepto la densidad. Aunque la densidad de algunos elementos aparecía sobre la web de *Rincón Educativo* existían grandes lagunas y para paliar este desperfecto se optó por extraer este valor de cada uno de los elementos de la tabla periódica de la [Wikipedia](#). A partir de las páginas anteriores construiremos los siguientes campos:

- **N\_atómico.** Se recopila el número atómico del elemento.
- **Elemento.** Forma abreviada del elemento, es decir, del hidrógeno aparecerá una H.
- **Nombre.** Nombre completo del elemento.
- **Masa.** Masa del elemento expresada en “umas”.

- **Periodo.** Periodo del elemento en la tabla periódica.
- **Grupo.** Grupo del elemento en la tabla periódica.
- **Estructura.** Estructura cristalina más estable en la que se encuentra.
- **Conf\_electronica.** Configuración electrónica más estable del elemento. Sobre algunos elementos, además de sus últimas capas, aparece entre corchetes un gas noble del que toman su configuración inicial.
- **Descubrimiento.** Año o años de descubrimiento. En algunos elementos aparece la “Prehistoria” como época en la que fue descubierto.
- **Densidad.** Nos muestra la densidad de cada uno de los elementos en  $\text{kg/m}^3$ .

Cabe destacar que no todas las columnas están completas, por ejemplo, no todos los elementos poseen grupo y, pese a la corrección realizada sobre la densidad seguiremos teniendo algunos elementos vacíos. Estos elementos vacíos serán representados mediante ‘-’.

## AGRADECIMIENTOS

Los datos han sido recopilados de un apartado de la página web *Rincón Educativo*, concretamente en el apartado de [Tabla periódica](#). A través de estas líneas me gustaría agradecer la gran labor sobre la sociedad que realiza esta web, aportando recursos educativos de gran bagaje que facilitan la comprensión de los recursos teóricos vistos en las aulas. Por otra parte, el atributo densidad será extraído de cada uno de los elementos que aparecen en la tabla periódica de la [Wikipedia](#). Los datos de la tabla periódica encontrados en la Wikipedia son extraídos de la IUPAC, entre otra infinidad de libros.

## INSPIRACIÓN

Con la intención de recopilar información sobre algo conocido que me ayudase a entender mejor como aplicar el web scraping sobre una página, opte por recabar información sobre los elementos de la tabla periódica. Son numerosas las páginas web donde aparece una tabla periódica pero pocas de ellas son interactivas.

En un primer momento encontré una tabla periódica [dinámica](#) creada a partir de enlaces sobre la Wikipedia. No conforme con esta, seguí buscando hasta dar con una tabla dinámica donde cada uno de los elementos y características habían sido insertados en código “html” sobre la página web. Como ya he comentado la [Tabla periódica](#) localizada en *Rincón Educativo* sería la utilizada para extraer toda la información. Con el objetivo de enriquecer la práctica se añadiría un atributo a partir de la tabla interactiva presente en la [Wikipedia](#).

Esta primera toma de contacto con el *web scraping* ha servido para formar un archivo con los elementos de la tabla periódica y algunas de sus características. Este conjunto nos permite tratar los datos de múltiples formas para conocer un poco más las propiedades intrínsecas de los mismos. Por ejemplo, puede comprobarse la correlación entre ciertas variables, como la masa y la densidad o el número atómico entre otros. Además, mediante la creación de una línea temporal podremos visualizar de forma clara la época en la que fue descubierto cada elemento. Estas son algunas de las diversas aplicaciones que puede tener nuestra base de datos.

Básicamente, el objetivo fundamental era la obtención de un archivo con los elementos de la tabla periódica y algunas de sus propiedades más importantes con las que poder responder preguntas y sacar conclusiones relevantes.

## LICENCIA

La licencia escogida sobre el dataset será “*CC BY-SA 4.0 License*” ya que nos aporta una serie de ventajas a nosotros como creadores del dataset y puede ser de utilidad para terceras personas. Mediante esta licencia podrán realizarse cambios sobre la misma, por ejemplo, si en un futuro se quieren añadir más atributos, pero deberá adjuntarse el nombre del creador del dataset generado indicando los cambios realizados. Además, la licencia nos permite abrirnos sobre el mundo comercial y dar a conocer nuestro trabajo llegando a más público. Por otra parte, el tipo de licencia que adquiere un dataset tras su modificación será el mismo que el del dataset original, es decir, “*CC BY-SA 4.0 License*”, con lo que favoreceremos la creación de un contenido enriquecido al alcance de todo el mundo.

## CÓDIGO FUENTE Y DATASET

Sobre este mismo portal se encuentran el código (carpeta “Python”) y el dataset resultante (carpeta “csv”).

## RECURSOS CONSULTADOS

- Lawson, R. (2015, october). *Web Scraping with Python*. Birmingham: Packt Publishing Ltd.
- <https://docs.python.org/3/library/re.html> consultada por última vez el 14/04/2020.

## CUADRO RESUMEN DE LA ACTIVIDAD

Contribuciones	Firma
Investigación Previa	MMG
Redacción de las respuestas	MMG
Desarrollo del código	MMG