

Publicaciones científicas y libros de ciencia y tecnología

2 de abril de 2021

Índice

1. Presentación y objetivos	3
2. Competencias	3
3. Desarrollo	4
4. Referencias	6

1. Presentación y objetivos

En esta práctica se elabora un caso práctico orientado a aprender a identificar los datos relevantes para un proyecto analítico y usar las herramientas de extracción de datos.

Los objetivos concretos de esta práctica son:

- Aprender a aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios o multidisciplinares.
- Saber identificar los datos relevantes que su tratamiento aportan valor a una empresa y la identificación de nuevos proyectos analíticos.
- Saber identificar los datos relevantes para realizar un proyecto analítico.
- Capturar datos de diferentes fuentes de datos (tales como redes sociales, web de datos o repositorios) y mediante diferentes mecanismos (tales como queries, API y scraping).
- Actuar con los principios éticos y legales relacionados con la manipulación de datos en función del ámbito de aplicación.
- Desarrollar la capacidad de búsqueda, gestión y uso de la información y de los recursos.

2. Competencias

En esta práctica se desarrollan las siguientes competencias del Máster de Data Science:

- Capacidad de analizar un problema en el nivel de abstracción adecuado a cada situación y aplicar las habilidades y conocimientos adquiridos para abordarlo y resolverlo.
- Capacidad para aplicar las técnicas específicas de web scraping.

3. Desarrollo

1. Contexto

Cada día más personas están más interesadas en la lectura mediante contenido digital, es por esto por lo que la industria del eBook ha permitido ampliar los mercados. La industria de la investigación y la ciencia también están migrando a estas plataformas provocando una mayor difusión y cercanía con la población a temas que antes solo estaban en boca de la comunidad científica.

2. Descripción

Se ha seleccionado el sitio web de Springer, ya que ofrece una gama muy amplia libros en formato físico y digital de ciencia, tecnología, medicina y más; dado que es una de las mayores editoriales de libros y la segunda más grande a nivel mundial en publicaciones científicas.

Considerando que en este sitio web ofrece la posibilidad de adquirir libros en formato físico y digital se ha decidido analizar cuál es la diferencia de precios entre estos dos formatos he intentar conocer cuando es más conveniente comprarlos en un determinado formato. Además, se quiere analizar la frecuencia de publicaciones anuales en las diferentes temáticas que corresponden a una disciplina que contiene el portal.

3. Contenido del dataset

El dataset está basado en poder encontrar toda la información sobre un libro publicado en el sitio Springer, partiendo desde el nombre hasta el costo.

Campos del dataset:

Fec_Informacion: Fecha en la cual se ha consultado la información en el sitio web, el formato es dd/mm/yyyy hh:mm:ss

Nombre_Libro: Nombre del libro.

Autor (es): Autor (es) del libro.

Tema: Tema del que se trata el libro.

Categoría: Sección en la que aparece el libro dentro de la web.

Editorial: Editorial del libro.

ISBN: Identificador ISBN del libro

Anio_Copyright: Año del copyright del libro.

Numero_Paginas: Número de páginas del libro.

Precio_EBook: Precio en dólares del formato digital.

Precio_Hardcover: Precio en dólares del formato físico.

Precio_Softcover: Precio en dólares del formato tapa blanda.

Online: ¿Tiene acceso online?

Tipo_Formato: Identificador si el libro esta solo en formato digital (1), físico (2), online (3), 2 formatos (4) o varios formatos (5).

4. Agradecimientos

La información ha sido obtenida desde la web de Springer la cual consideramos de gran valor, ya que es que una de las más completas del ámbito científico con miles de libros publicados a la venta lo que nos ha ayudado a poder generar el dataset que responde a los cuestionamientos ya mencionados. Para la extracción de la información hemos usado el lenguaje Python combinado con técnicas de web scraping presentadas en el aula.

5. Inspiración

Hemos elegido estos datos porque responde fácilmente en qué formato de libro es más conveniente comprar de acuerdo con los precios.

Otra pregunta a responder es conocer la frecuencia de publicaciones en idioma inglés, de los tópicos que corresponden a la disciplina de Ciencias de la Computación que se han publicado entre 2015 y 2021.

6. Licencia

Se ha seleccionado la licencia MIT, ya que es una licencia permisiva breve y simple con condiciones que solo requieren la preservación de los avisos de licencia y derechos de autor. Los trabajos con licencia, las modificaciones y los trabajos más grandes pueden distribuirse bajo diferentes términos y sin código fuente.

Permisos

- Uso comercial
- Modificación
- Distribución
- Uso privado

Limitaciones

- Responsabilidad
- Garantía

4. Referencias

Springer, *About Springer* [en línea]. Actualizada: 2021. [Fecha de consulta: 02 abril de 2021]. Disponible en: <https://www.springer.com/la/about-springer>

Wikipedia, *Springer Science+Business Media* [en línea]. Actualizada: 2021. [Fecha de consulta: 02 abril de 2021]. Disponible en: https://es.wikipedia.org/wiki/Springer_Science%2BBusiness_Media