

Lección 42

Periodismo de datos y data scraping

Lección en línea

Propósito

En la presente Lección, se espera que los estudiantes pongan en práctica la metodología del periodismo de datos capturando datos abiertos de la Wikipedia que, una vez tratados, analizados y sectorizados, permitan al estudiante redactar una noticia a partir de su interpretación.

Secuencia para el aprendizaje

Conocimiento inicial (15 min)

Ampliación del conocimiento (60 min)

Transferencia del conocimiento (25 min)

Objetivos

Los estudiantes serán capaces de:

- Experimentar el proceso de análisis e interpretación de datos.
- Conocer y experimentar herramientas digitales de creación de gráficos on-line.
- Experimentar con herramientas digitales de raspado de datos.

Preparación

- Imprima una Infografía [Datos para un nuevo periodismo](#), Píldora informativa [Datos que son noticia](#), Guía de trabajo Wikipedia y el data scraping para cada estudiante.
- Asegúrese de que cada estudiante tenga su [Bitácora de trabajo](#).

Recursos

¡Atención!

Por favor, haga una copia de cada documento que planea compartir con los estudiantes.

- Infografía – [Datos para un nuevo periodismo](#)
- Píldora informativa – [Datos que son noticia](#)
- Guía de trabajo – Wikipedia y el data scraping

Vocabulario

- **Data scraping:** Extracción o captura de datos.
- **Wikipedia:** Enciclopedia libre y políglota disponible en la red.

Estrategia de aprendizaje

Conocimiento inicial (15 min)

Presentación de la Lección e infografía

Presentación de la Lección de creación de una noticia periodística basada en el raspado de datos o data scraping. Creación de grupos de trabajo y se presenta la guía de trabajo Wikipedia y el “data scraping”. Visualización en grupo de la infografía Datos para un nuevo periodismo.

Ampliación del conocimiento (60 min)

Generación de una noticia a partir del data scraping

Trabajo grupal de creación y redacción de una noticia basada en la recopilación y análisis de datos abiertos (open data). Para hacerlo, los estudiantes deben seguir los siguientes pasos:

- Selección de un tema y de los datos que usaremos para hacer una breve noticia ficticia.
- Uso de las hojas de cálculo de Google para hacer data scraping de los datos escogidos (se debe tener una cuenta de Gmail).
- Uso de la herramienta on-line <https://infogr.am/> para crear un gráfico de los datos recogidos.
- Redactar la noticia final que incluya la gráfica obtenida, un titular y varias hipótesis NO VERIFICADAS que expliquen los resultados de los datos mostrados.

Transferencia del conocimiento (25 min)

Presentación de las noticias

Presentación de las noticias generadas al resto de la clase.

Escribir en el diario y charla rápida

El acto de escribir en sus diarios sobre lo aprendido, respecto de si les pareció útil y de lo que sintieron, ayuda a sus estudiantes a fortalecer cualquier conocimiento que hayan obtenido hoy y servir como un resumen al que puedan recurrir en el futuro.

Sugerencias para el diario:

- ¿Sobre qué se trataba la Lección de hoy?
- ¿Cómo te sentiste durante la Lección?
- ¿Sabrías utilizar el data scraping para capturar datos desde internet?

Wikipedia y el data scraping

Wikipedia es la gran enciclopedia de internet y consultarla se ha convertido en un acto natural cuando se quiere saber algo, sin importar el tema ni el usuario. Desde estudiantes hasta periodistas, todos consultamos Wikipedia. Es la fuente de información digital más popular.

A pesar de que algunos ponen en duda la fiabilidad de su información por el hecho de que son los mismos usuarios quienes crean los contenidos, es una de las fuentes de información más visitadas de internet, donde podemos encontrar desde la evolución de la población mundial o los resultados de las últimas elecciones, hasta los goles que se hicieron en la liga, entre otros ejemplos. Pero es importante tener en cuenta la verificación de la información.

<https://es.wikipedia.org/wiki/Wikipedia:Verificabilidad>

Les proponemos experimentar con el periodismo de datos utilizando técnicas para capturar datos de páginas web (en este caso Wikipedia) y la herramienta infogr.am, para crear gráficos de forma rápida y sencilla, a fin de redactar una noticia corta a partir de ellos.

Cosas que deben saber

1. El data scraping (raspado de datos) con hojas de cálculo de Google

es una técnica que sirve para obtener los datos de una tabla de una página web para poder trabajar con ellos: cruzarlos, analizarlos, visualizarlos... las hojas de cálculo de Google Drive permiten aplicar esta técnica de un modo muy sencillo:

<https://support.google.com/docs/answer/3093339?hl=es&ref>

2. Infogr.am

Es una herramienta on-line para crear gráficos e infografías:

<https://infogram.com/>

Nuestra noticia

Paso 1 – Busquen información sobre un tema en Wikipedia

Deberán escoger un tema para redactar su noticia que esté basado en datos que se puedan encontrar en Wikipedia. Por ejemplo, si quieren analizar la evolución del público de la serie Lost (Perdidos), podrían acceder a:

<https://es.wikipedia.org/wiki/Lost>

Allí podrán ver las audiencias de la serie y cómo fueron evolucionando de una temporada a otra.

Temporada	Eva de emisión (ET)	Episodios	Estadísticas			Ranking	Disponibilidad en streaming
			Revisión	Revisión por episodio (Revisión/Episodio)	Revisión por hora		
Temporada 1	Septiembre, 2005	24	1.000.000	41.667	1.000.000	10	Netflix
Temporada 2	Septiembre, 2005	24	1.000.000	41.667	1.000.000	10	Netflix
Temporada 3	Septiembre, 2005	24	1.000.000	41.667	1.000.000	10	Netflix
Temporada 4	Septiembre, 2005	24	1.000.000	41.667	1.000.000	10	Netflix
Temporada 5	Septiembre, 2005	24	1.000.000	41.667	1.000.000	10	Netflix
Temporada 6	Septiembre, 2005	24	1.000.000	41.667	1.000.000	10	Netflix

Otros ejemplos de temas para analizar podrían ser: la evolución de medallas olímpicas de Chile en los últimos 50 años, cuáles han sido las películas más rentables de Pixar, los países con más Grammys latinos de la última década...

Paso 2 – Extraigan los datos en una hoja de cálculo de Google

Para capturar los datos en los que basaremos nuestra noticia (audiencias de la serie por años, medallas según la Olimpiada, recaudación por película, Grammys latinos por países...) vamos a usar el data scraping, una técnica que sirve para obtener los datos de una página web y poder trabajar con ellos: cruzarlos, analizarlos, visualizarlos... Para hacerlo usaremos una hoja de cálculo de Google, por lo que debemos iniciar sesión en una cuenta de Gmail o crear una nueva para todo el grupo.

Para capturar los datos de una tabla deben escribir en el apartado Escribir fórmula de una hoja de cálculo de Google la siguiente sintaxis:

=IMPORHTML (“páginas web” ; “table” ; número tabla)

Donde dice “página web” deben escribir el enlace completo y donde dice “número de tabla”, deben poner el número identificativo de la tabla que quieren capturar (1, 2, 3...), ya que en una misma página de Wikipedia puede haber diversas tablas. Si quisiéramos capturar las audiencias de la serie Lost por temporadas, la sintaxis completa sería:

=IMPORHTML (“https://es.wikipedia.org/wiki/Lost”; “table” ;3)

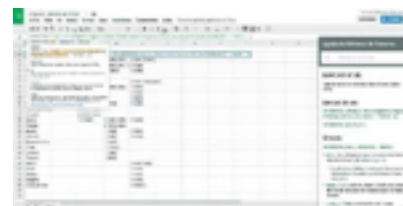
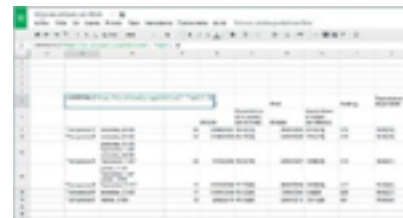
Si tienen dudas o dificultades, consulten el sistema de ayuda de las hojas de cálculo de Google o consulten el siguiente documento:

<https://support.google.com/docs/answer/3093339?hl=es&reftopic=3105411>

Una vez tengan los datos en su hoja de cálculo, limpien la tabla y dejen solamente los datos relevantes, eliminando aquellas columnas que necesiten. Para hacerlo, peguen los datos en una nueva tabla evitando repetir la sintaxis de captura de datos (ya que, si no lo hacen así, la tabla volverá a su estado original, aunque la modifiquen).

Paso 3 – Creen un gráfico en Infogr.am

Para crear un gráfico de los datos que han capturado, hagan clic en este enlace para acceder a la herramienta on-line <https://infogr.am>.



Una vez registrados, escojan el gráfico que mejor se adapte al contenido de lo que quieren mostrar. Peguen los datos de su hoja de cálculo en la tabla de Infogr.am y personalicen el gráfico resultante.

Paso 4 – Redactar un titular que resuman el gráfico

Escriban un titular para el gráfico que han creado insertando un bloque de texto en Infogr.am. El titular ha de ser breve, directo y conciso. En nuestro ejemplo de las audiencias de la serie Lost podríamos escribir:

La audiencia bajó en las últimas temporadas de “Lost”

Finalmente, pueden descargar el gráfico generado en JPG.

Paso 5 – Planteen hipótesis para explicar los hechos

A continuación, planteen diversas hipótesis que puedan explicar los hechos del gráfico que han generado. En nuestro ejemplo, podríamos argumentar que las audiencias de la serie Lost cayeron a medida que avanzaban las temporadas porque:

- Empezaron otras series de éxito
- La trama perdió interés
- Los actores populares dejaron la serie
- ...

Paso 6 – Redacten una breve noticia

Redacten una breve noticia que recoja el gráfico, el titular y una de las hipótesis planteadas en el paso anterior para explicar los hechos en un formato que les permita presentarla al resto de grupos de trabajo.

Recuerden que el periodismo real, todas las hipótesis que hicieron deberían citar las páginas y autores en los que se basan para hacer sus afirmaciones y poder comprobar así su veracidad.