



Bloc 5 • Les dades a l'entorn global

Fem

LA VIQUIPÈDIA I EL *DATA SCRAPING*



INTRODUCCIÓ:

La Viquipèdia és la gran enciclopèdia d'Internet, i consultar-la s'ha convertit en un acte natural quan es vol saber alguna cosa, sense importar ni el tema ni l'usuari. Des d'estudiants a periodistes, tots consultem la Viquipèdia. És la font d'informació digital més popular.

Tot i que hi ha qui posa en dubte la fiabilitat de la informació pel fet que són els mateixos usuaris els que creen els continguts, és una de les fonts d'informació més visitades a Internet on podem trobar des de l'evolució de la població mundial, els resultats de les darreres eleccions, o fins i tot els gols que es van marcar la darrera jornada de lliga. Però és important tenir en compte la **verificació de la informació**.

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

En aquesta activitat es proposa utilitzar tècniques per capturar dades de pàgines web, en aquest cas la Viquipèdia, així com l'eina Infogr.am per crear gràfics de manera ràpida i senzilla, amb l'objectiu de redactar després una notícia curta a partir d'ells.

OBJECTIUS:

- Analitzar les dades i saber descobrir-ne les més rellevants.
- Aprendre a utilitzar tècniques d'extracció i anàlisi de dades.
- Conèixer la millor manera de presentar la informació perquè resulti comprensible i atractiva.
- Redactar una notícia que sigui interessant.

CONCEPTES QUE HAS DE CONÈIXER:

1. *Data scraping* (raspament de dades) amb fulls de càlcul de Google

El *data scraping* és una tècnica per obtenir dades d'una pàgina web i poder treballar amb elles: crear-les, analitzar-les, visualitzar-les...

Els fulls de càlcul de Google permeten aplicar aquesta tècnica d'una manera molt senzilla.

<https://support.google.com/docs/answer/3093342/?hl=es&authuser=0>

LA VIQUIPÈDIA I EL DATA SCRAPING

2. Infogr.am

És una eina en línia per crear gràfics i infografies. <https://infogr.am/>

DESENVOLUPAMENT DE L' ACTIVITAT:

En aquesta activitat, haurem d'entrar a la Viquipèdia i buscar informació sobre algun tema que ens interessi per obtenir dades, analitzar-les, visualitzar-les i explicar-les en forma de notícia breu.

Per a aquesta pràctica, es necessita un compte de Gmail.

Els passos que hem de seguir són:

1. Buscar informació sobre un tema a la Viquipèdia
2. Extreure'n les dades en un full de càlcul de Google
3. Crear un gràfic a Infogr.am
4. Redactar un titular que resumeixi el gràfic
5. Plantejar una hipòtesi per explicar els fets
6. Redactar una notícia breu

I recorda verificar que la informació és de fonts fiables.

Per exemple, si volguessis analitzar l'evolució de l'audiència de la sèrie *Lost*:

1. Buscar informació a la Viquipèdia

Com a exemple, per analitzar l'evolució de la sèrie *Lost*, us proposem fer la prova d'importar la taula d'evolució de les audiències de la sèrie a partir de la seva entrada en Wikipedia:

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

Premios y nominaciones [editar]

Artículo principal: Anexo:Premios y nominaciones de Lost

Audiencias [editar]

En la siguiente tabla, se muestran las audiencias por temporadas (basado en una ponderación promedio total de espectadores por episodios, incluyendo las repeticiones) de la serie de ABC, *Lost*.

Temporada	Día de emisión (ET)	Episodios	Estreno		Final		Ranking	Espectadores en promedio
			Emisión	Espectadores en el estreno (en millones)	Emisión	Espectadores en la final (en millones)		
Temporada 1	miércoles, 20:00	25	23 de septiembre de 2004	18.65 ¹³	25 de mayo de 2005	20.75 ¹⁴	#15	15.69 ¹⁵
Temporada 2	miércoles, 21:00	24	21 de septiembre de 2005	23.47 ¹⁶	24 de mayo de 2006	17.84 ¹⁷	#15	15.50 ¹⁸
Temporada 3	miércoles, 21:00 (episodios 1-6) miércoles, 22:00 (episodios 7-23)	23	4 de octubre de 2006	18.82 ¹⁹	23 de mayo de 2007	13.86 ²⁰	#14	15.05 ²¹
Temporada 4	jueves, 21:00 (episodios 1-8) jueves, 22:00 (episodios 9-14)	14	31 de enero de 2008	17.77 ²²	29 de mayo de 2008	13.99 ²³	#17	13.40 ²⁴
Temporada 5	miércoles, 21:00	17	21 de enero de 2009	13.32 ²⁵	13 de mayo de 2009	9.43 ²⁶	#28	10.94 ²⁷
Temporada 6	martes, 21:00	18	2 de febrero de 2010	14.30 ²⁸	23 de mayo de 2010	13.57 ²⁹	#31	10.08 ³⁰

2. Extreure'n les dades en un full de càlcul de Google

Ara utilitzarem el data scrapping, una tècnica que serveix per obtenir les dades d'una pàgina web i poder treballar amb ells: crear-los, analitzar-los, visualitzar-los...

Obre un full de càlcul de Google (<https://docs.google.com/spreadsheets/u/0/>)

Afegeix a l'apartat de fórmula: `IMPORTHTML("https://ca.wikipedia.org/wiki/Lost"; "table" ;3)`

Explicació de la fórmula emprada: `IMPORTHTML("URL";"TABLE O LIST";NUMERO)`

- **IMPORTHTML** – Importar des de pàgina web
- **URL** – Adreça del lloc web, exemple: <https://ca.wikipedia.org/wiki/lost>
- **LIST** o **TABLE** – Escriu **LIST** si vols extreure una llista o **TABLE** si vols extreure una taula.
- **NOMBRE** - Indica el nombre d'ordre en què es troba la llista o la taula dins de la pàgina web que vols importar (En aquest cas és la tercera taula)

	Emissión	Espectadores en el estreno (en millones)	Final	Espectadores en la final (en millones)	Ranking	Espectadores en promedio		
Temporada 1	miércoles, 20:00	25	23/09/2004	18.65[13]	25/05/2005	20.75[14]	#15	15.69[15]

Utilitzar l'eina «Ajuda» en cas de dubte.

Ayuda de Editores de Documento

IMPORTHTML

Importa datos de una tabla o lista de una página HTML.

Ejemplo de uso

`IMPORTHTML("http://en.wikipedia.org/wiki/Demographics_of_India", "tabla", 4)`

`IMPORTHTML(A2, B2, C2)`

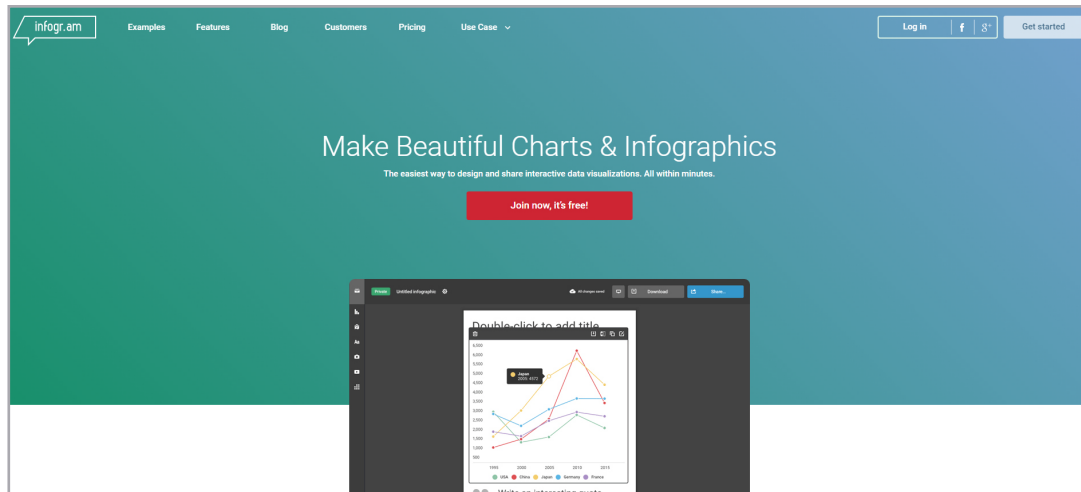
Sintaxis

`IMPORTHTML(url, consulta, indice)`

- **url**: URL de la página que se examinará, incluido el protocolo (por ejemplo, `http://`).
- El valor de **ur1** debe ir entre comillas o ser una referencia a una celda que contenga el texto adecuado.
- **consulta** puede ser "lista" o "tabla" en función

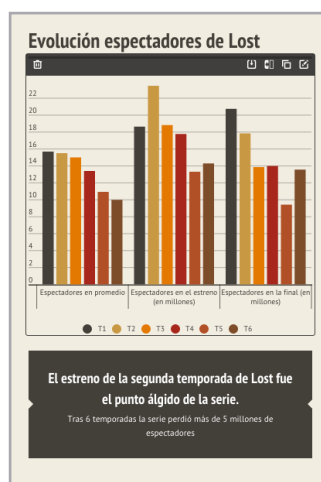
3. Crear un gràfic a Infogr.am

Fes clic en aquest enllaç per poder crear el gràfic: <https://infogr.am/>
Fes clic al gràfic que vols crear.



4. Redactar un titular que resumeixi el gràfic

<https://infogr.am/app/#/edit/a2d7dfa6-ad58-4c4f-9ce6-a7fba721ca9c>



5. Plantejar possibles hipòtesis per explicar els fets.

- Van començar altres sèries d'èxit
- La trama va perdre interès
- Els actors populars van deixar la sèrie

6. Redactar una notícia breu.

Recorda que, per redactar la notícia, has de citar les pàgines i autors en els quals et bases.