



Informe Taller 2

Parte 1 – Configuración Mongo DB Y Mozilla

The screenshot shows a Visual Studio Code editor with a terminal window open. The terminal displays MongoDB logs, including messages about saving checkpoints and session updates. Below the terminal, a Python script is visible, using BeautifulSoup to scrape HTML content from a browser page source.

• Robo 3T Inicial

The screenshot shows the Robo 3T application interface. On the left, there's a sidebar with a 'New Connection (5)' button and a list of connections: System, api-mongodb, config, and weather. The main area displays a promotional banner for Studio 3T, featuring the Robo 3T logo and a green button labeled 'Download Studio 3T'. Below the banner, there's a section titled 'Enjoy new features like:' followed by a list of features.

Need a more powerful GUI? Try Robo 3T's sibling: Studio 3T

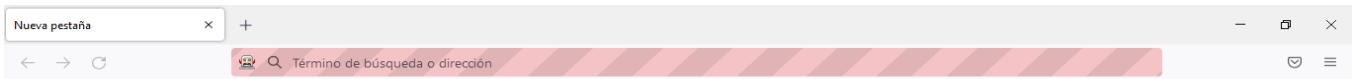
[Download Studio 3T](#)

Enjoy new features like:

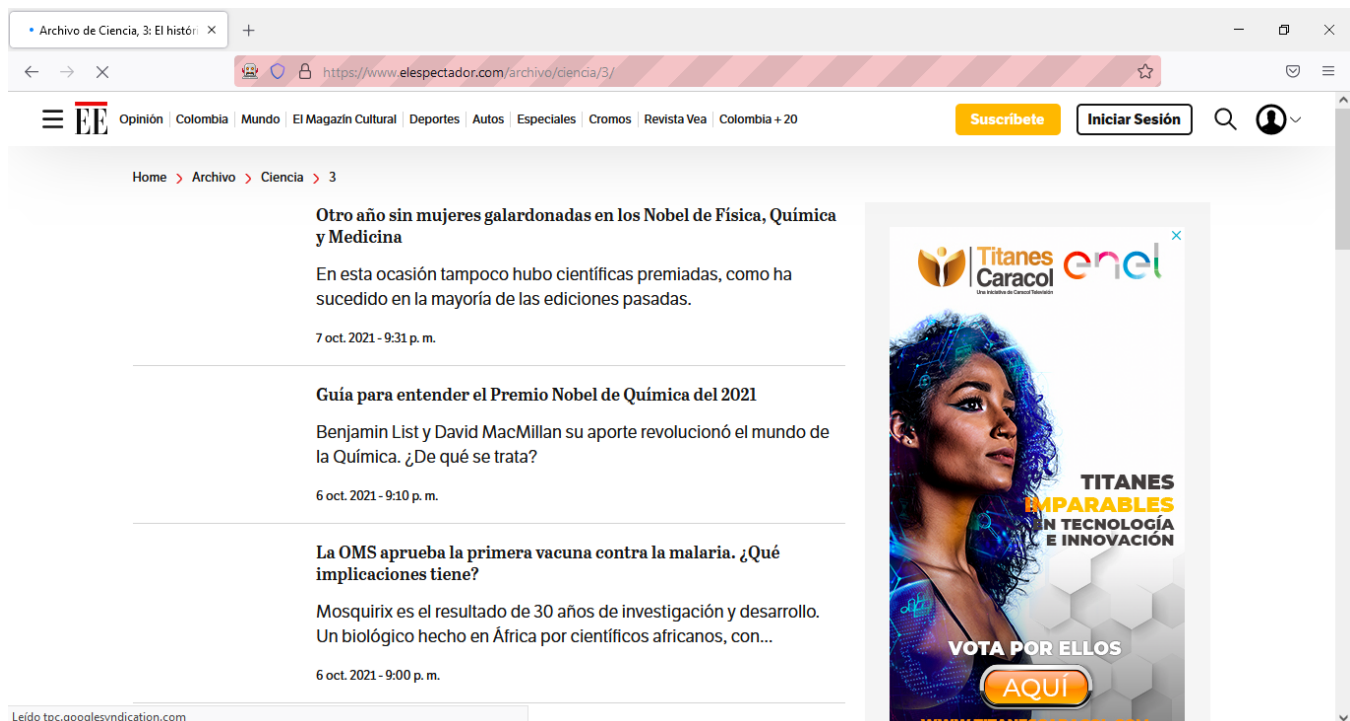
- Import and export CSV, JSON, SQL and more
- Generate flawless Javascript code from mongoshell and SQL
- Build and debug aggregation queries stage by stage
- Use Visual Query Builder to create powerful queries fast
- Write SQL queries direct to MongoDB
- Connect to as many connection presets as you like
- Explore schema stats visually and export documentation as CSV



Mozilla Inicial



- ***Se procede a realizar la descarga de las 5 categorías solicitadas, cabe resaltar que se realizó solo de 8 noticias por categoría debido a que el proceso consumía bastante maquina y los tiempos de ejecución eran muy demorados. Entre algunos ejemplos tenemos***





Archivo de Deportes, 3: El histó... X

← → X

https://www.elespectador.com/archivo/deportes/3/

Opinión Colombia Mundo El Magazín Cultural Deportes Autos Especiales Cromos Revista Vea Colombia + 20

Suscríbete Iniciar Sesión

Home > Archivo > Deportes > 3

Cuando Maradona debutó como futbolista profesional

El 20 de octubre de 1976, hace exactamente 45 años, lo hizo con la camiseta de Argentinos Juniors ante Talleres de Córdoba.

20 oct. 2021 - 7:29 p. m.

Bayern Múnich es líder solitario de su grupo en la Champions League

El equipo alemán goleó al Benfica de visita y se catapultó como líder solitario del grupo E de la Champions League.

20 oct. 2021 - 4:58 p. m.

Muriel y Zapata jugaron en derrota de Atalanta ante Manchester United

El equipo inglés protagonizó una remontada y ganó 3-2 con un tanto del portugués Cristiano Ronaldo.

20 oct. 2021 - 4:31 p. m.

Leido buy.tinypass.com

Archivo de Tecnología, 4: El histó... X

← → X

https://www.elespectador.com/archivo/tecnologia/4/

Opinión Colombia Mundo El Magazín Cultural Deportes Autos Especiales Cromos Revista Vea Colombia + 20

Suscríbete Iniciar Sesión

Home > Archivo > Tecnología > 4

Los datos de 1.500 millones de usuarios de Facebook están a la venta, alerta ESET

De acuerdo con la compañía de ciberseguridad, se trata de datos públicos recopilados por medio de scraping.

6 oct. 2021 - 6:16 p. m.

Es oficial: Windows 11 ya se puede descargar a nivel mundial

Microsoft ya inició la distribución de la nueva versión de su sistema operativo basado en la nube.

6 oct. 2021 - 3:10 p. m.

Twitch confirmó hackeo tras denuncias de masiva filtración de datos

La popular plataforma de videojuegos de Amazon, dio cuenta este miércoles de un hackeo tras la exposición en línea de datos...

6 oct. 2021 - 2:31 p. m.

vars.hotjar.com




EE Archivo de Política, 7: El históric... X

← → ↻ https://www.elspectador.com/archivo/politica/7/

EE Opinión Colombia Mundo El Magazín Cultural Deportes Autos Especiales Cromos Revista Vea Colombia + 20

Suscríbete Iniciar Sesión


Home > Archivo > Política > 7



Avanza en el Congreso proyecto para eliminar barreras en acceso a anticonceptivos

La iniciativa es de la senadora Victoria Sandino (Comunes) y busca contribuir a la garantía de los derechos sexuales y...


12 oct. 2021 - 7:37 p. m.



Gustavo Petro y su historia en la guerrilla del M-19

El candidato presidencial Gustavo Petro decidió narrar uno de los temas más controvertidos de su vida: su paso por la guerrilla...


12 oct. 2021 - 5:10 p. m.



Paz, elecciones y Venezuela, temas claves de la reunión entre Duque y Almagro

El presidente Iván Duque se reunió con Luis Almagro, secretario general de la Organización de Estados Americanos (OEA), en...

12 oct. 2021 - 4:01 p. m.



Titanes Caracol enel

Una iniciativa del Consejo Nacional

TITANES IMPARABLES EN TECNOLOGÍA E INNOVACIÓN

VOTA POR ELLOS

AQUÍ

WWW.TITANESCARACOL.COM


EE Archivo de Salud, 8: El históric... X

← → ↻ https://www.elspectador.com/archivo/salud/8/

EE Opinión Colombia Mundo El Magazín Cultural Deportes Autos Especiales Cromos Revista Vea Colombia + 20

Suscríbete Iniciar Sesión


Home > Archivo > Salud > 8



Colombia recibió 339.300 dosis de vacunas de Pfizer

Aterrizó un nuevo lote de vacunas contra el Covid-19.


29 sept. 2021 - 4:49 p. m.



El alquiler de vientres no es solo un negocio

En Colombia el alquiler de vientres necesita una legislación más precisa, aseguran investigadores de la UNAL.


29 sept. 2021 - 4:17 p. m.




En fotos: Así fue la marcha por al despenalización del aborto en Bogotá

Este 28 de septiembre se llevó a cabo la Marcha por el Aborto Libre Feminista y Antirracista en la ciudad de Bogotá.

29 sept. 2021 - 3:38 p. m.



El carbón provocará 264.900 muertes prematuras antes de 2030



Titanes Caracol enel

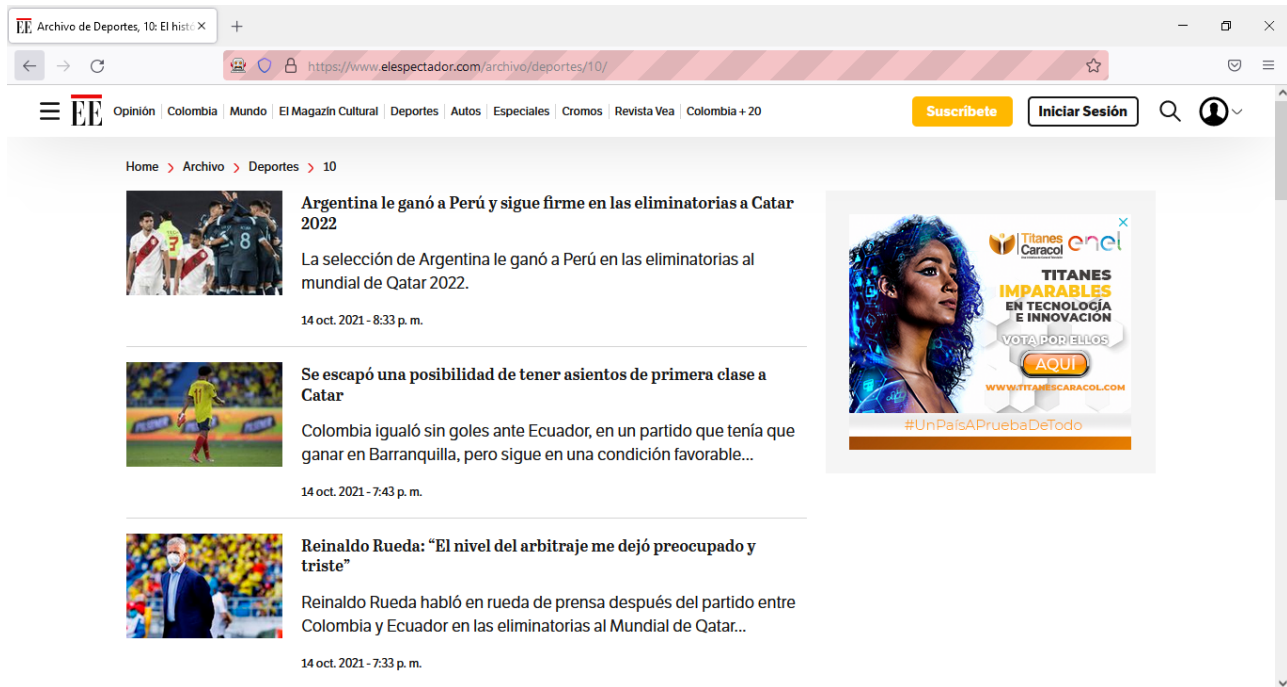
Una iniciativa del Consejo Nacional

TITANES IMPARABLES EN TECNOLOGÍA E INNOVACIÓN

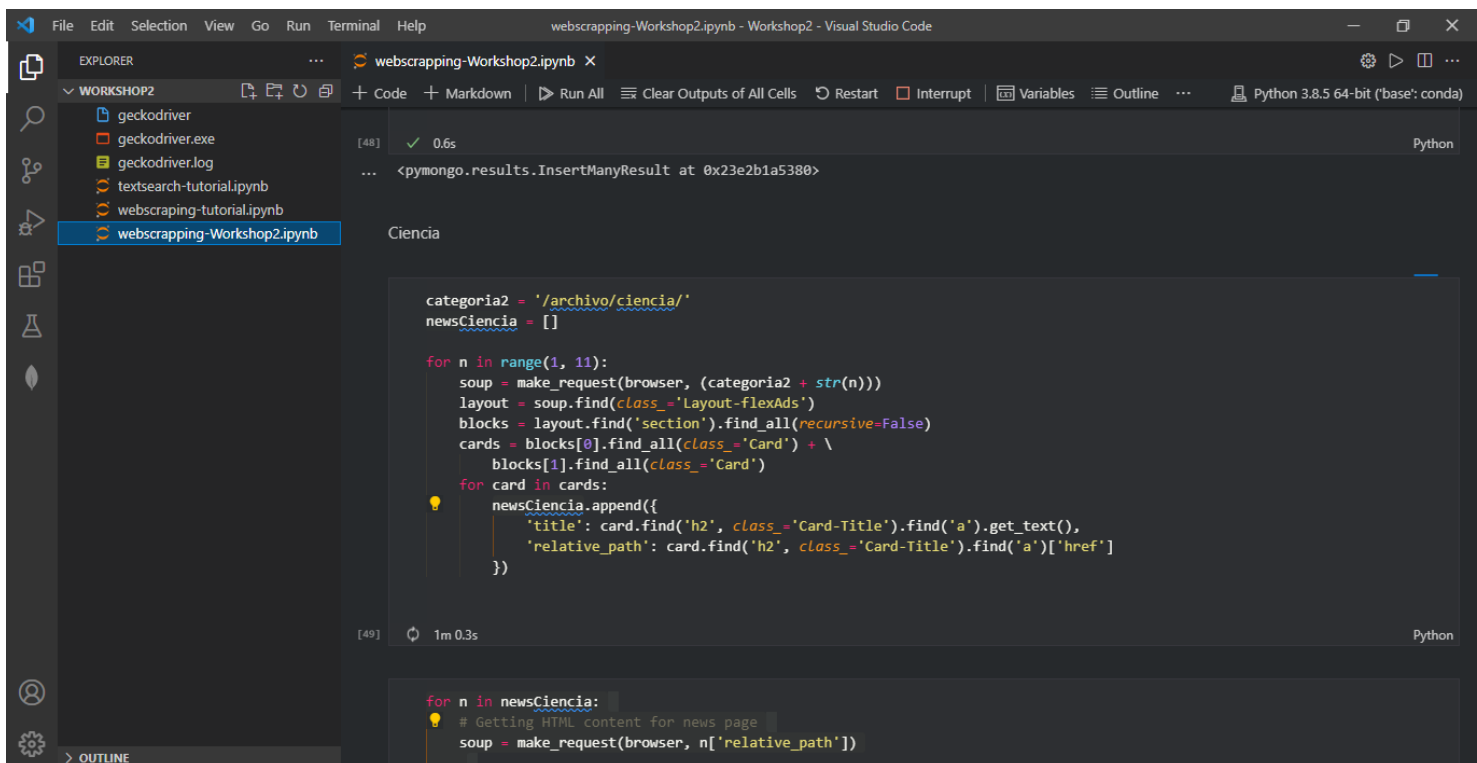
VOTA POR ELLOS

AQUÍ

WWW.TITANESCARACOL.COM



- **Método de descarga para las 5 categorías. Cabe resaltar que el proceso total fue de unos 10 minutos.**



David Leonardo Barrera
David Esteban Zamora
Materia – Big Data and Analytics
Docente Fabian Peña
2021 - II



UNIVERSIDAD
EL BOSQUE

Posterior a esto se realizo la carga de la información generada por la página del espectador.

The screenshot shows a Jupyter Notebook with the following code and output:

```
df = pd.DataFrame(newsPolitica)
df.head(8)
```

The output is a DataFrame with 8 rows and 7 columns: title, relative_path, datetime, author, summary, full_text, and _id. The data represents news articles from various sources.

	title	relative_path	datetime	author	summary	full_text	_id
0	Denuncian ante Corte Suprema a 71 congresistas...	/politica/denuncian-ante-corte-suprema-a-71-co...	22 oct. 2021 - 3:51 p. m.	Redacción Política	Los congresistas deberán responder por el deli...	Días después de que el Congreso, en el marco d...	617331302574ba43522d1110
1	Sergio Fajardo y Alejandro Gaviria: ¿debe o no...	/politica/sergio-fajardo-y-alejandro-gaviria-d...	22 oct. 2021 - 1:18 p. m.	Redacción Política	Los precandidatos presidenciales cruzaron trin...	Cada vez se hace más imperativo abordar a los ...	617331302574ba43522d1111
2	Jennifer Arias influyó en el voto de un repres...	/politica/jennifer-arias-influyo-en-el-voto-de...	22 oct. 2021 - 12:02 p. m.	Redacción Política	El representante Anatolio Hernández de la U. ...	Las redes no perdonan ni olvidan y, aunque ya ...	617331302574ba43522d1112
3	"Bases de datos del DANE no son confiables": r...	/economia/bases-de-datos-del-dane-no-son-confi...	22 oct. 2021 - 8:16 a. m.	Redacción Economía	El registrador nacional, Alexander Vega, dijo ...	Durante la entrega de tres nuevas registradur...	617331302574ba43522d1113
4	Gobierno, a rendir cuentas sobre sobre atenció...	/politica/gobierno-a-rendir-cuentas-sobre-sobr...	22 oct. 2021 - 6:00 a. m.	Redacción Política	El Congreso citó a nueve entidades del Estado ...	El próximo miércoles 27 de octubre se tiene pr...	617331302574ba43522d1114
5	También para Congreso: ya van 129 comités que ...	/politica/tambien-para-congreso-ya-van-129-com...	21 oct. 2021 - 10:27 p. m.	Redacción Política	La Registraduría precisó que no se exigirán má...	No solo es un fenómeno que se presente con mir...	617331302574ba43522d1115
6	En 2022, Gobierno podrá tomar cuentas bancaria...	/politica/en-2022-gobierno-podra-tomar-cuentas...	21 oct. 2021 - 12:58 p. m.	Redacción Política	Así lo aprobó el Congreso en el proyecto del P...	En la discusión sobre el Presupuesto General d...	617331302574ba43522d1116
7	Pandora Papers: director de la DIAN	/politica/pandora-papers-director-de-la...	21 oct. 2021 -	NaN	NaN	NaN	617331302574ba43522d1117

Se utilizo NLTK para realizar el conteo.

The screenshot shows a Jupyter Notebook with the following code and output:

```
import nltk
nltk.download('stopwords')

from nltk.corpus import stopwords

#Frequent words used in spanish
stop_words = set(stopwords.words('spanish'))
stop_words = stop_words.union(set(['aprobó', 'Congreso']))

#Print spanish stop words
print(stop_words)
```

The output is a list of Spanish stop words, including: 'estada', 'fuera', 'suyos', 'se', 'fueron', 'hubimos', 'estos', 'tuviera', 'hubiesen', 'también', 'fueseis', 'esos', 'hubiésemos', 'mucho', 'lo', 'estuvieses', 'estén', 'todo', 'contra', 'hubiese', 'sobre', 'sin', 'haya', 'tuvieses', 'tenga', 'algunas', 'muy', 'tendrías', 'ante', 'nuestro', 'esto', 'tuvieras', 'como', 'hubieras', 'tuvieses', 'vuestro', 'ese', 'tendríamos', 'somos', 'has', 'tenemos', 'habida', 'nuestra', 'estéis', 'tuvisteis', 'estuvimos', 'haya', 'le', 'otra', 'os', 'tenida', 'él', 'hubieseis', 'eso', 'nada', 'tened', 'tendría', 'esa', 'muchos', 'con', 'aprobó', 'fui', 'tiene', 'habidos', 'teníais', 'serían', 'a', 'estaríamos', 'tengáis', 'hubieran', 'una', 'estarían', 'su', 'habrías', 'habido', 'sería', 'algo', 'fuera', 'estaría', 'hubiéramos', 'estuviese', 'tuyos', 'les', 'fue', 'tienen', 'están', 'del', 'habría', 'erais', 'estuvieras', 'tengo', 'estás', 'hubisteis', 'me', 'tuvo', 'desde', 'e', 'mío', 'tienes', 'durante', 'tuyo', 'entre', 'estarás', 'esas', 'tendremos', 'mis', 'sean', 'estuvisteis', 'nosotros', 'fueses', 'los', 'más', 'mi', 'todos', 'seamos', 'estaban', 'tenías', 'mi', 'sus', 'fuésemos', 'suyas', 'un', 'te', 'estés', 'habré', 'serías', 'seas', 'sea', 'tuviéramos', 'hayáis', 'sentidos', 'estuvieseis', 'está', 'seremos', 'estuvo', 'suya', 'éramos', 'estas', 'estuvieron', 'algunos', 'nuestros', 'sentida', 'habrán', 'estaba', 'poco', 'seáis', 'suyo', 'cuando', 'tuvierais', 'tuvieron', 'hubierais', 'habíamos', 'nuestras', 'tengas', 'vosotros', 'estuvieran', 'he', 'fuerais', 'seré', 'ha', 'cual', 'o', 'fuese', 'tendrás', 'de', 'estado', 'eras', 'tenidas', 'tuvieran', 'habrían', 'ellas', 'tuvieseis', 'estaréis', 'estabas', 'estadas', 'habían', 'tú', 'estuviésemos', 'estados', 'habríais', 'tengamos', 'estemos', 'sentid', 'son', 'mía', 'estábamos', 'porque', 'habíais', 'estaremos', 'estuve', 'será', 'habríamos', 'tuviésemos', 'habremos', 'hasta', 'sintiendo', 'estuviesen', 'sois', 'es'.

[illegible]