

Desarrollo de Software

Evaluación práctica 2do Bimestre

Materia: Análisis de Datos.

Fecha: 21-agosto-2022.

Tema: Análisis y preparación de datos.

Nivel: Tercero

Instrucciones:

1. Extraer la información de la página web <https://news.ycombinator.com> según la Figura 1, donde en la parte izquierda se muestra los datos que se desea imprimir en jupyter y a la derecha el segmento HTML a los que deberá hacer referencia para la extracción de datos.



1. Coral makes comeback on Great Barrier Reef (bauckesburygazette.com.au)
93 points by WaitWaitWha 2 hours ago | hide | 21 comments

2. Old jokes (dynomight.net)
248 points by Gadiguibou 7 hours ago | hide | 176 comments

3. All about QBasic and QuickBasic (qbasic.net)
16 points by susam 1 hour ago | hide | 7 comments

4. Using grape harvest dates to estimate summer temperature over 650 years (tamino.wordpress.com)
99 points by guerby 4 hours ago | hide | 40 comments

5. The OG Social Network: Other People's Websites (jim-nielsen.com)
227 points by headalgorithm 7 hours ago | hide | 100 comments

6. Show HN: Penumbra, a perceptually optimized color palette based on natural light (github.com/nealmckee)
43 points by solaslualth 3 hours ago | hide | 16 comments

7. C99 doesn't need function bodies, or 'VLAs are Turing complete' (lemon.rip)

```
<tbody>
  <tr class="athing" id="32348567"></tr>
  <tr>
    <td colspan="2"></td>
    <td class="subtext"> == $0
    <span class="score" id="score_32348567">93 points</span>
    " by "
    <a href="user?id=WaitWaitWha" class="hnuser">WaitWaitWha</a>
    <span class="age" title="2022-08-04T21:32:26">
      <a href="item?id=32348567">2 hours ago</a>
    </span>
    <span id="unv_32348567"></span>
    " | "
    <a href="hide?id=32348567&goto=news">hide</a>
    " | "
    <a href="item?id=32348567">21&nbsp;comments</a>
  </td>
```

Figura 1. Sección de la página web a mostrar.

En la Figura 2, se muestra un ejemplo de los datos a mostrar en jupyter, la misma que refleja los campos score, número de horas y número de comentarios de la página web.

	score	horas	# comentarios
0	84	1	6 comments
1	100	5	44 comments
2	42	3	2 comments
3	26	2	14 comments
4	121	3	32 comments
5	36	4	[[discuss
6	86	8	35 comments
7	401	8	93 comments
8	247	13	113 comments
9	93	7	41 comments
10	56	6	36 comments

Figura 2. Datos a mostrar en jupyter.

2. Utilice lo datos obtenidos a manera de Dataframe y prepárelos. En la Figura 3, se muestra un ejemplo del Dataframe resultante.
- Elimine datos desconocidos en caso de tenerlos.
 - Añada una nueva columna de nombre título.
 - Añada filas con el título en la columna del punto anterior.
 - Esto debe realizarlo desde Jupyter con las diferentes funciones revisadas en clase.

id	score	horas	# comentarios	título
0	84	1	6 comments	Los 4 chiflados
1	100	5	44 comments	Los superfantásticos
2	42	3	2 comments	Mujer maravilla
3	26	2	14 comments	Comunidad de los anillos
4	121	3	32 comments	Magnificencia
5	36	4	[[discuss	
6	86	8	35 comments	Mundo de fantasía
7	401	8	93 comments	El hombre araña
8	247	13	113 comments	El perfume
9	93	7	41 comments	Cambios inaceptables
10	56	6	36 comments	Confucion

Figura 3. Ejemplo Dataframe con información lista.

3. Una vez tenga listo el Dataframe, crear una tabla que permita guardar los datos que permanecen almacenados en memoria. Asegúrese que la tabla tenga una pk, como generarla queda a criterio del estudiante.
4. Manipulación de registros de la BDD
- Ingrese en la tabla de la BDD 2 registros por teclado.
 - Actualice un registro de la BDD, ingrese los datos por teclado.
 - Elimine un registro en la BDD, acorde al dato de la pk ingresado por teclado.
 - Realice una consulta a la BDD, ingresando un dato por teclado.

RUBRICA:

1. Web Scraping (3 puntos).
2. Preparación de datos a partir de Dataframe (2,5 puntos).
3. Creación de la BDD acorde al Dataframe del punto dos (2,5 puntos).
4. Manipulación de registros de la BDD ingresando datos por teclado (3 puntos).

Tiempo: 1 hora y 30 minutos.