

Sintaxis y Semántica de los Lenguajes

Curso K2054

TRABAJO PRÁCTICO INDIVIDUAL N° 2

Web Scraping Financiero

Legajo: 163698-4

Alumno: Larramendy, Juan Camilo

Mail: jlarramendy@frba.utn.edu.ar

Repositorio:

https://github.com/larramendyjuan/ssl_k2054_larramendy

Usuario Github: larramendyjuan

Docente: Ing. Pablo Damián Mendez

Contenido

Estrategia de resolución..... 3

 Punto A 3

 Punto B 3

 Punto C 3

Casos de prueba 4

 Punto A: 4

 Punto B: 5

 Punto C: 5

Estrategia de resolución

Punto A

La estrategia utilizada fue leer línea por línea el archivo html, usando la función `fgets`, y buscar si la línea leída contiene parte de la fila que se debía tomar, usando la función `strstr`.

En primer lugar, se busca si la fila es de Contado (buscando un elemento `td` que contenga la palabra 'Cdo.'), y luego si la fila tiene variación negativa (a través del estilo `css` utilizado en la fila).

Si la línea cumple con ambos requisitos, se busca el índice del inicio del nombre de la especie (buscando la propiedad `data-order`) y luego la longitud del nombre (buscando el método `on-click`), a través de la función `get_index`. A partir de estos dos datos, se obtiene el nombre a partir de una función llamada `substr`, para obtener la subcadena. Por último, se imprime el nombre de la especie en pantalla. Se repite el proceso para todas las líneas restantes.

Punto B

Al principio de la ejecución, se imprime la cabecera en el archivo `csv` que se va a generar.

La estrategia utilizada fue leer línea por línea el archivo html, usando la función `fgets`, y buscar si la línea leída contiene parte de la fila que se debía tomar, usando la función `strstr`.

En primer lugar, se busca si la fila es de Contado (buscando un elemento `td` que contenga la palabra 'Cdo.').

Una vez en esta línea, se busca nuevamente el nombre de la misma forma que en el punto A, y luego se va buscando el fin y el comienzo de un elemento `dt`, en cada caso buscando la ocurrencia que interesa (ya que hay columnas que no se necesitan para el listado), a través de la función `get_indexes`. Una vez que se tienen los índices, se obtiene la subcadena de cada dato y luego se van imprimiendo en el archivo `csv`. Se repite el proceso para todas las líneas restantes.

Punto C

Al principio de la ejecución, se imprime la cabecera de la tabla en el archivo html que se va a generar.

La estrategia utilizada fue leer línea por línea el archivo html, usando la función `fgets`, y buscar si la línea leída contiene parte de la fila que se debía tomar, usando la función `strstr`.

En primer lugar, se busca si la fila es de Contado (buscando un elemento `td` que contenga la palabra 'Cdo.').

Una vez en esta línea, se buscan nuevamente los datos requeridos de la misma forma que en el punto B. Como se debían comparar los precios, se utilizó la función `atof` para convertir las cadenas en números de tipo `float`. Antes se utilizaron las funciones `removeChar` para eliminar el punto de la cadena (separador de millares) y `replaceChar` para reemplazar el separador decimal (coma por punto). Si el precio de compra y de venta es menor al de apertura, se imprime la especie, dentro de una fila de la tabla con el

color de fondo verde. Se repite el proceso para todas las líneas restantes. Al final se imprime el cierre de la tabla.

Casos de prueba

Para el archivo estático que se encuentra en el repositorio, llamado “**bolsar.html**”, se obtienen los siguientes resultados. En los tres casos se comprobó que los resultados son correctos.

Punto A:

```
Lista de especies con porcentaje de variacion negativo
ALUA
BBAR
BMA
BYMA
CEPU
COME
CRES
CVH
EDN
GGAL
HARG
LOMA
MIRG
PAMP
SUPV
TECO2
TGN04
TGSU2
TRAN
TXAR
VALO
YPFD
```

Punto B:

	A	B	C	D	E	F
1	Especie	Precio de cor	Precio de ver	Apertura	Precio Máxim	Precio Mínimo
2	ALUA	60,3	62,5	60	61,2	60
3	BBAR	187,35	1.400,00	193,25	193,25	187,95
4	BMA	245	302	253,5	253,7	249,35
5	BYMA	750	805	758	758	747
6	CEPU	42,1	45,7	42,5	44,05	42,1
7	COME	4,38	4,9	4,39	4,44	4,37
8	CRES	93,5	127	100	100	94
9	CVH	345	375	350	358,5	350
10	EDN	38	43	40,5	41,75	40
11	GGAL	115	155	148,9	150	145,2
12	HARG	180	183	185	185	183
13	LOMA	238	260,75	270	270	255,85
14	MIRG	2.215,00	2.228,00	2.228,00	2.228,00	2.218,00
15	PAMP	105,2	115	109,8	109,8	108
16	SUPV	50	72	69,7	70,2	69
17	TECO2	174	260	180,75	180,75	174
18	TGNO4	55,9	57,9	59	59,5	56,9
19	TGSU2	154	162,9	162,9	162,9	152
20	TRAN	26,1	37	32,85	32,85	31,5
21	TXAR	85,5	85,6	87,6	87,6	84,9
22	VALO	22,55	25,5	22,5	23,1	22,5
23	YPFD	752	805	774,25	790	770

Punto C:

Especie

HARG

LOMA

TGNO4

TXAR

Para el archivo estático que se encuentra en el repositorio, llamado “**SSL-TP2.html**”, se obtienen los siguientes resultados. En los tres casos se comprobó que los resultados son correctos.

Punto A:

```
Lista de especies con porcentaje de variacion negativo
ALUA
HARG
```

Punto B:

Especie	Precio de compra	Precio de venta	Apertura	Precio Máximo	Precio Mínimo
ALUA	65,2	72	67,8	67,8	63
BBAR	253	1.400,00	251,1	262,7	247,85
BMA	330	359,7	340	357	335,5
BYMA	843	848	820	850	820
CEPU	55,4	70	55	60	54
COME	4,33	5,4	5	5,15	4,92
CRES	98,5	105	98	99,85	96,8
CVH	409	430	405	414,5	405
EDN	54	61,5	56,2	61	56,2
GGAL	100	204,3	189	202	189
HARG	180	182	182	182	178,5
LOMA	238	274,5	253,4	257,2	252
MIRG	2.550,00	3.248,00	3.140,00	3.249,00	3.115,00
PAMP	105,2	150	130	132,5	125,5
SUPV	89,55	90,65	85	90	85
TECO2	178,8	260	184	191,4	184
TGNO4	75	82,5	78,8	85,8	77
TGSU2	184	197	184,2	195	184,2
TRAN	26,1	49,8	46	49	44,5
TXAR	89	102,5	96,2	101,75	96,2
VALO	23	26	23,6	24,35	23,6
YPFD	1	995	933,5	973,5	925

Punto C:

Se genera archivo vacío.

Especie