

División de Gestión y Difusión de la Información

XIII Jornada Expania

Santiago de Compostela, 19 de mayo de 2016

Trabajando con las utilidades de SFX

Caso práctico 3.2.4 Servicio de acceso a sumarios en Journal TOCs

José Luis Galán Cabilla

Unidad de Biblioteca

ÍNDICE

| | | |
|-------|--|-----------|
| 1 | Objetivo | 1 |
| 2 | Dificultad | 1 |
| 3 | Descripción | 1 |
| 4 | Solución | 1 |
| 4.1 | Creación de un Target Service local | 1 |
| 4.2 | Preparación del fichero de carga de los portfolios | 4 |
| 4.2.1 | Limpiar los datos | 4 |
| 4.2.2 | Obtener los ISSNs | 5 |
| 4.2.3 | Incluir estado de activación | 6 |
| 4.2.4 | Crear el texto de la nota en HTML | 6 |
| 4.2.5 | Guardar el fichero de carga en formato TSV | 8 |
| 4.3 | Carga de los portfolios en el DataLoader | 9 |
| 4.4 | Comprobación de resultados | 10 |
| 4.4.1 | Informes | 10 |
| 5 | Conclusión | 12 |

1 Objetivo

Utilizar **DataLoader** para crear un servicio de acceso a los sumarios de un grupo de revistas de un ranking por medio del **parser Generic** a nivel de **target service** e incluir código HTML para resalta campos locales.

2 Dificultad

Alta.

3 Descripción

Se acaba de publicar un artículo con un meta-ranking de revistas de economía:

Wohlrabe, Klaus (2016): *Taking the Temperature: A Meta-Ranking of Economics Journals*. CESIFO Working Paper No. 5726

En base a este ranking se quiere crear un servicio para el acceso a los sumarios de las revista en **JournalTOCs** que incluya una nota con la la posición de cada revista en el ranking. Los datos de partida de los que se dispone son los que recoge el listado que aparece en la tabla 3 del artículo (p. 24-30).

| Table 3: The Meta-ranking | | | | |
|---------------------------|----------------------------------|----|----|----|
| Rank | Journal | HM | AM | # |
| 1 | Quarterly Journal of Economics | 1 | 1 | 38 |
| 2 | Journal of Political Economy | 2 | 2 | 38 |
| 3 | Econometrica | 4 | 4 | 38 |
| 4 | American Economic Review | 3 | 3 | 38 |
| 5 | Journal of Economic Literature | 5 | 9 | 38 |
| 6 | Journal of Finance | 6 | 8 | 38 |
| 7 | Review of Economic Studies | 8 | 5 | 38 |
| 8 | Journal of Financial Economics | 7 | 12 | 38 |
| 9 | Journal of Economic Perspectives | 9 | 6 | 38 |
| 10 | Journal of Monetary Economics | 12 | 10 | 38 |

4 Solución

4.1 Creación de un Target Service local

El primer paso es crear un **Target local** para contener los *portfolios* de las revistas con los sumarios que se quieren cargar. Para ello hacer los siguiente:

- Seleccionar **Target / Add New Target** en el *KB Manager*
- Se abre la ventana **Add Target**
- Escribir **#JTOC_LCL¹** en **Public Name**
- Escribir *JournalsTOCs* en **Target**
- Seleccionar **getTOC** en **Services**
- Marcar *Active* en **Activation Status**

¹ SFX añade el sufijo **_LCL** a los target locales. Además es una buena práctica poner un sufijo para que todos los target locales se puedan localizar más fácilmente.

- Escribir *Servicio de acceso a JournalTOCs del meta-ranking de revistas de economía* en **General description**
- Seleccionar *JOURNAL* como **Target Type**
- Pulsar el botón **Submit**

Una vez creado el target hay que configurar el **Service getTOC** para que contenga los enlaces a los sumarios.

En la página [Customise de JournalTOCs](#) se puede encontrar la documentación sobre las URLs que se pueden utilizar para enlazar a sus sumarios.

I only want to be able to link the user to the JournalTOCs webpage straight from the journal in my e-journal A-Z list or from my OPAC, how do I construct a URL which will get the user to this page?
 Linking to journal TOCs hosted by us is free to anyone. You don't need to register for a customisation to link TOCs directly from your web pages. For example to link to the TOCs of PNAS you can use any of the following URLs:
<http://www.journaltoCs.ac.uk/index.php?action=tocs&journalISSN=0027-842>
<http://www.journaltoCs.ac.uk/index.php?action=tocs&journalISSN=0027842>
<http://www.journaltoCs.ac.uk/index.php?action=tocs&issn=0027842>
<http://www.journaltoCs.ac.uk/index.php?action=tocs&journalissn=0027-842>

Todas estas URLs constan de una base seguida del ISSN de la revista. Este tipo de URLs permiten la utilización del **parser Generic** a nivel del **Target Service**². Sólo hay que declarar la estructura de la URL en el **parse param** del **target service**.

Utilizando una de las bases que utiliza Journal TOCs el **parse param** sería la siguiente:

² Sobre los enlaces ver el punto 2.2 del [Manual del Taller](#).

IF(ISSN) "http://www.journaltoocs.ac.uk/index.php?action=tocs&journalISSN=" ISSN

De esta forma se está comunicando a SFX que complete el enlace añadiendo el ISSN de cada *portfolio* al final de la URL.

Para configurar el enlace en el *service* hacer lo siguiente:

- Localizar el Target, ir a Servicios y **editar el Servicio**
- Seleccionar *getTOC* como **Target Service**
- Marcar *Active* en **Activation Status**
- Escribir *Generic* en **Parser**
- Escribir en **Parse Param:**
IF(ISSN) "http://www.journaltoocs.ac.uk/index.php?action=tocs&journalISSN=" ISSN
- Pulsar el botón **Submit**

The screenshot shows the 'Edit Target Service' window in SFXAdmin. The browser address bar shows the URL: `sfx.greendata.es/sfxadmin/sfxtst41/sfxctrl.cgi?sid=70f62dc1bfd4082558995b68dc7bef41&funcio...`. The window title is 'SFXAdmin » sfxtst41 » Edit Target Service'. The interface has a red header bar with 'Main' and 'Additional details' tabs. Below this is a dark grey bar with the text 'Edit Target Service - #JTOC_LCL - getTOC'. The form contains the following fields and controls:

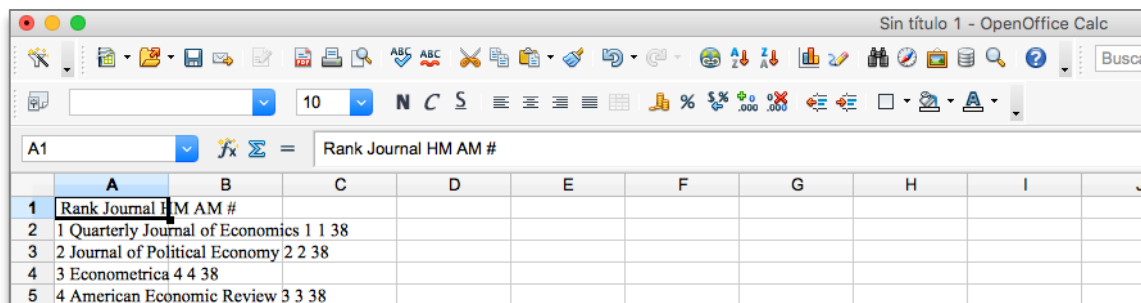
- Service:** A dropdown menu showing 'getTOC'.
- Public Name (global):** An empty text field with a '+ Add local' button.
- Description:** A large empty text area.
- Activation status:** Radio buttons for 'Active' (selected) and 'Inactive'.
- Site down:** A checkbox for 'Site down date' with a value of '0000-00-00 00:00:00' and a 'Site down reason' text field.
- Object Lookup (global):** A checked checkbox with a '+ Add local' button.
- Is free ?** An unchecked checkbox.
- Parser (global):** A dropdown menu showing 'Generic' with a '+ Add local' button.
- Parse Param (global):** A text field containing 'IF(ISSN) "http://www.journaltoocs.ac.uk/index.php?action=tocs&journalISSN=" ISSN' with a '+ Add local' button.
- AutoActive:** A text box containing the message: 'This target is not an aggregator package and the default AutoActive setting is 'No'. You can overwrite this setting by using the dropdown below.' Below this is a dropdown menu showing 'Aggregator Settings'.
- AutoUpdate:** An unchecked checkbox.
- Displayer (global):** An empty text field with a '+ Add local' button.
- Use proxy:** An unchecked checkbox.
- Crossref Supported:** A checkbox with a red question mark icon, currently unchecked.
- Threshold (global):** An empty text field with a '+ Add local' button. Below this field are 'Compose' and 'Clear' buttons.
- Authentication note:** An empty text area.
- General note:** An empty text area.
- Internal description:** An empty text area.

At the bottom of the window are two buttons: 'Close Window' on the left and 'Submit' on the right.

4.2 Preparación del fichero de carga de los portfolios

4.2.1 Limpiar los datos

La tabla con la información del ranking está en formato pdf, por ello al copiarla en la hoja de cálculo esta no reconoce los campos y pone cada fila en una única celda.



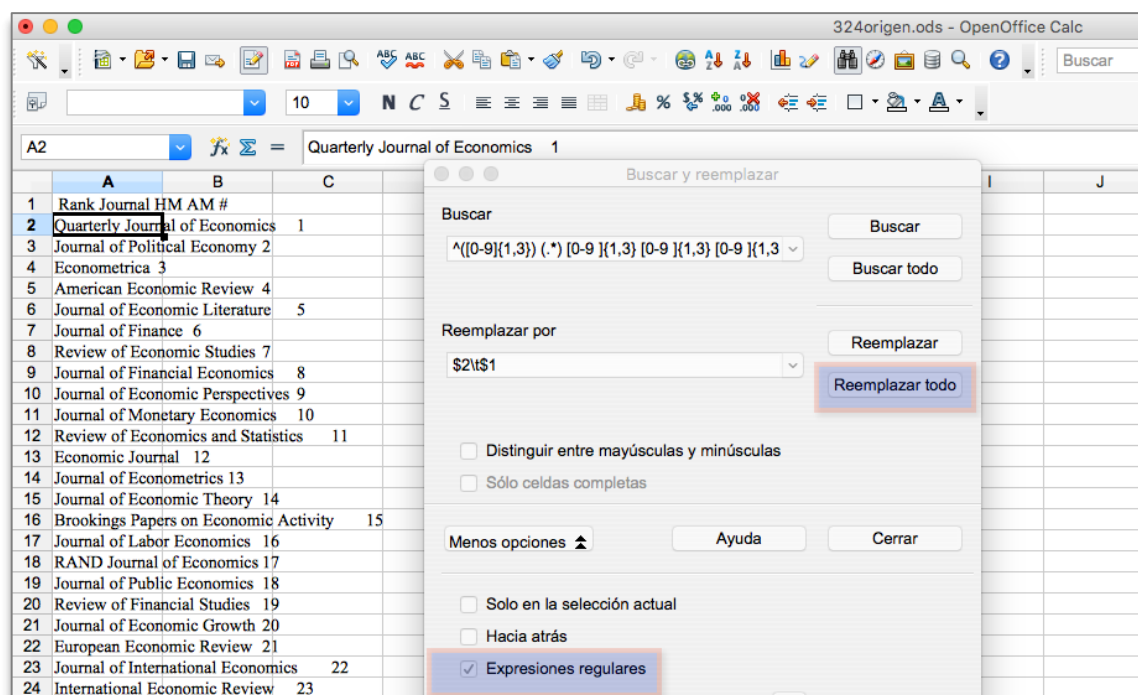
| | A | B | C | D | E | F | G | H | I | J |
|---|--------------|--------------------------------|----|---|----|---|---|---|---|---|
| 1 | Rank Journal | HM | AM | # | | | | | | |
| 2 | 1 | Quarterly Journal of Economics | 1 | 1 | 38 | | | | | |
| 3 | 2 | Journal of Political Economy | 2 | 2 | 38 | | | | | |
| 4 | 3 | Econometrica | 4 | 4 | 38 | | | | | |
| 5 | 4 | American Economic Review | 3 | 3 | 38 | | | | | |

En casos como este, para limpiar los datos se pueden utilizar **expresiones regulares** para buscar y reemplazar. *Calc* incluye esta funcionalidad. Excel no la incluye pero si la mayoría de los editores de texto (*TextPad* por ejemplo).

Las **expresiones regulares**³ son una herramienta muy potente para el tratamiento de los datos pero se requiere un buen conocimiento y práctica y mucho cuidado en su utilización para evitar desastres. Además aunque los principales meta-caracteres que se utilizan son comunes, las distintas implementaciones presentan algunas variaciones.

Para dejar solo los títulos de las revistas separados por un tabulador del número de ranking se pueden utilizar las siguientes expresiones regulares en *Calc*:

- Buscar: `^([0-9]{1,3}) (.*) [0-9]{1,3} [0-9]{1,3} [0-9]{1,3}$`
- Reemplazar: `$1\t$2`



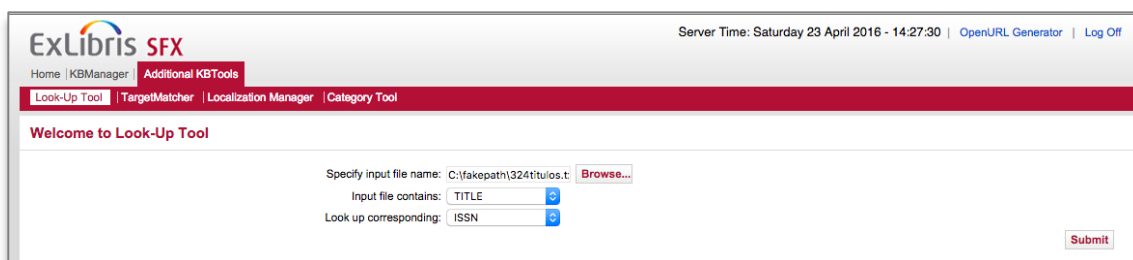
³ Ver el punto 2.5.2.2 del [Manual del Taller](#).

Para terminar de limpiar los datos hay que borrar las filas que no contiene información (cabecera, pies de páginas) y guardar el fichero en formato TSV (**324titulos.txt**)

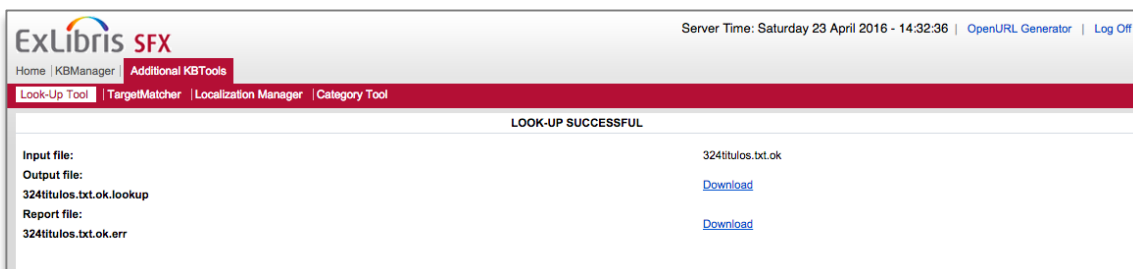
4.2.2 Obtener los ISSNs

Para poder cargar los portfolios se necesitan los ISSNs de las revistas. Se pueden obtener utilizando el fichero anterior con los títulos (**324titulos.txt**) y la herramienta **Look-Up Tool**⁴. Para ello hacer lo siguiente:

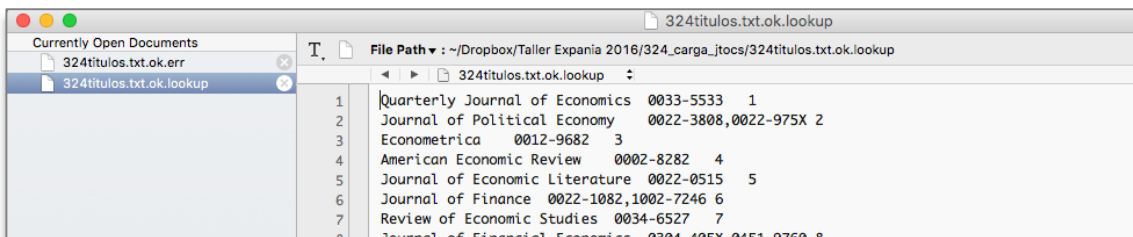
- Abrir **Look-Up Tool**.
- Buscar y selecciona **324titulos.txt** como **fichero de entrada**
- Seleccionar **TITLE** en **Input file contains**
- Seleccionar **ISSN** en **Look up corresponding**
- Pulsar el botón **Submit**



Al completar la búsqueda SFX devuelve dos ficheros:



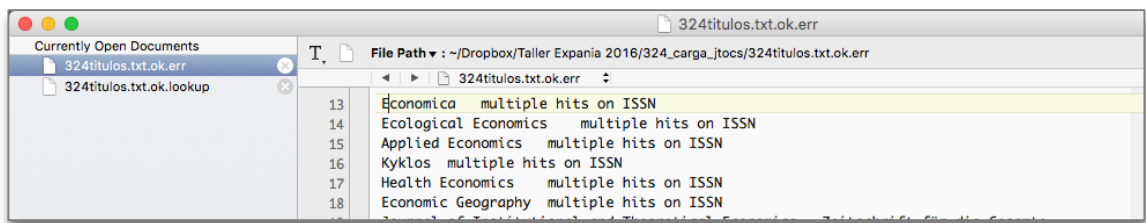
- **324titulos.txt.ok.lookup**. Contiene los títulos seguidos de los ISSNs que ha localizado SFX



| Line | Journal Title | ISSN | Count |
|------|--------------------------------|----------------------|-------|
| 1 | Quarterly Journal of Economics | 0033-5533 | 1 |
| 2 | Journal of Political Economy | 0022-3808, 0022-975X | 2 |
| 3 | Econometrica | 0012-9682 | 3 |
| 4 | American Economic Review | 0002-8282 | 4 |
| 5 | Journal of Economic Literature | 0022-0515 | 5 |
| 6 | Journal of Finance | 0022-1082, 1002-7246 | 6 |
| 7 | Review of Economic Studies | 0034-6527 | 7 |
| 8 | Journal of Financial Economics | 0304-405X, 0451-9760 | 8 |

- **324titulos.txt.ok.err**. Contiene los títulos de los que no se ha encontrado ISSNs o que tienen múltiples ISSNs

⁴ Es una de las *Additional KBTools* de SFX.



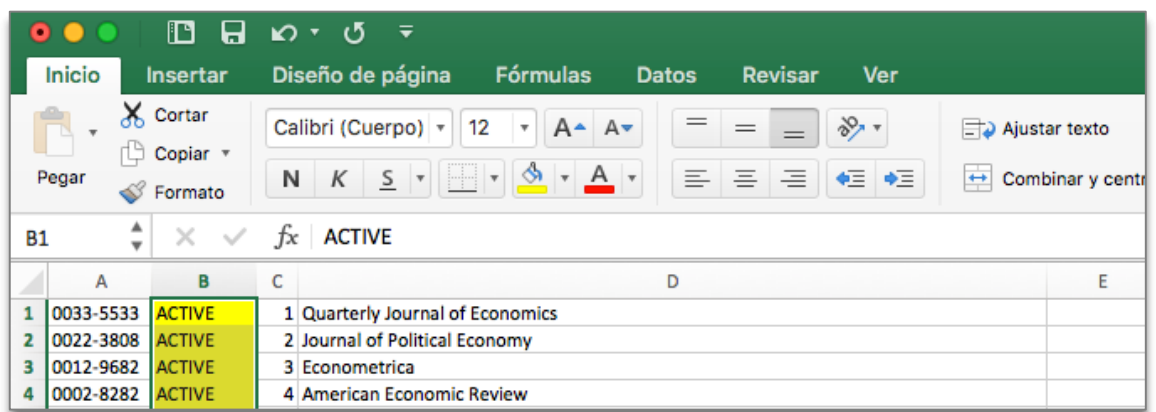
Revisando y completando los títulos de este último fichero se obtienen los datos necesarios para la carga (**324issn.txt**).

4.2.3 Incluir estado de activación

Abrir el fichero **324issn.txt** en la hoja de cálculo y guardar como **324carga** (**324carga.ods**, **324carga.xlsx**).

Mover la columna A con los títulos a la derecha de la columna B que contiene el orden en el ranking.

Insertar una columna a la derecha de la columna A y rellenarla con **ACTIVE**.



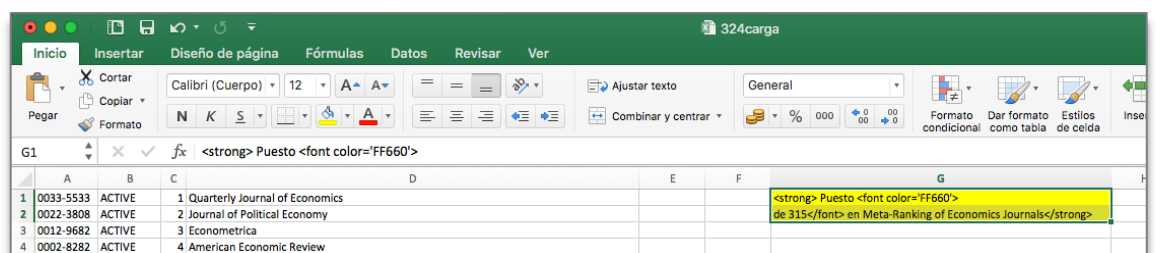
4.2.4 Crear el texto de la nota en HTML

En la columna vacía G1 poner:

Puesto

En la columna vacía G2 poner:

de 315 en Meta-Ranking of Economics Journals



En la celda vacía E1 componer la siguiente fórmula:

=**\$G\$1**&**C1**&" "&**\$G\$2**

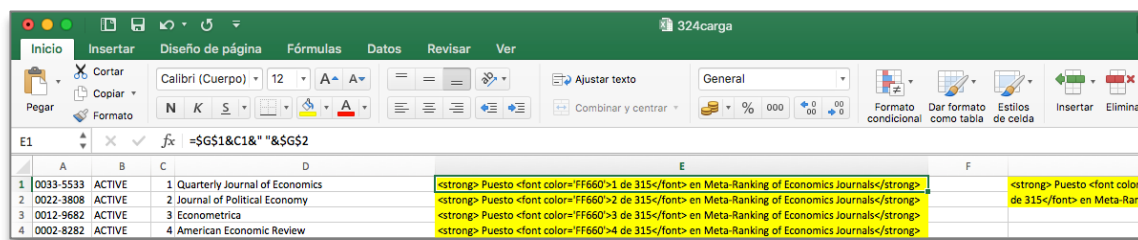
El significado de esta fórmula es el siguiente

= Signo de inicio de fórmula/función en **Calc** y en **Excel**
\$G\$1 Referencia absoluta al principio del texto en HTML
& Operador de unión
C1 Referencia relativa a la celda C1 que contiene puesto en el ranking
& Operador de unión
" " Espacio en blanco de separación
\$G\$2 Referencia absoluta a la celda G2 que contiene el fin del texto en HTML

Tras escribir la fórmula y pulsar retorno el resultado aparece en la celda E1⁵:

** Puesto 1 de 315 en Meta-Ranking of Economics Journals**

Mediante copiar o arrastrar trasladar la expresión de la celda E1 al resto de celdas de la columna E. Como la referencia en la fórmula a los puestos de la columna C es **relativa**, estos cambian con cada fila. En cambio el inicio y el fin del texto HTML es fijo ya que las referencias a la celda G1 y G2 son **absolutas**.

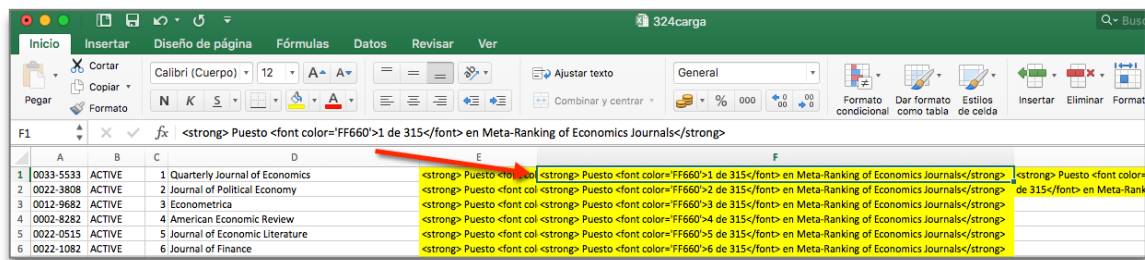


Aunque en cada celda de la columna E se muestren los resultados, lo que contiene son fórmulas tal como se muestra en la **barra de fórmulas** de la celda. Si se borra alguna de las celdas o columnas que intervienen en el cálculo de la fórmula se obtiene un error o un cambio no deseado en el resultado. Para evitarlo hay que convertir las fórmulas a valores mediante la opción de **pegado especial**:

- Seleccionar todas las celdas de la columna E y con el botón derecho del ratón seleccionar **Copiar**
- Situar en la celda F1 y con el botón derecho del ratón seleccionar **pegado especial**
- En la ventana que se abre elegir **Texto sin formato** o **Valores**

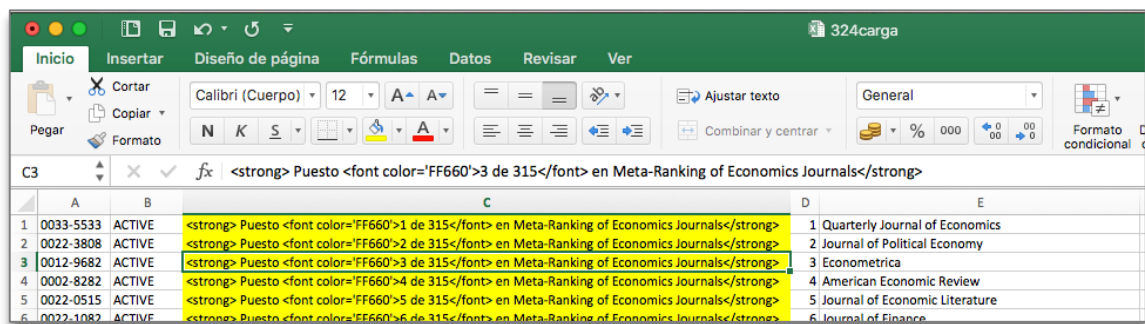
Como resultado en la columna F se tienen ahora los valores de cada celda tal como se puede apreciar también en la barra de fórmulas.

⁵ Si al pulsar retorno sigue apareciendo en la celda la fórmula en lugar del resultado esto es debido al formato de celda. Cambiar el formato de la celda a **General** y editar la fórmula quitando y volviendo a poner el signo igual del principio de la fórmula.



Ahora se puede borrar ya la columna G y E que hemos utilizado para componer el texto en HTML.

Finalmente, insertar una columna a la izquierda de la columna C y trasladar a la nueva columna C el contenido de la columna F que contiene el texto de la nota en HTML.



4.2.5 Guardar el fichero de carga en formato TSV

MUY IMPORTANTE: los nombres de los ficheros que se utilizan en el *DataLoader* no pueden contener espacio y es aconsejable utilizar sólo caracteres alfanuméricos y guiones bajos⁶.

Para guardar una hoja de *Calc* en formato TSV hacer lo siguiente:

- Seleccionar **Archivo / Guardar como**
- Seleccionar la **ruta de destino del fichero**
- Dar un **nombre al fichero que no contenga espacios**
- Seleccionar **Texto CSV**
- Marcar **Editar configuración de filtros** para poder cambiar los separadores de coma a tabuladores
- Pulsar el botón **Guardar**
- Se abre una ventana informando que se puede perder el formato al guardar en archivo de texto CSV. Pulsar en **Mantener el formato actual**
- Se abre la ventana **Exportar a un archivo de texto**
- Seleccionar **UTF-8** como **codificación**
- Seleccionar **{Tabuladores}** como **delimitador de campo**
- **Dejar en blanco** el **campo delimitador de texto**
- Pulsar el botón **Aceptar**

Para guardar una hoja de *Excel* en formato TSV hacer lo siguiente:

- Seleccionar **Archivo / Guardar como**

⁶ Además, para evitar problemas es aconsejable utilizar sólo caracteres alfanuméricos y guiones bajos.

- Seleccionar la **ruta de destino del fichero**
- Dar un **nombre al fichero que no contenga espacios**
- Seleccionar **Texto delimitado por tabulaciones** en el desplegable de **tipo de archivo**.
- Pulsar el botón **Guardar**
- Pulsar **Aceptar** en la ventana que se abre avisando que se pueden perder algunas características del libro al guardarlo como texto delimitado por tabulaciones.
- Para evitar problemas de codificación⁷ **abrir el fichero en el editor de texto y guardarlo como UTF-8**.

Fichero de carga en formato TSV: [324carga.txt](#)

4.3 Carga de los portfolios en el DataLoader

Para cargar los portfolios locales en SFX hacer lo siguiente:

- Revisar con el editor de texto que el fichero de carga está correcto
- Abrir el **DataLoader**
- Seleccionar la pestaña **Portfolios**
- Escribir **#JTOC_LCL-getTOC** en **Select Target + Service**
- Buscar y seleccionar **324carga.txt** como **fichero de carga**
- En **Specify content of the input file** seleccionar cada columna según las columnas definidas en el fichero de carga:
 - **Column 1 (Primary Key)** = Col. A = **ISSN**
 - **Column 2** = Col. B = **ACTIVATION_STATUS**
 - **Column 3** = Col. C = **GENERAL_NOTE**
- Marcar **Add Portfolios** como **Load Type**
- Pulsar el botón **Submit**

ExLibris SFX

Server Time: Saturday 23 April 2016 - 16:46:16 | [OpenURL Generator](#) | [Log Off](#)

Home | KBManager | **KBTools** | Statistics | Troubleshooting

DataLoader | Export Tool | Collection Tool

Welcome to DataLoader

Portfolios | Update Objects | Add/Remove Objects

Specify input file/target service

Select Target + Service: #JTOC_LCL-getTOC

Specify input file name: C:\fakepath\324carga.tx [Browse...](#)

Specify input file format: ☒ Dataloader TXT format ☐ KBART format

Specify content of the input file

Select the type of each column in your input file:

| Column 1 (Primary Key) | Column 2 | Column 3 |
|------------------------|-------------------|--------------|
| ISSN | ACTIVATION_STATUS | GENERAL_NOTE |
| Column 4 | Column 5 | Column 6 |
| Column 7 | Column 8 | Column 9 |

Specify load type

Load Type

- ☒ Add Portfolios
- ☐ Update Portfolios
- ☐ Remove Portfolios
- ☐ Activate target(s) / target_service(s)
- ☐ Report Mode - Do Not Update the Database

[Submit](#)

⁷ Excel guarda en codificación ANSI y en las versiones actuales no se puede cambiar la codificación al guardar.

4.4 Comprobación de resultados

Una vez terminada la carga *DataLoader* presenta los resultados. En este caso parece que todo ha sido correcto ya que se trataba de añadir 315 nuevos portfolios y esto es lo que muestra el resumen de la carga:

Summary: Added 315 portfolios, updated 0 portfolios

The screenshot shows the 'DataLoader Successful' page in the ExLibris SFX interface. The page header includes the ExLibris SFX logo and navigation links like Home, KBManager, KBTools, Statistics, and Troubleshooting. The main content area displays the following information:

- Mode:** (empty)
- Input file:** 324carga.txt
- Summary:** Added 315 portfolios, updated 0 portfolios for.
- Download links:** Download, Download, Download, Activation

Additional text at the bottom of the summary section states: 'An analysis of the data file can be found in: 324carga.txt.ok.err', 'A 'cleaned' copy of the data file can be found in: 324carga.txt.ok', 'A log of the dataloader run can be found in: dataloader.log', and 'If you want to activate added/updated target(s) target_service(s)'.

Además se puede comprobar que efectivamente se han cargado los portfolios locales consultándolos a través del target local.

The screenshot shows the 'List of Portfolios for #JTOC_LCL' page in the ExLibris SFX interface. The page header includes the ExLibris SFX logo and navigation links like Home, KBManager, KBTools, Statistics, and Troubleshooting. The main content area displays the following information:

- Available:** 315 portfolios
- Database:** sfxst41
- Table:** A table with columns: #, Service, Object, Title, Threshold, Mod., Loc., Act., Actions. It lists several portfolios, including 'The American Economic Review', 'American journal of agricultural economics', 'The American journal of economics and soc...', and 'Applied economics'.

Si se accede a los portfolios se puede comprobar la nota resaltada en HTML y el funcionamiento de los enlaces que dirigen a los sumarios en JournalTOCs⁸.

The screenshot shows the ExLibris SFX interface with the 'Fuentes' section highlighted. The 'Fuentes' section displays the following information:

- Fuente:** Applied economics [0003-6846]
- Texto completo:** Texto completo disponible vía EBSCOhost Business Source Corporate Plus
- Sumarios:** Sumario disponible en JournalTOCs Report Issues

A red arrow points from the 'Sumarios' section to the 'JournalTOCs' website. The 'JournalTOCs' website is shown with the following information:

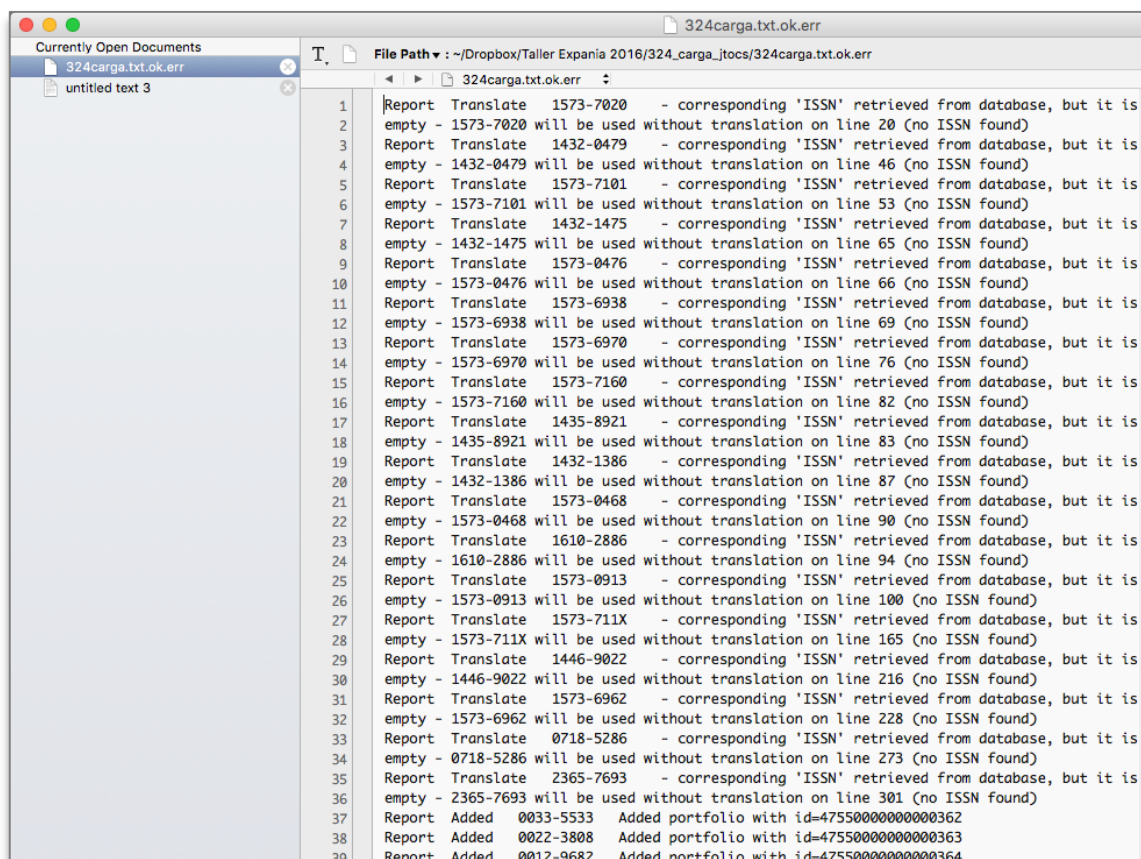
- Welcome to JournalTOCs:** Where researchers keep up-to-date
- Applied Economics:** [ISSN: 0003-6846] [34 followers] Follow
- Hybrid journal:** (It can contain Open Access articles)
- ISSN (Print):** 0003-6846 • ISSN (Online): 1466-4283
- Published by:** Taylor and Francis

4.4.1 Informes

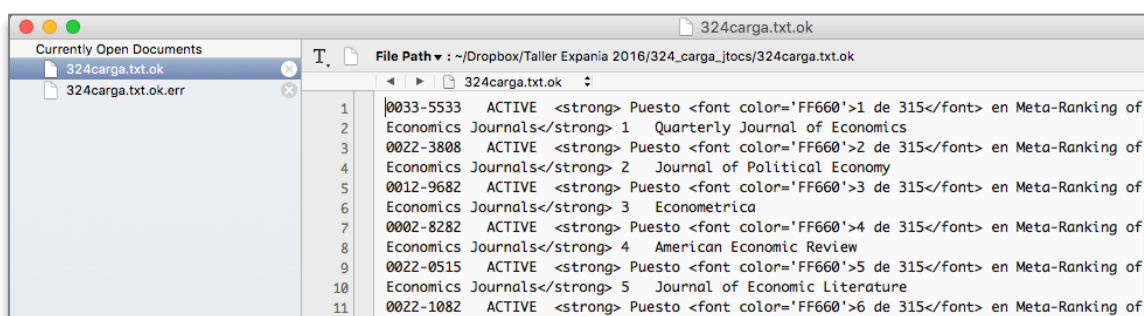
Como resultado de la carga se generan los siguientes informes:

⁸ Para garantizar bien el funcionamiento del servicio habría que revisar los enlaces para arreglar o desactivar aquellas revistas que no cuentan con sumarios en JournalTOCs.

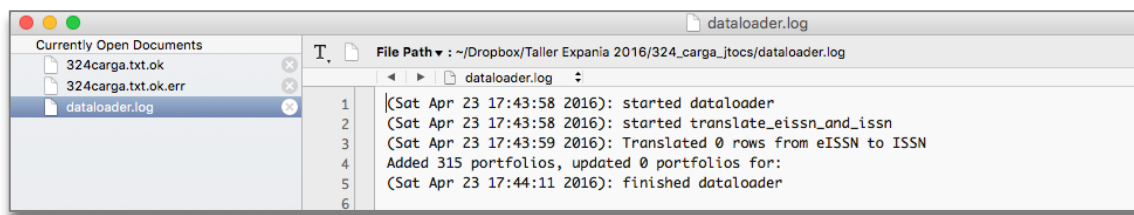
- **Análisis de los resultados de la carga:** [324carga.txt.ok.err](#). Este informe avisa de los errores, en este caso ISSNs no encontrados en SFX, y en qué línea del fichero de carga se han producido. Con esta información se puede revisar y tratar de arreglar o desactivar estos portfolios con problemas.



- **Copia del fichero de carga:** [324carga.txt.ok](#)



- **Fichero log del DataLoader:** [dataloader.log](#). Incluye información de la carga: fecha y hora de inicio, resumen de resultados y fecha y hora de finalización.



5 Conclusión

Aunque es algo laborioso, utilizando el *DataLoader* se pueden configurar servicios adicionales que conecten con otras fuentes de datos mediante el **parser Generic** si estas fuentes utilizan URLs estructuradas que permitan el enlazado modificando sólo el identificador (ISSN o ISBN normalmente) del recurso.

Un servicio con una configuración semejante a la de este caso es el de [Scimago Journal Ranking - Target de SFX](#) que aparece en el [Manualillo de SFX](#).

Además, por medio de las notas locales se puede incluir información adicional, en este caso el ranking de las revistas, y resaltarla usando código HTML.