

## Creación y personalización de una colección de documentos digitales basada en Greenstone (Parte 2)

### Creating and personalizing digital documents collections in Greenstone (Part 2)

**Jorgelina Jiménez Miranda**

Licenciada. Profesora Auxiliar. Centro de Cibernética Aplicada a la Medicina (CECAM). Universidad de Ciencias Médicas de La Habana. E-mail: [gelina@cecam.sld.cu](mailto:gelina@cecam.sld.cu)

---

#### RESUMEN

La presentación y organización de las colecciones construidas con el sistema Greenstone están dotadas de una notable flexibilidad que confiere un valor añadido al acceso a la información según los intereses, características y necesidades de los usuarios finales. En el artículo *Creación y personalización de una colección de documentos digitales basada en Greenstone*, se describió el diseño personalizado de una colección compuesta de archivos .doc y .pdf en lo relativo a mecanismos de búsqueda y navegación, por una parte, y cambios en algunos elementos de su página *Acerca de*, por la otra, los cuales se lograron a partir de disímiles cadenas de formatos correspondientes a los clasificadores y la macro extra.dm. Con el fin de profundizar en la familiarización de las ventajas de este sistema y ampliar el abanico de posibilidades que ofrece, en la segunda parte de este trabajo, se dio continuidad a la personalización de una colección similar sobre la base de la aplicación de otras cadenas de formato que operan modificaciones inmediatas en la disposición de los datos así como en la apariencia y organización de la información contenida en las diferentes páginas generadas por este sistema. Tal conjunto de posibles personalizaciones ilustra, en alguna medida, las alternativas de configuración de las salidas de esta aplicación las cuales, además, se ejecutan en forma expedita.

**Palabras clave:** Greenstone, creación de colecciones digitales, personalización de colecciones digitales, archivos .pdf, archivos .doc, cadenas de formato.

---

## ABSTRACT

The presentation and organization of collections built with Greenstone system are endowed with a remarkable flexibility adds value to information access in the interests, characteristics and needs of end users. Article Creation and customization of a collection of digital documents based on Greenstone, described the custom design consisting of a collection of files. Doc and. Pdf with regard to navigation and search mechanisms on the one hand, and changes in some elements of their About page, on the other, which were obtained from dissimilar formats for strings and macro classifiers extra.dm. To deepen the familiarity of the advantages of this system and expand the range of possibilities offered in the second part of this work, we continued to customizing a similar collection on the basis of the application of other format chains operating immediate changes in the provision of data as well as the appearance and organization of the information contained on the pages generated by this system. Such a set of possible customizations illustrates, to some extent, the alternative configuration of the outputs of this application which also run in an expeditious manner.

**Key words:** Greenstone, creating digital collections, customizing digital collections, .pdf files, .doc files, format strings.

---

## INTRODUCCIÓN

La flexibilidad, facilidad de uso, así como la libre disponibilidad de Greenstone le otorgan a este software una capacidad de adaptación a las innumerables necesidades de los usuarios en diversos escenarios y por consiguiente, dotan a esta herramienta informática de un valor inestimable para los países en vías de desarrollo.

La aplicación de cadenas de formatos diversas es una de las vías ofertadas por este sistema para lograr una configuración particular de las salidas de sus diferentes páginas (*Acerca de, índices, clasificadores y resultados de búsqueda*).

En el artículo **Creación y personalización de una colección de documentos digitales basada en Greenstone. Parte I** se reprodujo el procedimiento aplicado en una colección de documentos .doc y .pdf para configurar, en alguna medida, tales páginas mediante macros y cadenas de formato, aunque sólo se abordaron cierto número de adaptaciones por razones de espacio. Por tal motivo, en el presente trabajo, se retoma el tema para proporcionar nuevas sugerencias en cuanto a la personalización de colecciones de archivos .pdf y .doc. Por tanto, la finalidad de este artículo consiste en continuar la ejemplificación de disímiles configuraciones de las páginas de Greenstone mediante la aplicación de nuevas cadenas de formato las cuales determinarán apariencia y facilidades particulares en función de las necesidades de los usuarios de este sistema.

## DESARROLLO

El punto focal de la organización de las bibliotecas digitales gravita en torno de las colecciones cuyo propósito fundamental es el acceso a la información contenida en ellas.<sup>1</sup> En lo que al programa Greenstone concierne, éste se ha concebido para prestar auxilio a los usuarios en la organización y construcción, en forma muy rápida, de colecciones de información sencillas -por defecto-, aunque por las facilidades brindadas a través de la interfaz del usuario son sumamente configurables en múltiples niveles distintos en un breve lapso. Algunas de ellas son:<sup>2,3</sup>

- definición de índices de búsqueda puestos a disposición de los usuarios,
- estructuras de navegación adaptables a las necesidades de los usuarios,
- cambios en las macros,
- modificaciones en las cadenas de formato,
- presentación particularizada de los ítemes desde la pantalla al personalizar alguna o cada una de las páginas que Greenstone genera,
- utilización de macros dinámicas,
- personalización de la fase de "construcción" de las colecciones,
- modificaciones en las características del sistema de ejecución.

Por tanto, llegado el punto en que el usuario de Greenstone ha creado una colección por defecto, es decir, después de obtener una colección en forma básica, puede tomar la decisión de modificar la estructura lograda a partir de su construcción por omisión. Tales transformaciones se pueden concretar por distintas medios,<sup>2</sup> algunos de los cuales se esbozan a continuación:

### **Enriquecimiento de los metadatos lo cual en la práctica implica asignar valores específicos a cada uno de los metadatos**

En la concepción de las bibliotecas digitales Greenstone está implícito el empeño de añadir valor a las colecciones creadas y almacenadas en ellas. Una de los medios existentes para alcanzar esta meta radica en el enriquecimiento de las colecciones con metadatos de alta calidad sobre los cuales descansan actividades de búsqueda y navegación de manera uniforme y útil.<sup>1</sup>

Los metadatos son información adicional acerca de los documentos originales tales como los autores o creadores, título, editorial, palabras clave o descriptores, lugar y fecha de publicación, entre otras y brindan un resumen muy sucinto de cualquier recurso de información. Describen varios atributos de los objetos de información y les proveen de significado, contexto y organización.<sup>4</sup> Sirven para identificar, localizar, recuperar y acceder a esos documentos. Posibilitan una gestión eficaz y eficiente de la información, porque le confieren una organización estructurada a diferencia de Internet; por consiguiente, la descripción normalizada de los documentos o recursos electrónicos mediante conjuntos de metadatos constituye una solución para el mejoramiento de la relevancia o pertinencia de los resultados en el proceso de búsqueda y recuperación en Internet o en bibliotecas digitales.<sup>5</sup> En Greenstone hay disponibles diferentes conjuntos de metadatos: Dublin Core (dc.) de cobertura internacional, Development Library Subset Example Metadata (dls.) y Extracted Greenstone Metadata (ex.). Otros pueden crearse también a la medida de las necesidades.

Entre los conjuntos de metadatos citados, el Dublin Core se considera el más interesante pues responde a la norma ISO Conjunto de Metadatos Dublin Core, ampliamente aceptada como referente a nivel mundial.

Para organizar una biblioteca digital Greenstone, es necesario la intervención de los metadatos que son información estructurada referente a los recursos de información, la cual se refleja en la presentación y apariencia de la interfaz del usuario.<sup>2</sup> Los metadatos son la materia prima de las colecciones que permite configurar las distintas salidas del sistema por lo que puede afirmarse que la característica definitoria de las bibliotecas digitales es el uso de los metadatos en calidad de materia prima. La asignación de metadatos en Greenstone resulta decisiva y está asociada a la elección concreta de los índices de búsqueda y de clasificadores contemplados en las colecciones, así como otras facilidades.<sup>2</sup>

En lo que a metadatos se refiere, es conveniente subrayar que la arquitectura de Greenstone comprende a un editor de metadatos denominado Greenstone Editor Metadata Sets, una aplicación Java que facilita la creación de conjunto de metadatos definidos por el usuario a fin de que, con posterioridad, estén disponibles para su asignación a las colecciones.<sup>6</sup>

El panel *Enriquecer* es la opción del *Greenstone Librarian interface (GLI)* que ofrece la posibilidad de asignar metadatos a los documentos incluidos en las colecciones.

### **Asignación de conectores según el formato de los archivos originales procesados y configuración subsiguiente cuando se precise**

Los conectores son pequeños programas escritos en lenguaje de programación Perl que transforman los documentos originales (.pdf, .doc, .ppt u otros) a un formato normalizado XML y codificados en Unicode lo que asegura la capacidad de la aplicación para tratar cualquier idioma. Todos proceden de un conector básico denominado BasPlugin que realiza las operaciones necesarias para crear un nuevo documento en el formato de archivo Greenstone con el cual trabajará el sistema.<sup>7,8</sup>

Por tanto, es posible plantear que los conectores analizan los documentos importados y extraen los metadatos y debido a ello, en el diseño de índices o clasificadores hay una asociación entre ambos. Por ejemplo, el conector HTML convierte los metadatos que están implícitos en el formato del clasificador de títulos entre etiquetas <title></title>.<sup>8</sup>

La selección y configuración de los conectores desempeñan un papel clave durante el procesamiento de los documentos incluidos en las colecciones a tal punto que si estos procesos no son adecuados algunos documentos pueden ser rechazados. Asimismo, el orden en que se cargan los conectores igualmente puede originar problemas. Algunos autores recomiendan situar GreenstonePlugin y HTMLPlugin en el encabezamiento de la lista.<sup>6</sup>

Conviene subrayar que su revisión detenida reviste interés en lo que a los conectores heredados se refiere, pues en algunos casos, sirven de guía para configurar los metadatos asignados que permanecerán de manera definitiva en las colecciones dado que pueden sugerir los plugins y argumentos con los cuales funciona, correctamente, tal colección.

Desde el *GLI*, panel *Diseño*, opción *Plugins de documentos*, es posible asignar los conectores (plugins) a las colecciones Greenstone.

### **Determinación de los índices de los formularios de búsqueda que también se representan mediante metadatos de acuerdo con las necesidades de los usuarios**

Los índices de búsqueda figuran entre los mecanismos disponibles en Greenstone para recuperar, de manera muy rápida, información en volúmenes potencialmente elevados.

Estos índices especifican cuáles son las partes de las colecciones por las que pueden efectuarse las búsquedas o consultas. Se materializan en los formularios de búsqueda simple o avanzada, entre otros, y se determinan mediante la asignación de los metadatos *título*, **autores**, **fechas**, **materias**, **organizaciones** o también se definen a partir de determinado nivel, es decir, permiten realizar las consultas a texto completo, a nivel de capítulos, de párrafo o incluso de los títulos de las secciones. Igualmente, se pueden configurar índices que permitan buscar por varios campos a la vez, por frases largas o párrafos completos. Tal como se explicó al iniciar este tópico, la personalización lograda se debe a la asociación con metadatos.<sup>9</sup> Reviste importancia destacar que la búsqueda a texto completo en todos los documentos de una colección es una de las facilidades básicas que, por defecto, suministra Greenstone en todas las colecciones:<sup>9</sup> este tipo de búsqueda se ejecuta en forma muy eficiente gracias a los potentes motores (*MG*, *MGPP* y *Lucene*) disponibles en este sistema.

Para diseñar los formularios de búsqueda de las colecciones Greenstone, se precisa utilizar el panel *Diseño*, opción *Índices de búsqueda*.

### **Determinación de los tipos de clasificadores necesarios para crear las estructuras de navegación que permitirán la exploración de las listas jerárquicas generadas según los metadatos asignados y las necesidades de los usuarios**

Entre las funciones de búsqueda de Greenstone figuran los clasificadores. Estos son funciones que ofrecen una ordenación clasificada de todos los documentos comprendidos en una colección de acuerdo con criterios establecidos en el momento de construcción de ésta.<sup>6</sup> Tal como sucede con los índices, los clasificadores pueden organizarse según el título, los autores, palabras clave, fechas, o cualquier otro metadato por el cual se haya determinado su diseño.<sup>6</sup>

En las colecciones de Greenstone pueden apreciarse clasificadores que cubren una cierta gama, a saber, los mencionados en el párrafo anterior y también los clasificadores, *Phind*, *Navegar*, *Collage* y *Hierarchy*.

En este caso, también los metadatos desempeñan un papel relevante en la configuración de los clasificadores.

De igual modo, desde el panel *Diseño*, pero con la opción *Clasificadores de exploración*, se configuran las colecciones en cuanto a mecanismos de navegación.

## Configuración de las cadenas de formatos obtenidas tras la creación de la colección por defecto con vistas a que la visualización de los mecanismos de búsqueda y exploración obedezcan a las características deseadas por el usuario final

Otro de los factores imprescindibles a la personalización de las colecciones consiste en la utilización de las cadenas de formato (instrucciones o comandos); éstas controlan la apariencia de gran número de aspectos de las páginas.<sup>8</sup> Esos aspectos se pueden agrupar en 2 tipos distintos:

- 1- Elementos de la página que muestran documentos o partes de documentos.
- 2- Listas generadas por los clasificadores y formularios de búsquedas.

Las cadenas de formato son la vía seleccionada para la ejercitación en el presente trabajo.

Al experimentar con las cadenas de formato, durante el proceso de configuración de una colección, los cambios se suceden rápida y sencillamente puesto que estas cadenas se interpretan en el mismo instante en que se visualizan las páginas, es decir, tales páginas no se encuentran almacenadas previamente en Greenstone, sino que se generan a la hora en que se crean, configuran o modifican las cadenas de formato<sup>8</sup> y se muestran en pantalla las salidas del programa.

Tal hecho demuestra cuán potentes son las cadenas de formato, aunque por otra parte, son complejas y difíciles de aprender.<sup>10</sup> En lo fundamental son código Html. Constan de varios elementos (Cuadro 1):

Funciones	GL1	GL2	GL3	GL4	Search	Document
Navigationbar						
Componentes	VList		Hlist			
Variables	[link]	[dls .Title]	[ex. Source]			
Código Html	<td>					
Condiciones	{ If }	{Or}				

### Significado de algunos de sus elementos

**(text)** Muestra el texto del documento.

**(parent)** Se refiere al documento "padre".

**(icon)** Ícono de carpeta o página Web

**(srcicon)** Ícono de documento en formato Word o Pdf

**(link) (icon) (/link)** Enlace al ícono del documento Html. La expresión link denota un documento Html.

**(link) (Title) (/link)** Enlace con el documento en su versión Html, donde el hipervínculo se crea a partir del título del documento.

**(scrlink)(scricon)(/scrkink)** Crea un enlace etiquetado por el ícono estándar del documento original (Word o Pdf) y no por el título del documento.

**{Or} (dls. Title),(dc. Title), (ex.Title), Untitled** Muestra el que esté disponible de los 3, o sea, entre los metadatos dls., dc. o ex. y enlaza con el título del que esté disponible, o sea, el título del conjunto de metadatos que se haya utilizado para crear o configurar la colección.

En las cadenas de formato, estos elementos se combinan para conformar las órdenes o comandos que el sistema ejecutará.

Como se conoce, Greenstone es capaz de procesar un considerable número de formatos de gran diversidad. Ahora bien, durante el procesamiento, los archivos de tales formatos - .doc, .pdf u otro-, se convierten a Html mientras el archivo original se almacena como un archivo asociado. Debido a ello, en las diversas pantallas por defecto de Greenstone (clasificadores, resultados de búsqueda), se observan 2 íconos (Html y el ícono del documento original desde donde se hipervincula el texto (Figura 1).



**Figura 1.** Visualización de dos íconos en la salida de una colección de archivos.doc y .pdf

En cambio, si tan solo se desea visualizar el original, se debe suprimir la línea correspondiente:

**<td valign="top">(link)(icon)(/link)</td>**

Si se examina la cadena de formato por defecto de Greenstone en el párrafo siguiente resaltado en azul, es posible comprobar su complejidad, sin embargo, ese elevado nivel de complejidad resulta necesario para formular una expresión razonable en cualquier condición, o sea, que resulte válida para todos los casos y además permita mantener la compatibilidad con condiciones heredadas de versiones anteriores del programa Greenstone.<sup>6</sup>

Para trabajar en función de las cadenas de formato, se debe elegir el panel *Formato*, opción *Formatos*. Como primer paso, cuando se selecciona la función *Todas las características* conjuntamente con el componente afectado *VList*, se visualiza la siguiente cadena (Cuadro 2).

```
<td valign="top">[link][icon][/link]</td>
<td
  valign="top">[ex.srclink]{Or}{[ex.thumbicon],[ex.srcicon]}[ex./srclink]</td>
<td valign="top">[highlight]
  {Or}{[dc.Title],[exp.Title],[ex.Title],Untitled}
  [/highlight]{If}{[ex.Source],<br><i>([ex.Source])</i>}</td>
```



Hechas las precisiones necesarias en lo concerniente a los aspectos teóricos básicos previstos para este tema, se está en condiciones de acometer el ejercicio de personalización de una colección Greenstone de archivos .pdf y .doc mediante las cadenas de formato, luego de una construcción por defecto; tal ejercitación es el objetivo trazado en este artículo el cual contribuirá a la profundización del conocimiento de esta aplicación informática y de las facilidades que brinda en cuanto al acceso y organización de la información.

### **Recomendación**

En sentido general, resulta conveniente realizar determinadas comprobaciones antes de integrar a los documentos en su correspondiente carpeta mediante el panel *Reunir*. Éstas son:

- Asegurarse de que los archivos no estén protegidos por ninguna contraseña de seguridad.
- Verificar que su estructura sea correcta, por ejemplo, la presencia de tabla de contenido bien jerarquizada mediante los marcadores (bookmarks) en el caso de los .pdf.
- Comprobar que todas las fuentes estén incrustadas en el caso de los .pdf.

### **Construcción básica de la colección**

Considerando la explicación detallada de la construcción básica en Greenstone de una colección de documentos en formato .pdf y .doc en la primera parte de este tema, esa fase del trabajo se presenta en forma resumida a fin de recordar solamente ciertos puntos importantes, retomados para dar continuidad a la modificación de la colección.

Como resultado de la construcción básica de tal colección, se logró un diseño inapropiado para los fines perseguidos dado que se obtuvieron:

1. *Índices* que contemplaban la búsqueda a texto completo, por el metadato *título* y por el metadato *fuentes*. Estos 2 últimos son heredados del documento original y no satisfacen las necesidades requeridas. Los metadatos heredados resultan baratos porque el sistema los extrae automática y rápidamente de los documentos, sin embargo, no son confiables en el sentido de que no resuelven todas las necesidades ni cumplen requisitos más ambiciosos. Por consiguiente, deben asignarse metadatos manualmente, acción que torna confiables los resultados, pero encarece la operación. Ello significa que la mejor opción es intermedia, por tanto, se crea la colección en forma básica y luego se mejora mediante la asignación de metadatos Dublin Core de manera manual.<sup>11,12</sup>

Las imágenes siguientes muestran los índices logrados por omisión que en una etapa posterior serán objeto de personalización (Figuras 2 y 3):





Figura 2. Índice *títulos* por defecto de la colección.



Figura 3. Índice *fuentes* por defecto de la colección.

2. *Clasificadores de título y fuente*, solamente, los cuales se consideran insuficientes para la navegación de las colecciones de modo tal que aporten la información demandada por los usuarios. En un momento posterior, se eliminaron éstos para añadir otros enriquecidos por metadatos. Mediante las figuras 4 y 5, es posible observar el ícono que enlaza con la versión html al cual se elimina por su deficiente calidad (en ocasiones) y el nombre del archivo, dato insustancial a los efectos de la explotación de la colección y por demás antiestético.



Figura 4. Clasificador CL1 de título por defecto desplegado.

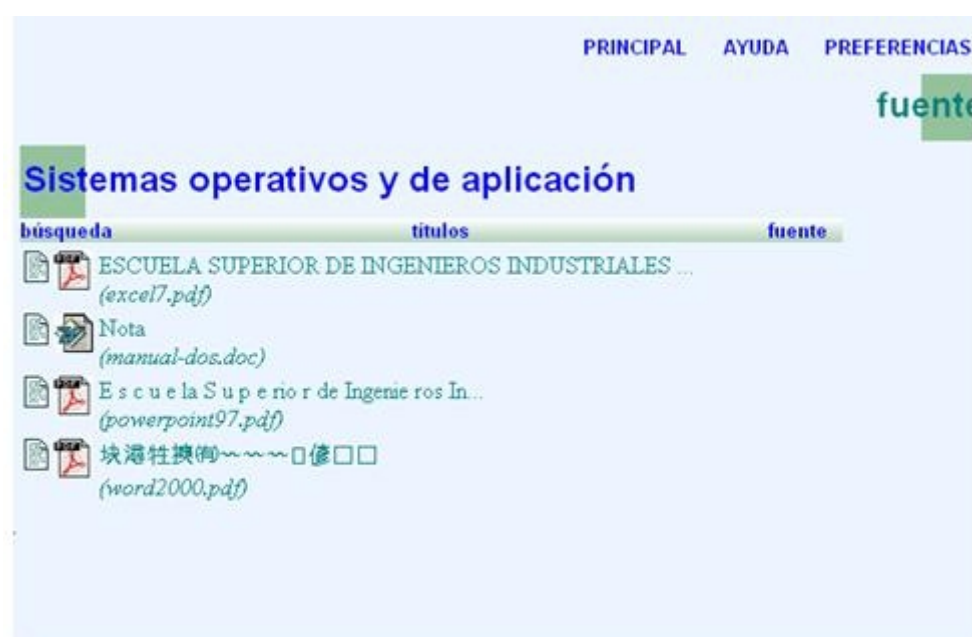


Figura 5. Clasificador CL2 de fuente por defecto desplegado.

Estas figuras visualizan los clasificadores *títulos* y *fuentes* desplegados. En ellas se aprecian, con claridad, errores en el título del último documento y en los anteriores, ausencia de uniformidad en la presentación de los datos debido a la insuficiencia de la construcción básica de la colección.

### Reconstrucción de la colección a partir de nuevos índices y clasificadores enriquecidos con metadatos

Con vistas a modificar el diseño de la colección desde el *GLI*, se decidió establecer índices de búsqueda por *autor*, *título*, *palabras clave* y a nivel de *capítulo* y *texto completo*. En cuanto a los clasificadores, se eligieron de *títulos*, *autores*, *palabras clave* y *fechas* con listas *VList*, *AZCompact List* y *Date List* respectivamente. Con la

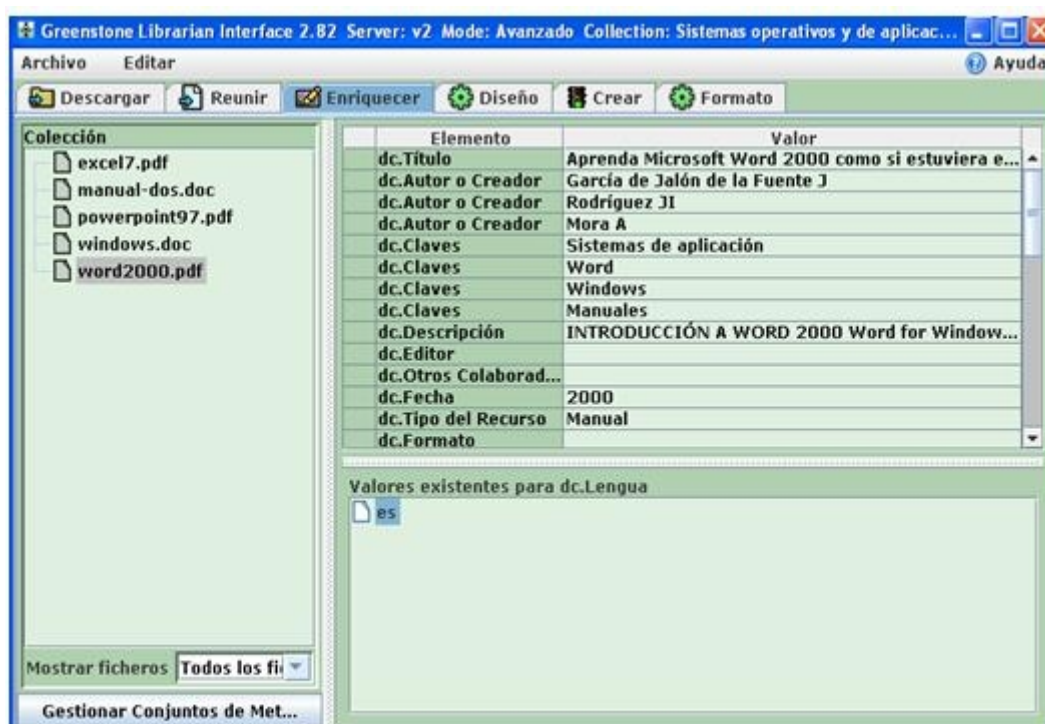
intención de reproducir los pasos acometidos, se utilizaron los paneles *Enriquecer*, *Diseño* y *Formato* con las opciones convenientes, explicadas más adelante. Los detalles correspondientes se exponen a continuación mediante los procedimientos abreviados puesto que en la primera parte de este artículo se abordaron paso a paso:

Desde el panel *Enriquecer*, se asignaron los valores de cada metadato comprendido en el conjunto Dublin Core necesarios para esta colección. Los metadatos seleccionados estaban relacionados con los tipos de índices y clasificadores que se deseaba visualizar.

Durante este paso, se recomienda registrar los datos completos y en forma adecuada, sobre todo, no olvidar consignar todos los autores y todas las palabras clave pese a la laboriosidad requerida, pues resultan imprescindibles a los efectos de la calidad de la colección. Omitir datos o teclearlos con descuido implica pérdida de información o de calidad.

El llenado de los registros en cuanto a reglas de entrada de los autores, notación de las palabras clave, fecha e idioma dependen de las políticas y normas implantadas en cada institución.

En la figura 6 se pueden apreciar los correspondientes a uno de los documentos procesados.



**Figura 6.** Asignación de metadatos Dublin Core a un documento de la colección mediante el panel Enriquecer.

Desde el panel *Diseño*, opción *Plugins de documento*, se eliminaron los conectores innecesarios para esta colección y se mantuvieron los imprescindibles a su buen funcionamiento, por ello, se conservaron únicamente, *GreenstonePlugin*, *HTMLPlugin*, *WordPlugin* y *PDFPlugin*, además de los ubicados debajo de la línea divisoria que nunca deben eliminarse tal como se presentan en la figura 7 que son *MetadataXMLPlugin*, *ArchiveInfPlugin* y *DirectoryPlugin*.



**Figura 7.** Plugins o conectores asignados manualmente a la colección.

Aunque la presencia de todos los conectores no afecta el diseño de la colección, se recomienda eliminar aquellos que no resultan indispensables dado que la aplicación puede trabajar mejor si se adjudican tan solo los requeridos. Por consiguiente, aunque no es estrictamente necesario retirar los conectores redundantes, sí constituye una buena práctica incluir, solamente, los que se necesiten y con ello evitar la inclusión accidental de documentos dispersos.<sup>12</sup>

Desde el panel *Diseño*, opción *Índices de búsqueda*, se suprimieron los índices contruidos en forma básica *title* y *source* y se mantuvo el *text*. Este último no debe suprimirse nunca pues determina las búsquedas a texto completo y debe figurar en primer lugar siempre en la relación de conectores (plugins) asignados. En el lugar de los reemplazados, se adicionaron otros 3: *títulos*, *autores* y *palabras clave*.

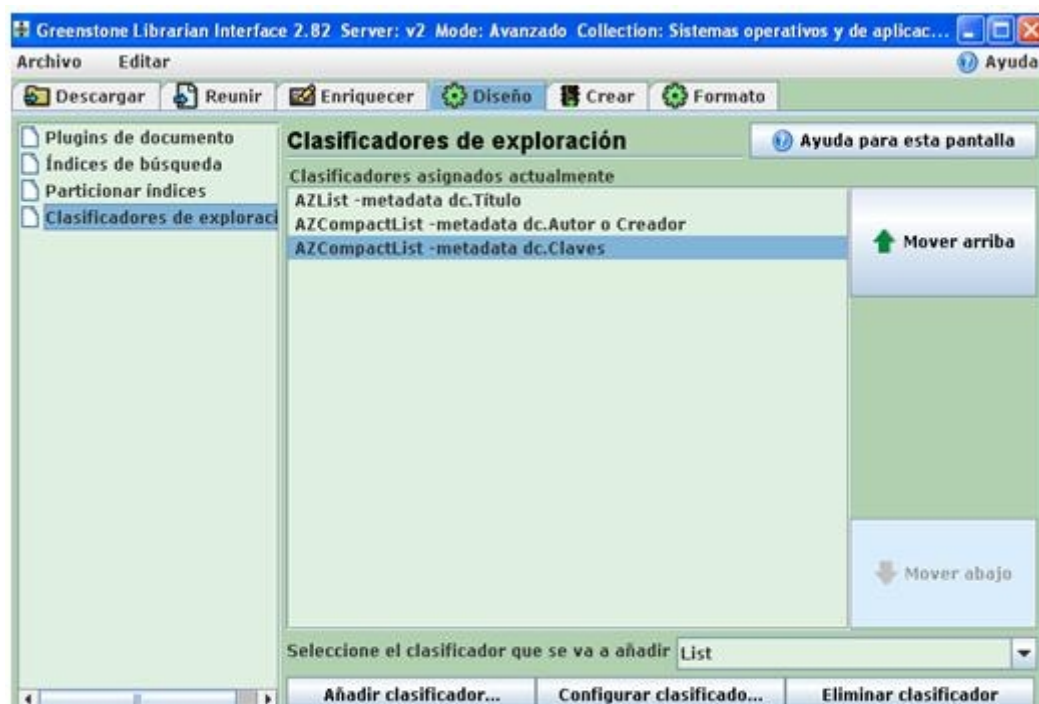
La figura 8 permite apreciar tal configuración:





**Figura 8.** Asignación de índices de búsqueda manualmente sobre la base de metadatos Dublin Core.

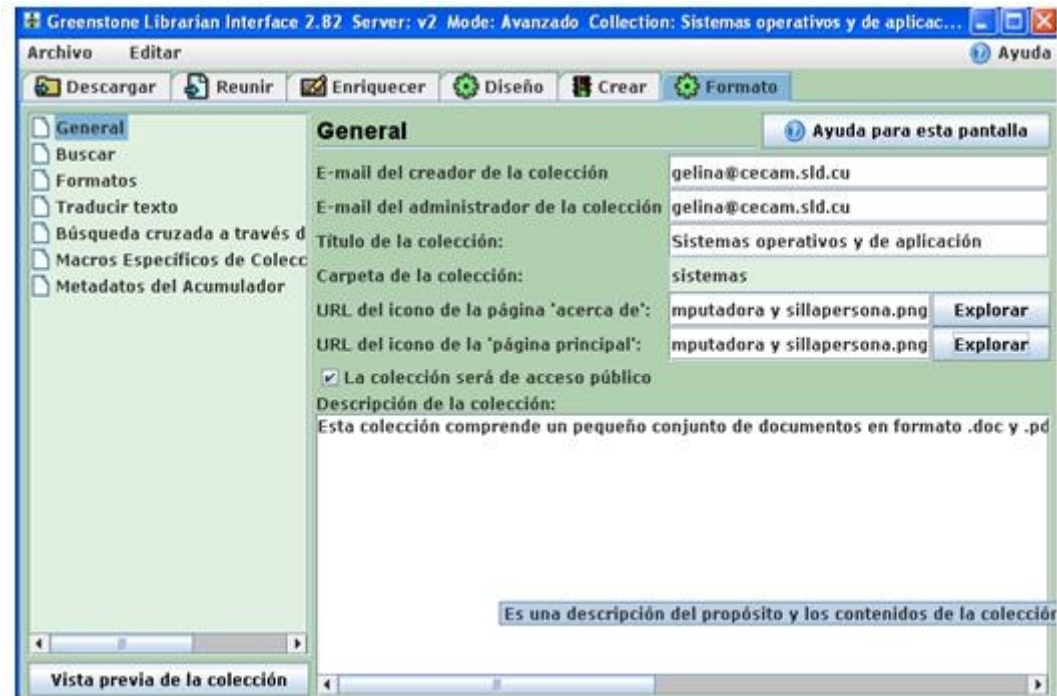
Desde el panel *Diseño*, opción *Clasificadores de exploración*, se eliminaron los clasificadores *List*, *title* y *source* construidos por defecto y se sustituyeron por otros de *título*, *autores* y *palabras clave* configurados manualmente por medio de metadatos del conjunto Dubin Core (Figura 9). Se utilizaron clasificadores del tipo *AZList* y *AZCompactList*.



**Figura 9.** Clasificadores de título, autores y palabras clave diseñados gracias a los metadatos Dublin Core.

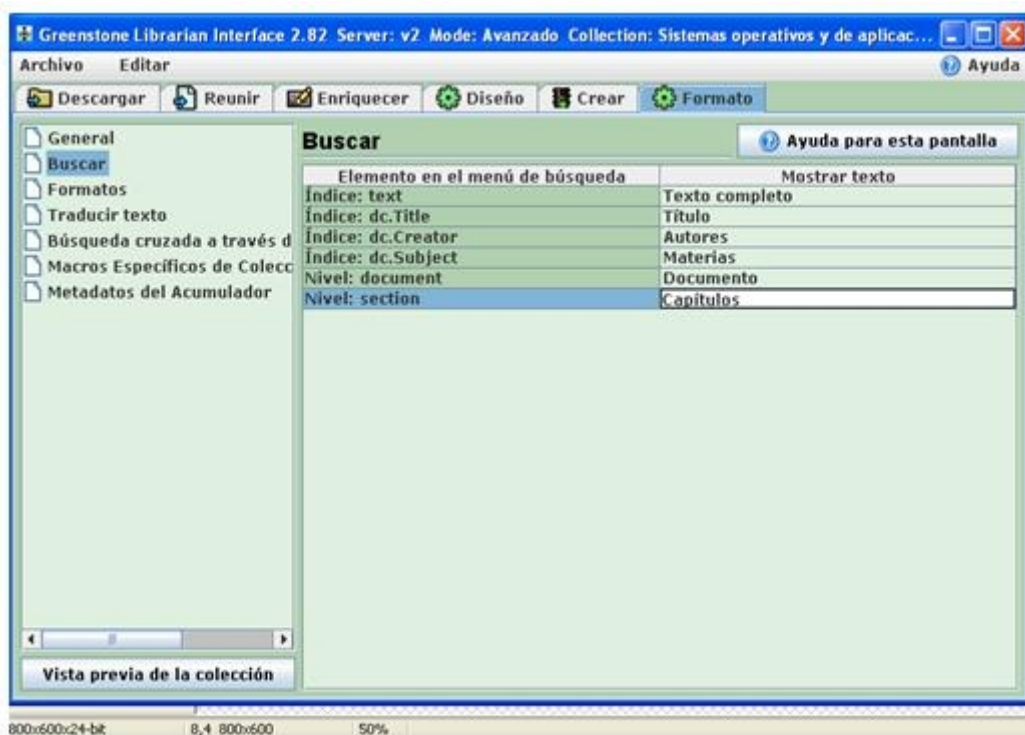
## Otras personalizaciones sobre la base del panel *Formato*

Desde el panel *Formato*, opción *General*, llenar los campos disponibles en el lado derecho del panel según se precise, excepto *Carpeta de la colección* que lo asigna el sistema Greenstone automáticamente y no es modificable (Figura 10). También desde esta opción se puede adjudicar una imagen a la colección tanto en la página principal de Greenstone como en la página *Acerca de* correspondiente.



**Figura 10.** Campos del panel Diseño, opción General llenados según necesidades.

De modo similar se procedió con el panel *Diseño*, opción *Buscar* y se logró visualizar los índices asignados manualmente en español. Las operaciones efectuadas en este sentido se aprecian por medio de la figura 11:



**Figura 11.** Sustitución de los nombres de índices por defecto por nombres en español de acuerdo con las necesidades del usuario final.

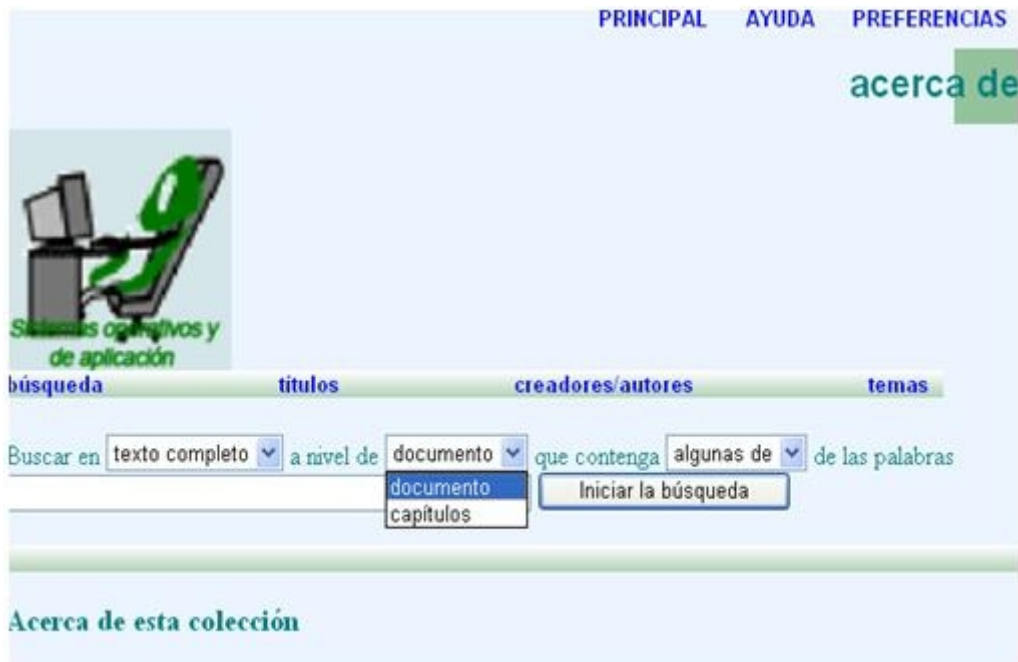
Luego de ejecutar estos pasos, se reconstruyó la colección desde el panel *Crear*, botón *Crear colección*.

Con el propósito de comprobar los resultados logrados, se examinaron todas las salidas de Greenstone. Partiendo de la página *Acerca de* de esta colección, se pueden verificar los ajustes operados en el formulario de búsqueda obtenido cuando se da un clic sobre el botón *Búsqueda*. Las figuras 12 y 13 ofrecen las imágenes correspondientes:



**Figura 12.** Índices por los que es posible buscar.





**Figura 13.** Niveles permitidos para realizar la búsqueda.

Desde la página *Acerca de* correspondiente a la colección, se muestra el clasificador *Títulos* (CL1) obtenido gracias a la utilización de un *VList* que determina una sencilla relación de títulos con sus respectivos íconos luego de la reconstrucción. Note que todos los títulos de los documentos se visualizan correctamente, aunque aparece el doble ícono y el nombre del archivo (Figura 14).



**Figura 14.** Clasificador Títulos (CL1) reconstruido.

Desde la página *Acerca de* correspondiente a la colección, al oprimir el botón *autores/creadores*, el clasificador *Autores* (CL2) ofrece la apariencia que muestra la figura 15. A diferencia del anterior en este figuran íconos de estanterías los cuales se logran al aplicar el clasificador *AZCompactList*.



Figura 15. Clasificador Títulos (CL2) reconstruido.

El clasificador *CL3* igualmente se creó a partir de *AZCompactList* que agrupa a los documentos cuyos temas coinciden.

Desde la página *Acerca de* igualmente se revisó la página de resultados. Éstos se observaron al pulsar el botón *Búsqueda* (o *Buscar*), teclear el término **sistemas de aplicación** y dar clic en el botón *Iniciar búsqueda*. La figura 16 ofrece la apariencia de la lista de resultados obtenida:



Figura 16. Página de resultados obtenida tras teclear el término *sistemas de aplicación*.

## Personalización de la colección mediante las cadenas de formato

Antes de abordar este tema se recuerda, en forma somera, el procedimiento inherente al manejo de las cadenas de formato dado que en el artículo **Creación y personalización de documentos digitales en Greenstone. Parte 1** se ofrecieron explicaciones al respecto. Para alterar el contenido de éstas se procede así:

- Activar el panel *Formato* del *GLI*.
- Dar clic sobre la opción *Formatos* en el lado izquierdo del panel.
- En el panel derecho, pulsar la lista desplegable correspondiente a la opción *Elegir función* para seleccionar la deseada.
- Debajo, seleccionar el componente afectado que corresponda al caso.
- Clic en el botón *Añadir formato* para activar el cuadro de texto correspondiente a *Cadena de formato Html* donde es posible modificar las cadenas de formato.
- También se encuentran disponibles los botones *Insertar variable*, *Deshacer*, *Rehacer*, *Eliminar formato* y *Volver a los parámetros*.

Una rápida inspección a todas las páginas de esta colección posibilita la comprobación de aspectos aún modificables entre los que figuran el doble ícono de los trabajos y el nombre del archivo, elementos que pueden suprimirse sin afectar el funcionamiento de la colección.

Tales circunstancias determinaron un nuevo intento de configuración de la colección, en esta oportunidad, a partir de las cadenas de formato. Éstas se modifican desde el panel *Formato*, opción *Formatos*.

Resulta conveniente aclarar que en esta ejercitación se aplicaron cambios ensayados en el artículo antes mencionado, pero mediante cadenas diferentes: se trata de la eliminación del nombre del archivo del documento original y de la inserción del número de ítemes en los clasificadores. Hecha esta precisión, es posible describir las modificaciones experimentadas.

El primer cambio ejecutado consistió en trabajar con la cadena de formato del clasificador *CL1* con el propósito de eliminar la primera línea cuyo efecto determina el enlace al ícono de la versión *Html* del documento. Se seleccionó como componente afectado *VList* y la función **CL1**. Esta acción se previó con la intención de acceder al trabajo mediante su versión original la cual, por lo general, posee una calidad superior a la lograda al visualizar el documento como una página Web.

La línea eliminada fue:

**<td valign="top">(link)(icon)(/link)</td>**

Con esta acción se logra la desaparición del ícono que enlaza con la versión en *Html* del documento, pero solo es aplicable en los clasificadores que no presentan estanterías, es decir para los *AZList*, en este caso, aplicados al metadato *title*. Para los *AZCompactList* es preciso utilizar otro procedimiento que se explicará más adelante.

La cadena de formato resultante se muestra en el cuadro 3.

```
<td valign="top">[ex.srclink]{Or}{[ex.thumbicon],[ex.srcicon]}[ex./srclink]</td>
<td valign="top">[highlight]
{Or}{[dc.Title],[exp.Title],[ex.Title],Untitled}
[/highlight]{If}{[ex.Source],<br><i>([ex.Source])</i></td>
```

En la figura 17 se aprecia la modificación lograda:



**Figura 17.** Clasificador CL1 en el que se suprimió el ícono de la versión Html del documento original.

Con la finalidad de hacer desaparecer el nombre del archivo del documento, que se encuentra ubicado debajo del título correspondiente, se borró la última línea de la cadena de formato de la función *CL1* y *VL1st*. A continuación se reproduce esa línea:

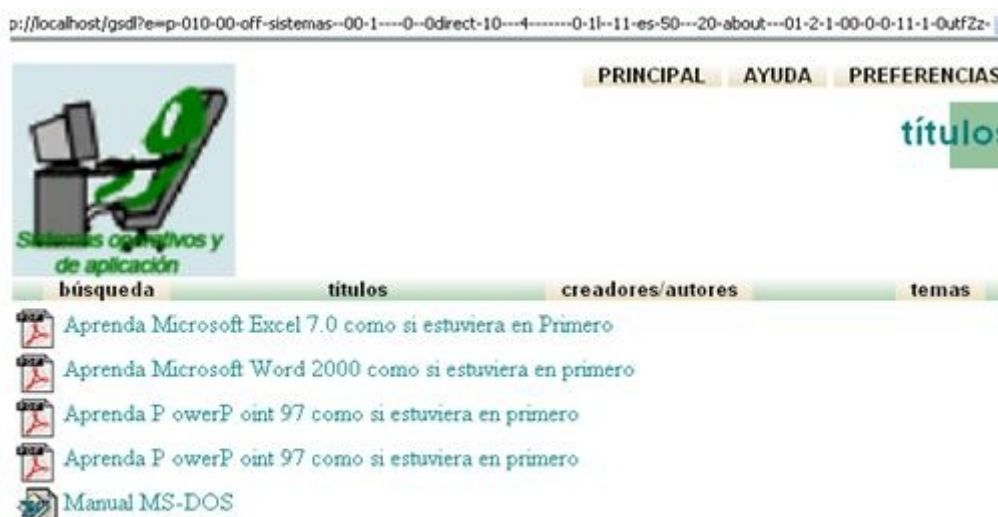
```
((/highlight){If}{(ex.Source),<br><i>((ex.Source))</i></td>)
```

que al suprimirse se obtiene una expresión como la que sigue en el cuadro 4:

```
<td
valign="top">[ex.srclink]{Or}{[ex.thumbicon],[ex.srcicon]}[ex./srclink]
</td>
<td valign="top">[highlight]
{Or}{[dc.Title],[exp.Title],[ex.Title],Untitled}
```

Tal cambio permitió la supresión del nombre del archivo del documento original, hecho que confiere una apariencia más profesional a la página debido a la eliminación de datos carentes de función y al uso más racional del espacio disponible en cada página. Estos procedimientos son válidos también para las funciones *Search* y *DateList*. La figura 18 posibilita la observación del nuevo formato de presentación.





**Figura 18.** Formato de presentación modificado para el CL1 de la colección reconstruida.

Este paso se repitió con el clasificador *CL2*, pero como se trata de un *AZCompactList*, se debió aplicar otra cadena de formato,<sup>13</sup> pues si se procediera eliminando la primera línea también sucedería lo mismo con el ícono de la versión .pdf o .doc y no se podría acceder al texto original sino a la versión .html que en el caso de los .pdf generalmente es defectuosa. Por tanto, se aplicó otra cadena de formato (Cuadro 5) para mantener el enlace a la versión pdf del documento desde la estantería de este clasificador (*CL2*):

```
<td valign=top>{If}{[srcicon],[srclink][srcicon][srclink],  
[link][icon][link]}</td>  
<td valign="top">[highlight]  
{Or}{[dc.Title],[exp.Title],[ex.Title],Untitled}  
</td>
```

con el consiguiente resultado que es posible observar en la figura 19.



**Figura 19.** Formato de presentación modificado para el CL2 de esta colección.

En la cadena de formato antes aplicada también se eliminó la última línea que determina la aparición del nombre del archivo original.

Observe que en esta cadena de formato se introduce la condicional `{If}` delante de variables Greenstone que contemplan tanto (*srcLink*) como (*link*) por cuanto el sistema ofrece la posibilidad de elegir entre ambos enlaces y lo practica con (*srcLink*) que es quien enlaza con la versión .pdf del documento.

Existe otra cadena de formato aplicable para CL2<sup>13</sup> con el mismo objetivo y que arroja resultados similares, los cuales pueden verse en el cuadro 6 y en la figura 20:

```
<td valign=top>{If}{[[FileFormat]] eq
"PDF",[srcLink][srcIcon][[/srcLink],[link][icon][[/link]]]</td>
<td valign="top">[highlight]
{Or}{[[dc.Title],[exp.Title],[ex.Title],Untitled]}
```



Figura 20. Segunda cadena de formato válida para el clasificador CL2.

Se recomienda, a los lectores, ejecutar las 3 alternativas (en la parte 1 se ofreció otra diferente) con el propósito de comprobar por sí mismos el alcance y consecuencias de la aplicación de cada una de esas variantes.

Cuando a los efectos del usuario final resulta provechoso adicionar los datos concernientes a los autores, las materias tratadas por los documentos y la fecha del documento dentro del clasificador *títulos* con vistas a su posterior visualización, se agregan varias líneas al final de la cadena de formato de un clasificador *AZList* o *AZCompactList*. Tales líneas se visualizan en el cuadro 7:

```
{If}{[[dc.Creator]],<br><b> Autor: [dc.Creator]</b></i>}
{If}{[[dc.Subject]],<br><b> Temas: [sibling:dc.Subject]</b></i>}
{If}{[[dc.Date]],<br><b> Fecha: [format:dc.Date]</b></i>}
```

En este caso, se insertaron, también, varios elementos en cada línea. Esos elementos se resaltan gráficamente en el párrafo anterior. No obstante, resulta pertinente comentar algunos aspectos. Observe que se adicionaron las palabras autor, temas y fechas seguidas de dos puntos (:).

En la primera línea se aplicó el metadato *Creator* del conjunto Dublin Core y el efecto alcanzado fue la aparición del primer autor debajo del título de cada documento mientras que para visualizar todas las palabras clave empleadas se insertó la variable *sibling*: al metadato *Subject*.

En el caso del metadato *Date* del conjunto Dublin Core se le antepuso la variable *format*: cuyo efecto fue el registro de la fecha de la manera prevista en Greenstone.

Las figuras 21 y 22 muestran el efecto logrado en un caso en el clasificador correspondiente a los autores (CL2) y el siguiente a los títulos (CL1).



Figura 21. Clasificador Autores desplegado.



Figura 22. Clasificador Títulos desplegado.

Por otra parte, en presencia de estanterías generadas por clasificadores *AZCompactList* resulta beneficioso conocer cuántos documentos están agrupados,



por ejemplo, por un mismo autor o materia según el diseño de la colección. En este caso la adición de ese dato es particularmente interesante y se obtuvo al añadir una nueva línea a la cadena de formato donde se utiliza la variable (*numleafdocs*):

**{If}{(numleafdocs),<td><i>((numleafdocs))</i></td>}**

Esta expresión es aplicable a todos los clasificadores del tipo *AZCompactList* que en esta colección son los *CL2* y *CL3*. Al emplearlo en el clasificador de temas *CL2* se obtuvo lo siguiente (Cuadro 8):

```
<td valign="top">[link][icon][link]</td>
<td
valign="top">[ex.srclink]{Or}{[ex.thumbicon],[ex.srcicon]}[ex./srclink]
</td>
<td valign="top">[highlight]
{Or}{[dc.Title],[exp.Title],[ex.Title],Untitled}
[/highlight]{If}{[ex.Source],<br><i>([ex.Source])</i></td>
{If}{[numleafdocs],<td><i>([numleafdocs])</i></td>}
```

que determinó este efecto que se muestra en la figura 23:



The screenshot shows a web application with a navigation bar at the top containing 'PRINCIPAL', 'AYUDA', and 'PREFERENCIAS'. Below the navigation bar is the title 'creadores/autores' in green. On the left side, there is a logo of a computer and a chair with the text 'Sistemas cognitivos y de aplicación'. The main content area displays a table with four columns: 'búsqueda', 'títulos', 'creadores/autores', and 'temas'. The table lists several authors and their corresponding number of works in parentheses.

búsqueda	títulos	creadores/autores	temas
Brazález Guerra A	(2)		
Larzabal A	(1)		
Mora A	(1)		
Girón Legorburu I	(1)		
Calleja J	(1)		
Fosch J	(1)		
García de Jalón de la Fuente J	(3)		
García I	(1)		

**Figura 23.** Clasificador de Creadores/autores con especificación del número de trabajos por autor.

En el artículo precedente a éste se aplicó una variante que consistió en la inserción del número entre paréntesis detrás de la estantería en lugar de figurar al final como sucede en la figura 23.

Es necesario dar clic en las estanterías para revisar los títulos subordinados a cada autor donde por supuesto aparecerán, además del título, el autor, las palabras clave y la fecha de publicación del documento según el efecto de la cadena de formato utilizada en ese caso (Figura 24).



**Figura 24.** Clasificador CL2 desplegado donde se visualiza el primer autor y las palabras clave.

La próxima personalización consistió en aplicar etiquetas Html para colorear el fondo de los resultados.<sup>6</sup> Ello se materializó mediante la inserción de una línea al clasificador *CL1* (debajo de la primera) como la que sigue:

**<td valign=top bgcolor="#FFFFCC">**

La cadena de formato resultante se muestra en el cuadro 9:

```
<td valign=top>[ex.srclink][ex.srcicon][ex./srclink]</td>
<td valign=top bgcolor="#FFFFCC"><b>[link][dc.Title][/link]</b><br>
<i>[dc.Creator]</i></td>
```

con la que se obtuvo la apariencia que muestra la figura 25:



**Figura 25.** Aplicación de cadena de formato que permite colorear la columna donde aparecen los títulos y autores.

Otra transformación que es posible practicar consiste en la presentación de los datos en dos columnas al añadir una etiqueta Html en la segunda línea; este código es

**<td valign="top">(highlight)**

el cual determina la disposición de los datos en la forma mencionada (Cuadro 10).

```
<td
valign="top">[ex.srclink]{Or}{[ex.thumbicon],[ex.srcicon]}[ex./srclink]
</td>
<td valign="top">[highlight]
{Or}{[dc.Title],[exp.Title],[ex.Title],Untitled}<td
valign="top">[sibling:dc.Subject]
[/highlight]
{If}{[ex.Source],<br>{Or}{[sibling:dc.Creator],[sibling:ex.Creator]}
{If}{[numleafdocs],<td><i>([numleafdocs])</i></td>}
```

El formato de presentación resultante puede apreciarse en la figura 26:



búsqueda	títulos	creadores/autores	temas
Aprenda Microsoft Excel 7.0 como si estuviera en Primero	Sistemas de aplicación Windows Hojas de cálculo Excel Manuales	García de Jalón de la Fuente J Gofí Lasheras R Aguirre Suárez O Sarriegui Domínguez JM Girón Legorburu I Brazález Guerra A	
Aprenda Microsoft Word 2000 como si estuviera en primero	Sistemas de aplicación Word Windows Manuales	García de Jalón de la Fuente J Rodríguez JI Mora A	
Aprenda Power Point 97 como si estuviera en primero	Sistemas de aplicación Windows Power Point Manuales	García de Jalón de la Fuente J Rodríguez JI Funes P Brazález Guerra A Larzábal A Calleja J García J	
Manual MS-DOS	Sistemas operativos MS-DOS Manuales		

Figura 26. Formato de presentación de los datos en 2 columnas.

Compare la figura 26 con las 21 y 22 a fin de corroborar las transformaciones operadas en las páginas correspondientes en cuanto a presentación de los datos e íconos.

En caso de que se desee hacer aún más explícita la información ofrecida en la columna de la derecha, es posible introducir encabezamientos que permitan identificar más claramente a los autores y palabras clave. Para ello se anteponen los términos *Autores:* y *Temas:* delante de los metadatos correspondientes de la forma que se muestra a continuación:

En la figura 27 se aprecia el efecto surtido:



búsqueda	títulos	creadores/autores	temas
Aprenda Microsoft Excel 7.0 como si estuviera en Primero	Temas: Sistemas de aplicación Windows Hojas de cálculo Excel Manuales	Autores: García de Jalón de la Fuente J Gofí Lasheras R Aguirre Suárez O Sarriegui Domínguez JM Girón Legorburu I Brazález Guerra A	
Aprenda Microsoft Word 2000 como si estuviera en primero	Temas: Sistemas de aplicación Word Windows Manuales	Autores: García de Jalón de la Fuente J Rodríguez JI Mora A	
Aprenda Power Point 97 como si estuviera en primero	Temas: Sistemas de aplicación Windows Power Point Manuales	Autores:	

Figura 27. Encabezamientos de los títulos y autores.

Si se aplicaran etiquetas Html convenientes (<b> y /b; <u></u>), sería posible resaltar en negritas o subrayar los encabezamientos Temas y Autores.



Por otra parte, desde las funciones disponibles en el panel *Formato*, opción *Formatos*, también puede configurarse el formulario de búsqueda para realizar búsquedas avanzadas por campos.<sup>14</sup> La función que permite este cambio es *Search* cuya cadena es *plain*, *form*. Basta suprimir *plain* y la coma(,) para operar una transformación radical. La cadena obtenida es:

### form

cuya salida se visualiza en esta forma (Figura 28):

**Figura 28.** Formulario determinado por la cadena de formato form.

Este tipo de formulario ofrece la posibilidad de ejecutar búsquedas más complejas al especificar valores para los campos disponibles (índices) y requiere la utilización del motor de búsqueda MGPP. Tal formulario puede transformarse desde el botón *Preferencias* cuando se selecciona la opción búsqueda avanzada, acción con la cual es posible utilizar operadores booleanos.

El cambio operado es el que se visualiza en la figura 29:

**Figura 29.** Formulario anterior transformado desde el botón *Preferencias*

## CONCLUSIONES

Las cadenas de formato aplicadas al reproducir este ejercicio de personalización de las colecciones permiten la profundización del conocimiento del sistema Greenstone y ofrecen la posibilidad de adentrarse en el tema de la configuración de tales colecciones a la medida de las necesidades de cada usuario final, en este caso específico, documentos .pdf y .doc. Con tales acciones se logra acceder a la información almacenada en forma más eficiente.

## REFERENCIAS

1. Witten IH., Don KJ, Dewsnip M, Tablan V. Text mining in a digital library. J Dig Libr. August; 4(1):56-9. Disponible en: <http://www.cs.waikato.ac.nz/~ihw/publications.html>. (Consultado: 2 mar 2011).
2. Witten IH. Customizing digital library interfaces with Greenstone. TCDL Bull. Summer 2003. Disponible en: <http://www.ieee.tcdl.org/bulletin/v1n1/summer2003.html>. (Consultado: 23 feb 2010).
3. Witten IH. Creating and customizing digital library collections with the Greenstone Librarian Interface. Disponible en: <http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp97.pdf>. (Consultado: 2 mar 2011).
4. FAO. Perspectiva general de los estándares de metadatos (curso en CD-ROM). En: La digitalización y las bibliotecas digitales del Repertorio de recursos para la gestión de información. (Roma?: IMARK?) 2006.
5. Méndez Rodríguez E. Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales (crítica de libro). Gijón: Ediciones Trea, 2002. Disponible en: <http://digital.csic.es/bitstream/10261/12202/1/308.pdf>
6. Tramullas J, Garrido P. Bibliotecas digitales: Greenstone. En: Software libre para servicios de información digital. Madrid: Pearson Prentice Hall, 2006. Disponible en: <http://eprints.rclis.org/6145/1/Capítulo10.pdf>. (Consultado: 2 mar 2011).
7. Witten IH, Boddie S, Thompson J. Visión general de Greenstone. Guía del usuario. Nueva Zelanda: Universidad de Waikato, 2006. Disponible en: [http://www.greenstone.org/manuals/gsdl2/es/Html/User\\_es\\_index.Html](http://www.greenstone.org/manuals/gsdl2/es/Html/User_es_index.Html). (Consultado: 2 mar 2011).
8. Bainbridge D, McKay D, Witten IH. Explotación óptima de sus documentos. En: Guía del programador. New Zealand: University of Waikato. Digital Library, 2004. Disponible en:

[http://www.greenstone.org/manuals/gsd12/es/Html/Develop\\_es\\_index.Html](http://www.greenstone.org/manuals/gsd12/es/Html/Develop_es_index.Html). (Consultado: 11 oct 2009).

9. University of Waikato. Digital Library Project. Ayuda: 6.3 Índices de búsqueda. En: Greenstone Digital Library Software (programa informático). ver 2.82 New Zealand: University of Waikato, 2008.

10. Aprender acerca de formatos y de macros. En: Ejercicio tutorial de Greenstone: modificado para versión 2.70. Disponible en: [http://wiki.greenstone.org/gsd/doc/tutorial/es/format\\_and\\_macros.htm](http://wiki.greenstone.org/gsd/doc/tutorial/es/format_and_macros.htm). (Consultado: 2 mar 2011).

11. Una colección de archivos de Word y PDF. En: Ejercicio tutorial de Greenstone modificado para versión 2.70. Disponible en: [http://wiki.greenstone.org/gsd/doc/tutorial/es/Word\\_PDF\\_collection.htm](http://wiki.greenstone.org/gsd/doc/tutorial/es/Word_PDF_collection.htm). (Consultado: 2 mar 2011).

12. Una colección bibliográfica. En: Ejercicio tutorial de Greenstone: modificado para versión 2.70. New Zealand: University of Waikato. New Zealand Digital Library Project, 2007. [http://wiki.greenstone.org/gsd/doc/tutorial/es/bibliography\\_collection.htm](http://wiki.greenstone.org/gsd/doc/tutorial/es/bibliography_collection.htm). (Consultado: 2 mar 2011).

13. Customizing collections from GreenstoneWiki: how do I suppress the link to the Greenstone text version of a document? En: Greenstone FAQ. Disponible en: [http://wiki.greenstone.org/wiki/index.php/Customizing\\_collections#How\\_do\\_I\\_suppress\\_the\\_link\\_to\\_the\\_Greenstone\\_text\\_version\\_of\\_a\\_document.3F](http://wiki.greenstone.org/wiki/index.php/Customizing_collections#How_do_I_suppress_the_link_to_the_Greenstone_text_version_of_a_document.3F). (Consultado: 2 mar 2011).

14. Bibliographic collections. En: Greenstone Tutorial Exercises: modified for Greenstone version: 2.82. Disponible en: [http://wiki.greenstone.org/wiki/gsd/doc/tutorial/en/bibliography\\_collection.htm](http://wiki.greenstone.org/wiki/gsd/doc/tutorial/en/bibliography_collection.htm). (Consultado: 2 mar 2011).

Recibido: 7 de julio de 2011.

Aprobado: 10 de octubre de 2011.