

Nombres descriptivos de archivos de colecciones digitales: propuesta con elementos del sistema de clasificación Dewey.

Lara Pacheco, Clemente Gonzalo

Lic. en Bibliotecología. clemente@dgb.unam.mx

Biblioteca digital UNAM. Tel: 56223984. Edificio Anexo DGB, Ciudad Universitaria, D.F., c.p.04510. México

Resumen

Se presenta una propuesta de nombramiento de archivos electrónicos con nombres descriptivos, basados en el esquema de clasificación Dewey, para facilitar la administración de colecciones digitales, especialmente en bibliotecas digitales.

Palabras clave: nombramiento de archivos; nombres descriptivos; administración de archivos electrónicos; bibliotecas digitales, clasificación Dewey.

Abstract: A proposal of file naming using significant names in digital libraries based on DDC scheme is presented. It is aimed to collections produced by libraries, like digital born and digitized items.

Key words: file naming; significant file names; file management; digital library, DDC classification

Nota: Agradezco las observaciones y comentarios hechos al presente texto por el Ing. Alberto Castro Thompson.

Introducción

El crecimiento de los recursos digitales de información ha comenzado a dejar ver sus beneficios, por ejemplo, en las áreas de educación e investigación académica. Este incremento ha arrojado diferentes retos, entre ellos, el control de las colecciones digitales emergentes. Dentro de esta actividad intervienen diferentes tareas, una de las cuales es la asignación de nombres o identificadores exclusivos para cada documento o ítem dentro de una colección digital (por ejemplo, el archivo que resulta de la digitalización de una foto o de la página de un libro o el libro completo en versión digital).

Para abordar satisfactoriamente esta actividad es necesario diseñar estrategias viables y duraderas para mantener controlada la generación de objetos digitales en colecciones tanto pequeñas como de gran tamaño. Tradicionalmente un libro cuenta con identificadores propios (que en ambientes digitales podemos llamar metadatos) como título, autor, editorial, ISBN, etc. Para diferenciarlo de otros títulos u otras ediciones, incluso, en la biblioteca se le asigna un código alfanumérico tanto para localizarlo físicamente en los estantes cuanto para diferenciarlo y darle un lugar en el espectro de las áreas del conocimiento (Medicina, Matemáticas, Geografía, etc.).

El caso de los archivos electrónicos que conforman obras como libros, revistas, etc., merecen consideraciones particulares¹. Conforme crece su número (en forma individual y como un documento completo: libro, revista, etc.) dentro de una colección, se hace más difícil identificarlos unívocamente, y la práctica demuestra que no es suficiente con asignarles un número progresivo o aleatorio. En este punto es necesario diferenciar entre nombre de archivo **individual** y nombre de archivo **del documento**. El archivo individual es el que resulta de la página escaneada de una revista, mientras que el archivo del documento es el de la revista completa.

La presente propuesta de nombramiento descriptivo² de archivos es aplicable a documentos digitales de origen y digitalizados (escaneados) que no cuenten con un elemento diferenciador único que pueda servir como nombre de archivo del documento (libro, revista, tesis, etc., completos) o de los archivos individuales (páginas de documentos, imágenes sueltas, etc.), según sea el caso. El elemento identificador puede ser, por ejemplo, el número de adquisición, control, inventario, etc., de un ítem de la colección. Para que pueda ser utilizado como nombre de archivo, necesariamente debe ser único e irrepetible pues es una condición indispensable para establecer hipervínculos entre hipertextos y objetos digitales.

En la práctica hemos observado que el documento digital que no recibe un nombre único y descriptivo dentro de una colección de obras digitales demanda la generación de un nombre con significado para su mejor administración y consulta.

Es importante tener en cuenta que el nombre identificador de cada archivo es el elemento que permite vincular al usuario con el ítem propiamente dicho; de alguna manera, lo que hace en las bibliotecas la signatura topográfica del lomo de los libros impresos. Conviene recordar que existen sistemas de administración de objetos digitales que permiten ingresar elementos que se refieran al tamaño, resolución, fecha de procesamiento, características,

¹ Identifiers for digital resources. [En Línea]. <<http://memory.loc.gov/ammem/award/docs/identifiers.html>> [Consulta: enero 2008].

² Se entiende por "nombre descriptivo" una nomenclatura que semánticamente remite a un concepto consensado.

etc., un poco a la manera de los metadatos administrativos³. Estos elementos contribuyen a robustecer la información en torno a un objeto digital y facilita su manejo y explotación en un sistema de información.

Nombres descriptivos y no descriptivos⁴

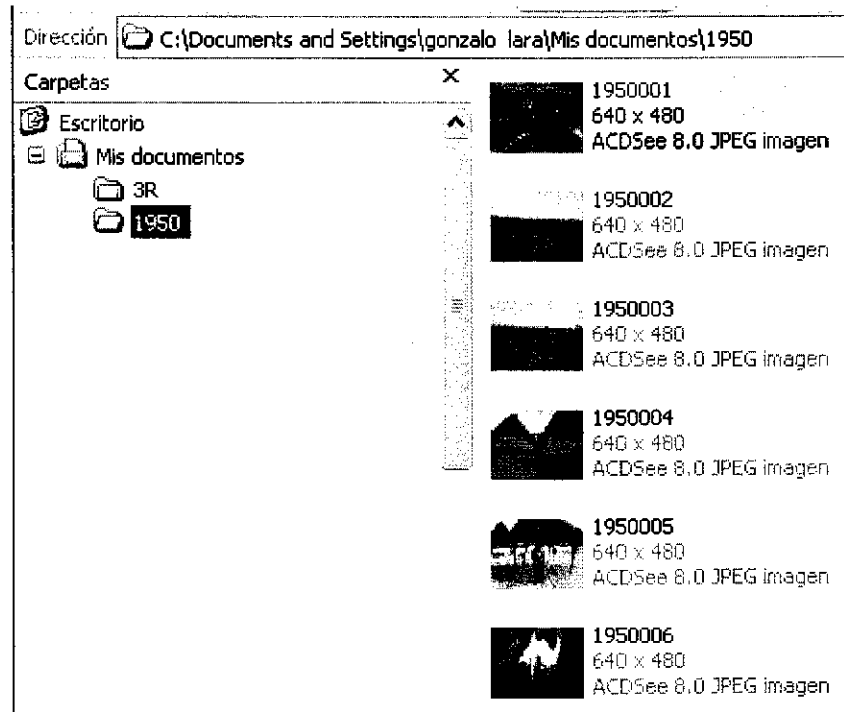
El nombre de un archivo puede contener caracteres alfabéticos, numéricos o una combinación de ambos. La cadena que lo componga podrá tener significado lógico (fecha, siglas, palabra, frase, claves locales, etc.), guardar relación con el contenido del archivo (iniciales, primeras letras, tema, etc.) o ser simplemente una secuencia de caracteres. Es común usar palabras y frases completas en archivos individuales personales o con contenido textual, gráfico, etc., pero el nombramiento de archivos en una biblioteca digital o una entidad que administra información digital requiere soluciones de mayor alcance y duración.

Los nombres **descriptivos** tienen significado para el administrador de la colección digital, es decir, se refieren a un concepto, referente, etc. Al ojear las carpetas en las que están almacenados es posible identificar su contenido. Por ejemplo, si una colección de documentos está separada por años, se pueden utilizar cuatro números como nombre de la carpeta y también pueden ser parte del o los archivos individuales (véase Fig. 1).

³ **Metadatos administrativos:** información necesaria que permite a un sistema de administración de información manipular los objetos digitales. Puede incluir información sobre cómo fue procesado digitalmente, su formato de almacenamiento, etc. (comúnmente se les conoce como metadatos técnicos), características, etc. Asimismo, contiene información sobre derechos de autor y licenciamiento y datos concernientes a la preservación a largo plazo de dichos objetos (metadatos de preservación). Véase *Metadata in the Oxford Digital Library (ODL)*. En: <http://www.odl.ox.ac.uk/metadata.htm> Consultado en noviembre 2007

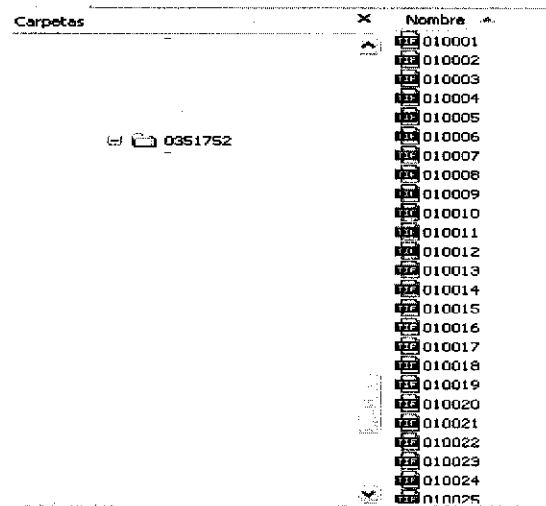
⁴ File naming. [En Línea]. *Technical Advisory Service for Images*.
<<http://www.tasi.ac.uk/advice/creating/filenaming.html>> [Consulta: octubre 2007].

(Fig. 1. Nombres descriptivos)



Los nombres **no descriptivos** carecen de significado. Un ejemplo típico son los archivos individuales producto de una digitalización masiva. Esta asignación de nombre se debe al equipo con el que se trabaja, pues los escáneres para proceso masivo asignan automáticamente un nombre a cada archivo que generan, aunque opcionalmente se pueden preconfigurar otros caracteres.

(Fig. 2 Nombres no descriptivos)



Nombres para archivos fuente (*master*) y archivos de consulta

Un documento digital puede contar con dos archivos, uno fuente, con todas sus características completas (tamaño, resolución, etc.) y uno de consulta (el que ve el usuario final), que regularmente ha sufrido cambios, ajustes, alteraciones o alguna otra modificación a fin de tener una presentación que permita que se consulte ágilmente. El nombramiento de dos archivos que se refieren al mismo documento debe diferir, pero también mantener elementos en común. El archivo fuente podría contar con un carácter que indicaría justamente que se trata de la versión fuente (**f**). Por ejemplo: **1950001f.pdf**. Mientras que el archivo de consulta puede prescindir de un distintivo o incluirlo, dependiendo de las necesidades del proyecto de digitalización.

Sugerencias para el nombramiento de archivos digitales

Manejar grandes cantidades de objetos siempre requiere de un método para hacerlo de manera sencilla, práctica y funcional. Dar orden a centenas, miles o millones de objetos es necesario y no siempre sencillo. Por ejemplo, para los sistemas operativos de hace algunos años sólo se podían utilizar ocho caracteres⁵ en el nombre de un archivo. Actualmente esta práctica ha quedado rebasada porque las cadenas de nombres pueden contener mayor

⁵ BROWN, Bruce. Find that file: Special Report - Getting Organized (hard disk files). [En Línea]. Lotus, June 1989, v5. no. 6, p 88.

número de caracteres. Brown⁶ señalaba que en el nombramiento de archivos personales los primeros cuatro caracteres se refirieran a las iniciales que ayudaran a identificar a una persona o tema al que aludiera el archivo; y sugería que los cuatro restantes se utilizaran para referirse a una fecha u otro concepto significativo, todo ello a fin de evitar confusiones y perder información. De estas sugerencias hechas hace ya varios años, podemos aprovechar actualmente lo que se refiere a formar el nombre del archivo con elementos que le den sentido, a fin de facilitar su administración.

Algunas organizaciones como la biblioteca del Congreso de los Estados Unidos⁷, puntera en la administración de colecciones de bibliotecas, y la Digital Library Federation, aconsejan que la asignación de los nombres a los archivos digitales obedezca a criterios de la propia institución o entidad que desarrolle un proyecto de digitalización. En otras palabras, apelan a que se utilicen criterios locales. Consideran que mientras el nombre sea funcional y práctico para quien lo asigne y lo use, no habrá problema. Sin embargo, con la idea de normalizar la administración y facilitar el intercambio de recursos internacionalmente y entre sistemas de información, estas mismas organizaciones sugieren la opción de seguir algunos lineamientos básicos estandarizados.

La Universidad de Michigan y la Joint RLG/TASK Force on Digital Preservation⁸ recomiendan considerar lo siguiente para el nombramiento de un archivo digital:

- Procurar ajustarse a la norma de nombramiento ISO 9660⁹ (una norma que define un sistema de archivo útil en una variedad de sistemas operativos).
- Establecer un nombramiento de archivos convencional y las extensiones que posteriormente se generen de él. (Esto quiere decir que un archivo puede tener

⁶ *idem*

⁷ Identifiers...op cit.

⁸ SNAVELY, Cory. Handle Service Conventions and Practices. [En Línea]. *Digital Library production Service*. 2006. <<http://www.umdl.umich.edu/docs/arch/general/handle-conventions.html>> [Consulta: febrero 2008]

⁹ Joliet Specification: CD-ROM Recording Spec ISO 9660:1988. [En Línea]. 1995. <<http://bmr.c.berkeley.edu/people/chaffee/jolspec.html>> [Consulta: octubre 2007]

A pesar de que se refiere a una ISO hecha para discos compactos, ejemplifica claramente la práctica adoptada generalmente para el nombramiento de archivos asentados en formatos de diferentes tipos. Al respecto, conviene revisar también la especificación *Joliet* (<http://bmr.c.berkeley.edu/people/chaffee/jolspec.html>), extensión de la ISO 9660, para conocer las posibles adaptaciones a esta norma.

derivaciones, copias, respaldos, etc., y estas variaciones deben mantener rasgos del nombre fuente (por ejemplo: 0754123**b** deriva de 0754123)

- Basar los nombres en números de acceso, de control, de inventario, de sistema o en algún otro identificador único al interior de una colección de objetos.
- Evitar el uso descontrolado de mayúsculas y minúsculas y de signos diacríticos.

Según Cornell University¹⁰ conviene observar las siguientes recomendaciones:

- Utilizar una estructura de nombramiento de los archivos que sea compatible con cualquier sistema operativo y con los medios de almacenamiento que se planea utilizar.
- Utilizar extensiones de archivo estándares para diferentes tipos de archivos.
- No sobrecargar directorios con demasiados archivos.
- En el caso de colecciones grandes, dejar que sea el propio software de digitalización el que asigne nombres individuales seriados.
- Nombrar previendo el crecimiento de la colección.

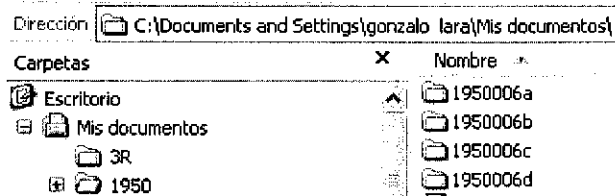
Por otra parte, el proyecto Building Digital Collections¹¹ de la Biblioteca del Congreso de los Estados Unidos recomienda:

- a) que el nombre de un archivo no rebase los 8 caracteres (numéricos, alfabéticos o ambos), por facilidad de manejo y por compatibilidad con el nombramiento de archivos en diferentes plataformas (ISO 9660).
- b) seguir la fórmula **8.3**: ocho caracteres más la extensión del archivo indicada por tres caracteres: jpg, tif, gif, pdf, bmp, etc.
- c) si se prevén extensiones o variaciones del mismo archivo generar un directorio y subsecuentes, si se considera necesario, partiendo de la misma estructura más un carácter secuencial al final (véase Fig. 3).

¹⁰ Cornell University. Moving theory into practice: Digital Imaging Tutorial. [En Línea]. <<http://www.library.cornell.edu/preservation/tutorial/technical/technicalC-01.html>> [Consulta: diciembre 2007]

¹¹ Identifiers...*op.cit.*

(Fig. 3. Nombres descriptivos con derivaciones)



Propuesta de la Biblioteca digital-UNAM¹² para el nombramiento de archivos de documentos digitales (libros, revistas, etc.).

La presente propuesta se basa en la utilización del sistema de clasificación decimal Dewey¹³ aplicado a partir del contenido intelectual del documento a nombrar. Consiste en estructurar un nombre *descriptivo* construido a partir del esquema más un número secuencial y la extensión del archivo.

Supongamos que se tiene el siguiente documento:

Creating a virtual library : a how-to-do-it manual for librarians¹⁴

1. El primer paso es identificar el área general a la que pertenece el documento dentro del esquema Dewey, que para este título sería *000*, pues es en esa área en la que se localiza lo referente a las bibliotecas.

- 000 Computadoras, información y generalidades
- 100 Filosofía y psicología
- 200 Religión
- 300 Ciencias sociales
- 400 Idiomas
- 500 Ciencias puras
- 600 Tecnología
- 700 Bellas artes y recreación
- 800 Literatura
- 900 Geografía e Historia

¹² <http://bidi.unam.mx>

¹³ Sistema de clasificación internacional empleado en el ámbito bibliotecario y que organiza el conocimiento de manera sistemática <http://www.oclc.org/dewey/>. Esquema general en <http://www.tnrilib.bc.ca/dewey.html#000>

¹⁴ *Creating a virtual library : a how-to-do-it manual for librarians*. Edited by Frederick Stielow. New York : Neal- Schuman, 2003. Clasificación Dewey: 025/.00285/5276

2. A estos elementos agregamos una secuencia numérica y una extensión:

000 00125.pdf

Donde: 000 = Generalidades en Dewey
0025 = secuencia numérica

3. Opcionalmente, para especificar el contenido de la obra podemos emplear la segunda o tercera tabla del mismo esquema, dependiendo del nivel de profundidad que dicten las necesidades del proyecto.

En este caso pasamos al siguiente nivel de especificidad:

Clasificación Dewey	Descripción
025 (tomado de la segunda tabla)	Funciones de las bibliotecas

El nombre del archivo, siguiendo la estructura anteriormente expuesta, sería: **025 00001.pdf**.

Donde 026 = tópico de la segunda tabla Dewey
00001 = secuencia numérica

4. Siguiendo las recomendaciones de organismos antes mencionados, podemos diseñar una tabla local basada en el mismo esquema Dewey u otros similares. Con esto conseguimos un mayor nivel de especificidad.

- 00 Administración
- 02 Historia
- 03 Ensayos, discursos, conferencias, ponencias
- 05 Aplicación de tecnología (software)
- 06 Consorcios, asociaciones, grupos de trabajo, etc
- 07 Empresas distribuidoras, arrendadoras, proveedoras, etc., de contenido
- 08 Proyecciones, tendencias
- 09 Normas, estándares, políticas, prácticas y lineamientos

5. Si partimos de que el título, *Creating a virtual library...*, cabe en el rubro 09 de esta tabla. Y ahora el nombre del archivo nos queda de la siguiente manera:

025 09 001.pdf

Donde:

025 = Bibliotecas por tema específico

09 = Normas, estándares, políticas, prácticas y lineamientos

001 = secuencia numérica asignada al archivo

Otros ejemplos

a) Para un documento de Antropología

Nombre: **AA 301 001.htm**

Donde :

AA = clave alfabética (de significado local, de colección, etc.)

301 = sociología y antropología (2ª tabla de Dewey)

001 = secuencia numérica

b) Para imágenes de Astronomía

Nombre: **520 00001.xml**

Donde:

520 = Astronomía (2ª tabla de Dewey)

00001 = secuencia numérica asignada al archivo

Comentarios finales

La selección de 8 dígitos, más una extensión de 3 caracteres, según lo recomienda la ISO 9660, debe tomarse como sugerencia, sobre todo considerando las capacidades de los actuales sistemas operativos. Por ejemplo, la Direction du Patrimoine del ministerio de cultura de Francia, por medio de su Inventario General¹⁵, utiliza en el nombramiento -descriptivo- de los archivos de sus acervos, de 10 a 14 caracteres. Por otro lado, un estudio sobre los hábitos en el nombramiento de archivos por parte de usuarios de computadoras personales¹⁶ ha revelado que un promedio típico de nombres de archivos va de los 9 a los 10 caracteres, incluyendo subcadenas formadas por separadores como el guión bajo (v.gr.: tarea_Feb07.doc). Esto nos recuerda que es lícito utilizar un nombramiento que responda a las necesidades particulares de cada colección de documentos, pero no deben perderse de vista las ventajas de mantener un patrón y observar prácticas de uso internacional.

Antes de elegir unas cuantas cifras o letras que compondrán el nombre del archivo, conviene considerar un crecimiento no previsto de la colección. Por ejemplo, se pueden utilizar 4 dígitos para la secuencia numérica del nombre, previendo un crecimiento de hasta 9,999 objetos. Si se rebasa esta cifra, siempre se pueden agregar los dígitos necesarios, flexibilizando la recomendación de la fórmula 8.3.

Como otros elementos, el nombramiento de los objetos, su definición e identificación, son fundamentales para el buen funcionamiento de la administración de los objetos digitales de una colección. Arms¹⁷ recuerda que, como lo hemos venido señalando, es importante que cada nombre de cada objeto sea único; y no sólo eso, sino que además sean inmunes a cambios de tecnología, de repositorios (mudanza de servidores) o de cualquier agente que pueda dificultar su recuperación. El factor humano es otro del que hay que cuidar los

¹⁵ Projet de gestion des images numériques à l'Inventaire général : recommandations techniques, mayo 1999. [En Línea]. (Apartado 4.) <<http://www.culture.gouv.fr/culture/nrt/numerisation/fr/technique/inventaire.htm>> [Consulta: enero 2008]

¹⁶ HEATHER Chapman. The File Naming Habits of Personal Computer Users. [En Línea]. A Master's paper for the M. S. in I. S. degree. March, 1999. 85 h. <<http://neoref.ils.unc.edu:8080/2475.pdf>> [Consulta: enero 2008]

¹⁷ ARMS, William Y. Key Concepts in the architecture of the digital library. [En Línea]. *D-Lib Magazine*, July 1995. <<http://www.dlib.org/dlib/July95/07arms.html#junewya4>> [Consulta: noviembre 2007]

nombres: deben establecerse políticas que garanticen que un administrador, jefe de proyecto, etc., no altere los nombres de los objetos sin una muy bien justificada razón.

Para garantizar la singularidad de los nombres, puede hacerse uso de un software dedicado a esta tarea. La Corporation for National Research Initiatives¹⁸ ha establecido un servidor que atiende este tipo de necesidades. Lo importante en el nombramiento de archivos es mantener nombres únicos para cada archivo u objeto digital que faciliten su administración, consulta y recuperación.

¹⁸ *Idem*

FUENTES CONSULTADAS

1. ARMS , William Y. Key Concepts in the architecture of the digital library. [En Línea]. *D-Lib Magazine*, July 1995. <<http://www.dlib.org/dlib/July95/07arms.html#junewya4>> [Consulta: noviembre 2007].
2. BROWN, Bruce. Find that file: Special Report - Getting Organized (hard disk files). *Lotus*, June 1989, v5. no. 6, p 88.
3. Digital Library Federation. Benchmark for Faithful Digital Reproductions of Monographs and Serials. Version 1. [En Línea]. *The Digital Library Federation Benchmark Working Group (2001-2002)*, December 2002. <<http://www.diglib.org/standards/bmarkfin.htm>> [Consulta: noviembre 2007].
4. File naming. [En Línea]. *Technical Advisory Service for Images*. <<http://www.tasi.ac.uk/advice/creating/filenaming.html>> [Consulta: octubre 2007].
5. GUENTHER, Rebecca. Naming Conventions for Digital Resources, *January 3, 1996*. [En Línea]. <<http://lcweb.loc.gov/marc/naming.html>> [Consulta: octubre 2007].
6. HEATHER Chapman. The File Naming Habits of Personal Computer Users. [En Línea]. A Master's paper for the M. S. in I. S. degree. March, 1999. 85 h. <<http://neoref.ils.unc.edu:8080/2475.pdf>> [Consulta: enero 2008].
7. Identifiers for Digital Resources. [En Línea]. *The Library of Congress, National Digital Library Program* <<http://memory.loc.gov/ammem/award/docs/identifiers.html>> [Consulta: enero 2008].
8. Joliet Specification: CD-ROM Recording Spec ISO 9660:1988. 1995. [En Línea]. <<http://bmrc.berkeley.edu/people/chaffee/jolspec.html>> [Consulta: octubre 2007].
9. Cornell University. Moving theory into practice: Digital Imaging Tutorial. [En Línea]. <<http://www.library.cornell.edu/preservation/tutorial/technical/technicalC-01.html>> [Consulta: diciembre 2007].
10. Projet de gestion des images numériques à l'Inventaire général : recommandations techniques, mayo 1999. [En Línea]. (Apartado 4.) <<http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/inventaire.htm>> [Consulta: enero 2008].
11. SNAVELY, Cory. Handle Service Conventions and Practices. [En Línea]. *Digital Library production Service*. 2006. <<http://www.umdl.umich.edu/docs/arch/general/handle-conventions.html>> [Consulta: febrero 2008].