

## TITULO

# IDENTIFICADORES DIGITALES: UNA HERRAMIENTA QUE APOYA LA RECUPERACIÓN DE INFORMACIÓN

## AUTORES

Andrés Corrales Rubiano, Candy Yazmín López Herrera

## ABSTRAC (Ingles / Español)

Today, because of the continuing explosion of information, it is necessary that the centers of organizing information evolve in the sense of being mere repositories of printed information to real centers of digital information "global." To achieve these information centers should use all mechanisms provided by new technologies to meet the information needs of a community increasingly demanding.

Within this framework of ideas, libraries and information centers are not immune to this challenge, since much of the documents of interest within his collection is not limited to those tangible resources deposited in the information centers, but extends a large number of electronic resources that are available, in most cases, remotely via the Internet.

Faced with such a situation information centers tend to develop a set of strategies that focus their attention towards a new horizon of digital techniques that allow organizing this mass of information, in order to facilitate access to and hence their recovery. Through this article are some resources that are used to support the electronic information retrieval.

## ABSTRAC

Hoy en día, debido a la continua explosión de la información existente, es necesario que los centros de organización de la información evolucionen en el sentido de ser simples repositorios de información impresa a verdaderos centros digitales de información "global". Para lograrlo estos centros de información deben utilizar todos los mecanismos provistos por las nuevas tecnologías con el fin de satisfacer las necesidades de información a una comunidad cada vez más exigente.

En este marco de ideas, las bibliotecas y centros de información no escapan a este desafío, puesto que gran parte de los documentos de interés dentro de su colección no se limitan sólo a aquellos recursos tangibles depositados en los centros de información, sino que se amplía a

un gran número de recursos electrónicos que son accesibles, en la mayoría de los casos, en forma remota a través de Internet.

Ante tal situación los centros de información tienden a desarrollar una serie de estrategias que centran su atención hacia un nuevo horizonte de técnicas digitales que permiten a organizar esta gran masa de información, con el fin de facilitar el acceso y por ende su recuperación. A través del presente artículo se presentan algunos recursos que son utilizados como apoyo de la recuperación electrónica de información.

## IDENTIFICADORES DIGITALES

Un identificador es una serie de secuencias cortas de caracteres alfanuméricos que identifican recursos en la red: documentos, imágenes, archivos, servicios, cajas electrónicas, y otros recursos. Hacen que éstos sean disponibles bajo una variedad de protocolos<sup>1</sup> y métodos de acceso, tales como HTTP<sup>2</sup> (Protocolo de Internet), FTP (Protocolo de transferencia de archivos), y correo del Internet. Sus funciones principales son:

1. Proporcionar datos acerca de la identificación del documentos,
2. Desarrollar la forma de localizarlos, y finalmente
3. Recuperar la información que contiene para generar nuevo conocimiento.

### URI (Uniform Resource Identifier)

El Identificador Uniforme de Recursos. Es un sistema global que condensa la Dirección (URL) y el nombre (URN) del recurso para identificarlo dentro de la red y de esta forma tener mayor efectividad en su recuperación. Es una cadena corta de caracteres que identifica de manera única un recurso<sup>3</sup>. Es de resaltar que normalmente estos recursos son accesibles a través de una red o sistema.

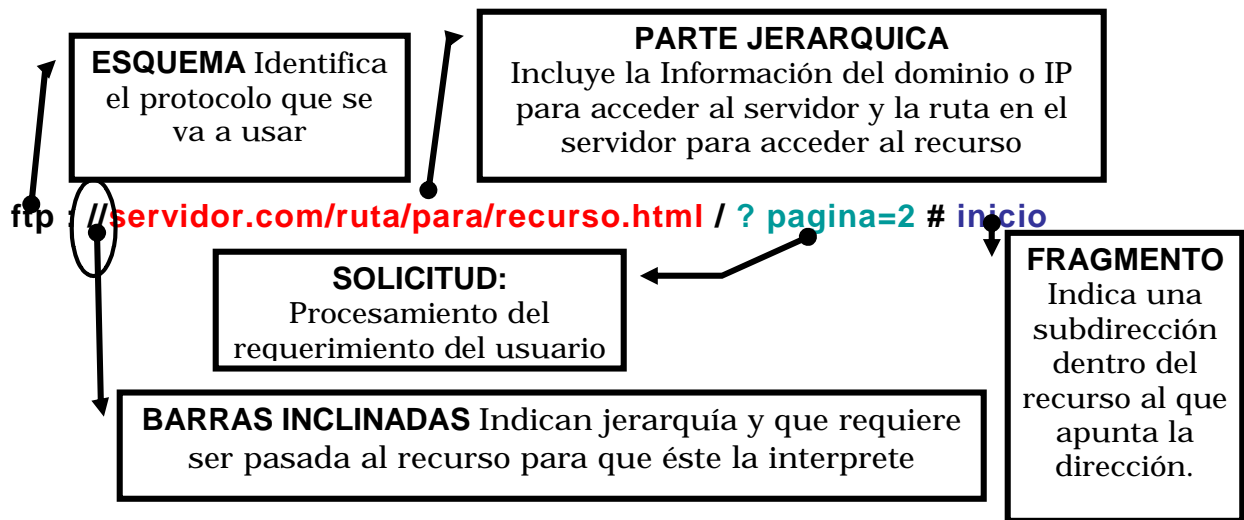
El siguiente esquema ilustra la estructura de un URI:

---

<sup>1</sup> “Conjunto de reglas que controlan la secuencia de mensajes que ocurren durante una comunicación entre entidades que forman una red” WIKIPEDIA. Protocolos. [SF] disponible en línea en: <http://es.wikipedia.org/wiki/Protocolo>.

<sup>2</sup> (HyperText Transfer Protocol) “es el protocolo de alto nivel del World-Wide Web que rige el intercambio de mensajes entre clientes y servidores del Web” ADELL, Jordi. La Internet como telaraña: el world-wide web. España. 1994. Disponible en línea en: <http://www.uv.es/~biblios/mei3/Web022.html>

<sup>3</sup> Entiéndase “recurso” como: servicio, página, documento, archivo de programa, dirección de correo electrónico, enciclopedia, etc



Un URI no se limita a recuperar recursos Web, sino por el contrario tienen la capacidad de llegar al nivel de especificidad de identificar personas, conceptos, etc.

Los URI son administrados y controlados por la IETF (Internet Engineering Task Force) o Grupo de Tareas de Ingeniería de Internet, organización de técnicos encargados de las tareas de ingeniería de telecomunicaciones y principalmente de Internet.

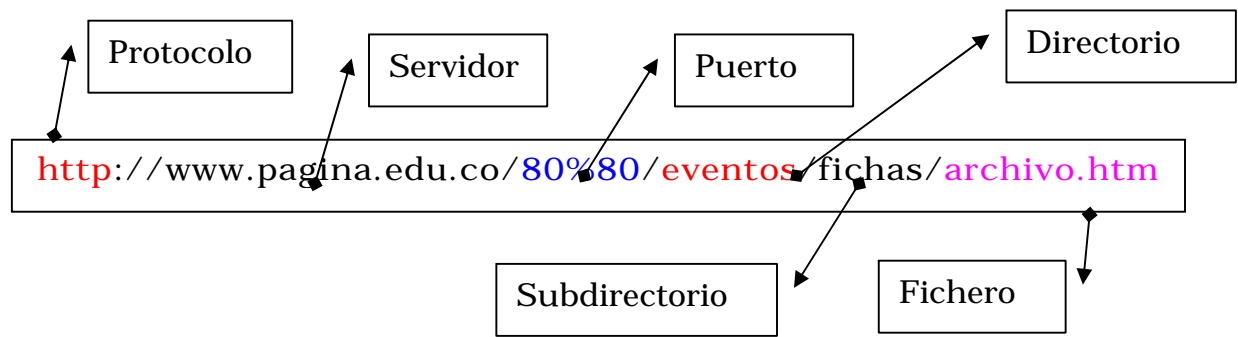
URL (Uniform Resource Locator),

Localizador uniforme de recursos de Internet. (También conocido como la "Dirección Electrónica" del recurso). Es una secuencia de caracteres de acuerdo a un formato estándar, que se usa para identificar la localización de los recursos en la Web.

Por decirlo de cierta forma crea la conexión entre el recurso y el Servidor, brindando información sobre donde se encuentra un recurso y cómo se puede acceder a él.

Mediante esta cadena de caracteres se asigna una dirección de carácter único a cada uno de los recursos de información disponibles en la Web, de esta forma permite que el navegador los encuentre y los muestre de forma adecuada. El URL combina el nombre del ordenador que proporciona la información, el directorio donde se encuentra, el nombre del fichero y el protocolo a usar para recuperar los datos.

El formato general de un URL es:



La diferencia con respecto al URI, radica básicamente el URL se limita a identificar recursos Web, asigna las coordenadas del servidor donde se encuentra el recurso, mientras que el URI identifica el recurso independientemente de donde se encuentre sea en Internet o en un sistema o en una red local y apunta directamente al recurso por medio de una subdirección o FRAGMENTO dentro del recurso al que apunta la dirección., independientemente del servidor donde se encuentre. Por decirlo en pocas palabras todas las URL son URI, pero más cortas.

URN (Uniform Resource Name),

Nombre Uniforme de Recursos. Es el nombre con el que se distingue como único un recurso dentro de la red global. El URN realiza la identificación digital basándose en cuatro características sustanciales:

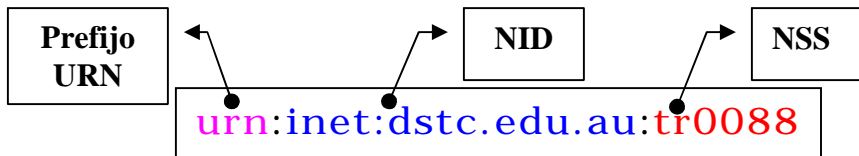
- Son globales y Únicas: Tienen alcance mundial, Sirven bajo cualquier plataforma tecnológica o cualquier sitio geográfico y se puede tener la certeza que nunca se va a repetir dentro del universo de la red.
- Accesibles por medio de la red: El nombre es valido o genera utilidad solo para utilización de recursos en red.
- Persistentes: Se pretende que la vida útil del URN no sea otorgado de manera temporal, sino por el contrario que pervivan en el tiempo como algo inherente al recurso.
- Independencia: Este mecanismo no depende de la información de su localización de la ubicación del recurso (Servidor o Sitio Web).

La sintaxis de un URN esta compuesta por tres bloques, separados por dos puntos (":"):

- Secuencia de Caracteres "urn:": Al inicio le indica al sistema que se trata de una URN.

- NID: (Namespace Identifier) Nombre único registrado para el sistema identificador
- NSS: (Namespace Specific String) Cadena específica de espacio de nombre

El control de los URN esta a cargo de la IANA<sup>4</sup> que llevan el registro global de los NID, la sintaxis normalizada es la siguiente:



## DOI

Es un identificador de objetos Digitales que proporciona algo más allá de una localización, pues se trata de un sistema por el cual se asigna un “nombre” único de identificación de documentos electrónicos o en otras palabras, de recursos digitales dentro de la red. Proporciona un sistema para la identificación persistente o en otras palabras que tenga la propiedad por la que la información del sistema puede persistir (sobrevivir) durante el tiempo que sea requerida. Esto con el fin de ser recuperable en el tiempo. Dicho nombre consiste en un número permanente que identifica a puede ser por ejemplo una revista online, o artículo, imagen, citación, applet, script u cualquier objeto digital.

Este sistema se caracteriza por:

- ✓ Identificar o nombrar de forma única una pieza de contenido electrónico
- ✓ Servir como un vínculo persistente y estable a la localización de ese contenido en el Web.
- ✓ Definir un nombre para buscar una dirección en un directorio
- ✓ Servir para el contenido digital lo mismo que el ISBN para lo hace para los libros impresos, incluyendo datos como:
  - ✓ número de identificación,
  - ✓ los metadatos asociados y
  - ✓ un sistema para proveer o localizar el objeto digital.

La sintaxis es la siguiente:

---

<sup>4</sup> Internet Assigned Numbers Authority, organismo que regula los números asignados en Internet

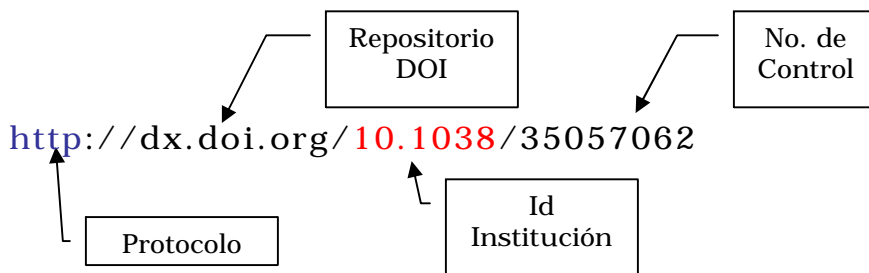


Figura 1.3

(Para adquirir este identificador es necesario contactar con su proveedor – INTERNATIONAL DOI FUNDATION y su precio por registrar el prefijo es de 1000 dólares más un canon anual dependiendo la cantidad de prefijos que se hayan registrado).

#### HANDLE

Según diversos autores “Handle” puede tener las siguientes connotaciones:

Diccionario de informática y Telecomunicaciones:

“1. Identificación corta que se usa para acceder a cierto recurso (de identificación más larga). En general dicha identificación es asignada por el sistema, no por el usuario.”

“2. Cuando se selecciona un objeto gráfico en aplicaciones que las aceptan, son asas los pequeños cuadros que cuando se seleccionan y desplazan permiten mover todo el objeto gráfico”

Diccionario de computación:

“1. En gráficos por computador, ubicación en una imagen gráfica que puede moverse para darle una nueva forma, por lo general es un pequeño cuadrado.”

“2. Nombre o número temporal asignado a un archivo, fuente o tipo u otro objeto. Por ejemplo un sistema operativo puede asignar un número secuencial a cada archivo que abre como una forma de identificación y rastreo de este”

Diccionario para usuarios de computadoras e Internet:

“1. Identificador, controlador de tamaño.”

“2.En administración de memoria, Canal de acceso a un bloque de memoria extendida. El parámetro /Numhandles=Num se usa para informar a hymen.sys el número de identificadores que debe controlar.”

“3.En Windows, la pila global consta de todas las tareas, llamadas objetos que han recibido memoria y donde cada objeto tiene asignado un identificador.”

## Diccionario de Internet y Redes de Microsoft

### Descriptor (Handle)

“1.En la comunicación el línea como en los salones de charla y tableros de anuncios electrónicos, es el nombre que una persona utiliza para identificarse. El descriptor es comparable a un alias o un apodo y es similar a los utilizados en las radios de banda ciudadana.”

“2.Un identificador alfanumérico univoco de hasta 10 caracteres asignado por INTERNIC a los nombres de dominio, contactos y registro de red en su base de datos de nombres de dominio. El descriptor NIC se utiliza como una forma abreviada de localizar registros y garantizar la precisión de la base de datos. También llamado descriptor NICEn administración de memoria, Canal de acceso a un bloque de memoria extendida. El parámetro /Numhandles=Num se usa para informar a hymen.sys el número de identificadores que debe controlar.”

En fin, muchas son las vertientes que tiene este sistema, pero en pocas palabras podemos definir que handle tiene las siguientes clases fundamentales:

1. Bajo el ambiente de Windows, cada vez que se abre o ejecuta una pantalla, el sistema operativo asigna un número de identificación temporal al objeto y de esta forma controla la asignación de un espacio de memoria temporal para mantenerlo abierto de tal forma que pueda ser visualizado y manipulado por el usuario.
2. En la comunicación interactiva e Internet es un código alfanumérico de 10 dígitos que se asigna a cada nombre de dominio.
3. Finalmente y es el que nos interesa es un sistema por el cual se asigna un código identificador único de recursos. Este identificador sirve

como recuperador de la información de ubicación del recurso y permite su enlace de una manera más directa desde Internet.

Actualmente su administración esta a cargo del CNRI y su uso esta condicionado en cuanto al manejo que se le de (personal o Público)

Handler

Manipulador. Componente de los handles y se trata de una rutina de software que realiza una determinada tarea. Por ejemplo, cuando se detecta un error, se llama a un manipulador de error para recuperarse de esa condición.

Sirve para manejar identificadores de objetos de forma independiente de su implementación y permite cambiar fácilmente la implementación de los identificadores (Sea compuesto por caracteres alfanuméricos (integer) o cadenas de texto (String), etc.) hacia cualquier tipo básico o clase primitiva, sea sencilla o compuesta.

## LOS IDENTIFICADORES DIGITALES COMO RECURSOS PERSISTENTES

Un sistema de persistencia dentro del ambiente de programación, proporciona la abstracción o el “alquiler” de un espacio que se destina para que los objetos digitales (Documentos, Imágenes, graficas, etc.) tengan la propiedad de coexisten en el mundo digital indefinidamente. Esta propiedad de los objetos puede servir como base para un desarrollo más fácil e integrado de otras herramientas, como sistemas de gestión de bases de datos orientados a objetos.

La función principal de los Recursos persistentes radica en la interoperabilidad (término a menudo traducido como inteorperatibilidad, del inglés interoperability) es la condición mediante la cual sistemas heterogéneos pueden intercambiar procesos o datos<sup>5</sup>.

Se pretende que la vida útil de un identificador persistente ha de ser permanente. Es decir, que será único a nivel mundial para siempre, y bien puede ser utilizado como una referencia a un recurso mucho más allá de la vida útil de los recursos que se indican o de cualquier autoridad de nombres que participan en la asignación de su nombre

---

<sup>5</sup> WIKIPEDIA.ORG. Interoperatibilidad. Disponible en línea en: <http://es.wikipedia.org/wiki/Interoperabilidad>



## APLICACIÓN PRÁCTICA DE LOS IDENTIFICADORES DIGITALES DENTRO DE LA RECUPERACIÓN ELECTRÓNICA DE INFORMACIÓN.

En el tema de recuperación de información electrónica, los identificadores digitales juegan un papel importante, puesto que utiliza cada uno de los recursos anteriormente descritos como llaves que amarran los recursos con la necesidad del usuario.

Una de las aplicaciones que ayuda a ilustrar el uso de los identificadores digitales lo constituyen los “buscadores Web” que es un software complejo que conforme a algunos criterios específicos busca en una base de datos o repositorio de documentos en la Web, representando los datos resultantes del contenido intelectual de los documentos o recursos disponibles en la red permitiendo así la consulta por medio de sus índices, de los que luego aplica un algoritmo para resolver las búsquedas y recuperar sólo aquellos documentos que contienen las palabras clave o parámetros acordados introducidos en la búsqueda.

Ahora bien, el tema de identificadores digitales se evidencia de una forma más clara examinando las diferentes clases de buscadores Web:

- Directorios o Índices: Son aplicaciones controladas de forma manual que administra o gestiona grandes bases de datos que contienen la referencia de URIs<sup>6</sup> de recursos en Internet y los clasifica en categorías temáticas.

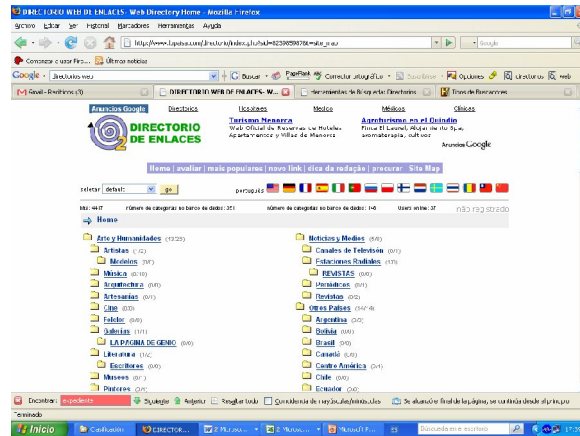
Los directorios utilizan dos formas de capturar la información, en la primera forma son los productores de los recursos quienes envían directamente su información a la organización que realiza el directorio para que ellos realicen todas las gestiones de inserción de los respectivos vínculos (Captura Pasiva) y la otra la realiza directamente la organización que administra el directorio quienes buscan y capturan las URLs (Captura Activa)

Este tipo de buscador presenta como problema que al ser operado de forma manual la actualización es lenta y además sumado a que intentan por medio de esta técnica abarcar todos los recursos de Internet hacen que se poco eficiente.

La siguiente figura ilustra la disposición de la información:

---

<sup>6</sup> Direcciones de páginas, títulos, descripciones, etc,



Según se observa los diferentes enlaces se realizan por medio de vínculos que contienen la referencia de las URLs de las Páginas Web categorizadas.

Sin embargo, puesto que se utilizan las URLs como identificador digital cuando el recurso es movido la referencia que inutilizada y el vínculo se pierde.

- **Motores de Búsqueda:** Es un software más robusto que recopila URLs y las indiza de forma automatizada, utilizando una serie algoritmos avanzados de búsqueda que analizan las páginas que tienen en su memoria y recuperan resultados más específicos.

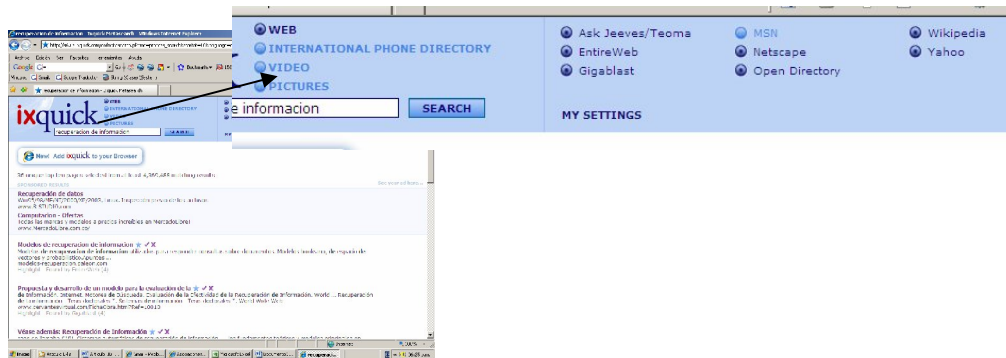


En la gráfica anterior se utilizó como ejemplo el motor de búsqueda de “GOOGLE” que es uno de los más populares del ciberespacio.

La captura de la información se realiza a través de “Arañas” o “Spiders” que son agentes digitales o algoritmos automatizados que rastrean las URLs de la Web a través de enlaces o URLs predeterminadas para luego ir extendiéndose y multiplicándose en forma de araña, de esta forma por lo menos una vez al mes se recopilan los datos y se mantiene actualizada la base de datos de forma continua.

Metabuscadores: Aplicación que aunque no posee una base de datos propia se comunica, enlaza y traslada de forma simultánea los parámetros de búsqueda proporcionados por el usuario en motores de búsqueda o directorios externos, de esta forma compila los resultados de la búsqueda de una forma más global y así recuperar los recursos más relevantes de acuerdo a las necesidades expresadas. Para tal fin los metabuscadores utilizan toda la tecnología de identificación de recursos disponibles.

El siguiente gráfico visualiza el vínculo que realiza un metabuscador, en él se visualiza como a través de ellos se puede recuperar información utilizando varios motores de búsqueda:



En la gráfica anterior, se tomo como ejemplo el metabúscador de “IXQUICK” uno de los más populares de Internet.

Todos ellos utilizan la tecnología de identificación digital. Sin embargo en el caso de los recursos Web, la parte de URLs pero presenta como problema que si se mueve el recurso se pierde inmediatamente el índice que lo enlaza con el buscador.

Otro ejemplo que puede ilustrar la utilización de identificadores digitales enfocado a la recuperación de información, se visualiza en el hecho que dentro del concepto de BIBLIOTECA DIGITAL la recuperación se realiza por medio de una base de datos que enlaza los recursos disponibles en red por medio de llaves digitales (URIs) que le permiten al sistema mostrarlos en pantalla.

Puesto que la biblioteca digital se apoya en la recuperación de metadatos que son información descriptiva acerca de un objeto o recurso, en otras palabras es la etiqueta que brinda información acerca del contenido del

recurso, mediante este mecanismo permite a los diferentes algoritmos de recuperación de información extraer los enlaces o descriptores claves fuente para capturar los recursos requeridos. Ahora bien, los nombres son los que proporcionan una guía que identifica de manera única los objetos digitales que se vinculan en la biblioteca o colección y es precisamente allí donde entran a operar los metadatos que por decirlo de cierta forma son comparables al ISBN de los libros en una biblioteca tradicional, es por ello que se puede citar los recursos, recuperar de ellos información y para este caso crear vínculos entre objetos y / o recursos.

En cuanto a herramientas disponibles aplicables se pueden encontrar los “handles” y los “DOI” que son aseguran la persistencia de los recursos y sobre todo la confiabilidad de su ubicación.

#### Key Words

Handle, identificador de recursos digitales, recuperación de información, metadatos, DOI, Recursos persistentes, URL, URN, URI, metabuscadores, motores de búsqueda, Directorios,

#### Bibliografía

1. MORENO MARTIN, Arturo. Diccionario de informática y telecomunicaciones. 1ª. Ed. Barcelona. 2001. P 296
2. PFAFFENBERGER, Bryan. Diccionario para usuarios de computadoras e Internet. México. 1996. p 229
3. FREEDMAN, Alan. Diccionario de Computación. 7ª Ed. Bogota. 1996. P 253
4. MICROSOFT Corp. Diccionario de Internet y Redes de Microsoft. Madrid. 2003. P
5. RUSSINOVICH, Mark. Handle v3.20. 2006. Disponible en línea en: <http://www.microsoft.com/latam/technet/sysinternals/ProcessesAndThreads/Handle.aspx>
6. HANDLE.NET. Services: Global Handle Registry® 2006. Disponible en línea en: <http://www.handle.net/introduction.html>
7. PWeb Online. Preguntas Más Comunes – FAQs. 2001. Diponible en línea en: <http://www.pweb.com.ar/faq/04dominio/dominio21.htm>

8. ESPAÑA, Universidad de Alicante. Handle: identificador persistente. [¿2006?]. Disponible en línea en: <http://www.ua.es/rua/ayuda.html#handle>
9. MARMONTI, Emiliano. Linking Tecnology (introducción). Memorias 38ª Reunión de Bibliotecarios ABSGRA. Buenos Aires. Abril 2006. Disponible en línea en: <http://www.cnea.gov.ar/cac/ci/7-marmonti.pdf>
10. MERLO VEGA, José Antonio. [En línea - .htm] Salamanca. 2001. Disponible en: <http://exlibris.usal.es/merlo/escritos/bilbao2.htm>
11. LAMARCA Lapuente, María Jesús Normalización, localización, e Identificación. Disponible en línea en: <http://www.hipertexto.info/documentos>
12. OCLC. Acerca de OCLC. Facilitando el acceso a la información. Disponible en línea en: <http://www.oclc.org/americalatina/es/about/default.htm>
13. ÁLVAREZ Gutiérrez, D y otros. Persistencia en sistemas operativos orientados a Objetos. Ventajas para los sistemas de gestión de Bases de datos. Oviedo. [SF] Disponible en línea en: <http://www.di.uniovi.es/ioos/publications/oodbms/corunia96.pdf>
14. DODERO, Juan Manuel y otro. Patrón Handler. [En línea - .Pdf]. Madrid. 2003. Disponible en: [http://peterpan.uc3m.es/docencia/p\\_s\\_ciclo/pa/web/apuntes/handler.pdf](http://peterpan.uc3m.es/docencia/p_s_ciclo/pa/web/apuntes/handler.pdf)
15. HISPAMEDIOS. URL y URI. [En línea - .html]. 2007. Disponible en: <http://www.hispamedios.com/articles/id34-url-y-uri>
16. PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE. Estructura de una dirección electrónica (URL). [En línea - .Xml] Santiago. 2007. Disponible en: [http://www.puc.cl/sw\\_educ/gnosis/H/gnosish1.htm](http://www.puc.cl/sw_educ/gnosis/H/gnosish1.htm)
17. QUITIAN Fino, Freddy y otro. Identificadores de Recursos Digitales. [En línea - .doc]. Georgia. 1996. Disponible en: <http://hdl.handle.net/123456789/271>
18. MÉNDEZ Rodríguez, Eva. MERLO Vega, José Antonio. Localización, identificación y descripción de documentos Web: tentativas hacia la normalización. [En línea - .Pdf]. Madrid. [SF]. Disponible en:

[http://www.uag.mx/eci/infosource/Articulos/Profinfo/organizacion\\_conocimiento.pdf](http://www.uag.mx/eci/infosource/Articulos/Profinfo/organizacion_conocimiento.pdf)

19. ARMS, William. Uniform Resource Name. [En línea - .Xml]. Georgia. 1996. Disponible en: <http://www.dlib.org/dlib/february96/02arms.html>
20. GUTIÉRREZ, Cristian. URI y URL. [En línea - .Xml]. Santiago. 2007. Disponible en línea en: <http://www.tejedoresdelweb.com/307/article-5671.html>
21. CLEVELAND, Gary. Bibliotecas Digitales: Definiciones, aspectos a considerar. Y Retos. . [En línea - .pdf]. Santiago. 2007. Disponible en: <http://redalyc.uaemex.mx/redalyc/pdf/285/28540207.pdf>
22. MERCOVICH, Eduardo G. Bases en la Web. [Buenos Aires?] 1996. [En línea: .htm]. Disponible en: <http://www.gaiasur.com.ar/infoteca/varios/urls.html>
23. RUIZ Portanova, Noelia. Recuperación de Información en la web. Madrid. 2006. [En línea: htm]. Disponible en: <http://recupinformacion.webcindario.com/buscadores%20web.html#motores>
24. GARCÍA, María Alfonso. Directorios o índices de búsqueda. Madrid. 2006 [En línea: . htm] Disponible en: [http://es.geocities.com/directorio\\_sari/](http://es.geocities.com/directorio_sari/)
25. MASADELANTE.COM. ¿Cómo funciona un motor de búsqueda? Alicante. 2007. [En línea: . htm] Disponible en: <http://www.masadelante.com/faq-como-funciona-motores-de-busqueda.htm>

VILLEGAS Quezada, Carlos. Análisis comparativo de herramientas informáticas para “minería de texto” y sus posibilidades de aplicación en el análisis de documentos de educación a distancia seleccionados en el web. [Madrid?]. 2006. [En línea: .htm] Disponible en: <http://www.uned.es/catedraunesco-ead/villegas/apartado2.htm>