

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE FILOSOFÍA Y LETRAS
DIVISIÓN DE ESTUDIOS DE POSGRADO

**LA ORGANIZACIÓN Y REPRESENTACIÓN DE LA INFORMACIÓN
POR MEDIO DE SISTEMAS DE CLASIFICACIÓN
AUTOMATIZADOS EN REDES**

TESIS
QUE PARA OBTENER EL GRADO DE
MAESTRA EN BIBLIOTECOLOGÍA

PRESENTA:

LEONOR GARCÍA URBANO

Asesor Mtro. Ramiro Lafuente López



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

TABLA DE CONTENIDO

Introducción	Págs.
Capítulo 1	
<i>Aspectos generales sobre el uso y disposición de la información en la Web</i>	
Influencia de la tecnología en las nuevas formas de distribución de la información.	1
El manejo de la información en la Web, desde un enfoque social y comercial.	6
Influencia de la Web en la percepción informativa del individuo.	12
Capítulo 2	
<i>Herramientas de organización, representación y acceso a la información en la Web</i>	
La biblioteca a través de los procesos de organización y disposición de la información en la Web.	21
Estructura y operatividad de las bases de datos documentales.	28
Estructura y operatividad de los índices y directorios.	34
Capítulo 3	
<i>Tendencias sobre la organización, representación y disposición de la información en la Web</i>	
Dificultades en la definición de los recursos en la Web.	45
<i>Metadatos</i> para la gestión y descripción de recursos en la Web.	51
Propuesta para el desarrollo de la Web Semántica.	56
Capítulo 4	
<i>Viabilidad de los sistemas de clasificación documental en la Web</i>	
Sobre la naturaleza de los sistemas de clasificación documental.	65
Los sistemas de clasificación documental a través de algunos de sus proyectos en la Web.	73
Conclusiones	84
Bibliografía	
Anexo 1: Modelo Dublín Core	
Anexo 2: Clasificaciones documentales relevantes	

INTRODUCCIÓN

La Internet ha contribuido a modificar las circunstancias en las que se promueve el intercambio de la información. Representa un nuevo espacio en el que se diluyen las nociones de tiempo y distancia cuando se navega por la Web, se transfieren archivos o se usa el correo electrónico. Es una puerta abierta a todo un mundo informativo, tal como si, al andar por la calle, uno pudiera tener la capacidad de escucharlo o verlo todo a voluntad, sin importar las distancias. En consecuencia, la Internet se ha establecido bajo un entorno atractivo, dinámico y deslumbrante y, precisamente por ello, en ocasiones resulta difícil visualizar las repercusiones reales de sus alcances tecnológicos, informativos y sociales.

Dentro de la Internet, la Web se describe como un espacio cambiante, heterogéneo, efímero y caótico, debido a la gran cantidad y diversidad (en formatos y contenidos) de información, que por ella aparece, desaparece y fluye cotidianamente. Debido a estas características, para el individuo común, resulta difícil la tarea de localizar y recuperar la información que requiere con la finalidad de atender una necesidad o resolver un problema determinado. Más aún, a pesar de las alternativas que los sistemas de información ofrecen en la Web, con mucha frecuencia, también es posible observar que el individuo no sabe cómo localizar la información que necesita o no encuentra un camino relativamente fácil para hallar la información que requiere.

A partir de estas observaciones, es factible determinar que, entre la necesidad del usuario para recuperar cierto tipo de información y la incapacidad de un sistema para hacer recuperable la

información que el usuario necesita, existe un estado crítico sobre el acceso a la información.

Frente a los problemas que manifiestan las nuevas formas de representación, organización, disposición y recuperación de la información en la Web, es conveniente hacer un alto en el camino para comentar que la biblioteca tiene ya siglos de experiencia en cuanto al tratamiento y organización de la información. Se ha mantenido en pie como una institución que integra y canaliza la información con un alto sentido de responsabilidad y compromiso sobre el uso social de la misma y, justamente por ello, resulta poco comprensible que los sistemas de información de las bibliotecas no se hayan logrado proyectar con un mayor nivel de incidencia en la Web.

Debido a los planteamientos anteriormente expuestos, se justifica un trabajo de investigación que permita analizar, de manera general, por qué la Web no ha podido tener la capacidad de ofrecer una respuesta eficiente y satisfactoria a todas las posibles inquietudes informativas de sus usuarios. Así, para atender a esta pregunta, es necesario resolver también las siguientes interrogantes: i) ¿qué aspectos influyen en la determinación de los sistemas de información en la Web?, ii) los esquemas y principios, bajo los cuales operan los sistemas de información en la Web, ¿son funcionales?, iii) ¿cuál es la tendencia y utilidad de los sistemas de información en la Web? Con base en estos cuestionamientos, se contemplan la hipótesis y la metodología siguientes.

A partir de los principios, esquemas y tecnologías desarrollados para tener acceso a la información contenida en la Web, es posible identificar las perspectivas y tendencias, que se presentan para la organización y recuperación de la información digital, y determinar la

viabilidad de las mismas para la creación de principios y fundamentos que permitan el desarrollo de nuevas formas de organización y acceso a la información digital.

El presente trabajo se apoyará en una metodología de investigación cualitativa, que se sujetará al análisis documental e interpretativo acerca de las formas de representación y organización de la información para facilitar su acceso desde la Web. Dicho análisis se aplicará, además, para obtener una visión general sobre los alcances de los sistemas de información a partir de las propuestas, dificultades y la propia naturaleza de los esquemas normativos y principios de organización y recuperación bajo los cuales operan dichos sistemas, en general, con el propósito de comprender su viabilidad.

Un primer problema que se manifestó en esta investigación fue determinar cómo identificar y abordar los aspectos que influyen en la estructura que mantienen los sistemas de información en la Web. Luego se consideró que, fuera de los aspectos tecnológicos, si el uso y manejo social que se pretende brindar a un recurso informativo, determina el diseño de un sistema de información, entonces, sería posible encontrar las respuestas en el entorno social bajo el cual se gestan dichos recursos. Pero realizar una investigación exhaustiva, abarcando todos los posibles aspectos que confluyen para condicionar las formas de acceso a la información y determinar, con ello, la utilidad de los sistemas informativos, sería excesivo. Por otro lado, excluir el enfoque social, dificultaría la tarea de comprender e interpretar el comportamiento de estos sistemas.

Debido a lo anterior, sólo se considerarían algunas de las manifestaciones más relevantes que se han definido en el entorno

social de la información, es decir, la influencia tecnológica, el comercio electrónico, los supuestos ideológicos de la llamada Sociedad de la Información y los aspectos relacionados con la influencia por el uso de la Web sobre el individuo, así como el nivel cognitivo e intencional a partir del cual el individuo realiza una búsqueda.

Para definir la funcionalidad de los sistemas de información en la Web, se contemplaría la necesidad de destacar, por una parte, el trabajo que se realiza desde la biblioteca (y por extensión en la propia bibliotecología), para sistematizar la información en la Web y, por otra parte, se efectuaría un análisis con base en la descripción operativa que manejan las bases de datos documentales, los directorios y los buscadores de la Web para señalar sus ventajas y desventajas.

La exposición de las fortalezas y debilidades de dichos esquemas, también facilitaría la labor de justificar la intencionalidad con la que se desarrollan las propuestas y los esquemas normativos de la Web, por lo que también se optó por describir los más reconocidos o divulgados para completar este análisis.

Los principios bajo los cuales se determinan los sistemas de información en la Web permitirían contrastar, las alternativas propuestas para impulsar los sistemas de la Web, así como la propuesta que, para el mismo efecto, establecen los sistemas de clasificación documental. De tal modo que, es posible obtener los elementos argumentativos suficientes para valorar la utilidad y viabilidad de dichos sistemas, en el contexto de la organización y los retos del acceso a la información digital.

La presente investigación se desarrolla en cuatro capítulos. El primer capítulo, muestra algunas de las principales manifestaciones, que se han establecido en el entorno de la información, a partir de su uso dentro del contexto tecnológico, social, comercial y del individuo, para tratar de averiguar cuál es el sentido en el que se distribuye la información en la Web y, si este aspecto influye en la forma en que se utiliza y estructura la información.

En el segundo capítulo, se destaca la actividad que, desde la perspectiva de la bibliotecología, se realiza para organizar y difundir la información en la Web, con la finalidad de demostrar cuál es el papel que desde la disciplina se define para sistematizar la información. Asimismo, se señalan los aspectos técnicos y operativos de las bases de datos documentales, los índices y los directorios en la Web, a partir de los cuales se definen las ventajas y desventajas que ofrecen para el acceso a la información.

En el tercer capítulo, se señalan las dificultades que existen para reconocer e identificar un recurso de información documental en la Web y se especifican las alternativas que, desde el enfoque de los *metadatos* y algunos de los esquemas normativos más relevantes, se plantean para procesar, representar y localizar los recursos digitales. Se describe la propuesta de la Web Semántica, la cual se establece como una tendencia a partir de la Web actual, y se identifican los principios y el esquema sobre los que pretende operar para dar acceso a la información digital.

En el cuarto capítulo, se define la naturaleza de los sistemas de clasificación desde el punto de vista de la bibliotecología, se señalan los principios y propósitos bajo los cuales operan y se determina, a partir de los proyectos para utilizar los sistemas de clasificación

documental en la Web, cuál sido el impulso de éstos para la organización y recuperación de los recursos documentales digitales en la Web.

CAPÍTULO 1

ASPECTOS GENERALES SOBRE EL USO Y DISPOSICIÓN DE LA INFORMACIÓN EN LA WEB

Influencia de la tecnología en las nuevas formas de distribución de la información

La tecnología siempre ha jugado un importante papel en las formas de producir y difundir la información. Bajo esta premisa, es posible argumentar que la segunda mitad del siglo pasado puede ser reconocida como la cuna de una nueva etapa en la historia de la producción y difusión de la información, ya que, con el surgimiento de las llamadas *Tecnologías de la Información y de la Comunicación (TIC)*, se impulsó una revolución informativa, debido a que el uso social de dichas tecnologías, no sólo influyó en la transformación del quehacer sobre el manejo y difusión de la información, al alterar los modelos tradicionales para procesarla, registrarla y comunicarla, sino que, además, modificó las formas para buscar, consultar y utilizar dicha información.

A partir de las décadas que siguieron a la primera mitad del siglo pasado, es posible observar que no sólo comenzaron a producirse enormes volúmenes de información, a través de tecnologías cada vez más complejas y reducidas, con capacidades ascendentes para el procesamiento de la información, sino que también se promovió el uso masivo de grandes cantidades de información. Mas el impulso informativo generado por el desarrollo de estos nuevos soportes y canales de transmisión, prosperó dentro de un espacio tecnológico particularmente diferente: la Internet.

Sobre la Internet es conveniente señalar que ésta tuvo su origen en los años sesentas (a raíz de la *Guerra Fría*, la cual se reconoce como el periodo de tensiones entre Estados Unidos y el bloque soviético que siguió a la Segunda Guerra Mundial), cuando se difundió el miedo por otra guerra. Por tanto, el *Ministerio de Defensa* de Estados Unidos impulsó un proyecto de investigación que tenía la finalidad de mantener las comunicaciones en caso de estallar una guerra mundial. Entonces, se hizo necesario asegurar la supervivencia de un centro de telecomunicaciones ante un posible ataque de tal magnitud. La idea era crear una red compuesta por computadoras, en la que todos los nodos (o intersecciones) tuvieran la misma importancia, de tal forma que cada nodo permitiera que los datos siguieran distintas rutas hasta llegar a su destino sin que la destrucción de algún nodo afectara el tráfico del resto de los datos en general.

Con el tiempo, en 1969, la *Agencia de Proyectos de Investigación Avanzada (ARPA)*, del *Pentágono*, financió la primera red denominada *ARPANET*, la cual mantuvo un periodo de evolución hasta 1983 cuando nació realmente la Internet. *ARPANET* desapareció en 1989, el mismo año en que Tim Berners-Lee, investigador del *Centro de Investigación Física Europeo de Alta Tecnología (CERN)* en Ginebra, Suiza, elaboró una propuesta sobre un sistema hipertextual para Internet, es decir, el primer esbozo de la *World Wide Web (WWW)*, cuyas características esenciales perduran hasta nuestros días.

La Web, fue el último servicio incorporado a la Internet y comenzó a trascender desde 1993 con el lanzamiento de *Mosaic*, el primer navegador de dominio público compatible con *Unix*, *Windows* y *Macintosh*. Este hecho marcó el momento en que la Web se dio a conocer en el mundo, extendiéndose primero hacia las universidades

y laboratorios, y en cuestión de meses, entre el público en general. El éxito de la WWW fue tal que incluso los usuarios comenzaron a confundirla con la propia Internet.

Esta breve historia sobre el nacimiento de la Internet, permite entender algunas de sus principales características operativas, entre ellas, la flexibilidad para incrementar el número de nodos en la Web situación que fue aprovechada, en su mayor parte, por los empresarios, quienes invirtieron en este nuevo entorno para promover sus productos y servicios, generando una mayor cantidad y diversidad de la oferta informativa. Lo anterior, aunado a la creciente demanda propiciada por la gran capacidad de la Web para albergar y permitir que fluya la información, contribuyó a establecer las condiciones bajo las que se definieron los nuevos esquemas de distribución y recuperación de la misma. No obstante, la dificultad para localizar información útil y la recuperación de información irrelevante, se hicieron cada vez más evidentes, desencadenando una crisis informativa.

Pese a las dificultades mencionadas en el párrafo anterior, algunos sistemas en la Web se han beneficiado, gracias a la implantación de ciertos principios, métodos y el conocimiento aplicado de diversos especialistas de la informática, la lingüística, la bibliotecología, etcétera, quienes se han preocupado y ocupado por el diseño de dichos sistemas, para encauzar la recuperación de cierta información específica en la Web. Al respecto, conviene hacer un paréntesis para comentar que la influencia entre la tecnología y la práctica de la bibliotecología ha sido recíproca, porque tanto el campo de aplicación de la bibliotecología se ha dejado influir por el desarrollo de las TIC, como las TIC también se han adaptado a las necesidades de procesamiento y disposición de la información en la Web.

Con base en todo lo expuesto anteriormente, es posible observar que aunque la tecnología ha sido un factor determinante para promover el uso y manejo de la información, a través de una mayor distribución y disposición de la misma, la Web no ha podido resolver el asunto de encauzar la información para satisfacer cada una de las necesidades de sus usuarios. Y parece que ahí se encuentra uno de los focos del problema ¿es posible encauzar adecuadamente todo lo que existe en la Web? Si coincidimos en una respuesta negativa, entonces, ¿por qué pretender o insistir que todo en la Web deba ser recuperable?

Entre los problemas de acceso a la información en la Web, no sólo se encuentran los derivados de los grandes flujos de información (como la falta de normatividad y estandarización de sus recursos), sino la ausencia de procesos selectivos y de organización para poder valorarlos y encauzarlos. Por ello, el uso de ciertos sistemas en la Web, que sí contemplan dichos aspectos, ha prevalecido porque ofrecen una mayor capacidad de satisfacción a sus usuarios. Con ello no se pretende afirmar que el resto de los sistemas en la Web no sean funcionales, sino que, más bien, en comparación con aquéllos no logran el nivel de satisfacción que puede lograr, por ejemplo, un sistema de clasificación documental.

A pesar de la visión generalizadora de la Web para promover el acceso a toda su información, la biblioteca ha mantenido una clara posición y se ha establecido como una *unidad informativa*, selectiva e independiente, a través de la disposición de sus catálogos en línea o como entidad organizativa, por ejemplo, por medio de la denominada *biblioteca digital*. De ese modo, la biblioteca no sólo ha demostrado su capacidad de adaptación a las TIC, sino que también justifica la importancia de su funcionalidad, porque se ha mantenido sobre

ciertos principios tradicionales, relacionados con la selección, organización, estandarización y normalización de la información.

Por otra parte, y, desde un punto de vista más específico, el acceso informativo en la Web también se encuentra condicionado por otros aspectos tecnológicos y técnicos básicos, por parte del usuario, tales como la disponibilidad de una infraestructura tecnológica (al menos una computadora personal y una conexión a Internet), el conocimiento sobre los procedimientos para poder operar dicha tecnología, idealmente, la capacidad para intuir o saber dónde localizar la información necesaria (que se define con base en su propia experiencia para realizar búsquedas), así como la capacidad para definir o saber cómo plantear la información que busca (en términos de los sistemas que manejan dicha información en la Web).

Como se ha visto, la influencia de la tecnología ha propiciado una crisis y una revolución en el marco de la información. Pero, dentro del ámbito de la bibliotecología, cabe aclarar que estos términos aluden a conceptos completamente distintos a los establecidos desde un enfoque sociológico. La *revolución informativa* se refiere a la transformación de una serie de procesos que tienen que ver con el registro, almacenamiento y difusión de la información, mientras que la *crisis de la información* debe entenderse como la dificultad para localizar información relevante, frente a las necesidades informativas que establece un individuo, en función del entorno en que se desenvuelve. Más aún, el concepto de *crisis de la información* se encuentra ligado al de *redundancia informativa*, es decir, a la duplicidad de la información disponible en la Web, mezclándose con el ruido o el desperdicio informativo, lo que, como ya se ha comentado, determina la dificultad para hallar información útil a cierto tipo de necesidades.

El manejo de la información en la Web, desde un enfoque social y comercial

Para enfocar el tema sobre el manejo de la información en la Web, Natalia Volkow¹ sugiere visualizarla en tres dimensiones de las que, de acuerdo a las necesidades del presente trabajo, rescataré dos. La primera corresponde al *comercio electrónico*, en la que Internet representa el medio que crea las condiciones de un mercado virtual entre las empresas, o bien, entre los consumidores y las empresas, y la segunda dimensión incumbe a la *Sociedad de la Información* que se promueve como una nueva etapa evolutiva de la sociedad, donde la Web se ha gestado como un medio de acceso a un gran acervo informativo, pero ante cuyo desarrollo se interpone la *brecha digital*.

Comercio electrónico

Sobre el concepto del *comercio electrónico*, Roberto Velasco² comenta que éste puede ser comprendido a partir del uso de las *TIC* entre las empresas, o bien, entre vendedores y compradores, para apoyar la adquisición de bienes y servicios. Se trata de una forma de transacción comercial, en la que las partes involucradas interactúan de manera electrónica, en lugar de hacerlo por medio del trato físico directo como tradicionalmente se hacía (y se sigue haciendo). La Web ofrece un nuevo mercado virtual en el que los productores, los proveedores de bienes y servicios y, el individuo común, mantienen un contacto permanente a través de la oferta y la demanda informativa con fines comerciales y de consumo.

Por otra parte, el desarrollo del *comercio electrónico* ha permitido fortalecer la manera de hacer negocios, ya que han sido numerosas

¹ Volkow, Natalia, "La brecha digital, un concepto social con cuatro dimensiones", p. 2.

² Velasco, Roberto, *La economía digital: del mito a la realidad*, p. 12.

las empresas que han incorporado sus productos y servicios en la Internet propiciando, por una parte, la relación de negocios para o entre las empresas, y, por otro lado, ha generado entre sus usuarios y a través de diversas estrategias publicitarias, necesidades de uso y consumo donde antes no existían. Asimismo, *el comercio electrónico* también ha influido para que la infraestructura tecnológica y los servicios, que dicho comercio mantiene, hayan logrado un crecimiento importante.

A pesar de que el grado de penetración social de las nuevas tecnologías no ha sido homogéneo, los cambios generados por éstas son cada vez más evidentes, pues cada día es más común que las personas tengan en sus hogares una computadora para apoyar las actividades cotidianas y que las empresas adquieran más computadoras para realizar sus operaciones. Luego es cierto que la tecnología ha permitido facilitar ciertas cosas en la cotidianidad, pero también es verdad que, entre otras razones, el uso de la tecnología y sus servicios se han favorecido debido a las prácticas publicitarias.

La publicidad no sólo ha sido un factor clave para promover la distribución, el manejo y el uso de la información, también ha transformado el modo de ver a sus usuarios, porque los ha convertido en potenciales clientes y consumidores de los servicios y productos que promueve dicha publicidad por medio de la Internet. Desde la perspectiva del *comercio electrónico*, el usuario es un cliente y, como tal, atiende a sus necesidades sólo en función de los intereses comerciales de las propias empresas, y promueve el intercambio de un tipo de información utilitaria y efímera (porque es fácilmente desechable), lo cual resulta hasta cierto punto deshumanizante.

Estas nuevas prácticas de comercialización han invadido la Web y han fomentado el uso masivo de la información, generando nuevos y diversos problemas sobre el uso y manejo de la misma, tales como la cuestionada validez legal de las transacciones y contratos sin papel; la falta de protección para los consumidores en cuanto a la publicidad engañosa o no deseada; el envío de propaganda en forma indiscriminada a través de los correos electrónicos; la falta de confidencialidad en los datos personales que los usuarios suministran cuando realizan alguna compra, el fomento de la piratería, etcétera.

Sociedad de la Información

Sobre la *Sociedad de la Información*, es posible comentar que, en general, es considerada como la sucesora de la *Sociedad Industrial*. Uno de los primeros escritores que utilizó este concepto fue Fritz Machlup en su libro *The Production and Distribution of the Knowledge in the United States*, publicado en 1962. Pero, a pesar de la antigüedad de dicho concepto y de acuerdo a la literatura sobre el tema, podemos señalar que esta idea comenzó a florecer durante las primeras décadas de la segunda mitad del siglo XX.

Básicamente, los cambios referidos por la *Sociedad de la Información* aluden a las nuevas formas de generar la riqueza, las cuales se fueron trasladando del sector industrial al empresarial, porque, aunque durante la segunda mitad del siglo XX los procesos de industrialización fabril marcaron la pauta para el desarrollo de las sociedades, a principios del siglo XXI, este desarrollo ya se había afianzado en torno a un sector de servicios, basado en el manejo y disposición de la información entre las empresas y sus empleados. Dentro de este nuevo entorno, las *TIC* jugaron un papel crucial, convirtiéndose en el nuevo motor del desarrollo y del progreso social.

Incluso todavía en la *Declaración de los Principios de la Cumbre de la Sociedad de la Información*, llevado a cabo en Ginebra, Suiza, en el 2003,³ se reconoce la influencia económica-social de las TIC, cuando se indica que, éstas pueden considerarse como un instrumento eficaz para acrecentar la productividad, generar el crecimiento económico y fomentar la ocupación laboral. No obstante, para Ignacio Ramonet, el uso de estas tecnologías también ha mantenido otras consecuencias de índole informativa porque, en el nuevo entorno social de la información, las TIC han desencadenado tres características fundamentales:

*Sobreabundancia, frente a la escasez de los siglos pasados.
Extremada rapidez, en contraste con el ritmo lento de otros tiempos y,
Su valor es esencialmente comercial, porque la información es, antes que nada, una mercancía. Y como tal está esencialmente sometida a las leyes del mercado, de la oferta y la demanda, y no a otras reglas sociales como podrían ser las derivadas de criterios más cívicos o éticos.*⁴

Asimismo, Ramonet continúa su argumento sobre las repercusiones de lo que llama “*un sistema de producción superabundante de información*”, refiriéndose a la Web, cuando advierte que:

*Hoy estamos convencidos de que una información de tipo cuantitativo no resuelve los problemas que nos planteamos [...] La información debe tener un aspecto de orden cualitativo [...] que debiera pasar por dos cuestiones: la credibilidad y la fiabilidad.*⁵

³ Declaración de Principios, “*Construir la Sociedad de la Información, un desafío para el nuevo milenio*”. Disponible en Internet desde <http://www.itu.int/wsis/docs/geneva/official/dopes.html>. (Revisado el 6 de enero del 2006).

⁴ Ramonet, Ignacio, *El periodismo del nuevo siglo*. Disponible en Internet desde <http://www.lafactoriaweb.com/default-2.htm>. (Revisado el 6 de enero del 2006).

⁵ Ramonet, Ignacio, *op. cit.*

De lo anterior, es posible suponer que la abundancia informativa, la falta de confiabilidad y de credibilidad de la misma, son problemas fundamentales derivados de los recientes flujos de la información en la Web.

Bajo estas circunstancias, algunos escritores han comenzado a alejarse de las engañosas propuestas de la *Sociedad de la Información*, dando cuenta de los falsos universalismos en los que se ha pretendido circunscribir y alejándose del aparente optimismo que mantienen los patrones de desigualdad social. Por ejemplo, Herbert Schiller⁶ ha criticado el control y el poder de las corporaciones multinacionales sobre la distribución de las nuevas tecnologías y Noam Chomsky⁷ ha destacado la influencia negativa y el dominio político que, a partir de la propaganda, realizan algunos medios de comunicación. Pero una de las críticas más intensas versa sobre la denominada *brecha digital*, que según la *Organización para la Cooperación y el Desarrollo Económico (OCDE)*⁸ se define como la capacidad (o la incapacidad) de tener acceso a las TIC, a la Internet y al conocimiento para poder operar y utilizar esas tecnologías.

El término *brecha digital* se originó a finales de la década de los setentas y principios de los ochentas. Aunque no fue sino hasta la década de los noventas cuando fue reconocido por Estados Unidos para denotar el riesgo de marginar a los sectores sociales menos favorecidos en cuanto al acceso a la información en Internet. A partir de este planteamiento, surgió la inquietud de medir la *brecha digital* en términos de la relación porcentual de la población total que sí tiene acceso a la Internet y se generaron estadísticas en las que se

⁶ Schiller, Herbert, *Información y economía en tiempos de crisis*, 143 pp.

⁷ Chomsky, Noam, *Secretos, mentiras y democracia*, pp. 56-63.

⁸ Volkow, Natalia, *op. cit.* p. 1.

estableció una clara relación entre el acceso a la Internet y las características sociales, demográficas y económicas de la población, evidenciando la falta de oportunidad informativa que mantienen los grupos de mayor marginación socioeconómica. Asimismo, dejó al descubierto otro tipo de dificultades como las relacionadas con la lengua, pues la mayor parte de la información disponible en la Web se presenta en inglés con una notoria influencia de las instituciones públicas, privadas y de organizaciones no gubernamentales de Estados Unidos.

Debido a lo anterior, es posible exponer que, aunque la nueva tecnología transformó la manera de hacer negocios, ofertar productos, y otorgar servicios educativos, empresariales y de entretenimiento, por citar tan sólo algunos ejemplos, para promoverse como uno de los más sólidos instrumentos de la modernidad, el rechazo de algunos teóricos (como los mencionados anteriormente) puso al descubierto la otra cara de la moneda, al señalar las consecuencias negativas generadas por el uso de las nuevas tecnologías y el manejo tendencioso sobre los recientes flujos de información.

Incluso, coinciden en señalar que nuestra sociedad nos encamina hacia un mundo socialmente fragmentado y dividido en cuanto al uso y disposición de la información. La tesis de que la abundancia se ofrece a todos, sin jerarquías, para producir una sociedad informativa libre y solidaria, es falsa, pues a pesar de sus bondades, la fragmentación social de la información es una consecuencia negativa de las nuevas tecnologías, ya que en la Internet existe una clara separación entre quienes tienen la posibilidad de pagar por la disponibilidad de la información y los que no pueden hacerlo.

Influencia de la Web en la percepción informativa del individuo

Hay mucho escrito acerca de la influencia social de la información. Incluso, cuando se dice que estamos en la *era de la información*, es precisamente porque ésta se ha convertido en el símbolo social de nuestra época. Pero, justamente, el sentido de ese símbolo es el que forja una gran controversia (en cuanto a la fragmentación social y al uso comercial del que ya se hablaba anteriormente), porque alimenta, en el imaginario colectivo, la idea de que, a partir del uso de la información, es posible alcanzar un bienestar social común, el cual todavía no representa (y parece impensable que llegue a representar) una realidad para todos.

Uno de los hitos creados en esta *era de la información* es la Internet, ya que su uso ha permitido sacar provecho de ella con fines comerciales, políticos, sociales, etcétera, mientras que, por otro lado, el manejo que se ha establecido en torno al uso de la información también ha trastocado la identidad social del individuo. A ello ha contribuido el hecho de que los nuevos flujos de información se manifiestan a partir de una estructura propia, caracterizada por la velocidad, el acortamiento de distancias, el traspaso de fronteras y la abrupta visualización de diversas imágenes en un instante. Sobre esta última idea conviene aclarar que, así como en la televisión es posible pasar de un canal a otro para ver una novela, un noticiero o un evento deportivo, en la Web, también es factible abrir, simultáneamente, una página sobre turismo, otra de espectáculos, de modas, etcétera y que dentro de este juego de representaciones dispersas, el *usuario-espectador* de la Web se deja deslumbrar fácilmente por el ritmo que impone la tecnología y la fascinación que provoca lo virtual.

La influencia de la tecnología en el individuo suscita que la vida se torne aparente, rápida y emotiva. Así, los referentes individuales comienzan a disolverse para luego construirse y reorganizarse en función de los nuevos códigos simbólicos, provenientes de repertorios culturales, sociales y comerciales que, en la mayoría de los casos, son en mayor o menor medida ajenos a sus costumbres o se encuentran fuera de sus alcances, pero que son promovidos por los sistemas publicitarios y por el deseo del propio individuo de no quedar ajeno a la modernidad.

En este nuevo entorno, según Anthony Giddens,⁹ la ausencia de las nociones del tiempo y del espacio, también vulneran las relaciones sociales de los individuos en sus contextos locales de interacción. Es decir, aunque la Web con su potencial para transmitir información desde cualquier lugar en tiempo real, hace que cualquier persona con el ratón de una computadora en la mano pueda incursionar en un mundo de creencias, costumbres, valores, mentalidades o modas, de las regiones más distantes del mundo, también propicia que los espacios tradicionalmente compartidos como las plazas, parques, cafés, etcétera, ya no sean los únicos que conformen su identidad, sino una pluralidad de símbolos desarraigados que se incorporan en su cotidianidad y que circulan por el ciberespacio.

Asimismo, el uso de la Web ha provocado que el usuario la vea como una herramienta informativa, desplazando en muchas ocasiones a otras fuentes tradicionales impresas. Por ejemplo, tiempo atrás, el libro constituía el principal soporte del conocimiento y, para tener acceso a ese conocimiento, la lectura del libro se tornaba obligatoria. Sin embargo, ante el auge de las nuevas formas en que circula la información, la práctica de la lectura sobre el libro se ha ido

⁹ Giddens, Anthony, *Modernidad e identidad del yo*, p. 110.

desfavoreciendo sustancialmente, o tal vez sea necesario reconocer que este tipo de lectura ya sufría, de tiempo atrás, un deterioro social, pero que con el uso de las nuevas tecnologías, se agudizó y se hizo más evidente.

Por otro lado, el hecho de que la Web se haya convertido en una herramienta de consulta masiva, no obedece sólo a la funcionalidad que le ha otorgado la tecnología, sino a las necesidades establecidas de acuerdo al entorno en que se desenvuelve el individuo en su cotidianidad y al nivel cognitivo del mismo, pues, según Luis Villoro:

La gran mayoría de los conocimientos que utilizamos para orientar nuestra vida no alcanzan el nivel de un conocimiento científico o sabiduría. Pocos de nuestros pretendidos saberes tienen el rigor y la claridad de los enunciados científicos; pues en realidad suelen ser creencias más o menos razonables, basadas en razones más o menos controvertibles [...] en raras ocasiones nuestros conocimientos personales alcanzan la comprensión profunda de una auténtica sabiduría.¹⁰

Debido a ello, la gran cantidad de información que, en lo general, fluye por la Web, permite atender las necesidades cotidianas del individuo, porque en ella reconoce cierto tipo de información que le es útil para saber cómo determinar o tomar decisiones, y, así, resolver aspectos de su vida diaria. Asimismo, cabe mencionar que este tipo de información no siempre se caracteriza por ser fiable o creíble (como lo advierte Ramonet) para cubrir, por ejemplo, los requerimientos de un conocimiento científico o académico. No obstante, en la Web hay recursos confiables, como los catálogos o las bases de datos especializadas que cubren las necesidades de los perfiles académicos o científicos. Aunque la disponibilidad de esas fuentes es regularmente de acceso restringido, también existe la

¹⁰ Villoro, Luis, *Creer, saber, conocer*, p. 234.

posibilidad del acceso abierto para poder disponer de este tipo de recursos.

No obstante, las necesidades para conocer entre las comunidades científicas o académicas, en lo general, y los propios investigadores, en lo particular, se establecen con un claro distanciamiento frente a los requerimientos ordinarios del saber en el individuo común. Para orientar más ampliamente estos conceptos, Villoro comenta:

Saber no implica tener que examinar todas las alternativas que pudieran contravenir las razones con las que contamos, pero sí las necesarias para inferir que no hay otras. Sin embargo (en la vida diaria) el saber se usa con mayor laxitud, pues se aplica a situaciones en las que se concluye la objetividad de las razones, a partir de datos escasos y sin considerar las alternativas pertinentes. La percepción apresurada del entorno, las palabras de un amigo, la lectura de un aviso, etcétera, se toman como garantía de objetividad, sin comprobar su veracidad.¹¹

Mientras que, por otro lado, continúa diciendo:

[...] cuando decimos que deseamos conocer, en realidad queremos decir que (independientemente de los objetivos que nos proponamos alcanzar), tenemos la necesidad de saber cuáles son los medios para llegar al conocimiento. Aunque los intereses para saber sean de lo más variado, debe haber en todos una característica general por la que se derivan nuestros intereses en conocer y no sólo en creer algo. Debe haber un interés en encontrar razones objetivas, en dar con argumentos válidos, en rechazar razones insuficientes, en comprobar nuestros asertos (...) en fin, en continuar sin cejar nuestro cuestionamiento.¹²

Entonces, es posible considerar que si bien las prácticas del saber se encuentran evidentemente diferenciadas de los intereses por conocer, también es cierto que la distribución de la información de

¹¹ *Ibidem*, p. 166.

¹² *Id.*, pp. 259-260.

calidad, para atender este último tipo de necesidades, se halla claramente limitada hacia las masas, debido a las restricciones de la Web para tener acceso a la información académica o científica.

Por otra parte, es imperativo apuntar que, en la práctica, el exceso y rapidez con la que fluye la información genera una tendencia que consiste en creer que ciertos hechos o ideas adquieren legitimidad sólo al ser exhibidas en los medios masivos de comunicación, tal como sucedió con los acontecimientos del 11 de septiembre. Confunde al individuo quien llega a interponer o validar la creencia y/o el simple saber como conocimiento. Esta circunstancia, en la que el individuo tiende a subsumir el pensamiento crítico ante la vorágine informativa de la Web, puede llevarnos a considerar que existe una falta de conciencia sobre la importancia de entender el conocimiento y la esencia de interpretar cualquier información o dato. Debido a ello, se hace necesario distinguir entre estos tres elementos básicos, vinculados con la necesidad del saber y el conocer, y con el acceso mismo de la información en la Web.

Ahora bien, es posible reconocer un dato como información cuando pueden ser interpretado (según el entorno y el nivel cognitivo del usuario) como números, imágenes o signos. Por ejemplo, un símbolo químico puede no indicar nada para el individuo común, pero para un especialista o estudioso de la química, o un farmacéutico ese símbolo cobra un sentido. Por otro lado, a la información se le puede distinguir como dato, en un contexto determinado, tal puede ser el caso de una nota periodística. Es decir, la información vertida en una nota periodística puede dar cuenta de un aspecto de la realidad, a través de la descripción de un hecho, y éste puede ser considerado como un dato para interpretar o crearse un juicio acerca de la

realidad. En consecuencia es posible afirmar que la información no es neutra y está sujeta a interpretación.

Por su parte, el conocimiento está basado en la reflexión y en el procesamiento mental de la información. El conocimiento oscila entre el pensamiento conceptual y el creativo. Puede ser implícito, explícito, práctico o abstracto. Está individualmente almacenado en la mente de cada persona y, al mismo tiempo, es compartido con otras personas; por ello, puede ser creado y recreado a través de la interacción social. El conocimiento es consecuencia o resultado de un proceso de investigación activo, que permite interpretar la realidad con ciertos fines específicos, dentro de un contexto cultural, científico o social.

Sin embargo, la separación entre conocimiento e información en ocasiones puede ser tan sutil, que es difícil de discernir, porque se establece en una dualidad de planos que incluyen un nivel abstracto y un nivel tangible. Al respecto, Páez Urdaneta¹³ explica que la información se manifiesta como entidad subjetiva cuando se genera en la mente de las personas y, como entidad objetiva cuando se plasma sobre algún soporte o se transmite mediante algún proceso de comunicación. Bajo estas consideraciones, Páez Urdaneta continúa señalando que la información además transita o se ventila a través de dos momentos importantes: el primero cuando la mente humana la asimila, procesa e interpreta, para transformarla en conocimiento, y, el segundo, cuando la registra en forma documental.

La interpretación y la asimilación que se haga de la información no puede llevar nunca a una verdad inmutable, pues el conocimiento no es fijo, ni absoluto, es en realidad un proceso dinámico. De este

¹³ Páez Urdaneta, Iraset, *Gestión de la inteligencia, aprendizaje tecnológico y modernización del trabajo. Retos y oportunidades*, pp. 12-16.

modo, el acceso a la información sólo representa una fase preliminar para alcanzar el conocimiento, mientras que la forma en que la información se integra en el individuo constituye tan sólo una fase, a partir de la cual se genera el conocimiento.

Con base en lo expuesto, es posible apreciar que, en el contexto del acceso a la información de la Web, existen circunstancias sociales e individuales, a partir de las cuales se establecen las tendencias que giran en torno al diseño y construcción de los esquemas de organización y recuperación de la información digital. En otras palabras, es factible reconocer que existe una intencionalidad de uso, promovida por percepciones ideológicas, comerciales y del desarrollo mismo de la tecnología, a partir de las cuales, se definen los esquemas de organización y recuperación de la información.

Desde el enfoque social, por ejemplo, el comercio electrónico y las prácticas publicitarias, para fomentar el pago por el uso de los servicios y productos informativos, han convertido al usuario en un cliente y a la información en una mercancía. Lo anterior refuerza la idea, promovida por la llamada Sociedad de la Información, de alcanzar un bienestar económico-social común a partir del uso de la tecnología y la información. Estos aspectos, en su conjunto, han contribuido a fomentar que a la información se le atribuya un valor económico, de utilidad inmediata y efímera. Pues, desde el punto de vista comercial, parece no existir un genuino interés para abastecer de información a los usuarios con el fin de atender sus necesidades formativas, culturales o de investigación, sino de encontrar un cauce entre el producto o el servicio informativo y el cliente.

La fragmentación social que se deslinda de la falta de acceso a la tecnología y a la información, es otra parte de la realidad que se

define a través de la brecha digital. Entonces, es factible hacer notar que, entre la información y el individuo, el acceso a la información no sólo está condicionado por su conocimiento y manejo de la tecnología, sino por la capacidad económica del individuo y de la sociedad.

Desde la perspectiva del individuo, la utilidad inmediata atribuida a la información también propicia que la valide, sin hacer más investigación sobre la veracidad de la misma. De tal forma que la funcionalidad atribuida a la Web, como herramienta de distribución de la información, propicia el hábito de la simple consulta y desfavorece el hábito de la investigación a través de la lectura. Sin embargo, no por ello se pretende afirmar que la información que navega por la Web deje de ser funcional, pues coexisten niveles de información subordinados a distintas necesidades informativas. Dentro del trabajo realizado, es factible ubicar, por lo menos, dos: un nivel funcional inmediato y efímero de la información, que permite resolver aspectos de la cotidianidad en el individuo y un nivel funcional más perecedero, valorado por su apoyo al trabajo académico y de investigación.

De lo anterior es posible suponer que, en la Web, existen sistemas de recuperación basados en la indización y generación de taxonomías que fomentan el uso indiscriminado y utilitario de la información, mientras que, por otro lado, también se hallan esquemas de organización y localización basados en sistemas documentales, como los catálogos de las bibliotecas. No obstante, los sistemas comerciales tienen más presencia que los sistemas documentales y, debido a ello, muestran un mayor nivel de impulso que estos últimos, ya que incluyen complejos algoritmos para rastrear información a velocidad más alta y con capacidad grande de almacenamiento. En esos esquemas, no existe una preocupación latente para seleccionar,

preservar o congregar la información en unidades específicas de información, en otras palabras, no se da, en la Web, un impulso para contribuir al desarrollo o proyección de los sistemas documentales, ya que ello entorpecería el ritmo acelerado en que se pretende fijar y mantener el flujo de la información.

Llegados a este punto, vale la pena reconocer y rescatar la responsabilidad social y humana de la biblioteca y la propuesta funcional de los sistemas de clasificación documentales, en respuesta a la actividad formativa, cultural y de investigación entre los propios usuarios, ya que, para este tipo de población, las bases de datos especializadas y los catálogos de las bibliotecas siguen siendo herramientas fundamentales para satisfacer sus necesidades informativas. Por otro lado, si consideramos que la información es ante todo un *producto social* (ya que la construimos todos), entonces, como tal, la labor de organizarla y brindarle una forma de acceso para su difusión y recuperación, deberá seguir representando parte de la responsabilidad social del bibliotecólogo.

CAPÍTULO 2

HERRAMIENTAS DE ORGANIZACIÓN, REPRESENTACIÓN Y ACCESO A LA INFORMACIÓN EN LA WEB

La biblioteca a través de los procesos de organización y disposición de la información en la Web

Durante la segunda mitad del siglo pasado, la biblioteca se ajustó al avance de la tecnología y modificó su estructura para facilitar la administración de sus recursos y servicios, aunque ello no repercutió en la definición temática de sus colecciones, ni en los sistemas de organización adoptados para sus acervos. Sí modificó, en cambio, el enfoque tradicional de la biblioteca, pues a partir de sus nuevas características, surgieron otros conceptos para tratar de definirla, tales como *biblioteca electrónica, virtual y digital*; en algunos casos, porque se buscaba establecer cierta complementariedad entre las colecciones impresas y digitales y, en otros, porque se pretendía constituir como una nueva entidad informativa. Dentro de este proceso evolutivo, la biblioteca digital se estableció como el último peldaño tecnológico escalado dentro de la Web.

Actualmente, miles de bibliotecas de todo el mundo se han incorporado a la Web manteniendo el desarrollo de sus recursos documentales en apego a ciertos principios tradicionales de selección, control y organización. Sin embargo, la formación de dichas colecciones se ha visto influenciada por la tendencia que existe en publicar o reeditar digitalmente los recursos documentales. Ello ha provocado que dichas colecciones no se formen exclusivamente por documentos impresos, ya que ahora también incluyen recursos digitales que se ubican en la propia biblioteca o se consultan a través de algún acceso remoto.

La distribución de los recursos digitales regularmente se realiza a través de medios comerciales o gratuitos. Entre los medios comerciales, se difunde todo tipo de documentos, pero debido a su demanda cobran especial importancia las bases de datos especializadas y las revistas, mientras que los recursos gratuitos generalmente incluyen documentos de dominio público. Del grado de interés y calidad de estos recursos, depende que la biblioteca les preste atención y decida seleccionar aquellos que juzgue más importantes o útiles para sus usuarios.

Los modelos que las bibliotecas han generado para dar acceso a los recursos seleccionados en o desde la Web, se pueden tipificar de la siguiente manera¹:

- *Guías*: son listas ordenadas alfabéticamente, que permiten el acceso secuencial a la información por título, tema o por tipo de documento. En estos casos, los criterios de selección no son siempre explícitos, la descripción es muy elemental y no explota ningún mecanismo de recuperación en particular.
- *Base de datos*: el registro documental se incluye en una base de datos separada del catálogo y su acceso se gestiona a través del *software* que facilita su recuperación en la Web.
- *Recursos integrados al catálogo*: esta opción exige el mismo tratamiento detallado y normalizado que se realiza en todos los documentos de la colección. Diversas bibliotecas (particularmente en Estados Unidos) utilizan las etiquetas *MARC* como registro básico y único con el cual se alimenta el catálogo y las bases de datos.

¹ Estivill, Assumpció y Ernets Abadal, "Acceso a los recursos Web gratuitos desde las bibliotecas", p. 6.

Descartado el primero de los tres modelos anteriores, ya que se trata de una simple lista alfabética, cabe señalar que tanto los registros de las bases de datos, como los recursos de la Web, se integran al catálogo considerando ciertos criterios de selección y mantenimiento de las direcciones digitales, aunque la descripción de los recursos, el análisis del contenido y el sistema de recuperación adoptado difieren, en mayor o menor medida, en todos ellos según el modelo de sistematización adoptado.

Subject Gateways, *Information Gateways* y *Resource Discovery Systems*² son algunos de los términos que se utilizan comúnmente para designar a las bases de datos, que contienen descripciones de diversos recursos en la Web. El proyecto *Desire*³ define estas herramientas como pasarelas o portales temáticos, y algunos de éstos se han constituido gracias al esfuerzo cooperativo e interinstitucional de algunos organismos que, además, se han dedicado a generar políticas bien definidas para el crecimiento, actualización y mantenimiento de dichos recursos.

Sobre la presencia de los portales temáticos en la Web, John Kirriemuir⁴ hace referencia a 141 bases de datos de este tipo. La mayoría de éstas han sido creadas en el Reino Unido y en Estados Unidos. En general, los portales británicos se han diseñado y actualizado por bibliotecas universitarias. De ellos, 70 proyectos han sido financiados por la *Electronic Libraries (eLib)* desde 1995. Asimismo, en el Reino Unido, la *Resource Discovery Network (RDN)* también se ha encargado de agrupar diversas bases temáticas de

² La terminología indicada para este tipo de bases de datos es muy variada, aunque en español puede aparecer como *bases de datos de recursos Web*. Estas fuentes se encuentran dirigidas a comunidades académicas y científicas y se elaboraron con criterios selectivos de calidad a cargo de especialistas.

³ El proyecto *DESIRE* es un proyecto de la *Unión Europea*, creado para el desarrollo de portales temáticos. Ha permitido generar un marco de referencia común para respaldar diferentes servicios de información europeos.

⁴ Kirriemuir, Jhon, Susan Welsh y Martin J. Knigh "Cross-searching subjects gateways: the query routing and forward knowledge approach". Disponible en Internet desde <http://www.dlib.org/dlib/january98/01kirriemuir.html>.

datos, que han sido creadas por bibliotecas académicas y de investigación con el objeto de facilitar su consulta simultánea.

La *RDN* permite el acceso a diferentes bases temáticas de datos, tales como *Biome (salud y ciencias de la vida)*, *EMC (ingeniería, matemática e informática)*, *Humbul (humanidades)*, *PSIgate (física, química, ciencias de la tierra, historia de la ciencia, medio ambiente, etcétera)* y *Sosig (ciencias sociales, derecho y economía)*, por citar tan sólo algunos ejemplos. En el ámbito europeo, también destaca *Renardus*, creado en el 2000 y financiado por el programa de *Tecnologías para la Sociedad de la Información (IST)* de la *Unión Europea*. *Renardus* busca mejorar el acceso a los recursos culturales y científicos de la Web, a través de un servicio que permita realizar búsquedas temáticas como si se tratara de un catálogo colectivo.

Por otra parte, entre las bases de datos de Estados Unidos, destacan: *Agnic*, de la *National Agricultural Library*; *Medlineplus*, de la *National Library of Medicine* y la *Isaac Network*, financiada por la *National Science Foundation*. Esta última tiene el objetivo de investigar cómo enlazar diversas colecciones de *metadatos*, para formar una sola colección virtual. Debido a su cobertura, el proyecto *LANIC*⁵, *Latin American Network Information Center*, requiere una mención especial, pues se encuentra a cargo de la *Universidad de Texas* y tiene la misión de facilitar el acceso informativo en la Web para, desde y sobre América Latina. Este recurso fue implantado en 1992, pero comenzó a operar de manera continua desde 1994 hasta nuestros días.

Considerando lo expuesto anteriormente, es factible comentar que las bases temáticas de datos o bases de datos especializadas se han convertido en una de las herramientas más

⁵ [Http://lanic.utexas.edu](http://lanic.utexas.edu)

populares de la Web para el almacenamiento y disposición de la información, porque operan bajo ciertos principios normativos, selectivos y estructurales, que les permite mantener un alto nivel de calidad. No obstante, y debido a ciertas condiciones restrictivas para consultar dichos recursos, podría decirse también que se han convertido en la parte invisible de la Web.

La *Internet invisible* o *Intranet* se define como el conjunto de recursos disponibles únicamente a través de algún tipo de pasarela o formulario Web, que requiere el uso de una clave de acceso para poder consultarla. Al respecto, Isidro Aguillo⁶ comenta que existen sitios excluidos en forma deliberada, porque las páginas están protegidas por contraseñas o contienen un campo "*noindex*", que impide a un buscador o motor de búsqueda, seguir avanzando en la Web para encontrar más resultados. En otros casos, esta situación obedece a que las páginas incluyen formularios en los que es necesario registrarse y pagar para tener acceso al contenido, o también porque se encuentran conformadas por páginas que no pueden ser indizadas, debido a las limitaciones técnicas de los propios *buscadores*.

A causa de estas limitaciones, existen recursos digitales que no pueden ser indizados por los buscadores tradicionales de la Web. En consecuencia, la gran mayoría de las bases de datos, con las características mencionadas, se han convertido en verdaderas islas de calidad, ya que su invisibilidad para los directorios y buscadores, implica una seria dificultad para localizar y recuperar la información de los recursos documentales que albergan.

Para ejemplificar esta situación, se describen sucintamente los casos siguientes:

⁶ Aguillo, Isidro F., "Internet invisible o Infranet: definición, clasificación y evaluación", pp. 249-270.

- *Bases de datos documentales*

Éstas se encuentran disponibles regularmente como catálogos (OPAC) de bibliotecas. Y tal vez éste sea el caso más común debido a la gran cantidad de sitios de bibliotecas presentes en la Web. Aunque también existen bases de datos documentales creadas con fondos privados y que han hallado en la Web un magnífico sistema para su explotación, por ejemplo, *Medline* (Medicina), *AGRIS* (Agricultura), *ERIC* (Educación), etcétera.

- *Revistas digitales*

Se trata de bases de datos que incluyen el texto completo de las revistas y que sólo se pueden recuperar, con una previa identificación del registro, a través de un formulario de consulta y una clave de acceso. Originalmente, esta información no era “invisible”, pues la recuperación se hacía a través de índices o listados de referencia, mas el crecimiento de este tipo de recursos, como consecuencia de las necesidades de la comunicación científica, motivó que llegara a superar los millares de registros, razón por la cual dicho método de consulta quedó obsoleto.

Por lo pronto, el número de revistas disponibles en la Web es considerable y cabe esperar que, en los próximos años, la gran mayoría de los títulos impresos sobre papel, tengan una versión digital, además de los títulos que ya aparecen digitalizados. No obstante, cabe hacer notar la tendencia por la edición no formal, desde la publicación de *pre-prints*, hasta la creación de depósitos independientes en la Web que conviven con el formato de las revistas o las rechazan. Esto obedece a que, en el ámbito científico, la batalla de las revistas parece estar basada en torno a las políticas de comercialización establecidas por las editoriales.

El reconocimiento que se ha hecho de los contenidos de las bases de datos ha permitido el impulso de un gran esfuerzo para recuperar la información que existe detrás de la *Internet invisible*.

Poco a poco, esta labor se ha encauzado, gracias a la posibilidad de algunos directorios y buscadores que, en algunos casos, han logrado identificar y organizar los principales componentes de aquellos recursos de acceso limitado.

Estructura y operatividad de las bases de datos documentales

Una base de datos es un conjunto de información estructurada en registros y almacenada en un soporte electrónico. Cada registro constituye una unidad autónoma de información que puede estar señalada en diferentes campos o tipos de datos que se capturan en dicha base de datos. Las *Bases de Datos Documentales (BDD)* agrupan registros documentales en forma impresa, audiovisual, gráfica, sonora, o bien en formato digital. Dicha base de datos se crea y mantiene con el objetivo de resolver necesidades informativas para un colectivo (empresarial, institucional, académico, científico o de cualquier otro tipo). Es por ello que representa uno de los recursos más socorridos por las bibliotecas.

Debido a sus características, es posible tipificarlas en:

- *Bases de datos de texto completo*

Permiten un enlace directo al documento original, ya sea un documento iconográfico (fotografía, imagen de televisión, etcétera), o un documento textual digitalizado. La operación de búsqueda y consulta del documento se realiza sin salir del propio sistema de información. En este tipo de base de datos, normalmente la búsqueda se limita a los campos de referencia bibliográfica, aunque ocasionalmente también permite recuperar palabras del texto completo o del resumen.

- *Bases de datos referenciales*

Sólo incluye referencias al documento original, por tanto, para obtener dicho documento es necesario emplear otro servicio de recuperación, aunque algunas bases de datos ofrecen enlaces directos para obtener el original, cuando registran una fuente disponible en la Web.

En cuanto a las opciones de búsqueda, los sistemas de recuperación comúnmente las presentan a través de formularios que permiten realizar:

- *Búsquedas directas*

Las búsquedas directas son aquellas en las que se capturan directamente una o varias palabras sobre el espacio reservado para ello. Pueden distinguirse, a su vez, en dos tipologías. La primera es la *Búsqueda libre*, que se realiza sin tener en cuenta algún campo en particular de la estructura de la base de datos. Generalmente, el sistema busca el término o los términos solicitados en la totalidad de los campos que ofrece la propia base de datos. La segunda es la *Búsqueda por campos* en la cual, como su nombre lo indica, se hacen consultas sobre un campo o ciertos campos seleccionados previamente por el usuario, tales como: autor, serie, clasificación, descriptores, etcétera.

- *Búsquedas en índices*

En este tipo de búsquedas, el usuario visualiza un diccionario o índice alfabético de todas las entradas posibles a los campos, o selecciona un tipo de índice en concreto (índice de autores, descriptores, etcétera), de acuerdo a sus intereses. Dentro de esta modalidad, es posible distinguir dos opciones: *Índices por palabras* que incluyen *unitérminos*, es decir, conceptos expresados en una sola palabra (que se elaboran, por ejemplo, con todas las palabras extraídas del título o del resumen de los recursos documentales), los *Índices por frases* que tienen entradas

compuestas por varias palabras, por ejemplo, apellidos y nombre del autor, descriptores o encabezamientos de materia compuestos, que permiten distinguir el área geográfica, periodo, o algún otro dato relacionado o subordinado al tema principal. Esta clase de búsquedas también muestra correspondencia con la modalidad de búsqueda jerarquizada.

- *Búsquedas jerarquizadas*

Esta opción se realiza a partir de un concepto genérico, a través del cual es posible localizar no sólo los registros en los que aparece dicho término, sino en todos aquellos en los que figure algún concepto más específico de su campo semántico. Esta opción de búsqueda sólo es posible realizarla en aquellos sistemas de información que cuentan con lenguajes estructurados de clasificación o descriptores. Para precisar mejor la funcionalidad de dichos lenguajes, a continuación se realiza una breve descripción de ellos:

- *Encabezamientos y sub-encabezamientos de materia*

Es el sistema utilizado en los catálogos de las bibliotecas para describir los temas abordados en un documento. Se trata de términos procedentes de un lenguaje controlado establecido previamente (como *listas de encabezamientos de materia*). Cada entrada al índice se compone de una materia principal (*encabezamiento*), que puede estar acompañada, de facetas de materia como lugar, tiempo o forma (*sub- encabezamientos*) y que tienen la función de precisar o complementar la entrada principal.

- *Palabras clave*

Son términos simples o compuestos extraídos del texto original de los documentos, sin que estén presentes en algún lenguaje controlado. En un sistema de indización por palabras clave, pueden utilizarse tantas formas de expresar el mismo

concepto, como las utilizadas por los propios autores de los documentos originales (lenguaje natural).

➤ *Descriptores*

Son términos simples o compuestos extraídos de un lenguaje controlado que representan los conceptos utilizados por los autores, para evitar la ambigüedad del lenguaje natural. Regularmente dichos términos se obtienen de *tesauros*.

Regularmente, los sistemas de recuperación de las *BDD* cuentan con diferentes herramientas para construir una estrategia de búsqueda, tales como:

Operadores lógicos booleanos y truncado

Los *operadores booleanos* permiten combinar diferentes términos en una misma búsqueda. Esto se logra a través de las operaciones propias de la teoría de conjuntos o álgebra de *Boole*, que en general aplican la suma o unión (A o B), la resta o negación (A no B), e Intersección (A y B). El *truncado*, por su parte, consiste en la localización de términos que comparten la misma raíz, a través de un signo de truncado (*, o ?, o \$, que varía según el programa).

Para realizar una búsqueda en una *BDD*, es preciso tener en cuenta la polisemia, sinonimia y homonimia de los términos empleados, así como saber cuál es la estructura del sistema y utilizar estratégicamente los términos para rastrear la información requerida.

De acuerdo al tipo de contenido, formato y servicio que ofrecen las bases *BDD*, es posible distinguirlas y ejemplificarlas de modo general como sigue:

- Bases de datos *especializadas* o *multidisciplinarias*. Entre las especializadas es posible citar a *Eric (Educación)*, *Biosis (Ciencias de la vida)*, *Medline (Medicina)*, *Lisa (Ciencias de la información)*, *Psycoinfo (Psicología)*, etcétera. Entre las multidisciplinarias se encuentran: *Article@INIST (revistas y monografías)*, *Pascal (Ciencia y Tecnología)*, *Scisearch (Ciencia y Tecnología)*, *Inspec (Física, Ingeniería e Informática)*, etcétera.
- Bases de datos que se clasifican por algún *tipo de formato* en particular, tales como libros, revistas, tesis, literatura gris, etcétera. Por ejemplo: *Sigle (Literatura gris)*, *Dissertation Abstracts (Tesis)*, *Librunam (Libros)*, etcétera.
- Bases de datos que se distinguen por el *tipo de servicio* que otorgan, como las que acompañan una ficha bibliográfica con resumen como *Chemical Abstracts*, *FSTA (Food Science and Technology Abstracts)*, *Historical Abstracts*, *Sociological Abstracts*, entre otras.

En general, las *BDD* se incorporaron a la Web como parte de una evolución natural de los catálogos impresos utilizados tradicionalmente en las bibliotecas y que, como tales, consideran diversos principios de organización y control, que ofrecen las ventajas siguientes:

- Un adecuado desarrollo de las colecciones que integran y almacenan, a partir de ciertas políticas específicamente definidas para constituir una colección funcional en atención a una comunidad o un tipo de interés en particular.
- El empleo de sistemas de clasificación documental o la generación de vocabularios controlados, para realizar una descripción y organización más precisa de las colecciones,

permiten contextualizar un término dentro de otro mayor fijando una relación de jerarquía y subordinación temática, lo que abre las perspectivas de orientación para el usuario al momento en que realiza una búsqueda.

- Los sistemas de localización y recuperación de información de las *BDD* permiten identificar registros referenciales y, en ocasiones, recuperar los textos completos de dichos registros, bajo diversas opciones de búsqueda. Aspecto que atiende con cierta irregularidad los directorios y buscadores de la Web.

Finalmente, entre las desventajas de las *BDD*, tal vez la más preeminente se encuentra en el hecho de que regularmente mantienen un acceso restringido en la Web, ya que, para poder consultarlas, es necesario el uso de un *password* o el llenado de un formulario para dar entrada a los registros que albergan.

Estructura y operatividad de los índices y directorios

El crecimiento de la información que fluye por la Web ha fomentado la creación de herramientas que, empleando sistemas de operación parecidos a los de las *BDD*, ofrecen la posibilidad de realizar búsquedas en conjuntos de páginas Web. Incluso, todas esas herramientas recogen información de la Web y la almacenan en una base de datos. Sin embargo, a pesar de que las diferencias entre ellas se hacen cada vez menos evidentes, es posible en general, distinguirlas según su funcionamiento.

Los programas que localizan la información en la Web toman principalmente dos morfologías llamadas *directorios* o *índices* y

*motores de búsqueda o buscadores*⁷. La información que contienen los directorios normalmente es compilada y está organizada manualmente en una estructura jerárquica. Los usuarios recorren esta estructura para identificar y recuperar la información que necesitan. Mientras que los índices usan programas que van recavando la información de la Web y la organizan en una base de datos. Para cada petición informativa, los programas enlazan un documento a partir de dichas bases de datos.

Traugott Koch⁸ realiza una analogía entre la localización de la información dentro de un libro, y dentro de la Web, con el fin de fijar la diferencia entre los términos *Browsing Service* (directorios o índices) y *Search Service* (buscadores o motores de búsqueda). Los primeros, es decir, permiten revisar de forma sistemática las páginas Web con la ayuda de una guía estructurada de la información (similar a buscar ciertos capítulos en la tabla de contenido de un libro), mientras que los *buscadores o motores de búsqueda*, permiten llevar a cabo un proceso de localización con la ayuda de bases de datos creadas para tal propósito (como ocurre cuando se buscan palabras en los índices alfabético y temático, al final de un libro).

Cuando se inicia el proceso de búsqueda en un libro, se utiliza el índice alfabético, el cual indica para cada entrada las páginas de su localización. Algo parecido sucede con el empleo de ficheros inversos para identificar la información en la WWW, donde para cada entrada se indican las *URLs* de localización correspondientes a dicha entrada.

⁷ Barrer, Joe, "Una aproximación a Internet y a sus herramientas de búsqueda". Disponible en Internet desde <http://www.eduteka.org/pdfdir/BuscadoresBasico.pdf>

⁸ Koch, Traugott, *Internet Search Service*. Disponible en Internet desde <http://www.lub.lu.se/tk/demos/DO9603-meng.html>

No obstante, en algunos casos, todavía existe confusión sobre el uso del término *Index*, dado que, dentro del contexto de la Web, tal término en inglés sería sinónimo de *Search Service* y no de *Directory*, mientras que en español se utiliza el término índice como sinónimo de directorio. La explicación es que, si volvemos a la analogía del libro realizada por *Koch*, en español nuestro concepto de índice del libro tendría un equivalente en inglés a *Table of Contents*, mientras que *Index* sería en español el índice alfabético de un libro.

Después de realizada esta precisión en cuanto al uso de los términos *Index* y *Directory*, es posible continuar exponiendo que, los directorios o índices se exploran mediante navegación (*browsing*), aunque, como ya se ha comentado, los directorios permiten un acceso estructurado (generalmente en forma de *árbol hipertextual*) y atendiendo algún criterio de clasificación convencional. Entonces, el usuario puede descender por niveles de especificidad, hasta encontrar la categoría adecuada a sus intereses. Si alguna de esas categorías es seleccionada, aparece una nueva pantalla con sub-apartados que mantienen una relación de dependencia jerárquica con el nivel superior.

Los sub-apartados incluyen enlaces con páginas Web, que responden al encabezamiento o, en su caso, a nuevos sub-encabezamientos. La metodología para realizar búsquedas es semejante a la que se sigue en cualquier *BDD*, aunque las posibilidades de búsqueda varían de un índice a otro (truncado, búsquedas por campo, usando *operadores booleanos*, etcétera) dependiendo de las opciones de gestión del programa.

Tanto los directorios como los buscadores generan bases de datos de *URLs*, pero, por norma general, los directorios recuperan registros documentales y los buscadores sólo páginas. Luego, para registrar un sitio, se incorporan ciertos datos complementarios que

se capturan en forma manual o automática, visto que, entre más completa sea la información descriptiva, más útil será para obtener los resultados deseados durante la búsqueda.

Entre los datos solicitados por los formularios, existen algunos que se consideran como imprescindibles: título, *URL*, clasificación temática (regularmente es convencional) y resumen. Otros se consideran sólo como recomendables: palabra clave, localización, idioma, tipo de información (académica, comercial, etcétera) y tipo de propietario (organismos públicos, empresariales, personales, etcétera).

La selección de los recursos en los directorios se realiza generalmente por personas que aplican determinados criterios de pertinencia para evaluar las sedes Web y que toman en cuenta si merecen o no estar incluidas en su directorio. Las sedes seleccionadas se registran en forma manual, mientras que los motores de búsqueda solicitan únicamente la *URL* para registrar los sitios en forma automática. Entonces, el robot de búsqueda visita e indiza la página dada de alta, así como las páginas relacionadas con ésta para establecer los enlaces correspondientes en un momento determinado.

Por su parte, el lenguaje *HTML* permite el uso de etiquetas en *metadatos*, a través de las cuales se puede precisar información para identificar una página al momento de crearla. Los datos solicitados para registrar las etiquetas de *metadatos* son los mismos que se encuentran en los formularios de alta para los índices, esto es: título, *URL*, una clasificación temática, palabras clave, localización, idioma, tipo de recurso y propietario de la página.

A pesar de sus ventajas, el uso de las etiquetas de *metadatos* no se hace obligatorio para crear una página. Debido a ello, muchas

de las *URLs* lamentablemente no las incluyen. Ahora bien, las características, que, desde el punto de vista de la realización de una búsqueda se encuentran comúnmente en un *buscador* son:

- *Formularios de búsqueda: simples y avanzados.* El formulario simple está dirigido a personas con poca familiaridad sobre el uso de las bases de datos, mientras que el avanzado brinda la oportunidad de realizar búsquedas más complejas a usuarios expertos.
- *Herramientas de búsqueda:* éstas ofrecen la posibilidad de truncar las palabras, usar *operadores booleanos* y recuperar términos compuestos (utilizando comillas o paréntesis). Estas herramientas permiten diseñar estrategias para ajustar las condiciones de la búsqueda de acuerdo a las necesidades del usuario, tal como sucede en una *BDD*.
- *Clasificación temática:* mediante el empleo de un sistema clasificatorio, regularmente convencional, se facilita la localización de la información requerida, en virtud de que ubica los términos de consulta según el contenido semántico de los mismos.
- *Campos de búsqueda:* éstos se hallan regularmente detallados en el formulario avanzado.

Para comprender la operatividad y la forma como se estructuran los directorios, se ha seleccionado uno dentro de un estudio que analiza los principales buscadores, realizado por Ángeles Maldonado y Elena Fernández:⁹

⁹ Maldonado Martínez, Ángeles y Elena Fernández Sánchez, *Evaluación de los principales "buscadores" desde un punto de vista documental: recogida, análisis y recuperación de recursos de información.* Disponible en Internet desde www.fesabid.org/servicios/publicaciones/valencia1998.htm

Yahoo! <http://www.yahoo.com/>

Significado: “Yet Another Hierarchical Officious Oracle”.

Origen: fue creado por dos estudiantes de la *Universidad de Standford* (David Filo y Jerry Yang), quienes (en 1994) comenzaron a organizar temáticamente las direcciones en Internet que habían ido recuperando por su cuenta y las colocaron en la Web, sin saber que dicha actividad tendría gran acogida entre sus usuarios, pues a finales de 1994 registraron un millón de visitantes. Este hecho les hizo ver que tenían un gran negocio entre las manos y, en marzo de 1995, *Yahoo!* se constituyó como empresa.

Contenido: cobertura internacional con nodos locales que sirven a sus respectivos ámbitos geográficos en Europa, Asia y América.

Registro de páginas: la mayoría de los sitios son sugeridos por los propios usuarios, y más tarde son visitados y evaluados por un equipo de *Yahoo!* Existe un formulario de sugerencias donde se solicita la *URL*, el título y la descripción del registro a incluir. *Yahoo!* está organizado en categorías temáticas, por lo que el solicitante debe indicar las categorías en las que piensa se debe enmarcar el registro sugerido.

Búsqueda: ésta se puede realizar siguiendo dos procedimientos, a través de la clasificación temática, o bien, por medio de los formularios de búsqueda. En el primer caso, se puede descender por la jerarquía hasta visualizar los temas de interés. En el segundo, se teclean los términos en una ventana, debiendo elegir entre los índices o los recursos que contengan dichos términos. Si el resultado no es satisfactorio, existe la opción de preguntar en las páginas indizadas por el buscador *Inktomi*, o conectarse con otros buscadores (*Google*, *Altavista*, *Infoseek*, etcétera). El sistema admite truncar, buscar por términos compuestos y el uso de los operadores booleanos.

Resultados: en el caso de que la búsqueda haya sido realizada por Yahoo!, de cada sitio, incluido se visualiza el título con el enlace, una pequeña descripción y la clasificación a la que pertenece. Si la búsqueda fue realizada por *Inktomi*, la descripción corresponde a las líneas de las páginas. En general, los resultados se amplían en la medida en que los documentos contengan más palabras clave, o cuando el recurso pertenezca a una categoría que esté situada como una palabra clave más general.

Ventajas: el registro de los recursos, se realiza a través de un formulario que incluye la *URL*, el título y una breve descripción del contenido, por lo que permite recuperar cada uno de los elementos mencionados por separado. Entre los resultados obtenidos, es más común que pueda recuperar recursos documentales y no sólo páginas. El directorio de *Yahoo!* cuenta con más de dos millones de Webs, indizadas en la base de datos *Google* y permite a más de 200 millones de usuarios utilizarla mensualmente.

Desventajas: no puede mantener abiertas, paralelamente, las listas de los resultados de la búsqueda junto con las direcciones recuperadas. Además, los términos que maneja en su clasificación temática son escasos y muy generales.

Con respecto a los buscadores, éstos almacenan en su base de datos gran cantidad de páginas Web, localizadas por *robots* internos, cuya misión es saltar de una página Web a otra empleando hiperenlaces para rastrear la información. Luego, dicha información se indiza y se introduce en una base de datos, que después es explorada utilizando un motor de búsqueda. A este tipo de robots se les denomina *Spiders* (arañas) o *Web Crawlers*.

Cada *robot* de la Web utiliza un método distinto de búsqueda; de ahí que la información almacenada en sus bases de datos sea diferente. Asimismo, cuando un *robot* de búsqueda llega

a una página virtual comprueba si ha sido visitada anteriormente. En caso de confirmar tal situación, examina si ha sufrido modificaciones y la actualiza. Para comprender la operatividad de los buscadores, veamos otro caso, recuperado del estudio realizado por Maldonado y Fernández.¹⁰

Altavista <http://altavista.digital.com/>

Significado: “Una visión desde las alturas”.

Origen: Nació en 1995 y fue el primer servicio de texto completo de Internet. Durante la primavera de 1995, los científicos del Laboratorio de Investigaciones de *Digital Equipment Corporation* en Palo Alto, California, crearon una forma para almacenar las palabras de las páginas *HTML* de la Web en un índice sobre el que se podían realizar búsquedas. Así fue como surgió *Altavista*.

Contenido: posee cobertura internacional con nodos locales que sirven a sus respectivos ámbitos geográficos en Australia, Canadá, España, Estados Unidos, Japón y Suecia.

Registro de páginas: las páginas son recuperadas e indizadas por un *robot* que además tiene la capacidad de saltar a las páginas enlazadas. El registro se realiza únicamente con la dirección de la página, pero el sistema reconoce las etiquetas del título y palabras clave, así como palabras libres dentro del contenido de la página.

Búsqueda: presenta dos formularios de búsqueda semejantes. Las búsquedas se pueden acotar por idioma, por campos concretos (como dominios, servidores, *URLs*, ligas, imágenes y título), utiliza *operadores booleanos*, *truncamiento* de palabras, proximidad y localización de términos compuestos. La diferencia entre sus formularios reside en que, en el simple, las páginas seleccionadas

¹⁰ Maldonado Martínez, A y E. Fernández Sánchez, *op. cit.*, p. 6.

se muestran en un orden de relevancia predefinido por el propio sistema, mientras que, en el avanzado, el usuario determina el orden de las páginas recuperadas.

Resultados: en cada página recuperada, se visualiza el título de enlace, la URL y las primeras palabras del documento. En caso de reconocer etiquetas con *metadatos de descripción*, recupera el contenido completo de las páginas. Incluye el tamaño de cada página (en *bits*), el idioma y la fecha de la última modificación.

Ventajas: reconoce etiquetas en *metadatos*, permite utilizarlas y mostrarlas con criterios de relevancia. Además, ofrece la opción de realizar búsquedas por campos concretos: dominios, servidores, URLs, ligas, imágenes, título e idioma, presentando tanto los registros de los resultados como las páginas en forma simultánea.

Desventajas: la más importante es que carece de una clasificación temática. Actualmente, ha decaído considerablemente, al competir con otros buscadores con mejores cualidades, pero sigue siendo reconocido como uno de los más populares.

En general, las características básicas que distinguen a un directorio o índice de un buscador son las siguientes:

	Directorios	Buscadores
Selección de recursos	<i>Realizado por personas.</i>	<i>Principalmente de forma automática.</i>
Representación del contenido del documento	<i>Clasificación manual.</i>	<i>Indización automática.</i>
Representación de la consulta	<i>Implícita (mediante navegación por categorías).</i>	<i>Explícita (mediante palabras clave, operadores, delimitadores, etcétera)</i>
Presentación de los resultados	<i>Páginas creadas previamente a la consulta. Poco exhaustivos pero más precisos.</i>	<i>Páginas creadas de forma dinámica para cada consulta. Muy exhaustivos, pero poco precisos.</i>

Tanto los buscadores como los directorios ofrecen un esquema de estructuración semejante: una base de datos, un programa de indización, un *robot* de búsqueda y una interfaz. Así que, considerando sus características, ventajas y desventajas de operatividad, es posible comentar lo siguiente:

Las direcciones de los índices y buscadores regularmente se acompañan del título y de una descripción del contenido. En los índices, el resumen del contenido del recurso se realiza manualmente. Mientras que, en los buscadores, sólo se incluyen las primeras palabras de una página, las cuales pueden arrojar ciertas pistas sobre el contenido de la misma, pero que no consiguen orientar adecuadamente al usuario.

Los directorios e índices utilizan taxonomías con un orden temático demasiado artificial y perecedero, ya que los términos de clasificación son muy amplios y escasos, pues se establecen en función del interés general de sus usuarios y no propiamente sobre el contenido de los recursos.

Entre las ventajas de los buscadores, existe la capacidad para extraer información mediante la identificación de etiquetas en *metadatos*. Dichas etiquetas son incorporadas en algunas páginas para indicar el título, idioma, país, institución responsable, etcétera, e incluso llegan a representar el contenido de los recursos a través de palabras clave. Entonces, los motores de búsqueda que reconocen los *metadatos*, aprovechan esa información tanto para la presentación de sus resultados como para la recuperación de sus contenidos.

Cuando los buscadores recuperan información desde el texto libre, es posible apreciar, ocasionalmente, que los resultados pueden ser útiles, pero en general no son plenamente satisfactorios. En estas herramientas, las páginas almacenadas no

se clasifican temáticamente, por tanto impiden realizar búsquedas jerarquizadas o temáticas como punto de apoyo. Debido a ello, los diseñadores de algunos motores de búsqueda han incluido un índice temático para suplir en cierta forma dicha carencia.

En comparación con los índices, parece que los directorios brindan más posibilidades de orientación al usuario, ya que al elegir una categoría principal en la búsqueda, la estructura del directorio permite descender por aquellas subcategorías que más se aproximen al término elegido, hasta llegar al término preciso, donde se encuentran los enlaces a los recursos de la Web. Sin embargo, el usuario podría tener muchas dificultades para decidir lo que le interesa, cuando el programa muestre una lista demasiado extensa sobre los términos de consulta.

Algunos directorios indican en cada nodo cuántas referencias y subcategorías contiene en cada una de las bifurcaciones posibles desde el nodo actual. Esto ayuda al usuario a tener una visión general sobre el volumen y contenido del índice, aspecto que es muy difícil de determinar en un motor de búsqueda.

Entre las desventajas de los directorios, cabe destacar que éstos no disponen de criterios de calidad confiables para seleccionar los recursos que formarán parte de su base de datos. Asimismo, muestra ciertas deficiencias en la lógica, jerarquía y también en la cobertura de su terminología. En general, carecen de principios de descripción y selección homogéneos.

En la Web, los directorios y los motores de búsqueda también se enfrentan a los cambios promovidos por la propia dinámica informativa de la Web (es decir, la aparición y desaparición de páginas, la redirección de un recurso, etcétera), lo que también condiciona los resultados de una búsqueda. Ello obliga a que, para validar y seleccionar los recursos, el usuario

deba explorar nuevamente los resultados de una búsqueda, mediante la navegación directa sobre los resultados obtenidos. De tal modo es posible señalar que las herramientas de recuperación en la Web, son un medio y no un fin en sí mismas.

Finalmente, en la Web, el trabajo relativo a la gestión de la información para facilitar su acceso, se ha visto influido, en cierta medida, por el trabajo que se ha realizado tradicionalmente desde la biblioteca, sin conseguir el mismo éxito. Debido a lo anterior, existen ciertas tendencias para afrontar y atender las problemáticas que desde la Web actual aún no se han resuelto en materia de acceso a la información.

CAPÍTULO 3

TENDENCIAS SOBRE LA ORGANIZACIÓN, REPRESENTACIÓN Y DISPOSICIÓN DE LA INFORMACIÓN EN LA WEB

Dificultades en la definición de los recursos en la Web

La recuperación de los recursos digitales sigue siendo una preocupación latente en la gestión de la información moderna. Homogeneizar y normalizar dicha información ha sido el objetivo principal que han perseguido las bibliotecas a lo largo del tiempo, aunque en la *Web* el crecimiento masivo de los recursos disponibles en línea, ha desafiado a la par que ha enriquecido las labores relacionadas con el tratamiento de la información, particularmente las vinculadas con la descripción y tipificación de los recursos digitales.

En su momento, iniciativas pioneras como el formato *MARC* surgieron tratando de enfrentar el universo de la información señalado en *HTM*, pero, el posterior desarrollo sobre la teoría para la descripción de los nuevos recursos, demostró que la *Web* era un entorno informativo más complejo y diferente al señalado por la biblioteca tradicional, ya que la diversidad de los recursos en la *Web*, se encuentra constituida por un conjunto prácticamente inabarcable de objetos y/o entidades de información de todo tipo, codificada en todos los formatos posibles, tales como texto, imagen fija, video, voz, etcétera y, no sólo eso, también afectó la forma tradicional de apreciar y reconocer la información, ya que de lo tangible hubo la necesidad de abrir un nuevo espacio de reconocimiento a lo intangible en el mundo de los recursos digitales.

Para identificar y describir los recursos digitales, es preciso determinar primero qué entidades de información son susceptibles

de organizarse a partir de los esquemas o modelos propuestos en *metadatos*. Debido a ello, más adelante se abordará el concepto de *Document Like Objects (DLO)* (Documento como objeto de información), a partir del cual es posible reconocer las unidades mínimas de información en la Web, para luego acercarse a los procesos de normalización y descripción de dichos recursos.

Sobre el concepto de *metadatos*, es factible comentar que tienen por objeto facilitar el acceso a los recursos documentales de la Web, aportando, como su nombre lo indica, datos sobre los datos. El origen de esta tendencia o teoría también parte (como ya se había dicho) de la saturación informativa de la Web, del crecimiento exponencial de sus recursos en diversos formatos de codificación y del reconocimiento de las limitaciones que los buscadores mantienen para la recuperación de información útil.

Según Eva Méndez,¹ los *metadatos* son datos estructurados y legibles por máquina, a los que es posible atribuirles múltiples funciones, pues permiten realizar diversas gestiones con los recursos informativos de la Web. Dada la diversidad de estas funciones, es posible tipificarlas de acuerdo a las necesidades descriptivas y operativas de los recursos en tres grupos:

- *Metadatos descriptivos*. Permiten identificar los recursos de acuerdo a la forma como se generan los modelos de representación e indización en la Web, los cuales, a su vez, se distinguen en dos categorías: *Metadatos orientados al recurso*, que por su proximidad a la catalogación tradicional, son aquellos modelos que definen las características básicas de un *objeto de información digital* y *Metadatos orientados a la materia*, que suponen modelos semánticos de

¹ Méndez Rodríguez, Eva María, *El documento como objeto de información: metadatos y esquemas de organización de la información en la Web*, pp. 51-52. Disponible en: [Http://www.ect.juntaex.es/dgsi/Documentacion/Eva.pdf](http://www.ect.juntaex.es/dgsi/Documentacion/Eva.pdf)

representación por materias tales como sistemas de clasificación, tesauros, etcétera.

- *Metadatos administrativos* que incluyen datos técnicos para facilitar el tratamiento y gestión de las colecciones digitales en la Web.
- *Metadatos estructurados*, los cuales suministran información sobre la configuración interna de los recursos en un sistema y sus relaciones con otros recursos dentro de la Web.

No obstante, el problema fundamental de los *metadatos* reside en elucidar qué es exactamente un objeto de datos, entendido como una unidad lógica de información estructurada, o unidad mínima susceptible de recuperarse en función de los datos que la describen. En otras palabras, precisar qué es un objeto informativo en el contexto de la Web. Una propuesta para definir esta situación es el denominado *Document Like Object (DLO)*, término que surgió en el seno del primer taller realizado en *Dublin Core* Ohio, Estados Unidos, en 1995. Desde entonces, la expresión *DLO* ha sido ampliamente reconocida en la literatura sobre *metadatos* y sobre gestión de la información en la Web, aunque coexiste con otros términos semejantes como *DIO* “*Document-like Information Object*”, *EIO* “*Electronic Information Objects*”, *DDO* “*Digital Data Object*”, etcétera. Mas para no entrar en controversia sobre el uso de estos términos y la variedad de sus definiciones, baste decir que el *DLO* se define como el conjunto de elementos digitales que un programa reconoce y manipula como si fuese un documento único (a través de un *URI* o *URL*), independientemente del formato que mantenga y los formatos que contenga.

A pesar de esta alternativa para reconocer los objetos de información en la Web, en la práctica, todavía puede resultar complicado identificar qué objetos de información serán

determinados como tales, para someterlos a un proceso de descripción y representación independiente, ya que lo que se percibe como un objeto de información en la *Web*, puede estar formado por una combinación de varios archivos. Incluso el contenido de cada archivo puede variar considerablemente de un momento a otro.

Por ejemplo, al considerar el caso de una *Página Web*, es preciso señalar que ésta es, hasta cierto punto, una ilusión creada por un *browser*, es decir, en la pantalla se observa un despliegue que aparenta ser sólo una página, cuando en realidad ese despliegue está formado por varios componentes independientes. En un estudio realizado por *Grete Pasch*,² se analizó una colección de 300 *páginas Web* y, en cada una de ellas, se encontraron tres tipos principales de objetos: 2,453 imágenes, 3,766 archivos *HTML* y 27 programas. De las 300 páginas, 276 tenían dos o más componentes, o sea, que es relativamente raro encontrar páginas sólo con texto, y cada uno de sus componentes tiene una propia dirección *URL*, por lo que, en teoría, cada uno de ellos podría ser catalogado independientemente. Lo anterior genera la disyuntiva siguiente: ¿qué resulta más conveniente catalogar, la *Página Web* en general o cada uno de sus componentes por separado?

Para el catalogador de una biblioteca tradicional, esta labor se encuentra relativamente resuelta, ya que, por ejemplo, en el caso de un libro impreso, el bibliotecario determina si catalogará una obra en su conjunto o cada uno de sus componentes por separado, dependiendo de los intereses funcionales que le pretenda brindar a dicha obra dentro de una colección, lo mismo sucedería en la *Web*. Mas, en el caso de los *metadatos*, todos los objetos pueden llegar a ser utilizables, por tanto, lo mismo se interesa en

² Pasch, Grete, *los metadatos y el reto a las tradiciones bibliográficas*. Disponible en <http://www.ucol.mx/interfaces/interfaces2001/archivos/gretepasch.pdf>

cada uno de los objetos informativos de la *Página Web* como en la propia *Página Web* en su conjunto.

Debido a lo anterior, los modelos propuestos en *metadatos* han emprendido el camino de definir las diversas manifestaciones que un *DLO* puede contener. Cabe señalar que dichos modelos no son normas, en el sentido estricto de la palabra, aunque buscan establecerse como estándares de regulación sobre los recursos de la *Web*. Así, por ejemplo, en el caso del *Dublin Core*, el elemento que refleja un objeto informativo es *DC.Type* y para describir *documentos hipermedia* (esto es, varios objetos interrelacionados entre sí) incluye tres elementos básicos: *DC.Form*, *DC.Type* y *DC.Relation*. En general, *Dublin Core* establece valores para cada objeto de información de acuerdo a la tipología siguiente:

- *Colección (Collection)*. Incluye un conjunto de *items*. Este término indica que el recurso digital se describe como grupo.
- *Conjunto de datos (Dataset)*. Se trata de información codificada en una estructura definida, por ejemplo: listas, tablas y bases de datos.
- *Evento (Event)*. Es aquella información que tiene la característica de no ser persistente como exposiciones, conferencias, talleres, etcétera.
- *Imagen (Image)*. En este caso, el *DLO* es ante todo una representación visual o simbólica como, por ejemplo, todas las imágenes o fotografías, cuadros, dibujos, películas, etcétera.

- *Recursos interactivos (Interactive Resources)*. Recursos que requieren algún tipo de interacción con el usuario para su ejecución. Ejemplo: los *multimedia* para el auto-aprendizaje.
- *Servicio (Service)*. En la categorización de *Dublin Core*, un servicio es un sistema que proporciona una o más funciones de valor al usuario final, por ejemplo, un servicio de préstamo interbibliotecario.
- *Software*. Es un programa para computadora que se encuentra disponible para su instalación en otra máquina.
- *Sonido (Sound)*. Se trata del contenido de un recurso cuando está pensado para distribuirse como audio, por ejemplo, los formatos de música pregrabada.
- *Texto (Text)*. Son aquellos recursos cuyos contenidos son equiparables a un documento como libros, conferencias, revistas, periódicos, etcétera.

Dublin Core se propone como un conjunto de elementos básicos para describir y almacenar, de forma sencilla, los recursos digitales de la Web y, por sus características, es posible afirmar que surgió como parte de un proceso evolutivo de los métodos tradicionales implicados en la catalogación bibliográfica, para la descripción de los recursos en la Web. *Dublin Core* es tan sólo un ejemplo de los esquemas que procuran mantener la visibilidad y recuperación de los recursos digitales a partir de los formularios utilizados por los propios motores de búsqueda en la Web.

Metadatos para la gestión y descripción de recursos en la Web

La importancia de los modelos de *metadatos* consiste en agregar y extraer datos sobre la información publicada en la Web, para facilitar su localización y recuperación. En los modelos de *metadatos*, el valor de los datos que se relacionan con un *DLO* radica en añadir a la estructura de un recurso digital diversas etiquetas descriptivas que consideren la propiedad intelectual, la tipología, la temática del recurso, etcétera para contar con más y mejores alternativas de acceso a dicho documento.

La descripción de un recurso digital a través de los *metadatos*, se establece bajo la misma dinámica de funcionalidad que la catalogación tradicional, es decir, identificar cada *DLO* en la Web, proporcionar distintas formas de acceso a cada recurso y relacionar la información contenida en cada registro con otros recursos en la Web. Aunque los alcances y necesidades descriptivas de los recursos digitales son diferentes a los impresos, es posible observar ciertas semejanzas entre ambos esquemas, por ejemplo:

- *Localización* Si bien un registro del catálogo, tal como se muestra en una *OPAC (Online Public Access Catalog)* puede hacer referencia a su localización física en la biblioteca, un registro de *metadatos* lo hará en función de las localizaciones remotas del recurso (por ejemplo, *FTP, HTTP*, etcétera).
- *El formato del documento*. Los documentos digitales pueden presentarse en diferentes formatos dentro de un mismo sitio (*HTML, PDF, ASCII*, etcétera), o en distintas versiones sobre un mismo documento, lo que en los catálogos tradicionales correspondería a los variados formatos o ediciones de una misma obra.

- *Definición del nivel de detalle.* En los documentos bibliográficos, el catalogador define el nivel de análisis descriptivo al que se someterá un documento. Por ejemplo, si el documento será descrito como un todo o por cada una de sus partes. Del mismo modo puede ocurrir con una *Página Web* y con cada una de sus partes.

El fundamento de las diferencias entre la actividad catalográfica tradicional y el registro de *metadatos*, obedece a que los recursos se establecen en un paradigma de producción totalmente diferente y esta circunstancia es, precisamente, la que condiciona la forma de ver, tratar y desarrollar los esquemas de descripción y normalización de los recursos digitales. De este modo, la asignación de *metadatos* por un lado y la catalogación convencional, por otro, ofrecen una doble visión sobre el procesamiento al que serán sometidos los recursos digitales, pues , mientras la asignación de *metadatos* responde a una visión basada meramente en la gestión de datos, la catalogación lo hace bajo un enfoque de *control bibliográfico*.

Ahora bien, los recursos digitales pueden y deben ser parte de las colecciones de las bibliotecas modernas. Por ello, establecer sistemas robustos de acceso a esa información tiene una vital importancia en el incipiente desarrollo de las *bibliotecas digitales*. Desde hace algunos años, se han desarrollado modelos de *metadatos*, con distintos niveles de difusión para describir las colecciones digitales. No obstante, los diversos modelos de metadatos no son normas como tales, responden, más bien, a “proyectos formativos” o estándares, empleados por determinadas comunidades.

Analizaremos a continuación algunos de los modelos más reconocidos o utilizados:

Dublin Core³

Desde su nacimiento, *Dublin Core* ha evolucionado como un formato de registro descriptivo para el intercambio de información en la Web. Debe su nombre al lugar en el que se llevó a cabo el primer encuentro realizado sobre *metadatos* en Dublin, Ohio (Estados Unidos), en marzo de 1995. Y, desde entonces, ha evolucionado como el marco normativo para la interoperatividad entre conjuntos de *metadatos* (a través del empleo de sistemas de clasificación documentales y/o vocabularios controlados) y, para facilitar el desarrollo de conjuntos de *metadatos* específicos en una disciplina o comunidad dentro del marco de recuperación de la información.

El modelo de *Dublin Core* comprende quince elementos, que pueden usarse para describir distintos objetos de información digital (véase el Anexo 1). Los elementos de dicho modelo pueden interpretarse desde tres grandes áreas que proyectan el alcance de la información que los caracteriza y que intentan aproximarse a las formas como los usuarios buscan la información. Estos elementos son:

- *Los relacionados con el contenido de un DLO*, es decir: título, materia, descripción, fuente, lengua, relación y cobertura.
- *Los relacionados con la propiedad intelectual del recurso*: creador/autor, editor, otros autores/colaboradores y derechos de autor.
- *Los relacionados con la temporalidad del documento*, tales como: fecha de creación del documento, tipo de recurso, formato e identificador del recurso.

³ Véase <<http://www.dublincore.org/>>

Text Encoding Initiative (TEI)⁴

TEI surge en 1987 como un proyecto del área de las humanidades y fue promovido en un congreso de la *Association for Computer Humanities*, pero su publicación efectiva para la codificación de textos literarios se estableció desde 1994. Se trata de un modelo maduro, formado a partir de *SGML (Standard Generalized Markup Language)* para garantizar su estructura como un formato estándar con amplias capacidades para el marcaje, la indización y el intercambio de información textual.

A pesar de sus orígenes lingüísticos, las recomendaciones de la *TEI* sirven para la descripción de la información digital de todo tipo. Uno de los componentes más significativos de este modelo es la cabecera *Text Encoding Initiative Head (TEIH)*, que permite realizar una descripción bibliográfica detallada para cada texto codificado. Desde 1998, se ha trabajado una versión de la *TEIH* en *XML (Extensible Markup Language)*, que sin duda ampliará el espectro de aplicaciones futuras de este formato de descripción en la Web.

Resource Description Framework (RDF)⁵

La importancia de este modelo de descripción reside en que, por un lado, es promovido por el *Consortio Web* (mejor conocido como *3WC*) y, por otro lado, en que se trata de una aplicación que utiliza el *XML*, lo que permite generar un marco estándar (para la descripción de contenidos) de interoperatividad entre los distintos modelos de *metadatos*, como *Dublin Core*. Dicho modelo está basado en las iniciativas de la *3WC*, tales como el *Platform for Privacy Preferences (P3P)*, que sirve para salvaguardar la privacidad en la Web y otros trabajos previos en el ámbito de los modelos de

⁴ Véase para consulta de la guía <http://www.tei-c.org/Guidelines2/>; para consulta del *software* <http://www.tei-c.org/Software/>, y para consultar proyectos [<http://www.tei-c.org/Applications/>](http://www.tei-c.org/Applications/)

⁵ Véase [<http://www.w3.org/RDF/>](http://www.w3.org/RDF/)

descripción de contenidos como, por ejemplo, *Warwick Framework* del *Dublin Core*.

El modelo *RDF* parte de una construcción tripartita básica, que incluye *recursos*, *propiedades* y *sentencias*. En dicho modelo, un *recurso* es una entidad que puede ser referenciada con un *Identificador Único de Recursos (URI)*, una *propiedad* que define una relación binaria entre recursos y/o valores atómicos de los tipos de datos provistos por *XML* y una *sentencia* que especifica el valor de una propiedad para un determinado recurso.⁶

Como es posible observar, el impulso de los esquemas que emplean *metadatos* para la descripción de los contenidos en la *Web*, requiere del desarrollo de estándares y la interoperabilidad entre éstos, a partir de la normalización formal, para optimizar la recuperación de la información en la *Web*. Y aunque en la *Web*, proliferan los proyectos y tentativas de normalización para recuperar, compartir y hacer compatible la información, el desarrollo de dichos estándares o proyectos de normalización, denotan el estado incipiente que existe sobre el tema de la identificación y descripción de los registros documentales en la *Web*.

En realidad, el proceso de evolución formal de los estándares de descripción en la *Web*, es demasiado lento, como para ajustarse a las necesidades que plantea el nuevo entorno digital y, por otra parte, la falta de masa crítica sobre las aplicaciones de dichos estándares, dificulta la tarea de evaluar el éxito de los mismos. Bajo estas circunstancias, se torna fundamental, el papel que juegan las instituciones de estandarización como la *ISO*, las

⁶ Steffen Staab, *et al.*, "An Extensible Approach for Modeling Ontologies", citado por Silva Muñoz, Lydia, p. 2. Disponible en [Http://www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo_lydia.pdf](http://www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo_lydia.pdf)

encargadas del desarrollo de la Web como la W3C y las empresas que desarrollan el *software*, para asimilar estas tendencias.

Propuesta para el desarrollo de la Web Semántica

Junto al crecimiento informativo de la *Web*, la tecnología informática que la sustenta ha experimentando una evolución, desde las primeras tecnologías básicas como *HTML* y *HTTP* hasta las más recientes como *Java*⁷, *CGI*⁸, *JSP*⁹ y *XML*,¹⁰ tan sólo por citar algunos ejemplos. Estas tecnologías han influido y se han dejado influir por el propio ritmo que impone el desarrollo y transformación de la *Web*, esto es, por la generación dinámica de sus páginas, la mayor interactividad con el usuario, etcétera.

No obstante, entre las más recientes tendencias sobre las formas de concebir a la *Web* en un futuro, se encuentra la llamada *Web Semántica*, promovida por el propio inventor de la *Web* y presidente de la 3WC Barnes-Lee,¹¹ quien señala que la *Web Semántica* es un área que propone introducir descripciones explícitas sobre el significado de los recursos, para que las propias máquinas logren hacerse cargo del trabajo que ordinariamente realizan los usuarios cuando navegan e interactúan con los contenidos de la *Web*.

⁷Véase <http://www.java.com/es/>

⁸ CGI es una norma para establecer comunicación entre un servidor *Web* y un programa para interactuar con la *Web*. Véase: <Http://www.recursoescgi.com/html/gp7.html> o <Http://www.ok.cl/cgi/chap0/>

⁹ JavaServer Pages (JSP) es un *software* que sirve para generar, de forma dinámica, páginas *Web* en el servidor. Véase: <Http://www.java.com/es/download/manual.jsp>

¹⁰ XML es la sigla de eXtensible Markup Language (lenguaje de marcado extendible). Una de las principales funciones de este programa es separar la estructura del contenido en HTML y permitir el desarrollo de vocabularios modulares, compatibles con cierta unidad y simplicidad del lenguaje. Además, tiene otras aplicaciones, entre las que se destaca su uso como estándar para el intercambio de datos entre diversas aplicaciones. Véase: <Http://dat.etsit.upm.es/~abarbero/curso/xml/xmltutorial.html>

¹¹ Barnes-Lee, T., J. Hendler y O. Lassila, "The Semantic Web", p. 2.

La *Web Semántica* busca la posibilidad de compartir e integrar más fácilmente la información con los servicios de la *Web* actual, para sacar mayor provecho de los recursos y servicios disponibles. Aunque también obedece a la necesidad de dar mayor capacidad de representación a los recursos de la *Web*, para poder expresar y recuperar sus contenidos.

Actualmente, resulta difícil moderar las tareas relacionadas con la representación de la información en la *Web*, entre otras razones, porque los programas hacen que la semántica de los recursos se limiten casi exclusivamente al uso de palabras clave, cuya asignación suele ser plana e inconexa, lo que impide recuperar resultados con significados más elaborados. No obstante, la *Web Semántica* supone superar esas limitaciones mediante una estructura que permita anotar los recursos con un mayor significado, explícito preferentemente, para las máquinas. Además, se sujetaría a ciertos principios operativos de la *Web* actual, tales como compatibilidad, descentralización, crecimiento y uso de nuevos recursos no previstos en programas recientes.

Para cumplir el objetivo anterior, la *Web Semántica* ha rescatado del campo de la *Inteligencia Artificial* (y éste, a su vez, del campo de la *Filosofía*) la noción de *ontología*. Sobre la propuesta de este concepto, Gruber¹² señala que se trata de una taxonomía de palabras con atributos y relaciones, que proporciona una terminología consensuada para definir redes semánticas de unidades de información interrelacionadas, mientras que, por su parte, Qin y Pailing¹³ añaden que las ontologías tienen un alto grado de flexibilidad, pues admiten fórmulas de distribución y reutilización de diversos términos descriptivos en la *Web*.

¹² Gruber, T. R. "A Translation Approach to Portable Ontology Specifications", pp. 199-200.

¹³ Qin, J., Pailing, S., "Converting And Controlled Vocabulary into an Ontology: The Case of GAME". Disponible en [Http://informationr.net/ir/6-2/paper94.html](http://informationr.net/ir/6-2/paper94.html)

Según Antonio García,¹⁴ para construir formalmente una ontología, es necesario considerar los elementos básicos siguientes: *clases y subclases*, *slots* (en ocasiones también llamados roles que determinan las propiedades y las características de cada concepto en varios rangos) y las *facetas* (conocidas, además, como restricciones de roles, las cuales describen el tipo de valor y cualquier otra característica de un *slot*). En dicha estructura, los campos temáticos se organizan a través de facetas, y las relaciones que conectan los términos de una misma categoría se señalan mediante vectores internos. Estas relaciones requieren una definición tanto de las clases como del comportamiento de cada subclase para hacerlas operativas.

En cuanto a la definición de las clases y subclases dentro de una ontología se han realizado diversos estudios enfocados a identificar la presencia de las formas verbales en que se manifiesta el contenido de un recurso, partiendo del lenguaje natural. En estos estudios, los verbos pueden tipificarse en clases o por dominios, para la elaboración de modelos semánticos específicos. Aunque también existen otros métodos para la determinación de ontologías, por ejemplo, aquellos que tienen que ver con la investigación de las propiedades de los conceptos y sus relaciones; estructuras conceptuales asimilables en redes jerárquicas de trabajo, etcétera.

Ding y Foo¹⁵ señalan que estos métodos de trabajo parten habitualmente de los llamados *datos fuente* (es decir, vocabularios controlados, corpus de sentencias o texto libre), los cuales permiten elaborar conceptos (empleando técnicas como el análisis

¹⁴ García Jiménez, Antonio, "Instrumentos de representación del conocimiento: tesauros versus ontologías", p. 82. Disponible en: [Http://www.um.es/fccd/anales/ad07/ad0706.pdf](http://www.um.es/fccd/anales/ad07/ad0706.pdf)

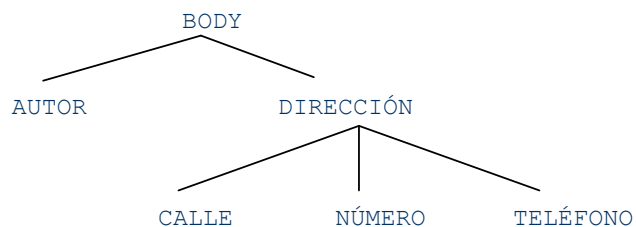
¹⁵ Ding Y. y Foo S. "Ontology Research and Development. Part 1. A Review of Ontology Generation", pp. 123-136.

sintáctico en forma manual o automática, procesamiento de lenguaje natural, etcétera) y la extracción de relaciones (tarea que normalmente se lleva a cabo en forma automática a través de algoritmos, aunque en ocasiones también se realiza manualmente). Así, la representación de una ontología contempla una estructura que se nutre de la elaboración de conceptos y el establecimiento lógico de relaciones entre ellos, considerando al *XML* como un lenguaje básico.

Si bien *XML* no fue expresamente pensado para definir ontologías, desde su aparición en 1998, se ha mantenido como el estándar más utilizado hasta hoy y, por tanto, se ha tomado en cuenta como base para la representación explícita de los datos y la estructura de los contenidos digitales de camino a la *Web Semántica*. Asimismo, *XML* ha permitido que los creadores de *Páginas Web* usen su propio conjunto de etiquetas de demarcación (*markup-tags*), las cuales pueden ser elegidas de manera que reflejen la semántica de un dominio. Por ejemplo:

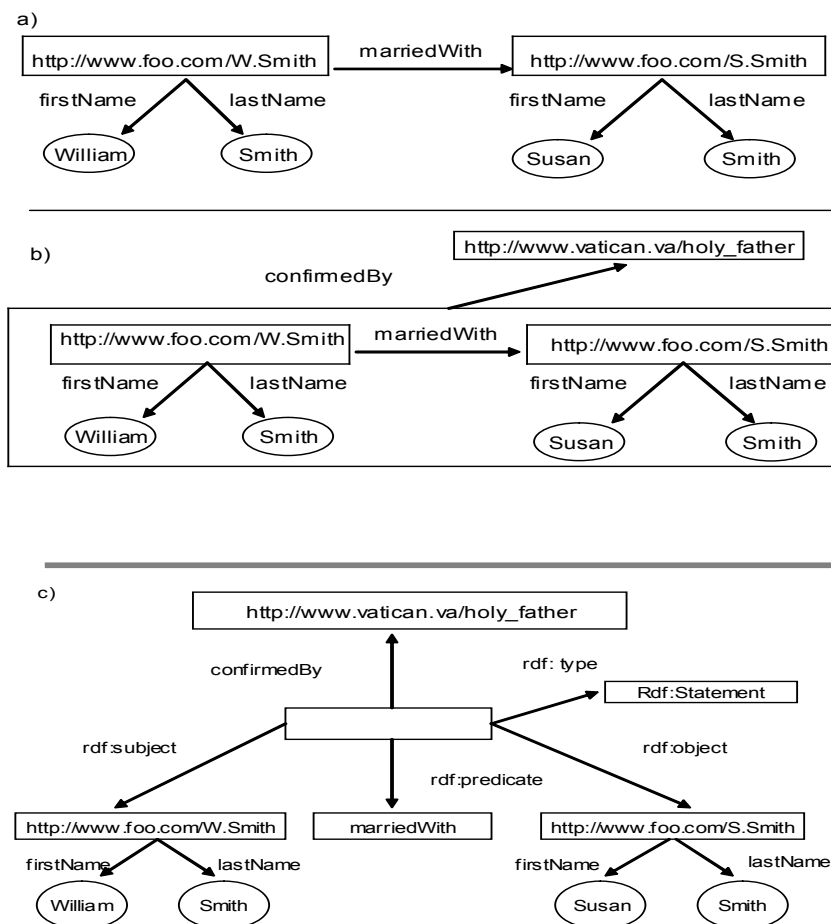
```
<BODY>
  Esta página fue escrita por
  <AUTOR> Pedro </AUTOR>
    <DIRECCIÓN>
      <CALLE> García Zúñiga </CALLE>
      <NÚMERO> 1234 </NÚMERO>
      <TELÉFONO> 57777777 </TELÉFONO>
    </DIRECCIÓN>
</BODY>
```

En este caso, las marcas `<AUTOR>` Y `</AUTOR>` denotan el comienzo y fin de los datos almacenados entre ellas (es decir, el *string*). Estableciendo una estructura con las siguientes relaciones:



El lenguaje *XML* puede proveer información semántica como subproducto, al precisar la estructura de un recurso, gracias a la *hipertextualidad*, de tal modo que la estructura y semántica en *XML* quede entrelazada. Pero, en la práctica, el lenguaje *XML* manifiesta ciertas limitaciones, porque ofrece un modelo que no distingue un mayor contenido dentro de las relaciones entre los objetos, ni jerarquías entre las clases.

En consecuencia, se han generado otros lenguajes como vehículo para la definición de ontologías. De ellos, el *RDF* se ha establecido como uno de los estándares más populares. Veamos el siguiente ejemplo:



En la parte (a) del esquema anterior, se definen dos recursos. Cada uno asocia una propiedad `FirstName` y `LastName` con los

valores reconocidos como *William* y *Susan Smith*. Ambos recursos tienen un *URI* y dichos identificadores se hallan relacionados por la propiedad `MarriedWith`, que señala que *William* está casado con *Susan*. En la parte (b), se muestra una relación que expresa una sentencia, la cual define, a su vez, una propiedad para el nuevo recurso obtenido. El ejemplo denota que el matrimonio de los *Smith* ha sido confirmado por el recurso del *Vaticano*. Y el modelo (c) ofrece el recurso predefinido por `rdf:statment` y las propiedades predefinidas `rdf:subject`, `rdf:predicate` y `rdf:object` para rectificar la sentencia inicial.

Dentro del proceso evolutivo de los lenguajes, se han desarrollado distintos programas cada vez más complejos para crear ontologías tomando en consideración el *RDF*. Entre los más reconocidos se encuentran: *Ontology Inference Language (OIL)* y *DARPA Agent Markup Language*, mejor conocido como *DAML*, creados en Europa y Estados Unidos, respectivamente. Con el tiempo, ambos lenguajes se fusionaron y surgió *Ontology Web Language (OWL)*, el cual representa un esfuerzo para enmendar los problemas y reunir las ventajas de *OIL+DAML*. *OWL* es un lenguaje complejo que opera bajo la misma división tripartita que *RDF*, pero con mayor capacidad de representación debido a la lógica descriptiva en que se sustenta. *OWL* se formuló con la capacidad expresiva de *RDF*.

Técnicamente, *OWL* se descompone en tres lenguajes de representación: *OWL Lite*, *OWL Dl* y *OWL Full*. *OWL Lite* es el nivel más simple de los tres y permite realizar clasificaciones con ciertas restricciones. *OWL Dl* tiene mayor expresividad que el nivel anterior y admite toda la normatividad estructural de *OWL*, aunque también con ciertas restricciones. *OWL Full* ofrece la misma expresividad que *OWL Dl* sin restricciones, pero no es de uso común, ya que es difícil que una computadora pueda soportar el razonamiento completo de cada una de sus funciones.

En general *OWL* permite realizar representaciones temáticas, a partir de un conjunto de recursos, mediante la definición de clases. Las clases y subclases se determinan a través de axiomas que, a su vez, se validan de acuerdo a ciertos indicadores de compatibilidad, intersección, complementariedad, diferencia, unión, equivalencia, restricciones de rango, dominio, etcétera. Aunque no es intención del presente trabajo indicar cómo opera exactamente *OWL*, baste comentar que contiene herramientas para la representación de contenidos en la *Web* muy avanzadas, la *Guía OWL*,¹⁶ ofrece diversos ejemplos que a detalle muestra cada uno de los elementos normativos de *OWL*.

Ahora bien, los lenguajes para la construcción de ontologías todavía se encuentran en pleno desarrollo, aunque su diversidad y presencia en la *Web* demuestran que ya existen suficientes elementos para iniciar la transición a la llamada *Web Semántica*.¹⁷ Sin embargo, esta transición podría considerar, al menos en un principio, cierta afinidad con la tecnología actual, es decir, idealmente debería conservar ciertos lenguajes compatibles con los navegadores actuales, por ejemplo, una alternativa podría consistir en conservar los documentos actuales en *HTML* y crear ontologías asociadas, debido al gran número de documentos disponibles en *HTML*.

Además, la construcción de la *Web Semántica* sugiere la necesidad de resolver el problema de determinar qué tipo de ontologías se establecerán para ciertas comunidades de usuarios, pues la implantación de las ontologías debe suponer que cada

¹⁶ Bechofer, Sean, et al. "OWL Web Ontology Language Reference". Disponible en [Http://www.w3.org/TR/2004/REC-owl-ref-20040210](http://www.w3.org/TR/2004/REC-owl-ref-20040210)

¹⁷ Se ha avanzado mucho sobre las herramientas, los estándares y la infraestructura necesaria para la *Web Semántica*. Asimismo, es posible reconocer un gran interés académico y corporativo sobre el tema, tal como lo refleja la *Internacional Semantics Web Conference*, que se lleva a cabo anualmente, auspiciado por la *3WC* y la revista *Journal of Web Semantics*.

parte de un sistema conlleva particularidades distintas, vinculadas con el uso que se brinda a la información.

La importancia de las ontologías se encuentra ligada a la capacidad para crear representaciones derivadas del desarrollo lógico, semántico y discursivo, que se realice sobre las características de los recursos, aunque es necesario advertir sobre la inconveniencia de las relaciones meramente asociativas, ya que éstas se realizan en menoscabo de las relaciones jerárquicas. En otras palabras, las funciones asociativas facilitan la construcción de un vocabulario y la falta de funciones jerárquicas dificultará la tarea de articular los términos relacionados dentro del mismo en un contexto dado. La eliminación de relaciones jerárquicas, en su nivel pragmático funcional, tendría efectos de alcance cuantitativo y cualitativo, en cuanto a la efectividad de la representación y recuperación de los contenidos de los recursos se refiere.

Entre las principales diferencias sobre las ontologías y el trabajo documental, es posible advertir que éstas se establecen primordialmente sobre los objetivos que cada una de ellas persigue y, en consecuencia, sobre la manera de interpretar y procesar la información. Es decir, mientras las ontologías se realizan con fines referidos fundamentalmente al uso de los recursos para fomentar el comercio electrónico de los mismos, la bibliotecología lo realiza buscando un *control bibliográfico* para apoyar el trabajo académico, científico y difundir la información de acuerdo a cierto tipo de intereses comunitarios.

Finalmente, según Kwasnik,¹⁸ a partir de la complementariedad entre el lenguaje natural y los lenguajes controlados, en términos de representación de la información, es

¹⁸ Kwasnik, B., "The Role of Classification in Knowledge Representation and Discovery", pp. 22-47.

posible crear ontologías como instrumentos que pueden derivar en tesauros o a la inversa. Por esta razón, el profesional de la bibliotecología podría jugar un papel importante en la creación de ontologías y utilizar las herramientas desarrolladas para la construcción de las mismas, con el fin de ofrecer una mayor proyección para dar, así, acceso a la información a través de esquemas de representación con los que habitualmente trabaja tales como los sistemas de clasificación documentales, tesauros, etcétera.

CAPÍTULO 4

VIABILIDAD DE LOS SISTEMAS DE CLASIFICACIÓN DOCUMENTAL EN LA WEB

Sobre la naturaleza de los sistemas de clasificación documental

La trascendencia de los sistemas de clasificación documental se encuentra justificada, primero, por los antecedentes históricos que los vinculan con la necesidad de clasificar el conocimiento y, después, por la visión y utilidad pragmática para la representación y organización de los documentos y sus contenidos. No obstante, una de las principales dificultades, en el desarrollo de los sistemas de clasificación, proviene de los límites establecidos por el concepto mismo de organización del conocimiento. Esto se debe a la imposibilidad espacio-temporal de crear una clasificación de carácter general del conocimiento, desde un punto de vista teórico, ya que las clasificaciones existentes tienen un carácter provisional en la naturaleza de las mismas y obedecen a parámetros sociales que los determinan desde su momento de creación.

La clasificación de las ciencias reparte las distintas disciplinas en clases, de modo que las ordena a partir de características afines que establecen cierta conexión entre ellas. La clasificación puede verse, entonces, como el resultado de hacer divisiones sobre subdivisiones sobre un conjunto de clases. Mas la invalidez de la organización del conocimiento se determina, precisamente, porque la organización o sistematización de las ciencias supone la estructuración de la realidad y del conocimiento, la cual también se modifica según las concepciones del mundo de aquellos que las realizan. De ahí, el carácter provisional de la clasificación científica, e incluso, de cualquier otro tipo de clasificación.

Rosa San Segundo¹ realizó un amplio estudio sobre la clasificación y, dentro de su obra, son destacables las ideas de los autores siguientes.

De acuerdo a Perec:

[...] cualquier orden prescribe y además nunca responde a criterios satisfactorios. Luego, la imposibilidad de una clasificación se establece entre lo alcanzable y lo inalcanzable, y precisamente en el ámbito de lo inalcanzable, es donde se fragua la idea del orden y el desorden. Ya que difícilmente el pensamiento se remite a un saber constituido, pues en todo caso, se trata de una acumulación de carácter intuitivo, no organizado.²

Aunque, en su obra, Perec también reconoce que las clasificaciones documentales representan una herramienta de auxilio para la memoria, ya que permiten interpretar y aproximarse al conocimiento en alguna fase de su desarrollo.

Por su parte *Durkheim* y *Maus*, citados por San Segundo, aseguran que:

[...] las clases sociales determinan las distintas estructuras de la clasificación del universo. Pues la clasificación es una elaboración abstracta del pensamiento, ya que señalan que, en la base de todas las concepciones clasificatorias laten diversas formas de organización subordinadas a factores económicos y políticos, que en conjunto, constituyen una organización social y que este entramado preside la organización lógica de las cosas.³

Ambos autores coinciden en señalar que las clasificaciones se forman según la organización social de la que surgen, ya que, para ellos, cada sociedad es un modelo dentro del cual se desenvuelve el pensamiento clasificatorio de las cosas.

¹ San Segundo, Rosa, *Sistemas de organización del conocimiento: la organización del conocimiento en las bibliotecas española*, pp. 2-30.

² *Ibidem*, p. 8.

³ *Id.*, p. 29.

Pese a las dificultades derivadas de la influencia social y el carácter temporal con el que se establecen las clasificaciones del conocimiento, cabe aclarar que los sistemas documentales no se constituyen del mismo modo y para los mismos fines que los sistemas de clasificación de las ciencias, aunque es necesario hacer hincapié en que deben su impulso a los esfuerzos realizados para clasificar el conocimiento científico.

La clasificación documental forma parte de la identidad en toda biblioteca, pues, en virtud de ella, la biblioteca establece su capacidad para definir un núcleo del conocimiento, a partir del cual cumple con ciertas funciones. Mas la práctica de clasificar, desde la perspectiva de la bibliotecología, no sólo involucra a los sistemas de clasificación y sus técnicas de aplicación, sino que también alude a la comprensión de fenómenos relacionados con los procesos de selección y evaluación, indispensables para determinar y medir la funcionalidad y operatividad con la que se manejan sus colecciones. En otras palabras, la determinación de intencionalidades sobre el uso y conformación de un acervo, contribuye a establecer los procesos de selección y el modelo de clasificación que será adoptado para organizar y controlar los documentos. Luego, la operatividad de dicho modelo deberá ser evaluado para poder precisar el nivel de confiabilidad que ofrece como sistema.

Dentro de nuestra disciplina, clasificar conlleva un ejercicio intelectual que debe realizarse con cierto nivel de abstracción para crear representaciones por cada documento y contenido. Con el fin de lograr dichas representaciones, es necesario interpretar y determinar los elementos más significativos del documento, tomando en consideración los diversos contextos en los que dicho documento se encuentra involucrado, tales como su valor social, la posibilidad de uso que se le adjudicará dentro de una comunidad y las formas en que circula y se encuentra disponible.

La construcción de representaciones bien puede desencadenar en la creación de símbolos (como los códigos de Dewey, LC, etcétera) o expresiones lingüísticas (provenientes de tesauros, encabezamientos de materia, etcétera) las que, por sí mismas, mantienen un sentido y un significado propios, pero que, al incorporarlas dentro de un sistema, cobran un sentido y un significado distinto, debido a las relaciones que establecen con otras representaciones. En otras palabras, la representación de un documento mantiene la intención de individualizarlo para identificarlo y, posteriormente, distinguirlo y recuperarlo, con base en sus relaciones de identidad temática y descriptiva con otras representaciones; entonces, cuando dicha representación se observa a partir del sistema en que opera, adquiere una connotación diferente, debido a las relaciones mismas del sistema documental y a los contextos culturales del medio donde se encuentra inmersa la colección.

Por otro lado, si se asume que el flujo de la información es de naturaleza caótica (fenómeno particularmente visible en la *Web*), entonces, es factible señalar que la clasificación documental tiende a disminuir el caos y a reducir la ambigüedad para generar un orden, a partir de las representaciones y relaciones temáticas y descriptivas de los documentos. Asimismo, los sistemas de clasificación documental contribuyen a establecer núcleos específicos de concentración y representación del conocimiento frente a la turbulencia de la información. Sin embargo, el impulso generado por los buscadores de la *Web*, para rastrear diversos recursos digitales por medio de la generación de taxonomías, es decir, concentrando diversos términos a partir de la frecuencia con la que aparecen dentro de sus propios contenidos textuales, sin el establecimiento de un esquema semántico, parece ubicar a las clasificaciones documentales en un nivel funcional intermedio, entre las clasificaciones de la ciencia y la generación de taxonomías dentro de la *Web*.

Ahora bien, la clasificación y la búsqueda de información documental son normalmente concebidas como un método por medio del cual los usuarios conocen acerca de los contenidos temáticos de los documentos, o la existencia de determinado tipo de ellos, a partir de un contexto dado. Mientras que en la *Web*, las taxonomías sólo permiten identificar aquellos recursos (sin tratarse específicamente de documentos) que concuerdan con el término solicitado, pero sin ubicarlo dentro de un contexto semántico y sin reducir la ambigüedad que propicia la sinonimia y la polisemia del propio término rastreado.

La relevancia de los sistemas documentales sobre las taxonomías, se debe a que los principios de selección y clasificación documental no se configuran únicamente para establecer un orden intencional de sus colecciones, sino que, además, constituyen una respuesta a la necesidad de incidir en el flujo mismo de la información. Pues el ejercicio de coleccionar y clasificar cualquier tipo de documento, guarda estrechos vínculos con la forma como se desenvuelve la circulación de los mismos.

En general, las concepciones teóricas acerca de la clasificación deben contribuir a generar principios que permitan el descubrimiento de nueva información, para que además de ser funcionales en la búsqueda, sirvan para sugerir y comprender de qué se trata un orden documental determinado, de tal manera que sea factible explorarlo para encontrar aquello que se busca y lo que no se desea encontrar específicamente.

La comprensión de la naturaleza de la clasificación debe redundar en conocimientos teóricos, que fundamenten el diseño de los sistemas de clasificación para la organización bibliotecaria. Pues es la idea del sistema, en su sentido lógico y relacional entre diversos elementos, la que mantiene unida y le brinda un significado a ese caos amorfo de la información, para que de ese

modo realmente pueda contribuir al desarrollo del conocimiento, en un ambiente que nos permita inferir preguntas a partir de un núcleo temático y organizado, y no derive en simples relaciones de datos o términos.

El ejercicio de replantear la clasificación, como objeto de estudio, exige no sólo que la formulación de enunciados tenga significados claros y específicos sino que sean entendidos en el mismo sentido por todos aquellos que los usan. Es decir, el desarrollo de criterios claros y objetivos de aplicación no bastan para ser útiles, deben reflejar uniformidades en el objeto, motivo de clasificación. Ello implica una clara definición del campo y fenómenos que abarca la clasificación para brindar los elementos necesarios de orientación en la organización documental, con objeto de que los resultados y consecuencias de su aplicación sirvan para poner en marcha procesos que modifiquen los conocimientos de una persona, aunque sea de manera superficial.

La comprensión teórica de la clasificación, en el ámbito bibliotecológico, permite rescatar los valores y finalidades humanas presentes en los procesos de selección y clasificación, con los cuales es factible subordinar las técnicas y las tecnologías al cumplimiento de estas finalidades y no a la inversa, como sucede cuando la tecnología se toma como modelo para determinar las formas y los contenidos de la clasificación. Asimismo, la utilidad de las clasificaciones documentales se debe a la funcionalidad de sus representaciones, toda vez que intentan reflejar los conocimientos y valores específicos de cada cultura, cada sociedad y cada época desde un punto de vista pragmático.

En general, el enfoque pragmático-funcional de las clasificaciones documentales se basa en el hecho de que pretende proporcionar el medio para reunir y organizar un gran número de documentos a partir de su descripción y su contenido. Bajo estas

condiciones, un sistema de clasificación documental ordena, a través de la asignación de un tema y una nomenclatura, los fondos documentales de un acervo, para generar herramientas (como los catálogos) que se encuentran regidos por un orden intencional, interrelacionado y estimulante para la investigación documental.

Por otro lado, en la constitución misma de las distintas clasificaciones documentales, es posible observar que prevalecen características o criterios que han determinado su adecuación y su desarrollo como sistema. Con el tiempo, esas características han establecido la base y las diferencias entre las clasificaciones documentales, de las que en el *Anexo 2* tan sólo se mencionará, de modo general, la estructura de aquellas que han cobrado mayor relevancia por su permanencia en la *Web*. Estos sistemas son: la Clasificación Decimal de Dewey, la Clasificación Decimal Universal y la Clasificación de la Biblioteca del Congreso (aunque por su funcionalidad la Clasificación Bibliográfica de Bliss y la Clasificación Colonada de Ranganathan podrían ser candidatas para influir en los procesos de sistematización documental en la *Web*). Sin embargo, es necesario subrayar que la permanencia de estos sistemas de clasificación, en realidad, ha derivado de su popularidad en la aplicación y difusión en reconocidas bibliotecas y amplios sectores bibliotecarios.

La intención de proyectar los sistemas documentales en la *Web* y la introducción de las nuevas tecnologías en la biblioteca, han motivado, en su conjunto, una radical revisión de los conceptos básicos de los problemas de clasificación documental y han contribuido a modificar las concepciones acerca de la naturaleza de los procesos para seleccionar y organizar los documentos digitales. Todos los procesos del análisis documental, encaminados a lograr una sistematización coherente de la representación y establecimiento de relaciones entre los contenidos temáticos de los documentos, han adquirido un nuevo sentido, al

pretender formalizar y convertir las representaciones de los documentos en signos, susceptibles de ser procesados por una computadora y dejar, a los algoritmos de los programas, las posibilidades de rastrear, relacionar y representar los contenidos de los documentos digitales.

A continuación se describe, brevemente, la trayectoria que se ha mantenido durante las últimas décadas, para incorporar y explorar los esquemas de clasificación documental más comunes en la *Web*, esto es, mediante el desarrollo de diversos programas experimentales que han logrado realizar la clasificación automática de diversos recursos digitales. Aunque es necesario reconocer que la clasificación automatizada, aún requiere de la actividad humana, es decir, el pensamiento abstracto para la determinación de las representaciones documentales y el entendimiento de las necesidades del entorno y sus usuarios, pues estos aspectos aún no han podido ser sustituidos, en su totalidad, por la inteligencia artificial o los sistemas expertos.

Los sistemas de clasificación documental a través de algunos de sus proyectos en la Web

Desde la incorporación de los primeros catálogos en la *Web*, ha permanecido latente la inquietud en demostrar el potencial de los principales sistemas de clasificación documentales y explorar sus posibilidades como instrumentos para organizar y mejorar el acceso a la información. En ese sentido, según Fernández-Molina,⁴ el primer proyecto importante que intentó evaluar las ventajas de la recuperación documental a partir del uso de un sistema documental fue desarrollado en 1968 por Freeman y Atherton.

⁴ Fernández-Molina, J. Carlos y Eduardo Peis, "Uso de esquemas de clasificación para mejorar las presentaciones de visualización de los catálogos en línea", p. 297.

Para llevar a cabo su proyecto, Freeman y Atherton crearon un fichero bibliográfico con base en la *CDU*, en el que los números, notas y referencias cruzadas constituían un índice referencial. Luego, para rastrear y recuperar los documentos, usaron un sistema interactivo denominado *Audacious*⁵ con el que los usuarios realizaban diversas búsquedas, empleando términos del lenguaje natural o utilizando los números topográficos de la *CDU* y los *operadores booleanos*. El sistema permitió, entonces, realizar el cruce entre los términos usados y los números *CDU*, para situar los documentos en su contexto jerárquico.

El éxito en los resultados obtenidos por Freeman y Atherton, les permitieron concluir que la *CDU* podría ser utilizada como lenguaje de indización en un sistema automatizado y que esta misma idea podría llevarse a cabo con otros sistemas de clasificación documental como la *DDC* y la *LC*.

Sin embargo, Pauline Atherton⁶ fue la primera en sugerir que el empleo de los sistemas de clasificación documental en la *Web*, podrían facilitar el acceso a la información mediante la disposición de catálogos en línea, y que los términos de las tablas clasificatorias proporcionarían el lenguaje necesario para recuperar los registros documentales en su contexto temático.

Durante los años ochentas, surgieron nuevas propuestas que intentaron demostrar el potencial de los sistemas documentales. Por ejemplo, en 1987, Karen Markey⁷ inició una investigación en la que se estudiaban las posibilidades de la *DDC* para facilitar el

⁵ *AUDACIOUS* se basó en una de las primeras versiones de *RECON*. Era un programa desarrollado por la NASA, para más información véase: Bourne, C. P. y Hahn T., *A History of Online Information Services, 1963-1976*, 122 pp.

⁶ Atherton, Pauline, *Books are for use: Final Report of the Subject Access Project to the Council on Library Resource*, citado por Fernández-Molina, J. Carlos, *op. cit.*, p. 297.

⁷ Markey, Karen, "Subject Searching Strategies for Online Catalogues through the Dewey Decimal Classification", citado por Fernández-Molina, J. Carlos, *op. cit.*, pp. 298-299.

registro y la recuperación de la información a través de sistemas interactivos en línea. Este proyecto, denominado *DDC Online Project* y patrocinado por *Online Computer Library Center (OCLC)*, permitió confirmar las ideas propuestas por Atherton, ya que, según Markey, al incorporar un sistema de clasificación documental en la *Web*, se generaba una herramienta funcional para apoyar el trabajo de búsqueda y se lograba un acceso más eficaz a la información.

Los trabajos de investigación realizados, durante esa década, dejaron de manifiesto no sólo la utilidad pragmática de las clasificaciones para la recuperación de la información, sino la necesidad de formalizar los registros en sus catálogos. Debido a ello, se creó el formato *USMARC para Códigos de Clasificación* en el que el registro de cada código contenía las notas de aplicación y las referencias a los números relacionados y obsoletos, así como los campos de encabezamientos por materia, descriptores y palabras clave para uniformar adecuadamente los registros.

Como es posible observar, los estudios realizados para valorar la funcionalidad de los sistemas documentales se apoyaron, principalmente, en la *DDC* y la *LC*, ya que fueron y siguen siendo los de mayor uso en Estados Unidos. Otro caso representativo, sobre este tipo de estudios, fue el proyecto *Dewey Retrieval System (DORS)*,⁸ creado en 1991, por la *Universidad de California* con el patrocinio de la *OCLC* y dirigido por los investigadores Songqiao Liu y Elaine Svenonius.

DORS surgió con la idea de generar un sistema prototipo basado en *DDC*, así como para probar el diseño y funcionalidad de su interfaz como catálogo en línea. Para el diseño de este proyecto se contempló, en primera instancia, analizar y determinar cuáles

⁸ Liu, S. y E. Svenonius "DORS: DDC Online Retrieval System", pp. 359-375.

eran las características que debería reunir un sistema de búsqueda y, a partir dicho análisis, se establecieron los objetivos siguientes para dirigir el proyecto: mejorar el vocabulario de uso en las formulaciones de búsqueda, facilitar la navegación entre la clasificación, realizar búsquedas por signatura topográfica, compatibilizar el sistema con otros sistemas de organización y facilitar el acceso por materias.

En la primera fase del proyecto, el vocabulario de entrada se extendió incorporando las tablas de clasificación de *DDC*; asimismo, se generó una opción para observar todas las jerarquías clasificatorias de dicho sistema, ya que se consideró que, cuanto más amplio fuera el campo de visualización, mejor entendimiento se podría obtener del entorno semántico del término solicitado, evitando cualquier desorientación. El *software* se desarrolló permitiendo que el usuario mantuviera, en todo momento, el control de la navegación y el acceso a la *DDC*, con la finalidad de contar con los elementos necesarios para evaluar la funcionalidad del sistema a partir de los resultados obtenidos en las búsquedas.

En la segunda fase del proyecto, se incorporaron al sistema los elementos siguientes:

- Una base de datos que contenía las tablas del grupo 700 (arte) de la *DDC* con 2,600 clases.
- Una base de datos con 2,992 registros bibliográficos proporcionada por la *OCLC*, que sólo contenía el título, la mención de responsabilidad y los encabezamientos de materia en *LC*.

Con base en los registros de las bases de datos, se extrajeron los términos más significativos para generar un índice, eliminando los términos no significativos mediante una lista de palabras

“vacías”, después se generaron cadenas que permitían “contextualizar” temáticamente los textos y establecer relaciones jerárquicas con el sistema documental. Durante esta fase Liu y Svenonius observaron que recuperaban una gran cantidad de documentos no deseados, por lo que atendieron el problema incorporando al programa las funciones siguientes: el *Scope* para ver las explicaciones o notas asociadas a cada clase; el *Browse* para tener acceso a las tablas de clasificación y observar los títulos de los documentos asociados a cada número de clasificación; y *Search* para permitir la navegación usando el índice en cadena, el número de clasificación y los encabezamientos de materia.

Los resultados obtenidos en *DORS* fueron interesantes, porque a partir del índice en forma de cadena se lograron ubicar temáticamente diversos términos, de acuerdo a las jerarquías del sistema documental. Además lograron que el usuario pudiera visualizar la secuencia completa de las clases del sistema donde se localizaba el término de búsqueda. No obstante, el sistema propuesto por *DORS* no fue muy bien aceptado.

A pesar de lo anterior, las observaciones e interpretaciones realizadas por Liu y Svenonius, durante el desarrollo de dicho proyecto, influyeron en subsecuentes estudios sobre las clasificaciones documentales en la *Web*. En 1995, Chan⁹ destacó algunas de esas observaciones para justificar las ventajas de los sistemas documentales, las cuales se resumen a continuación:

- Permiten visualizar, en pantalla, los registros temáticamente afines, independientemente de que éstos pertenezcan físicamente a otras colecciones o unidades de información.

⁹ Chan, Lois, “Classification, Present and Future”, pp. 5-17.

- Es posible ampliar o limitar una búsqueda, usando términos más genéricos o específicos, gracias a la capacidad de navegación implícita en la estructura jerárquica de un sistema documental, permitiendo, además, contextualizar los términos dentro de un campo semántico.
- Las nomenclaturas de clasificación, particularmente la numérica, pueden funcionar como un “*lenguaje-puente*” para salvar las barreras lingüísticas en amplias bases de datos y con registros en diferentes lenguas.

De acuerdo a estas argumentaciones, es posible subrayar que, la utilidad de las clasificaciones documentales proviene de su estructura jerárquica, sintética y relacional.

Durante los años noventas nuevos proyectos se iniciaron para valorar la utilidad de las clasificaciones en bases de datos disponibles en línea. De entre ellos, uno de los proyectos más destacados, *DESIRE*¹⁰ señala que la *DDC* representa la clasificación que ha logrado mayor proyección en la *Web*.

Sobre las ventajas de la *DDC*, Koch¹¹ rescata los aspectos siguientes:

- La popularidad de *DDC* sobre otras clasificaciones documentales contribuye a que resulte más familiar para los usuarios.
- La versión digital de la *DDC* la hace amigable para su implementación (aunque también existen ediciones digitales

¹⁰ Véase [Http://www.desire.org](http://www.desire.org)

¹¹ Koch, Traugott, “*Advantages and Disadvantages of Classification*”. Disponible en [Http://www.lub.lu.se/desire/radar/reports/D3.2.3/0](http://www.lub.lu.se/desire/radar/reports/D3.2.3/0)

de *CDU* y *LC*). La versión para *Windows* de *DDC* permite generar automáticamente cadenas de indización que se asocian fácilmente con los encabezamientos de materia y los códigos numéricos correspondientes.

En contraste, Koch destaca las desventajas siguientes:

- Habitualmente, la división del conocimiento no sólo responde a un punto de vista desfasado y subjetivo, susceptible de numerosas críticas, sino que, en muchas ocasiones, resulta ilógico, pues la prioridad de subordinar los temas a ciertas disciplinas, como primer criterio de división, fragmenta temas afines que en realidad corresponden clases diferentes.
- La rigidez de su estructura jerárquica, los hace poco aptos para incorporar nuevas áreas de conocimiento.

Para vencer estas dificultades, según Koch, una de las alternativas podría consistir en experimentar y rescatar las posibilidades de otros sistemas analítico-sintéticos como *Colon Classification* o *Bliss Bibliographic Classification*.

Por otro lado, es prudente recordar que el desarrollo del conocimiento se mantiene en un *continuum*, no es estático y, de ahí proviene el carácter provisional de las clasificaciones. Por tanto, las dificultades de las clasificaciones documentales sólo pueden ser atendidas a través de la actualización y la evaluación constante sobre la funcionalidad de las mismas.

En la *Web* existen algunos proyectos que, utilizando métodos de indización para agrupar diversas fuentes documentales temáticamente afines, en ciertas categorías semánticas y, utilizando métodos de cálculo de similitud entre las categorías propuestas y los encabezamientos de un sistema documental,

permiten fijar la clasificación y la nomenclatura de dichas fuentes. A continuación se ejemplifican brevemente algunos de esos proyectos:

NORDIC WAIS/WWW¹²

Financiado por *NORDINFO*, fue dirigido por la *University of Lund* y la *National Technological Library*, en Dinamarca, de 1993 a 1994. El propósito de este proyecto fue demostrar cómo un sistema de clasificación documental puede integrarse en la arquitectura *cliente/servidor* de la *Web* para rastrear bases de datos temáticas y clasificarlas. Asimismo, sirvió para definir el estado de arte de las bases de datos de las bibliotecas incorporadas en la *Web* y permitió experimentar con 700 bases de datos, utilizando la *CDU* como sistema clasificatorio.

El proyecto contempló el diseño de una base de datos denominada *WAIS*, la cual incluía una serie de ligas a diversas bases de datos (compiladas desde los protocolos *WWW*, *WAIS* y *Gopher*), un programa para la asignación automática de la clasificación y un árbol temático construido a partir de la *CDU*. La clasificación automática fue coordinada por Anders Ardö y Traugott Koch, quienes, con base en la descripción de las bases de datos (esto es, el título, el resumen, las palabras clave y los campos temáticos en los que venían definidas las propias bases de datos), extrajeron diversos términos y construyeron una lista de palabras clave.

En la medida en que la lista de palabras clave se iba contrastando con las tablas del sistema documental, se generó una lista de palabras “candidatas”, las cuales fueron revisadas una vez más, empleando un método basado en el valor acumulativo de las

¹² Véase [Http://www.lub.lu.se/W4/](http://www.lub.lu.se/W4/)

coincidencias obtenidas entre ambas listas, para definir la clasificación final. La dificultad primordial de este trabajo consistió en ajustar los valores de las clasificaciones para mejorar la calidad de la clasificación. Actualmente, la base de datos *WAIS* es de uso local, pero se espera que en un futuro sea de uso público.

GERHARD¹³

El proyecto *GERHARD* (*German Harvest Automates Retrieval and Directory*) fue dirigido por la biblioteca de la Universidad de Oldenburg y financiado por la *Deutsche Forschungsgemeinschaft* (Asociación Alemana de Investigación) en Zürich, Alemania. Su objetivo general consistió en crear un buscador que permitiera almacenar, automáticamente, los documentos de todos los servidores universitarios para después indizarlos y clasificarlos de manera automatizada. Dicho proyecto se llevó a cabo en dos fases: la primera de 1996 a 1998 y la segunda del 2001 al 2002. Los términos utilizados en su base de datos se encuentran en inglés, francés y alemán e incluye cerca de 60,000 entradas con diferentes relaciones entre ellas.

GERHARD opera con *Harvest*, que es un sistema muy conocido y reconocido entre las bibliotecas universitarias de Alemania. Este sistema permite establecer un modelo de búsqueda y recuperación distribuida a partir de dos componentes principales: *Harvest Gatherer* y *Harvest Broker*. *Harvest Gatherer* es la parte del programa responsable de la recuperación del texto completo y sus *metadatos*; regularmente visita varios sitios y construye una base de datos, que tiene la capacidad de incrementar sus registros, mientras que *Harvest Broker* es la parte del *software* responsable de entregar la información indizada.

¹³ Véase: [Http://gerhard.bis.uni-oldenburg.de/](http://gerhard.bis.uni-oldenburg.de/)

GERHARD opera de un modo similar al modelo de *Nordic WAIS/WWW*, pero con un método lingüístico más avanzado y con una tipología documental más heterogénea. Dicho método (coordinado por el *Institute of Computer Linguistics*, de la Universidad de *Osnabrück*) consiste en segmentar el contenido de los documentos de la *Web* en entidades, palabras y frases (esto es, genera un índice representativo de los documentos), para luego compararlos con un diccionario creado a partir de las definiciones de las notaciones *CDU*. El resultado es la propuesta de una clasificación para cada documento, que se pondera y ordena automáticamente, en función de la frecuencia de aparición de ciertas palabras. Las palabras más relevantes del documento se indizan junto a los códigos de clasificación, de tal forma que sea posible utilizar y visualizar en conjunto, el índice de términos de búsqueda y la estructura jerárquica de *CDU*.

El programa ofrece la posibilidad de integrar el índice y la estructura del Browsing para que el usuario pase de un documento a los resultados de búsqueda y a los documentos relacionados con la base de datos, de acuerdo al sistema documental. No obstante, en cuanto a la precisión temática de los artículos rastreados, *GERHARD* todavía muestra ciertas deficiencias. Actualmente sigue siendo una base de datos experimental de uso local.

SCORPIO¹⁴

Desde el 2003, *Scorpio* se constituyó como un proyecto de investigación, coordinado y financiado por la *OCLC*, cuyo propósito era demostrar las posibilidades y la funcionalidad de la catalogación automatizada a partir de la *DDC* y la *LC* en documentos digitalizados, a través de un programa que sería diseñado como herramienta de reconocimiento temático. Asimismo,

¹⁴ Véase: <http://purl.oclc.org/scorpion/>

en este proyecto, se consideró la descripción y asignación automática de *metadatos* de clasificación en *Dublin Core*.

Scorpio opera bajo un modelo de extracción, indización y correlación de términos con los sistemas de clasificación semejante a los proyectos anteriormente mencionados. Para ello, utiliza el *Electronic Support Systems (ESS)*¹⁵ de la *Electronic Dewey for Windows*. Este sistema sirve para comparar y explorar las relaciones entre las clases *DDC* y los encabezamientos de *LC*, considerando, especialmente, la naturaleza jerárquica de las relaciones conceptuales de *DDC*.

Durante el desarrollo de este proyecto, se observó que las relaciones entre los términos de *DDC*, los de *LC* y los términos documentales se establecían con ambigüedad, por lo que se diseñaron diferentes filtros para eliminar los registros más genéricos o específicos dentro de una clase determinada. Algunos de esos filtros utilizan dos algoritmos para ponderar y calcular la similitud entre los términos extraídos de los contenidos documentales y los términos utilizados por *DDC* y *LC*. El éxito de este programa ha sido tal, que actualmente se ofrece a través de la venta de licencias para la clasificación automática de documentos.

A partir de los trabajos mencionados, es posible señalar que existe una preocupación latente para desarrollar programas que permitan agilizar y facilitar los procesos de asignación de clasificaciones documentales en los recursos de la *Web*. Ello demuestra, por un lado, el interés que prevalece en las bibliotecas para rastrear y localizar información con un nivel más consistente en sus contenidos (bases de datos temáticas y documentos digitales) y para la organización de la misma. Por otra parte,

¹⁵ *ESS* es un programa que permite asignar encabezamientos de materia, en forma automática, a bases de datos documentales.

también demuestra la necesidad de generar sistemas de organización y recuperación, que permitan no sólo encauzar la información eficientemente hacia el usuario, sino brindar un contexto temático a la información seleccionada de la *Web*.

La viabilidad de los programas para la asignación automática de los sistemas documentales, como una herramienta de apoyo para la organización y el acceso a la información de los recursos documentales, puede ser demostrada a partir de los resultados obtenidos en los proyectos desarrollados en la *Web*. Tal como se muestra en la página principal de la *UCD Consortium*,¹⁶ en donde se mencionan los siguientes: *UCD Translation Bibliographical Survey* (2004), *Facet Analytical in Maning Knowledge Structure for Humanities*, *FAKTS* (2002-2003), *Music Information Resources Asisted Computer Library Exchange*, *MIRACLE* (1999-todavía en proceso), *Educators Access to Services in the Electronic Landscape*, *EASEL* (2000-2001), entre otros. Sin embargo, es necesario reconocer, que aún nos encontramos en una fase experimental o, tal vez, de transición, debido a que los resultados de dichos proyectos no son de uso público y por lo mismo no han logrado mayor presencia en la *Web*. Asimismo, cabe comentar que, a pesar de estos esfuerzos, existen diversos problemas pragmáticos para la asignación de la clasificación, que sólo podrían ser planteados desde perspectiva de las necesidades del usuario, el contexto, y desde un punto de vista teórico de la clasificación documental.

¹⁶ Véase: <http://www.Duch.orgresearch.htm>

CONCLUSIONES

Con base en el desarrollo de este trabajo, es posible señalar que la hipótesis establecida como eje de la presente investigación fue de utilidad para demostrar que, desde la perspectiva de la bibliotecología, los principios de selección, organización y evaluación con los que se administran los recursos documentales en la *Web* son funcionales, para atender los problemas derivados del acceso a la información, ya que a partir de estos principios es posible constituir núcleos informativos, temáticamente representativos, para cubrir cierto tipo de necesidades vinculadas con el quehacer cultural, educativo y de investigación documental.

En contraste, los buscadores de la *Web* han sido creados para rastrear, localizar y recuperar grandes cantidades de información, a través de la indización y la generación de taxonomías, pero, sin resolver los problemas derivados de la ambigüedad en el uso de los términos empleados, tales como la homonimia y la polisemia, así como los que surgen de la falta de un contexto que permita ubicarlos temáticamente. Debido a ello, los sistemas documentales ofrecen mayor precisión en los resultados de sus búsquedas, que los buscadores tradicionales, los cuales, además, funcionan para atender un tipo de información de uso más efímero.

Entre las tendencias para atender los problemas de organización y recuperación de la información en la *Web* es posible distinguir dos. Aquella que incorpora los sistemas documentales para clasificar de manera automática los recursos documentales (bases de datos y documentos digitales) de la *Web* y la llamada *Web Semántica* que, mediante el desarrollo de ontologías, pretende representar y recuperar cualquier tipo de información.

A partir de las formas como se proyectan ambas tendencias, es posible determinar que, mientras la biblioteca se encamina a trabajar con un sentido de compromiso y responsabilidad social sobre el uso y manejo de la información para contribuir a la difusión y desarrollo del conocimiento (mediante la investigación documental) la *Web Semántica* aborda los problemas derivados de la falta de acceso a la información, desde una perspectiva comercial, para tratar de garantizar el acceso de los distintos recursos informativos con sus posibles clientes o consumidores.

Por otra parte, existe cierta influencia de la bibliotecología hacia el método de indización empleado por los buscadores para rastrear la información, así como también los métodos empleados en la *Web* para crear ontologías podrían tener cierta influencia, dentro de nuestra disciplina, de forma recíproca, debido a su utilidad como herramientas de apoyo para la elaboración de tesauros. Mas, las limitaciones que ofrecen los esquemas comerciales, se deben, en general, a la falta de un trabajo selectivo, normativo y más consistente sobre los procesos de la información, actividades que son atendidas ampliamente por la biblioteca. No obstante, los esquemas propuestos por los buscadores, para rastrear la información en la *Web*, tienen mayor presencia que los catálogos de las bibliotecas. Ello obedece a distintas causas, pero, tal vez, la más preeminente sea la necesidad de tener acceso rápido a la gran cantidad de información que surge cotidianamente.

Para explicar las condiciones que han influido e influyen en las dinámicas y esquemas de acceso a la información en la *Web*, fue necesario reconocer el transitar de la biblioteca desde su adaptación a las nuevas tecnologías, en materia de organización y disposición de la información, así como ubicar los problemas vinculados con el acceso a la misma, desde los conceptos de

Sociedad de la Información, comercio electrónico, así como el contexto de interacción entre el individuo y la Web.

A partir del nuevo entorno tecnológico en el que se incorporó la biblioteca, comenzaron a replantearse los procesos para seleccionar, organizar y difundir la información. A pesar de ello, la biblioteca siguió manteniendo, de fondo, sus procesos de organización y tratamiento de la información documental. Bajo esas circunstancias incorporó, primero, sus catálogos generales en la Web, luego apoyó el desarrollo de bases de datos temáticas (*Subject Gateways*) hasta llegar a la fase experimental, en la que actualmente se muestran los proyectos vinculados con la clasificación automática de los recursos documentales de la Web.

En el contexto económico, se observó que en la Web existe una intención comercial para fomentar el uso y consumo de los recursos y servicios informativos. Dentro de este marco, las necesidades de los usuarios son atendidas en función de los intereses de las propias empresas. Mientras que la información es vista como una mercancía, cuyo valor es esencialmente comercial y, como tal, se encuentra sometida a las leyes de la oferta y la demanda del mercado y no a otro tipo de criterios, más éticos, de responsabilidad social.

Dentro del concepto de *Sociedad de la Información*, se destacó que no fomenta el desarrollo de una sociedad libre e igualitaria, como se ha manifestado en la mayoría de los discursos, ya que, actualmente, existe una fragmentación social determinada por las limitaciones económicas de ciertos sectores de la población para disponer de la tecnología necesaria que les permita tener acceso a la información desde la Web (véase el concepto de *brecha digital*). Asimismo, el manejo de la información en la Web dejó al descubierto otro tipo de dificultades, como las relacionadas con la

manipulación de los medios sobre las notas informativas y la lengua, pues la mayor parte de la información se presenta en inglés, con una notoria presencia de las instituciones públicas, privadas y organizaciones no gubernamentales de Estados Unidos.

En el contexto del individuo, se señaló que la influencia informativa de la *Web*, ha contribuido a disolver los referentes sociales que forman parte de la identidad de una persona, para luego reconstruirlos en función de repertorios culturales, sociales y comerciales ajenos a sus propias costumbres, y que esta relación se mantiene debido a las prácticas publicitarias y al deseo individual de no quedar ajeno a la modernidad. También se observó que el uso de la información se encuentra subordinado a las necesidades del individuo, y que éstas se ven reflejas de acuerdo al nivel formativo y al entorno en el que se desenvuelve el mismo.

Para distinguir las necesidades informativas, se hizo necesario reconocer las diferencias que existen entre el saber y el conocer. Ambos tipos de necesidades pueden ser atendidas ampliamente a través de las prácticas bibliotecarias, aunque no sucede lo mismo con los buscadores comerciales, pues ofrecen ciertas limitaciones para tener acceso a las fuentes documentales. Bajo esas circunstancias, es posible inferir que la funcionalidad de los buscadores se acerca más al hecho de promover el uso de un tipo de información utilitaria y efímera, vinculada con la necesidad cotidiana del saber, mientras que las bases de datos temáticas y los catálogos de las bibliotecas, por su parte, promueven un tipo de información más consistente en apoyo al desarrollo del conocimiento del individuo.

A pesar de las ventajas de los sistemas documentales, anteriormente señaladas, es necesario subrayar que los procesos

de revisión y evaluación sobre la operatividad y funcionalidad de dichos sistemas, representan parte del compromiso y la labor que los especialistas de la información deberán considerar para mantener la adecuada integración de sus sistemas.

Sin embargo, existe una preocupación latente generada por la gran variedad de formatos que han aparecido en la *Web* y por el desarrollo de modelos que permitan describir dichos recursos. Debido a ello, ha subsistido la inquietud de crear esquemas de estandarización que, a partir del uso de *metadatos*, faciliten su descripción para lograr cierta normalización y control sobre dicha producción; tal es el caso de modelos como el *Dublin Core*, *TEI* y *RDF*, entre otros, pero, el proceso de evolución de dichos estándares, ha sido lento, en comparación con los requerimientos que plantea el nuevo entorno de la información.

Aunado a lo anterior, la falta de una masa crítica sobre las aplicaciones de dichos estándares dificulta la tarea de evaluar su funcionalidad. En ese sentido, resulta fundamental el papel que juegan las instituciones como la *ISO* para asimilar y poner en marcha dichos modelos.

Finalmente, si consideramos que la información (en cualquiera de sus formatos de disposición) es ante todo un producto social (ya que la construimos todos), entonces, como tal, la labor de organizarla para su difusión y recuperación, sin realizar algún tipo de distinción, es parte del compromiso y la responsabilidad social que, los bibliotecarios, como especialistas de la información, debemos seguir manteniendo.

Anexo 1

Modelo Dublin Core

Elemento DC	Descripción y uso
Title	Nombre dado a un documento, usualmente por el autor.
Subject	Materias del recurso. Descriptores o frases que definen el contenido del <i>DLO</i> . Fomenta el uso de vocabularios controlados y sistemas de clasificación documental.
Description	Incluye el resumen en el caso de un documento textual o una descripción del contenido en el caso de un documento visual.
Source	Secuencia de caracteres utilizada para identificar unívocamente un trabajo a partir del cual proviene el recurso actual.
Language	Idioma del contenido del <i>DLO</i> .
Relation	Identificador de un recurso y de la relación que tiene con el objeto de información que se está describiendo. Por ejemplo, si el recurso representa una versión, una parte o está basado en algún otro tipo de documento.
Coverage	Cobertura especial y/o temporal del contenido del <i>DLO</i> . La cobertura espacial se refiere a una región física, preferiblemente extraída de una lista controlada. La cobertura temporal no se refiere a la fecha de creación o publicación del registro, sino a la temporalidad a la que alude el contenido.
Creator	Persona física o jurídica responsable de la creación del contenido del recurso.
Publisher	Entidad responsable de hacer que el recurso se encuentre en la Red (editorial, universidad o algún otro organismo).
Contributor	Persona u organización que haya tenido una colaboración o participación en la creación o contenido del recurso.
Rights	Referencia (por ejemplo, <i>URL</i>) sobre derechos de autor o enlace a un servicio que dará información sobre los términos y condiciones de acceso al recurso.
Date	Fecha en la que el recurso se puso a disposición de los usuarios en su forma electrónica actual.
Type	Categoría en cuanto al tipo de información que representa el <i>DLO</i> (Ejemplos: imagen, texto, software, etcétera)
Format	Indica el <i>software</i> y posiblemente también el <i>hardware</i> que se requiere para mostrar el documento.
Identifier	Secuencia de caracteres usados para identificar unívocamente un recurso. Para el caso de los recursos en línea, éste puede ser el <i>URL</i> y, para otros, puede ser el ISBN, ISSN, etcétera.

Anexo 2

Clasificaciones Documentales Relevantes

Clasificación Decimal Dewey (DDC)

Melvil Dewey formuló, en 1873, su modelo de clasificación decimal cuando tenía 22 años y lo realizó para el *Amherst College*, donde trabajaba. Este sistema fue diseñado con una visión pragmática, para atender las necesidades de un tipo determinado de biblioteca que pretendía ofrecer el acceso directo de los libros a todos sus usuarios. Sin embargo, la propuesta de Dewey no era nueva, tuvo sus antecedentes en el esquema desarrollado por William Torrey Harris y el catálogo de Natale Battezzati el “*Nuovo Sistema di Catalogo Bibliografico Generale*”. Harris tenía una formación filosófica y tomó, como base de su sistema, la clasificación propuesta por Francis Bacon, quien había subdividido la ciencia en las *tres facultades del espíritu*: la historia de la memoria, la poesía de la imaginación y la filosofía de la razón. Harris retomó esta clasificación y la estableció de la siguiente manera: ciencia, arte e historia, les añadió nuevas subdivisiones e integró una notación numérica al sistema. Pero el problema de estas clasificaciones era que no permitían la inserción de nuevas materias, ni tampoco una división apropiada de las mismas.

A pesar de lo anterior, Dewey estableció, por ejemplo, que el número asignado a un registro ya no tendría que implicar el desplazamiento de los libros en los estantes, sino que únicamente respondería a la relación de materias entre sí. Los números asignados a una disciplina serían decimales, entonces, cada nueva cifra añadida sería una subdivisión de la anterior, lo que permitiría incorporar un sinfín de subdivisiones sin alterar el orden previamente establecido; asimismo, el esquema y las tablas se

acompañarían de un índice, al que Dewey llamó relativo, en el cual cada término podría ser asociado con distintas disciplinas.

Dewey estableció una clasificación en el siglo XIX que, evidentemente, y, con el paso del tiempo, caducó, razón por la cual ha sido modificada, pues además, la distribución del conocimiento que plantean las disciplinas ha generado nuevas necesidades. En general, es posible mencionar que la *CDD* ha vivido tres periodos diferentes: el primero, durante la vida de Dewey, en el que los signos asignados a una materia no se podían alterar para evitar que la biblioteca desplazara físicamente sus acervos, el segundo, que va de la publicación de las ediciones 14 a la 17, en las que se realizaron diversas modificaciones sobre todas las materias, sin considerar sus resultados y, el tercero, en el que ya es posible remodelar una materia completa, dejando las demás intactas.

Debido a lo anterior, es posible comentar que, si bien en un principio el sistema *CDD* era exclusivamente enumerativo y expansivo, esto es, cada materia tenía un lugar en el esquema, al que podían subordinarse innumerables subdivisiones, con el tiempo y, gracias a la influencia de sistemas como el de Ranganathan y la *CDU* y, en general, debido a la influencia sobre la teoría de la clasificación, logró mantener un desarrollo facetado, esto es, nuevas materias comenzaron a crearse, a partir de la combinación de distintas notaciones o con la añadidura de auxiliares.

La *CDD* fue incorporada a una gran mayoría de las bibliotecas en Estados Unidos y después comenzaron a sucederle numerosas ediciones que impulsaron su desarrollo. Pese a las múltiples ediciones del sistema de Dewey, dos hechos determinaron su permanencia: primero, la adopción de este sistema en la Biblioteca del Congreso de Estados Unidos, para la elaboración de sus catálogos impresos y, segundo, su adopción en la *British National Bibliography* para la organización de sus acervos, lo que

permitió que penetrara en el ámbito nacional del Reino Unido, como modelo de control bibliográfico.

Para Rider,¹ la utilidad del *DDC*, se debe al hecho de haberse establecido como un esquema de clasificación temática, que permite la progresiva subordinación de clases hasta el infinito y a la incorporación de un índice relativo a las tablas. Mas, el alcance principal de este sistema fue obtenido por la difusión que lograron Paul Otlet y Henry La Fontaine, quienes además desarrollaron ampliamente las tablas y constituyeron un nuevo sistema clasificatorio, el *Decimal Universal*.

Clasificación Decimal Universal (CDU)

En 1895, Otlet y La Fontaine pidieron permiso a *Dewey* para traducir, ampliar y usar su clasificación con el propósito de crear un *Repertorio Bibliográfico Universal*. Durante ese mismo año, organizaron en Bruselas la “*Conferencia Internacional de Bibliografía*”, donde dieron a conocer los principios de dicho repertorio. A partir de entonces y bajo este novedoso proyecto, Otlet y La Fontaine crearon y promovieron la *Clasificación Decimal Universal*. Dicha clasificación apareció en 1905, como “*Manuel du Repertoire Bibliographique Universel*”.

La *CDU* contiene una notación numérica que estructura el conocimiento en diez divisiones y subdivisiones de diez en diez, en grupos temáticos infinitos. La estructura de la *CDU* es jerárquica, ya que sigue un orden que parte de lo general a lo particular, en donde las materias encuentran, si su actualización lo permite, un desarrollo decimal y jerárquico, en la siguiente secuencia: punto de vista, lugar, tiempo, forma,

¹ Rider, A. Fremont, “The Story of DC 1876-1951”, citado por San Segundo Manuel, Rosa, *op. cit.*, p. 79.

idioma y demás auxiliares. Asimismo, la *CDU* contiene un índice alfabético que es de utilidad cuando no se domina una materia.

Por otra parte, el *Instituto Internacional de Bibliografía (IIB)* realizó una extensa labor, para la difusión de la *CDU* en el ámbito internacional. En 1921, el *IIB* cambió su nombre a *Instituto Internacional de Documentación* y, a partir de 1938, se constituyó como *Federación Internacional de Documentación (FID)*. Este organismo ha venido ocupándose de la revisión, actualización y publicación de la *CDU* en sus distintas ediciones y todavía sigue siendo utilizado por varias organizaciones internacionales, entre ellas, la UNESCO que la emplea para sus bibliotecas y sus repertorios bibliográficos.

Clasificación la Biblioteca del Congreso (LC)

La *Biblioteca del Congreso* de Washington, Estados Unidos, fue establecida en 1800 y, dentro de sus orígenes, contó con dos clasificaciones: la primera fue publicada en su catálogo de 1812 y contaba con dieciocho grupos temáticos; la segunda fue incorporada posteriormente y correspondía a la *Biblioteca de Thomas Jefferson*, cuya catalogación y clasificación fueron diseñadas por el propio *Jefferson*, quien en 1815 publicó su catálogo organizado en cuarenta y cuatro grupos temáticos, bajo el título “*Catalogue of the Library of the United States*”. Este esquema sobrevivió hasta 1899, cuando el director de la biblioteca, Herbert Putman, decidió reclasificar la colección completa.

En aquel tiempo, existían dos clasificaciones de amplio reconocimiento: la *Clasificación de Dewey* y la *Clasificación Expansiva de Cutter*. Entonces, Putman resolvió considerar ambos esquemas para formular un nuevo sistema, declinando por los principios de Cutter. El nuevo sistema consistía en veintiún divisiones por materia, representadas

con una letra mayúscula del alfabeto latino. Cada grupo temático es independiente, lo que ha permitido que cada letra se halla publicado por separado.

En LC, cada materia se subdivide, a su vez, en materias expresadas por letras mayúsculas y números, sin valor decimal y, cada clase está provista de un índice alfabético independiente. Este índice y sus entradas incluyen nombres personales, geográficos y además hace uso de referencias. Este sistema permite una gran actualización, debido a que en realidad está compuesto por un compendio de clasificaciones individuales en cada una de sus letras, lo que le confiere el atributo de ser sistemática, elástica y expansiva. Mas, entre sus desventajas, es necesario reconocer que su empleo en otros países implica un desfase cultural, porque se encuentra impregnado de valores, conceptos y tópicos propios de la cultura norteamericana.

Clasificación Bibliográfica de Bliss (BC)

Una gran parte de la labor del norteamericano Henry Evelyn Bliss se enfocó en tratar de corregir los errores de la clasificación de Dewey y de la LC, incluso. El sistema de Bliss fue implantado en la *Biblioteca del College* en Estados Unidos, en 1902, aunque dicho esquema comenzó a publicarse a partir de 1910 y fue hasta 1953 cuando se concluyó la publicación de la edición completa, en cuatro volúmenes, a los que dedicó prácticamente toda su vida.

El sistema basa su organización en clases que intentan representar la “*totalidad del conocimiento*”. Las divisiones de las distintas disciplinas incluidas son lógicas y conceptuales. Y las diferencias entre ellas se basan en los distintos grados y transiciones de las formas del saber. Es decir, la división de cada disciplina se define en función de cuatro puntos de vista: filosófico, teórico, histórico y práctico. Asimismo, intenta en todo momento

conciliar el enfoque científico con el bibliotecario, pero considerando en todo momento las diferencias entre la clasificación del conocimiento y la documental.

El esquema de la Bliss se estableció a partir de dos visiones: una específica y otra general, de tal modo que intentó organizar el conocimiento en relaciones de subordinación (vertical) y de coordinación (horizontal). Asimismo, desarrolló una división del conocimiento en veintiocho clases principales, señaladas con una notación alfabética, las cuales pueden subdividirse en grados de subordinación y especialidad por cada una de las materias.

Bliss consideraba que la importancia de su sistema se basaba en la aplicación de cinco principios básicos: el primero es el *consenso*, en el que se indica que la unanimidad sobre la identificación de las clases principales debería ser determinada considerando la opinión de un grupo de científicos; el segundo es el de la *colocación cuidada*, que implica ordenar las clases temáticas de acuerdo a un criterio de proximidad; el tercero es el de *subordinación*, que hace referencia a la distribución de las especialidades por jerarquías; el cuarto es el de *localización alternativa*, que supone dotar de flexibilidad a las tablas de clasificación y permite colocar un tema en dos sitios y el quinto, es el de la *notación breve*, que hace referencia a que la notación debe ser sintética, expresiva y sistemática.

El sistema de *Bliss* logró ser tan reconocido como el sistema de *Dewey* y el *LC*, por lo que fue objeto de numerosas reimpressiones. También, tuvo cierta incidencia sobre el trabajo de Ranganathan y el mundo anglosajón debido a la aceptabilidad de los postulados teóricos y prácticos incluidos en su obra.

Clasificación Colonada de Ranganathan (CC)

Shiyan Ramarita Ranganathan fue el creador de una de las clasificaciones más destacadas de nuestro tiempo. Ranganathan trabajó como profesor de matemáticas en la Universidad de *Madrás* y, más tarde, fue bibliotecario en la misma Universidad para emprender dicho trabajo, fue enviado a estudiar a la *British Museum* y a la *Library School* de la *University of London* y, de regreso a *Madrás*, se dedicó a construir su sistema clasificatorio. La *CC* fue publicada en 1933 y, posteriormente, Ranganathan realizó seis ediciones más para complementar su obra. Aunque la séptima edición (publicada en 1971) sólo logró dirigirla parcialmente debido a que le sobrevino la muerte.

Resulta difícil describir la *CC* debido a su complejidad y amplitud. Sin embargo, es posible mencionar que se desarrolló bajo la exposición básica de que un tema debe mostrarse en correlación con una determinada escuela del pensamiento ya sea filosófica, política, científica, religiosa, etcétera. La estructura del sistema parte de una tabla de materias principales donde se incluyen las disciplinas, cada una de las cuales pueden subdividirse en facetas. La notación comprende veintiséis letras del alfabeto latino, diez números arábigos del 0 al 9 y diferentes letras del alfabeto griego. Este esquema se distingue porque incluye distintos signos de puntuación que mantienen una función de correlación, el más relevante es el *colon* (vocablo en inglés que designa dos puntos), del que proviene su nombre.

La tabla de las clases generales se encuentra definida por una combinación de categorías que constituyen la *personalidad*, *energía*, *espacio* y *tiempo* del saber, aunque estas concepciones son difíciles de comprender, en el mundo occidental, es posible ejemplificar cada una de ellas: la *personalidad*, indica la clave de las cosas. Así, en la astronomía, la *personalidad* sería determinada por el objeto de estudio, como la Tierra, la

Luna, el Sol, etcétera; la *materia* refleja el material constituyente de un objeto, p.e., en la manufactura de papel la fibra sería la *materia*; la *energía* es la acción, el procedimiento y la operatividad de las cosas; de este modo, en la mecánica, la *energía* podría ser el movimiento o la vibración; el *espacio* supone sencillamente una distribución territorial, y el *tiempo* permite ubicar la temporalidad de un objeto.

En general, se trata de un sistema analítico, porque permite atribuir distintas categorías a un documento y, es sintético, porque individualiza la representación de dicho documento, asignándole una notación y una clase específica con base en el contenido del mismo. Esta metodología analítica-sintética proporciona gran hospitalidad al sistema para incluir nuevos temas y permitir mayor independencia al catalogador. Debido a lo anterior, la técnica taxonómica de Ranganathan resulta ser una gran innovación en las clasificaciones documentales, pues rompe con los esquemas tradicionales de clasificación bibliográfica.

Sin embargo, la CC puede ser considerada más como un amplio ejercicio teórico de clasificación, que como un sistema meramente documental, ya que ha enriquecido la teoría de la clasificación y ha impulsado los estudios sobre ésta.

La terminología empleada en el sistema CC es original, precisa y única, aunque no ha tenido gran aceptación en las bibliotecas del ámbito europeo y americano, por razones culturales, sociales y políticas de diversa índole y, a causa de la compleja estructura del sistema.

BIBLIOGRAFÍA

Actas del III Encuentro de ISKO-España del 19 al 21 de noviembre de 1997: Organización del Conocimiento y Sistemas de Información y Documentación, Javier García Marco, Ed., Zaragoza, ISKO, 1999, 420 pp.

Aguillo, Isidro F. (2000), "Internet invisible o infranet: definición, clasificación y evaluación", *Gestión del conocimiento: retos y soluciones de los profesionales de la información. VII Jornadas Españolas de Documentación*, Bilbao, 19-29 de octubre, pp. 249-270.

Barlow, John Perry (1994), "A Taxonomy of Information", *Bulletin of American Society for Information Science*, June-July, pp. 13-17.

Barners-Lee, T, J. Hendler y O. Lassila (2001), "The Semantic Web", *Scientific American*, vol. 284 (5), mayo, pp. 34-43.

Barrer, Joe (2001), "Una aproximación a Internet y a sus herramientas de búsqueda", en EDUTEKA, *Tecnologías de la información y comunicaciones para enseñanza básica y media*, Biblioteca de la Universidad de California, Berkeley. Disponible en: <<http://www.eduteka.org/pdfdir/BuscadoresBasico.pdf>> (Revisado el 06 de enero del 2006).

Bell, Daniel (1976), *El advenimiento de la sociedad post-industrial: un intento de prognósis social*, Madrid, Alianza Editorial, 578 pp.

Castells, Pablo (2005), *La Web Semántica*. Disponible en: <<http://www.ii.uam.es/~castells/publications/castells-uclm03.pdf>> (Revisado el 06 de enero del 2006).

Cañón Vega, Nora (2002), *Sociología de la información: el rol de la información en la sociedad de hoy*, Colombia, Universidad de la Salle, 115 pp.

Codina, Lluís (2003), "Internet invisible y Web Semántica. ¿El futuro de los sistemas de información en línea?", *Tradumática: Traducció i Technologies de la Informació i la Comunicació*, vol. 2, noviembre. Disponible en: <<http://www.fti.uab.es/tradumatica/revista>> y <<http://www.bib.uab.es/pub/tradumatica/15787559n2a6.pdf>> (Revisado el 06 de enero del 2006).

Chomsky, Noam (1997), *Secretos, mentiras y democracia*, entrevista realizada por David Barsamian, trad. de Bertha Ruiz de la Concha, México, Siglo XXI Editores, 150 pp.

Dahlberg, Ingetraut (1993), "Why Knowledge Organization? Reasons for International Classification's Change of Name", *International Classification*, vol. 20, p. 1.

“Declaración de principios: construir la Sociedad de la Información, un desafío para el nuevo milenio”, documento presentado en la *Cumbre Mundial sobre la Sociedad de la Información*, Ginebra, 2003, Túnez 2004, el 12 de mayo del 2004. Disponible en Internet desde <<http://www.itu.int/wsis/docs/geneva/official/dop-es.html>> (Revisado el 06 de enero del 2006).

Ding, Y. y Foo, S. (2002), “Ontology Research and Development. Part 1, A Review of Ontology Generation”, *Journal of Information Science*, vol. 28 (2), pp. 123-136.

Estivill, Assumpció y Ernests Abadal (2000), “Acceso a los recursos Web gratuitos desde las bibliotecas”, *El Profesional de la Información*, vol. 9 (11), noviembre, 20 pp. Disponible en: <<http://www.swets.nl/sps/journals/iwe.html>> (Revisado el 06 de enero del 2006).

Figueroa Servin, Reynaldo Demetrio (2002), *Los principios de estructuras clasificatorias en un catálogo sistemático en el entorno hipertextual*, México, tesis de maestría en Biblioteconomía, UNAM, Facultad de Filosofía y Letras, 110 pp.

García Jiménez, Antonio (2004), “Instrumentos de representación del conocimiento: tesauros versus ontologías” en *Anales de Documentación*, núm. 7, pp. 79-95. Disponible en: <<http://www.um.es/fccd/anales/ad07/ad0706.pdf>> Revisado el 06 de enero del 2006).

Galdón, Gabriel (2001), *Introducción a la computación y a la información*, Barcelona, Ariel, 325 pp.

Giddens, Anthony (2000), *Modernidad e identidad del yo: el yo y la sociedad en la época contemporánea*, 3ª. ed., Barcelona, Península, 299 pp.

Goñi Camejo, Ivis (2000), “Algunas reflexiones sobre el concepto de información y sus implicaciones para el desarrollo de las ciencias de la información”, *ACIMED*, vol. 8 (3). Disponible en <http://www.bvs.sld.cu/revistas/aci/vol8_3_00/aci05300.htm> (Revisado el 06 de enero del 2006).

Gruber, T. R. (1993), “A Translation Approach to Portable Ontology Specifications”, *Knowledge Acquisition*, vol. 5 (2), pp. 199-220.

Kirriemuir, John, Dan Brickley, Susan Welsh *et al.* (1998), *Cross-Searching Subject Gateways: The Query Routing and Forward Knowledge Approach*. Disponible en: <<http://www.dlib.org/dlib/january98/01kirriemuir.html>> (Revisado el 06 de enero del 2006).

Koch, Traugott (1996), “Internet Search Service”, en DESIRE, *Development of a European Service for Information on Research and Education*, EU Project. Disponible en: <<http://www.lub.lu.se/tk/demos/DO9603-meng.html>> (Revisado el 06 de enero del 2006).

Krüger, Karsten y María Caprile (2005), “Evaluación de portales electrónicos de ciencias sociales: una propuesta conceptual”, *Revista Electrónica de Geografía y Ciencias Sociales*, vol. X (565), febrero, 21 pp. Disponible en: <<http://www.ub.es/geocrit/b3w-565.htm>> (Revisado el 06 de enero del 2006).

Kwasnik, B, (1999), “The Role of Classification in Knowledge Representation and Discovery”, *Library Trends*, vol. 48, pp. 22-47.

Lafuente López, Ramiro (1990), *De la naturaleza de los sistemas de clasificación, su aplicación en las ciencias sociales*, México, tesis de maestría en Bibliotecología, UNAM, Facultad de Filosofía y Letras, 204 pp.

--- (1992), “La clasificación como objeto de estudio en bibliotecología”, *Edición Conmemorativa del X Aniversario del Centro Universitario de Investigaciones Bibliotecológicas*, México, UNAM, Centro Universitario de Investigaciones Bibliotecológicas, pp. 43-53.

Lim, Edward (1999), “Pasarelas temáticas del sudeste asiático: análisis de sus métodos de clasificación”, *65th IFLA Council and General Conference Bangkok*, Thailand, 20-28 de agosto de 1999, 10 pp.

López Alonso, Miguel Ángel y Juan Mares Martín (1998), “La organización del conocimiento contenido en la información hipertextual de internet”, *Actas de la VI Jornadas Españolas de Documentación: Los Sistemas de Información al Servicio de la Sociedad*, FESABID, Valencia, 29-31 de octubre, 26 pp. Disponible en: <http://fesabid98.florida-uni.es/Comunicaciones/ma_lopez.htm> (Revisado el 06 de enero del 2006).

Markey, Karen (1986), “Class Number Searching in an Experimental Online Catalog”, *International Classification*, vol. 13 (3), pp. 142-150.

Maldonado Martínez, Ángeles y Elena Fernández Sánchez (1998), “Evaluación de los principales “buscadores” desde un punto de vista documental: recogida, análisis y recuperación de recursos de información”, *Actas de la VI Jornadas Españolas de Documentación: Los Sistemas de Información al Servicio de la Sociedad*, FESABID, Valencia, 29-31 de octubre, 26 pp. Disponible en <<http://www.fesabid.org/servicios/publicaciones/valencia1998.htm>> (Revisado el 06 de enero del 2006).

Marcella, Rita y Arthur Maltby (2000), *The Future of Classification*, London, Gower Publishing, 144 pp.

Méndez Rodríguez, Eva (2002), *Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales*, España, Trea, 429 pp.

--- (2002), *El documento como objeto de información: metadatos y esquemas de organización de la información en la Web*, pp. 49-62. Disponible en:

<<http://www.ect.juntaex.es/dgsi/Documentacion/Eva.pdf>> (Revisado el 06 de enero del 2006).

Montesinos, Antonio (1999), *La Sociedad de la Información e Internet: fundamentos, aptitudes y uso de la red*, Madrid, San Pablo, 183 pp.

Montiel, Edgar (2002), *El nuevo orden simbólico: la diversidad cultural ante la globalización*, Madrid, Fundación Mapfre Tavera y Secretaría de Cooperación Iberoamericana, 83 pp.

Neches, R. *et al.* (1991), “Enabling Technology for Knowledge Sharing”, *Al Magazine*, vol. 12 (3), pp. 36-56.

Páez Urdaneta, Iraset (1992), *Gestión de la inteligencia, aprendizaje tecnológico y modernización del trabajo. Retos y oportunidades*, Caracas, Consejo Nacional de Investigación Científica y Tecnológica, Instituto de Estudios del Conocimiento, Universidad Simón Bolívar, 253 pp.

Pastor, J. A. y Saorín T. (1997-8), “La escritura hipermedia”, *Cuadernos de Documentación Hipermedia*, núms. 6-7, pp. 221-238. Disponible en: <<http://www.ucm.es/info/multidoc/multidoc/revista/cuad6-7/saorin.htm>> (Revisado el 6 de enero del 2006).

Qin, J., Pailing, S. (2001), “Converting and Controlled Vocabulary into an Ontology: the Case of GAME”, *Information Research*, vol. 6 (2), enero. Disponible en: <<http://informationr.net/ir/6-2/paper94.html>> (Revisado el 6 de enero del 2006).

Ramonet, Ignacio (1999), *El periodismo del nuevo siglo*, trad. de Mirnaya Chabás, publicado originalmente en la revista *La Factoría*. Disponible en: <<http://www.lafactoriaweb.com/default-2.htm>> (Revisado el 06 de enero del 2006).

Ríos Hilario, Ana Belén (2003), *Nuevos horizontes en el análisis de los registros y la normativa bibliográfica*, España, Ediciones Trea, 2003, 165 pp.

Rodríguez Yunta Luis (2001), “Bases de datos documentales: estructura y uso”, en *La información especializada en Internet*, coord. Ángeles Maldonado. Disponible en: <<http://www.unav.es/dpp/documentacion/proteger/lryunta.pdf>> (Revisado el 06 de enero del 2006).

Roszkowski, M. y Lukas Christopher (1998), “A Distributed Architecture for Resource Discovery Using Metadata”, en *D Lib Magazine*, vol. s/n, junio. Disponible en: <<http://www.dlib.org/dlib/june98/scout/06roszkowski.html>> (Revisado el 06 de enero del 2006).

Rusell, Jane M. (2001), *La comunicación científica a comienzos del siglo XXI*. Disponible en: <<http://www.campus-oei.org/salactsi/rusell.pdf>> (Revisado el 06 de enero del 2006).

Russell Green, Andrew (2003), “Cambios actuales en el esquema de supuestos básicos de las prácticas catalográficas”, ponencia presentada en el *Seminario de Estudios sobre lo Visual en la Investigación Social*, Instituto Mora, 1 de septiembre. Disponible en: <<http://creativecommons.org/licenses/by-nc-sa/1.0/>> (Revisado el 06 de enero del 2006).

Saramago, José, Kenzaburo Oé, Armad Mattelart *et al.* (2002), *El mito Iinternet: doce artículos publicados*, en *Le Monde Diplomatique*, Chile, Aún Creemos en los Sueños, 90 pp.

San Segundo Manuel, Rosa (1996), *Sistemas de organización del conocimiento: la organización del conocimiento en las bibliotecas españolas*, Madrid, Universidad Carlos III de Madrid y Boletín Oficial del Estado, 317 pp.

Schwartz, Candy (2001), *Sorting Out the Web Approaches to Subject Access*, Connecticut, Ablex Publishing, 169 pp.

Schiller, Herbert (1984), *Información y economía en tiempos de crisis*, trad. de Idelfonso Pérez Rodríguez, Madrid, FUNDESCO, 143 pp.

Silva Muñoz, Lydia (2002), *Representación de ontologías en la Web Semántica*, 10 pp. Disponible en: <www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo_lydia.pdf> (Revisado el 6 de enero del 2006).

Tramillas Saz, Jesús (2001), *Recuperación de la información en Internet*, Madrid, RAMA Editorial, 232 pp.

Tovar Ramírez, Aurora (2005), “Internet 2: la investigación de la UNAM mediante el cómputo intensivo compartido”, *Acta Sociológica*, núm. 43, enero-abril, pp. 59-75.

Velasco, Roberto (2003), *La economía digital: del mito a la realidad*, Barcelona, Tusquets, 232 pp.

Villoro, Luis (2002), *Creer, saber, conocer*, 15ª ed., México, Siglo XXI Editores, 310 pp.

Volkow, Natalia (2003), “La brecha digital, un concepto social con cuatro dimensiones”, *Boletín de Política Informática*, núm. 6, México, INEGI, pp. 1-5. Disponible en: <<http://www.inegi.gob.mx/inegi/contenidos/espanol/prensa/contenidos/articulos/tecnologia/brecha.pdf>> (Revisado el 06 de enero del 2006).