

SOCIEDAD DE LA INFORMACIÓN: METADATOS Y FUTURO DE LA INTERNET EN LA RECUPERACIÓN DE INFORMACIÓN DE CALIDAD

Wilmer Arturo Moyano Grimaldo
Profesional en Ciencia de la Información – Bibliotecólogo
Jefe Area de Análisis de Información
Biblioteca Universidad de la Sabana
wilmer.moyano@unisabana.edu.co

1. Introducción

La sociedad de la información es un hecho desde hace ya unas cuantas décadas y aunque autores como Negroponte consideren que ya vamos en una etapa postinformacional e incluso otros ya la consideran la Sociedad del Conocimiento, lo cierto es que estamos inmersos en una explosión de información que ya alcanza, en algunos casos, el tratamiento de infoxicación. Este cambio social que se viene dando desde hace ya algún tiempo, se ha visto potenciado por la aparición y especialización de nuevas tecnologías de la información, que para nuestra profesión significan retos y desafíos que no se pueden ignorar, donde quizás la que mas nos ha cambiado el paradigma de organización del conocimiento ha sido la Internet, especialmente la WWW (World Wide Web). Una de estas herramientas son los llamados metadatos los cuales merecen gran consideración por el avance que han logrado en tan solo pocos años y por brindar soluciones sencillas y muy completas para el caos que existe dentro de la organización de información digital.

El presente artículo, tiene como objetivo hacer un recorrido que va desde el concepto de sociedad de la información hasta la tendencia propuesta por la W3C en cabeza de su director, el cual propone la creación de una Web Semántica; que potencie las funcionalidades de la ya existente a favor de las necesidades del hombre, esto con ayuda de herramientas como los metadatos, el XML (eXtensible Markup Language) o el RDF (Resource Description Framework).

2. Sociedad de la Información

Para nadie es un secreto que la sociedad actual ha sido denominada de la Información, esto definida a raíz de su fenómeno mas destacable: la *Información*, de la misma forma en que la industrialización definió la sociedad Industrial durante la mayor parte del siglo XIX y las primeras décadas del XX. Este aumento en el consumo de la información tiene su inicio a partir del gran desarrollo urbano el cual complicó los aspectos sociales de la población, lo que ocasionó que la sociedad necesitara conseguir “información” para suplir sus necesidades y mejorar su situación. Esto no quiere decir que es la primera vez que la información se usa dentro del desarrollo social de la sociedad, puesto que si se mira con un poco mas de detenimiento ya se utilizaba incluso en la edad media; no obstante si es la primera vez que se globaliza y se convierte en fuente fundamental de la productividad y el poder.

El aumento exponencial en el consumo de Información, se ha venido dando en todos los campos del conocimiento, desde el científico hasta el mismo industrial, e incluso en la vida cotidiana lo que ha desarrollado nuevos conceptos para la definición de productos y

servicios. Esto ha afectado de manera directa, a la documentación que contiene estos términos al imprimirle mayor validez económica, de tal manera que la información se ha convertido ahora en un factor clave en la riqueza de los pueblos, y por ende en el desarrollo industrial de los mismos. La información hace parte ahora del patrimonio de cualquier empresa porque permite la creación de conocimiento.

Si se mira con detalle, se puede decir que básicamente, la sociedad de la información se ha caracterizado por tres aspectos claves¹:

1. Mayor cantidad de trabajadores de la Información, como consecuencia del aumento en el consumo de la misma
2. Nuevas Tecnologías
3. Aumento del PIB en el sector de la información en los países desarrollados.

Moreiro² (1998) por su parte, considera que el gran auge de la información se debe a la aplicación tecnológica del lenguaje, a quien él considera “*el gran vehiculo de la información*”, puesto que las tecnologías de la información y la comunicación (TICs) han permitido mejorar la toma de decisiones en la vida diaria. De ahí, que la información con apoyo tecnológico, se pueda determinar como un aspecto clave para el éxito económico y el acceso a la cultura. Se podría afirmar, sin caer en el ridículo, que esta también es la era de las Tecnologías globales de la Información y no solo de la información.

El mundo actual se basa mas en un conocimiento a través de la imagen mas que de lo escrito, de ahí que surjan conceptos como “*cultura de la pantalla*” que vienen a irrumpir y a reemplazar en gran parte a la llamada “*cultura de la escritura*”. Esto conlleva a la creación de un mundo donde también prima lo inmaterial e intangible abriendo brechas para que los documentalistas y bibliotecólogos puedan ejercer su profesión en la transmisión de esta información. Las nuevas tecnologías, han roto con los esquemas propuestos en materia de información que se venían produciendo como consecuencia de la implementación de la imprenta; si bien, éste invento masificó y permitió una mayor difusión del conocimiento a partir del siglo XV, ahora las nuevas tecnologías de la información no solo son transmisoras de conocimiento, sino que también contribuyen a su organización y control, de forma paralela al aspecto documental. A esto es a lo que se denomina: Revolución de la información.

La unión de TICs, información y procesamiento tecnológico, se ha establecido como un pilar fundamental en la sociedad actual, especialmente desde una óptica político-económica; la información, es así, el mecanismo de poder de los países desarrollados sobre los subdesarrollados y en vías de desarrollo, manifestado en la hegemonía que tienen los primeros sobre los mercados mundiales. La información administrada, actualizada y precisa está llamada a reemplazar el dinero en tanto que las tecnologías sirvan de soporte a procesamiento, almacenamiento y transmisión de la misma, especialmente en formato electrónico por ser la de mas fácil manipulación.

A pesar de todo este preámbulo, el impacto social de las tecnologías es aun incierto pero ya su influencia es palpable, especialmente en la forma como algunos gobiernos ya están

¹ PINEDA, Juan Manuel. El rol del bibliotecólogo en la sociedad de la información

² MOREIRO GONZALEZ, José Antonio. Introducción al estudio de la Información y la documentación.

creando leyes que permitan regular la sociedad informacional, véase por ejemplo las leyes de comercio electrónico o mas recientemente las leyes antispam en Estados Unidos. Pero quizás de los ejemplos mas palpables ha sido el desarrollo de lo que se denomina gobiernos electrónico o e-government, cuyo fin es permitir el acceso de información de parte de los gobiernos de turno a la comunidad. Aún asi, para que estos proyectos lleguen a funcionar a cabalidad, se hace necesario que se creen políticas reales de información, que se facilite el acceso de la comunidad a las tecnologías necesarias y que se organice por parte de profesionales la información que se necesite.

De esta forma queda expuesto, que la base de la Sociedad de la información es precisamente la propia tecnología, como lo afirma el Doctor Joaquin Aguirre en su discurso pronunciado en julio de 1999³: “*La sociedad de la Información es la edad del bit*”, haciendo alusión a Negroponte y a su libro “*Ser Digital*” indicando que estamos viviendo una etapa donde se está pasando del átomo al bit, es decir de lo sólido a lo energético-informativo.

Retos digitales dentro de la Sociedad de la información

Si se hiciera una encuesta acerca de cual es la herramienta mas poderosa que ha surgido dentro de la Sociedad de la Información, en cuanto a transmisión y masificación de información se refiere, ganaría, sin lugar a dudas, la Internet. La razón, es que ha traído consigo cambios importantes y trascendentales en la manera de procesar la información y por ende ha transformado de manera casi radical el papel del profesional en información; esto se aprecia especialmente a raíz del nacimiento de la WWW a principios del decenio pasado. La Internet, como ya se acotó antes, al ser considerada como el medio de comunicación mas importante de los últimos años, ha sido considerada por muchos como la futura Biblioteca Universal, aunque realmente esté muy lejos de serlo; paralelo a esto, el cambio de pensamiento que ha generado ha sido inmenso, puesto que ha creado nuevas formas de aprehensión del conocimiento, de tal manera que esta nueva forma de percepción del mundo requiere que se generen igualmente nuevas formas de organización de la información que los diferentes usuarios consumen.

En vista de esto, el reto ahora, en palabras de Méndez⁴ (2002) es el de definir estratégicamente las funciones del bibliotecólogo siempre basado en la experiencia bibliotecaria tradicional, la cual no se puede desechar, pero siempre con miras a nuevos pilares o estructuras que permitirán el manejo integral de la información digital producto de los nuevos paradigmas de la información. Aquí es donde se hacen importantes los metadatos, porque la biblioteca y los bibliotecólogos deben trabajar con la Internet, no contra ella.

El estudio de los metadatos –continúa Méndez- y su aplicabilidad a futuro en proyectos como bibliotecas digitales, debe obedecer a tres cambios estructurales en lo que a manejo de información se refiere: el cambio de paradigma en los sistemas de información, la normalización en la Internet y la recuperación de la información presente en la Web. Por ahora, considera ella, la adaptación debe hacerse a dos niveles, el primero en cuanto a la

³ AGUIRRE ROMERO, Joaquín María. Las fronteras de la información en la era digital.

⁴ MENDEZ RODRIGUEZ, Eva María. Metadatos y recuperación de información (p. 18)

infraestructura de los protocolos de transmisión de información, labor a desarrollar mas por parte de los ingenieros que de los documentalistas y bibliotecólogos; y como segunda medida, la creación de una Infoestructura, que asegure no solo el acceso a la información sino su recuperación y organización. Es por eso que los metadatos, son solo una de las herramientas llamadas a colaborar en esta parte pero no es la única.

La tendencia actual, una vez desechada la creencia de la Internet como una biblioteca universal, es la que propone Tim Berners-Lee, el mismo creador de la WWW y actual director del consorcio W3C, la cual consiste en la implantación de lo que el llama: una Web Semántica. Esta, no es una nueva Web sino el desarrollo de las potencialidades de la existente, basada en el metalenguaje XML (Extensible Markup Language) y otras tecnologías como lo son la RDF (Resource Description Framework) y la iniciativa DCMI (Dublin Core Metadata Initiative). Con esto se espera que se aumenten las capacidades de almacenamiento de la información estructurada para que pueda ser comprendido por las máquinas desde el punto de vista semántico y así, lograr una mejor relación entre los computadores y las personas. Este tema se profundizará mas al final del texto.

3. Información digital

Al hablar de Era de la Información, de TICs y demás términos descriptivos de esta era, se hace imperativo hablar de los nuevos tipos de documento, que a la larga se conocerán y harán parte de la llamada Información digital y de los recursos digitales de información.

Para entender este concepto, se hace necesario aclarar lo que se entiende por digital y analógico. La gran mayoría de las señales que el ser humano recibe a través de sus sentidos, están representadas análogamente, es decir, por ondas; de tal forma que el sonido, por ejemplo, son ondas o vibraciones que se perciben por medio del oído; por otra parte, al hablar de digital, la representación de esa señal se hará a través de dígitos, o en otras palabras, números. Por todo esto, se puede concluir que⁵ *“una señal digital es la representación numérica de ondas sonoras o visuales”* las cuales, luego son interpretadas y traducidas a un formato análogo que permite su impresión o visualización por parte de los sentidos humanos. Si se aplican esto sencillos conceptos a la Ciencia de la Información, se encontrarán de igual forma, recursos en formato impreso (análogo) y recursos electrónicos (digitales).

Estos dos tipos de recursos, han permitido que el conocimiento se pueda almacenar, administrar y difundir de diferentes formas, lo cual se ha reflejado en las múltiples discusiones acerca de la muerte de los libros en papel y el posterior reinado de los e-books o libros digitales. Los documentos digitales, por su parte, de acuerdo a su origen pueden se pueden clasificar en:

- Documentos de nacimiento digital: representados por aquellos documentos que han sido creados directamente en este formato. Como ejemplo se puede tomar un documento en un procesador de palabras o una página web.
- Documentos digitalizados: o documentos cuyo nacimiento no tiene su origen en los computadores sino que son digitales luego de pasar por un proceso de digitalización.

⁵ ZULETA, Juan Antonio. La imagen digital sin misterios. p. 10

Como ejemplo se puede tomar un libro que luego de ser escaneado se convierte en un archivo en PDF.

Codina⁶ (2001) nombra tres características que -según afirma el- diferencian la información análoga de la digital:

- La computabilidad o la acción de ser “computable” o fácilmente procesado por un computador mediante un programa o algoritmo preciso. Esto se debe a que la información digital se puede representar como multimedial, al permitir, al mismo tiempo, el uso de texto, imagen y sonido; además, por la capacidad de interactuar con el usuario del a misma según sus necesidades, tal y como se puede observar en el proceso de recuperación de información en un sistema.
- La virtualidad o la realidad de que la información exista “sin las limitaciones propias de lo analógico”, o como lo afirma Negroponte la posibilidad de existir sin estar compuesto por átomos solo por bits. Esta existencia virtual, permite que un documento digital siempre esté disponible para cualquier usuario.
A pesar de esto, la virtualidad también posee desventajas frente a la realidad del documento análogo, en lo que se refiere a soporte documental, derechos intelectuales o degradación
- La capacidad o la posibilidad de no tener límites en la cantidad de contenido; en comparación, un documento análogo tiene limitaciones en este aspecto debido precisamente a su soporte documental.

Al aterrizar estos conceptos a la WWW, se ve que los documentos que ésta posee son más complejos que los análogos, esto debido a su carácter multimedial y a que en algunos casos contienen su propia descripción y catalogación implícita en su metalenguaje de creación (metadatos), por tal razón, a estos documentos se les ha denominado comúnmente “objetos de información” u “objetos digitales”, y es a ellos a los que la comunidad documental y científica denomina DLOs (**D**ocument **L**ike **O**bjects); donde los DLOs son unidades básicas informativas (texto, imagen, video, etc...) en un formato de codificación numérica (digital), para cuyo acceso son imprescindibles medios informáticos. De esta forma, la WWW estará llena de miles de DLOs que vienen siendo las unidades mínimas susceptibles de ser descritas a través de metadatos. Esto hace que el tratamiento documental de los DLOs, difiera en ciertos aspectos al de los documentos tradicionales, de tal forma que se debe tener en cuenta, además de los aspectos propios de todo tipo de documento, unos aspectos muy importantes para su análisis documental, tales como⁷:

- La localización, porque su ubicación física es prácticamente inexistente, y tiene diferentes modos de acceso a través de protocolos de intercambio de información como FTP o HTTP; además, puede tener múltiples localizaciones dentro de la Web.
- El formato de los documentos, porque un objeto digital, o DLO, puede estar conformado por diferentes formatos, incluso al mismo tiempo (html, pdf, txt, rdf, jpg, gif, etc.)
- Falta de estabilidad, porque la información digital es bastante inestable, en el sentido que puede cambiar de localización en la Web, puede actualizarse o simplemente puede desaparecer sin dejar rastro.

⁶ CODINA, Lluís. Las propiedades de la información digital.

⁷ MÉNDEZ RODRIGUEZ, ibid. P. 77-79

- Nivel de detalle (granularity), lo cual se refiere al nivel máximo de descripción al que se debe llegar en el tratamiento documental, es decir, si se debe mirar una página Web, un conjunto de páginas Web, las imágenes de la página Web, etc.
- Naturaleza de la ubicación de los datos, es decir, datos tales como restricciones, direcciones de contacto, además de la URI o ubicación electrónica, ya que la ubicación del documento no existe físicamente como en el caso de los documentos tradicionales, cuya localización es concreta (ej, en un estante)

4. Metadatos

Como bien se dijo al principio, y basado en las características antes mencionadas del tratamiento documental para los Objetos digitales, se ha afirmado que una de las principales herramientas creadas con el fin de tomar el control, organización y descripción de esta explosión de información digital, son los llamados Metadatos.

El concepto “metadato” nació en la década de los sesentas, y su creación se atribuye a Jack Myers, quien lo utilizó para describir datos correctamente⁸; actualmente, la definición mas sencilla de metadatos es la que los refiere como “datos sobre datos” de tal forma que lo que básicamente representarán será metainformación. A pesar de esto, tradicionalmente, el término se ha empleado con referencia a datos sobre indización y catalogación creados por profesionales en la ciencia de la información; no obstante, su definición estará subordinada tanto a su labor como al campo profesional que los maneje. Por tal razón, su papel va mas allá de lo simplemente descriptivo, puesto que también se usan en la elaboración de registros de adquisición, catálogos de exhibiciones, datos sobre utilización o incluso, información para la administración de información digital; todo esto en pro de la recuperación de información digital. De esta forma, los metadatos se convierten en componentes primordiales de las Bibliotecas Digitales.

En términos generales, tal y como lo afirma Ortiz-Repiso (1999)⁹ la palabra metadato, en el contexto de las bibliotecas, sin importar si estas son digitales o no, se utiliza comúnmente para referirse a información que:

- Proporciona (normalmente de forma breve) una caracterización de un recurso de información específico en las colecciones de la biblioteca
- Se almacena en los catálogos
- Se usa principalmente para ayudar al usuario a acceder a los recursos de información de su interés

Méndez (2002), por su parte, considera que definir los metadatos requiere tener en cuenta las diferentes tendencias de éstos, con el fin de limitar el concepto y acercarlo mas hacia lo que realmente representan:

⁸MILLER, Paul. Metadata for the masses, [consultado: 28 de octubre de 2001] <http://www.ariadne.ac.uk/issue5/metadata-masses/>

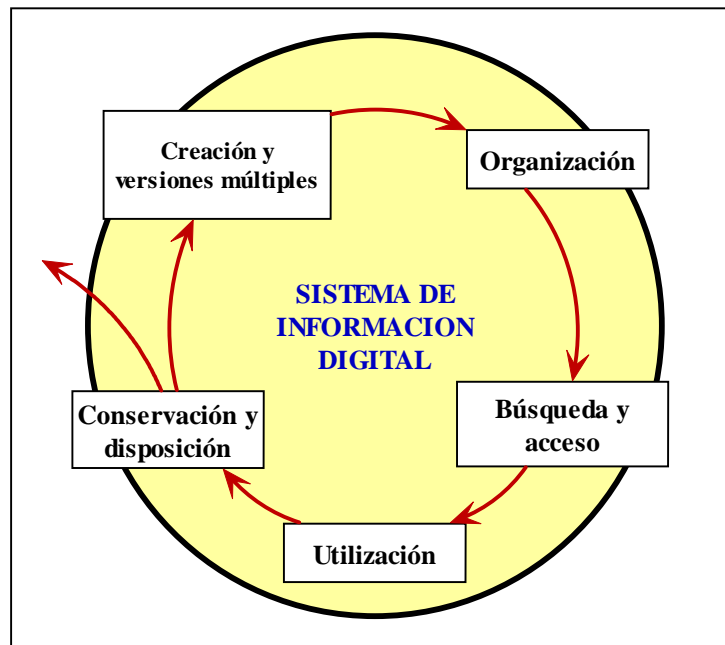
⁹ ORTIZ-REPISO JIMÉNEZ, Virginia. Nuevas perspectivas para la catalogación: metadatos versus MARC. *En: Revista Española de documentación Científica.*

“El término metadatos, alude a cualquier dato que ayuda en la identificación, descripción de recursos electrónicos en red. Son pues, datos que describen los atributos de un recurso, caracterizan sus relaciones, soportan su recuperación, gestión y uso efectivo, y existen en un contexto estrictamente electrónico”

Es por ello, que el contexto de los mismos variará de acuerdo a su área de utilización, por ejemplo, en archivística puede ser utilizado, para designar toda la información contextual, de procesamiento y de uso necesaria para identificar y documentar la aplicación, la autenticidad y la integridad de un registro en un sistema electrónico; de tal manera, que para entenderlo, se debe limitar a su alcance, contenido y campo de aplicación. De Igual forma, si se quiere llegar a una definición mas precisa, se debe tener en cuenta tres aspectos que son inmanentes en estos: su funcionalidad, su contexto y su interlocutor.

La diversidad de los metadatos y de la creación de los mismos por parte de cualquier persona, se ve manifestada en la flexibilidad de la inclusión de estos en cualquier etapa del ciclo de vida de los objetos digitales dentro del sistema digital (Fig. 1), ya sea haciendo parte de los propios objetos o unidos a estos por un hipervínculo de manera bidireccional. Mas sin embargo, y sin importar cuales sean las funciones de éstos, deben conservar una similitud entre los datos que contienen y el objeto del que los toman para que tengan validez documental.

Fig. 1 Ciclo de vida de los objetos dentro de un sistema digital



Fuente: Introducción a los metadatos: Vías a la Información Digital

Debido a la variedad y a la amplitud del concepto *Metadatos*, Gilliland-Swetland¹⁰ considera que éstos, se pueden dividir en 5 grandes categorías: administrativas,

¹⁰ Introducción a los metadatos: vías a la Información Digital. p 3.

descriptivas, de conservación, de uso y técnicos; de acuerdo a la funcionalidad de los mismos. (Tabla1).

Tabla1. Distintos tipos de metadatos y sus funciones

Tipo	Definición	Ejemplos
Administrativo	Metadatos usados en la gestión y administración de recursos de información	<ul style="list-style-type: none"> – Adquisición de información – Control de derechos y reproducciones – Documentación de requisitos legales – Información sobre localización – Criterios de selección para la digitalización – Control de las distintas versiones
Descriptivo	Metadatos usados para describir o identificar información sobre recursos	<ul style="list-style-type: none"> – Catalogación de registros – Ayudas para la búsqueda – Índices especializados – Relaciones hiperenlazadas entre recursos – Anotaciones de usuarios
Conservación	Metadatos relacionados con la preservación de recursos de información	<ul style="list-style-type: none"> – Documentación de recursos sobre la condición física – Documentación de acciones tomadas para conservar versiones físicas y digitales de los recursos, e.g., regeneración y migración de datos

Técnico	Metadatos relacionados con el funcionamiento de los sistemas o el comportamiento de los metadatos	<ul style="list-style-type: none"> – Documentación de hardware y software – Digitalización de la información, p. ej., formatos, ratios de compresión, ajustes. – Control del tiempo de respuesta de los sistemas – Autenticación y seguridad de los datos, p. ej., claves cifradas, contraseñas
Uso	Metadatos relacionados con el nivel y el tipo de uso de los recursos de información.	<ul style="list-style-type: none"> – Registro de exhibiciones – Seguimiento de usos y usuarios – Uso repetido del contenido e información sobre versiones múltiples.

Fuente: Introducción a los metadatos: vías a la información digital p. 3

En contraste con lo anterior, el Dr. Martínez Arellano¹¹, simplifica la función de los metadatos a únicamente tres, aunque igualmente, hace inferencia en que el papel de los metadatos va mas allá de la simple descripción:

1. Proporcionar una descripción de una entidad de información junto con otra información necesaria para su manejo
2. Proporcionar los puntos de acceso a esa descripción
3. Codificar esa descripción para facilitar su manejo por medio automatizados

De todos modos, la utilidad de los metadatos es muy extensa puesto que se pueden emplear en catalogación, recuperación de información en Internet, comercio electrónico, agentes de software inteligente, firmas digitales, valoración y clasificación del contenido, derechos de propiedad intelectual o política de preservación y privacidad del contenido Web entre otros, aunque su uso mas importante es en las bibliotecas digitales.

Méndez, también intenta demostrar la utilidad de los metadatos bajo las diversas visiones de los estudiosos en el tema, sin embargo, describe dos que son bastante importantes y que vale la pena resaltar:

En la primera visión, separa el concepto en “*todo metadatos*”¹² y en “*metadatos en sentido estricto*”, donde el primero recae en el hecho de que los metadatos, se han considerado en

¹¹ MARTÍNEZ ARELLANO, Filiberto Felipe. Metadatos: organización de recursos electrónicos, *En: Revista Interamericana de Nuevas tecnologías de la Información.*

algunos círculos profesionales, como funcionales y aplicables a cualquier tipo de información sin importar su soporte y si es digital o no; sin embargo, al observar el segundo concepto que propone, ella intenta demostrar que realmente los metadatos, son aplicables únicamente a recursos de información electrónicos, basados especialmente en la WWW; de tal manera que los metadatos solo existirán dentro de un ambiente meramente digital. Por esta razón considera que el procesamiento documental de los objetos digitales a través de metadatos, no se puede llamar “catalogación” en el sentido clásico del concepto; precisamente por la profundidad y los aspectos que pueden llegar a manejar los metadatos, que van mucho más allá de la simple descripción documental y temática propuesta por reglas como las RCAA2 o las ISBD; además por las características propias de la información digital que exigen un tratamiento diferente al de los materiales impresos o análogos.

La segunda visión que ofrece¹³, la basa en los conceptos de Burnett, Bor Ng y Park (Méndez, cit. P. 42) donde el concepto y alcance de los metadatos se evoluciona a partir de dos disciplinas: la biblioteconomía (o bibliotecología en el ámbito latinoamericano) y la Informática. De la primera se deriva la funcionalidad descriptiva de los metadatos y dentro de la segunda, el aspecto netamente de gestión de información. De esta forma se demuestra una vez más, que el concepto metadato, va más allá de lo puramente catalográfico y descriptivo para avanzar a planos más profundos de la organización y administración de la información electrónica.

Utilidad de los metadatos

El crecimiento desmedido de la WWW desde su aparición, ha potenciado el papel de los especialistas de la información al abrirle una amplia gama de posibilidades de trabajo a la par que jamás pensó. Todo esto se debe a que la Internet es una red inmensa y aún con muchos caminos por recorrer, a la que con justa razón Gubern¹⁴, compara con un rizoma, puesto que la información que ésta posee no tiene un centro, sino que se enlaza en relaciones asimétricas complementarias discrepantes; lo que ocasiona que un usuario al momento de realizar cualquier proceso de búsqueda en ésta, se aproxime a grandes experiencias entrópicas que le proporcionan cantidades inmensas de información. La causa de esto es que la WWW tal y como la pensó en su comienzo Tim Berners-Lee no fue diseñada con la idea de catalogar su contenido, y la mayoría de sus protocolos y estándares son simplemente para transmisión de datos, por lo tanto no sirven para la localización de recursos de información disponibles en la Web.

Los metadatos son, como ya se ha venido diciendo, una de las herramientas que más fuerza ha cogido en los últimos años destinadas a apoyar la organización del caos existente en la Web. Su función en todo este proceso se puede enfocar desde dos puntos de vista, la del sistema y la del usuario. En la primera facilita la interoperabilidad y la posibilidad de compartir datos dentro de las diferentes herramientas de búsqueda, y en la segunda, permite

¹² MÉNDEZ RODRIGUEZ, Op. Cit. P. P. 38-41

¹³ Ibid. p. 42-44

¹⁴ GUBERN, Roman. El eros electrónico p. 47

la recuperación del conocimiento, al suplir dos necesidades básicas de los usuarios, el acceso y la utilidad de la información.

No obstante, es importante resaltar que los retos profesionales clásicos no han desaparecido, solo ha cambiado el paradigma y las necesidades de descripción, administración y recuperación de la información, es decir, no ha cambiado el “por qué” solo se ha transformado el “como” se hacen las cosas y han aparecido otros medios. Ante esto, Gilliland-Swetland, comenta que los metadatos son importantes en el proceso de recuperación de información debido a que proporcionan:

- *Mas accesibilidad*, al mejorar significativamente la efectividad en la búsqueda, siempre y cuando se haga por medio de metadatos detallados y consistentes; además que facilitan la recuperación de la información debido a su capacidad de actuar en diferentes niveles, es decir, en el objeto mismo o en la colección donde se encuentra este.
- *Retención del contexto*, porque mantienen relacionados el objeto digital a la documentación que le acredita autenticidad, integridad estructural y exhaustividad.
- *Uso mas amplio*, al facilitar el acceso a la información a usuarios que se encuentran ubicados geográficamente en diferentes latitudes del globo terráqueo, algo que es muy útil en instituciones como los museos. Por otra parte, los metadatos permiten proporcionar información acerca de los diversos usos de los sistemas de información y sus contenidos, lo que los hace mas adaptables a las necesidades de los usuarios de los mismos y facilita el proceso de la toma de decisiones en las empresas.
- *Versiones múltiples*, al interrelacionar las diferentes versiones que puede tener un objeto digital, a medida que este evoluciona, al igual que lo relaciona con el objeto real o las copias impresas del mismo.
- *Cuestiones legales*, porque los metadatos permiten a los depositarios identificar los diferentes niveles de información sobre los derechos y las restricciones que afectan el uso de un objeto informático y de sus distintas versiones.
- *Conservación*, porque los metadatos garantizan que los objetos digitales perduren independiente de la evolución de las tecnologías, puesto que proporcionan información sobre su accesibilidad, sus aspectos técnicos o su mantenimiento.
- *Mejora de sistemas y economía*, porque la información técnica que proporcionan, permiten evaluar los sistemas y corregir sus falencias.

De esta forma, queda demostrado que los metadatos, debido a su gran cantidad de aplicaciones, independiente del contexto en el que se utilicen, serán unas de las herramientas llamadas a organizar la información digital en la Web y en otros campos donde la información digital se encuentre desorganizada.

El futuro de la Internet: ¿la Web Semántica?

Ya se ha visto a lo largo de este artículo que la Internet, y mas precisamente la WWW ha sido la tecnología que mas ha permitido la globalización y difusión de las ideas, datos, información y conocimiento dentro de la Sociedad de la Información. Sin embargo, tal y como se comentó al comienzo de este texto, mas exactamente dentro de los retos digitales, la tendencia actual, es alcanzar un punto de equilibrio dentro de la Web ya existente, donde

por medio de ciertos elementos, se logre mejorar la semántica y la relación entre el hombre y/o los computadores. A esto es a lo que Berners-Lee denomina Web Semántica.

Pero, ¿que es exactamente la Web Semántica? ¿Es posible alcanzarla? para entenderla, se hace necesario repasar un poco el origen de la actual WWW. A finales de la década de los 80s, y basado en ideas ya preexistentes, Tim Berners-Lee, un ingeniero que trabajaba para la Organización Europea de Investigación Nuclear, diseñó un sistema basado en el hipertexto (concepto creado por Ted Nelson en 1965) con el fin de compartir información científica dentro de su organización. Su sistema estaba de alguna u otra forma emulando la concepción que tuvo Vannevar Bush en los 40s con su invento llamado “Memex”. Berners-Lee, al aterrizar mejor sus ideas, se unió a Robert Cailliau y juntos crearon el primer Browser de la historia, al cual denominaron World Wide Web. Inmediatamente se encontraron con un inconveniente, los formatos de los documentos no podían ser leídos por todos los computadores y sistemas operativos del momento, de tal forma que diseñaron el HTML, un lenguaje derivado del SGML y crearon el protocolo de red http. De esta forma nació la WWW.

Para el año 1993, ya existían 50 servidores en el mundo que utilizaban la idea revolucionaria de Berners-Lee y en junio de ese año ya existían 130 sitios en Internet con esas características. El crecimiento fue tan sorprendente que para enero de 1997 la cifra de sitios en Internet alcanzaba unos 650 mil, lo que indica un crecimiento del 500 mil por ciento en 42 meses.¹⁵

Con este crecimiento tan descontrolado, los problemas rápidamente aparecieron, el lenguaje HTML inicial sobre el cual se escriben las páginas Web se convirtió en algo que cualquier persona podía manipular y acabó “degradándose” a un estándar sin normalización alguna (guerra Internet Explorer Vs. Netscape), aparecieron miles de sitios Web con información “basura” y la Internet se convirtió en un monstruo incontrolable. Viendo esto, Berners-Lee, creó el consorcio W3C y dentro de sus objetivos, propuso empezar a corregir muchas de las falencias que tenía su WWW inicial. Por tal razón, es que se empieza a vislumbrar el concepto Web Semántica, como una alternativa de trabajo en la Web corriente, para potenciar sus fortalezas y lograr una mejor representación del conocimiento a través de esta.

En palabras del propio Berners-Lee (2000), la Web Semántica se define como:

“una extensión de la Web ya existente, en donde la información se encuentra bien definida y entendible, mejorando la cooperación y mejorando la comunicación y cooperación entre el hombre y la computadora”¹⁶,

Con esta definición, él pretende demostrar que el futuro de la Web es este. Sin embargo por ahora esta solo se ha convertido en poco mas que una aspiración de su propio mentor, y no en una realidad palpable, puesto que aún está muy lejos la posibilidad de que los computadores entiendan semánticamente el contenido de la información existente en la

¹⁵ Op cit. Introducción a los metadatos. P. 11

¹⁶ Cita textual: “The semantic web is not a separate Web but an extension of the current one, in wich information is given well-defined meaning, better enabling computers and people to work in cooperation”

Web, y por ende emulen la forma en que los humanos lo hacen. La recuperación de la información por máquinas aún no se soluciona completamente las complejas necesidades de información del hombre.

La razón de la anterior aspiración se debe a varias causas, como primera medida, Berners-Lee considera que si las máquinas han sido creadas por el hombre deben servirle a cabalidad a este, de esta forma, la Internet debe ser muy rápida y de fácil acceso. Como segunda medida, el considera que debe mejorar la interacción y la comunicación entre el hombre y las computadoras, de tal manera que estas puedan analizar mas fácilmente las necesidades de información del primero. Esto solo será posible con la inclusión de textos con sentido en la Internet, es decir, la inclusión de textos con una semántica igual para el hombre y para la máquina.

Aquí pueden verse claramente dos puntos importantes que se deben trabajar a futuro para lograr una Web con sentido, primero un componente netamente informático que mejore la velocidad de conexión y la accesibilidad a la Internet, para lo cual se deben mejorar los protocolos de comunicación y la tecnología de conexión; y segundo un componente documental que se encargue de darle sentido a la información existente en formato digital en la Web. A pesar de esto, llama la atención que Berners-Lee no nombra en su discurso, la necesidad de crear políticas nacionales que permitan el acceso a la Internet para todas las personas, algo que es fundamental en países subdesarrollados y en vías de desarrollo, puesto que ya se ha visto las brechas informacionales y tecnológicas entre los diferentes pueblos. Si se mira con detenimiento, este concepto de lo informático y lo documental es muy parecido a lo expuesto anteriormente por Burnett, Bor Ng y Park referente a los metadatos y las ciencias de las cuales derivan.

Actualmente, ninguno de estos dos logros se ha concretado del todo, la velocidad de conexión y las facilidades técnicas han mejorado, pero no se encuentran al alcance de la mayoría de las personas (brechas informacionales = inforicos vs. infopobres), igualmente, la aplicación de la semántica de la información digital tampoco se ha generalizado para todos los documentos digitales. Todavía los motores de búsqueda recuperan información de acuerdo a la concordancia de los términos de consulta dentro del texto de los objetos digitales y no a la relevancia de los mismos; los spiders¹⁷ distan aún mucho de ser agentes inteligentes para la recuperación, además que la precisión de su funcionamiento depende en muchos aspectos de los “trucos” que utilizan los autores de textos en la Web para mejorar la recuperación de sus ideas (Vg. spamming, bombing). En conclusión, el reto de la Web semántica radica en proveer un lenguaje que exprese tanto reglas para los datos, como reglas para la representación del conocimiento, de tal forma que toda esta información pueda ser exportada a la Web.

Estructura de la Web Semántica

Para que la Web de Berners-Lee llegue a comportarse según su sueño, se requieren de algunas herramientas que permitan construirla y que le den la estructura necesaria:

¹⁷ Spiders se denominan a las herramientas que se encargan de ubicar la información de Internet para los motores de búsqueda.

1. Un lenguaje que estructure los objetos digitales sintacticamente, denominado XML (eXtensible Markup Language)
2. Un formato que estructure el significado de la información que poseen los objetos digitales (en conjunto con los metadatos asociados a el) denominado RDF (Resource Description Framework)
3. Un programa de computador que recupere la información existente basado en la Inteligencia Artificial, los denominados Agentes Inteligentes
4. Un conjunto de reglas que permitan a los Agentes Inteligentes moverse dentro de la Web con libertad y de acuerdo al perfil informacional del usuario que lo use, denominadas Ontologías

Para entenderlas un poco mas procederemos a explicarlas mas detenidamente.

El lenguaje que permite expresar las reglas para los datos y para la representación del conocimiento es un metalenguaje, el cual se denomina XML (eXtensible Markup Language), que nace como una evolución del SGML a finales del siglo pasado. Contrario a lo que se cree, no se surgió como un reemplazo del ya muy degradado HTML ni mucho menos para la construcción de páginas Web, sino como una herramienta estandarizada y flexible precisamente para la creación de redes semánticas en Internet y para el intercambio de información. Tiene como ventaja, que es de muy fácil uso, comparado con el SGML, y que sirve para la creación de otros lenguajes e incluso de los metadatos, de tal forma que es la herramienta principal para el soporte de estos. Si se compara su organización con la del HTML, se podrá entender el por qué del auge de este metalenguaje en el desarrollo de la Web Semántica y de las bibliotecas digitales, el cual radica básicamente en dos aspectos:

- La flexibilidad en el uso, creación y expansión.
- La posibilidad de creación de etiquetas de carácter semántico (descriptivo) y de acuerdo al contexto del creador; algo de lo que adolece HTML.

Los metadatos por su parte, como se ha venido diciendo hasta ahora a lo largo de este texto, son las herramientas llamadas a indizar, analizar, describir, administrar y recuperarán la información digital dentro de la Web, sin embargo, ante la existencia de diferentes alternativas, se vió la necesidad de crear una infraestructura que permitiera agrupar a estos en un marco con una semántica y una sintaxis única, esta estructura se denomina RDF (Resource Description Framework). Este segundo componente de la Web Semántica se basa en tres elementos: un sujeto, un objeto y una acción (nótese la similitud con la estructura de cualquier oración), los cuales pueden ser escritos usando etiquetas de XML, y se comporta de manera similar a una DTD pero de carácter mas semántico. El RDF garantizará que la máquina pueda entender la estructura y la organización de los metadatos en la Web. Actualmente, el RDF se usa casi como un estándar mas en la organización de las bibliotecas digitales, siendo así una herramienta de gran poder en la recuperación de información no solo en la Web.

Con estos elementos ya es mucho mas fácil para el crawler o buscador ubicar la información que necesita el usuario final, sin embargo, dado que la visión de Berners-Lee es que la Internet sea una estructura que básicamente sirva a las necesidades del hombre especialmente en su construcción de conocimiento, se hace necesario que aparezcan dos componentes mas, los agentes inteligentes y las ontologías.

El primero es básicamente un programa de computador que se adapta a las necesidades de información del hombre, para lo cual tendrá que tener la capacidad de inferir resultados para conseguir una respuesta mas efectiva y eficaz tanto en contenido como en tiempo de recuperación. Sin embargo, el agente por si solo no es capaz de identificar el camino a seguir dentro de la Web puesto que no tiene la capacidad de interactuar de la forma en que lo haría el cerebro humano o un agente humano. Por tal razón, surge el concepto de ontologías, las cuales llegarían a representar los caminos que debe seguir el agente inteligente para poder “atar” los diferentes núcleos de información existentes en el ciberespacio. De esta forma, -en palabras de Codina (2003)- si por ejemplo, un usuario necesita información referente a caballos, el agente gracias a las ontologías recuperará información no solo referente a estos animales sino quizás también referente a yeguas.

Por todo esto, y aunque el mismo Codina (2003) considera que aún se está muy lejos como para cumplir con el sueño de Berners-Lee, la Internet o la WWW se convertiría en una herramienta que facilitaría la difusión de grandes volúmenes de documentos y conocimiento, y quizás, como algunos mas temerarios consideran, se podrá empezar a hablar del concepto de Biblioteca Universal.

Conclusión

Sin importar si se vive o no en la Sociedad de la Información o del Conocimiento, lo cierto es que para nuestra profesión se han abierto puertas que permiten ingresar a mundos donde podemos empezar a aplicar nuestros conocimientos en el campo de la organización de información, o mas recientemente llamado Gestión del Conocimiento. Por esta razón, se hace necesario empezar a educarse y a involucrarse de lleno en las posibilidades que se generan con la aparición de la Información Digital.

El campo donde mas se está trabajando es precisamente el campo de la Internet, representada por la WWW, sin embargo las iniciativas a nivel Latinoamericano en cuanto a proyectos similares, como las bibliotecas digitales, aun son muy pobres y escasas. De tal forma que depende directamente de nosotros si queremos ir a la par de la evolución de la información y las nuevas tecnologías de la Información con el fin de empezar a romper las brechas informacionales que existen y empezar a crear conciencia dentro de los gobiernos sobre esta necesidades tan apremiantes.

Para finalizar me gustaría recordar que todos estos cambios y paradigmas en el manejo de la Información, en este caso de soporte digital, no dependen de las tecnologías sino del manejo que se dé a la misma por parte de profesionales capacitados. Solo asi y de esta forma la sociedad no se volverá dependiente de las TICs sino que dependerá mas de la información y el conocimiento generado. Como decía el titulo de un artículo muy clásico en bibliotecología: “ponga al frente a un buen bibliotecólogo no a un computador”.

Referencias bibliográficas

------. Introducción a los Metadatos: vías a la información digital. Murtha Baca, ed. Los Angeles: J. Paul Getty Trust, 1998

AGUIRRE ROMERO, Joaquin María. Las fronteras de la información en la era digital. En: Especulo. Revista de Estudios literarios, Año 4, No.12, julio – octubre 1999 [En línea] http://www.ucm.es/info/especulo/numero12/era_digi.html [Consultado en Febrero 21 de 2004].

BERNERS-LEE, Tim. Weaving the Web: the original design of the World Wide Web by it's inventor. New York: Harper Collins, 2001 246 p.

BERNERS-LEE, Tim; Hendler, James y LASSILA, Ora. The semantic Web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. En: Scientific American, No. 501, mayo 2001 [En línea] <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21> [Consultado en febrero 01 de 2004]

CODINA, Lluís. La Web semántica: una visión crítica. En: El profesional en Información. Vol. 12, No. 2. Año 2003. P. 149-152.

CODINA, Lluís. Las propiedades de la información digital. En: El profesional en Información. Vol. 10, No. 12. Año 2001 P. 18-25

MENDEZ RODRIGUEZ, Eva María. Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón: Trea, 2002. 429 p.

MOREIRO GONZALEZ, José Antonio. Introducción al estudio de la información y la documentación. Medellín, Universidad de Antioquia, 1998. 188 p.

ORTIZ-REPISO JIMÉNEZ, Virginia. Nuevas perspectivas para la catalogación: metadatos versus MARC. En: Revista Española de Documentación Científica. Vol. 22 N° 2, 1999.

PINEDA, Juan Manuel. El rol del bibliotecólogo en la sociedad de la información. En: Biblios, Año 2, No. 6 2000. [En línea] <http://www.documentalistas.com/web/biblios/articulos/6/B6-02.pdf> [Consultado en Febrero 18 de 2004]

VILCHE ROMAN, Carlos. La sociedad de la Información: ¿mito o realidad?. En: Biblios., Año 4, No 16. Julio – Diciembre 2003. [En línea] http://www.documentalistas.com/web/biblios/articulos/16/2003_023.pdf [Consultado en Febrero 18 de 2004]

ZULETA, Juan Antonio. La imagen digital sin misterios. Buenos Aires: Ediciones Fotográficas Argentinas, 2000. 89 p.