

## La Biblioteca Semántica

JUAN VOUTSSÁS MÁRQUEZ

*Centro Universitario de Investigaciones Bibliotecológicas,  
UNAM, México*

Las bibliotecas digitales están conformadas por tres elementos básicos: las colecciones —*obviamente digitales*—, los servicios —*por supuesto en red*— y los usuarios —*naturalmente diseminados a lo largo del ciberespacio*. De acuerdo con lo anterior, una biblioteca digital les ofrece a sus usuarios servicios en red a través de sus colecciones digitales, al margen de todo lo que pueda poseer en sus acervos tradicionales. Por lo mismo es fundamental saber desarrollar una colección digital adecuadamente. Para empezar, es necesario establecer que, al igual que en una biblioteca tradicional, no existe una única colección documental, sino que en realidad está formada por una variedad de colecciones parciales que engloban distintos tipos de materiales y formatos los cuales conformarán lo que la biblioteca y sus usuarios conocen como “la colección” en términos generales. De esta forma, al referirme a la “colección digital” de una biblioteca estoy diciendo que el conjunto de ellas conforman esa colección. Por lo mismo y, como en toda colección bibliográfica que se respete, debe establecerse un objetivo y una política de desarrollo para la colección —o como hemos explicado— colecciones; es decir, debe haber una *selección*. Ésta es una condición imperativa y que sin embargo es soslayada frecuentemente procediéndose a crear

colecciones digitales de manera directa, sin ninguna planeación. La importancia de esto es tan significativa que sin duda se vuelve el primer principio de la preservación digital.

Al igual que en una biblioteca tradicional, no puede desarrollarse adecuadamente una colección en el entorno digital si no se ha establecido previa y claramente cuál es el objetivo que se persigue al crear esa colección, a quién va dirigida, cuál es su nivel, alcance, cobertura, política de adquisición, permanencia y descarte de materiales, etcétera. Aquí es donde la capacitación y la experiencia del bibliotecario profesional es capitalizada por ser sumamente útil en el momento de establecer esa política de manera adecuada lo que, por supuesto, debe quedar por escrito en un documento al efecto como guía de largo alcance para la biblioteca. Es de suma importancia que en estos objetivos y esta selección, distingan cuáles partes de la colección son para distribuir, cuáles son para preservar y cuáles para ambas cosas.

El deseo de preservar proviene, como ya hemos establecido, de la voluntad de heredar los documentos a futuras generaciones. Según sea la fragilidad del soporte en el que se encuentre el documento, el número de copias existentes, su edad y su estado de conservación, decidiremos que tal documento puede estar en riesgo de destrucción inminente, y deberemos entonces preocuparnos de “fijarlo” en otro soporte para evitar, si no la destrucción del documento-soporte como se conoce, cuando menos sí la de su contenido. Tal sería, por ejemplo, el caso de periódicos de finales del siglo XIX o principios del siglo XX. En otros casos la preservación puede darse evitando su manipulación aunque su estado de conservación sea satisfactorio, para alargar la vida del soporte actual; por ejemplo el caso de un incunable: por lo general, si éste está más o menos conservado, su edad no lo pone en un riesgo inminente de destrucción, pero sabemos que cada vez que ese libro se abre y cierra y que sus páginas son pasadas, lo sometemos a un desgaste físico que si fuese frecuente, en algunas décadas, el esfuerzo mecánico, la exposición a la luz, etcétera, lo deteriorarían inexorable y gravemente, cosa que, por supuesto, no sería deseable. Ésta es la primera decisión de fondo que el bibliotecario debe tomar al momento de pensar en crear y operar colecciones digitales; digitalizar una colección es una decisión que depende de muchas cosas.

Las bibliotecas que sólo crean colecciones para distribuir carecen de responsabilidades de preservación a largo plazo, tienen una tarea un poco más sencilla, ya que el material efímero de esas colecciones no implica tanto trabajo adicional. Pero cuando el objetivo es la preservación, la responsabilidad crece proporcionalmente con el número de variables a ser consideradas. Estas variables obedecen a una serie de principios perfectamente establecidos en la actualidad y que deben ponderarse antes de desarrollar la colección para ver el grado en que debe integrarse cada principio a cada colección de la biblioteca.

Así, aunque los elementos de los documentos digitales están identificados no están agrupados de la misma forma. Por lo mismo, he tratado de establecer una resultante de lo que todo documento que se pretenda preservar debe cumplir y la he agrupado en mi versión de cinco principios generales:

- Selección
- Calidad
- Permanencia
- Accesibilidad
- Funcionalidad

El principio de la calidad es directamente proporcional al grado en que el documento digital emula al documento original o representado, y le otorga por tanto un valor de uso equivalente al de ese documento. Las características que conforman la calidad incluyen conceptos como la apariencia respecto de un original; es decir, ¿qué tanto se parece el documento digital a lo que pretende representar? Esta apariencia a su vez está definida por elementos tales como la resolución, o grado de minuciosidad que tienen los elementos que conforman el documento; tono, color e incluso textura; su escala de —preferencia uno a uno—, su secuencia original, su integridad, un identificador único, etcétera.

La integridad es sumamente importante. Algunos autores la consideran incluso como uno de los elementos básicos de la preservación, y para el mundo de la archivística es indispensable y fundamental. Este concepto se refiere al grado en que el documento digital refleja al

original, no tanto en su apariencia física, sino en su esencia. Un documento íntegro es el que refleja totalmente la esencia del original; es decir, no ha sido corrompido en su contexto: alterado, mutilado, interpretado, aumentado, recortado, deformado, etcétera; es confiable y por tanto aceptable. Su mensaje, autoría, fechas asociadas, lugares, etcétera, son en realidad las consignadas en el documento; es auténtico. En esencia, refleja de manera completa lo que el autor original estableció. Un objeto digital debe poder ser autenticable en al menos tres aspectos. Primero, el usuario debe poder determinar que el objeto es lo que pretende ser. Segundo, el usuario debe poder ser capaz de determinar el origen, estructura, versión, y porción del objeto. Y tercero, el usuario debe poder determinar que el objeto no está corrompido o alterado de forma intencional o accidental.

Los objetos digitales de calidad están nombrados con un identificador único y permanente de acuerdo con un esquema bien documentado. Nunca de acuerdo con una referencia absoluta a su nombre de archivo o dirección —como el caso de su URL— ya que éstos tienden a cambiar con el tiempo. En lugar de ello, un identificador estable debe poder ser resuelto o mapeado siempre a su dirección actual.

El principio de la *permanencia*, tiene que ver con el concepto de que el documento estará disponible durante un lapso considerable. Está asociado con su presencia, su seguridad y, por supuesto, con la duración de su soporte. Muchas personas tienden a confundirlo con el principio de la accesibilidad, el cual tiene que ver con que el documento, existiendo, pueda ser accesible. Son dos conceptos distintos; trataré de explicar brevemente esta diferencia.

Si tenemos en nuestras manos un disco fonográfico de acetato de 78 r.p.m., un cassette con una película “Beta”, una microficha de 4 X 6 pulgadas, una película en 8 mm., o un archivo en *Lotus 1-2-3* en un disquete de ocho pulgadas; un archivo de *Word 5.1* en una cinta magnética de computadora de media pulgada, 7 o 9 canales, y todos estos documentos se encuentran en perfecto estado de conservación, tenemos entonces un conjunto de documentos que han permanecido a lo largo del tiempo: existen sin duda. Han tenido “permanencia”. Pero una cosa es que existan, que hayan permanecido en buen estado, y otra es que podamos acceder a ellos. Para cada uno de ellos reque-

rimos de un accesorio tecnológico del cual dependemos para lograr esa acción. En el caso del archivo *Lotus*, además de un lector de disquetes de ocho pulgadas conectado a una computadora, necesitamos del programa *Lotus* o alguno que pueda acceder a ese archivo; doble requerimiento. Lo mismo sucede con el archivo en *Word* en una cinta magnética: necesitamos el dispositivo lector de la cinta y el programa que pueda abrir e interpretar el archivo, sin mencionar el sistema operativo, que para ese procesador de texto, era MS-DOS. Dependiendo de nuestra capacidad de disponer de esos artefactos, tendremos acceso a esos documentos. Habrá o no “accesibilidad”, independientemente de su permanencia.

Hoy en día ya se maneja el concepto de *archivo permanente*, el cual consiste en una serie de estrategias y técnicas tendientes a lograr que la arquitectura de los sistemas de archivos de información digital para preservación sea independiente de la tecnología usada para crearlos, precisamente para reducir este problema de la accesibilidad.

El concepto de *interoperabilidad* significa que el documento será accesible a lo largo de varias plataformas y programas de computadora. Esto es, el documento no debe estar atado a formatos específicos de un proveedor de *software*, a una marca de computadora o de cámara fotográfica, etcétera. Un ejemplo muy ilustrativo lo vemos cuando intentamos abrir un documento en una página Web y nos aparece el mensaje “este documento sólo puede verse en *explorer* versión tal...”. Ese documento no es interoperable. Un documento en verdadero formato html es visible en prácticamente cualquier navegador: *explorer*, *netscape*, *mozilla*, *firefox*, *ópera*, etcétera; incluso a lo largo de varias versiones de los mismos. El secreto de la interoperabilidad es preferir siempre los *estándares abiertos* y evitar los *estándares propietarios*, ya que a veces, en aras de una cierta comodidad de creación, nos atamos a una plataforma de equipo o de programas.

El último principio mencionado en la preservación es el de la *funcionalidad*. Hablando formalmente, la funcionalidad son *las características interconstruidas en una interface de búsqueda que determinan la facilidad con que los usuarios pueden formular búsquedas y obtener resultados*.<sup>1</sup> Tratemos de explicar este concep-

---

1 ODLIS, Online Dictionary for Library and Information Science.

to: su trasfondo está asociado a la facilidad con que los usuarios puedan acceder a un documento y a sus documentos relacionados. Pero esa facilidad depende de las capacidades que puedan alcanzar las interfaces de búsqueda de documentos, las que a su vez dependen de las características de información que puedan obtener del interior de los propios documentos. En conjunto, la funcionalidad se mide por la suma de las capacidades que tengan los documentos, programas, sistemas, etcétera; si observamos que una parte muy importante de esta funcionalidad proviene del propio documento, podemos entender la importancia de que un documento digital tenga funcionalidad, pues de otra forma se afectaría negativamente al resto de los componentes informáticos que accederán a él.

De ahí la importancia que tienen los metadatos de un “objeto digital”. Sin metadatos el documento digital no podría ser accedido en forma alguna. No tendría funcionalidad. Algunos metadatos sencillos van incrementando paulatinamente esa funcionalidad y ésta es la medida en que el documento puede ser recuperado y asociado con otros. Buenos metadatos nos van aproximando a una buena funcionalidad, como el grupo de metadatos “típico” o “estándar”: metadatos descriptivos, técnicos, estructurales y legales. Obviamente deseamos acercarnos a una óptima funcionalidad para que el documento tenga óptima recuperación y acceso.

Pero para poder entender más a fondo este concepto de “óptima funcionalidad” es necesario explicar primero qué son la *Biblioteca Semántica* y la *Web Semántica*.

La *Biblioteca Semántica* es la tendencia evolutiva de las bibliotecas; las bibliotecas digitales se están convirtiendo poco a poco en *bibliotecas semánticas* y ése será su siguiente estadio. Este concepto se basa a su vez en el principio de la *Web Semántica*. En una explicación sencilla, la *Web Semántica* es una *Web* extendida, dotada de mayor significado en la que cualquier usuario de Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Hasta ahora las páginas *Web* han estado hechas para ser leídas por seres humanos, no por máquinas. Al dotar a la *Web* de más semántica para sí misma, para sus programas y buscadores tendrá a su disposición más significado, y gracias a ello se-

rá posible obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, permitirá compartir, procesar y transferir información de forma sencilla. Esta *Web* extendida y basada en el significado no tan sólo humano sino computacional, se apoya en lenguajes universales que resuelven los problemas ocasionados por una *Web* carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante.

Una *biblioteca semántica* es una biblioteca digital cuyas colecciones están formadas por *documentos funcionales*. Un documento funcional, comienza con un objeto digital, como en una colección de biblioteca digital, documento que, como ya mencionamos, consta de una “reproducción digital fiel”, con sus características de calidad, permanencia, integridad, interoperabilidad, etcétera, pero no se queda en sólo algunos simples metadatos agregados para identificarlo. Un *documento funcional* es un objeto digital de calidad, semánticamente rico, el cual, además de los cuatro tipos de metadatos estándar y gracias a otros metadatos llamados de hipervínculo, ha sido definido y ligado de tal forma con otros documentos funcionales que lo han dotado de un óptimo reuso, búsqueda, integración, minería de datos e interoperabilidad, y lo han provisto también de una lista de sus contenidos que permiten las interfaces con otros programas externos que puedan explotarlos. No se trata de una inteligencia artificial mágica que les permita a las máquinas entender las palabras de los usuarios, es sólo la habilidad programada en una máquina para resolver problemas bien definidos, a través de operaciones bien definidas que se llevarán a cabo sobre datos existentes bien definidos.

Para obtener esa adecuada definición de los datos, la *Web Semántica* utiliza herramientas cuyas primeras versiones de prueba ya se han empezado a definir y utilizar, tales como el “marco de descripción de recursos” —*Resource Description Framework*— —*RDF*— y —*Web Ontology Language*— —*OWL*—, dos estándares que ayudan a convertir la *Web* en una infraestructura global con la que es posible compartir, y reutilizar datos y documentos entre diferentes tipos de usuarios. *RDF* proporciona información descriptiva simple sobre los recursos que se encuentran en la *Web* y que se utiliza, por ejemplo, en catálogos de libros, directorios, colecciones personales de música,

fotos, eventos, etcétera. OWL es una herramienta para desarrollar temas o vocabularios específicos que pueden asociar esos recursos. Lo que hace OWL es proporcionar un lenguaje que puede definir ontologías estructuradas para ser utilizadas a través de diferentes sistemas. Las ontologías son ni más ni menos que la definición moderna y para computadora del concepto de tesoro: incluyen las definiciones de conceptos básicos y subordinados en un campo determinado y la relación entre ellos, y se encargan de definir los términos utilizados para describir y representar un área de conocimiento. Estas ontologías son utilizadas por los usuarios, las bases de datos y las aplicaciones que necesitan compartir información específica; es decir, la de un campo determinado como puede ser el de las finanzas, medicina, deporte, bibliotecología, etc etcétera. Véase el diagrama de *Web Semántica en el Anexo*.

En resumen, la funcionalidad de un documento está dada por características inherentes de calidad aunadas a metadatos adecuados, sobre todo los de hipervínculo los cuales los relacionan con otros documentos semejantes —es decir, funcionales— de manera tal que el conjunto documental así relacionado *tenga un significado* procesable por computadora, y esto permita que los sistemas puedan darle a los documentos prelación, pertinencia, pertenencia, jerarquía, relación, etcétera, y lograr con ello una búsqueda inteligente, y por tanto una recuperación más precisa y útil para el usuario. Esto debe incrementar sensiblemente el reuso y hallazgo de esos documentos, la adecuada minería de datos y la integración, todo bajo un esquema de interoperabilidad entre sistemas.

En conclusión, dentro de la preservación digital, todo documento que se pretenda preservar debe cumplir lo más que sea posible con los cinco principios de los objetos digitales para ser preservados. En la medida en que cumpla con esto, ese documento será un buen objeto digital. Resumiendo, un **buen objeto digital deberá:**

- 1) Provenir de una política específica de selección; y por tanto cumplir con las prioridades y objetivos de la colección.
- 2) Tener calidad intrínseca: apariencia adecuada, resolución, tono escala, secuencia, integridad, comprensión sin pérdida, identificador único, etcétera.

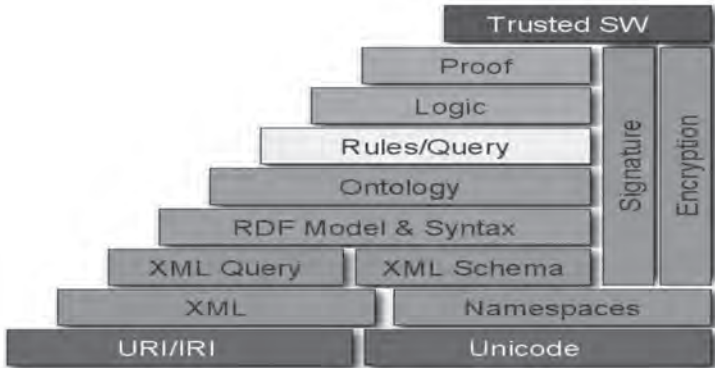


- 3) Ser permanente; esto es, debe existir la intención y responsabilidad de una persona o institución identificadas, de que ese objeto permanecerá accesible a lo largo del tiempo. Y deberá estar representado en un formato digital que soporte el uso actual para el que fue diseñado así como usos futuros, y ser susceptible de ser copiado a otros formatos.
- 4) Tener accesibilidad; esto es, la institución que lo custodia tendrá acceso a las tecnologías que le permitan acceder a ese documento, o bien el objeto digital habrá sido modificado, actualizado, emulado, etcétera a nuevas tecnologías al alcance de la institución. Tendrá por tanto interoperabilidad —será intercambiable entre plataformas—, habrá sido hecho bajo algún estándar o buena práctica reconocidos y sólo se alejará de ellos debido a razones claras y bien documentadas.
- 5) Será funcional; esto es, tendrá una riqueza notable de metadatos, tanto descriptivos como técnicos, estructurales, legales y de hipervínculo, de tal forma que será semánticamente rico, y por tanto estará ligado de tal forma con otros documentos funcionales de manera tal que permitirá su óptimo reuso, búsqueda, integración, minería de datos e interoperabilidad y les dará a otros programas, sistemas e interfaces una lista completa de sus contenidos.

Como habrá podido observarse con estos principios, es sumamente importante que toda biblioteca digital establezca con precisión sus objetivos para desarrollar sus colecciones. Las bibliotecas deben crear buenos objetos digitales, pero el concepto de lo que significa un buen objeto digital depende en mucho de para qué se crean esas colecciones y sus objetos. Por tanto, la primera decisión a tomar, como ya hemos mencionado, es establecer si vamos a crear objetos digitales para preservarlos, para distribuirlos, o para hacer ambas cosas, ya que como hemos empezado a establecer, las características y los costos de cada uno de ellos son sumamente distintas. Nunca debemos emprender proyectos de preservación con especificaciones de distribución ni viceversa. Las consecuencias pueden ser lamentables.

## ANEXO

Diagrama de la *Web Semántica*.



**Unicode** es una codificación del texto que permite utilizar y visualizar en pantalla los caracteres de diferentes idiomas. Esto permite expresar información en la Web Semántica en cualquier idioma.

**URI** es el acrónimo de *Uniform Resource Identifier*, o en castellano, *Identificador Uniforme de Recursos*. Es un poco confusa la diferencia entre URI y URL, por lo que diremos que un URL es un recurso que puede ser accedido via Internet.

**XML + namespaces + esquema xml + XML Query**; la capa más técnica de la Web Semántica. En esta capa se agrupan las diferentes tecnologías que hacen posible que los agentes puedan entenderse entre sí.

**Reglas / Query** - Reglas que deben aplicarse durante la búsqueda para favorecer la recuperación del documento.

**Modelo RDF + sintaxis**; directamente basada y apoyada en la capa anterior, esta capa define el *lenguaje universal* con el cual podemos expresar diferentes ideas en la Web Semántica.

**Ontología ó Lenguaje de Ontologías** - nos permiten extender la funcionalidad de la Web Semántica, y agregar nuevas clases y propiedades para describir los recursos.

**Lógica** - Lógica asociada a las búsquedas o recuperaciones del documento.

**Pruebas** - pruebas aplicables al documento.

**Software de confianza** - software confiable asociado para explotar el documento.

**Firma Digital** - firma digital que puede agregarse al documento para autenticarlo si se desea.

**Encriptado** - llaves de encriptado del documento si se desea, esto.

