

Web semántica, datos enlazados y bibliotecas

Semantic web, linked data and libraries

La sección editorial de la Revista Cubana de Información en Ciencias de la Salud pretende llamar la atención en esta ocasión sobre la importancia de los denominados datos enlazados¹ y su uso en las bibliotecas, archivos, museos y otras instituciones de la memoria. Existe una creciente producción académica sobre el tema, así como iniciativas y proyectos tanto en Cuba como en el resto del mundo² que necesitan ser visibilizados para la comunidad académica y profesional de la información. Puede decirse que el tema ha alcanzado "masa crítica" y esto es resultado también de una maduración teórica y tecnológica que hacen más viable la utilización de metodologías y herramientas de este tipo para resolver problemas concretos de gestión de la información en estas instituciones.

En el reciente Congreso Internacional de la Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA) realizado en Ohio, Estados Unidos, en agosto del año 2016, el grupo de interés sobre Web semántica³ y datos enlazados contó con una numerosa participación y llamó a desarrollar iniciativas y proyectos que aprovechen las oportunidades que ofrecen estos enfoques y tecnologías para incrementar el impacto de los datos bibliográficos en el espacio global de la Web. Este trabajo viene realizándose desde Congresos anteriores.⁴ De hecho, el aporte potencial de los datos bibliográficos y otras bases de datos resultado de la labor de bibliotecas, archivos y museos al proyecto de la Web semántica ha sido visualizado como un componente de la "espina dorsal" o "backbone" de la denominada Web 3.0, si se considera la calidad y el tratamiento profesional de estos datos.⁵

Web semántica y datos enlazados son expresiones estrechamente relacionadas y, como conceptos, forman parte de la producción teórica y técnica del consorcio WWW⁶ responsabilizado con liderar el avance de la infraestructura Web como plataforma abierta de información de alcance universal. Son propuestas originales de su fundador Tim Berners-Lee y referentes necesarios para cualquier sistema de información contemporáneo que se integra o pretende integrarse a la red.

La idea central detrás de la propuesta de la Web semántica es incorporar una capa de datos que recoja relaciones de significado entre los contenidos de la Web. A la red hiperenlazada de documentos se adiciona una capa de datos interenlazados que

declaren de manera explícita la semántica utilizada. Esto con el propósito de que esos datos puedan ser consumidos no solo por los humanos sino también, y sobre todo, por las máquinas, con las consiguientes posibilidades de procesamiento automatizado de esa información incluyendo técnicas de inteligencia artificial.⁷ Datos enlazados es el nombre que emergió a partir de un grupo de recomendaciones específicas del fundador de la Web sobre la forma de representar la información en esa red y que sería la base de concretar de manera descentralizada la visión estratégica propuesta.⁸

En el trabajo original que ha generado una enorme cantidad de contenido teórico y de propuestas prácticas, Berners-Lee insiste en la importancia de enlazar los datos y hacer esto con determinada formalización. Ve la Web semántica como datos enlazados usando un lenguaje formal, en este caso el RDF,⁹ que permite relacionar cosas entre sí con gran libertad pero explicitando la semántica utilizada. Esos enlaces deben realizarse utilizando identificadores uniformes de recursos, URI,¹⁰ que identifican a cualquier objeto o concepto. La propuesta podría resumirse como enlazar todo lo que pueda ser enlazado.

Las cuatro reglas básicas o expectativas de comportamiento para llevar adelante tamaño esfuerzo y que abrirían un camino concreto para avanzar en la visión de la Web 3.0 fueron entonces:⁸

1. Usar URI como nombres de las cosas.
2. Usar URI compatibles con el protocolo HTTP de manera que las personas puedan buscar esos nombres en la Web.
3. Ofrecer como respuesta información útil en esas direcciones usando estándares como el caso de RDF.
4. Incluir enlaces a otros URI para poder descubrir más cosas.

Las bibliotecas, archivos, museos y otras instituciones dedicadas a la preservación y disseminación del patrimonio cultural y la memoria de los pueblos se han distinguido por la producción sistemática y organizada de datos. Sus catálogos y bases de datos referenciales son tal vez los ejemplos más representativos de su esfuerzo en este sentido. Estas comunidades han desarrollado estándares y metodologías para describir y representar, de la mejor manera posible, sus colecciones de documentos y objetos. Muchos de estos catálogos están disponibles en Internet desde hace años y algunos han comenzado a ofrecer representaciones de sus datos en forma de datos enlazados. Un grupo importante de bibliotecas están integradas en consorcios bibliotecarios como la Online Computer Library Center (OCLC), que han producido catálogos colectivos de alcance mundial de manera colaborativa como es el caso de WorldCat (<http://www.worldcat.org/>) y herramientas como el registro internacional de autoridades conocido como VIAF (<http://viaf.org/>), que ahora forman parte del espacio de datos enlazados.

El aislamiento relativo de estos recursos de información y su existencia paralela con los principales motores de búsqueda como Google, u otros recursos de información de referencia como la Wikipedia, ha ido cambiando en la medida que han avanzado iniciativas y proyectos para incorporarlos a la web de datos. Esto ha abierto una perspectiva alentadora para lograr una mayor presencia de estos recursos aprovechando todo el potencial de la arquitectura de la Web. Aunque los impactos concretos de esta nueva visión se han demorado más de lo deseado y de lo que inicialmente fue prometido en este campo,⁷ en los últimos cinco años un grupo de bibliotecas nacionales, archivos y museos han dado importantes pasos para

transformar sus datos y exponerlos siguiendo los principios y metodologías mencionados. La biblioteca del Congreso de los Estados Unidos, la Biblioteca Nacional de Alemania, la Biblioteca Británica y muchas otras importantes bibliotecas del mundo ofrecen indistintamente sus registros de autoridades, sus catálogos bibliográficos y otros recursos siguiendo las cuatro reglas de los datos enlazados. La Biblioteca Nacional de Cuba, junto a la Biblioteca Central de la Universidad de La Habana y el Instituto de Historia de Cuba están enfrascados en un proyecto peculiar que trata de convertir los viejos catálogos de fichas impresas en datos enlazados, iniciativa que fue muy bien acogida en la mencionada reunión de IFLA.¹¹

Las demandas de nuevos conocimientos que exigen transformaciones como la mencionada deben estar estrechamente vinculadas a la práctica. El bombardeo permanente de información en Internet y la pérdida del sentido cuando no se tienen referentes claros de propósitos hace que muchas veces no se conozca incluso lo que debe ser aprendido. En este contexto, las publicaciones científicas pueden jugar un papel orientador. La mejor manera de aprender sobre estos temas es hacerlo como parte de procesos de construcción de soluciones concretas. Hay una interrelación muy dinámica entre el diseño de nuevas actividades y la adquisición de los conocimientos y las habilidades que se requieren. Es decir, que aprendizaje y práctica deben ir de la mano. Es lo que se ha denominado aprendizaje expansivo,¹² que ofrece una perspectiva teórica ajustada a las necesidades de estudio de procesos de construcción de sistemas de actividades en los que el aprendizaje se manifiesta de manera primaria como cambios en un objeto colectivo de la actividad. Teóricamente, el aprendizaje expansivo exitoso lleva a una transformación cualitativa de todos los componentes de un sistema de actividad.¹³

El tema de la Web semántica, y en particular la aplicación de técnicas de datos enlazados en las actividades bibliotecarias, de archivos y museos tiene un significativo potencial para estimular procesos de innovación que ayuden a dinamizar el sistema de actividades de estas instituciones y ofrecer un mejor posicionamiento de sus recursos de información y sus servicios en el contexto de la Web. El desarrollo de proyectos e iniciativas concretas para recuperar, organizar y posicionar los datos descriptivos de las colecciones de objetos culturales acumulados en bibliotecas, archivos y museos puede operar como una zona de desarrollo y revitalizar el papel de estas instituciones en la sociedad. Ese proceso puede estimular el aprendizaje y la innovación, la interacción de estas instituciones y sus profesionales entre sí y con otras instituciones de la información y las comunicaciones y con el espacio abierto de la Web e Internet en general. Ofrece también la posibilidad de incorporar visiones alternativas del mundo en la construcción de los discursos globalizadores a través de la forma en que se identifican y organizan los datos de la realidad y sus interpretaciones. Esta es una labor a la que no pueden renunciar las instituciones responsabilizadas con la preservación y difusión de la memoria y la cultura de las naciones y los pueblos.

PEDRO URRÁ GONZÁLEZ,

Profesor de la Universidad de La Habana, Cuba.

REFERENCIAS BIBLIOGRÁFICAS

1. LinkedData - W3C Wiki [Internet]. 2014 [citado 14 de octubre de 2014]. Disponible en: <http://www.w3.org/wiki/LinkedData>
2. LODLAM [Internet]. 2016 [citado 25 de septiembre de 2016]. Disponible en: <http://lod-lam.net/summit/>
3. Web semántica - Wikipedia, la enciclopedia libre [Internet]. 2016 [citado 25 de septiembre de 2016]. Disponible en: https://es.wikipedia.org/wiki/Web_sem%C3%A1ntica
4. BnF-IFLA 2014 Satellite Meeting - Linked Data in Libraries: Let's make it happen! - Program [Internet]. 2014 [citado 5 de octubre de 2014]. Disponible en: <http://ifla2014-satdata.bnf.fr/program.html>
5. Neubert J, Tochtermann K. Linked Library Data: Offering a Backbone for the Semantic Web. En: Lukose D, Ahmad AR, Suliman A, editores. Knowledge Technology [Internet]. Springer Berlin Heidelberg; 2012 [citado 11 de junio de 2016]. p. 37-45. (Communications in Computer and Information Science). Disponible en: http://link.springer.com.proxy.cc.uic.edu/chapter/10.1007/978-3-642-32826-8_4
6. World Wide Web Consortium (W3C) [Internet]. 2014 [citado 5 de octubre de 2014]. Disponible en: <http://www.w3.org/>
7. Byrne G, Goddard L. The strongest link: Libraries and linked data. Lib Mag [Internet]. 2010 [citado 24 de abril de 2015];16(11):5. Disponible en: <http://dialnet.unirioja.es/servlet/articulo?codigo=3745053>
8. Berners-Lee T. Linked Data - Design Issues [Internet]. 2006 [citado 18 de junio de 2014]. Disponible en: <http://www.w3.org/DesignIssues/LinkedData.html>
9. Yu L. The Building Block for the Semantic Web: RDF. En: A Developer's Guide to the Semantic Web [Internet]. Springer Berlin Heidelberg. 2014 [citado 16 de agosto de 2015]. p. 23-95. Disponible en: http://link.springer.com.proxy.cc.uic.edu/chapter/10.1007/978-3-662-43796-4_2
10. Identificador de recursos uniforme - Wikipedia, la enciclopedia libre [Internet]. 2016 [citado 25 de septiembre de 2016]. Disponible en: <https://es.wikipedia.org/wiki/URI>
11. Biblioteca Nacional "José Martí". Catálogo de Libros y Folletos-Oficial. Versión digitalizada [Internet]. 2016 [citado 16 de julio de 2016]. Disponible en: <http://localhost/cgi-bin/item.py?idficha=bnjmsculyfof0010001>
12. Engeström Y. Expansive learning at work: Toward an activity theoretical reconceptualization. J Educ Work [Internet]. 2001 [citado 25 de septiembre de 2016];14(1):133-156. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/13639080020028747>
13. Engeström Y, Sannino A. Studies of expansive learning: Foundations, findings and future challenges. Educ Res Rev [Internet]. 2010 [citado 25 de septiembre de 2016];5(1):1-24. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1747938X10000035>