

Wikidata y Scholia como herramientas para la gestión de datos bibliográficos y la concreción de una Web Semántica

Ghiglione, Agustina

aghiglione@fahce.unlp.edu.ar

Mesa 1: Trabajos de adscripción

Cátedra: Tratamiento automático de la información II (2022)

Palabras clave: FUNDACIÓN WIKIMEDIA, DATOS ENLAZADOS, RDF, WIKIDATA, SCHOLIA, WEB SEMÁNTICA.

Resumen ampliado

El trabajo de adscripción desarrollado para la cátedra de Tratamiento Automático de la Información II, se vertebró en primer lugar, sobre algunas de las innovaciones aparejadas por el desarrollo de la denominada Web Semántica. En palabras de Tim Berners-Lee, el creador de la Word Wide Web, la Web Semántica “no es una web separada sino una extensión de la actual, donde la información está dotada de un significado bien definido, los ordenadores están mejor capacitados y las personas trabajan en colaboración” (Beners-Lee, 2006). En este punto, es importante mencionar, que la concreción de esto solamente es posible mediante la publicación de datos estructurados con esquemas de metadatos como RDF (*Resource Description Framework*) que garanticen el enlace de la información y su interoperabilidad.

En segundo lugar, a fin de contar con ejemplos prácticos sobre datos estructurados y su importancia para la Web Semántica, se consideraron iniciativas de la Fundación Wikimedia. Esta es una organización sin fines de lucro fundada el 20 de junio de 2003 a partir de la fusión de Wikipedia y Nupedia. En la actualidad, su principal misión es lograr reunir y desarrollar contenido educativo bajo licencias libres o de dominio público. Estrechamente relacionado con esto, su visión se sintetiza en la capacidad de “imaginar un mundo en el que todo ser humano pueda compartir libremente la suma de todo el conocimiento” (Wikipedia, 2022). Siempre bajo estos preceptos, se han desarrollado a lo largo de los diecinueve años de la Fundación diversos proyectos, con la finalidad de difundir el conocimiento de forma libre.

En el marco de esta adscripción nos centramos en el proyecto Wikidata y en la herramienta desarrollada para gestionar la información bibliográfica que allí se encuentra, Scholia. Por un lado, Wikidata es una base de datos enlazados, libre, colaborativa y multilingüe, que sirve de base de datos secundaria (Wikidata, 2022). Es decir, que no sólo registra datos sino también sus fuentes. Tiene la particularidad, además, de ser compatible con otros sitios, servicios y bases, aunque estos no sean de la Fundación Wikimedia. Por otro lado, Scholia es la herramienta desarrollada para “manejar” o “gestionar” la información bibliográfica científica que se encuentra alojada en Wikidata. Permite visualizar de forma gráfica tópicos como la correlación entre autores y artículos, cantidad de citas de un/a autor/a o revista en determinado periodo, coautoría dentro y fuera de las instituciones o incluso en revistas específicas.

Es posible afirmar, entonces, que ambas iniciativas se basan en los principios y tecnologías de la Web Semántica, como lo es RDF, contribuyendo a su concreción. En paralelo, a partir de la introducción de datos de forma automática, se proporcionan perfiles académicos e institucionales en base a información confiable cargada por la propia comunidad. Asimismo, la visualización gráfica que puede realizarse desde Scholia sirve para ser integrada en las diversas páginas o utilizarse para el análisis de datos y la toma de decisiones.

Explicado lo anterior, el primer objetivo de la adscripción fue realizar un estudio exploratorio con el fin de relevar la bibliografía disponible sobre la temática y los proyectos específicos realizados. En este sentido, y desde los lentes de nuestro campo disciplinar, se destacan los trabajos de Pavarini da Luz *et. al.* (2019) y Hernández-Cazorla, Sánchez y Rodríguez Herrera (2019). En el primero de ellos los autores se ciñen estrictamente a analizar los modelos de recuperación de la información que pueden emplearse en Wikidata, no tanto así como a la gestión de los datos allí disponibles. A pesar de esto, el trabajo brinda definiciones conceptuales sencillas de comprender para aproximarse a la cuestión de los datos estructurados, RDF y Wikidata. El segundo de los trabajos seleccionados, es producto de un proyecto concreto realizado en el Instituto Universitario de Análisis y Aplicaciones Textuales (IATEXT) de la Universidad de Las Palmas de Gran Canaria, España. El mismo tenía como objetivo, por un lado, demostrar el potencial de Wikidata y Scholia, y por otro lado, procesar de forma automática la producción científica de la institución para mejorar su visibilidad e interoperabilidad. A primera vista la bibliografía académica disponible es reciente y escasa,

aunque carecemos en la actualidad de estudios bibliométricos que respalden la afirmación. Por esta razón fue de gran utilidad la información obtenida desde la página oficial de la Fundación Wikimedia.

En cuanto al segundo objetivo de la adscripción, se planteó la importancia de elaborar instructivos cuyo objetivo fuera que los alumnos pudieran comprender el funcionamiento de las herramientas. A partir de ellos, se llevó a cabo una actividad en la cual se les propuso a los alumnos trabajar con artículos de revistas disponibles en el Portal de Revistas de la Facultad de Humanidades y Ciencias de la Educación, e introducir sus metadatos de forma automática utilizando QuickStatements en Wikidata. Luego de la curaduría correspondiente, se procedía a visualizar la información procesada en Scholia.

Dada la escasa bibliografía sobre aplicaciones prácticas de las iniciativas, la actividad fue de gran importancia para observar y documentar, tanto ventajas como desventajas. En líneas generales, los alumnos lograron concretar lo planificado y al haber realizado los procesos desde un usuario de Wikidata unificado, creado únicamente para la cátedra, es posible “mapear” los resultados. Teniendo en cuenta que se logró realizar la carga de los recursos de forma automática, podría afirmarse que una de las grandes ventajas es la interfaz gráfica diseñada para Wikidata, así como también, la conexión con QuickStatements que facilita la extracción de metadatos y la descripción con RDF, sin tener la necesidad de agregarlos y vincularlos manualmente. No se presentaron mayores dificultades respecto al uso, lo que no significa, que no haya cuestiones para mejorar. Sobre esto último, se hace necesario mencionar la necesidad de que dentro de la comunidad Wikipedista, se trabajó en conjunto para encontrar una forma de vincular de forma automática las citas y los documentos citados. Ya que hasta el momento puede realizarse pero únicamente de forma manual, lo cual es poco operativo.

En conclusión, las líneas de trabajo que se abren en relación tanto a Wikidata como a Scholia son varias y se encuentran aún vacantes de explorar en profundidad. Nuestro rol como profesionales en el desarrollo de las iniciativas, así como también de la Web Semántica y por qué no, otras tecnologías, se encuentra de igual manera en discusión. A pesar de ello, esto no debería ser un impedimento para comenzar a trabajar sobre lo existente, teniendo en cuenta las ventajas de disponer de información bibliográfica institucional de una forma sencilla y abierta.

Bibliografía

- Berners-Lee, T. (2006). *Linked Data*. Recuperado de <https://www.w3.org/DesignIssues/LinkedData.html>
- Hernández-Cazorla, I., Ramírez Sánchez, M. y Rodríguez-Herrera, G. (2019). Wikidata, WikiCite y Scholia como herramientas para un corpus de datos bibliográficos enlazados. Curación y estructuración de la producción científica de los investigadores del IATEXT. *Prima*, 40, 78-87. Recuperado de <https://ojs.letras.up.pt/ojs/index.php/prismacom/article/view/6529/6086>
- Luz, L. P. da, Coneglian, C. S. y Santarem Segundo, J. E. (2019). Tecnologias da web semântica para a recuperação da informação no wikidata. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, 17, e019003. Recuperado de <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8651791>.
- Wikidata: Scholia (12 de octubre de 2022). En *Wikidata*. Recuperado de <https://www.wikidata.org/w/index.php?title=Wikidata:Scholia/es&oldid=1568079501>
- Fundación Wikimedia (12 de octubre de 2022). En *Wikipedia*. Recuperado de https://es.wikipedia.org/w/index.php?title=Fundaci%C3%B3n_Wikimedia&oldid=146546064