

Clasificación automatizada usando Datos abiertos enlazados. Un estudio sobre la clasificación facetada y Wikidata

Automated classification using Linked Open Data. A case study on faceted classification and Wikidata

Carlo Bianchini & Stefano Bargioni

Cataloging & classification quarterly, ISSN 1544-4554, Vol. 59, n. 8, 2021, p. 835-852

Es difícil encontrar números de clasificación interesantes para una biblioteca occidental en la Parte 3 de la Clasificación Colonada (CC). Una Parte 4 dedicada a los clásicos occidentales podría ayudar a difundir el uso de la CC fuera de India. Su construcción llevaría mucho tiempo y solo sería útil si se pudiera realizar a gran escala de manera automatizada, rápida y eficiente. Dos estudios previos demuestran que un proceso automatizado de clasificación basado en clasificaciones facetadas como la CC y en datos disponibles tan consistentes y completos como los Datos abiertos enlazados (LOD) es factible. En la CC el proceso de clasificación de obras literarias puede transformarse en un algoritmo. La primera faceta determina el lenguaje, la segunda representa la forma literaria de la obra, la tercera al autor de la obra literaria (o al autor sobre el que trata la obra) y la cuarta la manifestación de la obra. Se eligió Wikidata para el desarrollo del proceso automatizado por tres razones principales: la disponibilidad de una herramienta para la creación automática y semiautomática de datos adicionales,

tales como números de clasificación; el uso del software Wikibase, que permite la creación de bases de conocimiento colaborativas; y su difusión en el mundo bibliotecario. Además, se puede enriquecer Wikidata con preferencias y gadgets, y existe la disponibilidad de bots, que pueden realizar ediciones sin participación humana. El gadget CCLitBox se creó como una herramienta de clasificación automatizada. Su objetivo principal es ejecutar un algoritmo para la construcción automatizada de los números de clasificación del autor y la obra del campo de Literatura. La herramienta contiene dos tipos de evaluación de consistencia, que permiten comprobar la presencia de datos relevantes sobre el ítem en Wikidata y verificar que el ítem no tiene todavía un número dentro de la CC. La plataforma Wikidata ofrece un entorno cooperativo, de manera que cuando faltan datos para la creación de una clasificación de autor o de obra, pueden ser fácilmente añadidos por el usuario interesado. Los datos de clasificación pueden exportarse de manera sencilla y ser reutilizados en cualquier formato proporcionado por SPARQL (JSON, TSV, CSV y HTML). Además, como cualquiera puede añadir y editar datos en Wikidata, una pequeña biblioteca puede grabar los datos de sus autores y obras, usar el gadget y obtener el número de clasificación de la CC deseado para reutilizarlo en su catálogo y aplicarlo en sus colecciones de manera gratuita.

<https://www.tandfonline.com/doi/full/10.1080/01639374.2021.1977447>