# MAchine Readable Cataloging to MAchine Understandable Data with Distributed Big Data Management

Kumar Sharma, Ujjal Marjit & Utpal Biswas

*Journal of Library Metadata*, ISSN 1937-5034, Vol. 18, n. 1, 2018, p. 13-29

En los últimos años las bibliotecas han usado las tecnologías de la web semántica para permitir que la información basada en datos pueda ser procesada directamente por máquinas. Los intentos de transformar los datos han evolucionado desde los formatos MARC hacia RDF. Almacenar datos bibliotecarios en formato RDF enriquece los enlaces combinados y la reutilización de recursos en la web. Además, la máquina puede interpretar los recursos bibliotecarios de manera comprensible gracias a la riqueza de fuentes semánticas. Las actuales estrategias reposan en un entorno de nodo único, pero fallan cuando se encuentran con grandes volúmenes de datos entrantes. Algunos de los registros bibliográficos en formatos MARC 21 tienen un gran tamaño que no pueden procesar las herramientas tradicionales de gestión de datos y requiera áreas de almacenamiento más amplias. Estos datos requieren una seria atención por los sistemas que realizan tareas en paralelo. En este artículo proponemos una estrategia distribuida para convertir los datos de bibliotecas de patrimonio cultural en formato RDF usando Apache Spark y Hadoop. Describimos el proceso de conversión de datos de formatos MARC 21 para datos bibliográficos en RDF y mostramos informes preliminares sobre la velocidad de proceso y análisis de almacenamiento. La ejecución del proceso de conversión se mejora en términos de tiempo de proceso y tamaño de almacenamiento.

Traducción del resumen de la propia publicación