

Semantic enrichment for enhancing LAM data for supporting digital humanities. Review article

Marcia-Lei Zeng

El profesional de la información, ISSN 1699-2407, Vol. 28, n. 1, 2019, p. 1-35

El papel de las bibliotecas, archivos y museos (LAM) en el apoyo a la investigación y educación en humanidades digitales (DH) ha sido ampliamente reconocido en los últimos años. En la investigación de DH, la parte difícil no suele estar en las etapas de limpieza de datos, análisis, visualización y sintetización. La etapa más desafiante es esencialmente al principio, cuando se decide qué y cómo se pueden recopilar datos al tratar con materiales históricos. Para aquellos investigadores que necesitan estos datos históricos que no se pueden obtener a través de la web, estos mismos y los recursos de información proporcionados por los LAM tienen un valor extraordinario. Este artículo presenta una serie de métodos y esfuerzos de enriquecimiento semántico que pueden aplicarse a los datos LAM en varios niveles, después de la explicación básica de un conjunto de conceptos clave, métodos y enfoques que se explican a través de tipos de datos que deben mejorarse, categorizados principalmente como datos estructurados, semiestructurados y no estructurados. El artículo concluye con los puntos de referencia recomendados por el W3C (2017) *Mejores prácticas de datos en la Web*, que identifican los objetivos finales para los datos LAM: comprensión, procesabilidad, capacidad de descubrimiento, posibilidad de reutilización y efectividad, confianza, conectividad, accesibilidad e interoperabilidad. Con el rápido desarrollo del campo de las humanidades digitales, las demandas más grandes e inteligentes de datos de patrimonio histórico y cultural, que por lo general no se puede obtener a través del rastreo web, han centrado la atención en los datos LAM, el tesoro de toda la sociedad. En las secciones de "Discusión" dedicadas a cada uno de estos tipos de datos que aparecen en el grueso del artículo, se resumen las características típicas y los enfoques, y se dan recursos para las discusiones posteriores. Las tecnologías de la información y las nuevas tecnologías semánticas que soportan cada uno de los procesos de enriquecimiento semántico han avanzado rápidamente. Se pueden encontrar informes de diversos experimentos, herramientas y problemas en publicaciones de una amplia gama de dominios, y con extraordinaria profundidad. No importa qué productos finales resulten del enriquecimiento semántico de datos LAM, lo más probable es que se unan a los que ya existen en la web. En conclusión, este estudio se centra principalmente en la publicación de datos, en lugar del consumo de los mismos, y está orientado hacia los que están disponibles a través de la web. Proporciona las mejores prácticas relacionadas con la publicación y el uso de estos datos en la web, para que sean detectables y

comprensibles para los humanos y las máquinas, mientras se facilita un ecosistema autosostenido.

Resumen realizado por José María Amate Sánchez