

# Developing and image-based classifier for detecting poetic content in historic newspaper collections

Lorang, Elizabeth, y otros.

*D-Lib Magazine*, ISSN 1082-9873, Vol. 21, n. 7/8, 2015

El equipo del proyecto *Image analysis for discovering archive* (Aida) está investigando el uso de análisis de imagen para identificar contenido poético en los periódicos históricos. El proyecto busca tanto aumentar el estudio de la literatura (llamando la atención sobre la magnitud de la poesía publicada en los periódicos y haciendo que la poesía sea más accesible) como avanzar en el trabajo sobre el uso de imágenes digitales para facilitar el descubrimiento en las bibliotecas digitales y otras colecciones digitalizadas. Este artículo, destinado tanto para una amplia gama de bibliotecas digitales como para público interesado en colecciones digitales de humanidades, describe la primera etapa del trabajo, que se lleva a cabo a través de un clasificador de aprendizaje automático para identificar contenido poético. Se procesan más de 7 millones de imágenes de la página *Chronicling Latina*, el archivo de acceso público del National Digital Newspaper Program of the United States. Los valores de precisión y recuperación alcanzados durante ésta primera etapa (90,58%; 79,4%) y análisis (74,92%; 61,84%) son alentadores. Además, en el proceso de formación, el clasificador ha ayudado a vislumbrar los desafíos potenciales, incluyendo los presentados por diferentes calidades de diseño y diagramación de los periódicos originales y por daño y degradación causados por el paso del tiempo, así como las introducidas durante la microfilmación y la posterior digitalización. Por lo tanto,

además de la exploración de análisis de imagen para el descubrimiento de archivo, éste proyecto también explora las posibilidades y limitaciones de las estrategias de preservación y acceso. Éste trabajo, además, analiza los resultados preliminares, que apoyan la factibilidad y viabilidad para la detección de contenido poético en las colecciones de periódicos históricos. Por último, mientras que el equipo del proyecto Aida ha comenzado con un solo estudio de caso para explorar el análisis de imágenes como estrategia de descubrimiento, un objetivo más amplio del proyecto es avanzar en el procesamiento de imágenes y análisis de imágenes como una metodología dentro de las humanidades digitales y de las comunidades de bibliotecas digitales.

Resumen elaborado por la Sección de Documentación  
Bibliotecaria