

A workflow management system to feed digital libraries: proposal and case study*

Ángeles S. Places¹, Antonio Fariña¹, Miguel R. Luaces¹,
Óscar Pedreira¹ y Diego Seco²

¹ Universidade da Coruña, Laboratorio de Bases de Datos

² University of Concepción, Department of Computer Science, Chile

{asplaces, fari, luaces, oscar.pedreira}@udc.es, dseco@udec.cl

Resumen. La construcción de una biblioteca digital de documentos antiguos implica, además de aspectos técnicos relativos a la implementación, de aspectos relacionados con el proceso de digitalización de grandes colecciones de documentos. Los documentos antiguos normalmente son delicados y necesitan ser manejados con extremo cuidado. Además, el pobre estado de conservación y el uso de fuentes tipográficas irreconocibles dificultan enormemente el proceso de reconocimiento automático y, por lo tanto, requieren de una profunda revisión por parte de personal que realice las correcciones necesarias en el texto. Esto hace que la participación de expertos en el proceso de digitalización sea obligatoria (y costosa).

En este trabajo, se presenta un sistema para la gestión del flujo de trabajo asociado a la digitalización de grandes colecciones de documentos antiguos. Por un lado, describimos el proceso de digitalización y mostramos una herramienta que soporta todas sus fases y tareas. Por otro lado, presentamos un caso de estudio en el que mostramos como nuestro sistema de control de flujo de trabajo fue aplicado a la digitalización de más de 10.000 documentos provenientes de revistas del siglo XIX. Por último, mostramos la biblioteca digital resultante, centrándonos en sus aspectos tecnológicos más relevantes.

Palabras clave: bibliotecas digitales, recuperación de texto, sistema de control de flujo de trabajo.

Publicación: A workflow Management System to feed digital libraries: proposal and case study. *Multimedia tools and Applications*, 75(7): 3843–3877, 2016 (DOI: doi:10.1007/s11042-014-2155-3).

* Este artículo ha sido financiado parcialmente por: Ministerio de Economía y Competitividad (PGE, CDTI, and FEDER) refs. [TIN2016-78011-C4-1-R, TIN2015-69951-R, TIN2013-46238-C4-3-R, ITC-20161074, ITC-20151305, ITC-20151247, IDI-20141259]; Xunta de Galicia (co-financiado por FEDER) ref. [ED431G/01]; y European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 690941 (project BIRDS).