

La búsqueda del santo algoritmo

Datación automática de manuscritos medievales

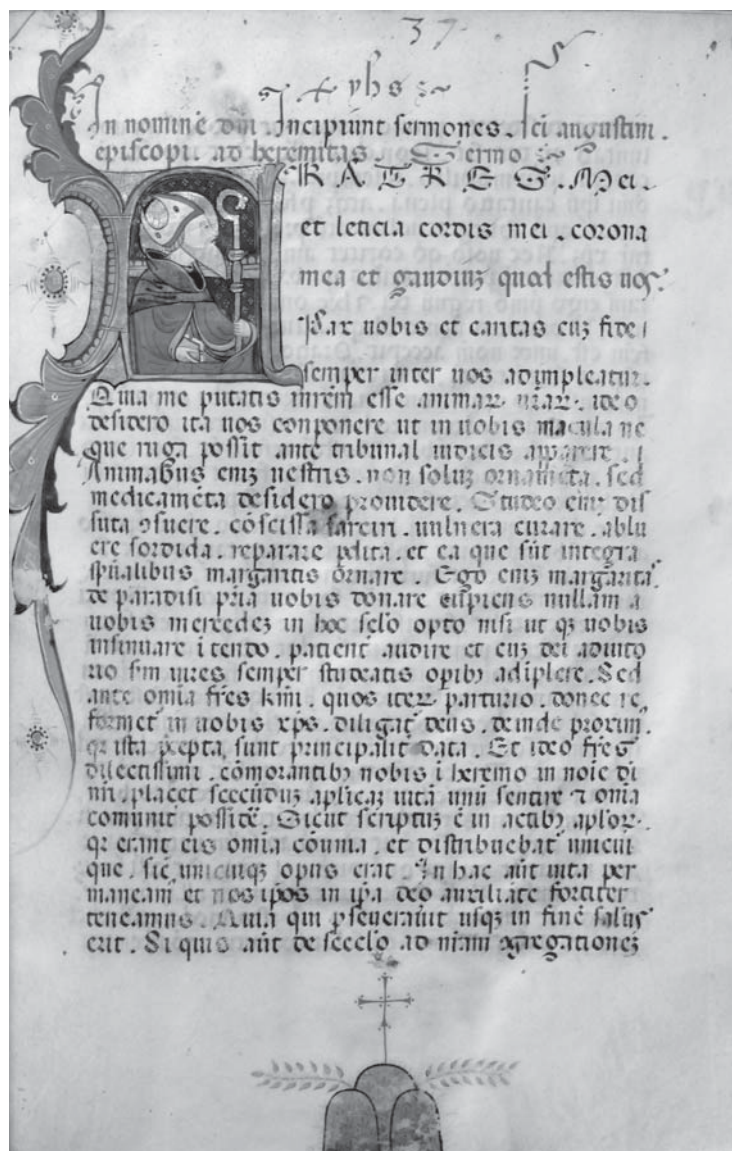
GRUPO PINAKES (GUILLERMO BLÁZQUEZ MARTÍN, CARLOS CRISTÓBAL MEDINA GARCÍA, ALBERTO RAMOS ALONSO)

Se ha publicado recientemente en la revista "The Annals of Applied Statistics" un trabajo del proyecto DEEDS (Documents of Early England Data Set) en el que se desarrollaba un instrumento para la datación de manuscritos. Desde su creación en 1975 el proyecto DEEDS del Centro de Estudios Medievales de la Universidad de Toronto, no solo se ha preocupado de la digitalización de los documentos con los que trabaja, sino que con el avance de las tecnologías se ha explotado al máximo la información que en ellos se encuentra a través del análisis informático con la creación de una base de datos de personas, lugares, ocupaciones y propiedades reseñadas en ellos. De esta forma se ha obtenido una base de datos de gran utilidad para los investigadores. En la actualidad el número de documentos con los que trabaja el Centro es 10.000 aproximadamente, de los cuales 3353 han sido utilizados para la implementación del instrumento.

Las fechas de estos documentos están acotadas entre el siglo XI y principios del siglo XIV en Inglaterra, para ser más precisos entre 1066 y 1307. La mayor parte de ellos no están datados pues por los cambios introducidos en la chancillería real en el reinado de Guillermo I "El Conquistador" se mandó redactar los documen-

tos de la época sin marca cronológica alguna, siguiendo la tradición normanda. Estos manuscritos de tipo jurídico se redactaban en latín, lengua oficial de la administración en aquel momento. En su mayoría tratan de derechos de propiedad, sobre todo ingleses y en menor medida galeses y escoceses. Ante tal situación, el proyecto DEEDS, con Michael Gervers a la cabeza, decide abordar esta problemática con la esperanza de determinar su fecha, beneficiando así a cerca de un millón de documentos del mismo periodo. Pero, ¿cómo han logrado crear un sistema que permita la datación de manuscritos? DEEDS ha desarrollado unos algoritmos que realizan la datación de forma automática. Para ello planteó varios métodos posibles basados en coincidencias con manuscritos previamente datados, en la aparición de expresiones o palabras que solo se usaron en un periodo de tiempo bastante delimitado o en la predicción estadística mediante regresión. Teniendo como objetivo alcanzar una precisión de datación que permita acotar la fecha de los documentos en torno a unos +/- 3 años.

Estos métodos de datación también pueden ser adaptados para su utilización en documentos de cualquier otro periodo y también para otro



tipo de aplicaciones como la identificación de los autores de los textos o detectar falsificaciones. Destacando entre ellos el método de prevalencia máxima como el de mayor fiabilidad. ■