
Catalogación automatizada de incunables con la herramienta CACIN

Hala Neji

Javier Noguerras-Iso

Javier Lacasta

Instituto de Investigación en Ingeniería de

Aragón (I3A) –

Universidad de Zaragoza

Manuel José Pedraza-Gracia

Instituto de Patrimonio y Humanidades (IPH) –

Universidad de Zaragoza

Palabras clave: Análisis y reconocimiento de documentos; aprendizaje automático; Incunables; metadatos; catálogo

Keywords: Document analysis and recognition; machine learning; incunabula; metadata; catalog

El estudio de la producción bibliográfica de un momento y ámbito específico ofrece una panorámica del conocimiento y el pensamiento en ese momento y ámbito. Identificar, por tanto, una edición coadyuva a la mejora de nuestro conocimiento y proporciona valores, incluso económicos, a los ejemplares de la misma.

En el periodo incunable y postincunable, el diseño y la elaboración material de los tipos metálicos para componer un libro es un proceso manual que lleva a cabo el propio taller de impresión (Zappella, 2001). Esta característica condiciona tanto el aspecto como la morfología de los incunables, porque cada taller posee unos tipos particulares que dan como resultado unos caracteres específicos.

Desde el inicio de la incunabulología, en el siglo XVII, se persiguió más la localización de ediciones y ejemplares que la identificación de sus productores (Bradshaw, 1889; Barbieri, 2008). Sin embargo, los problemas derivados de la existencia de mútilos y membra disjecta, de la propia idiosincrasia de la producción del impreso, de los incunables sine notis, y de la ausencia de una sistematización científica dieron lugar a la elaboración de una metodología identificativa. A finales del siglo XIX, Robert Proctor aplica un método cuantitativo destinado a la identificación del taller que imprimió el incunable, mediante la medida de la huella que dejan en el papel veinte líneas de texto impreso con la misma fundición desde un punto de la línea uno hasta el mismo punto de la línea veintiuno, añadiendo tras esa dimensión una “R” para un tipo redondo o romano o una “G” para uno gótico (Proctor, 1905). De esta manera un tipo adquiere tanto

una identificación como una denominación con sentido: 84R o 162G, por ejemplo. Unos años después Konrad Haebler aplica, además, un método comparativo sustentado en la morfología de las dos letras más significativas: “M” para tipografía gótica y “Q” para la romana. Entre 1905 y 1924 publicó un muestrario de casi 4.000 “M”, “Qu” y otros signos; el *Typenrepertorium der Wiegendrucke* disponible en la actualidad a través del portal web de la Biblioteca Estatal de Berlín (Haebler, 1905-1924). La combinación de ambos sistemas ha dado lugar a una metodología que permite: a) denominar una tipografía; b) identificarla; y c) adscribirla a un taller específico (Rial Costas, 2012).

El objetivo de esta investigación es que el método Proctor-Haebler pueda ser aplicado de forma automática a ejemplares incunables y postincunables, evitando errores, superando la identificación en remoto y adecuándolo a muestras que presenten dificultades de aplicación.

En 2020 creamos un primer prototipo de software, desarrollado en Python, para automatizar el método Proctor-Haebler para la identificación de tipografías góticas (Lacasta et al., 2022). Este software incluye un algoritmo para detectar bloques de líneas continuas que se apoya en el uso de la biblioteca Tesseract para la segmentación de líneas y caracteres. El software también integra dos redes neuronales convolucionales para la detección de letras “M” y la identificación de la forma específica de “M” que se está utilizando. Con la información de la altura del bloque y la forma de la letra “M”, el software consulta un almacén de tripletas RDF, que replica los contenidos del *Typenrepertorium der Wiegendrucke*, para recuperar las tipografías que se corresponden con la altura y forma de “M” detectada, y los talleres de impresión que crearon esa tipografía.

Durante 2024 hemos mejorado este prototipo inicial, que solo podía ser utilizado en un entorno de pruebas y con conocimientos medio-altos de configuración de sistemas informáticos, para integrarlo en un conjunto de herramientas más robusto y escalable denominado CACIN, acrónimo de “Computer Aided Cataloguing of INcunabula” (Neji et al., 2025). CACIN se compone de varias herramientas para administradores y usuarios finales que colaboran con el objetivo de catalogar incunables con asistencia automatizada para la identificación de tipografías. La Figura 1 muestra un diagrama de despliegue con la arquitectura de CACIN donde se pueden observar tres partes diferenciadas: el software de entrenamiento (Training Software); la herramienta para los administradores (Admin frontend/backend); y la herramienta para los usuarios finales (Final users frontend/backend). Tanto la herramienta para los administradores como la herramienta para los usuarios finales son aplicaciones web organizadas en tres niveles: un frontend con la interfaz de usuario; un backend con la lógica de negocio y servicios de apoyo; y una base de datos.

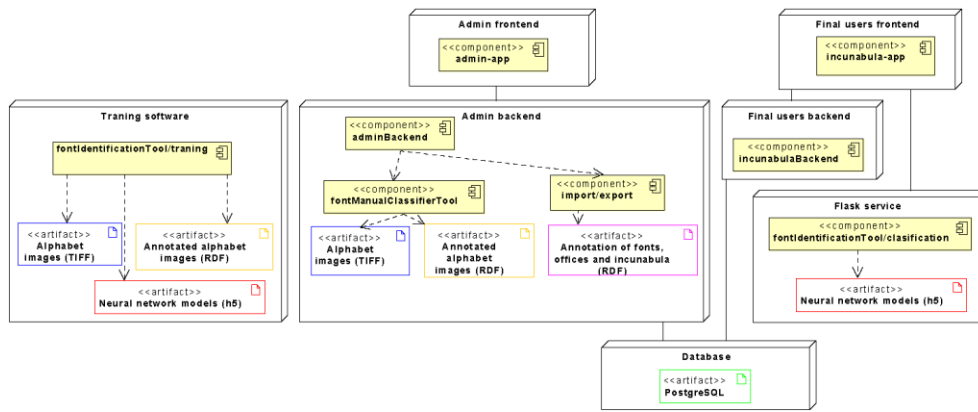


Figura 1. Diagrama de despliegue con la arquitectura del software. Fuente: Neji et al., 2025.

La herramienta Training Software se encarga de entrenar las redes neuronales para la detección de las letras “M” y las formas concretas de “M” tomando como entrada las imágenes del alfabeto de las distintas tipografías que se quieren reconocer y una anotación de la posición de la letra que se corresponde o no con la letra “M” en cada imagen. Para facilitar la anotación de las imágenes de cada alfabeto, la herramienta de administración proporciona una aplicación web donde que permite segmentar de forma automática los caracteres en las imágenes de los alfabetos (facilitados por el Typenrepertorium der Wiegendrucke para la mayoría de las tipografías), modificar manualmente esta segmentación si no ha sido correcta, y marcar las áreas correspondientes a letras especiales. La herramienta de administración también permite importar y exportar la información del Typenrepertorium der Wiegendrucke en formato RDF y almacenarla en una base de datos que será consultada posteriormente por las herramientas para los usuarios finales.

CACIN: Computer Aided Cataloguing of Incunabula

Lista de imprentas

Información de la imprenta

Buscar imprentas por nombre.

Filtros

Ubicación: Valencia X

Año de fundación: Año de fundación

Año de cierre: Año de cierre

Limpiar filtros Buscar

ID	Nombre de la imprenta	Nombre alternativo	Año de fundación	Año de cierre	Ubicación
of1083	Drucker des Officium BMV (GW 13442)				Valencia
of1088	Miguel Albert				Valencia
of1657	Philipp Vizlant				Valencia
of1091	Christoph Cofmann		1496	1500	Valencia
of1087	Lope de la Roca	Lope de la Roca und Peter Trincer	1495	1497	Valencia
of1084	Nikolaus Spindeler		1490	1500	Valencia
of1086	Johann Rosembach				Valencia
of1456	Jaime de Vila				Valencia
of1080	Lambert Palmart	Lambert Palmart für Gabriel Luys de Arinyo y Lope	1474	1494	Valencia
of1085	Peter Hagembach und Leonhard Hutz	Peter Hagembach und Leonhard Hutz für Jacobus de Villa	1493	1496	Valencia

Número de resultados a mostrar: 10

1 2

Figura 2. Interfaz de consulta de talleres de impresión de la herramienta para usuarios finales.

La herramienta orientada a los usuarios finales es una aplicación web para la catalogación de incunables que además de las típicas operaciones de un catálogo para el alta, baja, modificación y eliminación de registros, incorpora también funcionalidades para seleccionar una imagen representativa de una página de un incunable e invocar al software de reconocimiento de tipografías (nodo Flask service en la Figura 1). En la Figura 2 se puede ver un ejemplo de la interfaz de usuario de la aplicación para consultar los talleres de impresión disponibles en la base de datos y ubicados, por ejemplo, en Valencia. En la Figura 3 se puede observar un ejemplo de la de la interfaz de usuario para la edición de un incunable donde se ha invocado al software de reconocimiento y se muestran las posibles tipografías y talleres de impresión que se podrían asignar, en este caso el taller de Paul Hurus en Zaragoza y una tipografía gótica con altura de 104 mm para bloques de 20 líneas y la forma “17E” para la letra M.

Para demostrar la funcionalidad ofrecida por CACIN a los expertos en incunables, se ha instalado la aplicación web de catalogación en la dirección <https://cacin.iaaa.es/>. La base de datos del catálogo se ha alimentado con la información de 1.626 talleres de impresión, 6.169 tipografías, y 577 incunables. En cuanto a las redes neuronales para la identificación de letras góticas especiales integradas en este despliegue de la aplicación web, estas redes han sido entrenadas para el reconocimiento de 107 tipografías góticas que utilizan 39 tipos distintos de formas de la letra “M”. En el futuro, se estudiará cómo conseguir que estas redes puedan reconocer un mayor número de tipografías, y escalar el sistema para el posible reconocimiento de las tipografías romanas utilizando la letra “Q” como letra especial.

Agradecimientos

Esta publicación es parte del proyecto de I+D+i T59_23R financiado por el Gobierno de Aragón.

Bibliografía

- BARBIERI, E. (2008), *Haebler contro Haebler: appunti per una storia dell'incunabolistica novecentesca*, Milano, Università Cattolica del Sacro Cuore.
- BRADSHAW, H. (1889), *Collected Papers*, Cambridge, Cambridge University Press.
- HAEBLER, K. (1905–1924), *Typenrepertorium der Wiegendrucke*, Leipzig, Verlag von Rudolf Haupt, 5 vols. Disponible en línea: <https://tw.staatsbibliothek-berlin.de/>
- LACASTA, J., NOGUERAS-ISO, J., ZARAZAGA-SORIA, F. J., & PEDRAZA-GRACIA, M. J. (2022), «Tracing the origins of incunabula through the automatic identification of fonts in digitised documents», *Multimedia Tools and Applications*, 81(28), pp. 40977–40991.
- NEJI, H., NOGUERAS-ISO, J., LACASTA, J., & PEDRAZA-GRACIA, M. J. (2025), «CACIN: Computer Aided Cataloguing of Incunabula», preprint, SSRN. <https://dx.doi.org/10.2139/ssrn.5201154>
- PROCTOR, R. (1905), *Bibliographical Essays*, introducción de A. W. Pollard, New York, Burt Franklin.
- RIAL COSTAS, B. (2012), «El Sistema Proctor-Haebler y el estudio de las letrerías en las impresiones góticas incunables», en N. FERNÁNDEZ RODRÍGUEZ & M. FERNÁNDEZ FERREIRO (coords.), *Literatura medieval y renacentista en España: líneas y pautas*, Salamanca, Universidad de Salamanca / SEMYR, pp. 855–864. En línea: <http://repositoriodigital-lasemyr.es/index.php/rd-ls/catalog/book/51>
- ZAPPELLA, G. (2001), *Il libro antico a stampa: struttura, tecniche, tipologie, evoluzione*, Milano, Editrice Bibliografica, 2 vols.