



ACELERANDO LA TRANSICIÓN DE LOS REGISTROS DIGITALES A LOS ARCHIVOS NACIONALES MEDIANTE LA INTELIGENCIA ARTIFICIAL

"El proyecto "AI for Digital Selection" tiene como objetivo conocer mejor las herramientas de IA que podrían utilizarse para llevar a cabo la valoración y selección de los "montones digitales" de documentos, correos electrónicos, conjuntos de datos y otros tipos de información que se conservan en toda la Administración."

Los Archivos Nacionales

Más detalles estarán disponibles en el próximo informe redactado por los Archivos Nacionales titulado: "El uso de la IA para la selección digital en la Administración: una evaluación de las soluciones del mercado que utilizan el aprendizaje automático para seleccionar los documentos de archivo digitales para su conservación permanente".



RETO

El gobierno del Reino Unido necesitaba encontrar una forma eficaz de gestionar el primer lote de documentos del inicio de la era digital. Estos documentos debían depositarse en los Archivos Nacionales en virtud de la norma de 20 años para la conservación de registros de valor histórico.



SOLUCIÓN

Iron Mountain aprovechó la última tecnología en Inteligencia Artificial (IA) y Aprendizaje Automático (ML) para entrenar a su sistema en el reconocimiento de registros candidatos a la preservación permanente, la detección de duplicados para su eliminación, la extracción de entidades y el análisis de archivos.



RESULTADOS

Las capacidades de IA y ML permitirán a los departamentos gubernamentales transferir con confianza los registros de forma mucho más rápida y eficiente que trabajando con el mismo volumen de materiales sin recurrir a estas herramientas.

EL RETO DEL GOBIERNO

“El primer lote de documentos del gobierno del Reino Unido del inicio de la era digital está a punto de ser liberado al público bajo la regla de los 20 años. El uso de la IA y el ML permitirá a los departamentos gubernamentales transferir con confianza los registros de manera mucho más rápida y eficiente que trabajar con los mismos volúmenes de material sin estas herramientas”.

Ed Irving, Director de Desarrollo de Negocio de la Administración Central, Iron Mountain

Todos los departamentos gubernamentales del Reino Unido son responsables de revisar sus registros. Los que se identifican como de valor histórico deben ser transferidos a los Archivos Nacionales una vez que cumplen 20 años de antigüedad, de acuerdo con los requisitos de la Ley de Registros Públicos de 1958. Un paso clave en este proceso es la Revisión de Sensibilidad. Esto asegura que los registros transferidos no contienen ninguna información sensible que pueda causar daño a otros o riesgo para la reputación del gobierno del Reino Unido.

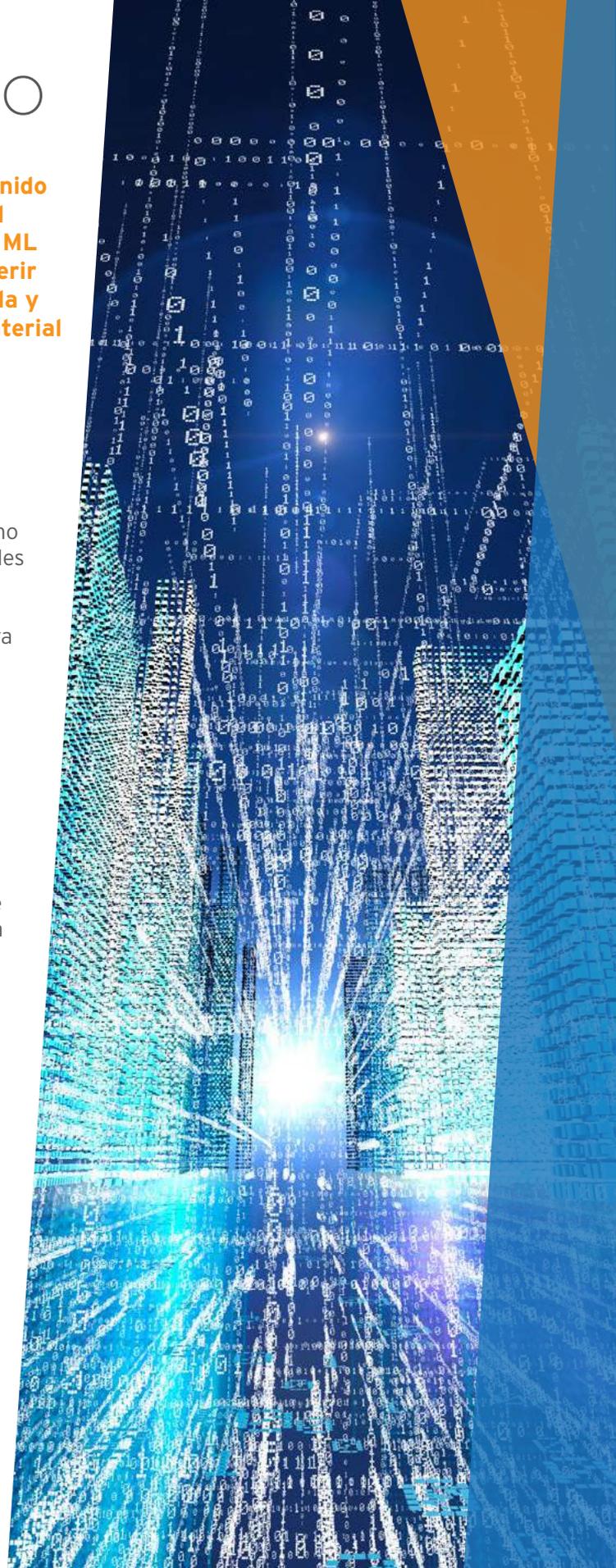
El inicio de la era digital

La mayoría de los principales departamentos gubernamentales del Reino Unido empezaron a trabajar en formato digital en torno al 2004; algunos departamentos y organismos sólo han existido en la era digital (aunque siguen creando algo de papel). Esto significa que muchos de los registros sujetos a la Ley de Registros Públicos ahora sólo existen en formato digital. Estos incluyen una amplia gama de información, desde conjuntos de datos estructurados hasta correos electrónicos, documentos de Word y hojas de cálculo.

La escala y variedad de esta información digital significa que la identificación y selección de los registros de valor es casi imposible utilizando el enfoque tradicional de “sólo mano de obra humana”. Los departamentos gubernamentales están empezando a estudiar 3 cuestiones clave:

- **¿Pueden las tecnologías de inteligencia artificial (IA) o de aprendizaje automático (ML) acelerar y simplificar el proceso de clasificación y revisión de los archivos digitales?**
- **¿Con qué eficacia puede esta tecnología clasificar los registros de valor y los datos efímeros o los archivos duplicados que no tienen valor?**
- **¿Es factible ingerir todo en un solo sistema una vez seleccionado?**

En apoyo del reto gubernamental, los Archivos Nacionales han puesto en marcha el proyecto *AI for Digital Selection*. El objetivo del proyecto es conocer mejor las herramientas de IA existentes que podrían utilizarse para realizar la evaluación y la selección de los “montones digitales” de documentos, correos electrónicos, conjuntos de datos y otros tipos de información que posee la Administración.





DESARROLLO DE IA PARA LA SELECCIÓN DIGITAL

“Dada la importancia y la complejidad del reto, los Archivos Nacionales decidieron llevar a cabo un programa piloto para garantizar que las herramientas de procesamiento fueran capaces de alcanzar los niveles de precisión y control requeridos. También necesitaba un socio con un historial probado de gestión segura y eficiente de proyectos grandes y complejos.”

Los Archivos Nacionales llevaron a cabo una revisión de las herramientas disponibles para identificar entre 3 y 5 con el fin de probarlas en profundidad con un conjunto de sus propios registros corporativos. El plan consistía en averiguar cómo se comportan las herramientas a la hora de identificar los registros que deben ser seleccionados para su conservación permanente y los que no. Además, los Archivos Nacionales querían ayudar a los departamentos del gobierno en el uso de la IA para la selección. Más concretamente a identificar dónde se podían incorporar estas técnicas en el proceso o flujo de trabajo de la selección de documentos de archivo digitales para su transferencia a los Archivos Nacionales.

Una relación bien establecida

Iron Mountain ya tiene una larga y consolidada relación con los Archivos Nacionales, así como con algunos de los departamentos más grandes del gobierno, habiendo facilitado la transferencia de registros en papel a escala durante varios años. La empresa también ofrece los servicios InSight™, que incluyen el procesamiento inteligente de documentos combinado con el análisis de contenidos, la gestión de datos y experiencia en la gobernanza de la información con Aprendizaje Automático (ML) e inteligencia artificial (AI).

Iron Mountain fue seleccionada para participar en el proyecto *AI for Digital Selection*, con el fin de comprender la eficacia de la IA en el proceso de selección digital.

Gestión de múltiples formatos de archivo

Las herramientas de Iron Mountain son capaces de manejar una amplia selección de formatos de medios, lo que era esencial ya que había más de 100 tipos de archivos diferentes para procesar en el estudio de prueba de concepto, incluyendo audio, vídeo y documentos de texto. Algunos de los formatos más oscuros fueron excluidos de la prueba, pero el número real de expedientes que quedaron fuera del estudio fue pequeño: menos del 10%. Los archivos incluidos en la prueba de concepto se cargaron en la plataforma de servicios de contenidos de Iron Mountain para su análisis.

¿EN QUÉ CONSISTIÓ LA SOLUCIÓN?

Determinar lo que hay que transferir

Como parte del proyecto, los Archivos Nacionales proporcionaron a Iron Mountain conjuntos de datos etiquetados y no etiquetados para testar las capacidades de IA intrínsecas a InSight™ en la identificación de registros relevantes para los criterios de selección.

Eliminación de datos sensibles

Incluso en el caso de los archivos que se ponen a disposición del público, tiene que haber una fase en la que la información sensible que no está autorizada para ser publicada pueda ser eliminada. En un mundo basado en el papel, la redacción de la información es un proceso físico que se lleva a cabo como parte de una revisión dirigida por personas: los documentos se marcan o editan a mano. Las herramientas de tecnología forense digital adoptan un enfoque diferente: ofuscan palabras o frases, codificando el contenido cuando sea necesario.

Pasos para el éxito

Para la prueba, Iron Mountain cargó primero los 17.000 documentos de prueba en el almacenamiento en la nube de Google. Los documentos se procesaron con tecnología de reconocimiento óptico de caracteres para hacerlos totalmente reconocibles. A continuación, la plataforma de procesamiento inteligente de documentos InSight™ los clasificó en 20 categorías utilizando el procesamiento del lenguaje natural (PNL), un proceso de software que puede descifrar el contenido de un documento y los matices contextuales de la lengua utilizada. Esto permitió a la plataforma extraer con precisión la información contenida en los archivos y crear asociaciones en el conjunto de la muestra para garantizar que una búsqueda fácil y precisa fuese posible.

Las capacidades de aprendizaje automático de InSight™ permitieron al equipo del proyecto entrenar el modelo en un proceso iterativo a lo largo del proyecto. Al final, la herramienta obtuvo una puntuación F1 superior al 85%. También se obtuvieron los siguientes resultados:

- Se identificaron los duplicados para eliminarlos
- Se identificaron los registros candidatos a la conservación permanente
- Se extrajeron entidades que incluyen organizaciones y personas
- Se realizaron análisis de expedientes, incluido el resumen de contenido, el resumen de edad, etc.
- Puntuaciones medias de Precisión y Recall

Prueba de concepto con éxito

InSight™ proporcionó el nivel de funcionalidad requerido, la clasificación de documentos y la detección de duplicados. Iron Mountain se complace en apoyar a los Archivos Nacionales con este programa de interés nacional. La solución puede utilizarse ahora en toda la administración para acelerar la transición de los documentos de todos los departamentos gubernamentales a través de vías directas de adquisición de marcos.

**DESCUBRE OTRAS
FORMAS EN LAS QUE
TE PODEMOS AYUDAR**

VER
Otros casos de Iron
Mountain como éste



CONTÁCTANOS
800 1 IRONMEX
ironmountain.com/mx

**ÚNETE A LA
CONVERSACIÓN**

Síguenos en las redes
sociales y mantente al día
con las ideas y los avances
empresariales y tecnológicos

2022 Iron Mountain (UK) PLC. Todos los derechos reservados. Iron Mountain, Iron Mountain InSight, Iron Mountain Connect y el diseño de la montaña son marcas comerciales o registradas de Iron Mountain Incorporated en los Estados Unidos y otros países y se utilizan bajo licencia. Todas las demás marcas comerciales y registradas son propiedad de sus respectivos dueños.

