



DIGITALE AKTEN DANK KI SCHNELLER IN DAS NATIONALARCHIV ÜBERFÜHREN

„Im Rahmen des Projekts *AI for Digital Selection* sollte ermittelt werden, welche bereits vorhandenen KI-Tools verwendet werden könnten, um das „digitale Chaos“ aus Dokumenten, E-Mails, Datensätzen und anderen Informationstypen in verschiedenen Ministerien zu prüfen und eine Auswahl zu treffen.“

The National Archives UK

Details werden im angekündigten Bericht „Using AI for Digital Selection in Government: an evaluation of marketplace solutions using machine learning to select digital records for permanent preservation“ des Nationalarchivs UK veröffentlicht.



HERAUSFORDERUNG

Großbritannien suchte nach einer effizienten Methode für den Umgang mit den ersten digitalen Dokumenten, die nach der 20-Jahre-Regel zur Bewahrung historisch wertvoller Akten in das Nationalarchiv überführt werden mussten.



LÖSUNG

Mit den neuesten Entwicklungen im Bereich der künstlichen Intelligenz (KI) und des maschinellen Lernens (ML) trainierte Iron Mountain seine Systeme so, dass potenziell historisch wertvolle Akten identifiziert, zu entsorgende Duplikate erkannt, Entitäten extrahiert und Dateien analysiert werden können.



ERGEBNISSE

Mit KI und ML können die Ministerien ihre Unmengen an Akten viel schneller und effizienter aussortieren und überführen, als dies mit manuellen Prozessen der Fall wäre.

EINE HERAUSFORDERUNG FÜR DIE REGIERUNG

„Nach der 20-Jahre-Regel müssen in Großbritannien nun erstmals digitale Regierungsdokumente für die Öffentlichkeit freigegeben werden. Mit dem richtigen Einsatz von KI- und ML-Lösungen können die Ministerien ihre Unmengen an Akten viel schneller und effizienter aussortieren und überführen, als dies mit manuellen Prozessen der Fall wäre.“

Ed Irving, Business Development Director, Central Government, Iron Mountain

Alle britischen Ministerien sind dafür verantwortlich, regelmäßig ihre Akten zu prüfen. Diejenigen, bei denen ein historischer Wert festgestellt wird, müssen nach dem Public Records Act von 1958 (PRA) in das Nationalarchiv (The National Archives) überführt werden, wenn sie ein Alter von 20 Jahren erreichen.

Ein wichtiger Schritt in diesem Prozess ist die Sensibilitätsprüfung. Dabei wird sichergestellt, dass die überführten Akten keine Informationen enthalten, die anderen Schaden zufügen oder den Ruf der britischen Regierung beschädigen könnten.

Der Anbruch des digitalen Zeitalters

Um 2004 fingen die britischen Ministerien an, auf digitales Arbeiten umzustellen, und manche Ministerien und Behörden sind so jung, dass sie von Anfang an größtenteils digital waren. Das bedeutet: Viele der jetzt dem PRA unterliegenden Akten existieren nur in digitaler Form. Dazu gehören ganz unterschiedliche Informationen, von strukturierten Datenbeständen über E-Mails bis hin zu Word-Dokumenten und Excel-Tabellen.

Aufgrund der Menge und Vielfalt dieser digitalen Informationen ist es fast unmöglich, allein mit den bisherigen manuellen Methoden herauszufinden, welche Akten wertvoll sind. Deshalb stellen sich nun drei zentrale Fragen:

- **Können künstliche Intelligenz (KI) bzw. maschinelles Lernen (ML) die Klassifizierung und Prüfung digitaler Akten beschleunigen und vereinfachen?**
- **Wie gut kann diese Technologie zwischen wertvollen Akten und wertlosen kurzlebigen Daten oder Duplikaten unterscheiden?**
- **Ist es möglich, die gesamte Auswahl in ein zentrales System einzuspeisen?**

Um diese Herausforderung zu meistern, rief das Nationalarchiv das Projekt *AI for Digital Selection* ins Leben. Damit sollte ermittelt werden, welche bereits vorhandenen KI-Tools verwendet werden könnten, um das „digitale Chaos“ aus Dokumenten, E-Mails, Datensätzen und anderen Informationstypen in den verschiedenen Ministerien zu prüfen und eine Auswahl zu treffen.





KI-ENTWICKLUNG FÜR DIE DIGITALE AUSWAHL

„Angesichts der Bedeutung und Komplexität der Herausforderung entschied sich das Nationalarchiv für ein Pilotprogramm, um sich zu vergewissern, dass die Tools die erforderliche Genauigkeit und Kontrolle bieten können. Außerdem wurde ein Partner mit einer nachgewiesenen Erfolgsbilanz bei der sicheren und effizienten Durchführung großer, komplexer Projekte benötigt.“

Das Nationalarchiv prüfte die verfügbaren Tools, um drei bis fünf auszuwählen, die anschließend mit einer Stichprobe eigener Dokumente intensiver getestet werden sollten: Wie gut würden die Tools die für eine dauerhafte Aufbewahrung infrage kommenden Akten erkennen? Außerdem wollte das Nationalarchiv die Ministerien bei der Nutzung von KI für die Auswahl unterstützen. Dafür musste ermittelt werden, wo genau diese Techniken im Auswahlprozess eingesetzt werden können.

Jahrelange Geschäftsbeziehung

Iron Mountain hatte schon seit mehreren Jahren für das Nationalarchiv sowie einige Ministerien im großen Maßstab [Papierakten ausgelagert](#). Außerdem bietet das Unternehmen für die intelligente Dokumentenverarbeitung die [Lösung InSight™](#) an, eine Content Services Platform.

Iron Mountain wurde ausgewählt, sich am Projekt *AI for Digital Selection* zu beteiligen, um herauszufinden, wie effektiv KI im digitalen Auswahlprozess sein kann.

Umgang mit mehreren Dateiformaten

Die Tools von Iron Mountain unterstützen verschiedenste Medienformate, was bei dieser Proof-of-Concept-Studie äußerst wichtig war: Mehr als 100 verschiedene Dateitypen mussten verarbeitet werden, darunter Audio- und Videodateien sowie Textdokumente. Einige besonders exotische Formate wurden in dieser Phase noch nicht berücksichtigt, doch diese machten nur 10 % der Dateien aus. Die in die Studie einbezogenen Dateien wurden anschließend zur Analyse in die [Content Services Platform](#) von Iron Mountain geladen.

WELCHE LEISTUNGEN UMFASSTE DIE LÖSUNG?

Zu überführende Akten identifizieren

Für das Projekt stellte das Nationalarchiv Iron Mountain Datenbestände mit und ohne Labeling bereit. Anhand dieser Daten wurde nachgewiesen, dass die KI-Funktionalitäten von InSight™ in der Lage sind, Akten nach bestimmten Auswahlkriterien zu identifizieren.

Sensible Daten entfernen

Auch bei Dateien, die für die Öffentlichkeit freigegeben werden, muss es einen Schritt zum Entfernen sensibler, vertraulicher Informationen geben. Auf Papier ist das Zensieren ein physischer Prozess, bei dem man Informationen im Rahmen einer Prüfung manuell schwärzt. IT-Forensiktools gehen anders vor - Wörter oder Sätze werden verschleiert, bei Bedarf wird mit Scrambling gearbeitet.

Schritte zum Erfolg

Für das Pilotprojekt lud Iron Mountain zuerst die 17.000 Testdokumente in einen Storage-Bucket in der Google Cloud. Die Dokumente wurden mit optischer Texterkennung (OCR) bearbeitet, damit sie vollständig durchsuchbar sind. Die InSight™ Plattform für intelligente Dokumentenverarbeitung klassifizierte sie anschließend in 20 vordefinierte Kategorien. Dafür wurde NLP (Natural Language Processing) eingesetzt, ein Softwareprozess, der den Inhalt eines Dokuments und die Textzusammenhänge erfasst. Dadurch konnte die Plattform die in den Dateien enthaltenen Informationen exakt extrahieren und zwischen den Dateien in der Stichprobe Verknüpfungen herstellen, um eine aussagekräftige Suche zu ermöglichen.

Mit den Machine-Learning-Funktionalitäten von InSight™ konnte das Projektteam in einem iterativen Prozess nach und nach das Modell trainieren. Am Ende erreichte das Tool einen F1-Score von über 85 %. Weitere Ergebnisse:

- Zu entsorgende Duplikate wurden erkannt.
- Für die dauerhafte Aufbewahrung infrage kommende Akten wurden ermittelt.
- Entitäten wie Namen von Organisationen und Menschen wurden extrahiert.
- Dateien wurden analysiert: Inhaltzusammenfassung, Alter usw.
- Die durchschnittliche Genauigkeit und Trefferquote wurden ermittelt.

Erfolgreicher Proof of Concept

InSight™ konnte alles liefern, was in Bezug auf Funktionalität, Dokumentenklassifikation und Duplikaterkennung erforderlich war. Iron Mountain freut sich, The National Archives bei diesem Programm von nationalem Interesse unterstützen zu dürfen. Es wurde ein Rahmenvertrag abgeschlossen, über den nun ganz unkompliziert alle Ministerien die Lösung nutzen können, um Dokumente zügig zu überführen.

MEHR ERFAHREN

[Mehr lesen über die Content Services Platform Insight](#)
[Weitere Kundenprojekte erkunden](#)

Folgen Sie uns

[IronMountain.de](#)
[IronMountain.co.at](#)
[IronMountain.ch](#)