



DIGITALE INFORMATIE SNELLER ARCHIVEREN DANKZIJ ARTIFICIAL INTELLIGENCE

“De eerste lading Britse overheidsdocumenten van twintig jaar oud staat op het punt om openbaar te worden. Dankzij AI en machine learning kunnen overheidsinstanties dezelfde hoeveelheid materiaal veel sneller en efficiënter overdragen aan The National Archives.”

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

Meer details staan in het rapport van The National Archives: ‘Using AI for Digital Selection in Government: an evaluation of marketplace solutions using machine learning to select digital records for permanent preservation.’



UITDAGING

De Britse overheid wilde efficiënt omgaan met de eerste lading documenten uit het begin van het digitale tijdperk. Volgens de 20-jaar-regel voor het bewaren van gegevens met historische waarde moesten deze documenten naar The National Archives.



OPLOSSING

Iron Mountain maakte gebruik van de nieuwste ontwikkelingen in artificial intelligence (AI) en machine learning om het systeem te trainen in het herkennen van permanent te bewaren documenten en te verwijderen duplicaten, het extraheren van entiteiten en het analyseren van bestanden.



RESULTAAT

Dankzij AI en ML kunnen overheidsinstanties dezelfde hoeveelheid documenten veel sneller en efficiënter overdragen aan The National Archives.



DE UITDAGING VOOR DE BRITSE OVERHEID

"Het doel van het project *AI for Digital Selection* was om meer te weten te komen over bestaande AI-tools die gebruikt kunnen worden voor de beoordeling en selectie uit de 'digitale stapel' documenten, e-mails, datasets en andere soorten informatie binnen de Britse overheid."

The National Archives

Alle Britse ministeries zijn verantwoordelijk voor het controleren van hun documenten. Hebben documenten historische waarde? Dan moeten deze zodra ze twintig jaar oud zijn, worden overgedragen aan The National Archives (TNA), volgens de bepalingen van de Public Records Act 1958. Een belangrijke stap in dit proces is een sensitiviteitsonderzoek. Dit garandeert dat de overgebrachte archiefstukken geen informatie bevatten die anderen kan schaden of die de reputatie van de Britse overheid in gevaar kan brengen.

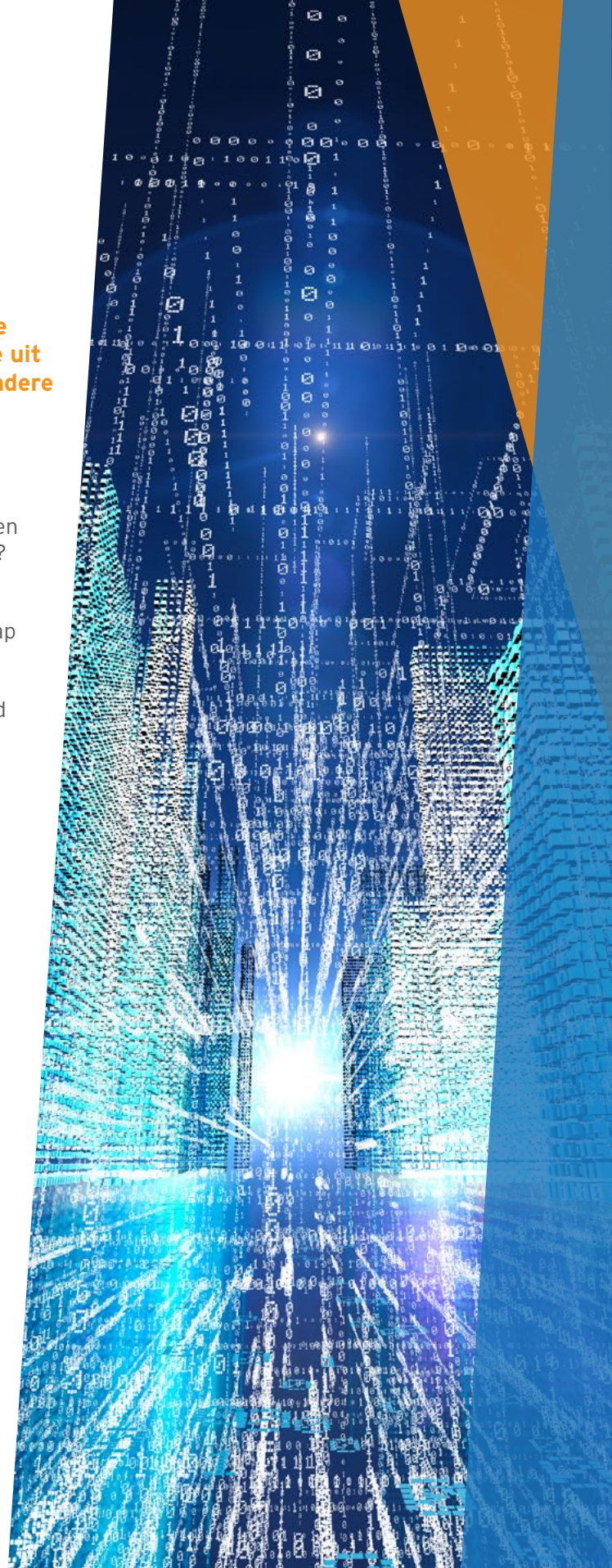
Het begin van het digitale tijdperk

De meeste Britse ministeries gingen rond 2004 digitaal werken. Sommige departementen en instanties hebben zelfs alleen maar in een digitaal tijdperk bestaan. Dit betekent dat veel van de gegevens die onder de PRA vallen nu alleen nog digitaal zijn. Het gaat om een breed scala aan informatie: van gestructureerde datasets tot e-mails, tekstdocumenten en spreadsheets.

Door de omvang en de verscheidenheid van deze digitale informatie is het bijna onmogelijk om alleen met menselijke inspanning alle waardevolle bestanden te identificeren en te selecteren. Ministeries zitten daarom met deze drie vragen:

- **Kunnen technologieën als artificial intelligence en machine learning het classificatie- en beoordelingsproces van digitale gegevens versnellen en vereenvoudigen?**
- **Hoe goed is deze technologie in het sorteren van waardevolle bestanden en vluchtige data of duplicaten zonder waarde?**
- **Is het mogelijk om na selectie alles in één systeem op te nemen?**

Ter ondersteuning van de uitdaging bij de overheid startte TNA het project *AI for Digital Selection*. Het doel was om meer te weten te komen over bestaande AI-tools die gebruikt kunnen worden voor de beoordeling en selectie uit de 'digitale stapel' documenten, e-mails, datasets en andere soorten informatie binnen de Britse overheid.





AI EN DIGITALE SELECTIE

“Om er echt zeker van te zijn dat de verwerkingstools de vereiste niveaus van nauwkeurigheid en controle konden bereiken, besloot The National Archives een pilot uit te voeren. Ze hadden ook een partner nodig met aantoonbare ervaring in het veilig en efficiënt aanpakken van grote, complexe projecten.”

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

The National Archives zocht AI-tools in de markt en onderwierp er drie tot vijf aan een grondige test met een paar van hun eigen bedrijfsdocumenten. Het plan was om na te gaan hoe de tools presteren bij het identificeren van documenten die wel en niet permanent bewaard moeten worden. Daarnaast wilde TNA ministeries helpen bij het gebruik van AI voor selectie. Daarbij werd nagegaan waar deze technieken in het selectieproces van de te archiveren stukken pasten.

Iron Mountain en The National Archives

Iron Mountain ondersteunt al jaren de overdracht van vele papieren documenten voor zowel TNA als enkele van de grootste Britse ministeries. Hierbij wordt ook InSight™ ingezet, waarbij de expertise van Iron Mountain op het gebied van contentanalyse, gegevensbeheer en informatiegovernance wordt gecombineerd met machine learning en artificial intelligence. Om inzicht te krijgen in de effectiviteit van AI in het proces van digitale selectie mocht Iron Mountain meedoen aan het *AI for Digital Selection*-project.

Beheer van meerdere bestandsformaten

In het proof-of-conceptonderzoek moesten meer dan 100 verschillende bestandstypen worden verwerkt, waaronder audio, video en tekstdocumenten. Sommige onbekende formaten werden tijdens de proef buiten beschouwing gelaten, maar uiteindelijk viel minder dan 10% van de bestanden buiten de studie. Vervolgens werden de bestanden in het onderzoek voor analyse in het content services platform van Iron Mountain geladen.

WAT HIELD DE OPLOSSING IN?

Vaststellen wat moet worden overgedragen

Om de AI-mogelijkheden van InSight™ te demonstreren bij het identificeren van relevante bestanden, voorzag The National Archives Iron Mountain van gelabelde en ongelabelde datasets.

Verwijdering van gevoelige gegevens

Zelfs voor bestanden die openbaar worden, moet er een fase zijn waarin gevoelige informatie kan worden verwijderd. In een wereld vol papier is dat redigeren van informatie een menselijke handeling: documenten worden met de hand gemarkeerd of bewerkt. Met digitale forensische technologie gaat dat anders: woorden of zinnen worden waar nodig vervormd en verhuld.

Stappen naar succes

Voor de proef laadde Iron Mountain eerst de 17.000 testdocumenten in Google Cloud-opslag. Om ze volledig doorzoekbaar te maken werden de documenten verwerkt met OCR-technologie (optische tekenherkenning). Het documentverwerkingsplatform InSight™ deelde ze vervolgens in twintig vooraf bepaalde categorieën in via natuurlijke taalverwerking. Dit softwareproces ontcijfert de inhoud van een document en de contextuele nuances van de gebruikte taal. Daardoor werd alle informatie nauwkeurig uit de bestanden gehaald, ontstonden er verbanden tussen de data en werd zoeken makkelijker.

Dankzij de machine learning-mogelijkheden van InSight™ kon het projectteam het model trainen in een herhalend proces. Uiteindelijk behaalde de tool een F1-score van meer dan 85%. Dit waren andere resultaten:

- Duplicaten opgespoord om ze te verwijderen
- Gegevens voor permanente bewaring geïdentificeerd
- Entiteiten waaronder organisaties en personen geëxtraheerd
- Bestandsanalyse inclusief inhoudsoverzicht, leeftijdsoverzicht enz.
- Gemiddelde scores voor precisie en recall

Snellere overdracht in de toekomst

InSight™ zorgde voor het vereiste niveau van functionaliteit, documentclassificatie en duplicaatdetectie. We zijn dankbaar The National Archives te kunnen ondersteunen met dit programma van nationaal belang. De oplossing kan nu binnen de gehele overheid worden gebruikt om de overgang van documenten van alle afdelingen te versnellen.

MEER WETEN?

LEES
Andere
klantcases



**NEEM
CONTACT OP**
0800 272 44 33
www.ironmountain.nl



PRAAT MEE

Volg ons op social media
en blijf op de hoogte van
zakelijke en technologische
inzichten en ontwikkelingen

© 2022 Iron Mountain Incorporated. Alle rechten voorbehouden. Iron Mountain, Iron Mountain InSight, Iron Mountain Connect en het ontwerp van de berg zijn gedeponeerde handelsmerken van Iron Mountain Incorporated in de VS en andere landen. Alle andere handelsmerken en gedeponeerde handelsmerken zijn eigendom van hun respectieve eigenaren.



UITDAGING



OPLOSSING



RESULTAAT