



EN PÅSKYNDAD DIGITAL OMSTÄLLNING FÖR RIKSARKIVET MED HJÄLP AV ARTIFICIELL INTELLIGENS

“AI for Digital Selection-projektet syftar till att öka kunskapen om de befintliga AI-verktyg som kan användas för att genomföra bedömningen och urvalet av det “digitala berg” av dokument, e-postmeddelanden, datauppsättningar och andra typer av information som finns inom regeringen.”

Riksarkivet

Mer information kommer att finnas tillgänglig i den kommande rapporten författad av Riksarkivet med titeln: “Using AI for Digital Selection in Government: an evaluation of marketplace solutions using machine learning to select digital records for permanent preservation.”



UTMANING

Den brittiska regeringen behövde hitta ett effektivt sätt att hantera den första batchen dokument som tillkom i början av den digitala eran. Dessa dokument skulle placeras i Riksarkivet enligt regeln om att dokument som är över 20 år gamla och har ett historiskt värde ska bevaras.



LÖSNING

Iron Mountain utnyttjade den senaste utvecklingen inom artificiell intelligens (AI) och maskininlärning (ML) för att träna sitt system att känna igen kandidatposter för permanent bevarande, upptäcka dubletter för radering, extrahera enheter och tillhandahålla filanalyser.



RESULTATET

AI- och ML-funktioner kommer att göra det möjligt för myndigheter att på ett tryggt sätt överföra register mycket snabbare och mer effektivt än om de hade gått igenom samma mängd material utan att använda dessa verktyg.

UTMANINGEN FÖR REGERINGEN

“Den första batchen av brittiska regeringsdokument som tillkommit i början av den digitala eran är på väg att göras tillgänglig för allmänheten i enlighet med 20-årsregeln. Att kunna använda AI- och ML-funktioner på ett framgångsrikt sätt kommer att göra det möjligt för myndigheter att tryggt överföra register både snabbare och mer effektivt än om de hade gått igenom samma volymer material utan dessa verktyg.”

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

Alla brittiska myndigheter ansvarar för att granska sina register. De som identifieras ha ett historiskt värde måste överföras till Riksarkivet när de är 20 år gamla enligt kraven i Public Records Act 1958 (PRA). Ett viktigt steg i denna överföring och digitalisering är att genomföra en känslighetsgranskning. Detta säkerställer att överförda register inte innehåller någon information som kan skada andra eller riskera anseendet för den brittiska regeringen.

Början på den digitala eran

De flesta av de större brittiska regeringsdepartementen började gå över till att digitalisera sin verksamhet runt 2004, medan vissa avdelningar och byråer endast har existerat under den digitala eran (även om de ibland även använder sig av papper). Detta innebär att många av de register som omfattas av PRA nu endast finns i ett digitalt format. Dessa inkluderar ett brett utbud av information – från strukturerade datauppsättningar till e-postmeddelanden, Word-dokument och kalkylblad.

Omfattningen och variationen av denna digitala information innebär att det i princip är omöjligt att identifiera och välja ut vilka register som är värdefulla endast med hjälp av mänsklig arbetskraft. Regeringsdepartementen har därför börjat titta på tre nyckelfrågor:

- **Kan tekniker som artificiell intelligens (AI) eller maskininlärning (ML) påskynda och förenkla klassificerings- och granskningsprocessen för digitala register?**
- **Hur effektivt kan dessa tekniker sortera ut viktiga register samt flyktiga data eller duplicerade filer som inte längre behövs?**
- **Är det möjligt att mata in allt i ett enda system när de viktiga registren väl har valts ut?**

För att hjälpa regeringen satta TNA ihop projektet *AI for Digital Selection*. Projektet syftar till att öka kunskapen om befintliga AI-verktyg som kan användas för att bedöma och sortera den “digitala högen” av dokument, e-postmeddelanden, datauppsättningar och andra typer av information som finns inom hela regeringen.





UTVECKLING AV AI FÖR DIGITAL SORTERING

“Med tanke på utmaningens betydelse och komplexitet beslutade Riksarkivet att genomföra ett pilotprogram för att säkerställa att alla bearbetningsverktyg kunde uppnå den höga noggrannhet och kontroll som krävs. Det behövdes också en partner med en dokumenterad erfarenhet av att hantera stora, komplexa projekt på ett säkert och effektivt sätt.”

TNA genomförde en granskning av de tillgängliga verktygen för att kunna välja ut tre till fem stycken som de kunde testa på djupet med en uppsättning egna företagsregister. Planen var att ta reda på hur verktygen klarar av att identifiera vilka register som bör eller inte bör väljas ut för permanent bevarande. Dessutom ville TNA hjälpa myndigheterna att använda AI för sorteringsprocessen. Detta innebar ett arbete av att identifiera var dessa tekniker kunde implementeras i processen eller arbetsflödet för att de sedan skulle kunna välja ut vilka digitala register som skulle överföras till Riksarkivet.

Ett väletablerat partnerskap

Iron Mountain har redan en lång och väletablerad relation med TNA, liksom flera av den brittiska regeringens största avdelningar, efter att ha hjälpt dem med digitalisering och att överföra en mängd pappersregister under ett antal år. Företaget erbjuder också InSight™-tjänster, vilka inkluderar intelligent dokumenthantering där man kombinerar Iron Mountains innehållsanalys, datahantering och informationsstyrningsexpertis med funktioner som maskininlärning (ML) och artificiell intelligens (AI). Iron Mountain valdes ut för att delta i projektet *AI for Digital Selection* för att förstå effektiviteten hos AI gällande digital sortering.

Möjlighet att hantera flera filformat

Iron Mountains verktyg kan hantera ett brett urval av medieformat, vilket var viktigt i detta fall eftersom det fanns fler än 100 olika filtyper som skulle bearbetas i proof-of-concept-studien, till exempel ljud-, video- och textbaserade dokument. Några av de mest obskyra formaten åsidosattes själva testet, men det faktiska antalet filer som inte ingick i studien var väldigt lågt - lägre än 10 %. Filerna som ingår i proof-of-concept laddades sedan upp i Iron Mountains innehållstjänstplattform för att analyseras.

VAD INNEBAR LÖSNINGEN?

Fastställa vad som behövde överföras

Som en del av projektet tillhandahöll TNA Iron Mountain med märkta och omärkta datauppsättningar för att demonstrera de AI-funktioner som finns i InSight™ för att identifiera vilka register som är relevanta för sorteringsprocessen.

Ta bort känslig information

Även när det gäller filer som görs tillgängliga för allmänheten måste det finnas en tidpunkt då känslig information som inte är godkänd för utlämnande kan tas bort. I en pappersbaserad värld sker redigeringen av information av en fysisk process som är en del av en mänskligt ledd granskning - dokumenten markeras eller redigeras alltså för hand. Digitala kriminaltekniska verktyg använder sig av ett annat tillvägagångssätt - de döljer ord eller meningar och krypterar innehåll där det behövs.

En väg mot framgång

Under testet laddade Iron Mountain först upp de 17 000 testdokumenten i Googles lagringsmoln. Dokumenten bearbetades med hjälp av optisk teckenigenkänning för att göra dem fullt sökbara. InSight™ Intelligent Document Processing-plattformen klassificerade dem sedan i 20 fördefinierade kategorier med hjälp av naturlig språkbehandling (natural language processing - NLP) - en programvaruprocess som kan uttolka innehållet i ett dokument och de kontextuella nyanserna i det språk som används. Detta gjorde det möjligt för plattformen att på ett korrekt sätt extrahera informationen i filerna och skapa associationer för exempeluppsättningen för att säkerställa att det var möjligt att genomföra en meningsfull sökning.

InSights™ maskininlärningsfunktioner gjorde det möjligt för projektteamet att träna modellen i en iterativ process under projektet. I slutändan uppnådde verktyget en F1-score på över 85 %. Man kunde även se följande resultat:

- Dubletter kunde identifieras och tas bort
- Kandidatregister kunde identifieras för permanent bevarande
- Enheter inklusive organisationer och personer kunde extraheras
- Filanalys inklusive innehållssammanfattning, ålderssammanfattning m.m.
- Genomsnittliga Precision- och Recall-scores

Bevisat framgångsrikt koncept

InSight™ levererade den nivå av funktionalitet, dokumentklassificering och dubblettdetektering som krävdes. Iron Mountain är glada att kunna hjälpa Riksarkivet med detta arbete som är av nationellt intresse. Lösningen kan nu användas inom hela regeringen för att påskynda övergången av dokument från alla myndigheter via enkla ramupphandlingsförfaranden.

Se andra sätt som vi kan hjälpa dig på

TA DEL AV
Andra fallstudier från Iron Mountain



KONTAKTA OSS
+46 8 55 10 2030
[IRONMOUNTAIN.COM/SE](https://www.ironmountain.com/se)

DELTA I DISKUSSIONEN

Följ oss på sociala medier och håll dig uppdaterad om affärsmässiga och tekniska insikter och utvecklingar inom branschen

© 2022 Iron Mountain (UK) PLC. Alla rättigheter förbehållna. Iron Mountain, Iron Mountain InSight, Iron Mountain Connect och designen på berget är varumärken eller registrerade varumärken som tillhör Iron Mountain Incorporated i USA och andra länder och används under licens. Alla andra varumärken och registrerade varumärken tillhör respektive ägare.



UTMANING



LÖSNING



RESULTATET