

ACCÉLÉRER LA TRANSITION DE DOSSIERS NUMÉRIQUES VERS LES ARCHIVES NATIONALES À L'AIDE DE L'INTELLIGENCE ARTIFICIELLE

Le projet « Al for Digital Selection » a pour but d'étudier les outils à base d'intelligence artificielle (IA) existants qui permettent d'évaluer et de sélectionner le volume numérique de documents, courriers électroniques, ensembles de données et autres types d'informations détenues au sein du gouvernement. »

Les Archives nationales britanniques

Plus de détails seront disponibles dans le prochain rapport rédigé par les Archives nationales et intitulé « Utiliser l'IA pour la sélection d'archives numériques au sein de l'Administration: évaluation des solutions de Machine Learning pour sélectionner les documents numériques à conserver de manière permanente. »



DÉFI

Le gouvernement britannique avait besoin d'un moyen efficace pour traiter le premier lot de documents remontant au début de l'ère numérique. Ces documents devaient être confiés aux Archives nationales dans le cadre de la règle des 20 ans, pour la préservation des archives présentant une valeur historique.



SOLUTION

Iron Mountain a tiré parti des tout derniers progrès en matière d'intelligence artificielle (IA) et de Machine Learning (ML) pour entraîner son système à reconnaître des archives candidates à une conservation permanente, à détecter les doublons, à supprimer, à extraire les entités et à fournir une analyse des documents.



RÉSULTATS

Une utilisation à bon escient des ressources d'IA et de ML permettra aux administrations de transférer les archives en toute confidentialité et de travailler plus vite et plus efficacement sur le même volume de documents traités sans ces ressources.









LE DÉFI DU GOUVERNEMENT

« Le premier lot de documents du gouvernement britannique remontant au début de l'ère numérique est sur le point d'être rendu public dans le cadre de la loi des 20 ans. Une utilisation à bon escient des ressources d'IA et de ML permettra aux administrations de transférer les archives en toute confidentialité et de travailler plus vite et plus efficacement sur le même volume de documents traités sans ces ressources. »

Ed Irving, Directeur du développement de l'Activité, Administration centrale Iron Mountain

Toutes les administrations britanniques sont chargées d'examiner leurs archives. Celles identifiées comme ayant une valeur historique doivent être transférées aux Archives nationales (The National Archives - TNA) dès lors qu'elles datent de 20 ans, conformément à la loi de 1958 sur les archives publiques (Public Records Act - PRA).

Une étape-clé de ce processus est l'analyse de sensibilité. Cette analyse garantit que les archives transférées ne contiennent aucune information susceptible de nuire à des tiers ou de créer un risque réputationnel pour le gouvernement britannique.

L'aube de l'ère numérique

La plupart des principaux ministères du gouvernement britannique ont commencé à passer au numérique vers 2004. Certains départements et agences n'ont d'ailleurs jamais existé que sous la forme numérique, même s'ils continuent d'utiliser du papier dans une certaine mesure. Ce qui veut dire que bon nombre des archives concernées par la loi PRA n'existent désormais qu'au format numérique. Il s'agit d'un large éventail d'informations provenant aussi bien d'ensembles de données structurées que de courriers électroniques ou encore de documents bureautiques.

L'ampleur et la richesse de ces informations numériques font qu'il est quasiment impossible d'identifier et de sélectionner les archives de valeur en faisant intervenir uniquement des agents humains. Les ministères commencent à s'intéresser aux trois questions principales suivantes :

- Les technologies basées sur l'intelligence artificielle (IA) ou sur le Machine Learning (ML) permettent-elles d'accélérer et de simplifier le processus de classification et d'analyse des archives numériques ?
- Comment ces technologies font-elles concrètement la distinction entre des archives de valeur et des données éphémères ou des documents en doublon qui n'ont aucune valeur?
- Est-il possible de tout intégrer dans un seul système une fois la sélection effectuée ?

Pour aider le gouvernement à relever ce défi, les Archives nationales britanniques ont lancé le projet intitulé *AI for Digital Selection* (« *Utiliser l'IA pour la sélection d'archives numériques au sein de l'Administration* »). Ce projet a pour but d'étudier les outils à base d'intelligence artificielle (IA) existants qui permettent d'évaluer et de sélectionner le volume numérique de documents, courriers électroniques, ensembles de données et autres types d'informations détenues au sein du gouvernement.













DÉVELOPPER L'IA POUR LA SÉLECTION D'ARCHIVES NUMÉRIQUES

« Étant donné l'importance et la complexité du défi, les Archives nationales britanniques ont décidé de lancer un programme pilote pour s'assurer que les outils de traitement soient capables d'atteindre les niveaux requis en matière de précision et de contrôle. Les Archives nationales ont également eu besoin d'un partenaire à la réputation éprouvée en matière de réalisation fiable et efficace de gros projets complexes. »

Les Archives ont donc lancé une évaluation des outils disponibles afin d'en identifier entre trois et cinq à tester en profondeur sur un ensemble de leurs propres archives. L'objectif était d'évaluer la manière dont ces outils parviennent à distinguer les archives à sélectionner à des fins de conservation permanente de celles qui ne doivent pas être conservées. En outre, les Archives nationales britanniques souhaitaient aider les différents ministères à utiliser l'IA à des fins de sélection. Ceci impliquait d'identifier où ces techniques pouvaient être intégrées au processus ou au flux de travail de sélection des documents numériques à transférer aux Archives nationales.

Une relation durable

Iron Mountain entretient depuis longtemps une relation constructive avec les Archives nationales ainsi qu'avec certains des plus importants ministères du gouvernement britannique, notamment en accompagnant le transfert des archives papier à grande échelle depuis de nombreuses années. L'entreprise propose aussi des services InSight® parmi lesquels le traitement intelligent des documents qui associe l'expertise d'Iron Mountain en matière d'analyse de contenu, de gestion de données et de gouvernance de l'information avec des ressources de Machine Learning (ML) et d'intelligence artificielle (IA).

Iron Mountain a été retenue pour participer au projet « *Utiliser l'IA pour la sélection d'archives numériques au sein de l'Administration* » pour comprendre l'efficacité de l'IA pour la sélection numérique.

Gestion de nombreux formats de documents

Les outils Iron Mountain peuvent traiter un large éventail de formats de supports, un critère essentiel dans la mesure où plus de 100 types de documents différents devaient être traités dans le cadre de la démonstration de faisabilité, notamment des documents audio, vidéo et texte. Certains formats parmi les plus obscurs ont été mis à l'écart dans le cadre de l'essai, sachant que leur nombre reste faible, à savoir moins de 10 %. Les documents évalués dans le cadre de cette preuve de concept ont ensuite été chargés sur la plateforme de services de gestion de contenu Iron Mountain à des fins d'analyse.









Enjeux:

Déterminer ce qui doit être transféré

Dans le cadre de ce projet, les Archives nationales britanniques ont fourni à Iron Mountain des ensembles de données labellisées et non labellisées pour s'assurer que les ressources IA d'InSight® permettent bien d'identifier les documents répondant aux critères de sélection.

Supprimer les données sensibles

Même pour les documents rendus publics, les informations sensibles non autorisées à être diffusées doivent être supprimées à un moment ou un autre. Dans un monde fait de papier, anonymiser des informations est un processus physique mené dans le cadre d'une analyse réalisée par des humains sur des documents repérés ou édités à la main. Les outils basés sur une technologie numérique scientifique ont, eux, une approche différente en masquant des mots ou des phrases, voire en brouillant le contenu le cas échéant.

Les étapes vers le succès

Dans le cadre de cette évaluation, Iron Mountain a commencé par charger les 17 000 documents à tester dans un compartiment de stockage Google Cloud. Les documents ont été traités via la technologie de reconnaissance optique des caractères (OCR) qui les a rendus pleinement consultables. La plateforme de traitement intelligent des documents InSight® les a ensuite classés en 20 catégories prédéfinies en s'appuyant sur le traitement automatique du langage naturel, un processus logiciel permettant de déchiffrer le contenu d'un document ainsi que les nuances contextuelles de la langue utilisée. La plateforme a ainsi pu extraire avec précision les informations contenues dans les documents et créer des associations avec l'échantillon de test pour vérifier qu'une recherche efficace était possible.

Grâce aux fonctionnalités de Machine Learning d'InSight®, l'équipe de projets a pu entraîner le modèle sur un processus itératif tout au long du projet. Au final, l'outil a obtenu un F1-score supérieur à 85 %. Les autres avantages obtenus sont les suivants :

- · Identification des doublons à détruire
- · Identification des documents candidats à une conservation permanente
- Extraction d'entités (entreprises et individus)
- · Analyse des documents, y compris une synthèse du contenu, un résumé de l'âge, etc.
- Scores moyens de précision et de rappel

Validation de principe réussie

InSight® a fourni le niveau requis de fonctionnalités, de classification des documents et de détection des doublons. Iron Mountain est ravie d'accompagner les Archives nationales britanniques dans le cadre de ce programme d'intérêt national. Désormais, la solution peut être utilisée par l'ensemble du gouvernement pour accélérer la transition des documents provenant de tous les ministères au moyen d'un approvisionnement direct.

DÉCOUVREZ NOS AUTRES SOLUTIONS

EN SAVOIR PLUS SUR

CONSULTEZ la transformation numérique d'autres études de avec Iron Mountain cas Iron Mountain



CONTACTEZ-NOUS 0800 215 218 www.ironmountain.com/fr

SUIVEZ-NOUS

Suivez-nous et restez informés de nos actualités.

©2022 Iron Mountain (UK) PLC. Tous droits réservés, Iron Mountain, Iron Mountain InSight, Iron Mountain Connect et le logg en forme de montagne sont des margues commerciales ou déposées d'Iron Mountain orated aux États-Unis et dans d'autres pays et sont utilisés sous licence. Toutes les autres marques commerciales ou déposées sont la propriété de leurs détenteurs respectifs







