clearswift
RUAG Cyber Security

Reconnaissance optique de caractères (OCR) Clearswift

Suite aux milliards de dossiers perdus à cause des fuites de données en 2017 et 2018, les entreprises doivent s'assurer de pouvoir analyser de manière exhaustive tout le contenu et tous les fichiers qui circulent via la messagerie électronique. Et aujourd'hui ce contrôle ne peut plus se limiter au texte.

Produits

OCR est un module optionnel tarifé pour les produits suivants :

- Clearswift SECURE Email Gateway (SEG)
- Clearswift ARgon for Email
- Clearswift SECURE Exchange Gateway (SXG)

Support

Clearswift fournit en standard un support mondial 24 heures sur 24, 7 jours sur 7 ainsi que des options supplémentaires pour la formule premium.

Problématique métier

Votre entreprise convertit-elle régulièrement des documents Excel ou Word au format PDF ? Utilisez-vous une imprimante / scanner / photocopieur multifonction pour numériser les documents papier, les convertir en fichiers PDF puis les stocker ou les envoyer par courrier électronique ? Pour numériser un document, une image doit être créée pour chaque page et chaque image est stockée dans un fichier PDF.

Aujourd'hui, la plupart des entreprises utilisent ces méthodes de traitement de l'information au quotidien. Mais dans une perspective de prévention des fuites de données (DLP), les documents PDF circulent au sein et en dehors de l'entreprise avec une impunité relative car les solutions DLP classiques ne sont pas en mesure de détecter les informations sensibles que contiennent ces fichiers, à savoir le texte contenu à l'intérieur des images stockées dans le PDF.

D'autant plus que ce risque de fuite de données ne concerne pas que les fichiers PDF, mais tous les types de formats d'image, y compris les captures d'écran et les images de type JPG, BMP, GIF, PNG et TIFF qui sont intégrées à d'autres fichiers notamment Microsoft Office.

Méthode d'analyse du contenu sensible

La reconnaissance optique de caractères (OCR) est un processus qui consiste à détecter et extraire du texte d'un fichier image, d'une image intégrée à un document électronique ou d'un scan d'un document.

Le processus OCR examine l'image contenant du texte et crée du texte éditable sur ordinateur en analysant les points minuscules (pixels) qui forment ensemble une image de texte. Aussi, alors que le moteur OCR numérise les pixels, il crée ce qu'il pense être une lettre. Voir la Figure 1 avec des pixels qui deviennent des lettres puis forment du texte. Le résultat est ensuite comparé à des caractéristiques spécifiques et à un alphabet pour reconnaître la lettre correspondante.

Figure 1 : des pixels qui deviennent des caractères ou des lettres puis du texte

Les lettres sont ensuite associées en mots après avoir localisé les espaces, la ponctuation et les retours à la ligne, et les mots sont vérifiés dans un dictionnaire pour devenir du texte. Le texte extrait est ensuite traité par le moteur d'inspection approfondie du contenu de Clearswift pour déterminer si des informations sensibles sont affichées dans l'image. Si tel est le cas, une procédure est lancée. C'est ainsi que par exemple une pièce jointe avec une image contenant des données sensibles est bloquée.

Powered by Nuance

NUANCE

À propos de Clearswift

Clearswift est plébiscitée dans le monde entier pour ses solutions qui permettent aux entreprises de protéger leurs informations sensibles, de collaborer en toute sécurité et de développer leur activité. Notre technologie unique est une solution de prévention des fuites de données simple et « contextuelle » qui limite le risque d'interruption de l'activité tout en permettant aux entreprises de garder le contrôle total et d'avoir une visibilité complète sur leurs informations critiques à tout moment.

Clearswift est présent dans le monde entier grâce à des sièges régionaux en Europe, Asie-Pacifique et aux États-Unis. Clearswift anime un réseau de partenaires de plus de 900 revendeurs à travers la planète.

Pour plus d'informations :

www.clearswift.fr

Nous contacter

Clearswift Ltd.
1310 Waterside
Arlington Business Park
Theale, Reading
Berkshire RG7 4SA
Royaume-Uni

E: info@clearswift.com
T: +44 118 903 8300

Déploiement

Nos toutes dernières solutions de sécurité Clearswift SECURE Email Gateway (SEG), SECURE Exchange Gateway (SXG) et Argon for Email peuvent être enrichies d'une nouvelle option tarifée pour l'OCR afin d'atténuer les risques de fuite de données via les images. Multilingue, cette offre peut être facilement utilisée par des entreprises d'envergure mondiale qui communiquent dans plusieurs langues.

L'utilisation de dictionnaires multilingues réduit le nombre de faux positifs et augmente le taux de reconnaissance. L'architecture sous-jacente de la passerelle SEG a toujours été de nature évolutive avec possibilité d'avoir plusieurs instances assemblées en mode 'peering' pour garantir évolutivité et disponibilité.

L'introduction de la technologie OCR s'appuie sur cette évolutivité pour pouvoir ajouter des instances supplémentaires et faire face au traitement additionnel engendré et maintenir la performance générale de la solution.

Fonctionnalités

Conçu pour s'adapter à des déploiements en entreprise, le système fournit :

- Un contrôle granulaire des politiques qui permet d'utiliser OCR en entrée et sortie et en fonction du trafic (de tel expéditeur vers tel destinataire)
- Le scan des images dans les documents (tels que les fichiers bureautiques classiques comme MS Office, PDF et LibreOffice), des images en pièces jointes ou embarquées dans des pièces jointes compressées
- Le support de plus de 20 formats de fichiers image dont JPEG, PNG, BMP, GIF et TIFF
- Des performances élevées avec des images numérisées en moins d'une seconde (JPEG 300 Ko par exemple)
- Le support de 48 langues ainsi que le chinois, le japonais et le cyrillique
- Une taille d'image de 16x16 pixels minimum et de 8400 x 8400 pixels maximum
- Un angle d'inclinaison maximum de 15° et une rotation de 180°
- Une police de caractères de 8 pour le texte
- Le support de la plupart des combinaisons de couleurs pour le texte et l'arrière-plan (tant que les couleurs sont suffisamment différentes)

Coût

La reconnaissance optique de caractères (OCR) est une option tarifée pour les produits Clearswift SECURE Email Gateway, SECURE Exchange Gateway et ARgon for Email. Son coût est fonction de votre installation et de vos besoins de traitement.

Contactez-nous dès aujourd'hui pour en savoir plus.

clearswift
RUAG Cyber Security