

Clearswift Optical Character Recognition (OCR) Regelbasiertes Erkennen Von Sensitiven Inhalten in Bilddateien Und Screenshots

Aufgrund von Datenschutzregularien müssen Unternehmen sicherstellen, dass sie in der Lage sind, per E-Mail verschickte Inhalte gründlich zu untersuchen. Inzwischen beschränkt sich das jedoch nicht nur auf Text, sondern auch auf Bilddateien und Screenshots verborgene sensitive Inhalte.

Geschäftsprozesse

Konvertiert Ihr Unternehmen regelmäßig Excel- oder Word-Dokumente in das PDF-Dateiformat? Haben Sie einen multifunktionalen-Fotokopierer/Drucker/Scanner, mit denen Papierdokumente gescannt, in PDF konvertiert und gespeichert oder weitergegeben werden? Vielleicht müssen gesendete Dokumente in elektronischer Form gespeichert oder weitergegeben werden, so dass die Dokumente innerhalb von Sekunden gescannt, in ein PDF konvertiert und automatisch per E-Mail versendet werden. Das heißt, wenn ein Dokument gescannt wird, wird für jede Seite ein Bild erstellt und die Bilder werden in der PDF-Datei gespeichert.

Die meisten Unternehmen nutzen heute eine oder beide Methoden als Standard im Tagesgeschäft. Aus der DLP-Perspektive könnten PDF-Dokumente jedoch innerhalb und außerhalb der Organisation verschickt werden, da herkömmliche DLP-Lösungen die in diesen Dateien enthaltenen sensiblen Informationen nicht erkennen können: den Text, die Bilder und das PDF.

Dieses Risiko des Datenverlustes gilt nicht nur für PDFs, sondern für alle Arten von Bildformaten, einschließlich Screenshots oder Bilder (z. B. JPG, BMP, GIF, PNG und TIFF), die in andere Dateien wie Microsoft Office eingebettet sind.

Wie Können Wir Nach Sensiblen Inhalten Scannen?

Optische Zeichenerkennung (OCR) ist der Prozess der Erkennung und Extraktion von Text aus einer Bilddatei, einem in einem elektronischen Dokument eingebetteten Bild oder des Scan eines Dokumentes.

PRODUCT SUMMARY

Produkte

OCR ist ein optionales Modul für die folgenden Produkte:

- Clearswift Secure Email Gateway (SEG)
- Clearswift ARgon für E-Mail
- Clearswift Secure Exchange-Gateway (SXG)
- Clearswift Secure Web-Gateway (SWG)

Support

Clearswift bietet standardmäßig einen globalen 24x7-Support mit zusätzlichen Optionen für Premium-Support.

Der OCR-Prozess untersucht das Textbild und erstellt aus dem Scannen der winzigen Punkte (Pixel), die zusammen ein Bild des Textes bilden, den computerlesbaren Text. Während die OCR-Engine also die Pixel scannt, baut sie etwas auf, von dem sie glaubt, dass es ein Buchstabe ist, siehe Abbildung 1: Aus Pixeln werden Buchstaben, die zu Text werden. Dieser wird dann durch eine Reihe von Muster- und Buchstabenvergleichen einem entsprechenden Buchstaben zugeordnet.

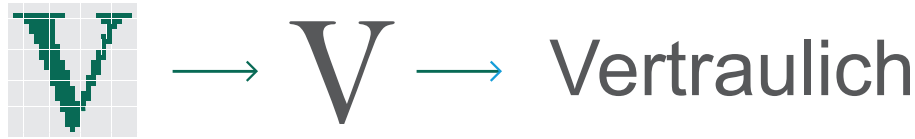
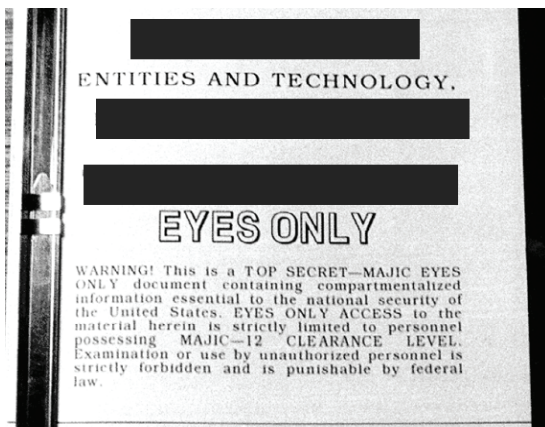


Abbildung 1: Pixel werden zu Buchstaben und zu Text

Die Buchstaben werden dann kombiniert, indem geprüft wird, wo die Leerzeichen, Satzzeichen und das Ende einer Zeile vorkommen und die Wörter werden mit Sprachwörterbüchern abgeglichen, um die entsprechenden Wörter zu erkennen. Der extrahierte Text wird anschließend durch die Deep Content Inspection (DCI) und der Policy-Engine von Clearswift verarbeitet, um festzustellen, ob im Bild sensible Informationen angezeigt werden. Mit unserer DCI-Engine, die standardmäßig bis zu 50 Ebenen tief rekursiert, kann das Bild in eine Excel-Tabelle eingebettet sein, die in ein Word-Dokument eingebettet ist, welches dann über ein an eine E-Mail angehängtes ZIP-Archiv weitergegeben wird. Der extrahierte Text von der Clearswift Adaptive Data Loss Prevention (ADLP)-Funktionalität verarbeitet. Zum Beispiel wird der Anhang mit dem Bild, das sensible Daten enthält, blockiert.

Eine weitere Verbesserung der OCR-Analyse ermöglicht die Schwärzung von Text in Bildern, wobei nur die Informationen entfernt werden, die gegen die Richtlinien verstoßen, indem ein schwarzer Kasten über die Wörter gezeichnet wird. Clearswift erkennt das Bild, analysiert es und entfernt alle sensiblen Informationen, so dass die "sichere" Datei an den Empfänger weitergeleitet werden kann, selbst wenn das Bild in ein Word-Dokument oder eine Zip-Datei eingebettet wurde.



Bereitstellung

Unsere Sicherheitslösungen Clearswift Secure Email Gateway (SEG), Secure Exchange Gateway (SXG), ARgon for Email und Secure Web Gateway (SWG) verfügen über eine Kostenoption für OCR, um das Risiko von Datenverlusten durch Bilder zu verringern. Die Lösung unterstützt mehrere Sprachen, so dass sie problemlos von globalen Organisationen eingesetzt werden kann, die mit mehr als einer Sprache arbeiten.

Die Verwendung von sprachspezifischen Wörterbüchern reduziert die Anzahl der Fehlalarme und erhöht die Erkennungsrate. Die Architektur hinter SEG war schon immer skalierbar, wobei mehrere Instanzen zusammengeschaltet werden können, um sowohl Skalierbarkeit als auch Verfügbarkeit zu gewährleisten.

Die Einführung der OCR-Option macht sich die Skalierbarkeit zunutze, indem mehr Instanzen hinzugefügt werden können, so dass der Gesamtdurchsatz des Systems beibehalten werden kann, obwohl mehr Verarbeitung stattfindet.

Funktion

Das System ist für die Skalierung auf Unternehmensanwendungen ausgelegt und bietet:

- Granulare Richtlinienkontrolle - erlaubt OCR-Nutzung nach Richtung und nach Absender/Empfänger-Kombination
- Kann eingebettete Bilder, an Nachrichten angehängte Bilder, Bilder in komprimierten Anhängen und eingebettete Bilder in gängigen "Office-Dateien" wie MS Office, PDF und LibreOffice scannen
- Hohe Leistung, Bilder werden <1s gescannt (Beispiel JPEG 300Kb)
- Unterstützung von 48 Sprachwörterbüchern (einschließlich Chinesisch, Japanisch und Kyrillisch)
- Minimale Bildgröße 16x16 Pixel; Maximum 8.400 x 8.400 Pixel

- Maximaler Schrägstellungswinkel 15°; 180°-Drehung unterstützt
- Maximum angle of skew 15°; 180° rotation supported
- Unterstützt 8pt Text
- Unterstützt die meisten Farbkombinationen Text / Hintergrund (ausreichende Farbtrennung ist erforderlich)

Kosten

Optical Character Recognition (OCR) ist eine kostenpflichtige Option für die Produkte Clearswift Secure Email Gateway, Secure Exchange Gateway, ARgon for Email und Secure Web Gateway (SWG) und für das Secure ICAP Gateway im Laufe des Jahres verfügbar. Die Kosten sind abhängig von Ihrer Installation und den Anforderungen an den Verarbeitungsdurchsatz.

Kontaktieren Sie uns noch heute für weitere Informationen.



Über Fortra

Fortra ist ein Cybersicherheits-Unternehmen wie kein zweites. Wir erschaffen eine einfachere und solidere Zukunft für unsere Kunden. Unsere bewährten Experten und unsere breite Palette integrierter und skalierbarer Lösungen bringen Ausgewogenheit und Kontrolle in Unternehmen auf der ganzen Welt. Bei Ihrer Reise zu mehr Cybersicherheit sind wir Ihr Wegbereiter und Ihr unermüdlicher Verbündeter auf jeder Etappe. Erfahren Sie mehr auf fortra.com/de.