Intelligente „Ordnungs-Hüter“

#Intelligentes#Archiv, #Intelligente#Dokumente, #Automatisierung, #Selbstlernende-Systeme, #Kontext-Daten, #Content-Daten, #Aufbewahrungsfristen

Benjamin Schröder ist Leiter R&D Team der **KGS Software GmbH**. Beim digitalen Archivspezialisten mit Hauptsitz in Neu-Isenburg werden Daten und Dokumente aus SAP mittels schlanker Software migriert und archiviert. Mit „tia®“ – the intelligent archive – hebt KGS Archivierung auf eine neue, intelligente Technologie, die auch andere Applikationen anbindet. Seit 2005 zertifiziert KGS für die SAP weltweit ArchiveLink® und ILM-Schnittstellen und ist globaler SAP Value Added Solutions Partner.

www.kgs-software.com

Marie Kondo und ihr Kleiderschrankprinzip haben den Kampf gegen ungetragene Kleidungsstücke in unseren Haushalten lang aufgenommen. Überträgt man ihre Prinzipien auf Unternehmen, stellt sich die Frage, wie es sich dort mit den vielen ungenutzten Daten- und Dokumenten auf riesigen Speichersystemen verhält. Zwar gestalten sich digitale Use-Cases des Aufräumens durchaus komplexer, aber die Resultate sind nicht weniger erleichternd. Auch muss ein rechtlich relevantes Dokument anders ‚gelagert‘ werden als eine Alltagsnotiz und es gibt unterschiedliche gesetzliche Aufbewahrungsfristen und unternehmensspezifische Regelungen oder Praxishandhabungen. Jede Frist, jedes Kriterium – alle diese Informationen sind beim „Aufräumen“ zu beachten, um am Ende nicht doch Wichtiges unwiederbringlich zu verlieren.

Dennoch: Unkritische Dokumente, die jahrelang nicht geöffnet wurden sollten zum Löschen vorgeschlagen oder automatisch und transparent auf einen günstigeren Speicher umgezogen werden – ganz abgesehen davon, dass einige Daten aufgrund gesetzlicher Vorschriften sogar zwingend gelöscht werden müssen. In allen Fällen braucht es intelligente Lösungen, die alle Trigger identifizieren, daraus lernen und selbstständig agieren.

Das intelligente Archiv

Das Archiv der Zukunft muss Muster selbst erkennen und vorausschauend Optimierungsvorschläge bringen. Übervolle Archive sollten reorganisiert, Vorschläge zum Löschen gemacht, teure Speicherplätze entlastet und günstiger Speicher umfänglich genutzt werden. Dies alles jedoch nicht wahllos, sondern diversen Informationen folgend, die unmöglich jedem einzelnen

Dokument „aktiv“ via Metadaten oder gar durch manuellen Aufwand des Bearbeiters mitgegeben werden können.

Auf dem Weg zu einer vollständigen Automatisierung der Prozesse benötigt das intelligente Archiv genau diese Informationen, die es lernen lassen und Handlungen auslösen. Die spannendsten Informationen sind dabei häufig diejenigen, die ein Dokument selbst sozusagen „inaktiv“ erlebt und die es inhaltlich ausmachen. Ein lernfähiges System kann diese Daten aufnehmen und auf deren Grundlage vielleicht sogar besser – sehr wahrscheinlich jedoch verlässlicher und unfehlbarer – handeln als der Mensch.

‚Autonomie‘ ist das Ziel

Vom autonomen Fahren hat heute jeder eine Vorstellung. Doch es sind nicht die Straße oder der Zielort, die das Auto zum autonomen Fahren befähigen, sondern das Auto selbst. Dieser Gedanke lässt sich auch auf das Thema Dokumentenarchivierung übertragen: Weder die Software-Anwendung, in der das Dokument erstellt und bearbeitet wird, noch der Speicher, in dem das Dokument archiviert wird, können die Archivierung steuern.

Eine dritte Ebene, so wie es häufig praktiziert wird – ein DMS (Dokumenten Management System) oder ECM (Enterprise Content Management System) – wäre ebenso eine übergeordnete Intelligenz. Diese bringt erstens weiteren „Systemballast“ mit sich und raubt zweitens ebenfalls Flexibilität. Der Grund: Systeme gehen, aber die Dokumente bleiben. Das Szenario: Bei einem Anwendungswechsel muss meist jede Querverbindung und Datenreplikation zwischen den Anwendungen neu konzipiert, jedes Dokument migriert und gegebenenfalls jede ID neu vergeben werden. Der Austausch einer Instanz wird zum Großprojekt für die IT.

Wie wird ein Dokument intelligent?

Das intelligente Dokument ist die Lösung für diesen Flickenteppich der Systeme. Dokumente tragen diverse Informationen in sich, wie beispielsweise den Content selbst, aber auch Meta- und kontextbezogene Daten. Die Metadaten etwa stammen aus den führenden Anwendungen und werden dem Dokument mitgegeben. Beim Abschalten einer Anwendung gehen diese Daten im Falle des intelligenten Dokuments nicht mehr verloren. Vielmehr machen sie das Dokument zu einer von der Anwendung unabhängigen, autonomen Instanz, die jederzeit in anderen Anwendungen angezeigt und bearbeitet werden kann.

Wie wird ein Dokument intelligent? Durch ein System, das die Daten des Dokuments an geeigneter Stelle erfasst, ausliest und weitergibt, um die „freien“ Dokumente sozusagen in die richtigen Bahnen zu lenken, wie tia – the intelligent archive – von KGS. Schnittstellentechnisch setzt das intelligente Archiv dabei auf den Content Management Interoperability Services-Standard (CMIS). Anders als etwa die SAP ArchiveLink-Schnittstelle ermöglicht er die Verwaltung von Metadaten; eine weitere dafür geeignete Schnittstelle ist OData - verbreitet vor allem im SAP-Umfeld, aber auch bei Anwendungen wie MS SharePoint. Damit wird in der Folge ein Single Point of Truth (SPoT) für Dokumente jeder Art geschaffen. Ganz gleich, in welcher Anwendung gearbeitet wird – auf die Dokumente kann von überall zugegriffen werden.

Komplexe Daten managen die Archivierung

Diese Fähigkeiten lassen sich über Statistics realisieren. Zunächst beschreibt Statistics nur die Disziplin, die die Erfassung, Organisation, Analyse, Interpretation und Darstellung von Daten betrifft. Es ist die Basis für regelbasierte Operationen und ebnet den Weg zu einem selbstlernenden System (Machine Learning) und den vorausschauenden Services. Metadaten sind dabei nur eine der Quellen, aus denen Statistics schöpft. Die Kontext- und Content-Daten sind das eigentliche Gold, nach dem ein intelligentes Archiv schürft. Ein Beispiel sind Nutzungsdaten, von wem, wann und wie oft wurde das Dokument zurück aus dem Archiv auf den Bildschirm geholt.

Technisch funktioniert dies über sogenannte Data-Lakes. Alle Daten werden erfasst und in entsprechende Lakes weggeschrieben. Mit Zugriff darauf können Analysen und diverse Operationen durchgeführt werden. Dabei werden Daten erst strukturiert oder gegebenenfalls umformatiert, wenn sie konkret benötigt werden. Neben text- oder zahlenbasierten Daten können Data-Lakes auch Bilder, Videos oder andere Datenformate aufnehmen. Ein Content bezogenes Beispiel aus dem Personalbereich und dem Datenschutz könnte demnach sein: Ex-Mitarbeiter X verlangt eine Löschung aller Fotos mit seinem Konterfei von Betriebsfeiern und offizieller Nutzung. Nun muss entsprechend auf Bilder- und Videodaten zurückgegriffen werden aus denen entsprechende Inhalte extrahiert werden.

Viele große und mittelständische Unternehmen haben bereits Data-Lakes bzw. ganze Datensysteme für ihre Business-Intelligenz-Anforderungen aus anderen Bereichen im Einsatz. Die autonome Archivierung zeichnet sich dadurch aus, dass sich die Archiv-Management-Fähigkeiten auch flexibel in diese bestehenden Systeme von zum Beispiel Google, AWS und SAP einfügen lassen.

Effizienz bei der Daten- und Dokumentenmigration

Das Thema Migration von Daten und Dokumenten kommt in nahezu jedem Archivierungsprojekt zum Tragen. Die meisten Unterlagen eines Unternehmens liegen bereits an einem Ort – also in dem abzulösenden System – und müssen schnell und unbemerkt vom Anwender umgezogen werden. Das wird häufig als der kritischste Teil des Projekts betrachtet. Die Angst vor Stillstand in den Abteilungen oder Produktionsketten aufgrund von fehlendem Dokumentenzugriff oder einer starken Verlangsamung der Anwendungen ist groß. Ebenso die Frage, ob Daten verloren gehen oder allzu viele korrupte Dokumente zum Vorschein kommen, treibt die Verantwortlichen um.

Im Zweifel sind diese Bedenken berechtigt, umfassen solche Umzüge doch meist hunderte Millionen Dokumente und zweistellige Terrabytes, was zu langen Laufzeiten der Migrationsprojekte führt und Ressourcen bindet. Auch hier heißt das Prinzip, weg von der manuellen hin zur autonomen Steuerung, um mittels statistischer Möglichkeiten Projekte schneller sowie performanceneutral abzuwickeln und Fehler zu vermeiden. Die Entwickler von tia arbeiten beispielsweise an einem „Gaspedal", das bei großen Dokumentenmigrationen selbstständig erkennt, welcher Migrationsumfang bei der aktuellen Last der Unternehmenssysteme sinnvoll ist. So können Migrationen stattfinden, von denen der Anwender nichts merkt und das nahezu autonom bei maximaler Effizienz.

Konkret orientiert sich tia dabei nicht an den vom Menschen assoziierten Auslastungsszenarien, sondern beginnt die Migration auf Basis einer ständigen Analyse der Zugriffszeiten. Wie lange dauert das Holen, Verarbeiten und Wegschreiben – also der Weg von der Quelle bis zum Ziel. Daraus wird abgeleitet, wie viele gleichzeitig stattfindende Operationen in der entsprechenden Situation sinnvoll sind, und man erhält so eine überzeugende Performanz des Gesamtlayers.