

Wertschöpfung durch relevantes Wissen: Methoden wichtiger als Tools

Wenn Unternehmen immer wüssten, was sie schon alles wissen, könnten sie ihren Umsatz nachhaltig steigern. Die Realität bei der Datenhaltung in vielen Organisationen ist aber, dass sie sehr grosse Speicher aus strukturierten und unstrukturierten Daten in unterschiedlichen Systemen unterhalten. Vor allem Unternehmen, die ihre Geschäftsprozesse agiler gestalten sowie End to End digitalisieren wollen, müssen alle ihre Informationen verfügbar machen. Dabei helfen Enterprise Search & Analytics-Lösungen (ESAL). Sie basieren auf KI- und Machine Learning-fähigen Suchplattformen, die integriert in Collaboration-Umgebungen in Sekundenbruchteilen kontext- und rollengerechte Suchergebnisse liefern.

Die Crux bei ESAL liegt aber weniger in der Auswahl eines Tools. Entscheidend sind Konzeption, Methoden der Modellierung und Aufbereitung, die Security sowie das wirklich relevante Suchergebnisse auch in den Kontext der jeweiligen Geschäftsprozesse des Nutzers passen. Marc Hoffmann, Senior Consultant Search & Analytics, DTI Schweiz AG, erläutert, worauf es beim Konzept und der Implementierung von ESAL ankommt.

Häufig betreiben die Fachabteilungen eigene Dokumentenmanagementsysteme, Speichersysteme oder unterstützende Informatik-Plattformen, die teilweise unverbunden nebeneinander her bestehen. So arbeitet der Vertrieb z.B. mit einem Customer Relations Management-Tool, das Marketing organisiert sich mit Marketing Automation Tools und das Management wiederum holt sich regelmässig Daten, um diese in seinen Enterprise Resource Planning- (ERP) oder Business Analytics-Tools (BA) zu verarbeiten und auszuwerten. Der Grund für diesen Wildwuchs der Systeme sind historisch gewachsene IT-Landschaften, die unkonsolidiert mittlerweile mehr Aufwände bereiten, als dass sie einen Wertbeitrag für die



Marc Hoffmann, Senior Consultant bei der DTI Schweiz AG verfügt über mehr als 20 Jahren praktische Erfahrung in den Bereichen Enterprise Search und Content Analytics. Nach seinem Abschluss in «Bachelor of Science in Information Technologies» und einem berufsbegleitenden Nachdiplomstudium «Betriebswirtschaft und Management» war Marc Hoffmann unter anderem als Entwickler, Projektleiter und Consultant bei der DTI Schweiz AG tätig.

Kollaboration in einem agil arbeitenden Unternehmen leisten. Das führt dazu, dass unstrukturierte Daten die in einem nicht gepflegten Datensilo häufig gar nicht gefunden werden können, weil sie nicht für Suchprozesse aufbereitet sind. Gleiches gilt für relevantes Wissen, das sich in externen Informationsquellen wie Datenbanken von Branchenverbänden, weltweiten Patentenbehörden sowie öffentlichen und Wissenschaftsinstitutionen befindet. So ist es kein Wunder, dass Mitarbeiter in vielen Unternehmen oft Stunden aufwenden müssen, um an gesuchte Informationen zu gelangen. Hier helfen ESAL, Suchprozesse zu beschleunigen, sicherer zu gestalten und damit einen nachhaltigen Wertbeitrag zu leisten.

Integration und Auswertung aller Datenquellen und Dateiformate

Wer verstehen möchte, welche Fähigkeiten und Vorzüge ESAL in einem Unternehmen entfalten, muss sich zunächst mit den Grundfunktionen beschäftigen.

ESAL sind in der Lage, über integrierte Konnektoren auf relevante Informationen und Daten innerhalb und ausserhalb eines Unternehmens zuzugreifen, um diese für eine Aufbereitung eines Such-Indexes abzugreifen und zu klassifizieren. Solche Quellen können beispielsweise File-Systeme- Ablagen, Datenbanken, Customer Relationship Management- und Enterprise Content Management-Systeme, gescannte und OCR erkannte Dokumente, oder aufbereitete Sprach-, Bild- und Video-Dateien sein.

Um aus einer Vielzahl von Datenquellen allerdings relevante Suchergebnisse zu finden, kommen wie bei einer ESAL für den Such- und Identifikationsprozess komplexe Algorithmen für semantische, linguistische und phonetische Analysen zum Einsatz, um Inhalte zunächst zu erfassen, zu klassifizieren und zu indexieren. Das Problem dabei ist, dass die Daten aus den verschiedenen Quellen nur teilweise verarbeitbare Inhalte wie Fliesstexte, Titel oder Metadaten enthalten. Video-, Bild- oder Audiodateien sowie verpackte oder geschützte Dateien wie zip lassen sich ohne Aufbereitung und einer automatischen Auswertung mit sinnvollem Inhalt indexieren. Häufig müssen zunächst Texterkennungsprogramme Dokumente auslesen, Audioaufzeichnungen in Text (Speech to Text) umgewandelt werden, bevor sie überhaupt für eine Auswertung bzw. Indexierung bereitstehen. Und bei der Auswertung mittels OCR (Optical Character Recognition) oder Speech to Text-Umwandlung passieren Fehler. Zudem existieren unterschiedliche Schreibkonventionen, es werden Namen und Begriffe in unterschiedlichen Varianten geschrieben oder in vielen Sprachen unbekannte Umlaute und Zeichensätze.

Erst durch Volltexterkennung und Korrektur sowie Integration aller Quellen in einen Such-Index sowie allfällige automatische Verschlagwortung werden Suchprozesse über die komplette Datenbasis effizient. Effiziente ESAL bringen die Infor-

mationen zudem in den Kontext der Geschäftsprozesse und können somit sehr genau auf Suchanfragen die relevanten Ergebnisse zusammenstellen. Zudem helfen ESAL durch Authentifizierung der Nutzer sowie eine Rechte- und Zugriffsverwaltung dabei, dass Mitarbeiter nur Suchergebnisse von Inhalten erhalten, für die sie auch entsprechende Befugnisse im Quellsystem erteilt bekommen haben.

Konzept und Modellierung von ESAL entscheiden über Erfolg

Auf die Zuordnung der Relevanz von Suchergebnissen, deren Kontextualisierung zu den jeweiligen fachbereichsabhängigen Prozessen sowie der Rechte- und Zugriffsverwaltung kommt es bei ESAL-Projekten also massgeblich an. Es reicht nämlich nicht, einfach ein ESAL-Produkt zu installieren, die Datenspeicher anzuschliessen und dann zu schauen, wie es funktioniert. Wichtiger und lange vor der Kaufentscheidung für ein Tool ist es von zentraler Bedeutung, ein ESAL-Konzept zu erstellen und eine Modellierung vorzunehmen, welche Datenquellen wie eingebunden werden, welche Relevanz bestimmte Datenarten und Dateiformate haben und welche Dateien überhaupt für welche Entscheidungsebenen zugänglich sein sollen.

In der Praxis zeigt sich auch häufig, dass Unternehmen zwar eine bestimmte Herausforderung wie beispielsweise ein DSGVO-konformes Datenhandling lösen wollen, aber die gesamte Dimension eines ESAL für ihre Compliance erst im Projektverlauf erkennen. Es gibt auf dem Markt einige Produkte, auf welchen eine ESAL Lösung aufgebaut werden kann, welche unterschiedliche Stärken und Schwächen haben. So zeigt sich typischerweise erst bei der Modellierung und der Konzeption eines ESAL-Projektes, welche Plattform die am besten geeigneten Funktionalitäten und Stärken bietet, die zu lösende Anforderungen umzusetzen.

Oft sind auch bestehende Tools oder Quellen der Fachabteilungen zu integrieren, für die bestimmte Konnektoren oder Schnittstellen erst noch programmiert werden müssen. Und ein hochaktueller Aspekt ist, dass oftmals Bestandssysteme bereits in die Cloud verlagert wurden, einige Da-

teien wie Kundenstammdaten und Verträge aber aus Sicherheitsgründen nur On-Premise gespeichert und verarbeitet werden dürfen. Hier muss die Modellierung auf jeder Ebene der Datennutzung eine Lösung finden, die unterschiedlichen Anforderungen gerecht wird. Damit Kunden ihre eigene sensiblen Kundendaten nicht in einer Cloud-Lösung für die Suche aufbereiten lassen müssen, sind in bestimmten Fällen ESAL-Lösungen On-Premise einer Cloud-Lösung vorzuziehen, die also stationär auf eigenen Servern läuft.

Die zu indexierenden Datenquellen werden typischerweise in einem initialen Schritt aufbereitet und indexiert und im Folgenden nur noch Aktualisierungen verarbeitet. Je nach Bedarf erfolgt dies einmal täglich oder auch häufiger, abhängig von der benötigten Aktualität oder auch der Datenmenge. Diese Aufgabe übernehmen Konnektoren, die unterschiedlichste Drittsysteme, Datenbanken, Filesystemquellen aber auch Webinhalte erschliessen. Für solche Datenquellen beherrschen Konnektoren zudem die inkrementelle Indexierung, also die automatische Indexierung von Datenänderungen.

Für Drittsystem oder öffentliche Quellen ist es zudem von entscheidendem Vorteil, dass diese über eine offene Architektur uns Schnittstellen wie ReST (Representational State Transfer) verfügen. ReST ist eine Schnittstelle für verteilte Systeme, mit dem heute bereits viele Inhalte abgegriffen werden können und somit eine Maschine zu Maschine-Kommunikation unterstützen.

Die über die Konnektoren gesammelten Daten bereitet das ESAL vor der Indexierung, also bei der Überführung in den Suchindex, auf. Über Filter (Conversion) werden die unterschiedlichen Dateiformate (Word, Excel, PowerPoint, PDF, HTML, RTF, TEXT, ZIP) in ein einheitliches, lesbares Format für den Index überführt. Zudem wird über Entity Extraktionen, Regeln oder Integration von externen Systemen versucht, die Daten automatisch um zusätzliche Metadaten anzureichern oder zu klassifizieren. Bei einigen ESAL werden für die Klassifizierung von Inhalten Künstliche Intelligenz (KI) und Machine Learning (ML) eingesetzt. So kann eine KI dafür trainiert werden, spezifische Inhaltstypen zu identifizieren, typischerweise für Binärdaten wie z.B. Bilder, Videos

welche keine «lesbaren» Informationen enthalten. Für die Nutzer ist es zudem häufig hilfreich, wenn ein ESAL bei der Aufbereitung auch Thumbnail-Previews für Dokumente erstellen kann. Bei der Spracherkennung und sprachlichen Aufbereitung der zu indexierenden Inhalte kommen Funktionen wie Tokenisierung, Synonyme, Spell Checking, Lemmatisierung, Phonetic, Natural Language Processing (NLP) für die jeweilige Sprache zum Einsatz. Und ist ein Text formal korrekt aufbereitet, muss er natürlich auch klassifiziert werden. Hierfür kommen Technologien wie Entity Extraction, KI oder ML zum Zuge.

So werden Inhalte «verstanden» und mittels den obigen Technologien klassifiziert und mit zusätzlichen Informationen angereichert. Diese Klassifizierung und Anreicherung kann dann sowohl bei der Relevanz für bestimmte Geschäftsprozesse als auch beim Ranking der Suchergebnisse genutzt werden.

Datenaufbereitung entscheiden über Relevanz und Sicherheit

Die Modellierung einer ESAL und ihrer Funktionen im Suchindex bilden damit also das Kernstück der Suchfunktionen und letztlich den Erfolg des Systems. Und weil es eine Schlüsselfunktion in einer ESAL bildet, sollten bei der Modellierung bereits analysiert werden, mit welchem Tool eine spezifische Aufgabe am besten gelöst werden kann.

Denn der Such-Plattform hat die Aufgabe, die Daten intern so zu strukturieren und aufzubereiten, dass selbst in extrem grossen Datenmengen und bei komplexen sowie sehr spezifischen Suchen, die relevanten Resultate innerhalb kürzester Zeit bereitstehen. Basierend auf diesem erstellten Such-Index ist es auch möglich, den Nutzer aktiv über neue Inhalte, die die vordefinierten Kriterien erfüllen, automatisch zu informieren. Auch das Ranking der Ergebnisausgabe ist ein Qualitätsfaktor von ESAL. Je nach Nutzergruppe im Unternehmen können die Anforderungen an das Ranking der Resultate unterschiedlich ausfallen. Deshalb muss ein ESAL auch die Möglichkeit bieten, das Ranking und die Security-Richtlinien basierend auf den Quellsystemen an die verschiedenen Geschäftsprozesse der Nutzer anzupassen.

Dies sollte idealerweise schon bei der Modellierung berücksichtigt und bestimmte Nutzergruppen identifizieren werden. Für sie sind Rollen zu definieren und die Relevanz von Suchergebnissen an den Anforderungen ihrer Geschäftsprozessen auszurichten. Sucht beispielsweise ein Vertriebsmitarbeiter nach einem Produkt, sollten ihm die Produkte zusätzlich angezeigt werden, die auch andere Kunden gekauft haben. Ein Techniker aber braucht die Ersatzteilliste zu diesem Produkt, und die HR-Abteilung sucht bei einem Namen in der Regel eine Personalakte. Die Schadenabteilung einer Versicherung aber sucht Verträge, die eine Person mit dem Namen abgeschlossen hat oder Begünstigte in einem Vertrag benannt hat usw.

Zudem ist auch entscheidend, welche Suchergebnisse ihnen überhaupt eingeblendet werden. Ein Benutzer darf nur diejenigen Inhalte angezeigt bekommen, auf welche er in den Quellsystemen auch berechtigt ist. So sollte beispielsweise nur der Einkaufsleiter einen Rahmenvertrag mit einem Anbieter einsehen können, der Einkaufsachbearbeiter aber die einzelnen Verträge des täglichen Geschäfts. Dokumente, die bestimmten Hierarchiestufen vorbehalten sind, werden dadurch unbefugten Mitarbeitern erst gar nicht angezeigt. Dies wird durch eine korrekte Berechtigungsvergabe im Quellsystem erreicht, welche durch die ESAL-Lösung berücksichtigt wird. Da jedes Quellsystem und jeder Kunde eigene Security-Umsetzungen hat, ist es zwingend, diese bereits bei der Konzeption zu berücksichtigen. So wird deutlich, dass das Gesamtkonzept eines ESALs entscheidend für dessen Fähigkeit ist, die wirklich relevanten Suchergebnisse zu identifizieren und von den jeweils nutzerabhängig weniger relevanten Ergebnissen zu unterscheiden.

Fazit: Hoher Wertbeitrag und schnelles ROI

Es gibt also viele Nutzerdimensionen, Stellschrauben und Funktionen, die bereits bei der Modellierung einer ESAL vorab zu adressieren sind. Es ist deshalb auch wichtig, dass sich Unternehmen vor der Entscheidung für eine Technologie bzw. einer Plattform die Zeit nehmen, ein Gesamtkonzept zu erstellen, das die Interessen und Belange der künftigen Nutzergruppen, aber auch

der technischen Aspekte beschreibt. Erst auf dieser Grundlage sollte eine Modellierung erfolgen, an deren Ende die Entscheidung für die Technik steht.

Ein solches Vorgehen gewährleistet am Ende auch, dass eine ESAL seinen erwarteten hohen Wertbeitrag auch tatsächlich leistet. Aber wer sich die Mühe dieses Prozesses macht, erhält am Ende einen schnellen Return on Invest. Denn die stundenlangen Suchzeiten der Mitarbeiter werden auf Sekundenbruchteile schrumpfen und damit viel Zeit freisetzen, zu denen sich die Nutzer ihren eigentlichen Aufgaben widmen können.

Über die DTI

Die DTI Schweiz AG ist seit über 25 Jahren unabhängiger Systemintegrator und Experte rund um die Themen Datenanalyse, Texterkennung und -verarbeitung sowie Datenmanagement und Suche. Sie integriert individuelle, hochflexible Lösungen basierend auf modernsten Technologien und Produkten. Ein Consulting- und Umsetzungsteam kümmert sich um die individuellen Bedürfnisse des Kunden, berät produktneutral und begleitet Umsetzung sowie Roll-out.

Bei Fragen kontaktieren Sie uns gerne

Tel: +41 71 929 77 77

Mail: info@dti.ch

Web: www.dti.ch