

Finden statt suchen – dank Enterprise Search & Analytics

# Mit ESAL die Wertschöpfung steigern

*Strukturierte und unstrukturierte Daten in unterschiedlichen Systemen und großen Speichern – das ist die Realität bei der Datenhaltung in vielen Organisationen.*

*ESAL-Lösungen (Enterprise Search & Analytics) können hier Abhilfe schaffen.*

VON MARC HOFFMANN, SENIOR CONSULTANT SEARCH & ANALYTICS  
BEIM DATENMANAGEMENT- UND -ANALYSE-EXPERTEN DTI SCHWEIZ

Nur selten ist transparent, wo überall die wirklich relevanten Informationen lauern. Unternehmen, die ihre Geschäftsprozesse agiler gestalten sowie End-to-End digitalisieren wollen, müssen jedoch alle Informationen verfügbar machen. ESAL-Lösungen können dabei helfen – sie beruhen auf 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 – und dass wirklich relevante Suchergebnisse auch in den Kontext der jeweiligen Geschäftsprozesse des Nutzers passen. Worauf kommt es nun bei Konzept und Implementierung von ESAL an?

Häufig betreiben die Fachabteilungen eigene Dokumentenmanagement- und Speichersysteme oder unterstützende Informatik-Plattformen, die teilweise unverbunden nebeneinander bestehen. So arbeitet der Vertrieb beispielsweise mit einem Customer Relations Management Tool, das Marketing organisiert sich mit Marketing Automation Tools, und das Management wiederum holt sich regelmäßig Daten, um sie in seinen Enterprise Resource Planning oder Business Analytic Tools (ERP/BA) zu verarbeiten und auszuwerten. Der Grund für diesen Wildwuchs der Systeme sind historisch gewachsene IT-Landschaften, die unkonsolidiert mittlerweile mehr Aufwand bereiten, als dass sie einen Wertbeitrag für die Kollaboration in einem agil arbeitenden Unternehmen leisten. Das führt dazu, dass allzu oft Daten in einem ungepflegten Datensilo gar nicht gefunden werden können, weil sie nicht für Suchprozesse aufbereitet sind. Gleiches gilt für relevantes Wissen, das sich in externen Infor-

mationsquellen wie Datenbanken von Branchenverbänden, weltweiten Patentbehörden sowie öffentlichen und Wissenschaftsinstitutionen befindet. Und 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 will, 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 die Aufbereitung eines Suchindex abzugreifen und zu klassifizieren. Solche Quellen können beispielsweise File-Ablage-, 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 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 Audioda-

teien sowie verpackte oder geschützte Dateien wie ZIP lassen sich ohne Aufbereitung und automatische Auswertung mit sinnvollem Inhalt nicht indexieren. Häufig müssen zunächst Texterkennungsprogramme Dokumente auslesen oder Audioaufzeichnungen in Text (Speech

Anzeige



**emlix**  
embedded linux systems

### Industrial Linux Solutions

- Customized BSPs
- Qt Applications
- Automated Testing
- Processes & Qualification
- Lifecycle Maintenance

[www.emlix.com](http://www.emlix.com)

to Text) umgewandelt werden, bevor sie überhaupt für eine Auswertung beziehungsweise Indexierung bereitstehen. Und bei der Auswertung mittels OCR (Optical Character Recognition) oder Speech-to-Text-Umwandlung passieren Fehler. Zudem existieren unterschiedliche Schreibkonventionen, werden Namen und Begriffe in unterschiedlichen Varianten geschrieben. Hinzu kommen Zahlendreher, in vielen Sprachen unbekannte Umlaute oder Zeichensätze. Denn erst durch Volltexterkennung und Korrektur sowie Integration aller Quellen in einen Suchindex sowie notwendigerweise automatische Verschlagwortung werden Suchprozesse über die komplette Datenbasis effizient.

Effiziente ESAL bringen die Informationen obendrein 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 bekommen, für die sie auch über entsprechende Befugnisse im Quellsystem verfügen.

*Konzept und Modellierung von ESAL entscheiden über Erfolg*

Bei ESAL-Projekten kommt es also maßgeblich auf die Zuordnung der Relevanz von Suchergebnissen, deren Kontextualisierung zu den jeweiligen fachbereichsabhängigen Prozessen sowie die Rechte- und Zugriffsverwaltung 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 von zentraler Bedeutung ist es, 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 Daten-Handling lösen wollen, aber die gesamte Dimension eines ESAL für ihre Compliance erst im Projektverlauf erkennen.

Es gibt auf dem Markt einige Produkte, auf denen eine ESAL aufgebaut werden kann, die aber unterschiedliche Stärken und Schwächen haben. So zeigt sich typischerweise erst bei der Modellierung und der Konzeption eines ESAL-



Marc Hoffmann, DTI Schweiz

„ESAL können helfen, die Suche nach bestimmten Informationen zu beschleunigen und sicherer zu gestalten.“

Projekts, welche Plattform die am besten geeigneten Funktionen bietet, um bestimmte Anforderungen umzusetzen. Zudem sind auch bestehende Tools oder Quellen der Fachabteilungen zu integrieren, für die Konnektoren oder Schnittstellen erst noch programmiert werden müssen. Und ein hochaktueller Aspekt ist, dass oftmals Bestandssysteme bereits in die Cloud verlagert wurden, einige Dateien 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 eigenen sensiblen Kundendaten nicht in einer Cloud-Lösung für die Suche aufbereiten lassen müssen, ist in bestimmten Fällen eine On-Premise-ESAL-Lösung, die stationär in eigenen Servern läuft, einer Cloud-Lösung vorzuziehen. Die zu indexierenden Datenquellen werden typischerweise in einem initialen Schritt aufbereitet und indexiert; im Folgenden werden 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, File-System-Quellen, aber auch Webinhalte erschließen. Für solche Datenquellen beherrschen Konnektoren die inkrementelle Indexierung, also die automatische Indexierung von Datenänderungen.

Für externe Quellen ist es zudem entscheidend, dass sie über eine offene Architektur und Schnittstellen wie ReSt (Representational State Transfer) verfügen. ReSt ist eine Schnittstelle für verteilte Systeme, mit der viele Inhalte abgegriffen werden können und die somit eine Machine-to-Machine-Kommunikation unterstützt. Die über die Konnektoren gesammelten

Daten bereitet die ESAL vor der Indexierung auf, also bei der Überführung in den Suchindex. Über Filter (Conversion) werden die unterschiedlichen Dateiformate (Word, Excel, PowerPoint, PDF, HTML, RTF, TEXT, ZIP) in ein einheitliches, lesbares Format für den Indexer überführt. Zudem wird über Entity-Extraktionen, Regeln oder Integration externer Systeme 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 lässt sich eine KI dafür trainieren, spezifische Inhaltstypen zu identifizieren, typischerweise für Binärdaten wie beispielsweise Bilder und Videos, die keine „lesbaren“ Informationen enthalten. Für Versicherungen ist es häufig relevant, Bilder auszuwerten und Schadenereignisse analysieren zu lassen. 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 sowie 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 Techniken wie Entity Extraction, KI oder ML zum Zuge. So werden Inhalte „verstanden“ und mittels der obigen Techniken klassifiziert und mit zusätzlichen Informationen angereichert. Diese Klassifizierung und Anreicherung lässt sich dann sowohl bei der Relevanz für bestimmte Geschäftsprozesse als auch beim Ranking der Suchergebnisse nutzen.

*Datenaufbereitung entscheidet über Relevanz und Sicherheit*

Die Modellierung einer ESAL und ihrer Funktionen im Suchindex bilden damit also das Kernstück der Suchfunktionen und letztlich die Grundlage für den Erfolg des Systems. Und weil sie eine Schlüsselfunktion in einer ESAL ist, sollte bei der Modellierung bereits analysiert werden, mit welchem Tool eine spezifische Aufgabe am besten gelöst werden kann. Denn die Suchplattform hat die Aufgabe, die Daten intern so zu strukturieren und aufzubereiten, dass selbst in extrem großen Datenmengen und bei komplexen und sehr spezifischen Suchen die relevanten Resultate innerhalb kurzer Zeit bereitstehen. Auf Basis des erstellten Suchindexes 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 einer ESAL.

Je nach Nutzergruppe im Unternehmen können die Anforderungen an das Ranking der Resultate unterschiedlich ausfallen. Deshalb muss eine ESAL auch die Möglichkeit bieten, das Ranking und die Security-Richtlinien auf Basis der Quellsysteme an die verschiedenen Geschäftsprozesse der Nutzer anzupassen. Dies sollte idealerweise schon bei der Modellierung berücksichtigt und bestimmte Nutzergruppen identifiziert 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 dem Produkt. Und ein Einkäufer braucht die Informationen, wie viel das Produkt bei welchem Anbieter kostet. Oder die HR-Abteilung sucht bei einem Namen meist eine Personal-

akte. Die Schadenabteilung einer Versicherung aber sucht Verträge, die eine Person unter ihrem Namen abgeschlossen hat oder in denen sie Begünstigte benannt hat.

Entscheidend ist zudem, welche Suchergebnisse bestimmten Personen überhaupt angezeigt werden. Ein Benutzer darf nur diejenigen Inhalte eingeblendet bekommen, für die 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, die von der ESAL berücksichtigt wird. Weil jedes Quellsystem und jeder Kunde eigene Security-Umsetzungen hat, ist es zwingend, diese schon bei der Konzeption zu berücksichtigen. So wird deutlich, dass das Gesamtkonzept einer ESAL entscheidend für deren Fähigkeit ist, die wirklich relevanten Suchergebnisse zu identifizieren und

von den jeweils nutzerabhängig weniger relevanten Ergebnissen zu unterscheiden.

*Hoher Wertbeitrag und schnelles ROI*

Es gibt also viele Nutzerdimensionen, Stellschrauben und Funktionen, die schon bei der Modellierung einer ESAL vorab zu berücksichtigen sind. Es ist deshalb auch wichtig, dass sich Unternehmen vor der Entscheidung für eine Technik oder eine Plattform die Zeit nehmen, ein Gesamtkonzept zu erstellen, das die Interessen und Belange der künftigen Nutzergruppen, aber auch die technischen Aspekte beschreibt. Erst auf dieser Grundlage sollte eine Modellierung erfolgen, an deren Ende die Entscheidung für die Technik steht. Ein solches Vorgehen stellt letztlich auch sicher, dass eine ESAL ihren erwarteten hohen Wertbeitrag 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, in der sich die Nutzer ihren eigentlichen Aufgaben widmen können. (ak) ■

Die digitale Transformation betrifft die Tätigkeit von Unternehmen in fast allen Branchen.

