

Wertschöpfung aus dem eigenen Datenbestand

Was RAG leistet und was es dafür braucht

WORUM ES GEHT

Ungenutzte Erkenntnispotenziale in Ihren Daten

Das Ablegen von Daten kostet heute fast nichts. In vielen mittelständischen Unternehmen wachsen deshalb Bestände, für die niemand zuständig ist: Verträge, Prüfberichte, Betriebsanweisungen, Projektakten, Ticketverläufe. Wir nehmen wahr, dass oft nur ein Bruchteil der abgelegten Daten überhaupt abgerufen, geschweige denn sinnvoll genutzt wird. Der Rest kostet Speicher und Rechenzeit.

Woran liegt das? Um aus unstrukturiertem Text die Angaben zu gewinnen, auf die es fachlich ankommt, braucht es Menschen, die das Geschäft verstehen. Diese Fähigkeit skaliert nicht mit Speicherkapazität. Dazu kommt, dass für große, aus vielen Quellen zusammengelaufene Bestände oft keine Verantwortlichkeit definiert ist. Ohne Zuständigkeit pflegt niemand die Datenbestände, gibt ihnen keine Bedeutung.



Zwischen dem, was ein Unternehmen an Daten besitzt, und dem, was daraus an Erkenntnis und Mehrwert gewonnen wird, öffnet sich eine wachsende Lücke.

Wird sie geschlossen, ändert sich die Perspektive, neue Möglichkeiten entstehen. Denn: Die Abhängigkeit von Wissen und Erfahrung einzelner Kollegen reduziert sich. RAG erschließt den ganzen Datenbestand, jederzeit und mit Fundstelle. Welche unserer Verträge weichen von der Haftungsobergrenze ab? In welchen Anlagen steckt die Bauteilcharge, die vorige Woche ausgefallen ist? Welche Zusage haben wir vor sechs Jahren gegeben, an die sich heute niemand mehr erinnert? Die Antworten liegen bereits im Unternehmen, bezahlt und abgelegt. Was fehlt, ist der Zugriff. Ihn zu schaffen ist Ihre Entscheidung.

Ein Datenbestand ist so viel wert, wie sich aus ihm erschließen lässt.

Drei Begriffe für dieses Papier

Bestand. Ein abgegrenzter Teil Ihrer Daten mit einer Zuständigkeit und gemeinsamen Zugriffsregeln. Der engere Schnitt ist fast immer der bessere: die Rahmenverträge eines Geschäftsbereichs statt aller Verträge des Konzerns.

Unstrukturiert. Alles, was nicht in Zeilen und Spalten steht: PDF, Bild, Aufnahme, freier Text. In den meisten Unternehmen die größte und am wenigsten erschlossene Masse.

RAG. Retrieval Augmented Generation. Das Verfahren, das einen unstrukturierten Bestand befragbar macht, ohne ihn für jede Frage einzeln aufzubereiten. Ein Sprachmodell, also ein KI-Modell für Text, formuliert die Antwort allein aus Stellen Ihres Bestands. Trainiert wird dabei nichts: Ihre Inhalte bleiben unter Ihren Zugriffsregeln, was die Frage der Datensicherheit entschärft.

Zur Veranschaulichung greift dieses Papier durchgehend auf denselben Bestand zurück: die Verträge eines mittelständischen Unternehmens. Für Prüfberichte, Anlagendokumentation oder Projektakten gilt dasselbe.

DAS RISIKO

Was initial zu prüfen ist

Der Preis für RAG ist stark gefallen. Das Projektrisiko ist jedoch geblieben, unabhängig von der eingesetzten Technologie. Drei fachliche Voraussetzungen sollten erfüllt sein.

Die erste ist ein **aufbereiteter Bestand**. Zwischen der Rohablage von Daten und deren Auswertung liegt eine wichtige Aufbereitungsstufe: Dubletten werden entfernt, Formate vereinheitlicht, jedes Dokument wird seinem Vorgang zugeordnet und mit einer Vertraulichkeitsstufe versehen. Der Aufwand hierfür zahlt sich aus, da ein wichtiges Fundament geschaffen wird: Der aufbereitete Bestand steht jeder weiteren Auswertung zur Verfügung.

Die zweite ist eine **gelebte fachliche Eigentümerschaft**. Eine Führungskraft im Fachbereich verantwortet den Bestand, mit festen Aufgaben in der Data Governance: Sie legt fest, was die Inhalte fachlich bedeuten, sie gibt die Qualität frei, und sie entscheidet, wer welche Teile des Bestands sehen darf. Die Rolle hängt an einer Funktion und hat eine benannte Vertretung, denn diese Entscheidungen fallen laufend an. Jede Frage an den Bestand hat einen Adressaten und jede Antwort einen Maßstab.

Die dritte ist eine **Frage, deren Antwort einen Unterschied macht**: eine Auskunft, die heute niemand geben kann, ein Risiko, das heute niemand überblickt, eine Verhandlung, die heute ohne Faktenlage geführt wird. Eine solche Frage ist initial ausreichend, denn sie liefert den Maßstab für Aufwand und Qualität des ersten Vorhabens. Im laufenden Betrieb des Systems kommen weitere Fragen hinzu, ohne dass die Aufbereitung noch einmal anfällt. Darin liegt der eigentliche Ertrag.

Die drei Punkte gelten je Bestand und nicht kumuliert. Beginnen Sie deshalb dort, wo die Voraussetzungen zutreffen oder gut zu erfüllen sind. Bei Verträgen trifft das in der Regel zu, bei gesammelten Rohdaten selten.

Damit ist die Grundlage gelegt. Was danach über den Erfolg entscheidet, ist die Steuerung der Umsetzung: ein Vorgehen in Phasen, Entscheidungen an festen Punkten und eine Abnahme, die der Fachbereich führt. Grundlage und Steuerung zusammen machen ein Vorhaben erfolgreich, und beides gemeinsam aufzubauen ist der Gegenstand professioneller Beratung.

DAS VORGEHEN

In Phasen bauen, an Entscheidungspunkten steuern

Dieser Abschnitt beschreibt das Vorhaben, der folgende das System. Ein RAG-Vorhaben kann in fünf Phasen gegliedert werden, zwischen denen vier Entscheidungspunkte stehen. Wir empfehlen, den Zuschnitt so zu wählen, dass jede Phase für sich einen Wert liefert und dass ein Abbruch nach jeder Phase möglich bleibt, ohne dass bis dahin Erreichtes verloren geht. Freigegeben wird immer nur die nächste Phase. Budgetrisiken sind dadurch deutlich reduziert.

Die erste Phase klärt den Anwendungsfall. Sie hinterlässt eine Frage mit beziffertem Nutzen und einem benannten Eigentümer. Diese Beschreibung behält ihren Wert auch dann, wenn das Vorhaben hier endet, denn sie ist die Grundlage für jeden anderen Weg zur selben Frage.

Die zweite Phase stellt die Datenreife her und entwirft die Verarbeitung. Hier fällt die Entscheidung, wie der Bestand in Abschnitte geschnitten wird, welche Angaben jeder Abschnitt über seine Herkunft mitbekommt und mit welchem Verfahren seine Bedeutung in Zahlen umgerechnet wird. Der folgende Abschnitt beschreibt diese Schritte im Einzelnen. Die Zuständigkeit ist dabei geteilt: Der Fachbereich bestimmt Zuschnitt und Angaben, die IT wählt das Verfahren.

Die dritte Phase ist der Pilot. Nutzer aus dem Fachbereich arbeiten mit dem System, die Suche wird gegen die vorher festgelegten Fragen justiert und die Qualität der Antworten gemessen. Die vierte Phase härtet: Rechtefilter, Abgleich, Protokollierung und die Einordnung nach dem EU AI Act werden zu Aufgaben mit Terminen. Die fünfte Phase ist der Betrieb.

Die Entscheidungen der zweiten Phase sind der bleibende Teil des Vorhabens. Der durchsuchbare Index ist eine aus den aufbereiteten Texten abgeleitete Kopie. Wird das Umrechnungsverfahren später durch ein besseres ersetzt, muss der gesamte Bestand daraus neu berechnet werden. Deshalb bleiben die aufbereiteten Texte erhalten und werden nicht durch den Index ersetzt.

Der Aufwand verteilt sich ungleich. Die erste Phase kostet wenige Personentage. Den weitaus größten Teil trägt die zweite, weil dort der Bestand aufbereitet und die Verarbeitung gebaut wird. Die dritte und die vierte liegen deutlich darunter, der Betrieb ist laufender Aufwand. Die absolute Größenordnung hängt am Zuschnitt und wird in der ersten Phase bestimmt.

Vor jeder weiteren Phase steht ein Entscheidungspunkt. Dort entscheiden Geschäftsführung und Fachbereich, ob die nächste Phase beginnt. Bewertet wird dabei das Vorhaben im Ganzen, Gesamtnutzen gegen Gesamtaufwand. Eine isolierte Betrachtung der jeweils nächsten Phase führt in die Irre. Die strukturierte Bewertung und die Entscheidungsvorlage liefert die Projektleitung. Damit fällt die Entscheidung fünfmal statt einmal, und jedes Mal auf besserer Grundlage als zuvor.

Von zehn Anwendungsfällen der ersten Phase erreichen erfahrungsgemäß zwei den Betrieb. Das ist das gewünschte Ergebnis. Die acht übrigen haben wenige Personentage gekostet und nicht das Budget eines Jahres.



Ein Entscheidungspunkt ist die Gelegenheit, ein Vorhaben günstig zu beenden. Wer sie auslöst, zahlt später den vollen Preis.

DAS ERGEBNIS

Fragen werden zutreffend und belegt beantwortet

Das Ergebnis eines RAG-Vorhabens ist ein Bestand, den jede berechtigte Person befragen kann. Die Antwort kommt aus den eigenen Unterlagen, nennt ihre Fundstelle und liegt in Sekunden vor. Fragen, die bisher offenblieben, weil niemand dafür einen ganzen Bestand durcharbeiten konnte, lassen sich damit beantworten.

Dahinter stehen vier Verarbeitungsschritte, die in der zweiten und dritten Phase des Vorgehens umgesetzt wurden. Die folgenden Beispiele beziehen sich auf den Vertragsbestand.

Einmal je Dokument: Zerlegung und Umrechnung

Der erste Schritt zerlegt. Jedes Dokument wird entlang seiner Gliederung in Abschnitte geschnitten, damit ein Sachverhalt als Ganzes auffindbar bleibt. Ein Wartungsvertrag über vierzig Seiten ergibt so beispielsweise sechzig Abschnitte. Jeder von ihnen bekommt seine Herkunft mitgegeben: aus welchem Vertrag er stammt, für welchen Kunden dieser gilt und wer ihn sehen darf. Diese Zusatzangaben heißen Metadaten.

Der zweite Schritt rechnet um. Zu jedem Abschnitt wird seine Bedeutung als feste Reihe von Zahlen berechnet. Abschnitte mit ähnlichem Sinn erhalten ähnliche Zahlen, auch bei völlig anderer Wortwahl.

Bei jeder Frage: Suche und Antwort

Der dritte Schritt sucht. Die Frage wird auf demselben Weg umgerechnet und mit dem Bestand verglichen. Wer im Beispiel nach der Kündigungsfrist fragt, findet so auch den Abschnitt, der von der Beendigung des Vertragsverhältnisses spricht. Parallel läuft die klassische Stichwortsuche, denn wer eine Vertragsnummer eingibt, meint genau diese. Beide Ergebnisse werden zusammengeführt.

Der vierte Schritt antwortet. Das Sprachmodell erhält die gefundenen Abschnitte und formuliert daraus die Antwort. Im Beispiel lautet sie: drei Monate zum Quartalsende, nachzulesen in Paragraph 12 des Wartungsvertrags mit der Kundennummer 4711. Das Modell nennt also die Stelle, aus der die Antwort stammt, und es verlässt die gefundenen Abschnitte nicht.

Für das Unternehmen ändert sich damit die Arbeit der Rolle, die den Bestand fachlich verantwortet. Sie legt abschließend fest, welche Antwort als richtig gilt, und sie prüft dort, wo eine Antwort Folgen hat. Ihre Kriterien wirken daher auf jede Frage, ohne dass sie jede Frage selbst beantworten muss.

AUS UNSERER ERFAHRUNG

Was über den Erfolg entscheidet

Der Aufbau eines RAG-Systems ist technisch überschaubar. Was Vorhaben erschwert, liegt fast immer außerhalb der Technik. Vier Muster sehen wir immer wieder. Diese lassen sich vermeiden, wenn sie zu Projektbeginn benannt und geregelt werden.

Wo Vorhaben ins Stocken geraten

Anwendungsfall ohne Eigentümer. Ein Vorhaben startet aus technischem Interesse. Niemand im Fachbereich hat benannt, welche Erkenntnis künftig vorliegen soll und was sie wert ist. Damit fehlt der Maßstab für jede spätere Entscheidung.

Texterkennung unterschätzt. Ein erheblicher Teil der Dokumente liegt als Scan vor, mit Unterschriften, Stempeln und handschriftlichen Zusätzen. Die Texterkennung dafür ist aufwändig und fehlt regelmäßig in der Aufwandsschätzung. Alle weiteren Schritte arbeiten auf dem erkannten Text: Was hier nicht sauber gelesen wird, steht später nicht im Index und ist nicht auffindbar.

Bewertung nach Eindruck. Ein System, das flüssig formuliert und Fundstellen nennt, wirkt überzeugend, auch wenn es in einem von fünf Fällen die falsche Stelle zitiert. Ohne vorher festgelegte Fragen mit den zugehörigen Fundstellen bleibt die Trefferquote unbekannt. Beurteilt wird dann der Eindruck.

Rechte und Aktualität als Nebensache. Ein zentraler Index hebt die Zugriffsrechte der Ablage aus. Er findet die inhaltlich ähnlichste Stelle, gleich ob der Fragende sie sehen darf, und gibt ohne Rechtefilter Gehälter und Konditionen jedem preis, der danach fragt. Der Index muss zudem laufend mit dem Bestand abgeglichen werden. Fällt dieser Abgleich aus, altert er unbemerkt, und es braucht eine benannte Person, bei der die Störungsmeldung ankommt.

Was erfolgreiche Vorhaben anders machen

Sie beginnen klein. Ein Bestand, eine Frage. Der Zuschnitt bleibt klein genug, dass ein Ergebnis in Wochen vorliegt. Die Erweiterung folgt danach.

Sie messen von Beginn an. Vor dem ersten Test entsteht ein Bewertungsset: reale Fragen, zu denen der Fachbereich die richtige Antwort und die richtige Fundstelle vorher festhält. Daran lässt sich jede Änderung messen, auch im laufenden Betrieb.

Sie klären die Rechtsfragen früh. Die Einordnung nach dem EU AI Act und die Beteiligung des Betriebsrats gehören in die vierte Phase, bevor das System in den Betrieb geht. Nutzer müssen erkennen können, dass sie mit einem KI-System sprechen. Sind Beschäftigtendaten betroffen, verschiebt sich die Einordnung.

Sie verlangen, dass das System seine Grenzen meldet. Findet es zu einer Frage keine eindeutige Regelung im Bestand, sagt es das, statt eine Antwort zu formulieren. Dieses Verhalten wird technisch umgesetzt, an eigens dafür gewählten Fragen geprüft und ausdrücklich abgenommen. Ohne diese Abnahme geht es beim ersten Nachjustieren verloren.

Sie planen den zweiten Fall mit ein. Er beginnt wieder bei der ersten Phase, läuft dann aber auf vorhandenem Fundament und erfahrungsgemäß mit etwa halbem Aufwand. Wer das von Beginn an einplant, bewertet den ersten Fall nicht isoliert.

Sie führen die Veränderung. Der Fachbereich übergibt das Suchen an das System und gibt dafür die Kriterien vor, nach denen eine Antwort als richtig gilt. Die IT-Leitung entscheidet über Rechtemodell und Abgleich.

DER NUTZEN

Ihre Potenziale

Der Preis einer einzelnen Frage fällt auf einen Bruchteil. Darin liegt der eigentliche Hebel. Bisher lohnte sich eine Frage an einen unstrukturierten Bestand nur, wenn sie ein eigenes Aufbereitungsprojekt rechtfertigte. Alle übrigen Fragen wurden nie gestellt. Fällt diese Schwelle, wird ein Bestand zum ersten Mal durchgängig befragbar. Vier Potenziale folgen daraus.

Fragen werden beantwortet, die bisher niemand gestellt hat. Welche unserer Verträge weichen von der Haftungsobergrenze ab? Wo haben wir Verfügbarkeiten zugesagt, die wir heute nicht mehr halten? Welche Prüfberichte nennen denselben Mangel, den wir gerade zum dritten Mal beheben? Solche Fragen verlangen heute, dass jemand den gesamten Bestand durcharbeitet. Deshalb werden sie nicht gestellt, und die Antwort fehlt genau dann, wenn sie zählt.

Risiken werden kumuliert sichtbar. Pönalen, Haftungsgrenzen, automatische Verlängerungen und ungenutzte Preisanpassungen liegen verstreut im Bestand. Heute fällt jedes davon einzeln auf, meist zu spät. Als überschaubare Menge werden sie zur Grundlage für Nachverhandlung, für die Versicherungsdeckung und für die Vorbereitung einer Finanzierung.

Auskunftsfähigkeit wird zum Verhandlungsvorteil. Wirtschaftsprüfer, Banken und Zertifizierer verlangen Auskünfte unter Zeitdruck. Dasselbe gilt für den Datenraum einer Transaktion. Wer sie mit Fundstelle in Stunden liefert, verkürzt Prüfungen und verhandelt aus einer anderen Position.

Das Unternehmen weiß, was es weiß. Bisher hängt das Wissen über einen Bestand an wenigen Erfahrenen. Mit ihrem Wechsel geht es verloren, und mit ihrem Kalender ist es getaktet. Ein befragbarer Bestand überführt dieses Wissen in das Eigentum des Unternehmens.

Alle vier haben eine gemeinsame Bedingung. Eine Erkenntnis, die nur ein Dashboard füllt, bleibt wirkungslos. Wert entsteht, wenn sie dort ankommt, wo entschieden wird, beispielsweise als Hinweis im CRM oder als Aufgabe im Vertragsteam. Dieser Prozess ist in den meisten Unternehmen unterentwickelt. Technologie erschließt den Bestand. Über den Ertrag entscheidet, wie das Unternehmen die Antworten prozessiert.

WAS DIE ERSCHLIESSUNG EINBRINGT			
Neue Fragen beantwortet Was ein eigenes Projekt gekostet hätte, kostet jetzt eine Frage.	Risiko kumuliert sichtbar Haftung, Pönalen und Fristen werden steuerbar.	Auskunft in Stunden Prüfer, Banken und Datenräume, beliefert mit Fundstelle.	Das Unternehmen weiß, was es weiß Wissen übersteht jeden Wechsel und jede Vertretung.

Der Wert eines Datenbestands bemisst sich daran, wie viel Erkenntnis sich belegbar aus ihm ziehen lässt und wie viel davon im Prozess ankommt. Was darüber hinaus liegt, ist Speicherplatz.

IHR NÄCHSTER SCHRITT

Verorten Sie sich

Vier Fragen zeigen, wo Ihr Unternehmen heute steht.

Prüfen Sie Ihre Ausgangslage

- 01** Welche Frage an unsere Dokumente hilft dem Geschäft, und was ist ihre bessere Beantwortung wert?
- 02** Welche Funktion ist fachlicher Eigentümer dieses Bestands, und wer vertritt sie?
- 03** Steht zu jedem Dokument fest, zu welchem Vorgang es gehört und wer es sehen darf?
- 04** Welche Erkenntnis fließt heute in unsere Prozesse zurück?

Ihre Antworten sind der Ausgangspunkt. Wo Sie einsteigen, hängt davon ab, welche Frage bei Ihnen offen ist. Steht der Anwendungsfall in Frage, bildet eine RAG-Potenzialanalyse den Einstieg: ein bewerteter Anwendungsfall mit Nutzen, Aufwand und Datenanforderungen. Steht der Bestand in Frage, beginnt es mit einem Datenreife-Check. Zuschnitt und Beteiligte klären wir gemeinsam in einem Deep-Dive-Workshop.

Kühl & Barz begleitet mittelständische Unternehmen bei der Erschließung ihrer Datenbestände mit KI. Sprechen Sie uns an.