

Uwe Crenze

Big Content – und die Rolle von Enterprise Search & Analytics

Daten, Inhalte, Informationen, Wissen – egal wie wir diese Begriffe abgrenzen, wir haben genug davon. Die unterschiedlichsten Datenarten werden in einer Fülle von verschiedenen Systemen verwaltet. Strukturierte Daten in relationalen Datenbanken, Dokumente in ECM-Systemen, E-Mails verbleiben im Mail-Server oder werden in ein Archiv ausgelagert und so weiter. All diese Systeme sind gut gefüllt. Die meisten davon mit unstrukturierten Inhalten. Deshalb bietet sich in diesem Zusammenhang in Abgrenzung zu Big Data der Begriff Big Content an.

Die Analyse und Verdichtung umfangreicher, aber weitgehend homogener Daten (z. B. Logfiles) erfordert ein anderes Herangehen als der aggregierte Zugriff auf heterogenen Content in einer Vielzahl von Systemen. Infolgedessen zielen Big-Data-Lösungen mehr in Richtung Business Intelligence (BI), während die Durchdringung von Big Content in erster Linie auf Exploration und Metadaten-gestützter Suche basiert.

→ Von Enterprise Search zur Smart-Data-Lösung

Suchen hat seine Ursache nicht in Unordnung, sondern im Umfang der vorhandenen Informationsmenge. Big Content impliziert Suchen. Bei Darin Stewart, Research Director bei Gartner liest sich das so: „Enterprise Search is the Engine of Big Content“. [1]

Eine ganzheitliche Betrachtung der Informationslandschaft einer Organisation ist per se ein Big-Content-Thema. Enterprise Search sollte deshalb von vornherein als eine Big-Content-Lösung angesehen werden. Eine Suchinfrastruktur für umfangreiche und komplexe Informationsquellen bildet gleichzeitig das Fundament für viel weiterreichende Lösungsszenarien. Die Inhaltsanalysen einer Suchlösung helfen nicht nur die Suche zu verbessern,

sondern sind der ideale Ansatz für die Erkundung der Informationslagerstätten und die Nutzung der Enterprise Search für die Big-Content-Exploration.

Eine Big-Content-Lösung muss über ein agiles Content Processing verfügen, ohne die Inhalte auf den Quellsystemen neu crawlen zu müssen. Aus diesem Grund wird neben einem Suchindex eine Metadatenbasis benötigt, in der gecrawlte Inhalte abgelegt, zyklisch analysiert und mit neuen Metadaten angereichert werden können. Dies ist die Phase der Veredelung der jetzt gebündelt verfügbaren Inhalte. Weitere Analyseschritte bringen Informationsbeziehungen zutage, mit deren Hilfe Inhaltsobjekte zu einer neuen Qualitätsstufe aggregiert werden. Auf diese Weise kommt Struktur in den Informationsdschungel. Inhalte werden in einem Kontext auffindbar – nicht nur durch Stichwort-basiertes Suchen, sondern durch systematische Exploration. Die Metadatenbasis wird im Zusammenspiel mit dem Suchindex zur wertvollen Wissensdatenbank und zur Grundlage für die Überwachung der Information Governance.

Für eine solche Metadatenbasis ist ein Cluster-Filesystem wie Hadoop [2] mit darauf aufsetzenden MapReduce-Algorithmen aufgrund der damit verbundenen Latenzen nicht geeignet. Trotz der

großen, zu analysierenden Datenmengen muss eine Suchlösung eine echtzeitnahe Indexaktualität besitzen. Außerdem besitzen Big-Data-Lösungen eine Komplexität, deren Administration und Betrieb neben dem eigentlichen Anwendungshintergrund eine technische Herausforderung darstellt und deren Wirtschaftlichkeit oft nicht gegeben ist.

→ Search & Analytics im Enterprise Information Management

EIM-Lösungen befinden sich im Wandel. Viele Anbieter bringen die erweiterte Marktpositionierung ihrer Lösungen durch neue Produktbezeichnungen und Produktbeschreibungen zum Ausdruck. Typische Formulierungen sind Content Analytics & Enterprise Search oder Unified Information Access. Auf der anderen Seite haben die Anwender erkannt, dass große Informationsmengen ohne eine leistungsfähige Suche nicht beherrschbar sind. Suchlösungen werden deshalb immer stärker nachgefragt. Dabei geht der Blickwinkel mehr in Richtung Wissensmanagement als dass er durch Contentmanagement bestimmt wird. Ein Alarmsignal für die ECM-Branche, die Suche oft noch als Angriff auf ihre Produktphilosophie sieht, aber dazu unten mehr.

Von einer modernen Suche werden heute vielfältige semantische Funktionen erwartet und sie sollte darüber hinaus einen kollaborativen Ansatz unterstützen. Durch die rasch anwachsenden Informationsmengen und den damit höheren Stellenwert der Suche im Arbeitsprozess treten zusätzliche Fragen in den Vordergrund, wie die Verfügbarkeit der Suche und ihre Skalierbarkeit. Spätestens dann sind die Anforderungen an die IT-Infrastruktur ein Thema. Die Gesamtkosten einer Suchlösung werden schließlich nicht nur durch Lizenzkosten bestimmt.

Der Ressourcenbedarf einer Enterprise Search kann theoretisch dadurch gesenkt werden, dass die zentrale Suche über einen Federator Suchergebnisse der internen Suchen der verschiedenen Information-Management-Lösungen zusammenfasst. Insbesondere Archivhersteller propagieren solche Lösungen, damit der Archivinhalt nicht noch einmal indexiert werden muss. Dieser Ansatz ist aus mehreren Gründen nicht

zielführend. Werden Ergebnislisten unterschiedlicher Suchlösungen miteinander gemischt, ist die Ermittlung der Relevanz eines Suchtreffers nicht mehr möglich. Auch die Qualität, die Suchsyntax und die in den Suchen vorhandenen Metadatenfelder sind unterschiedlich.

Ein solcher Ansatz ist aus diesen Gründen lediglich als Notlösung zu betrachten. In einer solchen Situation ist es hilfreich, wenn die für eine Enterprise Search benötigten IT-Ressourcen auch für andere Aufgaben mit genutzt werden können. So kommt es zu den sich am Markt herausbildenden Mischlösungen aus Enterprise Search mit anderen Information-Management-Produkten bzw. zur Funktionserweiterung von Enterprise-Search-Produkten.

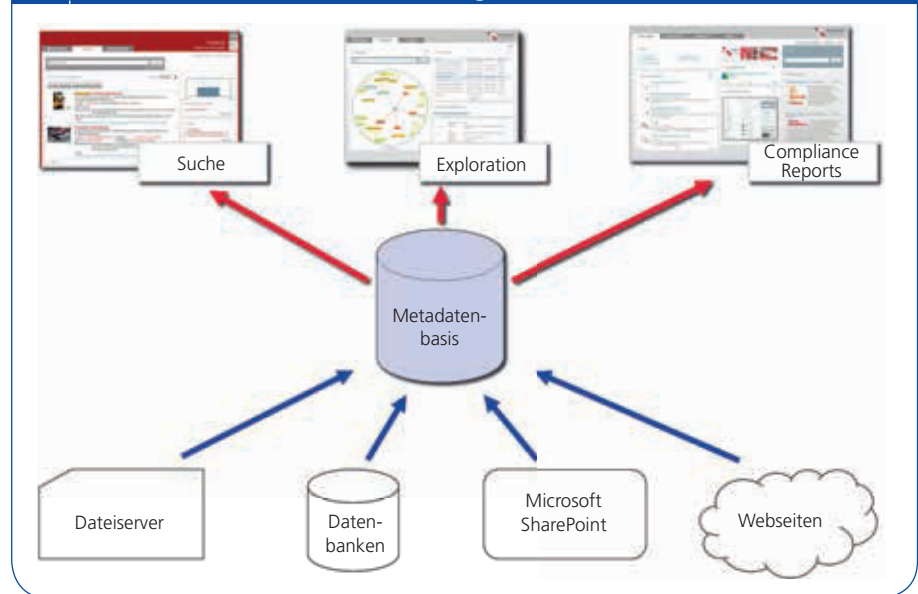
Da Suchen meist mehr ist als das stringente Nachschlagen eines bekannten Fakts, besitzen Enterprise-Search-Lösungen bereits ausgeprägte explorative Funktionen. Deshalb sind z. B. anwendungsübergreifende Reporting-Lösungen und Werkzeuge zur Unterstützung der Information Governance eine naheliegende Ergänzungsfunktion für eine innovative Suche.

Als Benutzerschnittstelle für die personalisierten Informationsströme einer Smart-Data-Lösung sollte ein Social Dashboard zum Einsatz kommen, dass die redaktionelle Bearbeitung und das Teilen von zusammengestellten Rechercheergebnissen gestattet und das als Single-Point-of-Access in Form eines Wissensportals dient.

Ein immer wieder anzutreffendes Paradoxon ist die Antipathie der Hersteller von Enterprise-Information-Management-Lösungen gegenüber Suche. Anstatt sich endlich dieses Thema zu eigen zu machen, beschwören sie den Konflikt zwischen der Notwendigkeit, Ordnung herzustellen, und der mutmaßlich zur Unordnung erziehenden Suche. Dabei wird das immense Potenzial von Suchlösungen als einfache und universelle Zugriffslösung verkannt.

Einfache Ablagestrukturen allein sind nicht ausreichend für das schnelle Auffinden von Informationen. Wissensmanagement benötigt ein dem konventionellen Informationsmanagement übergeordnete Verknüpfung aller in Zu-

→ Metadaten-basierte Anwendungsszenarien



sammenhang stehenden Informationen. Diese Verknüpfungen müssen sich am Unternehmensvokabular ausrichten und repräsentieren letztlich das Wissensmodell der Organisation.

→ Fazit

Analytics & Search sind die Kerntechnologien für die Erschließung großer Informationsmengen und die Gewährleistung ihrer Zugänglichkeit im Rahmen eines umfassenden Wissensmanagements. Nicht nur in Hinblick auf die Durchsuchbarkeit der verteilten Inhalte, sondern in erster Linie für die Gewinnung, Analyse und Aggregation der Inhalte.

Aufbauend auf diesem funktionellen Fundament kann eine moderne, skalierbare Enterprise Search & Analytics-Lösung einen wertvollen Beitrag auf folgenden Gebieten leisten:

- Steigerung der Produktivität von Wissensarbeitern durch schnellen und personalisierten Zugriff auf alle geschäftsprozessrelevanten Informationen,
- Prozessunterstützung durch Suche-basierte InfoApps,
- Überprüfung der Information-Management-Compliance durch Monitoring- und Reporting-Apps,
- wichtige Informationen durch Anreicherung mit Metadaten und Querbeziehungen zu anderen Informationen noch wertvoller machen sowie

- Förderung der Zusammenarbeit durch das Teilen von Rechercheergebnissen.

→ Anmerkungen

- [1] <http://blogs.gartner.com/darin-stewart/2013/05/13/enterprise-search-is-the-engine-of-big-content/>
- [2] Distributed Open Source Computing Platform <http://hadoop.apache.org/>.

→ Der Autor



Dr. Uwe Crenze ist seit 1993 geschäftsführender Gesellschafter der in Dresden ansässigen interface:business Unternehmensgruppe. Nach Berufsausbildung und Informatikstudium promovierte er auf dem Gebiet der Simulation hochintegrierter Schaltkreise. Im Unternehmensbereich interface:projects ist er insbesondere für die strategische Ausrichtung des eigenen Enterprise Search & Analytics Produktes intergator verantwortlich.

✉ crenze@wissensmanagement.net