

VOM SUCHEN UND FINDEN: **ENTERPRISE SEARCH UND WEBSITE SEARCH**



Ob im Web allgemein oder in Unternehmen: Die Menge an digital verfügbaren Informationen wächst laufend, auf praktisch allen Gebieten eines Unternehmens. Und sowohl auf der Unternehmens-Website als auch im Unternehmen intern gilt: Je mehr Informationen es gibt, desto wichtiger ist eine passende Suchlösung. Im Folgenden sollen Enterprise Search und Website Search näher erläutert werden.

„**Enterprise Search**“ ist als Schlagwort mittlerweile verbreitet – gemeint ist häufig das Verfügbarmachen heterogener Inhalte aus unterschiedlichen Quellen mittels einer Suchapplikation. Damit lässt sich das Aufgabenfeld von Suchtechnologie-Spezialisten als Disziplin von „Big Data“ begreifen: Eine große Menge von Informationen, die in alle Richtungen nach unterschiedlichen Regeln und Geschwindigkeiten wächst, lässt sich nur schwer in starre Hierarchien pressen. Damit liefern Such-Applikationen den optimalen – weil flachen und leistungsfähigen – Zugriff auf die unterschiedlichen Arten von Informationen.

Ein Beispiel hierfür ist die klassische Website mit Inhalten aus unterschiedlichen Quellen, z. B:

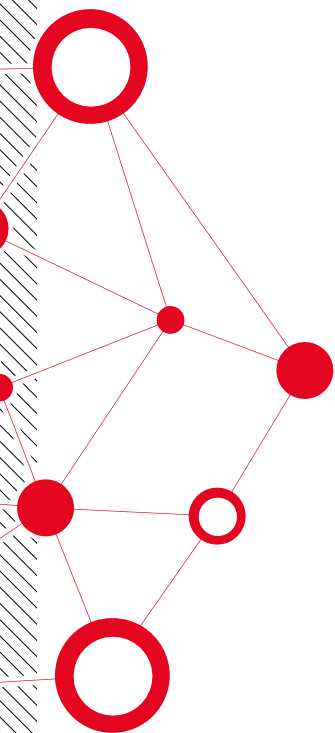
- Redaktionelle Inhalte, über statische Seiten oder CMS-Systeme gepflegt
- Social Media Applikationen, user generated content (Foren, Ratings, Testberichte usw.)
- Onlinekatalog-Inhalte oder Webshops (Produktdaten, Sortimente, Preise, Marketing-Texte, Verfügbarkeiten)
- Dritt-Applikationen, z.B. Händlersuchen oder Leistungskataloge
- Download-Inhalte, die unabhängig sind vom Produktsortiment
- Lose angebundene Websites, die z.B. zur selben Unternehmensgruppe gehören

Diese Vielfalt von Informationen bietet nicht nur für die Suchmaschinenoptimierung und die Gewinnung von organischem Traffic eine große Chance, sondern auch für organisationsinterne Suchlösungen. Um diese Chancen zu verstehen, lohnt es sich zunächst einmal, die Mechanismen einer Suchapplikation genauer unter die Lupe zu nehmen:

Im **Internet** (also von den klassischen, Crawler-basierten Suchmaschinen) wird jedes zu durchsuchende Dokument wie eine Website behandelt:

- Die eingegebenen Suchbegriffe werden entweder in der URL, den Meta-Feldern oder dem Inhalt der Seite gefunden.
- Je nach Fundort, Entfernung zueinander oder Übereinstimmung wird ein sogenannter Score errechnet, der die Relevanz des Treffers definiert. Üblicherweise kommt an dieser Stelle noch der individuelle und nur teilweise transparente Ranking-Algorithmus des Suchmaschinenbetreibers hinzu - meist wird hier versucht, per Algorithmus die Parameter zur berücksichtigen die auch für den menschlichen User relevant sind.
- Ziel von SEO-Maßnahmen ist es daher, in Abhängigkeit vom Berechnungs-Algorithmus den Inhalt für bestimmte Suchwörter besonders „wertvoll“ zu machen und so den Score nach oben zu treiben.

Eine sogenannte Normbildung sorgt zudem dafür, dass die Häufigkeit des Suchworts ins Verhältnis zur Dokumentlänge gesetzt wird. SEO-Marketer optimieren deshalb die Keyword Density eines Artikels.



Organisationsinterne Suchlösungen für Informationsbestände aus unterschiedlichen Quellen unterliegen komplexeren Anforderungen als Internet-Suchmaschinen.

So reichen bei der Suche auf einer Unternehmens-Website (oder jeder beliebigen Enterprise Search Applikation) die drei üblichen Arten der Gewichtung (URL, META-Felder, Inhalt) nicht aus:

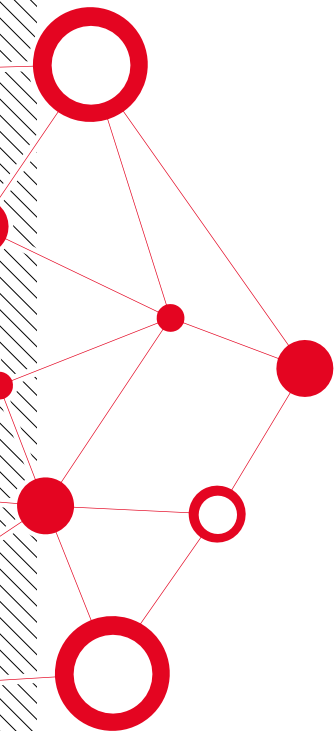
- Oft möchte man den Inhalt aus einer bestimmten Quelle (ggf. sogar innerhalb des Datensatzes) anders gewichten als die Treffer in den anderen Quellen. Sollte nicht ein Ergebnis aus dem Shop wichtiger sein als eines aus dem Archiv der Pressemitteilungen?
- Ebenso ist sinnvoll, Treffer in bestimmten Bereichen des Inhalts, wie zum Beispiel dem Produktnamen, besonders zu werten.
- Im Erstellprozess für redaktionelle Inhalte ist die Länge eines Textes meistens eher vom jeweiligen Redakteur als von der tatsächlichen Wichtigkeit abhängig.

Schon aus diesen Gründen bietet sich der Einsatz einer konfigurierbaren Lösung an, die diesen Faktoren Rechnung trägt.

Da die Daten zumeist aus Systemen kommen, die deutlich mehr als den dargestellten Inhalt verwalten, ergeben sich enorme Möglichkeiten – bis hin zur Aufnahme versteckter Ranking-Kriterien in die Suche, sodass z.B. innerhalb eines Shops die Produkte mit der geringsten Retourenquote (oder dem höchsten Deckungsbeitrag) relevanter dargestellt werden als andere.

Gleichzeitig sind die datenliefernden Systeme häufig, speziell im Mehrsprachen-Umfeld, unterschiedlich aktuell: Social Media Systeme sollten den Suchindex laufend befeuern, während z.B. der Onlinekatalog nur wöchentlich aktualisiert wird und das CMS im 12-Stunden-Rhythmus Updates liefert.

Im **Frontend** wiederum sind die Erwartungshaltungen, auch durch die großen E-Commerce Portale und Suchmaschinen, schon definiert: Einfache und prominente Suche, schnelle Reaktionszeit, Facettenbildung auf die relevanten Attribute eines Bereichs.



Über ein sinnvoll implementiertes Tracking lassen sich anhand von Klickpfaden für lernende Suchen, aber auch über einfache Hits/noHits Auswertungen laufend Optimierungen an der Suche planen – denn der Content auf einer Website ändert sich laufend, die SEO-Maßnahmen ebenfalls und ggf. auch das Produkt-Portfolio. Und so ist eine gut funktionierende Suche weniger eine einmalige Technologie-Umsetzung, sondern ein System, das über die initiale Umsetzung hinaus Aufmerksamkeit verlangt.

Gerne wird vergessen: Die Sucheingabe auf der Website ist ein wichtiger Kanal, über den der Benutzer direkt mit dem Unternehmen kommuniziert und die einzige Möglichkeit, um vorgegebene Navigationsstrukturen zu umgehen. Kleines Gedanken-Experiment: Wann waren Sie zum letzten Mal auf amazon oder einem anderen eCommerce Portal, ohne die Such-Funktion zu benutzen? Eine Suche ist die einfachste Möglichkeit, sich in einem Universum unterschiedlicher Inhalts-Strukturen und Klassifizierungen zurecht zu finden.

In der Technologie-Auswahl stößt der technische Umsetzer, sofern keine fertige Black-Box Lösung eingesetzt werden soll, schnell auf die beiden großen Player SOLR und Elastic Search, die beide auf dem Lucene Index basieren. Die Vor- und Nachteile beider Frameworks sollen hier nicht diskutiert werden – die Auswahl mag sich anhand technischer Gegebenheiten und Anforderungen orientieren, wohl auch was die technische Vorliebe, den Betrieb und die Systemlandschaft angeht.

Mit beiden Frameworks ist ein zeitgemäßes und skalierbares Setup möglich, das eine zukunftsfähige Suchlandschaft bedingt:

- Hohe Performanz und Geschwindigkeit, sowohl beim Indizieren (Near Realtime-Indexing) als auch im Antwortverhalten.
- Gutes Handling großer Datenmengen
- Stabile, einfache und skalierbare Betriebsmöglichkeiten
- Setup anhand von Feeds und Webservices, die flexibel Daten an den Suchindex liefern.
- Manipulation des Scorings, Boostings
- Implementierung zusätzlicher Funktionen, z.B. Unschärfe-Suche, Facettenbildung, Synonym-Interpretation, Teilwortsuche, Berücksichtigung von Stoppwörtern etc.

Der einfache Betrieb und das robuste Handling macht ein solches Setup auch attraktiv für die Umsetzung großer Produktkataloge: So setzt Kittelberger media solutions mit der advastamedia® suite mittlerweile auf einen nativen SOLR XML Export, dessen Struktur sich direkt in einen entsprechend konfigurierten Suchserver importieren lässt. Die komplette Datenpersistenz eines Kataloges inklusive aller Medienobjekte lässt sich damit sehr einfach, performant und skalierbar abbilden.

Fazit: Die Umsetzung einer potenten Such-Applikation auf der Website oder innerhalb des Unternehmens bietet große Chancen, um heterogene Inhalte einfach zugänglich zu machen und zu gewichten.

Mit aktuellen Frameworks lassen sich die meisten Anforderungen moderner Enterprise Search Applikationen gut abdecken. Ebenso wichtig wie das technische Setup ist das Management der Inhalte. Denn: Jede Suche basiert auf Inhalten! Speziell der Bereich der Website Suche bietet auf diese Weise eine große Chance, über die Suche die UX wie auch die Conversion zu optimieren.

Das Urheberrecht für diesen Artikel liegt bei der Kittelberger media solutions GmbH. Sollten Sie Teile hiervon verwenden wollen, wenden Sie sich bitte an uns.

Bayernstraße 8
D-72768 Reutlingen
Tel: +49 7121 6289-0
info@kittelberger.de

