

„Es ist enorm schwierig, über heterogene Strukturen hinweg einen Überblick zu erhalten“

Wie können Informationen für den Gebrauch im Unternehmen verfügbar gemacht werden? Ein Interview mit Dr. Cathrin Senn, Taxonomy Consultant, Factiva / Dow Jones

Umschreiben Sie bitte kurz das Tätigkeitsfeld von Factiva.

Factiva®, ein Unternehmen von Dow Jones, ist der größte News-Aggregator der Welt und bearbeitet jeden Tag mehr als 200000 Nachrichten und Wirtschaftsinformationen aus rund 10000 Quellen in 22 Sprachen und Dutzenden Formaten. Um den weltweit 1,6 Millionen Kunden eine nutzerfreundliche Abfrage der insgesamt 400 Millionen Dokumente anzubieten, haben wir über die letzten 25 Jahre eine eigene patentierte Taxonomie entwickelt und 2001 ein Taxonomy Services Team gegründet, das Kunden bei Projekten in den Bereichen Metadaten, Information Retrieval und Information Architecture beratend zur Seite steht.

Was versteht man eigentlich unter Wissensmanagement?

Wissensmanagement oder Knowledge Management befasst sich mit der Verwaltung von Inhalten in Form von Daten, Informationen und Wissen, wobei Wissen in expliziter oder impliziter Form (tacit knowledge) vorhanden sein kann. Eine der wichtigen Fragen des Wissensmanagements im Bereich von explizitem Wissen ist: Wie kann ich Inhalte greifbar, auffindbar, mitteilbar machen, so dass sie dem Unternehmen zeitgerecht und in effektiver Form zur Verfügung stehen und dabei die Informationsflut eindämmen?

Wie hat sich der Fokus beim Wissensmanagement mit den Jahren verändert?

Aus der Sicht expliziter Inhalte wurde das Verwalten von Information und Wissen früher vor allem mit Archiven, Bibliotheken und papierernen Ablagen in Verbindung gebracht. Inhalte wurden vor allem manuell erschlossen und abgelegt. Das explosionsartige Anwachsen elektronischer Inhalte erschwert oder verunmöglicht diese manuelle Pflege. Um der Informationsflut Herr zu werden, braucht es gewisse Automatismen, die jedoch durch menschliches Eingreifen gesteuert und angepasst werden sollten.

Besonderes Gewicht liegt heute auf den semi- und unstrukturierten Inhalten, die in der Vergangenheit gegenüber den strukturierten Daten in Datenbanken eher vernachlässigt wurden - auch weil die entsprechenden Werkzeuge fehlten, um sie zu erfassen. Zu diesen semi- und unstrukturierten Inhalten gehören z.B. E-Mails, Textdateien und textbasierte Informationen in Datenbanken, aber auch Bilder oder Videodateien. Viele dieser Inhalte werden in „Silos“ getrennt aufbewahrt, das heißt in unterschiedlichsten technischen wie auch inhaltlichen Speichermedien (Mailboxen, persönliche und gemeinsam genutzte Ordnerablagen, Intranets, kollaborative Plattformen, Lotus Notes Datenbanken, Sharepoint-Seiten etc.). Es ist enorm schwierig, über diese heterogene Struktur hinweg einen Überblick zu erhalten und geeignete Zugriffsmechanismen zu schaffen, um diese Informationen gezielt aufzufinden.

Wie löst man das Problem der Isolation von Inhalten?

In gewissen Bereichen macht es durchaus Sinn, Informationen getrennt von anderen aufzubewahren und abzufragen. Es kann aber auch Inhalte geben, die isoliert vorliegen und mit einem übergreifenden Abfragesystem einfacher verfügbar gemacht werden können. Der Endanwender braucht so nicht mehrere Suchen in verschiedenen Systemen abzusetzen, sondern kann sich mit einer Abfrage zu einem bestimmten Thema einen Überblick über die vorhandenen Informationen verschaffen. Oft werden solche Abfragesysteme über das firmeneigene Intranet zur Verfügung gestellt.

Die Abfragefunktion selbst sollte den Bedürfnissen der Endanwender gerecht werden und damit nach Möglichkeit verschiedene Retrieval-Vorgehensweisen zulassen. Die Suchfunktion erlaubt die Eingabe von gezielten Stichworten, während eine Navigationsstruktur ein Durchstöbern und Erforschen von Inhalten ermöglicht und auch in komplementärer Weise zur Suche verwendet werden kann, um Suchresultate zu gliedern und leichter überschaubar zu machen. Eine weitere Möglichkeit, Inhalte auffindbar zu machen, besteht in der Kanalisierung von Informationen basierend auf den Präferenzen und Rollen der Endanwender (Personalisierung). Das strukturierende Konzept, das diese drei Retrievalarten unterstützt oder erst ermöglicht, sind intelligent verknüpfte Metadaten, z.B. in der Form von Taxonomien (hierarchisch organisierte

kontrollierte Begriffe) oder auch Ontologien (kontrollierte Begriffe organisiert in definierten semantischen Beziehungen). Diese Metadaten und ihre Beziehungen unter einander sollten insbesondere auf die Unternehmensziele, die entsprechenden Inhalte und die Bedürfnisse der Anwender abgestimmt sein und dabei kontinuierlich angepasst werden. Die Verknüpfung der Metadaten mit dem Content (Kategorisierung der Information) kann dabei auf manuelle oder automatisierte Weise erfolgen, je nach Kontext des Projekts.

Einfach gesagt, ist die Aufgabe von intelligent verknüpften Metadaten, z.B. in Form einer Taxonomie oder Ontologie, die relevanten Themen und Inhalte des zu durchsuchenden Contents abzudecken und intelligent zu gruppieren. Dies ermöglicht, Informationen gezielt und einheitlich auffindbar zu machen und usergerecht aufzubereiten.

Wie geht man bei der Erstellung einer Taxonomie in einem Retrieval Projekt konkret vor?

Vereinfacht gesagt, sind drei Schritte wichtig:

1. Die Analyse. Zu allererst steht die Frage, welche Ziele das Unternehmen verfolgt und wie ein konkretes Retrieval Projekt in einem bestimmten Bereich diese Ziele unterstützen kann. Ganz wichtig ist es dann, den konkreten Benutzer in den Vordergrund zu stellen und zu verstehen, welche Informationsbedürfnisse vorhanden sind und wie diese abgedeckt werden können. Welches sind also die geschäftskritischen und -relevanten Inhalte, wo werden diese gespeichert und in welchem Schritt des Arbeitsprozesses abgerufen, mit welchen Metadaten sind die Inhalte bereits versehen? Welche Abfragebedürfnisse haben die Anwender?
2. Das Design. Anhand der erhobenen Analysedaten wird dann die Taxonomie erstellt. Eine Taxonomie aufzubauen, kann aufwändig sein, deshalb sollte man keine perfekte Struktur anstreben, sondern einen pragmatischen Ansatz verfolgen. Wichtig ist, den richtigen Detaillierungsgrad der Struktur zu finden. Die wichtigsten Inhalte müssen abgedeckt werden können, aber ist die Taxonomie zu detailliert, führt das dazu, dass zu wenige Inhalte mit einer Kategorie assoziiert werden können; umgekehrt produziert eine zu oberflächliche Taxonomie zu viele Treffer. Neben vielen anderen Aspekten ist auch zentral, die Taxonomie endanwendergerecht ins Benutzerinterface zu integrieren, damit diese die Abfrage möglichst effizient und effektiv unterstützt, ohne dem Endanwender bei der Abfrage hinderlich zu sein.
3. Die Pflege. Diesen Schritt sollte man von Anfang an im Auge behalten. Eine Taxonomie muss immer wieder den sich wandelnden Bedürfnissen des Unternehmens und der verwendeten Sprache angepasst werden. Dafür müssen die benötigten Ressourcen schon in der Anfangsphase des Projektes sichergestellt werden.

In welche Richtung werden sich Metadaten und die Strukturierung von Information weiterentwickeln?

Einen interessanten Trend stellen die so genannten Folksonomies dar. Dies sind vom User (= „Folk“) erstellte Metadaten, mit denen Inhalte versehen werden, ohne dass die zu Grunde liegende Begriffswelt zentral verwaltet wird, wie dies bei kontrollierten Vokabularen, also z.B. Taxonomien, der Fall ist. Solche vom User verwendeten Schlagwörter finden vor allem im Internet Verwendung, z.B. auf <http://del.icio.us>, einer Plattform zur Verwaltung von Weblinks. Solche Initiativen, die auf dem Web 2.0-Gedanken basieren, bringen auch Lerneffekte für das Wissensmanagement in Unternehmen, weil sie aufzeigen, wie Inhalte von Endanwendern miteinander verknüpft werden und welche Begriffe dafür verwendet werden. Vom User generierte Metadaten gibt es auch in Unternehmen, z.B. in Form von persönlichen E-Mail-Ablageordnern. Beim Aufbau einer Taxonomie kann es durchaus sinnvoll sein, sich diese Strukturen bei den Endanwendern anzusehen und bestehende Begriffe oder Strukturen für die Taxonomie weiter zu verwenden.

Sicher ist, dass die Strukturierung und Filterung von Information in Zukunft nicht weniger wichtig wird. Ob dabei Maschinen oder Menschen im Vordergrund stehen, lässt sich schwer abschätzen, wahrscheinlich ist aber, dass es eine Kombination von beidem sein wird. Menschen erkennen inhaltliche Zusammenhänge besser, Maschinen sind objektiver und schneller, wenn es darum geht, vorgegebene Strukturen und die damit verknüpften Inhalte aus einem Informationspool herauszufiltern. Immer wichtiger wird beim zunehmenden Informationsvolumen die Aufbereitung und Visualisierung von Inhalten und den damit verbundenen inhaltlichen Verknüpfungen. Abb. 13.3 zeigt, wie die Resultate einer News-Suche graphisch aufbereitet und strukturiert dargestellt werden können. Es werden dabei einerseits ad hoc wichtige Begriffe aus der Nähe des Suchbegriffs in den Dokumenten herausgefiltert (1), aber auch basierend auf bereits bestehenden

Unternehmens-, Branchen- und Thementaxonomien die Suchresultate nach den wichtigsten Verknüpfungen aufgegliedert (2). Im letzteren Fall wird nicht der ganze Taxonomiebaum angezeigt, sondern es werden nur die wichtigsten Taxonomiebegriffe in Bezug auf die konkrete Suche dynamisch angezeigt, so dass der Enduser nicht mit Informationen überflutet wird und doch eine Möglichkeit erhält, die grosse Menge an Suchresultaten mit einem Klick auf einen der Taxonomiebalken (3) weiter einzuschränken.

Welche Strategien kann man persönlich anwenden, um die Informationsstrukturierung des eigenen Arbeitsplatzes zu verbessern?

Eine klar umrissene Struktur zum Ablegen von z.B. E-Mails und elektronischen Dokumenten unterstützt klar die Wiederauffindbarkeit. Diese Ordnerstrukturen sollten in regelmässigen Abständen überprüft und den Bedürfnissen entsprechend angepasst werden. Ein regelmässiger Clean-up der entsprechenden Inhalte in den Ordnern ist dabei natürlich genauso wichtig. Wie auch bei den Taxonomien sollten die Begriffe selbst klar und eindeutig sein und nicht zu viele Zeichen enthalten. Von Sonderzeichen ist eher abzuraten. Insbesondere sollte darauf geachtet werden, dass nicht zu grosse Listen von Ordnern entstehen. Nach Möglichkeit sollten weitere Ebenen in der Struktur eingeführt werden, sobald eine Ebene mehr als einen Bildschirm ausfüllt oder die Anzahl von 7 bis 14 Foldern überschreitet. Was dabei als akzeptabel empfunden wird, hängt natürlich auch von den persönlichen Präferenzen des Users ab

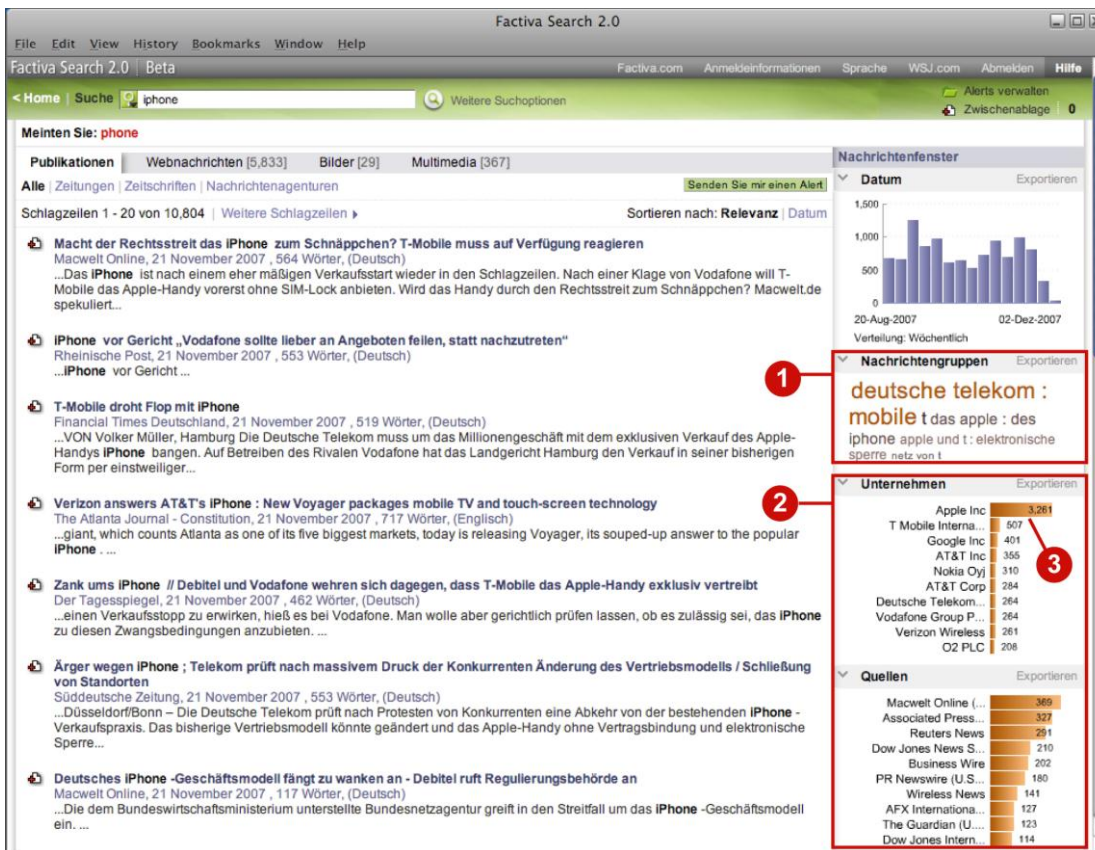


Abb. 13.3. Beispiel für die strukturierte Visualisierung von Suchergebnissen