

Aus: **Uszkoreit, H.** (1999): Sprachtechnologie für die Wissensgesellschaft: Herausforderungen und Chancen für die Computerlinguistik und die theoretische Sprachwissenschaft. In: F. Meyer-Krahmer und S. Lange (Eds.), Geisteswissenschaften und Innovationen, Physica Verlag.

### 3.1 Das Wesen der Information

Die Grundlage der Informationsgesellschaft ist eine nahezu unbeschränkte Verfügbarkeit von digitalisierter Information. Ab einer bestimmten Menge zugänglicher Information ist diese Verfügbarkeit jedoch rein theoretischer Natur, denn die Quantität der Information und die Bandbreite der Netzverbindung steht in keinem Verhältnis zur begrenzten Aufnahmekapazität des Benutzers. Im Gegensatz zu herkömmlichen Datenbanken ist die Information auf dem WWW oder in großen Text-Datenbanken viel weniger vorstrukturiert. Auf der anderen Seite sind die relevanten inhaltlichen Strukturen natürlich weitaus komplexer. Durch die Digitalisierung von großen Teilen des menschlichen Wissen (z.B. digitale Bibliotheken, Filmarchive etc.) wird dieses Problem noch zunehmen. Herkömmliche Techniken der Datenverarbeitung reichen daher nicht aus, um die Informationsflut zu kanalisieren. Der Druck auf den Ausschaltknopf mag im Privatbereich einen wirksamen Ausweg bieten, der wertvolle Freizeit rettet, wenn die Frustration über den Suchaufwand die Neugier übersteigt. In der Wirtschaft wird sich das Problem nicht durch Abstinenz lösen lassen, denn wenn auch nur ein Mitbewerber die Mühe auf sich nimmt, relevante Information schneller zu gewinnen, gerät der Verweigerer ins Hintertreffen.

Bevor ich die Aufgaben und Technologien des Informations- und Wissensmanagements näher beschreibe, möchte ich doch einige Grundbegriffe klären, um die beiden Modewörter etwas konkreter und faßbarer zu machen. Insbesondere werden das die Konzepte *Daten*, *Information* und *Wissen* sein, die wir im Alltag frei durcheinanderzuwerfen scheinen aber bei näherer Betrachtung doch eigentlich sehr konsistent verwenden. Ich will hier mit den Begriffen *Daten* und *Information* beginnen; eine Konkretisierung des Begriffs *Wissen* werde ich im Abschnitt 4. anbieten.

Was ist also der Unterschied zwischen Daten, Information und Wissen? Was unterscheidet Datenbanken von Informationsbanken und Wissensbanken. Sind Informationen keine Daten? Ist Wissen nicht auch Information? In der Tat können die gleichen gespeicherten Symbole in bestimmten Fällen zugleich Daten, Information als auch Wissen sein.

Daten sind Zeichenkonfigurationen, die gespeichert und verarbeitet werden können. Mit gespeicherten Daten können wir viele Dinge tun, ohne daß wir wissen, wofür sie eigentlich stehen und warum sie gespeichert wurden. Datenschutz, zum Beispiel, bezieht sich in erster Linie auf die Daten. Datenschutz ist kein Wissensschutz. Wenn ein Arzt mit schützenswerten personenbezogenen Daten arbeitet, muß er diese Informationen nicht vergessen, wenn er nach Hause geht. Der Datenschutz verbietet zwar, daß er Patienteninformationen ungeschützt auf dem Familien-PC installiert. Sollte er jedoch Details aus den Daten seiner Patienten im Wirtshaus erzählen, verletzt er nicht den Datenschutz sondern ganz einfach die ärztliche Schweigepflicht.

Ein viel schwierigerer Begriff ist der der Information. Sehr verschiedene Dinge werden als Information bezeichnet. Was haben zum Beispiel geheimdienstliche Informationen, dunkle Wolken am Himmel, die ein Gewitter ankündigen, und das Genom des Regenwurms gemeinsam? Der Verweis auf die Bedeutungsfunktion von Zeichen führt hier in die Irre, denn viele Arten von Information, wie z.B. die Wolken am Himmel sind zwar Anzeichen, aber keine bedeutungstragenden Zeichen. Informationen sind jedoch immer Muster, die die Zustandsübergänge eines zustandsverändernden Systems beeinflussen. Nicht die Übergänge werden durch Information bewirkt, sondern lediglich die Auswahl zwischen mindestens zwei Folgezuständen. Dabei kann einer der möglichen Folgezuständen durchaus gleich dem Ausgangszustand sein. Vorbedingung ist, daß das System bei der Auswahl auf die möglichen Muster eingestellt ist. Das klingt nicht nur sehr abstrakt, das ist auch sehr abstrakt. Daß diese Begriffsbestimmung dennoch richtig ist, mögen die folgenden Beispiele verdeutlichen:

- Teile der DNS-Molekülketten sind Information, weil sie die Ontogenese steuern, das heißt, weil sie zum richtigen Zeitpunkt eine Entscheidung bei der Ausprägung von artspezifischen oder individuellen Merkmalen beeinflussen.
- Die Häufungen von Wasserdunst am Himmel wären keine Information ohne die Lebewesen, die die dräuenden Wolken als Vorzeichen eines Gewitters deuten könnten.
- Für die antiken Auguren waren gewisse Eigenschaften der Flugformation von Zugvögeln Information, die für uns keine Information mehr sind.

Somit ist Information immer Information in Bezug auf ein bestimmtes System. Für uns Menschen dient Information in der Regel als Basis für bewußte Entscheidungen. Aber auch unbewußte Entscheidungsvorgänge werden durch Information beeinflußt. Der gesteuerte Zustandsübergang kann außerdem die Veränderung oder Erhaltung einer Stimmung sein. In diesem Sinne sind natürlich auch Sonaten oder abstrakte Ornamente Information.

Die Informationsgesellschaft wird ihren Bürgern den nahezu unbegrenzten Zugang zu Information ermöglichen, unbehindert durch Mengenprobleme sowie räumliche und zeitliche Distanz. Das bedeutet natürlich, daß jedes Mitglied dieser Versorgungsgemeinschaft nur einen winzigen Bruchteil der angebotenen Information nutzen kann und wird. Es bedeutet auch, daß es sehr viel Information geben wird und muß, die nie genutzt wird. Wer sollte denn sicher vorhersehen können, welche Information wirklich gebraucht werden wird. Wenn die oben vorgenommene Begriffsklärung stimmen soll, birgt der Begriff "ungenutzte Information" eine Kontradiktion.

An dieser Stelle muß der Sprachwissenschaftler die sinnvolle Definition des Informationsbegriffs vor seiner Aufweichung bewahren, indem er darauf hinweist, daß wir in unserer Alltagssprache oft den erwarteten Gebrauch von Objekten zur Grundlage ihrer Bezeichnung machen. So wie ein in Acrylharz gegossener Glückspfennig oder Dagobert Ducks erster selbstverdienter Dollar Zahlungsmittel sind, auch wenn sie nie zum Bezahlen eingesetzt werden, so wie auch ein nie benutzter, im Museum ausgestelltter Keltenkamm immer ein Gebrauchsgegenstand war und

bleiben wird, so bezeichnen wir in einem erweiterten Sinn auch potentielle aber ungenutzte Information als Information.

## Aufgaben und Probleme des Informationsmanagements

Das Ziel des Informationsmanagements besteht darin, potentielle Information bestmöglich nutzbar zu machen und somit in wirkliche Information zu verwandeln. Die praktische Aufgabe ist die Verwaltung und Nutzbarmachung von sehr großen Informationsmengen. Diese soll dem einzigen Zweck dienen, die Information den Entscheidungsträgern genau dann zu liefern, wenn sie benötigt wird, und sie überdies so zu präsentieren, daß sie den Nutzer bei seiner Entscheidung effizient und effektiv unterstützt.

Dafür wird die Information

- akquiriert,
- kategorisiert,
- gefiltert,
- zusammengeführt,
- strukturiert,
- dem Benutzer zugeführt und
- adäquat präsentiert.

Die Akquisition geschieht zum Beispiel durch das Abonnement von Zeitungen, Agenturmeldungen, Informationsdiensten, durch die Beschaffung von elektronischen Büchern, Gesetzestexten, Berichten oder Patentschriften, durch das Einscannen von Korrespondenz, durch Anfragen an spezialisierte Datenbanken und durch die gezielte Suche auf dem WWW.

Im Gegensatz zur Verwaltung von Information in einer herkömmlichen Datenbank ist das Informationsmanagement beim WWW oder großen Textdatenbanken mit Problemen konfrontiert, die sich aus der Repräsentation der Information ergeben.

Die problematischen Eigenschaften der Informationsbestände sind:

- **Distributivität:** Die Information liegt verteilt auf verschiedenen Maschinen.
- **Heterogenität:** Die heterogene Natur der Information ergibt sich durch die Vielzahl der Dokumentformate, durch Multilingualität, Multimedialität, Multimodalität, und durch verschiedene Konventionen für die Darstellung von Metainformation.
- **Unstrukturiertheit:** Es gibt weder eine einheitliche Klassifikation, noch eine einheitliche interne Strukturierung der Dokumente. Auch finden wir keine durchgängige und schon gar nicht eine verlässliche Hypertextverknüpfung der Dokumente.
- **Redundanz:** Sehr viele Informationen liegen in mehreren Dokumenten vor.

Ein Teil dieser Probleme läßt sich mit vorhandenen technischen Möglichkeiten lösen. Das WWW ist eine Lösung für die Distributivität, die sich im täglichen Einsatz bewährt. Für die Vielfalt an Dokumentformaten gibt es Konvertierungsprogramme und Austauschformate. Die vorhandenen technischen Lösungen für die Bewältigung der Multimedialität (verschiedene Medien: Sprache, Bilder, Klänge) und Multimodalität (z.B. herunterladbare Filme vs. Realzeitübertragung) werden ständig verbessert. Neue Hard- und Software, vor allem aber mächtigere Repräsentationssprachen und Protokolle für das WWW sind in Vorbereitung.

Heute nur ansatzweise lösbar sind alle Probleme, die eine inhaltliche Verarbeitung der Dokumente erfordern. Dazu gehören die Unstrukturiertheit der Dokumente und die Redundanz der Information. In Datenbanken sind Informationen, die relevant für den Benutzer sind, so strukturiert, daß auch die Maschine sie als Information verarbeiten kann. Redundanz läßt sich somit vermeiden. Die Maschine kann die Daten sortieren, auf ihre Konsistenz überprüfen und auf Suchanfragen hin dem Benutzer präsentieren.

Im Information Management hat man es hingegen vorwiegend mit Informationen zu tun, die sich nicht in das Datenmodell einer Datenbank einpassen lassen. Daher werden sie auch als unstrukturiert bezeichnet. Die Text- und Multimediadokumente sind aber alles andere als unstrukturiert. Die Information in den Dokumenten ist in Wahrheit viel stärker strukturiert als die Datensätze in Datenbanken. Jedes Dokument hat eine Dokumentstruktur, die die Bestandteile des Dokuments klassifiziert und zueinander in Beziehung setzt. Die Bestandteile können Überschriften, Spalten, Textfelder, Bilder, Seitenzahlen und Fußnoten sein. Texte bestehen z.B. aus Kapiteln, Absätzen und Sätzen. Die Textstruktur stellt die Beziehungen zwischen diesen Teilen her. Jeder Satz hat eine Struktur, die die Hierarchie der Phrasen und Wörter des Satzes repräsentiert. Wörter setzen sich aus Stämmen, Vorsilben oder Endungen zusammen.

Dokumente, Texte, Sätze und Wörter haben jedoch nicht nur eine syntaktische, sondern auch eine semantische Struktur. Die semantische Struktur ordnet den syntaktischen Einheiten ihre Bedeutungen zu und stellt zusätzliche Beziehungen zwischen Teilen her, die explizit oder implizit aufeinander verweisen. So stellt sie den inhaltlichen Zusammenhang zwischen Sätzen und Absätzen her und verbindet Anaphern mit ihren Referenten. Aus der semantischen Struktur folgt, wie sich die Bedeutungen von Einheiten aus den Bedeutungen ihrer Teile ergeben.

Das Problem für das Informationsmanagement ergibt sich daraus, daß bisher nur der Mensch die vielschichtigen Strukturen der Dokumente erkennen und verarbeiten kann. Eine inhaltliche Verarbeitung der Dokumente, wie sie für eine effektive Kategorisierung und Filterung vonnöten ist, erfordert aber die Erfassung der syntaktischen und semantischen Strukturen. Daher arbeiten viele Presse- und Informationsdienste nach wie vor sehr personalaufwendig mit Fachkräften, die die Information sichten, kategorisieren, filtern und verdichten.

Überall dort, wo Entscheidungen von großer wirtschaftlicher oder politischer Tragweite von der Qualität der Information abhängen, kann heute auf die menschliche Kompetenz nicht verzichtet

werden. Im Bundespresseamt wird die sogenannte Kanzlermappe, in der dem Bundeskanzler die wichtigsten Pressemeldungen des Tages vorgelegt werden, jeden Morgen von Fachkräften zusammengestellt. Auch die Sichtung der rund um die Uhr eintreffenden Meldungen der Nachrichtenagenturen und deren Kategorisierung für die Verteilung an die Ressorts wird im Schichtbetrieb durch gut ausgebildete Fachleute mit langjähriger Erfahrung erledigt. Ein Redakteur vom Dienst überfliegt jede eingehende Agenturmeldung auf dem Bildschirm und verteilt sie innerhalb weniger Sekunden durch Tastendruck an die relevanten Empfänger in den Ministerien und anderen Bundesbehörden. Nach seiner Schicht stellt er ein Bulletin zusammen, das einen Überblick über die wichtigsten Nachrichten der vergangenen Stunden gibt.

Viele große Firmen und öffentliche Verwaltungen haben sich eigene Informationsdienste geschaffen, die nach genauen Vorgaben relevante Informationen aus den Medien zusammensuchen. Viele andere beauftragen externe Anbieter mit dieser personalintensiven Aufgabe. Die Menge der Information, die durchgesehen werden muß, nimmt jedoch ständig zu. Mindestens genauso schnell wächst auch aber die Anzahl der Nutzer von Informationsdiensten. Die Suchmaschinen auf dem WWW und kostenlose elektronische Informationsdienste ermöglichen heute den mehr oder weniger gezielten Zugang zu vielfältigen Quellen und stellen somit eine Art Grundversorgung her. Auch die Software für die Verwaltung der selbsterzeugten Information ist erschwinglich geworden. Durch die Verwendung der WWW Technologie in internen Netzen, sogenannten Intranets, kann ein Unternehmen sich sehr schnell eine gemeinsame Informationsbasis schaffen.

So scheint es zunächst, als würde die Information immer billiger. Die Kosten für die Replikation und Transmission der Daten können heute vernachlässigt werden. Somit fallen die Kosten für die Bereitstellung der Information nur einmal an, ganz gleich wieviele Nutzer diese Information findet. Bei Millionen von Nutzern läßt sich auch eine teure Bereitstellung durch ein einziges kleines Werbebanner finanzieren. Doch der Schein trügt. Was billiger wird, ist tatsächlich nur die Information für Entscheidungssituationen, die millionenfach vorkommen. Die meisten Entscheidungssituationen mit großer wirtschaftlicher oder politischer Tragweite sind jedoch Einzelsituationen, das heißt, sie sind einzigartig in der Zusammenstellung und Gewichtung ihrer Kriterien. Die Information für solche Entscheidungen ist nach wie vor sehr teuer. Zwar sind die Bestandteile der relevanten Information leichter und kostengünstiger erhältlich, die Kosten für die Filterung und adäquate Zusammenstellung steigen jedoch mit der Menge der vorhandenen potentiellen Information. Es gab eine Zeit, in der die Medien, die durchgesehen werden mußten, teurer waren als die Arbeitskraft, die die Sichtung und Filterung vornahm. Diese Zeiten sind natürlich lange vorbei. Qualitativ hochwertige Information wird auch in der Zukunft nicht umsonst erhältlich sein. Ohne die Sprachtechnologie werden die Kosten aber mit der Menge der zu berücksichtigenden Information wachsen.