

6 Anwendungsbeispiele Online-gestützter Betriebswirtschaft

6.1 Elektronische Informationsdienste in globalen Netzwerken

Von Prof. Dr. Wolfgang G. Stock

Im Matthäus-Evangelium findet man eine frühe Formulierung des Wissensmanagement:

10 Die Jünger kamen zu Jesus und fragten: „Warum sprichst du in Gleichnissen, wenn du zu den Leuten redest?“

11 Jesus antwortete: „Euch hat Gott die Geheimnisse seines Planes erkennen lassen, nach dem er schon begonnen hat, seine Herrschaft in der Welt aufzurichten; den anderen hat er diese Erkenntnis nicht gegeben.“

12 Denn wer viel hat, dem wird noch mehr gegeben werden, so dass er übergenug haben wird. Wer aber wenig hat, dem wird auch noch das wenige genommen werden, das er hat.

13 Mit diesem Grund rede ich in Gleichnissen, wenn ich zu ihnen spreche. Denn sie sehen zwar, aber erkennen nichts; sie hören zwar, aber verstehen nichts. ...“

Der Bezug in Matthäus 13,12 ist die Erkenntnis, also Wissen. Wer viel Wissen hat, dem wird noch mehr gegeben; wer wenig Wissen hat, der verliert im Laufe der Zeit auch noch das wenige. Diese Lektion der Bibel hat auch für Unternehmen und Verwaltungen heutiger Tage Gültigkeit. Mitentscheidend für den Erfolg einer Institution ist seine Wissensbasis; diese muss groß genug sein, um weiteres relevantes Wissen zu finden, zu verarbeiten und so nutzbringend in die Unternehmens- und Verwaltungsprozesse einfließen zu lassen.

6.1.1 Total Knowledge Management

6.1.1.1 Externes Wissen in der betrieblichen Informationswirtschaft

Wissen (im Sinne von Informationsinhalt, neudeutsch „Content“) kann nach seiner Herkunft zwei Quellen haben: Es wird entweder im Unternehmen von dessen Mitarbeitern kreiert („internes Wissen“) oder befindet sich im Umfeld des Unternehmens („externes Wissen“). Der systematische Umgang mit internem Wissen wird üblicherweise als „Knowledge Management“ bezeichnet (vgl. Probst et al. 1998; Sveiby 1998). Häufig werden in betrieblichen Informationssystemen die externen Wissensspeicher außer Acht gelassen. In diesen Wissensspeichern, die Anfang des 21. Jahrhunderts zunehmend in elektronischer Form (und damit als elektronischer Content - E-Content) vorliegen, finden sich wissenschaftliche Informationen nahezu aller Fächer, technische Informationen (u.a. vollständig alle Informationen über Patente), Unternehmensdossiers (recht komplett für börsennotierte Unternehmen, aber auch über einige Millionen weiterer Firmen), Volltexte der Nachrichten aus Tausenden von Tages- und Wirtschaftszeitungen sowie aus Presseagenturen, Informationen über aktuelle Ausschreibungen und Fördermaßnahmen, juristische Informationen zu gesetztem Recht genauso wie zu grundsätz-

lichen Urteilen, biographische Informationen zu Führungskräften bis hin zu mehreren Millionen wirtschaftsstatistischer Zeitreihen (vgl. Stock 2000a). Es ist zeit- und personalintensiv, ja zum Teil sogar unmöglich, in einem Unternehmen Wissen zu reproduzieren, das schon – irgendwo – vorliegt. Neuentdeckungen bereits bekannten Wissens haben stets den unangenehmen Flair der „Wiedererfindung des Rades“. Zur Einstimmung sei nur ein Beispiel im Kontext technischen Wissens genannt: 1999 kamen am Deutschen Patent- und Markenamt insgesamt rund 94.000 Ideen zur Anmeldung, die die Anmelder für neuartige Erfindungen hielten. Im gleichen Jahr wurden jedoch nur 15.000 Patente erteilt. Für die Mehrzahl der sog. „Erfindungen“ – die Differenz ist seit Jahren ähnlich hoch wie 1999 – hat das Patentamt kein Patent erteilt oder die Anmelder haben erst gar keinen Prüfantrag gestellt – dies sind vielfach Formen völlig unökonomischer Doppelforschung, resultierend aus mangelnder Berücksichtigung externen Wissens.

Es ist demnach ein rationales ökonomisches Verhalten, die Gesamtheit des externen Wissens in die betriebliche Informationswirtschaft einzubeziehen. Die Bereitstellung erfolgt punktgenau: Externes Wissen kommt – just in time – bei konkreten Informationsbedarfen des Unternehmens, seiner Abteilungen und seiner Wissenschaftler ins Haus, um mit dem internen Wissen integrativ zu verschmelzen. Die Gesamtheit von internen und externen Informationen hilft, „kritische Situationen“ zu meistern. Solche Situationen sind

- Entscheidungsvorbereitungen
- Wissenslücken
- Frühwarninformationen.

In den ersten beiden Fällen ist der Informationsmangel bekannt; Entscheidungen wollen durch möglichst umfassende Informationen vorbereitet, Wissenslücken (z.B. bei Forschungs- und Entwicklungsprojekten) gestopft werden. Im dritten Fall weisen erst die Informationen darauf hin, dass überhaupt eine kritische Situation eingetreten ist. Das Frühwarnsystem signalisiert Gefahren des Wettbewerbs, etwa den Markteintritt neuer Wettbewerber oder beachtenswerte Änderungen bei Lieferanten oder Abnehmern. Das „richtige“ interne Wissen (einschließlich der „richtigen“ Wissensträger, also der Mitarbeiter) kombiniert mit dem „richtigen“ externen Wissen zeigt Chancen für das Unternehmen auf und warnt gleichzeitig vor Risiken. Die Beachtung beider Wissensquellen – internes Wissen auf der einen und externes Wissen auf der anderen Seite – wollen wir mit „Total Knowledge Management“ etikettieren. Der interne Aspekt des Total Knowledge Management wird u.a. im Beitrag von Matthias Fank in den Mittelpunkt gestellt (siehe Kapitel 4.2); wir konzentrieren uns hier auf das externe Wissen und seine Einbeziehung in betriebliche Informationssysteme. Schwerpunkt der Betrachtung ist das in elektronischen Speichern vorliegende Wissen, also dasjenige Wissen, das via Internet oder anderer Datennetze online erreicht werden kann. Da sich die Anzahl an Datensätzen im World Wide Web sowie in noch größerem Umfang in den kommerziellen Online-Archiven im zweistelligen Milliardenbereich bewegt, ist eine Recherche externen Wissens vergleichbar mit dem berühmten Suchen nach einer Stecknadel im Heuhaufen. Es bedarf eines elaborierten Fachwissens,

sowohl die passenden Wissensspeicher ausfindig zu machen als auch darin optimal zu suchen und zu finden.

6.1.1.2 Quellen elektronischer Informationen

6.1.1.2.1 Wissen und Information

Zunächst sei ein kurzer terminologischer Exkurs gestattet: Was heißt „Information“ und was ist „Wissen“? Die lateinische Quelle von „Information“ ist *informare*, also „in eine Form bringen“ bzw. „bilden“. Wenn Information die „Form“ ist, was ist der „Inhalt“, der mittels dieser Form bewegt wird? Seit der frühen Neuzeit lässt sich die Verwendung des Informationsinhalts als „Wissen“ verfolgen. „Informieren“ bedeutet im passiven Sinne über etwas kundig gemacht werden sowie im aktiven Sinne jemanden über etwas in Kenntnis setzen. Das „Etwas“ ist jeweils Wissen. Wissen ist damit statisch und liegt in Wissensspeichern. Erst durch die Information wird Wissen dynamisiert. „Information ist Wissen in Aktion“, formuliert zutreffend Rainer Kuhlen (1996, S. 42).

Je nach Art des Wissensspeichers unterscheiden wir „subjektives Wissen“ und „objektives Wissen“. Das subjektive Wissen ist stets an eine Person gebunden. „Eine Person P weiß, dass p“, bedeutet, dass unsere Person P der Überzeugung ist, dass der Sachverhalt p wahr ist, und p ist in der Tat wahr. Die epistemische Logik kennt jedoch noch weitere Prädikate, etwa Meinen, Glauben oder Lügen, bei denen weder die Überzeugung noch die Wahrheit zutreffen müssen. Aber die Inhalte dieser unsympathischen Formen menschlicher Erkenntnis können auch in Aktion gesetzt werden und erscheinen als Informationen. Damit sind Informationen keineswegs identisch mit in Bewegung gesetztem subjektiven Wissen, sondern umfassen mehr bis zum Zerrbild des Wissens in der Lüge. Andererseits enthält das subjektive Wissen Momente, die die Person gar nicht explizit als ihr Wissen wahrnimmt. Dieses „implizite“ Wissen lässt sich demnach nur schwer in Bewegung setzen, sprich via Information weitergeben.

Objektives Wissen ist grundsätzlich von Personen unabhängig und befindet sich auf (irgendeinem) physikalischen Träger, z.B. auf Papier (wie dieses Buch) oder in elektronischen Dateien. Im Sinne von Karl R. Popper mit seinen drei „Welten“ – I: Materie, II: Geist (menschliche Psyche), III: Produkte des menschlichen Geistes – befindet sich das objektive Wissen in Welt III. Das derart fixierte Wissen kann jederzeit reaktiviert werden, so dass verständlich ist, wenn sich Unternehmen bemühen, subjektives Wissen der Mitarbeiter in objektives Wissen zu transformieren: Auch wenn der Mitarbeiter persönlich nicht greifbar ist (weil er an einem anderen Standort arbeitet oder die Firma bereits verlassen hat), so ist es doch sein Wissen. Voraussetzung ist, dass die Rückverwandlung in subjektives Wissen klappt, dass jemand da ist, der das objektive Wissen versteht und in der Lage ist, Aktionen daraus abzuleiten. Dies gilt für die Speicher internen Wissens genauso wie für die des externen Wissens. Das Ziel der Umwandlung von objektivem Wissen via Information in handlungsrelevantes subjektives Wissen ist die Bereitstellung des richtigen Informationsinhalts (objektiven Wissens):

- zum richtigen Zeitpunkt
- am richtigen Ort
- im richtigen Umfang
- in der richtigen Form
- mit der richtigen Qualität,

wobei „richtig“ stets heißt, dass der Empfänger der Information das übertragene objektive Wissen adäquat in subjektives Wissen umwandeln kann.

6.1.1.2 Wissensspeicher World Wide Web

Wenn wir gerade gesagt haben, dass Information außer Wissen noch andere Inhalte wie Annahmen oder Lügen überträgt, so trifft dies in besonderem Maße auf das World Wide Web zu. Wir können uns ohne eingehende Prüfung nie sicher sein, objektives Wissen vor uns zu haben. Unternehmen werden auf ihren Internetseiten kaum Negative präsentieren; hier geht es um kommunikationspolitische Maßnahmen, nicht um Wahrheit. Wir sollten aber trotzdem nicht an den Inhalten des World Wide Web vorbeigehen, da durchaus nützliches Wissen vorliegt – eingebettet in einen Haufen völlig unnützer Informationen. Einige Formen objektiven Wissens liegen exklusiv im Web vor (beispielsweise Working Papers wissenschaftlicher Institutionen, Unternehmensnachfolgebörsen oder Datenbanken öffentlicher Ausschreibungen), so dass die betriebliche Informationswirtschaft bei der Recherche externer Informationen nicht um das Internet herumkommt. Eine zentrale Rolle bei der Suche nach Informationen im Web spielen seit etwa 1995 Suchwerkzeuge und – etwas später – Portale, die mehr oder minder komplette Datenbasen und mehr oder minder gute Recherchewerkzeuge zur Nutzung offerieren. Da Internetrecherchen zwar kostenfrei, aber durchaus sehr zeitaufwendig sind, ist die Kenntnis des Leistungsumfangs der Suchsysteme und Portale wesentlich, will doch die richtig formulierte Anfrage bei den richtig ausgewählten Systemen gestellt werden.

6.1.1.3 Wissensspeicher Online-Archive

Die Branche der kommerziellen elektronischen Informationsdienste arbeitet seit etwa 1970 daran, qualitativ geprüfte Informationen und damit (möglichst) objektives Wissen (und nur dieses) auf dem Informationsmarkt anzubieten. Die Wertschöpfungskette dieser Branche besteht aus Datenbasisproduzenten, Datenbank-anbietern sowie Informationsvermittlern. Die Produzenten erstellen eine oder mehrere Datenbasen in fachlich überschaubaren Gebieten entweder auf der Basis eigener Inhalte (etwa Unternehmensdossiers wie Hoppenstedt, Zeitungsartikel wie die F.A.Z. oder Zeitreihen wie Eurostat) oder sie produzieren Nachweisinformationen mit Hinweisen auf (meist) wissenschaftlich-technische Literatur. Datenbankanbieter bedienen entweder den elektronischen Offline-Bereich (CD-ROM und DVD) oder die Online-Welt. Wegen der oftmals sehr großen Berichtszeit der Online-Informationsanbieter bezeichnet man diese auch als „Online-Archive“. Online-Archive bündeln die Produkte der Informationsproduzenten innerhalb einer einheitlichen Retrievaloberfläche. Sie bieten mehrere hundert bis mehrere tausend Einzeldatenbanken in einem System an. Dabei reicht das Spekt-

rum der Datenbanken von Bereichen wissenschaftlicher Grundlagenforschung (etwa INSPEC für die Physik) bis hin zu alltäglichen Informationen (wie z.B. die ARGETRA mit Zwangsversteigerungsterminen deutscher Amtsgerichte). Teilweise vermarkten Datenbasenproduzenten ihre Produkte auch unabhängig von Online-Archiven eigenständig in Datennetzen. Man spricht dann von „Independents“. Die Inhalte der gespeicherten Informationen sind häufig mittels dokumentarischer Hilfsmittel (wie Klassifikation, Thesaurus oder Abstracts) thematisch erschlossen. Ohne Kenntnis der eingesetzten Hilfsmittel ist kaum ein sinnvolles Retrieval möglich. Da zudem die Suchsysteme der Online-Archive mitunter etwas kompliziert zu bedienen sind, haben sich als letztes Glied der Wertschöpfungskette die Informationsvermittler etabliert, die im Kundenauftrag Informationen recherchieren. Als wichtige verwandte Branche sind die Dokumentlieferdienste anzusehen. Bei Nachweisdatenbanken wissen wir nur, dass eine passende Publikation existiert, die in einem zweiten Schritt beschafft werden muss. Auch „Document Delivery Services“, die häufig von Bibliotheken betrieben werden, sind online erreichbar und liefern einen bestellten Zeitschriftenartikel entweder in seiner elektronischen Version, scannen einen gedruckten Text ein und senden diesen über E-Mail oder verschicken schlicht eine Fotokopie.

Mit den (derzeit kostenlosen) Suchwerkzeugen im World Wide Web sowie den (kostenpflichtigen) Online-Archiven (und den verwandten Branchen) haben wir die Quellen beisammen, in denen - online zugänglich über die globalen Netzwerke - objektives Wissen „schlummert“ und auf das von jedem Ort der Welt und zu jeder Zeit zugegriffen werden kann.

6.1.1.3 Informationsbedarf

Ausgang jeder Informationsbeschaffung ist der Informationsbedarf eines Unternehmens oder eines seiner Teile. Streng abzugrenzen ist der objektive Informationsbedarf vom subjektiven Informationsbedürfnis von Personen. Informationsbedarf und -bedürfnis können weit auseinander klaffen. Mitarbeiter sind manchmal nicht in der Lage einzuschätzen, was sie alles *nicht* wissen, aber in einer bestimmten Situation wissen sollten. Der Informationsbedarf wird mittels sozialwissenschaftlicher Erhebungsmethoden auf unterschiedlichen Ebenen (einzelner Mitarbeiter, Abteilung, Vorstand) eruiert. Zu unterscheiden ist der Ad-hoc-Informationsbedarf vom andauernden Informationsbedarf. Ersterer wird durch *eine* umfassende Recherche in allen zugänglichen Quellen befriedigt, letzterer durch das Anlegen eines Informationsprofils, nach dem in regelmäßigen Abständen (in der Regel via E-Mail) die jeweils aktuellen Informationen geliefert werden.

Die Differenz zwischen dem Informationsbedarf eines Nachfragers und den tatsächlich erhaltenen Informationen ist durch das Vorhandensein von Informationsbarrieren zu erklären (vgl. Engelbert 1976, S. 59 ff.). Auf einer eher psychologischen Ebene liegt die Bewusstseitsbarriere. Für den Nutzer oder auch das gesamte Unternehmen gibt es nichts Unbekanntes, man meint alles zu wissen (und würde sich nicht trauen zuzugeben, dass es sehr wohl Lücken gibt). Die Menge der übermittelten Informationen ist hier natürlich null, da kein Informationsbedürfnis

artikuliert wird. Entgegensteuern kann man mit dem Aufbau von „Informationsbewusstsein“, indem man anhand von Proberecherchen zeigt, was alles machbar ist.

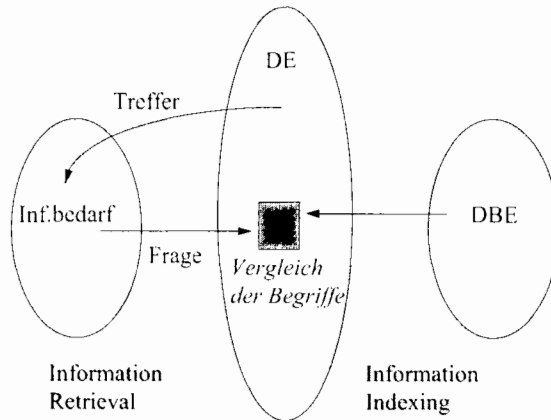
Die Zugangsbarriere verhindert Informationsflüsse aus gewissen Quellen, weil das Unternehmen entweder technisch nicht in der Lage ist, auf globale Netzwerke zuzugreifen, oder weil Suchwerkzeuge im Web oder in Online-Archiven nicht bekannt sind. Eine besonders häufig angetroffene Form ist die Finanzierungsbarriere. Informationen aus Online-Archiven kosten Geld und Internetrecherchen verschlingen Arbeitszeit (und damit auch wieder Geld). Aus diesem Grunde verzichten manche Institutionen auf wertvolle Informationen, anstatt den Nutzen (der sich allerdings nur schwer monetarisieren lässt) zu bedenken. Ein Zwangsneurotiker macht sich „seine“ Welt so klein, dass er kein Bedürfnis an externen Informationen hat. Ein Unternehmen, das im globalen Wettbewerb steht, kann sich ein solches Verhalten nicht leisten. Je internationaler eine Firma ausgerichtet oder je forschungsintensiver die Branche ist, desto größere Etats sind für Informationsbeschaffungen zu kalkulieren. Barrieren liegen zudem bei Fremdsprachen, die ein Nutzer nicht beherrscht, und auch bei Fachsprachen, insofern deren Terminologie nicht richtig gebraucht wird. Letztlich müssen Barrieren durch Mängel im Informationsvermittlungsprozess selbst beachtet werden, die durch Unkenntnis der richtigen Quellen, Hilfsmittel (wie Thesauri), Abfragesprachen usw. entstehen. Hier hilft die Rekrutierung einschlägig ausgebildeten Personals.

6.1.1.4 Indexierungs- und Retrievalverfahren

Wie arbeiten elektronische Informationsdienste, und wie können wir mit diesen Systemen arbeiten? Ausgang sei irgendein (Text-)Dokument, etwa ein wissenschaftlicher Artikel oder eine Bilanz. Beim Input dieses Dokuments in eine Datenbank wird der Inhalt – das enthaltene Wissen – durch Begriffe beschrieben und abfragbar gespeichert. Solcherart Inhaltsabbildung heißt „Information Indexing“. Das ursprüngliche Dokument ist die sog. „dokumentarische Bezugseinheit“ (mit einer Einheitenbildung nach festgelegten gleichbleibenden Regeln), das Resultat der Indexierung ist die „Dokumentationseinheit“. Eine Indexierung geschieht entweder mittels automatisierter Verfahren (automatisches Indexing) oder durch die Leistung menschlicher Indexer (intellektuelles Indexing). Hilfsmittel und Werkzeuge sind sog. „Dokumentationsmethoden“ (wie Thesaurus, Klassifikation usw.) sowie informationslinguistische und -statistische Verfahren (u.a. zur Bildung von Rangordnungen nach Relevanz).

Beim Information Retrieval formuliert der Nachfrager den Informationsbedarf unter Kenntnis der beim Indexing eingesetzten Werkzeuge an das Informationssystem. Die Systemleistung ist ein Vergleich der beim Retrieval verwendeten Begriffe mit den beim Indexing vergebenen Termen. Wir reden vom „exact matching“, wenn die Terme zeichengenau übereinstimmen, und vom „best matching“, wenn die Dokumente relativ zur Suchfrage in eine Reihenfolge nach Relevanz geordnet werden. Online-Archive arbeiten in der Regel nach Exact-Match-, Suchmaschinen im World Wide Web meist nach Best-Match-Verfahren.

In Abbildung 1 sehen wir die eminente Bedeutung der „richtig“ gewählten Begriffe beim Information Indexing und beim Retrieval.



DBE: dokumentarische Bezugseinheit, DE: Dokumentationseinheit

Abb. 1: Information Indexing und Information Retrieval

6.1.1.5 Informationsbeschaffung just in time

Ein Unternehmen wird sicherstellen wollen, dass eine bestimmte Information im Moment des Bedarfs zu suchen und zu finden ist. Es ist angesichts der Massen an externen Informationen nicht sinnvoll, eigene Bestände an solchem Wissen „auf Verdacht“ anzuhäufen und entsprechend Bibliotheken oder Datenbanken zu bestücken. Vielmehr muss im Bedarfsfall – just in time – derjenige Ausschnitt der Gesamtmenge des elektronisch vorhandenen Weltwissens zur Verfügung stehen, der für unsere Firma relevant ist. Hierzu bedarf es einiger Vorbereitungen, von denen die Bereitstellung der technischen Infrastruktur (Intranet mit Internet-Anschluss) die einfachste Aufgabe ist. Über die Erfassung des Informationsbedarfs im Hause ist abzuklären, welche Informationsressourcen (Datenbanken, Informationsanbieter, Internet-Suchwerkzeuge) den Bedarf befriedigen. Im Bereich der kommerziellen Informationswirtschaft ist es üblich, Nutzungsverträge abzuschließen. Da jede Datenbasis und jedes Retrievalsystem speziellen Regeln folgen, sind detaillierte Informationen darüber (etwa Datenbankbeschreibungen, eingesetzte Hilfsmittel, Listen mit Retrievalbefehlen und natürlich Preise) zu dokumentieren. Letztlich ist das Know-how der Nutzung elektronischer Informationsdienste dem Personal zu vermitteln. Es ist ein sinnvoller Ansatz, zweigleisig zu verfahren: Informationsbasiswissen hat jeder Mitarbeiter, unterstützt wird das Unternehmen durch das Spezialwissen ausgewiesener Information Professionals.

Das Basiswissen ist an jeden Mitarbeiter zu vermitteln, der im Unternehmen ein Informationsbedürfnis artikuliert, so dass sich unser Mitarbeiter anhand einiger weniger Wissensspeicher (die sein Fach betreffen und die er kennt) im Rahmen eines gewissen Budgets selbst bedienen kann. Im Sinne eines Pull-Dienstes wird

der Mitarbeiter aktiv, wenn ad-hoc ein Fall von Informationsbedürfnis auftritt; im Rahmen eines Push-Dienstes wird sein andauernder Informationsbedarf durch einen Profildienst periodisch befriedigt, so dass er laufend über Stand und Änderung auf „seinem“ Gebiet up to date ist. Die Organisation dieses Push-Services geschieht entweder via E-Mail oder über ein personalifizierbares „Enterprise Information Portal“ direkt über die Homepage des Mitarbeiters. Information Professionals schulen und coachen die Mitarbeiter in allen Fragen der Informationsbeschaffung, richten für die Mitarbeiter die Informationsprofile ein, organisieren die Personalisierung von Homepages im Unternehmensportal und bearbeiten schwierigere Recherchefälle.

„Just in time-Information“ bedeutet, folgende drei Aspekte zu beherrschen:

- Personalisiert - als Profildienst - dienen externe Informationen innerhalb eines Push-Services dazu, jederzeit „current awareness“ bei Mitarbeitern und Abteilungen zu schaffen. Die persönlichen Informationsprofile werden in Zusammenarbeit mit Information Professionals definiert.
- Zur Befriedigung des Ad-hoc Informationsbedürfnisses eines Mitarbeiters bedient sich dieser selbständig in einer (überschaubaren) Menge externer Datenbanken (Pull-Service).
- Information Professionals (in der Rolle als „Searcher“ oder „Researcher“) bieten für die Mitarbeiter bei komplexen Ad-hoc-Informationsbedarfen sowie bei Recherchen bei brisanten Problemen, wo es auf fehlerfreies Retrieval ankommt (z.B. bei Suchen zum Stand der Technik), weitere interne Dienstleistungen an.

6.1.1.5.1 „Berry picking“

Angeichts der Vielzahl und Vielfalt elektronischer Wissensbestände wäre es verfehlt anzunehmen, in nur einer Quelle alles zu finden, was den Informationsbedarf befriedigte. Außer bei einfachen Faktenfragen ist es vielmehr so, dass diverse Datenbanken bei unterschiedlichen Online-Archiven sowie mehrere Suchwerkzeuge im Internet angesprochen werden müssen. Im Modell des „Berry picking“, vorgestellt von Marcia J. Bates (1989), durchsucht man die Wissenspeicher wie einen Wald beim Beerenpflücken. Der Wald wird erst verlassen, wenn der Korb voll ist, egal, ob die Beeren von einigen wenigen Stellen stammen oder mühselig im gesamten Wald zusammengeklaut werden müssen. Man beginnt seine Recherche mit derjenigen Quelle, die einschlägig erscheint. Eine erste Treffermenge kann dazu führen, dass der Suchende Anregungen zum modifizierten Suchen in derselben Datenbank bekommt, etwa passendere Deskriptoren oder Notationen. Es wird sich zusätzlich zeigen, dass die Datenbank alleine den Bedarf nicht stillen kann. Weitere Datenbanken werden angewählt. Nach der Suche in einer Datenbank mit wissenschaftlichen Artikeln bekommt man z.B. den Hinweis, dass auch Patente zum Thema vorliegen. Folglich wechselt man in eine Patentdatenbank. Oder wir erfahren in einem Artikel einer Nachrichtendatenbank, dass Zeitreihen zum Suchgegenstand geführt werden. Nun müssen wir eine ökonomische Recherche

anschließen. Oder wir lernen ein Portal kennen, in dem Preisinformationen über bestimmte Produkte gesammelt werden.

Wir können die Informationsbereitstellung durchaus mit einem Puzzlespiel vergleichen. Jede angesprochene Quelle steuert einige Teile bei; das Gesamtbild ergibt sich erst aus systematischem Stöbern und Suchen in diversen Wissensspeichern - und dem geschickten Kombinieren der Einzelinformationen.

6.1.1.5.2 Informatrische Recherchen als Knowledge Mining

Bislang hatten wir Suchen im Blick, die in elektronischen Wissensspeichern einzelne Dokumente suchen. Diese Dokumente sind so, wie sie gefunden werden, auch in die Datenbank eingegeben worden. Nunmehr wenden wir uns gewissen *Mengen* von Dokumenten zu, die als Ganzes zu qualifizieren sind, etwa alle Patente eines Unternehmens oder alle Forscher zu einem Thema. Solcherart Informationen erhält man durch sog. „informatrische“ Recherchen, Suchläufe, die Informationsmengen „messen“ und mittels Kennwerten darstellen. Wir haben hier eine Form von Data Mining vor uns, die auf die - in der Regel - nur schwach strukturierten Daten in externen Wissensspeichern angewandt wird. Da meist textbasierte Daten verarbeitet werden, handelt es sich bei dieser Variante des Knowledge Mining um Text Mining. Es haben sich folgende vier Ansätze informatrischer Verdichtung von Einzelinformationen als nützlich erwiesen:

- Rangordnungen (z.B.: Unternehmen nach der Anzahl der Patente in einem Technikgebiet)
- Zeitreihen (z.B.: Anzahl der Patentanmeldungen eines Unternehmens pro Jahr im Laufe der letzten zwei Jahrzehnte)
- semantische Netze (z.B.: Netzwerke wissenschaftlicher Institutionen als Bild der gemeinsamen Publikationen)
- Informationsflussgraphen (z.B.: Stellung eines Unternehmens bei der Nutzung technischer Informationen: Technologieführer oder Nachahmer).

Das Berrypicking-Modell von Marcia J. Bates muss demnach um das Knowledge Mining ergänzt werden. Ergeben sich innerhalb einer Datenbank Hinweise auf in sich kohärente Mengen von Dokumenten, so muss von der Rechercheform „Suche nach Dokumenten“ auf die informatrische Rechercheform umgeschaltet werden. Auch wenn wir mit großen Dokumentmengen konfrontiert werden – etwa bei einem Rechercheauftrag der Art: alle Patente von Siemens –, ist es sinnvoll, informatrische Verdichtungsmethoden einzusetzen anstatt Tausende von Patentdokumenten auszugeben.

6.1.1.6 Relevanz, Recall und Precision

Wenn die betriebliche Informationswirtschaft ein Informationsprofil bedient oder den ad-hoc Informationsbedarfs eines Mitarbeiters deckt, soll das „richtige“ Wissen vermittelt werden. „Richtig“ ist das Wissen genau dann, wenn es relevant ist. Bei Relevanzbeurteilungen geht es vor allem um zwei Fragen:

- Haben wir *alle* Informationen gefunden, die handlungsrelevantes Wissen beinhalten?
- Haben wir *nur* solche Informationen gefunden?

6.1.1.6.1 Frage- und Antworttypen

Die Bestimmung der Relevanz von Informationen ist bei verschiedenen Informationstypen unterschiedlich. Betrachten wir zum Beispiel einige Informationsbedarfe, die sich in folgenden Fragen wiederfinden:

Frage typ A

- Welchen Umsatz hatte Unternehmen X im Monat Januar 1999 in der Region Y?
- Wo liegt der Schmelzpunkt von Kupfer?
- Wie schloss der Dollarkurs gestern an der Frankfurter Börse?
- Wie viele Patente hält unser Wettbewerber X AG in diesem oder jenem Technikfeld?

Frage typ B

- Wie hängen - nach Meinung von Marketingfachleuten - Dienstleistungsmarketing und Qualitätsmanagement zusammen?
- Wie ist das Unternehmen Z AG zu bewerten? (Wir wollen es ggf. übernehmen.)
- Wie beschreiben Marktforscher das Konsumklima für ausländischen Wein in Ungarn? (Wir wollen einen neuen Auslandsmarkt erobern.)

Frage typ A zielt auf die Übermittlung einer Fakteninformation ab. Der zugrunde liegende Informationsbedarf ist konkret. Frage typ B lässt sich nicht durch die Angabe *eines* Faktums befriedigen. Das Informationsproblem wird hier durch die Übermittlung einer mehr oder minder großen Sammlung von Informationen befriedigt – unser gerade eingeführtes „Puzzlespiel“. Die Relevanz einer Information zur Beantwortung einer Frage vom Typ A ist exakt zu bestimmen. Entweder beantwortet die gefundene Information unsere Frage oder sie tut es nicht. Etwas anderes liegt bei der Antwort auf eine Typ-B-Frage vor. Die Antwortmenge sammelt in der Regel Dokumente aus diversen Quellen. Hier haben wir mehrere „Treffer“ und ggf. zusätzlich informatrische Ergebnisse, die jeweils Aspekte dessen beantworten, was gefragt wurde. Bei solchen Informationspaketen als Antwort kann die Beurteilung der Relevanz durchaus vage ausfallen.

6.1.1.6.2 Recall und Precision

Die Informationswissenschaft hat zur Bestimmung der Qualität von Antwortmengen die beiden Aspekte „Recall“ und „Precision“ kreiert. Wir gehen von drei Mengen aus, wobei

- $a =$: gefundene relevante Treffer

- b =: nichtrelevante Dokumente, die in der Treffermenge enthalten sind (Ballast)
- c =: relevante Dokumente in der Datenbank, die nicht gefunden wurden (Verlust).

Recall (Vollständigkeit) errechnet sich als Quotient aus der Anzahl der gefundenen relevanten Datensätze und der Gesamtzahl der relevanten Dokumente in der Datenbank:

$$\text{Recall} = a / a+c.$$

Precision (Genauigkeit) ergibt sich als Quotient aus der Anzahl der gefundenen relevanten Dokumente und der Gesamtzahl der gefundenen Dokumente:

$$\text{Precision} = a / a+b.$$

Die Precision ist – sehen wir von Randunschärfen bei der Bestimmung der Relevanz ab – genau messbar. Recall ist demgegenüber ein reines Konstrukt, da der Wert c nicht messbar ist. Woher weiß ich, was ich *nicht* gefunden habe? Gäbe es Algorithmen, den Verlust zu benennen, würde ich als Rechercheur diese auch einsetzen und keinen Verlust erleiden. Leider gibt es solche Algorithmen außerhalb von experimentellen (überschaubaren) Datenbanken nicht, der Rechercheur wird immer mit Verlust zu kämpfen haben und es wird der Qualitätsmessende nie wissen, wie groß c tatsächlich ist.

Das Konstrukt „Recall“ ist in Verbindung mit der „Precision“ jedoch in einem Gedankenmodell nützlich. Versuchen wir, die Recallquote zu erhöhen, so sinkt gleichzeitig die Quote der Precision, d.h. wir erkaufen uns einen besseren Grad an Vollständigkeit mit einem schlechteren Grad an Genauigkeit, sprich mit Ballast. Versuchen wir umgekehrt, die Genauigkeit zu erhöhen, also Ballast möglichst völlig zu vermeiden, verringern wir die Vollständigkeit. In diesem Wechselspiel von Vollständigkeit und Ballast muss der Recherchierende ein optimales Ergebnis herstellen – eine Technik, die sich algorithmischen Verfahren entzieht und deren Güte entscheidend von den Erfahrungen und der „Kunst“ des Rechercheurs abhängt.

Ein Spezialproblem sind Null-Treffer-Recherchen. Der Recall hat den Wert „0“, wenn $c > 0$ und $a = 0$; der Recallwert ist nicht definiert, wenn $c = 0$ und $a = 0$. Die erstgenannte Möglichkeit signalisiert eine schlechte Recherchequalität, der zweite Fall durchaus nicht, denken wir nur an Neuigkeitsrecherchen im Patentbereich. Null Treffer bei der Frage nach einem technischen Sachverhalt bedeuten, dass – auch im patentrechtlichen Sinne – nichts Vergleichbares vorliegt, wir demnach technisches Neuland betreten und in der Folge auch mit einer Patenterteilung rechnen können. Eine entsprechende Recherche wäre demnach sehr erfolgreich. Eine solche (korrekte) Null-Treffer-Recherche im Patentbereich stellt ein hochrelevantes und präzises Ergebnis dar, das die Ausrichtung der Forschung und Entwicklung eines Unternehmens nachhaltig beeinflussen wird. Null Treffer können aber auch bedeuten, dass falsch gesucht wurde. Fehler sind zwar immer peinlich; in diesem

Fall können sie verheerende Folgen haben. Hier signalisiert die Null in der Tat die schlechteste Qualitätsstufe.

Bei Typ-A-Fragen liegt ein qualitativ einwandfreies Ergebnis dann vor, wenn sowohl Recall als auch Precision „1“ ergeben. Bei Typ-B-Fragen ist wegen des gegenläufigen Zusammenhangs von Recall und Precision jeweils die „1“ kaum möglich; sie zu erreichen, wäre der „Heilige Gral“ der Rechercheure.

6.1.2 Retrievalsysteme

Jede Datenbank besteht aus den beiden Aspekten

- Datenbasis *und*
- Retrievalsystem.

Jeder Recherchevorgang ist zwangsläufig durch beide Aspekte beeinflusst. Die Datenbasis besteht aus den Informationsinhalten (Content), die gespeichertes Wissen in Bewegung setzen – wenn sie denn aufgespürt werden. Das Retrievalsystem ermöglicht den erforderlichen Zugriff auf die Inhalte. In den folgenden zwei Abschnitten zerlegen wir diese Gesamtheit analytisch in die beiden Komponenten, um uns detailliert Klarheit über die jeweiligen Arbeitsweisen sowie unserer Arbeit mit den beiden Komponenten zu verschaffen. Wir beginnen mit den Retrievalsystemen, die wir – wie schon oben angesprochen – in die „Welt“ der Suchwerkzeuge im Internet und die „Welt“ der kommerziellen Online-Archive unterteilen.

6.1.2.1 Suchwerkzeuge im Internet

Das Internet bietet mehrere Services, u.a. E-Mail, Newsgroups, File Transfer und das World Wide Web. Öffentlich zugängliche und allgemein – unter Umständen – brauchbare Informationen liefert vor allem das World Wide Web. Da die Anzahl der Seiten im Web in ständiger Bewegung ist, lässt sich der Umfang an abfragbaren Informationen nur annähernd schätzen. Wenn wir uns am Umfang und Überschneidungsbereich der größten Suchmaschinen orientieren, beträgt die Zahl der einzelnen Webseiten Mitte 2001 mindestens zwei Milliarden – Tendenz anhaltend steigend. „Auf gut Glück“ anhand der URL zu suchen, ist so gut wie aussichtslos. Wichtig ist eine optimale Startseite, von der aus über das Verfolgen von Links („Browsing“) oder das systematische Suchen („Searching“) nach Content gefahndet wird. Derzeit finden wir vier Arten von Suchwerkzeugen im World Wide Web vor (vgl. Stock/Stock 2000/2001): Web-Kataloge (wie Yahoo! oder das Open Directory-Projekt) erschließen die Inhalte von Sites intellektuell, Suchmaschinen (wie Google, Northern Light, AltaVista oder Alltheweb) indexieren Webseiten automatisch und halten große Datenbasen vor, Metasuchmaschinen (wie MetaCrawler oder Oingo) besitzen keine eigene Datenbasis, sondern greifen mittels suchsystemübergreifender Retrievalwerkzeuge auf die Inhalte anderer Suchwerkzeuge zurück. Portale bieten letztlich außer Suchwerkzeugen weitere nützliche Funktionen (wie Personalisierung oder E-Mail) an.

6.1.2.1.1 Wie arbeiten Web-Kataloge?

Web-Directories katalogisieren Web-Sites ähnlich wie große Universalbibliotheken Bücher ordnen. Als Ordnungsbasis dient – ebenfalls wie in Bibliotheken – ein Klassifikationssystem, das alle Aspekte menschlichen Wissens berücksichtigt. In der Welt der Bibliotheken haben sich im Laufe der Jahrzehnte von Forschung und Entwicklung mächtige Systeme herauskristallisiert (etwa die „Dewey Decimal Classification“ *DDC*), die mittels einer überschaubaren Menge von Klassen (bei der *DDC* rund 30.000) mit geschickten Regeln große Wissensspeicher (u.a. die Bestände der British Library oder der Library of Congress) inhaltlich erschließen. Erstaunlicherweise haben alle großen Web-Kataloge auf die bibliothekarischen Vorarbeiten verzichtet und eigene Werkzeuge entwickelt. Der Versuch, das Klassifikationsrad neu zu erfinden, ist dabei nicht gelungen. Einer viel zu großen Menge an Klassen (beim Open Directory-Project Mitte 2001 mehr als 320.000) steht eine viel zu kleine Menge erschlossener Web-Sites gegenüber. Yahoo! kommt 2001 auf gut 1 Mio. Einträge, Open Directory auf rund 2,5 Mio., damit liegen beide Systeme bei unter 1 bzw. ½ Promille des gesamten Web.

Der kleine Teil der verzeichneten Sites ist bei Web-Katalogen intellektuell geprüft, so dass wir von einer Filterfunktion der Systeme ausgehen können. Yahoo! arbeitet mit festangestellten Mitarbeitern, das Open Directory-Project mit ehrenamtlichen Editoren, die jeweils für gewisse Klassen zuständig sind. Die Unterschiede in der Organisationsform erklären die Größendifferenz beider Datenbanken.

Aus Nutzersicht bequem ist das Browsen durch die jeweiligen Klassifikationssysteme. Auf jeder Hierarchieebene liegen die Links (mit kurzem Titel versehen) zu den Treffern sowie die Unterbegriffe der nächstniedrigeren Ebene (im Vorgriff betrachte man Abbildung 3). Bei der Eingabe von Suchtermen arbeiten die Systeme ausschließlich die Klassenbezeichnungen und die Titel der Links ab (und nicht die kompletten Texte der Dokumente wie die Suchmaschinen). Verknüpft man zwei Terme mit dem Boole'schen UND-Operator, so findet Open Directory nur direkte Treffer (beide Terme müssen in der Klassenbezeichnung oder im Titel vorkommen), während Yahoo! cleverer auch indirekte Treffer ausweist (ein Term im jeweiligen Text, der andere Term kann auch in den Klassenbezeichnungen der Unterbegriffe vorkommen).

Suchsysteme der Web-Kataloge sollten nur dann eingesetzt werden, wenn man am Beginn seiner Recherche steht und möglichst breit einsteigen muss, um sich erst im Browsen über die Spezifika seiner Suche klar zu werden. Gezielte Suchen sind weitaus besser bei Suchmaschinen durchzuführen.

6.1.2.1.2 Wie arbeiten Suchmaschinen?

Allen Suchmaschinen im World Wide Web ist gemein, dass sie keinerlei intellektuelle, sondern ausschließlich automatische Inhaltserschließung betreiben. Der Automatismus betrifft auch das Anlegen der Datenbasis (es sei denn, Webmaster melden ihre Seiten beim jeweiligen System an). Grundidee aller Einsammelroboter (wie z.B. Scooter von AltaVista) ist das Verfolgen von Links in bereits be-

kannten Dokumenten. Alle dem System noch nicht bekannten Adressen werden aufgesucht, die Dokumente auf den eigenen Rechner kopiert und sodann inhaltlich erschlossen. Prinzipiell nicht gefunden werden Webseiten, auf die kein einziger Link verweist. Viele Suchmaschinen haben Probleme mit Dokumenten, die nicht im (im WWW üblichen) HTML-Format vorliegen. So weist etwa kaum eine Suchmaschine (Ausnahme: Google) PDF-Dokumente nach.

Jede Suchmaschine verwendet andere Verfahren zur Inhaltserschließung. Zum Einsatz kommen Methoden der Informationslinguistik (so etwa Isolation einzelner Terme, Stoppwortlisten, Wortstammanalysen, Phrasenerkennung) und der Informationsstatistik (beispielsweise Zählungen von Worthäufigkeiten, dokumentenspezifische Wortgewichtung, Gewichtung nach Position im Text, inverse Dokumenthäufigkeit, Wortabstand; eingehende und ausgehende Links; vgl. Stock 2000a, S. 147 ff.). Grob gesagt, sorgt die Informationslinguistik dafür, dass über einen Abgleich zwischen Suchfrage und Dokumenten die Treffer überhaupt gefunden, und die Informationsstatistik dafür, dass die Treffer in eine Reihenfolge nach Relevanz bzgl. der Suchfrage gebracht werden.

Empirische Untersuchungen von Nutzeranfragen haben ergeben, dass eine durchschnittliche Suchanfrage im Web bei rund 1,5 Termen liegt - viel zu wenig „Futter“ für Suchmaschinen auf der Anfrageseite. Auf der Seite der Datenbasis liegen demgegenüber Hunderte von Millionen Dokumente. Wie soll eine Suchmaschine da optimal ein Best Match generieren? Eine spannende Idee verfolgt AltaVista. Ein erster Suchschritt erbringt zentral passende Dokumente zum Thema. Nun werden die Links nach hinten und nach vorne verfolgt und die verlinkten Dokumente in einen „Nachbarschaftsgraphen“ eingestellt. Alle Dokumente dieses Graphen werden automatisch indexiert und die wichtigsten Terme in eine Rangordnung gebracht. Diese Termliste ergibt den Anfragetopic. Da nunmehr genügend Anfrageterme vorliegen, hat die Suchmaschine genügend Material zum Abgleich mit der Gesamtdatenbasis. Über mehrere Schritte von „Beschneidungen“ der Treffermenge erhält der Nutzer sein Ergebnis. Durch die Erweiterung der Suchfrage kann es durchaus vorkommen, dass Webseiten gefunden werden, in denen die ursprünglichen, vom Nutzer eingegebenen Worte in den Treffern gar nicht vorhanden sind. Dies steigert zwar den Recall, kann aber auch zu Lasten der Precision gehen.

Für nahezu alle Suchmaschinen gilt das entsprechende Web-Dokument als dokumentarische Bezugseinheit. Eine beachtenswerte Ausnahme ist Google. Für diese Suchmaschine haben Links als konstituierendes Merkmal von Hypertexten eine besondere Bedeutung. Deshalb ist es verständlich, dass auch den Ankertexten (den Texten über den Links) große Bedeutung beigemessen wird. Umfang der dokumentarischen Bezugseinheit bei Google ist der gesamte Text einer Webseite sowie der Text aller Anker, die auf unsere Seite verweisen. Ankertexte werden demnach doppelt zugeordnet.

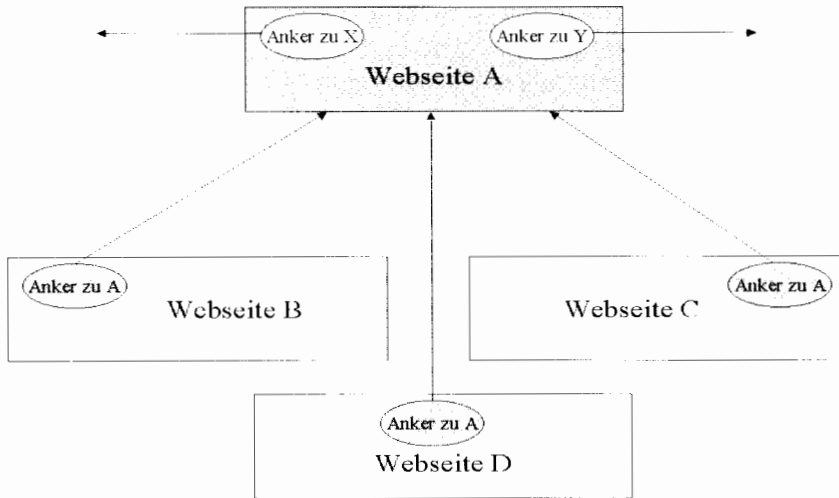


Abb. 2: Dokumentarische Bezugseinheit bei Google

In Abbildung 2 sehen wir die dokumentarische Bezugseinheit zur Seite A. Die Texte der Anker zu X und Y gelten als Teil von A, aber auch von X und Y, die (unserer Abbildung gemäß) noch gar nicht in der Datenbasis vorliegen. Zu A gehören auch die Ankertexte von B, C und D, unter denen Links auf A liegen. Dies hat drei Vorteile: Erstens hat Google mit den Ankertexten bereits einen Teil der Gesamtdokumentationseinheit im Speicher und kann die Seite schon nachweisen (dies erklärt z.T. den großen Umfang der Datenbasis), zweitens hat Google mit der Anzahl der Links auf noch nicht erschlossene Seiten einen Indikator auf die Relevanz dieser Seiten (und kann diese dann in genau dieser Reihenfolge aufsuchen und kopieren) und drittens sind gerade Ankertexte oftmals exakte Beschreibungen von Webseiten bzw. beschreiben Seiten, die selbst überhaupt keinen Text enthalten (z.B. Bilder).

Es gibt eine Reihe hochwertiger Suchmaschinen im Web mit jeweils individuellen Vorzügen: AltaVista (mit seiner Anreicherung der Suchanfrage), Northern Light (mit Clusteranalyse zur Einschränkung großer Treffermengen im Mensch-Maschine-Dialog sowie mit einer beachtenswerten Push-Dienst-Funktion), FAST mit der Suchmaschine Alltheweb (mit besonderen Stärken bei der Suche nach Nicht-Text-Dokumenten wie Bildern, Videos oder Musik) und Google (mit besonders großer Datenbasis). Bei optimaler Ausnutzung der Suchfunktionalität von drei Suchmaschinen (Google, FAST und entweder AltaVista oder Northern Light) lassen sich - einer sog. Availability-Study zufolge (vgl. Stock/Stock 2000/2001, I, S. 30) - 90% aller Dokumente des World Wide Web finden.

6.1.2.1.3 Wie arbeiten Meta-Suchmaschinen?

Meta-Suchmaschinen (z.B. MetaCrawler) verfügen über keine eigene Datenbasis, sondern greifen auf andere Suchwerkzeuge zurück. Es drängt sich der Gedanke auf, dass mittels Meta-Suchmaschinen ein besserer Recall zu erreichen ist. Leider

stimmt dies nicht. Gute Ergebnisse werden in der Regel nur erzielt, wenn die Suchsyntax der jeweiligen Systeme ausgereizt wird. Dies können Meta-Suchmaschinen nicht bewerkstelligen; sie sprechen die Suchmaschinen vielmehr im einfachsten Modus – quasi als kleinsten gemeinsamen Nenner – an. Meta-Suchmaschinen eignen sich demnach wohl für einen kurzen Überblick, jedoch nicht für detaillierte Suchen.

Anders liegt der Fall bei Meta-Suchmaschinen, die einen eigenen informationellen Mehrwert in die Suche einbringen. Das Suchwerkzeug Oingo greift derzeit auf die Datenbasis von Open Directory zurück und bietet zusätzlich ein semantisches Retrieval. Die Ebene des Wortes wird zugunsten der Ebene des Begriffs verlassen. Ist ein Wort mehrdeutig (z.B. „Java“), so fragt das System zurück: Meinst Du „Java“ als Kaffee, als Computerprogramm oder als Insel in Asien? Mit solch einer Homonymkontrolle lässt sich die Precision von Suchergebnissen eminent steigern.

6.1.2.1.4 Wie arbeiten Portale?

Internet-Portale bieten im Vergleich zu den „reinen“ Suchwerkzeugen eine Reihe weiterer Funktionen an. Gemäß Hermann Rösch (2001) sind insbesondere sieben Funktionen für ein Portal wesentlich. Es ist (1.) der einheitliche Einstiegspunkt, um im World Wide Web weitere Aktivitäten zu beginnen. Da Portale auf ein Massenpublikum abzielen, sind sie (2.) möglichst selbsterklärend und intuitiv anwendbar. Sie enthalten (3.) Suchwerkzeuge, meist eine Suchmaschine und ein Web-Directory in Kombination. Im Gegensatz zu den Suchwerkzeugen bieten Portale eigene Informationsinhalte an, z.B. aktuelle Nachrichten oder Börsenkurse; sie sind damit (4.) ein Content Provider. Ein wesentliches Merkmal ist (5.) die Personalisierung, indem der Benutzer sich mit seinem Informationsprofil identifiziert. Dies hat zwei Wirkungen. Zum einen bekommt der Kunde eine rudimentäre Form eines Push-Service angeboten (etwa „MyYahoo“), zum andern erhält der Portalbetreiber personenbezogene Daten über seinen Kunden. Viele Portale ermöglichen ihren Nutzern (6.) Formen elektronischer Kommunikation wie E-Mail-Accounts, Chatrooms bis hin zur Einrichtung einer eigenen Homepage. Einige Portale bieten (7.) zusätzliche Software an, im einfachen Fall z.B. Währungsumrechner oder Kalender, in elaborierten Varianten u.a. Programme zur vergleichenden Produktsuche mit direkten Preisvergleichen sowie Transaktionsprogramme zur Abwicklung von elektronischen Bestellungen.

Internetportale gibt es in den zwei typologischen Varianten des horizontalen und des vertikalen Portals. Horizontale Portale zielen auf größtmögliche thematische Breite ab, sie sind „vergleichbar mit der Angebotspalette eines Großkaufhauses und nicht der eines Fachgeschäfts“ (Rösch 2001, S. 30), aktuelle Beispiele sind Yahoo! oder Lycos. Vertikale Portale setzen auf themenspezifische Tiefe bei Vernachlässigung der Breite. Für eine spezifische Zielgruppe mit ihren entsprechenden Informations- und Produktbedarfen wird ein konkretes Marktsegment informationell präsentiert. Vertikale Portale existieren sowohl im B-to-B-Bereich (z.B. das deutsche Industrieportal „induport.de“ mit diversen Unter-Portalen) als

auch im B-to-C-Segment (etwa das Portal für Schwangere „babycenter.com“ nebst seiner Fortsetzung für Eltern kleiner Kinder als „parentcenter.com“).

Für ein Unternehmen ist die Nutzung von Portalen seiner Branche nahezu selbstverständlich, bekommt man doch so Informationen über Produkte aktueller und potenzieller Lieferanten, über Wettbewerber und deren Produktpalette, über aktuelle und potentielle Abnehmer sowie in den News der Portale über die Branche und ihr Umfeld – sprich: über die gesamte relevante Wertkette.

6.1.2.2 Kommerzielle elektronische Informationsdienste

Die Geschichte der kommerziellen elektronischen Informationsdienste ist mit über 30 Jahren bereits recht lang. Entsprechend ausgereift sind die eingesetzten Retrievalsysteme, so dass die meisten Systeme über eine beachtliche Palette von Befehlen zur Suche nach einzelnen Dokumentationseinheiten sowie zu informatrischen Recherchen verfügen (vgl. Poetzsch 1998; Stock 2000a; vom Kolke 1994). Nach der Art der abrufbaren Informationen können wir unterschiedliche Typen von Datenbanken bei kommerziellen Informationsanbietern ausmachen:

- bibliographische Datenbanken
- Volltextdatenbanken
- Faktendatenbanken
- statistische Datenbanken.

Bibliographische Datenbanken weisen ausschließlich das Vorhandensein von Literaturstellen nach. In deren Dokumentationseinheiten werden die formalen bibliographischen Aspekte so beschrieben, dass die dokumentarische Bezugseinheit identifiziert werden kann. Hinzu kommen inhaltsabbildende Momente, etwa Deskriptoren eines Thesaurus, Notationen eines Klassifikationssystems sowie in den meisten Fällen ein Abstract. Insbesondere die professionelle inhaltliche Beschreibung der wesentlichen Informationsinhalte der dokumentarischen Bezugseinheiten lässt eine zielgenaue thematische Suche zu. Nachteil dieser Datenbanken ist der Medienbruch im Anschluss an die Recherche, sofern der Volltext in einem zweiten Arbeitsschritt besorgt werden muss. *Volltextdatenbanken* bringen in der Dokumentationseinheit den gesamten Text der dokumentarischen Bezugseinheit. Tageszeitungen, Gesetzestexte, Grundsatzurteile, Patente, Artikel wissenschaftlicher Zeitschriften usw. sind so Wort für Wort abfragbar und als Ganzes abrufbereit. Volltexte können indexiert, d.h. wie die bibliographischen Nachweise inhaltlich erschlossen sein. Dies erleichtert ihr Auffinden enorm. In vielen Fällen findet jedoch keine intellektuelle Inhaltserschließung statt, so dass wir hier entweder auf besondere, am Volltext orientierte Retrievalbefehle oder auf die Möglichkeiten der automatischen Indexierung angewiesen sind. *Faktendatenbanken* lassen sich am besten mit Nachschlagewerken bzw. Handbüchern vergleichen. Hier wird nicht die Literatur zu einem Sachverhalt nachgewiesen, sondern das Faktum selbst. Die Bandbreite der Faktendatenbanken ist sehr groß und reicht von Unternehmensbeschreibungen über Bilanzen, biographische Informationen, Daten über Werkstoffe bis hin zu den Strukturformeln der organischen Chemie. *Statistische Datenbanken* sind Faktendatenbanken, die meist Zeitreihen enthalten.

Hauptunterschied zu den anderen Faktendatenbanken ist die Möglichkeit zu ökonomischen Berechnungen.

Einen guten Überblick über (nahezu) alle elektronischen Datenbanken verschaffen die Führer von *Gale*. In kurzen Abständen aktualisiert, beschreiben sie Online-Datenbanken sowie elektronische Offline-Datenbanken, u.a. CD-ROM (vgl. Gale 1999; Gale 2001).

Zur Abfrage einer kommerziellen Datenbank bedarf es der Kenntnisse, wie die Daten in einem Feldschema strukturiert sind und welche Hilfsmittel bei welchen Feldern zum Einsatz kommen. Wir müssen etwa wissen, wie in einem Feld, das Autorennamen aufnimmt, diese Namen angesetzt sind: mit allen Vornamen (AU=Wittgenstein, Ludwig Josef Johann), mit einem Vornamen (AU=Wittgenstein, Ludwig), mit den Initialen der Vornamen (AU=Wittgenstein, L.J.J.), ohne Komma zwischen Nach- und Vornamen (AU=Wittgenstein L) usw. Hierüber berichten „Bluesheets“. Oder wir müssen wissen, welche Hilfsmittel der Inhaltserschließung die Indexer benutzen: Normiert z.B. ein Thesaurus Themen wie „Unternehmensführung“ bzw. „Management“ mit dem Deskriptor „Management“, so wird eine Recherche nach „DE=Unternehmensführung“ keinen einzigen Treffer landen, obgleich - korrekt unter „DE=Management“ gesucht - diverse Datensätze vorliegen.

Es gibt eine Reihe von Standardbefehlen, die Datenbanken aufrufen, Suchen formulieren und Ausgabeformate definieren. Exemplarisch seien einige dieser Befehle aufgelistet:

- Aufruf einer Datenbank
- Aufruf mehrerer (gleichartiger) Datenbanken gemeinsam
- Suche in Feldern
- Blättern in Indices
- Suchtermfragmentierung (Truncation)
- mengentheoretische Operatoren (UND, ODER, NICHT)
- Abstandsoperatoren (Phrase, Nachbarschaft, Abstand von n Termen, im selben Satz, im selben Absatz usw.)
- Häufigkeitsoperator
- hierarchische Suche (Einbeziehung von Unterbegriffen)
- beim Aufruf mehrerer Datenbanken: Erkennen und Entfernen von Dubletten
- Bilden von Rangordnungen (bei informatrischen Suchen)
- Sortieren von Suchergebnissen
- Anzeigen von Suchergebnissen (in unterschiedlichen Ausgabeformaten mit jeweils unterschiedlichen Preisen).

Die Retrievalsysteme der einzelnen Anbieter bieten zwar einen vergleichbaren Umfang an Befehlen, verwenden jedoch meist unterschiedliche Syntax. Nehmen

wir an, wir suchten in einer Volltextdatenbank nach Artikeln über Julia Roberts im Film „Notting Hill“. Beim Anbieter Lexis-Nexis formulieren wir

HEADLINE:(“Julia Roberts” w/5 “Notting Hill”),

bei DIALOG wäre eine sinnvolle Variante

(Julia ADJ Roberts AND Notting ADJ Hill)/ti.

Da die Retrievalsysteme in der Regel nach dem Exact Match-Verfahren arbeiten, müssen sowohl die Suchsyntax minutiös beherrscht als auch die Suchargumente fehlerfrei eingegeben werden.

6.1.3 Informationsinhalte („E-Content“)

Wissen bedarf der Ordnung. Wollen wir gezielt auf Informationsinhalte zugreifen, machen wir von Ordnungssystemen Gebrauch. In Abbildung 1 sahen wir die Bedeutung der Begriffe beim „Match“, beim Abgleich der Suchargumente und dem vorhandenen E-Content. Professionelle Datenbanken setzen jeweils spezifische Ordnungssysteme als Hilfsmittel der Inhaltsabbildung ein. Wissen bedarf darüber hinaus der übersichtlichen Darstellung. Wenn wir mehrere Treffer aufgelistet bekommen, wollen wir als Nutzer möglichst auf dem ersten Blick sehen, welche davon für uns relevant sind. Datenbanken setzen hier auf Informationsverdichtung durch Methoden des Abstracting.

6.1.3.1 Werkzeuge der Inhaltsabbildung

Inhaltsabbildung geschieht mittels sog. „Dokumentationsmethoden“. Bei den eingesetzten Systemen zur Begriffsordnung haben sich die beiden Varianten Thesaurus und Klassifikation durchgesetzt. Jede Begriffsordnung verfügt erstens über eine terminologische Kontrolle des Vokabulars und zweitens über die Angabe der Relationen zwischen den Begriffen. Terminologische Kontrolle bedeutet, dass Begriffe und deren Bezeichnungen eindeutig aufeinander bezogen werden. Synonyme (Samstag und Sonnabend) werden unter genau einer Bezeichnung zu einer Klasse zusammengefasst, Homonyme (Kiefer als Knochen und Kiefer als Nadelbaum) durch erläuternde Zusätze getrennt. Die semantischen Beziehungen zwischen den Begriffen werden durch Hierarchie- und Assoziationsrelationen ausgedrückt. In Abbildung 3 haben wir ein kleines Begriffssystem vor uns, das neun Begriffe A bis I umfasst.

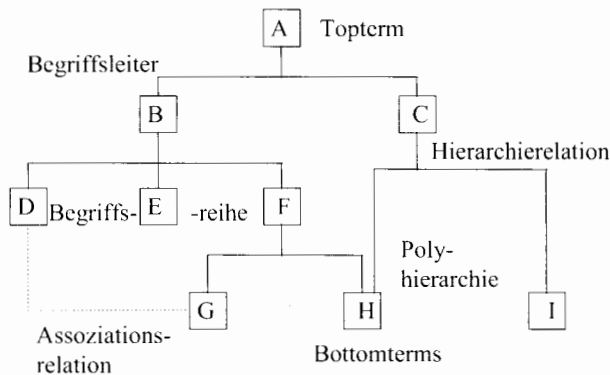


Abb. 3: Relationen in Begriffssystemen

Der Begriff A ist der oberste Begriff in der Hierarchie, ein Topterm, also ein Begriff von hoher Allgemeinheit. Im Bild von oben nach unten stehen auf den jeweils benachbarten Ebenen Unterbegriffe. B und C sind Unterbegriffe von A; D, E und F sind Unterbegriffe von B usw. In umgekehrter Reihenfolge spricht man von Oberbegriffen. Ober- und Unterbegriffe eines Begriffs fächern die Hierarchierelation auf. Diese hat zwei Ausprägungen, die jedoch in den meisten Begriffssystemen zusammenfallen. Einmal geht es um eine logische Sichtweise: Der Unterbegriff der Abstraktionsrelation schließt alle Merkmale des Oberbegriffs ein, hat aber mindestens ein weiteres Merkmal, das ihn von den nebengeordneten Begriffen, die auch Unterbegriffe desselben Oberbegriffs sind, eindeutig unterscheidet (Beispiel: Pkw und Lkw als Unterbegriffe von Kraftwagen). Zum anderen geht es um Ganzheiten und dessen Teile. Der Unterbegriff der Bestandsrelation repräsentiert jeweils einen Teil des übergeordneten Gegenstandes (Beispiel: Fahrgestell und Motor als Teile von Kraftwagen). Alle Paare, die in Hierarchierelation zueinander stehen, bilden in ihrer Gesamtheit vom Topterm bis zu den Bottomterms eine Begriffsleiter. In der Abbildung sind mehrere Begriffsleitern vorhanden, u.a. A - B - D oder A - B - F - G. Eine Begriffsreihe setzt sich aus Begriffen der selben Hierarchieebene zusammen, also z.B. D - E - F. Ein Begriffssystem, in dem alle Begriffe entweder keinen oder genau einen Oberbegriff haben, nennt man monohierarchisch. Systeme, bei denen ein Begriff mehrere Oberbegriffe haben kann, sind polyhierarchisch. Unser Beispiel ist demnach ein polyhierarchisches System (H hat die zwei Oberbegriffe C und F). Ein Beispiel möge die Nützlichkeit von Polyhierarchien verdeutlichen: Der Begriff Holzmöbelindustrie hat als Oberbegriff sowohl Holzindustrie als auch Möbelindustrie. Neben der Hierarchierelation kennt man bei Begriffssystemen weiterhin die Assoziationsrelation (bei uns zwischen D und G). Hier gehen alle Beziehungen zwischen zwei Begriffen ein, die unabhängig von Hierarchierelationen bestehen, u.a.

Kausalrelationen (Ursache - Wirkung), instrumentelle Relationen (Werkzeug - Anwendung des Werkzeugs) oder Ähnlichkeitsrelationen.

6.1.3.1.1 Klassifikation

Eine Klassifikation ist ein Begriffssystem, in dem die Begriffe durch kunstsprachliche Notationen (Buchstaben- oder Ziffernkombinationen) bezeichnet werden. So steht z.B. die Notation 72 in der NACE für Datenverarbeitung und Datenbanken und 72.2 (als einer der Unterbegriffe) für Softwarehäuser. Klassifikationen arbeiten meist nur mit der Hierarchierelation in monohierarchischer Struktur. In der Wirtschaftsdokumentation finden Klassifikationen vielfach Einsatz, wie folgende Beispiele zeigen:

- NACE (Nomenclature général des activités économiques dans les Communautés Européens): Systematik der Wirtschaftszweige zum Einsatz u.a. bei Wirtschaftsstatistiken in Europa
- NAICS (North American Industry Classification System): das amerikanische Gegenstück zur NACE
- IPK (Internationale Patentklassifikation): Technikklassifikation, nach der alle Patente weltweit inhaltlich erschlossen werden
- Wiener Klassifikation der figürlichen Darstellungen: Klassifikation der Graphikelemente von Bildmarken.

Ist eine Klassifikation in ihrer Notationsgestaltung strukturabbildend hierarchisch aufgebaut, so ergibt sich im Retrieval eine einfach durchzuführende Form hierarchischen Suchens, lassen sich doch durch Rechtstruncation Unterbegriffe mit in die Anfrage einbeziehen.

6.1.3.1.2 Thesaurus

Thesauri sind fachbezogene Begriffssysteme, die ihre Vorzugsbenennungen (Deskriptoren) aus der jeweiligen Fachsprache entnehmen und die sowohl Hierarchie- als auch Assoziationsrelationen verwenden, wobei die Deskriptoren polyhierarchisch geordnet sind (vgl. Aitchison et al. 1997). Im deutschsprachigen Raum hat sich für die wirtschaftswissenschaftliche Dokumentation der Standard-Thesaurus Wirtschaft (vgl. STW 1998; Stock, M. 1999) durchgesetzt. Mittels seiner Deskriptoren werden u.a. die elektronischen Datenbanken der großen Wirtschaftsforschungsinstitute HWWA, ifo und Institut für Weltwirtschaft inhaltlich erschlossen.

6.1.3.1.3 Informationsverdichtung durch Abstracts

Nach einer Suche in elektronischen Datenbanken erhält der Nutzer eine Titelliste der Treffer. Leider sind Titel oftmals wenig aussagekräftig. Die Entscheidung, ob ein Nutzer das volle Dokument braucht oder nicht, ist also allein mit dem Titel schlecht fundiert. Kommerzielle Informationsanbieter liefern deshalb zusätzlich ein Kurzreferat, ein Abstract, das den Inhalt der dokumentarischen Bezugseinheit in wenigen Sätzen wiedergibt. Das Verfassen von aussagekräftigen Abstracts ist eine Leistung der Informationsverdichtung, die für viele als Kunst gilt (vgl.

Cremmings 1982) und die sich weitgehend den Methoden automatischer, rein maschineller Bearbeitung entzieht. In vielen Fällen, insbesondere bei wissenschaftlicher Literatur, liegen Abstracts der Autoren vor, die von den Datenbankproduzenten übernommen werden. Hat die Vorlage kein Abstract oder ist ein Autorenabstract unzureichend, so lassen einige Informationsproduzenten eigene Kurzreferate erstellen. Suchmaschinen im Web setzen Abstracts nicht ein und verzichten damit auf jegliche professionelle Informationsverdichtung. Der Nutzer wird entsprechend mit einem großen Anteil an Informationsballast konfrontiert.

Gute Abstracts sind übersichtlich strukturiert, kurz und präzise, für den Fachmann leicht zu verstehen und referieren zentrale Momente des bearbeiteten Materials. Es sind zwei Formen von Abstracts zu unterscheiden. Indikative Abstracts berichten in einer Art Inhaltsverzeichnis über das Dokument (etwa: „...es geht um die Höhenlage von München...“), informative Abstracts referieren die wichtigsten Ergebnisse („...München liegt rund 530 m über NN...“).

Begriffssysteme (Thesauri, Klassifikationssysteme) helfen dem Nutzer elektronischer Informationsdienste, überhaupt die „richtigen“ Suchargumente zu finden und diese im Retrieval einzusetzen, Abstracts helfen dem Nutzer, das über die Deskriptoren bzw. Notationen gesuchte und gefundene Material zu sichten und nur bei den relevanten Treffern das gesamte Dokument anzufordern und einzusehen.

6.1.3.2 Typologie elektronischer Informationen

Wissen liegt in unterschiedlichen Kontexten vor. Der Content wirtschaftswissenschaftlicher Informationen ist anders aufbereitet als etwa Inhalte von Adressdatenbanken oder von ökonometrischen Zeitreihen. Wir skizzieren einige Typen von E-Content, die in der Praxis von Unternehmen oder Behörden eine Rolle spielen. Verdeutlicht werden die einzelnen Typen durch paradigmatisch ausgewählte Musterdatenbanken. Zusätzlich zu den Informationsressourcen im World Wide Web existieren mehrere tausend Datenbanken der kommerziellen Anbieter, deren Inhalte für Unternehmen potentiell relevant sind (für Details vgl. Stock 2000a).

6.1.3.2.1 Wirtschaftswissenschaftliche Informationen

Informationen zu wirtschaftswissenschaftlicher Literatur liegen recht vollständig in elektronischen Informationsdiensten vor. Die Informationsressourcen gehören entweder zu den bibliographischen Datenbanken oder zu einer Kombinationsform aus bibliographischer und Volltextdatenbank. Die in Deutschland gesammelte Wirtschaftsliteratur wird vor allem durch die vier bibliographischen Datenbanken ECONIS, HWWA, IFO und BLISS abgedeckt. Im internationalen Bereich dominieren bei den kommerziellen Anbietern ABI-Inform und EconLit, im World Wide Web die Ressourcen der NetEc-Gruppe.

Datenbank: IFOLIT

Nefiodow, Leo A.: *Informationsgesellschaft - Arbeitsplatzvernichtung oder Arbeitsplatzgewinne?* ifo Schnelldienst 47, Nr. 12 (1994), 25. April 1994, S. 11-19

Signatur: Z35(47)1994,12 (3)

Abstract: Der Bericht befasst sich mit der Frage, ob in einer Informationsgesellschaft Arbeitsplätze vernichtet oder gewonnen werden. Ausgehend von Basisinnovationen, die der Schlüssel zur Beseitigung der Arbeitslosigkeit sind, werden Strukturwandlungen in Wirtschaft und Gesellschaft hervorgerufen. Die langen Wellen, die die Konjunktur während entscheidender Basisinnovationen abbilden, münden derzeit in die fünfte Welle, die von der Informationstechnik getragen wird. Der Bericht stellt unter anderem die neueren Anwendungen der Informationstechnik, die Entwicklung der Telekommunikation, die Struktur der Informationswirtschaft und die Entwicklung der Beschäftigtenzahlen in den Industrienationen dar. Insgesamt wird festgestellt, dass die Quelle zukünftigen Wachstums nicht mehr vom Faktor Energie gebildet wird, sondern vielmehr vom Faktor Information. Die Industriegesellschaft wandelt sich zusehends zu einer Informationsgesellschaft. Zum Abschluss werden die Märkte näher betrachtet, die neue Arbeitsplätze in großer Zahl schaffen können. Dies sind der Medienmarkt, der Markt für Telekommunikationsdienste, die Anwendung der Informationstechnik in Organisationen und der Bildungsmarkt.

Deskriptoren: Informationsgesellschaft; Industriegesellschaft; Innovation; Beschäftigungspolitik; Rationalisierung; Subvention; Strukturwandel; Informationstechnik; Information; Telekommunikation; Forschung; Lange Wellen

Klassifikation: Arbeit/Beschäftigung/Lohn; Tertiärer Sektor/Dienstleistung; SG08; SG07

© ifo

Quelle: WISO II / ifo Institut

Abb. 4: Beispiel einer Dokumentationseinheit im Bereich wirtschaftswissenschaftlicher Informationen

Ein typisches Beispiel der IFO-Datenbank zeigt Abbildung 4. Gesucht wurde Literatur zum Strukturwandel in der Informationsgesellschaft. Der Treffer bringt die bibliographischen Angaben, die Signatur des Volltextes, Deskriptoren, Notationen sowie ein Abstract. Entscheidet der Nutzer anhand des Kurzreferats, dass er den Artikel lesen möchte, so muss dieser nunmehr in einem zweiten Arbeitsschritt beschafft werden. Zwei Wege stehen offen: Entweder benutzt man den Lieferdienst des Informationsproduzenten (in diesem Falle also – unter Angabe der Signatur – das ifo Institut) oder man wendet sich an einen umfassenden Document Delivery Service wie z.B. Subito. Der zweite Arbeitsgang wird umgangen, wenn der Datenbankproduzent direkt die Volltexte mit anbietet. So verfährt ABI-Inform.

6.1.3.2.2 Informationen aus der Marktforschung

Die Sekundärmarktforschung ist darauf angewiesen, alle bereits zum Thema vorliegenden Quellen zu sammeln. Im Bereich der Marktforschung dominieren Volltext- und Faktendatenbanken. Fakteninformationen wie eine spezifische Tabelle, eine Rangordnung (etwa eine Top-10-Liste von Unternehmen auf einem Markt)

oder eine Zahlenangabe liegen z.B. in der von Infratest Burke erstellten Datenbank FAKT vor. Volltexte nahezu aller wichtigen Marktforschungsinstitute sind beim Online-Archiv Profound recherchierbar; Ausgabeoptionen erlauben die Speicherung des gesamten Textes genauso wie den zielgenauen Zugriff auf ein Kapitel oder eine bestimmte Tabelle.

6.1.3.2.3 Wirtschaftsnachrichten

Ein Informationsbereich, wo die Informationsflut besonders hoch auftritt, sind die Nachrichten. Allein die Wirtschaftsnachrichten in Agenturmeldungen, in den Artikeln der überregionalen Zeitungen und Branchennewslettern betragen Schätzungen zufolge weltweit über 20.000 pro Tag. Die Datenbanken in diesem Wissensbereich gehören fast ausnahmslos zu den Volltextdatenbanken. Jede Agenturmeldung, jeder Zeitungsartikel wird im vollen Wortlaut gespeichert und durch inhaltserschließende Angaben komplettiert. Die Datensätze werden bei den Online-Archiven nicht, wie im Internet üblich, nach einer gewissen Zeit gelöscht, sondern verbleiben über Jahre oder Jahrzehnte in den Wissensspeichern. Datenbankbeispiele sind Factiva (ein gemeinsames Unternehmen von Reuters und Dow Jones) bei Agenturmeldungen, Handelsblatt (via GENIOS) oder F.A.Z. bei Zeitungen sowie die Ressourcen von Gale's Information Access Company bei den Newslettern.

6.1.3.2.4 Unternehmensinformationen

Informationsressourcen zu Firmeninformationen sind in der Regel als Faktendatenbanken aufgebaut. Grunddaten über Unternehmen bieten Datenbanken mit Unternehmenskurzdossiers. Für deutsche Unternehmen relevant sind Hoppenstedt (mit einer interessanten Schnittstelle zum Database-Marketing) und Creditreform. Die Eintragungen im deutschen Bundesanzeiger (u.a. das Handelsregister) sind im Wortlaut recherchierbar, Bilanzen der veröffentlichungspflichtigen Unternehmen einschließlich der betrieblichen Kennziffern sind online derart abgelegt, dass ein feldspezifischer Zugriff (z.B. nach Unternehmen mit einem Umsatz pro Mitarbeiter von über x Mio. DM) möglich ist. Bonitätsdatenbanken (u.a. von Creditreform oder von Dun & Bradstreet) versuchen eine Risikobewertung anhand von Zahlungsverhalten und anderen Kriterien. Für Deutschland liegen rund drei Millionen Dossiers vor; die international ausgerichtete Datenbank von Dun & Bradstreet enthält Angaben zu über 50 Millionen Unternehmen. Bei mittleren und großen sowie bei allen börsennotierten Firmen können wir davon ausgehen, in den Datenbanken Grunddaten der Unternehmen zu erhalten. Einen umfassenden Überblick zu Firmeninformationen, auch in Hinblick auf die Qualität der Informationsanbieter, liefern Bredemeier et al. (2001).

6.1.3.2.5 Risikoinformationen

Eine eher praxisorientierte Wissensdomäne sind aktuelle Informationen über Risiken in Städten oder Ländern der Welt. Gesetzt den Fall, ein Unternehmen möchte eine Niederlassung in einem bislang unbekanntem Land aufmachen, so ist die politische, ökonomische und kulturelle Situation abzuschätzen. Ist die politische Lage stabil? Wie hoch ist die Streikbereitschaft? Gibt es Beeinträchtigungen

der Arbeitsbereitschaft unter besonderen Bedingungen (Ramadan oder Fußballweltmeisterschaft)? Auch für den Fall einer Reise in eine unbekannte Stadt sind Informationen nötig. In welchen Gegenden ist man vor Überfällen sicher? Gibt es in der Reisezeit bei uns unbekannte Feiertage? Wie ist die lokale Krankenversorgung? Solcherart Risikoinformationen, stets als Volltexte erarbeitet, bieten spezielle Informationsproduzenten wie z.B. das Control Risks Information System.

6.1.3.2.6 Ausschreibungen - Geschäftsanbahnungen - Förderprogramme

Ebenso praktisch sind Informationsressourcen über Ausschreibungen, Kooperationsbörsen und Förderprogramme. In allen diesen Faktendatenbanken geht es darum, Informationen darüber zu streuen, wo mit welchen Leistungen Geld zu verdienen ist. In der Datenbank Tenders Electronic Daily werden tagesaktuell alle öffentliche Ausschreibungen ab einem gewissen Mindestumfang präsentiert. Die Förderdatenbank des deutschen Bundesministeriums für Wirtschaft unterrichtet über alle Fördermaßnahmen in Deutschland.

6.1.3.2.7 Gewerbliche Schutzrechte als Wirtschaftsinformationen

Technische Schutzrechte, Patente und Gebrauchsmuster, sind sowohl technische und juristische als auch ökonomische Informationen. Bei einer Wettbewerberanalyse ist es wichtig zu wissen, wie viele Patente der Konkurrent hält, in welchen Technikgebieten sie angesiedelt sind und wann sie ablaufen. Die elektronischen Informationsdienste zu Patenten sind in der Form bibliographischer Datenbanken nahezu komplett, d.h. es gibt zu jedem Patent auf der Welt einen (nach international einheitlichen Regeln erschlossenen) Nachweis. Große Patentdatenbanken sind World Patents Index von Derwent und Inpadoc vom Europäischen Patentamt, deutsche Patente finden sich in PatDPA und PatosDE. Lieferdienste der Volltexte sind bei Patentämtern oder speziellen Dienstleistern (wie Paton) angesiedelt.

Bei den nichttechnischen Schutzrechten kann davon ausgegangen werden, dass Informationen über Marken (Wort- wie Bildmarken) für alle wichtigen Länder komplett elektronisch vorgehalten werden. Bei den Wortmarken bieten einige Informationsdienste unscharfe (fuzzy) Suchen an, so dass auch phonetisch verwandte Terme (z.B. Mikrosaft bei einem Retrievalauf nach Microsoft) gefunden werden. Eine besonders elaborierte Software benutzt das Online-Archiv Questel-Orbit.

6.1.3.2.8 Beobachtung der Forschung und Entwicklung von Wettbewerbern und Technikgebieten

An einem Beispiel sei die Arbeitsweise einer informetrischen Recherche demonstriert. Wir gehen davon aus, dass Entwicklung und Stand der Forschung und die Entwicklung der Unternehmen auf einem Markt eruiert werden müssen. Unser Beispiel betrifft einen Teilmarkt des Maschinenbaus, die Verpackungsmaschinen. In einem ersten Rechenschritt wird bei einer internationalen Patentdatenbank nach der Notation des Technikfeldes gesucht (hier: IPK B65B) und die Treffermenge (rund 23.000 Patentfamilien) in eine Rangordnung nach anmeldenden Unternehmen gebracht. Ein zweiter Rechenschritt sucht jeweils nach unserer

IPK-Notation in Verbindung mit dem Unternehmensnamen für die top-platzierten Patentanmelder und gibt die Ergebnisse als informetrische Zeitreihe aus. Abbildung 5 zeigt einen Ausschnitt der Zeitreihe der Top 3 der Verpackungsmaschinenproduzenten, entstanden als informetrische Analyse der Datenbank World Patents Index beim Informationsdienst Questel-Orbit. Man sieht hier sehr anschaulich, wie im Laufe von gut zehn Jahren ein Unternehmen (Focke) die Technologieführerschaft übernommen hat.

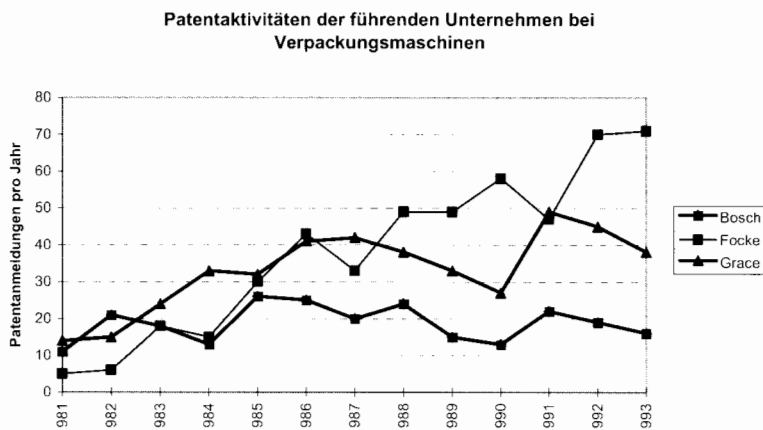


Abb. 5: Beispiel eines informetrischen Rechercheergebnisses

6.1.3.2.9 Wichtige Independents und Online-Archive

Die angesprochenen kommerziellen Datenbanken liegen entweder als Angebot von Independents im kostenpflichtigen Bereich des World Wide Web (so u.a. Hoppenstedt und Control Risks) oder sind Teil von umfassenderen Online-Archiven. Elektronische Informationsdienste mit Wirtschaftsinformationen sind u.a.:

- DataStar
- DIALOG
- DRI / Standard & Poors (Spezialgebiet: Zeitreihen)
- Factiva (Nachrichten)
- FIZ Technik (Technik)
- GBI / Gesellschaft für betriebswirtschaftliche Information
- GENIOS
- Juris (Recht)
- Lexis-Nexis
- Profound (Marktforschung)
- Questel-Orbit (Patente und Marken)

- STN (Wissenschaft).

Mit dem Informationsangebot dieser Online-Archive und dem einiger weiterer Independents dürften nahezu alle Informationsbedarfe von Unternehmen und Behörden im deutschsprachigen Raum befriedigt werden können. Unterstützend werden die Suchwerkzeuge und Portale des World Wide Web in die betriebliche Informationswirtschaft einbezogen.

6.1.4 Vertiefende Informationen zu elektronischen Informationsdiensten

Das Feld der elektronischen Informationsdienste in elektronischen Netzwerken ist ausgesprochen groß, nur schwach strukturiert, ständigen Veränderungen unterworfen und zudem im virtuellen Raum verstreut. *Einfach* ist der Umgang damit nicht. Die Aufgabe dieses Kapitels, das vielfach mehr Andeutungen denn komplette Lösungen anbietet, wäre erfüllt, wenn der Leser erkannt hat: So etwas brauche ich in meinem Unternehmen! Für diesen Leser geben wir abschließend Listen mit einschlägiger Literatur (zu fachspezifischen Zeitschriften und Newslettern), mit Messen und mit den Anschriften wichtiger Anbieter von Suchwerkzeugen im World Wide Web und kommerzieller Online-Archive an, so dass Einstiege zu vertiefenden Informationen möglich werden. Wir haben ja von Matthäus gelernt: Nur wer bereits genügend Wissen hat, dem wird weiteres Wissen gegeben.

Anhang:

Zeitschriften zu elektronischen Informationsdiensten

E-Content (ISSN 1525-2531)

Information Today (ISSN 8755-6286)

Information World Review (ISSN 0950-9879)

Link-Up (ISSN 0158-5460)

Managing Information (ISSN 1352-0229)

nfd - Information. Wissenschaft und Praxis (ISSN 1434-4653) - in deutscher Sprache

Online (ISSN 0146-5422)

Online and CD Notes (ISSN 0144-025X)

Online Information Review (ISSN 1468-4527)

Password (ISSN 0930-3693) - in deutscher Sprache

Searcher (ISSN 1070-4795)

Messen elektronischer Informationsdienste

Infobase - Messe Frankfurt; jährlich im Mai

Online Information - Olympia London; jährlich im Dezember

Anbieter elektronischer Online-Informationen

a) Anbieter von Suchwerkzeugen im Internet

AltaVista (www.altavista.com; ragingsearch.altavista.com)

FAST / All the Web (www.alltheweb.com)

Google (www.google.com)

MetaCrawler (www.metacrawler.com)

Northern Light (www.northernlight.com)

Oingo (www.oingo.com)

Open Directory Project (www.dmoz.org)

Yahoo! (www.yahoo.com)

b) Kommerzielle Online-Archive und Independents

Control Risks Information System (www.control-risks.de)

Dialog Corp. / DIALOG, DataStar, Profound (www.dialog.com)

DRI (www.dri.standardandpoors.com)

Factiva - Reuters, Dow Jones (www.factiva.com)

FIZ Technik (www.fiz-technik.de)

GENIOS (www.genios.de)

GBI (www.gbi.de)

Hoppenstedt (www.firmendatenbank.de)

Juris (www.juris.de)

Lexis-Nexis (www.lexis-nexis.com)

Paton (www.paton.tu-ilmenau.de)

Questel-Orbit (www.questel.fr)

STN (www.fiz-karlsruhe.de)

Subito (www.subito-doc.com)