

Robert Fugmann: Inhaltserschließung durch Indexieren

Intellektuelles Indexieren für Buchregister und Inhouse-Datenbanken

Robert Fugmann: Inhaltserschließung durch Indexieren: Prinzipien und Praxis. Frankfurt: Deutsche Gesellschaft für Dokumentation, 1999. - (Reihe Informationswissenschaft der DGD; 3). - VIII, 225 S. (+ 18 S. Register).

Was ist neu an diesem Buch? Leider nur recht wenig.

Fugmann legt dar, "daß schriftlich oder bildlich niedergelegtes Wissen einer besonderen Erschließung bedarf, um im späteren Bedarfsfall wiederauffindbar zu sein. Diese Erschließung muß einigen fundamentalen Prinzipien folgen, soll sie auf die Dauer ausreichend wirksam sein" (S. 1). Die Fundamentalprinzipien betreffen die Unterscheidung zwischen Begriff und sprachlicher Ausdrucksweise. Im Buch "werden die Gemeinsamkeiten von Datenbank- und Buchindexieren in den Vordergrund gestellt und aus der Perspektive der durch Indexieren zu erzielenden Voraussehbarkeit der gespeicherten Ausdrucksweisen für Begriffe betrachtet" (S. 2). Angestrebt wird ein Optimum an Retrievalqualität durch weitgehendes Vermeiden sowohl von Ballast als auch von Informationsverlust. Der Begriff der "Indexierung" wird definiert als "Zuweisen von inhaltskennzeichnenden Begriffsbenennungen zu Dokumenten zum Zweck des späteren Wiederauffindens" (S. 11).

Begriff und sprachlicher Ausdruck

Nach der Einführung wendet sich Fugmann im 2. und 3. Kapitel der prinzipiellen Unterscheidung zwischen Begriff und sprachlicher Ausdrucksweise zu. Der "Begriff" wird über die Menge der Merkmale bestimmt, die man über den entsprechenden Gegenstand aussagen kann. Individualbegriffe sind solche,

bei denen das Hinzufügen von weiteren Merkmalen keinen Sinn hat. Allgemeinbegriffe sind alle anderen. Da man hier sinnvoll Merkmale hinzufügen kann, lassen sich Allgemeinbegriffe zwangsläufig in (abstrakte oder partitive) Hierarchien einordnen. Die sprachliche Ausdrucksweise ist der kommunizierte Ausdruck eines Begriffes. Solche Ausdrucksweisen sind entweder lexikalisch oder umschreibend, paraphrasierend, gegeben. Kapitel 4 zeigt die Problematik der Verwendung natürlicher Sprache in Informationssystemen. Neben der Mehrdeutigkeit von Worten und der paraphrasierenden Ausdrucksweise sind vor allem die Allgemeinbegriffe ein Suchhindernis, müßte man doch deren Unterbegriffe mit in eine Recherche einbeziehen können.

"Indexsprachen"

Notwendig sind deshalb "Indexsprachen" (Kapitel 5). Ausführlich diskutiert wird in diesem Kontext die Präkombination, wobei Fugmann vor deren breitem Einsatz warnt. Unter- bzw. Oberbegriffe und verwandte Begriffe werden beim Thesauruseinsatz als "siehe-auch-Verweisung" eingesetzt, während die Nichtdeskriptor-Deskriptor-Relation als "siehe-Verweisung" etikettiert wird. Dieser von der Terminologie der Thesaurusnorm (DIN 1463 / Teil 1) abweichende Sprachgebrauch erklärt sich durch Fugmanns Wunsch, die "Indexsprache" auch für Buchregister anzuwenden.

Die Notationen von Klassifikationssystemen sind zwar mnemotechnisch nicht zu gebrauchen, bieten aber durch die Möglichkeit der Rechtsfragmentierung eine einfach durchzuführende hierarchische Suche.

Interessant sind die Ausführungen zu den Relationenindikatoren. Für Begriffe;

die Vorgänge ausdrücken, gibt es eine Reihe weiterer Begriffe, die assoziativ mit dem Vorgang verknüpft sind. Fugmanns Beispiel ist die "Korrosion" (S. 88 f.). Positiv verknüpfte weitere Aspekte werden durch "+", negative durch "-" benannt. Mit Zahlen ist die Richtung bezeichnet (etwa mit "2" die "Mittel" oder mit "7" entsprechende "Apparate". Damit kann Fugmann ganze Begriffsfamilien abbilden:

Korrosion+1
(Korrosion)
Korrosion-1
(Nichteintritt von Korrosion)
Korrosion+2
(korrosiv wirkendes Mittel)
Korrosion-2
(ausgeprägt nicht korrosiv
wirkendes Mittel)
oder
Waschen+7
(Waschmaschine).

Aktualität- und Spezifitätsmängel sind Aspekte, die für jede Indexsprache zu Problemen führen (Kapitel 6). Abhilfe wird durch Deskriptorkandidaten geschaffen, die der natürlichen Sprache zu entnehmen sind (Kapitel 7).

Die Praxis des Indexierens (Kapitel 8) zielt auf Voraussehbarkeit. Nach Fugmanns Ansicht ist Extraktionsindexierung von erheblichen Mängeln begleitet. Der Indexer darf nicht an der konkreten Ausdrucksweise eines Textes hängen, sondern muß die "Essenz wiederauffindbar zu machender Texte" erkennen (S. 103). Dies geschieht durch Zuteilungindexieren mittels der Terme der Indexsprache (sowie - falls nötig - der natürlchsprachigen Deskriptorkandidaten) nach dem Prinzip der "bestpassenden Begriffsbenennung" (S. 118).

Da Texte natürlchsprachig vorliegen und da Textproduktionen indeterminierte Prozesse sind, ist jede automa-

tische Übersetzung und jede automatische Indexierung "Illusion" (S. 124). Lediglich bei Individualbegriffen seien zufriedenstellende Ergebnisse zu liefern.

Buchregister

Jedes Buchregister sollte durch eine Indexsprache und eine vorgegebene Grammatik erstellt werden (Kapitel 9). Optimal wären alphabetisches und systematisches Register. Da dies in Print kaum machbar ist, schlägt Fugmann elektronische Buchregister vor, die ggf. "im Internet zur Verfügung gestellt werden" (S. 145). Am Beispiel des eigenen Buches führt unser Autor seine Variante eines thesaurusgestützten Buchregisters vor.

Indexierung für die Inhouse-Datenbank

Kapitel 10 gibt Ratschläge beim Aufbau einer (unternehmens-)internen Datenbank. Fugmann betont die Notwendigkeit einer Indexsprache für solche Datenbanken. Besprochen wird die Notwendigkeit unterschiedlicher Datenfelder (S. 184). Leider wird weder eine Methode der Datenstrukturierung besprochen noch ein Feldschema als Muster vorgestellt. Inhouse-Datenbanken sind sowohl Literaturdatenbanken (S. 187 ff.) als auch "Haushaltsdatenbanken" (S. 189 ff.). Letztere enthalten alle Dokumente, die in einem Privathaushalt oder in einem Unternehmen anfallen, angefangen von Protokollen über Verträge und Tagesordnungen bis zu Versicherungs- und Kraftfahrzeugakten.

Volltextspeicherung (Kapitel 11) ist für Indexierungszwecke ungeeignet; sie kann ausschließlich dazu dienen, "den kostspieligen und zeitraubenden Beschaffungsweg von Dokumenten drastisch abzukürzen" (S. 194).

Klagen

Kapitel 12 beklagt die "Vernachlässigung einer 'Informations-Philosophie'" (S. 197) und fordert u.a. die intellektuelle Indexierung im Internet (S. 198). Gestreift werden Finanzierungsmodelle für Indexierung und Informationsbereitstellung, wobei Fugmann dem heutigen Marktdenken eine Abfuhr erteilt und sich an die Maximen der 70er Jahre erinnert, wo im IuD-Programm die Informationspraxis als Staatsaufgabe festgelegt war. "(D)ie Sicherung einer geregelten Informationsbereitstellung (ist) eine wichtige Aufgabe der gesamten Gemeinschaft, die von der Nutzung vorhandenen Wissens direkt oder indirekt profitiert oder bald profitieren würde" (S. 204).

Kritik

Was ist neu in diesem Buch? Leider müssen wir feststellen: recht wenig. Letztlich erweist sich das Buch als nahezu überflüssig. Ein Leser, der die Schriften Fugmanns bereits kennt, wird kaum etwas Neues lernen. Natürlich ist es interessant zu erfahren, daß Indexieren auch etwas mit dem Anfertigen von Buchregistern zu tun hat. Allerdings hat eine gemeinsame Theorie zum Buchregister und zum Indexieren in Datenbanken ihre Tücken. Automatische Indexierung spielt beim Buchregister kaum eine Rolle; bei elektronischen Datenbanken ist sie eminent wichtig. Das komplette Ignorieren des Gesamtbereiches der automatischen Inhaltserschließung ist ein

zentraler Schwachpunkt des Buches. Fugmanns Ansatz ist nicht nur theoretisch unhaltbar, sondern auch durch praktische Beispiele faktisch widerlegt. Zu erinnern ist nur beispielsweise im Bereich der kommerziellen Online-Archive an "Freestyle" von Lexis-Nexis, "WIN" von Westlaw oder an Profound sowie im Bereich der Software für Inhouse-Datenbanken u.a. an "Global Text Retrieval" (GTR) oder an Produkte von Muscat bzw. von Verity. Letztlich müßten auch die Schwächen der Internet-Suchmaschinen (wie z.B. "AltaVista") nicht nur einfach behauptet (S. 198), sondern analysiert und evaluiert werden.

Ärgerlich ist, daß das Buch unter "Informationswissenschaft" firmiert. Neben einigen (durchaus zweifelhaften) theoretischen Vorgaben besteht das Buch vorwiegend aus praktischen Ratschlägen. Beispiel: Wie eignet man sich Kenntnisse zur Nutzung von Online-Datenbanken an? Antwort: "am Stammtisch" (S. 192). Dies ist sicherlich kein schlechter Rat, hat aber mit Wissenschaft wenig gemein.

Bei den Dokumentationsmethoden beschränkt sich Fugmann größtenteils auf die Erwähnung von Thesaurus und Klassifikation. Obwohl er in mehreren Stellen ein Datenbankfeld für "Zitiertes" fordert (u.a. S. 190), geht er nicht auf die Zitationsindexierung ein. Obwohl er die Grenzen von Dokumentationssprachen durchaus sieht und das Notieren von Deskriptorkandidaten vorschlägt, nennt er keine Methode (wie etwa die Textwortmethode), die textsprachlich arbeitet. Die Ausführungen zum führenden Regelwerk in deutschen Bibliotheken, den "Regeln für den Schlagwortkatalog" (RSWK) sind derart dürftig (S. 87 f.), daß sie besser ganz weggeblieben wären. Diese wenigen Aspekte verdeutlichen: Wir bekommen keinen Überblick zur Inhaltserschließung durch Indexieren. Auf ein brauchbares Lehrbuch zur Inhaltserschließung in deutscher Sprache müssen wir weiterhin warten.

Wolfgang G. Stock

| SICHERHEIT | | | |
|--|------------------|------------------|--|
| TCP/IP I Grundlagen | 19.-20.07.99/ M | 19.-20.11.99/ M | |
| TCP/IP II Aufbau | 21.-23.07.99/ M | 15.-17.11.99/ M | |
| Virtual Private Network | 19.-20.07.99/ M | 02.-03.11.99/ M | |
| Security Auditing | 22.-23.07.99/ K | 02.-03.09.99/ K | |
| Security Dynamics * | 26.-28.07.99/ M | 28.-30.09.99/ K | |
| Domain Name Service | 05.-06.08.99/ M | 16.-17.09.99/ K | |
| Stone Beat * | 05.-06.08.99/ M | 16.-17.09.99/ K | |
| SAP Security | 17.-20.08.99/ MH | 02.-05.11.99/ MH | |
| Aufbau eines Trust Centers | 26.-27.08.99/ M | 14.-15.10.99/ M | |
| Articon Certified Security Administrator (ACSA) | 20.-24.09.99/ M | | |
| Articon Certified Security Engineer (ACSE) | 26.-30.07.99/ M | 29.-03.12.99/ M | |
| Check Point Firewall I Administration I (CCSA) | 09.-11.08.99/ M | 27.-29.09.99/ M | |
| Check Point Firewall I Administration II (CCSE) | 08.-10.09.99/ M | 08.-10.11.99/ M | |
| Nokia Firewall | 14.-15.09.99/ M | 04.-05.10.99/ M | |
| Raptor Firewall | 14.-16.07.99/ M | 17.-19.11.99/ M | |
| Makroviren erkennen und bekämpfen | 13.-14.09.99/ M | 14.-15.10.99/ M | |
| Ausbildung zum Computer-Virenschutz-Beauftragten | 12.-13.10.99/ M | | |

*: neue Seminare, M: München, K: Köln, MH: Mannheim

Weitere Themen und Termine unter www.articon.de

ARTICON
Information Systems AG

Güterbergstraße 1 · 85737 Ismaning · Tel: 089/945 73 - 0
Fax: 089/945 73 - 449 · E-Mail: info@articon.de

SEMINARE