

Automatische Auswertungsmethoden (Information Retrieval)

Autor(en): **Wegmüller, F.**

Objektyp: **Article**

Zeitschrift: **Nachrichten VSB/SVD = Nouvelles ABS/ASD = Notizie ABS/ASD**

Band (Jahr): **53 (1977)**

Heft 6

PDF erstellt am: **21.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-771438>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

tionsgefüges, beispielsweise die Zentralstelle für maschinelle Dokumentation und das Lehrinstitut für Dokumentation. Der Wandel des, ohnehin stark traditionell verwurzelten, Bibliothekswesens zu modernen «Kommunikations- und Informationszentren» vollzieht sich dagegen sehr viel harziger. Umso bemerkenswerter sind die, von der Deutschen Forschungsgemeinschaft unterstützten, *Maßnahmen zu verbesserter Erschließung*: stärkere Gewichtung der Auskunftsdienste, «Verzeichnis laufend erscheinender Bibliographien» (1963), Karteien der Fortschritts- und Übersichtsberichte, Katalogisierung ausgewählter Zeitschriftenaufsätze (neu vor allem in Germanistik, Anglistik und Romanistik).

¹ Lexikon des Bibliothekswesens. Leipzig 1974/5, Sp. 163, 923.

² Heinr. Roloff: Lehrbuch der Sachkatalogisierung 4. Aufl. Leipzig 1975. S. 138.

³ Lexikon des Bibliothekswesens Sp. 525—28, 800/801 (Roloff).

⁴ Gottfr. Weilenmann: Moderne Registratur-Organisation. Bern 1973. S. 17.

⁵ Guide to Reference Material ed. 3. London 1972 ff. Guide to Reference Books ed. 9. Chicago 1976. Usw.

⁶ Vgl. Biological Abstracts gestern und heute — oder gar den Science Citation Index ...

⁷ Georg Schneider: Einführung in die Bibliographie. Leipzig 1936. S. 187. Zu Schneider als Bibliographietheoretiker vgl. Joris Vorstius, ZfB 47, 1930 S. 323 ff. insbesondere S. 324.

⁸ Heinr. Roloff und Georg Leyh, in Handbuch der Bibliothekswissenschaft 2. Aufl. (1952—65), Bd. 2 S. 242 und Bd. 3/2 S. 468.

⁹ Rolf Kluth: Gibt es eine Bibliothekswissenschaft? Zeitschrift für Bibliothekswesen und Bibliographie 17, 1970. S. 245/6.

Automatische Auswertungsmethoden (*Information Retrieval*)

F. Wegmüller

F. Hoffmann-La Roche & Co. AG, 4002 Basel

Zusammenfassung

Information Retrieval befaßt sich mit dem Wiederauffinden gespeicherter Informationen. Konventionelle und automatisierte Methoden werden verglichen und die Voraussetzungen für die automatische Durchführung, sowie die damit verbundenen Probleme behandelt.

1. Einleitung

Unter dem Begriff Information Retrieval (IR) werden Aktivitäten und Methoden zusammengefaßt, die das Auffinden von gespeicherten Informationen zum Ziel haben. IR kann durch die generelle Frage «welches Wissen existiert und wo ist es niedergelegt?» charakterisiert werden. IR vermittelt nicht direkt das Wissen selbst, sondern nur den Schlüssel dazu. Als Antwort werden aktuelle, umfassende und präzise Hinweise erwartet.

Wie jede andere Tätigkeit ist IR mit Aufwand an Zeit, materiellen und intellektuellen Mitteln verbunden und sollte einen Nutzen erbringen. Der Aufwand läßt sich erfassen, aber trotz allen Bemühungen, den Nutzen zu quantifizieren, konnte bis heute kein Maß für seine Berechnung gefunden werden. Daher gibt es auch keine reale Relation Kosten/Nutzen.

Dies gibt Anlaß, die konventionellen Methoden des IR den modernen, automatisierten Verfahren gegenüberzustellen. Dieser Vergleich kann sich u.a. beziehen auf: Art und Menge der verfügbaren Daten; Wirksamkeit der Recherchiermöglichkeiten in bezug auf Aktualität, sachliche Kriterien, Vollständigkeit und Präzision der Resultate; zeitlichen Aufwand für die Durchführung, Terminfrage; anteilmäßige Kosten für Personal, Daten und technische Einrichtungen; Kosten für Erstellung der Antwort und ihre Nachbearbeitung.

Bei der Verlagerung der Vorgänge vom konventionellen auf automatisiertes IR findet meistens auch ein Übergang der Arbeitslast vom Fragesteller (= Benutzer) zu einem Ausführenden (= Dokumentalisten) statt. Dabei entstehen zusätzliche Probleme z. T. psychologischer Natur: Hemmungen, über die Frage zu sprechen; fehlendes Vertrauen in bezug auf die Geheimhaltung; Ersatz der Intuition durch starre Maschinenalgorithmen; Wegfall des «Browsing» und damit der spontanen Assoziationen. Demgegenüber stehen seitens des Ausführenden: größere Gewandtheit im Umgang mit Daten, Methoden und technischen Einrichtungen, dadurch Zeitgewinn und umfassenderer Einsatz der Mittel; bessere Kenntnisse der Datenbestände und Suchstrategien, dadurch vollständigere und präzisere Ergebnisse. In diesem Sinne können wir von Vor- und Nachteilen des automatisierten IR sprechen.

Ein weiteres Problem entsteht dadurch, daß der Fragesteller bei der selbständigen Durchführung von Recherchen als Fachspezialist handelt, seine Fachkenntnisse und seinen Fachjargon einsetzen kann. Ein Dokumentalist kann die gleiche Aufgabe nur dann vollwertig übernehmen, wenn er über die entsprechenden Fachkenntnisse verfügt. Daher gibt es *den* Dokumentalisten nicht. Die Fähigkeit, sinnvolle Dokumentationsarbeit zu leisten, ist immer mindestens mit einer Grundausbildung in der gerade in Frage kommenden spezifischen Fachrichtung verbunden.

Erfahrungsgemäß bedient sich der Fachspezialist, der z.B. aktiv in der Forschung tätig ist, vorzugsweise der konventionellen Mittel, um selber IR zu betreiben und wird sich nur ausnahmsweise und widerwillig in die Metho-

den der automatisierten Möglichkeiten einarbeiten. Die ganze Entwicklung der modernen Dokumentationsmethoden hat in den letzten 20 Jahren stattgefunden und verlief häufig allzu getrennt von derjenigen der Wissenschaften, denen sie dienen können. Die Schuld an dieser Divergenz ist neben andern Faktoren vornehmlich auf Mängel und Lücken in den Ausbildungsplänen der Mittelschulen und der *éducation permanente* zurückzuführen.

Das Bestehen einer konventionellen Bibliothek und ihr Weiterbetrieb wird im allgemeinen als notwendig und selbstverständlich erachtet. Kredite stehen zur Verfügung, wenn auch nicht immer genügend, während für Dienstleistungen mittels automatisierter Methoden meistens noch herzlich wenig Verständnis aufgebracht wird. Die Beanspruchung eines automatisierten Literaturdienstes aber verursacht erfaßbare Kosten, für die kein Kredit vorhanden ist, obschon dadurch wertvolle Zeit eingespart werden kann und gegebenenfalls unnötige Doppelarbeit vermieden wird.

2. Voraussetzungen

Um IR betreiben zu können, müssen Voraussetzungen erfüllt sein, die banal und selbstverständlich tönen. Aber um alle Aspekte des Problems IR abzudecken, ist es manchmal gut, das Selbstverständliche, das meistens im Unbewußtsein angesiedelt ist, wieder ins Bewußtsein zu rufen.

2.1. Die Daten müssen existieren, mit andern Worten: Daten müssen auf einem physischen Datenträger dauerhaft fixiert und lesbar sein, z.B. als Schriftzeichen auf Papier oder photographischem Material, als Markierungen auf Lochstreifen, Lochkarten, photographischen oder magnetisierbaren Trägern. Gemeinsames Merkmal ist die Verwendung alphabetischer und numerischer Zeichen oder ihrer Äquivalente als geometrisch definierte Anordnung von Lochungen, optischen Dichteunterschieden oder magnetisierten Bezirken.

Daten finden sich aber auch als graphische Schaubilder, z. B. Kurven, in eigentlichen Bilddokumenten wie Mikroaufnahmen, Röntgenbildern, Landkarten, oder in abstrakter Form als Ideogramme. Für die Zwecke des maschinellen IR müssen derartige Daten linearisiert und in alphanumerische Ausdrucksweise transkribiert werden. Maßgebend ist dabei der Zeichenvorrat unserer technischen Einrichtungen, sowohl für die Speicherung, den Suchvorgang und die Ausgabe der Resultate. Daher müssen auch Zeichen fremder Alphabete, z. B. Griechisch oder Kyrillisch, transliteriert oder umschrieben werden. Transliteration ist zweckmäßig für lautgerechte Umsetzungen. Bedeutet dagegen ein Zeichen eines fremden Alphabets eine spezifische Qualität oder Dimension, würde Transliteration einen Informationsverlust herbeiführen. In solchen Fällen muß das Zeichen daher umschrieben werden, z.B. α = alpha.

2.2. *Die Daten müssen zugänglich sein.* Die Zugänglichkeit muß sowohl zeitlich, wie örtlich, technisch und begrifflich gesichert sein. Wir alle kennen den Ärger, wenn das Referateblatt beim Buchbinder oder der Index dazu noch nicht erschienen ist, wenn für den Mikrofilm kein geeignetes Lesegerät zur Verfügung steht, wenn ein 7-Track-Magnetband auf unserer 9-Track-Leseinheit nicht verarbeitet werden kann, andere Schreibdichten oder Magnetbandcodes vorliegen, oder wenn wir einen Text in einer uns nicht geläufigen Sprache vor Augen haben. Prinzipiell können alle diese Schwierigkeiten überwunden werden, bedingen aber oft einen erheblichen zusätzlichen Aufwand. Effizientes IR ist nur möglich, wenn sich die Daten in der unsern intellektuellen und technischen Mitteln unserer Umgebung entsprechenden Form vorfinden.

2.3. *Die Daten müssen erfaßbar sein.* Mit Originaldokumenten allein läßt sich IR kaum und auf jeden Fall nicht rationell betreiben. Ein Originaldokument enthält normalerweise und gemessen an den Erfordernissen des IR einen derart hohen Prozentsatz redundanter und irrelevanter Information, daß der Aufwand zum Auffinden des Relevanten unverantwortbar wird. Schon zu Beginn der wissenschaftlichen Ära ergab sich die Notwendigkeit der Kondensation der Informationen auf das Wesentliche in Form von Inhaltsverzeichnissen, Sachregistern, Katalogen und Bibliographien. Der gewaltige intellektuelle und technische Aufwand für die Erstellung der Sekundärliteratur machte es notwendig, nicht nur die Mittel und Ausrüstung für die Bewältigung der praktischen Arbeit bereitzustellen, sondern auch das Phänomen der Informationsvermittlung theoretisch und praktisch zu untersuchen. Allerdings ist der Standpunkt des Benutzers lange Zeit zu kurz gekommen. Es gab keine, und es gibt heute noch nur vereinzelte Benutzerorganisationen und die Benutzerforschung steht erst in den Anfängen.

Mit dem Eindringen der elektronischen Datenverarbeitung in die Dokumentationsmethode ergaben sich nun eine Reihe von Denkanstößen, die sich außerordentlich fruchtbar auf Gehalt und Gestaltung der Sekundärliteratur ausgewirkt haben: Systematisierung der Arbeitsabläufe, Standardisierung und Kategorisierung von Aussagen, Entwicklung von Thesauri, Beziehung von Semantik und Syntax, vieldimensionale Betrachtungsweisen und Kombinatorik und vieles Andere mehr.

Die Erfassung der Daten beruht primär auf der Selektion des Wichtigsten aus dem Gesamtangebot von Informationen und seiner Darstellung in geeigneter Form. Bereits in diesem Stadium des Werdegangs einer Dokumentation ist der Benutzer des daraus resultierenden IR-Systems einem anonymen Kollektiv von Bearbeitern ausgeliefert, das für ihn entscheidet, was wichtig und was unwichtig ist. Beispiele: Wahl der Zeitschriften aus dem Gesamtrepertoire, Artikelselektion aus diesen Zeitschriften, Differenzierung der Sachinformation in diesen Artikeln (Indextiefe), Darstellung der Sachverhalte in Form von Deskriptoren oder Notationen.

Die einfachsten IR-Systeme übernehmen aus ökonomischen Gründen von den Originaldokumenten nur Titel, Autoren und bibliographisches Zitat, ohne irgendwelchen intellektuellen Aufwand. Eine solche Erfassung ist rudimentär und liefert beim IR auch nur entsprechende Ergebnisse.

Die Anreicherung mit Deskriptoren und/oder Notationen und Abstracts stellt wesentlich höhere Anforderungen, gibt aber auch entsprechend bessere Resultate. Dieser Ausbau zeigt große Unterschiede von einer Datensammlung zur andern. IR ist daher nur sinnvoll und wirksam, wenn die gesamten Systemunterlagen, wie Zeitschriftenlisten, Selektionsregeln, Wortlisten, Thesauri, Code- und Notationstabellen usw. zur Verfügung stehen und gründlich beherrscht werden.

2.4. Die Methoden des Zugriffs müssen existieren. Die erfaßten Daten haben spezifische Aussagewerte auf unterschiedlichen Informationsniveaux. Numerische Angaben z. B. können so unterschiedliche Dinge wie Sequenznummern oder Jahrgang oder quantitative Maßangaben bedeuten. Autorennamen werden auch zur Charakterisierung von Sachbegriffen oder Vorgängen verwendet. Mittels definierter Anordnung der Daten gemäß ihrer Bedeutung werden auf dem Datenträger Formate gebildet, die mit Hilfe der Retrieval-(Such-)programme einen spezifischen Zugriff zu den einzelnen Datenkategorien gewähren.

Die im Handel durch Kauf oder Miete erhältlichen Datensammlungen unterscheiden sich in der Art und Zahl der Datenkategorien, wie auch in der Kennzeichnung der Datenfelder. Wenn daher Datensammlungen verschiedener Provenienz im gleichen IR-System verarbeitet werden sollen, müssen die unterschiedlichen Formate in ein Einheitsformat konvertiert werden, ein Umstand, der für den Austausch von Datenmaterial ein beachtliches Hindernis darstellt, umsomehr, als gegebenenfalls ein mit der Konvertierung verbundener Informationsverlust in Kauf genommen werden muß.

Der Suchvorgang besteht in der Feststellung der Kongruenz zwischen den Daten der Frage und den Daten des Datenvorrats im Speicher. Fragen bestehen normalerweise aus mehreren, manchmal hunderten von Frageterms, für die eine bestimmte logische Verknüpfung vorgeschrieben werden kann. Das Retrieval-Programm muß daher in der Lage sein, diese oft komplizierten und verschachtelten Verknüpfungen nach den Regeln der Boole'schen Algebra herstellen zu können. Als logische Operatoren werden am häufigsten UND, ODER, UND NICHT, gelegentlich auch KLEINER ALS und GRÖßER ALS verwendet. Zur rationellen Durchführung des Suchvorgangs ist es erforderlich, eine Suchstrategie für jede Frage zu entwickeln, die z. B. die statistischen Häufigkeitswerte der Frageterms zur Abkürzung des Arbeitsablaufs heranzieht.

Zudem spielt es eine Rolle, ob eine Datensammlung (File) sequentiell in der numerischen Reihenfolge der dokumentarischen Einheiten aufgebaut ist, oder ob ein invertiertes File zur Verfügung steht, in dem die Reihenfolge

durch die vorsortierten Klassifikationsmerkmale bestimmt ist. Die unterschiedliche Art des Zugriffs für die beiden File-Konstruktionen bedingt das Vorhandensein spezifischer Konfigurationen der Hardware und eine entsprechende Software. Für sequentielle Files kommen als Datenträger praktisch alle Speichermöglichkeiten, hauptsächlich Magnetbänder in Frage, während für invertierte Files Magnetplatten oder -trommeln benötigt werden, die den Direktzugriff (random access) ermöglichen.

Beim sequentiellen File kann der Inhalt der den Suchkriterien entsprechenden dokumentarischen Einheiten direkt ausgedruckt werden, während beim invertierten File nur die Nummern dieser Dokumente anfallen. Die für eine direkt interpretierbare Antwort notwendigen Angaben müssen anschließend an den Suchvorgang von einem Bibliographie-File geholt werden.

Beim invertierten File können die Suchkriterien selbst, z. B. Deskriptoren, die Zugriffsadresse darstellen. Deskriptoren können aber auch in einem in der Maschine gespeicherten Dictionary enthalten sein, wobei jedem Deskriptor oder einer begrifflich zusammengehörenden Deskriptorengruppe, die z.B. sämtliche Synonyme umfaßt, eine Nummer zugewiesen ist, die als maschineninterne Zugriffsadresse dient. Auf diese Weise kann das Synonymenproblem teilweise gelöst werden, setzt aber eine intellektuell aufwendige thesaurusmäßige Organisation des Deskriptorenbestandes voraus.

Von Bedeutung für die Suchstrategie ist die Verwendung der Truncation und der dummy characters. Bei der Truncation wird nur ein charakteristischer Wortstamm, bzw. Anteil eines Deskriptors gesucht, Präfix oder Suffix oder beides wird weggelassen und durch ein Spezialzeichen ersetzt. Der dummy character ist ein allgemeines Zeichen, das anstelle eines spezifischen Buchstabens in einem Wort eingesetzt wird und dadurch verschiedene Schreibweisen eines Begriffs in einer einzigen vereinigt. Beide Methoden können den Suchvorgang abkürzen oder unbewußt entgangene Begriffe in die Fragestellung einschließen.

Sequentielle und invertierte Files können im batch-Verfahren eingesetzt werden. Vorzugsweise werden die Retrieval-Programme so gestaltet, daß nicht jede Frage einzeln, sondern mehrere, je nach Maschinenkapazität auch hunderte von Fragen simultan verarbeitet werden können. Damit ergeben sich erhebliche Einsparungen an Computerzeit.

Moderne IR-Systeme benutzen invertierte Files on-line im real-time-Zugriff, d.h. daß auf jeden eingegebenen Programmbefehl unmittelbar eine Reaktion des Computers in Form einer Antwort oder Rückfrage oder Anweisung erfolgt. Auf grund dieser Angaben ist es möglich, den Ablauf des Fragevorgangs zweckmäßig zu steuern oder gegebenenfalls zu modifizieren und in dem sich ergebenden Dialog mit der Maschine den Suchvorgang und das Ergebnis zu optimieren. Die Eingabe der Operationsbefehle und der Frageterms erfolgt über die Tastatur eines Terminals, die Ausgabe der Antworten und Rückfragen über den Drucker und/oder den Bildschirm.

Retrospektive Recherchen verlangen eine sorgfältig abgewogene Fragestellung, die möglichst wenig Ballast, d. h. irrelevante Antworten, und keine Lücken (Missings) in der Antwort bewirkt. Da meistens die Zahl der nachgewiesenen dokumentarischen Einheiten relativ groß ist, wirkt sich der Ballast störend und zeitraubend aus. Bei Recherchen für current awareness (SDI), die laufend nur aus den Neuzugängen des Informationsmaterials durchgeführt werden, ist die Zahl der nachgewiesenen Einheiten relativ klein. Der Nachweis von Informationen über Randgebiete ist meistens erwünscht. Die Frage kann daher viel breiter angelegt werden und allfälliger Ballast fällt weniger ins Gewicht.

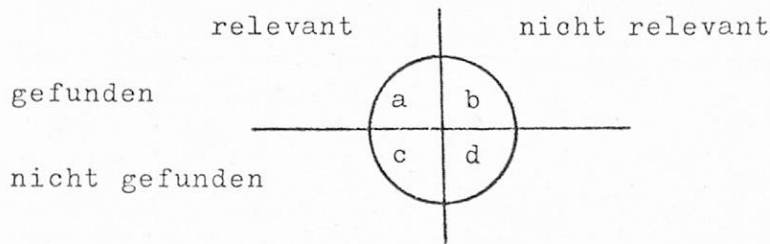
Die Vielfalt der durchzuführenden Programmschritte erfordert für die Fragenformulierung eine eindeutige Anfragesprache (Metasprache). Wie sich für die Programmiersprachen selbst eine Entwicklung von der maschinen- zur problemorientierten Sprache ergeben hat, ist auch eine Entwicklung von symbolischen zu natürlich-sprachlichen Elementen der Anfragesprachen festzustellen. Als weiterer Schritt ist die Vereinheitlichung der Anfragesprachen, die heute in den verschiedenen IR-Systemen in Anwendung sind, anzustreben.

2.5. Die aufgefundenen Informationen müssen eine Aussage beinhalten und interpretierbar sein. Der durchschnittliche Benutzer eines IR-Systems erwartet auf seine Frage eine Antwort, die ihm ohne großen Aufwand ermöglicht, zu beurteilen, ob er ein nachgewiesenes Originaldokument für seine Zwecke konsultieren soll oder nicht. Reine Referenznummernlisten können diesen Zweck nicht und reine Titellisten nur unvollkommen erfüllen. Die Aussagekraft wird wesentlich größer, wenn die Antworten Deskriptoren einschließen. Noch besser ist es, wenn Abstracts mitgeliefert werden.

Die Beurteilung des Wertes, oder besser gesagt des Nützlichkeitsgrades einer Recherche ist eine recht kritische Angelegenheit. Zur zahlenmäßigen Erfassung werden meist die Parameter Relevanz = Präzision und Recall herangezogen. Ihre Bedeutung und Berechnung geht aus der folgenden Darstellung (Seite 325) hervor.

Der Dokumentalist beurteilt z.B. eine Antwort als relevant, wenn sie den formalen und sachlichen Kriterien der Frage entspricht. Der Benutzer dagegen betrachtet Relevanz von seinem Kenntnisstand aus. Er kann daher Irrelevantes als interessant bezeichnen, Relevantes dagegen als nicht neu oder wertlos abtun. Zudem ist es bei umfangreichen IR-Systemen vollständig unrealistisch oder sogar unmöglich, den Recall-Parameter zu bestimmen.

Die pragmatische Beurteilung eines Resultats durch den Fragesteller ist bedeutsamer als jede Berechnung. Diese Beurteilung als Feedback, zusammen mit der Erfahrung des Dokumentalisten ist eine gute Hilfe für die Verbesserung der Qualität im Frage-Antwortspiel. Daher versenden einige Dokumentationsstellen mit der Antwort einen Fragebogen, um die Beurteilungen der Recherchen systematisch zu untersuchen.



- a: gefundene relevante Referenzen
 b: gefundene nicht relevante Referenzen = Ballast
 c: nicht gefundene relevante Referenzen = Missings
 d: nicht gefundene nicht relevante Referenzen

$$\frac{a}{a + b} : \text{Relevanz} = \text{Präzision} \quad \frac{a}{a + c} : \text{Recall}$$

3. Praktische Durchführung

Die folgende Zusammenstellung ist als Übersicht über die möglichen Vorgänge beim automatisierten IR gedacht. Sie ist eher eine Art «Checklist» als ein chronologischer Ablauf. Bei einzelnen Entscheidungen muß immer wieder auf vorherige Vorgänge im Sinne eines Regelkreises zurückgegriffen werden.

<i>Problem</i>	<i>Entscheidung</i>
Ist die Frage verständlich?	Rückfrage an Fragesteller, Klarstellung Beispiele typischer Publikationen
Ist die Frage zu allgemein?	Rückfrage an Fragesteller, Einengung?
Ist die Frage zu eng gefaßt?	Rückfrage an Fragesteller, Erweiterung?
Retrospektive Recherche	wie weit zurück? Letzte 3, 4, ... Jahre? fraktionierte Durchführung? eventuell Ergänzung, konventionell? im Anschluß an ein SDI-Profil?
Current awareness (SDI)	im Anschluß an eine retrospektive Recherche? genügt vorhandenes Makroprofil? spezielles Profil neu erstellen? kann Profil mit andern ähnlichen Profilen zusammengelegt werden? Laufzeit? sind während der Laufzeit Modifikationen zu erwarten?

<i>Problem</i>	<i>Entscheidung</i>
Welche Datensammlungen (Files) sind geeignet?	Auswahl nach Sachgebiet und innerhalb des Sachgebietes nach besonderen Spezifikationen (Art der Datenerfassung, Umfang der erfaßten Originalliteratur, Indextiefe, Patente/Zeitschriften)
Wo ist File zugänglich?	eigener Bestand fremde Datenbank direkt zugänglich (Terminal): SDC, Lockheed, SDS/ESA indirekt zugänglich: DOKDI
<i>Problem</i>	<i>Entscheidung</i>
Durchführung wie?	eigene EDV-Anlage eigenes Terminal Servicebureau Auftrag an fremde Dokumentationsstelle telefonisch schriftlich (besondere Formulare?)
Wo wird abgefragt? Bereichsabgrenzung (Kategorien, Datenfelder)	Autoren Bibliographie (Zeitschriftentitel, Zeitraum) Deskriptoren (Keywords, Aufsatztitel, Abstract) spezielle Indikatoren (z. B. Sprache) numerische Daten (Ordnungsnummern, quantitative Angaben) Codes Bit strings
Was wird abgefragt? Sachbegriffe Ergänzungsterms	Thematische Gruppen, Sections Deskriptoren (Ausdrücke in klarer Sprache) freies Vokabular (Synonyma-Problem) standardisiertes (gebundenes) Vokabular (Thesauri) gemischtes Vokabular (Codeless Scanning) Deskriptorenmodifikation Truncation (rechts, links, beidseitig) dummy characters Codes, Notationen DK Wiswesser Line Notation andere alphanumerische Codes Bit-Nummern Matrix-Darstellungen, Topologie numerische Daten benannte und unbenannte Zahlen Syntax-Indikatoren (Verknüpfungen) Links, Roles, Tagging, Gewichtung beizuziehen: Term-Statistik (Häufigkeit)

<i>Problem</i>	<i>Entscheidung</i>
Wie wird abgefragt? Logische Operatoren	AND, OR, NOT WITH (Vorkommen im gleichen Satz) Wortabstand kleiner als . . . , größer als . . . Gewichtung
Fragenformulierung Recherchenstrategie Kombination von <i>Wo, Was, Wie</i>	Verknüpfung von Frageelementen maschinenlesbar Kategorien Sachbegriffe logische Operatoren Anfragesprache (Metasprache)
Frageninput	Lochbelege, Lochkarten, Lochstreifen Terminal remote batch on-line (Dialog)
Antwortenoutput	Magnetband, z. B. für Subfiles Printer off-line on-line (Terminal) Bildschirm allein Bildschirm mit Druckwerk COM (Computeroutput on Microfilm) Print-Optionen nach Kategorien Formate endlos Papier Karten
Beurteilung des Output	formale Kriterien, Druckqualität (für Offset-Vorlagen), Reihenfolge der Datenfelder intellektuelle Kriterien Relevanz, Ballast Recall, Missings Feedback, Auswertung von Fragebogen
Nachbearbeitung	Konsultation gedruckter Abstracts Lieferung von Xeroxkopien der Abstracts Lieferung von Mikrofilm-Rückvergrößerungen der Abstracts Konsultation der gedruckten Originale Lieferung von Xeroxkopien der Originale Lieferung von Mikrofilm-Rückvergrößerung der Originale Redaktionelle Überarbeitung

Literatur

Lancaster, F. W.: Information Retrieval Systems. Characteristics, Testing, and Evaluation. John Wiley & Sons, Inc., New York (1968).

Salton, G.: Automatic Information Organization and Retrieval. McGraw-Hill Book Company, New York (1968)

Vickery, B. C.: Zur Theorie von Dokumentationssystemen. Ed.: Deutsche Gesellschaft für Dokumentation e. V., Verlag Dokumentation, München-Pullach (1970)

King, D. W. und Bryant, E. C.: The Evaluation of Information Services and Products. Information Resources Press, Washington, DC (1971)

Mikhailov, A. I., Cernyi, A. I. und Giljarevskij, R. S.: Grundlagen der wissenschaftlichen Information und Dokumentation. Westdeutscher Verlag, Köln (1970)

Chaumier, J.: Les techniques documentaires (Série Que sais-je? No. 1419). Presses Universitaires de France, Paris (1971)

Chaumier, J.: Systèmes informatiques de documentation. Entreprise Moderne d'Édition, Paris (1972)

Laisiepen, K., Lutterbeck, E. und Meyer-Uhlenried K.-H.: Grundlagen der praktischen Information und Dokumentation. Verlag Dokumentation, München-Pullach (1972)

Mauerhoff, G. F.: Selective Dissemination of Information. In: Advances in Librarianship, Vol. 4, 25—62 (Ed.: Voigt, M. J.). Academic Press, New York (1974)

Cuadra, A. C. (Ed.): Annual Review of Information Science and Technology. (Vol. 1, 1966 — Vol. 10, 1975). American Society for Information Science, Washington, DC.

Klawiter-Pommer, J. H. T. und Hoffmann, W. D.: Übersicht über die für den Leistungsvergleich mehrerer Literatur-Datenbasen wichtigsten Parameter. Nachr. Dok.27 (1976) Nr. 3, 103—108

Les techniques conventionnelles d'entreposage des supports d'information

T. J. Tanzler

Bibliothèque de l'École polytechnique fédérale de Lausanne
Av. de Cour 13, 1007 Lausanne

Résumé

Autrefois, le livre fournissait des connaissances, il était lu en entier. De nos jours, l'utilisation d'une bibliothèque ne cherche souvent qu'une information précise. L'exposé décrit la manière de placer les livres et autres supports d'information pour les rendre accessibles aux utilisateurs en partant du moyen âge à nos jours et en esquissant au passage certains problèmes techniques et