

BIBLIOTHEKSTECHNISCHE INFORMATIONEN

Persönliche Literaturdateien am Personal-Computer: Kriterien für Softwareauswahl und -einsatz

Nahezu schlagartig wurde in den letzten Monaten ein Themenbereich aktuell, der Bibliotheken und Bibliothekare ganz besonders berühren sollte: Gemeint ist die Verwendung von Datenbanksystemen auf Personal-Computern (PCs) zum eigenen, persönlichen Gebrauch. Davon betroffen sind in erster Linie Bibliothekskunden, also Wissenschaftler bzw. Studenten, die den naheliegenden Wunsch verspüren, die im Rahmen ihrer wissenschaftlichen Tätigkeit anfallenden Literaturdaten auf dem eigenen PC zu verwalten. Der Privatbesitz von PCs hat in der jüngsten Vergangenheit — besonders begünstigt durch die sogenannten „clones“ oder Taiwaner-PCs — in Hochschulkreisen epidemisch zugenommen, so daß der Umfang der interessierten Zielgruppe nicht unterschätzt werden darf. Die subjektiven Bedürfnisse reichen dabei vom Ersatz für den traditionellen „Zettelkasten“ mit Literaturzitaten und Anmerkungen, den bisher wohl jeder Dissertationskandidat anzulegen hatte, bis zur Führung einer kleinen, privaten Literaturdatenbank, welche über einen längeren Zeitraum hinweg verwendbar sein sollte und die Katalogisierung bzw. Erschließung des persönlichen Archivs von Artikelkopien, Sonderdrucken, Büchern usw. zu leisten hätte. Wer erfährt im Zeitalter der Textverarbeitung — immer mehr Wissenschaftler benutzen

längst keine Schreibmaschine mehr — Literaturstellen, die für mehrere Arbeiten verwendet oder zitiert werden sollen, schon gerne mehrmals? Wer sortiert sie schon gerne „händisch“, wenn dies der PC schnell und mühelos vermag? Und wer vergeudet schon gerne wertvolle Zeit, um ein verspätet hinzugekommenes Zitat in die bereits manuskriptfähig erstellte Literaturliste „einzuflicken“ und eventuell sogar einen neuen Seitenumbruch zu erstellen?

Entsprechend tönt es auch aus der aktuellen Firmenwerbung. Ein Beispiel: „Karl Marx verbrachte den größten Teil seiner Zeit auf der Suche nach Literatur und konnte sine Hauptwerk ‚Das Kapital‘ nur zur Hälfte fertigstellen. Ihm fehlte ein superschnelles Informationssystem wie XYZ. Auch Sie sollten nicht mehr wertvolle Wochen mit Registern und Indexverzeichnissen verbringen. Ob Literatur, Patente, Krankheitssymptome, Filme, Schallplatten, Bücher, Zeitschriften, Mitglieder, Fotos oder was sonst an Datenbeständen existiert, durchforstet XYZ und findet die Antwort auf Ihre Fragen in wenigen Sekunden.“

Tatsächlich hat sich gerade in letzter Zeit das Produktangebot für Textdatenbanksysteme stark verändert. Galt noch bis vor kurzem, wann immer das Gespräch die Problematik persönlicher Literaturdateien auf PCs berührte, die Verwendung gängiger PC-Datenbanksysteme wie z. B. dBase III — ungeachtet der fraglichen Eignung derselben für diesen Einsatzzweck — als das Optima

des Erreichbaren, so fällt es heute bereits schwer, sich im Softwareangebot überhaupt noch zurechtzufinden. Klangvoll sind meist die Namen der Programme, schier unbegrenzt die versprochenen Leistungen. Was leisten diese Produkte jedoch wirklich? Erfüllen sie die jeweiligen subjektiven Benutzervorstellungen und -ansprüche? Dies erfährt man meist erst dann, wenn sich die gepriesene „Volltextdatenbank“ bereits im Eigenbesitz befindet und trotz hohem Kaufpreis und gutem Zuspruch beispielsweise nicht gewillt ist, feldspezifisch zu suchen, sinnvolle logische Verknüpfungen zu realisieren oder die meist gar nicht so große Datenmenge in annähernd erträglicher Zeit zu recherchieren.

Die folgende Übersicht versteht sich vor diesem Hintergrund als Beurteilungshilfe, wenn es um die Auswahl von Datenbanksoftware für die Verwaltung bibliographischer Informationen auf PCs geht. Da vom Bibliothekar zunehmend eine Beratungsfunktion auch auf diesem Gebiet erwartet wird, wendet sich dieser Kriterienkatalog sowohl an Endbenutzer als auch an bibliothekarische Kollegen.

1. Entscheidungen, die vor der Erstellung einer persönlichen Literaturdatei zu treffen sind:

1.1. Festlegen der Dokumentationsbedürfnisse und -ziele:

Vor allen anderen Entscheidungen muß klar festgelegt werden, was überhaupt dokumentiert werden soll. Sollen z. B. nur Literaturzitate verwaltet werden oder soll mit derselben Software auch eine Adressenverwaltung oder eine Faktendokumentation (etwa Meßwerte, Fotos usw.) möglich sein?

- Um welche *Dokumentationseinheiten* wird es sich also handeln?
- Welche *Datentypen* werden dabei auftreten? (Textdaten vs. numerische Daten)
- Mit welchen *Größenordnungen* wird zu rechnen sein? Wie groß (ausgedrückt in Zeichen oder Bytes) wird jeweils eine Dokumentationseinheit

im ganzen sein? Wie groß werden die einzelnen Felder bzw. Kategorien (z. B. Titel, Quellenangabe usw.) der Dokumentationseinheit sein? Wie groß (ausgedrückt in Datensätzen oder „records“) wird letztlich die gesamte Datenbank sein?

1.2. Organisatorische Fragen:

- Welche *Datenquellen* werden für den Aufbau der Datenbank bzw. ihre Aktualisierung (das sogenannte „updating“) zur Verfügung stehen? In Frage kommen einerseits der manuelle Input, andererseits aber auch die Nutzung von Fremddaten (bzw. über „downloading“ aus Literaturdatenbanken).
- Ist an eine *sachliche Erschließung* der persönlichen Literaturdatei gedacht? Dieser Aspekt birgt besonders wichtige organisatorische Entscheidungen (Schlagwortsystem, Thesaurus, Klassifikation, Abstracts; aber auch: Kontinuität und Konsistenz der Sacherschließung), auf die hier im Detail nicht weiter eingegangen werden kann.

1.3. Rechner und sonstige Hardware

- Der Umstand, daß ein Rechner zur persönlichen Verfügung steht, beweist noch lange nicht, daß er für die Lösung eines Problems auch wirklich *notwendig* ist. Daher überlege man sich vor der Mechanisierung der persönlichen Literaturdatei gründlich, ob dies auch die beste Lösung ist, denn ein Grundsatz des EDV-Einsatzes lautet, daß die einmal getroffene Entscheidung zugunsten der EDV kaum mehr revidierbar ist.
- Häufig handelt es sich bei dem zur Verfügung stehenden PC um ein vorgegebenes Faktum. Sollte dies nicht der Fall sein, so ist zu empfehlen, ein *I-PC-kompatibles* Modell zu wählen (wobei das „I“ nicht etwa für einen Firmennamen, sondern für „Industriestandard“ steht), da dafür derzeit die größte Auswahlmöglich-

keit an Softwareprodukten gegeben ist.

- Bestimmte Programme setzen eine *Mindestgröße des Hauptspeichers* (z. B. 512 KB) voraus.
- Hinsichtlich des verfügbaren *Externspeichers* erscheint eine Festplatte als Voraussetzung für problemloses Arbeiten. Zwar betonen manche Softwareerzeuger, daß ihr Produkt auch mit Disketten allein verwendbar sei, doch zeigt die Erfahrung, daß gerade bei speicherplatzintensiven Literaturinformationen davon abzuraten ist.
- Weitere Aspekte (besondere Schnelligkeit des Rechners, Mehrplatzfähigkeit, Farbkarte bzw. -bildschirm) spielen eine vergleichsweise sekundäre Rolle, können aber im Einzelfall wichtig sein.
- Nicht genug betont werden kann hingegen — zwar durchaus unabhängig von der vorliegenden Problemstellung, aber deshalb für diese nicht weniger bedeutsam — die Wichtigkeit einer *Datensicherungskomponente*. Die regelmäßige Sicherung, z. B. auf einem sogenannten „streamer tape“, kann angesichts der Anfälligkeit von Disketten bzw. Platten im Schadensfall von unschätzbarem Wert sein. Der Verfasser dieses Beitrages weiß, wovon er spricht, zumal seine eigene Festplatte nach einem halben Jahr — ohne Vorwarnung — den Betrieb aufgab.

1.4. Anforderungen an die Datenbank im allgemeinen

Grundsätzliche Fragen, die aufgrund lokaler Gegebenheiten bzw. subjektiver Ansprüche und Präferenzen zu entscheiden sind, bevor man an die Erstellung einer persönlichen Literaturdatei geht, sind:

- Welche subjektive Wertigkeit haben *Schnelligkeit* und *Komfort* bei der Suche (dem sogenannten „retrieval“) sowie bei der Erstellung des Outputs (der Literaturverzeichnisse, Zitatlisten usw.)? Ist man gewillt, sich mit einfa-

cher Software zufrieden zu geben oder wird ein Produkt mit höheren Leistungsmerkmalen erwartet?

- Wird es sich bei der Datenbank um eine nur für den persönlichen Gebrauch erzeugte *Inhouse-Datei* handeln oder soll auch ein *externer Zugriff* möglich sein? Letzteres könnte beispielsweise bei Literatursammlungen von Universitätsinstituten der Fall sein, die aufgrund ihrer fachlichen Spezialisiertheit auch für andere Nutzer von Interesse sind.
- Wer soll(en) der/die *Benutzer* der Datenbank sein? Sollen neben dem individuellen Erzeuger der Datei z. B. auch Kollegen oder Sekretariatskräfte zugreifen können? Handelt es sich bei den Daten allfällig auch um schutzbedürftiges Material?

1.5. Software-Vorentscheidungen

- Kann (z. B. im Haus oder Institut) bereits *vorhandene* Software genutzt werden oder ist geeignete Software erst auszuwählen bzw. anzukaufen? Lohnt sich im ersteren Fall die mögliche Geldersparnis, falls die vorhandene Software lediglich sub-optimal ist?
- Welche *finanziellen Möglichkeiten* bestehen für den Kauf von geeigneter Software? Die Preispalette variiert von wenigen hundert Schilling bis zu Beträgen jenseits von öS 20.000,— oder 30.000,—!

2. Numerische versus bibliographische Daten

Wer beabsichtigt, bibliographische Daten mittels eines Datenbankprogrammes auf dem PC zu verwalten, sollte sich über die wesentlichsten Unterschiede zwischen diesem Datentyp einerseits und numerischen Daten andererseits im klaren sein. Insbesondere muß man sich bei der Softwareauswahl stets vor Augen halten, daß viele Programme primär für numerische Datendokumentationen entwickelt worden sind und bei der Verarbeitung von Literaturinformationen weniger geeignet sein können.

Die wichtigsten derartigen Charakteristika bzw. Unterschiede sind:

2.1. Strukturierungsgrad:

Numerische Daten sind üblicherweise stark strukturiert. Oft besteht der Inhalt jedes Datenfeldes nur aus einer einzigen Maßzahl. Ein Beispiel wären etwa die Kategorien einer Datei mit Angaben zu einzelnen Ländern: z. B.: Fläche in km², Einwohnerzahl, Verkehrsnetz in km, Prozentsatz der Hochschulabsolventen usw. Bibliographische Daten sind hingegen durch einen geringen Grad an Strukturiertheit gekennzeichnet. Als Beispiel halte man sich die möglichen Inhalte des Datenfeldes „Quellenangabe“ vor Augen, die etwa lauten können:

- Productivity in the Information Age: Proc. of the 46th ASIS Annual Meeting, vol. 20. Washington, D. C., Oct 2—6, 1983. Ed. by R. F. Vondran, A. Caputo, C. Wasserman, R. A. V. Diener. White Plains, N. Y.: Knowledge Industry Publ. 1983.

2.2. Häufigkeit von Kategorien:

Im Falle numerischer Dokumentationen tritt ein und dieselbe Kategorie meist nur einmal auf. Für bibliographische Daten ist hingegen das Faktum charakteristisch, daß ein multiples Auftreten bestimmter Kategorien möglich ist. Das wichtigste — aber nicht einzige — Beispiel ist das Datenfeld „Autor“.

2.3. Varianz der Feldlänge:

Numerische Daten verfügen üblicherweise über eine konstante bzw. fixe, das heißt klar vorher definierbare Länge des jeweiligen Datenfeldes (z. B. Zahl der Beschäftigten im Betrieb: maximal 5 Stellen). Für bibliographische Daten ist hingegen typisch, daß die Länge ein und desselben Datenfeldes von Dokumentationseinheit zu Dokumentationseinheit stark variieren kann. Man denke z. B. an die Kategorien „Titel“ oder „Quellenangabe“. Es wird daher von variabler oder flexibler Feldlänge gesprochen.

2.4. Größe der Kategorien:

Während im Falle numerischer Daten die einzelnen Felder überwiegend relativ kurz sind — Zahlen benötigen eben meist nur wenige Stellen — treten bei Literaturdokumentationen wesentlich längere Kategorien auf. Einzelne Felder — z. B. der Text von Abstracts — können durchaus auch überdurchschnittlich lang sein!

3. Beurteilungskriterien für die Software im Hinblick auf das Design einer persönlichen Literaturdatei

3.1. Dateistruktur:

- Maximal erlaubte Größe der Datenbank (beschränkt vs. unbeschränkt bzw. nur vom verfügbaren Speicherplatz abhängig);
- Anzahl der verwaltbaren Datenbanken unter einer Softwareinstallation (manche Produkte beschränken diese Zahl, z. B. auf sieben oder acht Dateien, während andere von einer solchen — durchaus störenden — Einschränkung absehen).

3.2. Dokumentationseinheiten (Datensätze, records):

- Maximal erlaubte Größe einer Dokumentationseinheit (beschränkt vs. unbeschränkt);
- Zahl der pro Datenbank erlaubten Datensätze (beschränkt vs. unbeschränkt; typische Beschränkung z. B. 65.000 records, aber zum Teil auch weit darunter);
- Speicherung der Datensätze mit variabler Länge? Softwareprodukte ohne diese Möglichkeiten benötigen für jede Dokumentationseinheit den maximal definierten Speicherplatz (unabhängig von der faktischen Satzlänge), was Dateien unnötig aufbläht und teures Speichervolumen kostet. Im Extremfall wird von manchen Programmen sogar Speicherplatz blockweise — das heißt für bestimmte Mengenklassen von records — im voraus reserviert, unabhängig davon, ob die entsprechende Fallzahl

von records überhaupt erreicht wird oder nicht.

3.3. Datenfelder (Kategorien):

- Maximal erlaubte Feldlänge (typische Beschränkungen: z. B. 240, 256, 1000 Zeichen);
- Erlaubte Anzahl von Feldern pro Datensatz (mit wieviel Kategorien kann für eine persönliche Literaturdatenbank das Auslangen gefunden werden?);
- Anzahl der indizierbaren, das heißt mit Hilfe von Registern im Schnellzugriff suchbaren Datenfeldern (Alternativen: alle Felder vs. eine beschränkte Anzahl, z. B. sieben);
- Inhaltlich frei definierbare vs. mit Kategoriebezeichnungen vordefinierte Felder (manche Softwareprodukte weisen unveränderbare Feldnamen, z. B. Autor, Titel, Quelle usw. auf und sind daher für andere als Literaturdokumentationszwecke schwer einsetzbar; von einem „Datenbank-Design“ kann in diesen Fällen eigentlich kaum gesprochen werden).

3.4. Änderungen der Dateistruktur:

Da erfahrungsgemäß viele Anwender zuwenig Vorarbeit in ein sorgfältiges Design ihrer persönlichen Datenbank investieren oder ihnen bestimmte Bedürfnisse erst nach einiger Erfahrung mit der erzeugten Datei bewußt werden, kommt den von der Software bereitgestellten Möglichkeiten zur Änderung einer bereits bestehenden Dateistruktur besondere Bedeutung zu. Dabei geht es z. B. um das Entfernen bzw. Hinzufügen einzelner Datenfelder, um Änderungen der Feldlänge oder der Indizierungsvorschriften oder auch um das Aufgliedern von bestimmten Feldern in neue Subfelder bzw. das Überführen einzelner Feldinhalte in andere Kategorien. Die angebotenen Softwareprodukte variieren stark hinsichtlich all dieser Möglichkeiten, aber auch im Komfort bzw. in der Leichtigkeit der Durchführung solcher Änderungen. Bei manchen Produkten

läuft man überdies bei der Änderung der Datenbankstruktur äußerste Gefahr, alle bisher erfaßten Daten unvermittelt zu verlieren!

4. Beurteilungskriterien für die Software im Hinblick auf Datenerfassung und Datenbankaufbau

4.1. Manuelle Eingabe:

- Für die manuelle Erfassung von Dokumentationseinheiten solltn Komfort und Leistungsfähigkeit eines guten Textverarbeitungsprogrammes erwartet werden können. Dies betrifft auch besonders die Korrekturmöglichkeiten während der Dateneingabe.
- Im Falle einer weniger leistungsfähigen Erfassungskomponente sollte das Programm ermöglichen, eine auf Benutzerseite bereits vorhandene, leistungsfähigere Textverarbeitungssoftware mittels einer vordefinierten Schnittstelle „anzuhängen“.
- Im Erfassungsmodul sollte weiters auch eine leichte Duplizierbarkeit bereits bestehender records (z. B. für die Erfassung ähnlicher Dokumentationseinheiten, die sodann nur korrigiert werden müssen) möglich sein.

4.2. Fremddatenübernahme:

Prinzipiell sollte die Software in der Lage sein, Daten, die bereits maschinenlesbar vorliegen, zu übernehmen. Dies könnten etwa Daten aus Online-Recherchen (downloading), aus Beständen von Arbeitskollegen oder auch aus bereits vorhandenen eigenen Dateien sein. Besonderes Augenmerk ist den erforderlichen Austauschformaten für eine derartige Datenübernahme zu schenken. Dabei ist u. a. zu fragen, welchen Aufwand es bedeuten würde, vorhandene Daten in ein solches Eingabeformat zu bringen.

4.3. Datenkorrektur:

Neben der Korrektur beim Erfassungsvorgang ist auch die Möglichkeit der späteren Korrektur von Datensätzen erforderlich. Die entsprechenden Editier-

funktionen sollten im Grundsatz jenen gleichen, die für den Erfassungsvorgang gelten.

4.4. Erfassungshilfen:

Einige Programme bieten hilfreiche Zusatzmodule für den Erfassungsprozeß an. Dies können etwa sein:

- Hilfen zur Rechtschreibung, z. B. in Form eines gespeicherten Wörterbuches mit automatischer Fehlerkorrektur bzw. Korrekturvorschlägen;
- Routinen zur Strukturprüfung, etwa Anzeigen bei Verletzung des Datenbank-Designs (z. B. Eingabe von unzulässigen Zeichen in bestimmte Datenfelder usw.);
- Vorspeicherung von Textbausteinen, die im Erfassungsprozeß auf Tastendruck abrufbar sind (z. B. immer wiederkehrende Datenelemente wie etwa Zeitschriftentitel; dies trägt übrigens auch zur Einheitlichkeit der Zitierweise bei).

4.5. Deutscher Zeichensatz:

Zu fragen ist schließlich auch, wie sich ein konkretes Softwareprodukt bei der Erfassung deutscher Zeichen wie der Umlaute bzw. des „ß“ verhält. Es kann heute erwartet werden, daß moderne Software in der Lage ist, diese Zeichen zu verarbeiten. Dies gilt sinngemäß natürlich auch für Symbole aus anderen Sprachkreisen (z. B. französische oder spanische), soweit es sich um sogenannte ASCII-Zeichen handelt.

5. Beurteilungskriterien für die Software im Hinblick auf Indizierung und Invertierung

Unter den Begriffen „Indizierung“ bzw. „Invertierung“ wird das Erstellen von datenbankinternen Stichwortregistern verstanden, die es ermöglichen, die so aufbereitete Information äußerst rasch (und zwar mehr oder weniger unabhängig von der Zahl der Dokumentationseinheiten) wiederzufinden. In diesem Zusammenhang kann die jeweilige Software nach einer ganzen Reihe von Kriterien bewertet werden:

- Ist die Zahl der indizierbaren Datenfelder unbegrenzt? (vgl auch oben 3.3.)
- Besteht im Falle von Programmen, die die gesamte Information indizieren, auch eine freie Wahlmöglichkeit, dies für bestimmte Felder oder bestimmte Datenelemente zu unterdrücken?
- Besteht die Möglichkeit, feldspezifisch unterschiedliche Indizierungsvereinbarungen zu treffen? Beispiele für verschiedene Varianten wären etwa: (a) automatisch alle Wörter des Datenfeldes; (b) alle Wörter, die manuell zur Indizierung hervorgehoben werden; (c) alle Wörter, die nicht zur Unterdrückung der Indizierung manuell hervorgehoben werden.
- Besteht die Möglichkeit, einen sogenannten „basic index“ aus mehreren Datenfeldern zu definieren? Darunter wird, analog zu professionellen Online-Informationssystemen, die Definition eines feldübergreifenden „Hauptregisters“ verstanden, in welchem vom Programm gesucht wird, sofern keine anderslautende Anweisung gegeben wird.
- Sind die (Sonder-)Zeichen, die vom Programm beim Datenbankaufbau bzw. bei der Indizierung als Kriterien für die Trennung von Wörtern bzw. zur Worterkennung verstanden werden, variabel definierbar?
- Was ist die maximale Länge eines Registerintrages? Entspricht dies den Erfordernissen eines deutschsprachigen (das heißt mit relativ langen Wörtern operierenden) Textdatenretrievals?
- Wieviel Speicherplatz wird für das Register bzw. die Indizes benötigt? Bestimmte Produkte erweisen sich in diesem Punkt als unvermutet platzaufwendig!
- Besteht die Möglichkeit, eine sogenannte „Stoppwortliste“ zu verwalten? Dies ist ein Verzeichnis von Wörtern, die bei der Erstellung der Register routinemäßig zu übergehen sind (z. B. Artikel, Präpositionen usw.).

- Besteht die Möglichkeit, eine „Positiv-Wortliste“ oder „Vergleichswortliste“ zu verwalten? Diese enthält jene Wörter, die bei der Indizierung als einzige zu berücksichtigen sind (z. B. eine Liste erlaubter Schlagwörter im Schlagwortfeld).
 - Ermöglicht das System die Verwaltung eines hierarchischen Thesaurus?
 - Wie verhält sich das System bei der Indizierung von Wörtern mit deutschen bzw. anderssprachigen Zeichen? Diese Frage ist unabhängig von der Datenerfassung solcher Wörter zu sehen!
 - Wird die Indizierung sofort und automatisch bei der Dateneingabe bzw. Datenkorrektur durchgeführt oder sind spezielle, zeitaufwendige Reorganisationsläufe der Datenbank erforderlich? Wieviel Zeit ist grundsätzlich für die Indizierung aufzuwenden?
6. *Beurteilungskriterien für die Software im Hinblick auf die Retrievalfunktionen*
- Gerade hinsichtlich der verfügbaren Suchmöglichkeiten und der sie steuernden Suchkommandos (Retrievalfunktionen bzw. Retrievaloperationen) gibt es beträchtliche Unterschiede zwischen den am Markt befindlichen Softwareprodukten. Der Interessent sollte an ein in Betracht gezogenes Produkt folgende Fragen richten:
- Wird mittels „Menütechnik“ (Vorgabe von Auswahlmenüs für jeden Arbeitsschritt) oder mittels einer Suchsprache (quere language) recherchiert? Wie entsprechen diese Alternativen den subjektiven Bedürfnissen, wie „angenehm“ wird die Abfragetechnik, auch in emotionaler Hinsicht, empfunden?
 - Welche Such- bzw. Antwortzeiten sind zu erwarten? Wie verändern sich diese bei größeren Datenmengen? Nicht selten hört man von minutenlangen (!) Retrievalzeiten, während andere Programme durchaus im Sekundenbereich bleiben.
 - Kann nur feldspezifisch bzw. nur feldübergreifend recherchiert werden oder ist beides gestattet? („basic index“ vs. „field qualification“)
 - Stellt das System die volle Boolesche Logik (OR, AND, NOT) für die Abfrage zur Verfügung? Diese Frage mutet weniger befremdlich an, wenn man weiß, daß auch bekannte Datenbankprogramme mitunter nicht in der Lage sind, ODER-Verknüpfungen durchzuführen.
 - Besteht auch die Möglichkeit, sogenannte „proximity operators“ (Distanzoperatoren) zur Definition eines gewünschten Wortabstandes zwischen Suchbegriffen — z. B. bei der Suche von Mehrwortphrasen — einzusetzen? Im Gegensatz zu professionellen Online-Informationssystemen ist dies bei Mikrocomputerprogrammen kaum zu erwarten, jedoch auch nicht auszuschließen.
 - Welche Möglichkeiten werden für die Recherche komplexer Formulierungen, etwa vom Typ „(A OR B OR C) AND (D OR E) NOT (F OR G)“ geboten? Ist dies überhaupt gestattet oder können immer nur einzelne Begriffe gesucht werden? Ist es im letzteren Fall wenigstens möglich, diese schrittweisen zu verknüpfen (z. B. ein vorhandenes Resultat durch Vereinigung mit dem nächsten zu vergrößern)? Wieweit wird die Recherche von Klammerungen („nesting“) unterstützt? Grundsätzlich kann an dieser Stelle gefragt werden: Erlaubt das System die individuell erforderlichen Recherchen (wenn vielleicht auch mühevoll oder aufwendig) oder sind bestimmte Abfragen prinzipiell nicht durchführbar?
 - Sind flexible Anfangs- bzw. Endmaskierung („truncation“) von Suchbegriffen bzw. fixe Maskierung („character masking“) gestattet? Haben diese Operationen eine gravierende Verschlechterung der Antwortzeiten zur Folge?
 - Können für numerische Datenfelder (z. B. Publikationsjahr) auch soge-

- nannte relationale Operationen („größer als“, „kleiner als“ „von — bis“ usw.) eingesetzt werden?
- Ist es möglich, bei der Suche auf frühere Rechenschritte zurückzugreifen oder wird bei jeder neuen Eingabe der vorhergehende Suchschritt „vergessen“?
 - Kann die sogenannte „Profiltabelle“ — also die Liste aller in einer Sitzung eingegebenen Suchwörter, logischen Verknüpfungen und Ergebnishäufigkeiten — beliebig am Bildschirm angezeigt werden?
 - Können die invertierten Dateien, also die Register der suchbaren Wörter, angezeigt bzw. aufgelistet werden? Werden dabei auch die Häufigkeiten ihres Auftretens bzw. die Nummern der betreffenden Dokumentationseinheiten angezeigt? Ist diese Auflistung nur feldspezifisch, nur global oder nach Wahl möglich?
 - Können bei der Wörterbuch-Anzeige bzw. bei der Suche auch Thesaurus-Relationen einbezogen werden?
 - Können Fragestellungen (Suchprofile) gespeichert und in späteren Durchläufen wiederverwendet werden? Dies kann gerade für persönliche Dateien von Interesse sein, da in diesen oft immer wieder dieselbe Fragestellung recherchiert wird.
 - Besteht die Option, bei der Recherche zwischen Groß- und Kleinbuchstaben zu unterscheiden? Dies wird in der Regel eher nicht gegeben sein, könnte jedoch im Einzelfall ein Anforderungskriterium darstellen.
 - Schließlich: Wie sind deutsche und fremdsprachige Spezialzeichen zu recherchieren? Kann bei der Recherche beispielsweise zwischen den Autoren „Müller“ und „Muller“ überhaupt unterschieden werden? Sind bestimmte Zeichen (z. B. „ø“, „ñ“ usw.) — wenn überhaupt — nur durch komplizierte

Tastenoperationen recherchierbar? Bestehen Diskrepanzen zwischen Input- und Recherchemöglichkeiten (z. B.: deutsche Umlaute zwar eingebbar, aber nicht recherchierbar)?

7. Beurteilungskriterien für die Software im Hinblick auf die Ausgabefunktionen

Literaturdaten auszugeben bedeutet einerseits, die gespeicherten bzw. recherchierten Informationen nach bestimmten Kriterien (z. B. Alphabet der Verfasser) zu sortieren, und andererseits, sie in bestimmter Form (z. B. gemäß den Anforderungen für eine Literaturliste in einem Zeitschriftenaufsatz) auf einem Ausgabemedium darzustellen.

7.1. Sortierfunktionen:

- Kann nach mehr als nur einem Datenfeld sortiert werden? (Beispiel: innerhalb desselben Autorennamens nach dem Publikationsjahr und — sofern immer noch mehrere Zitate vorliegen — nach dem Sachtitel).
- Besteht die Wahlmöglichkeit zwischen aufsteigender (z. B. Berger, Börner, Brunner, Burger) und absteigender (z. B. 1987, 1986, 1985, 1984) Sortierfolge?
- Ist die Gefahr feldinterner Konflikte beim Sortieren gegeben? (Beispiel: ein Programm, das stets nach dem im Alphabet erstgereihten Wort eines Datenfeldes sortiert, kann im Autorenfeld zwischen Familien- und Vornamen nicht unterscheiden und würde „Maier, Franz“ unter „F“ sortieren).
- Sind die Regeln für die Sortierfolge der einzelnen Zeichen frei vom Benutzer definierbar oder vom Programm irreversibel vorgegeben? Ein Programm, das beispielsweise nicht gestattet, Groß- und Kleinbuchstaben gleichwertig, das heißt „durcheinander“ zu sortieren, ist für bibliographische Zwecke ungeeignet!
- Wie verhält sich die Software bei der Sortierung von Umlauten usw.?

7.2. Anzeige- und Druckfunktionen

- Ist es möglich, Ergebnislisten sowohl über den Bildschirm bzw. den Drucker auszugeben als auch — z. B. zum „Einbau“ in eine wissenschaftliche Arbeit — auf eine Textdatei zu schreiben?
- Gewährleistet das Programm — auch eventuell durch entsprechende Definitionsmöglichkeiten — eine „druckreife“ Ausgabe, oder ist es stets erforderlich, die Ergebnisliste mittels eines Textverarbeitungsprogrammes entsprechend zu „polieren“, um Druckreife zu erlangen?
- Können gewünschte Ausgabeformate frei und flexibel definiert werden, sind bestimmte Standardformate vorgegeben, oder kann etwa nur im Eingabeformat ausgegeben werden? Wünschenswert wäre sowohl die Auswahl der benötigten Datenfelder als auch die freie Entscheidbarkeit darüber, wie die darin enthaltene Information in der Ergebnisliste angeordnet sein soll (z. B.: Titel an erster Stelle, Autor danach in neuer Zeile usw.).

8. Typen von Software für persönliche Literaturdateien

Abschließend sei kurz auf die Arten von PC-Datenbanksoftware eingegangen, die dem Interessenten für die Mechanisierung einer persönlichen Literaturdatei auf dem Markt angeboten werden.

8.1. File management programs

Unter dieser Bezeichnung werden relativ einfache und meist auch preisgünstige Programme angeboten, die man auch als „elektronische Karteikästen“ bezeichnet, da sie im wesentlichen bestrebt sind, herkömmliche Karteien am PC nachzubilden („Ausfüllen und Suchen elektronischer Karteiblätter“). Ihre Verwendung ist meist unkompliziert und leicht erlernbar. Die gebotenen Funktionen sind allerdings begrenzt, auch mit allzu sensationellen Suchzeiten darf nicht gerechnet werden. Als Beispiel sei das Programm FILING ASSISTANT

(von IBM) genannt, das — sofern nicht allzu hohe Ansprüche vorliegen — überraschend leistungsfähig ist.

8.2. Database management programs

Unter Datenbankprogrammen im engeren Sinn versteht man komplexere Ausformungen des oben genannten Typs. Programme dieser Art bieten in der Regel z. B. Boolesche Operatoren, Sortierung nach mehr als nur einem Datenfeld, Aufbau und gleichzeitige Verwendung mehrerer Indizes (invertierter Register), rasche Suche indizierter Information usw. Ihre Leistungsfähigkeit ist oft sehr groß und zum Teil auch durch zusätzliche Komponenten wie etwa eine eingebaute Programmiersprache weiter multiplizierbar. Nachteilig ist, daß diese Softwareprodukte meist für numerische Daten gedacht sind (optimale Eignung für strukturierte Daten) und gewisse, oft auch gravierende Probleme bei der Verarbeitung von bibliographischer Information aufweisen. Das bekannteste Beispiel dieses Typs ist das Programm dBase II/III (von Ashton Tate).

8.3. Indexing programs

Dies sind ebenfalls allgemeine Datenbankprogramme, die sich von den oben genannten durch besonders ausgebaute und für die Verwaltung bibliographischer Daten wichtige Registerfunktionen auszeichnen. Während z. B. mit dBase auch bei Indizierung des Datenfeldes „Titel“ nur der volle Wortlaut recherchiert werden kann (das heißt es muß der Wortlaut des Titels bekannt sein und der Beginn des Titels eingegeben werden), ist es für indexing programs selbstverständlich, jedes einzelne Wort eines Datenfeldes zu indizieren und damit unabhängig von der Wortfolge suchbar zu machen. Als Beispiele seien die Programme CARDBOX PLUS (von Business Simulations Ltd.) und MICRO-CAIRS (von Leatherhead Food RA) genannt.

8.4. Bibliographic programs

Unter dieser Bezeichnung ist Datenbanksoftware zu verstehen, die aus-

drücklich oder ausschließlich mit Zielrichtung auf die Speicherung und Suche bibliographischer Daten erstellt wurde. Meist zeichnen sich Produkte dieser Art durch verschiedene Spezialfunktionen aus, die auch im Rahmen dieses Beitrags genannt wurden, verfügen über eine gute Registertechnik und sind besonders flexibel in der Gestaltung des Outputs. Als Nachteil mag gewertet werden, daß sie für die Verwaltung von anderen als Literaturdaten oft weniger geeignet sind, was ihre Einsatzbereiche beschränkt. Dies kommt oft in der Präsenz einer fix vorgegebenen Datenbankstruktur mit nicht veränderbaren Feldnamen zum Ausdruck. Der Einsatzzweck manifestiert sich oft schon in den Programmbezeichnungen, z. B. LIDOS (Literatur-Dokumentationssystem; von Express Edition GmbH) oder REFSYS (von Elsevier-Biosoft).

Als Beispiel für ein interessantes nicht-kommerzielles Programm sei in diesem Zusammenhang MASULIST (Maschinelle Suche von Literaturstellen) genannt, ein vom Bibliotheks- und Informationssystem der Universität Oldenburg in Zusammenarbeit mit dem dortigen universitären Rechenzentrum erstelltes bibliographisches Datenbankprogramm. Es wird an externe Interessenten für eine Schutzgebühr von DM 50,— (inkl. Handbuch) abgegeben und kann trotz einiger Restriktionen als erstaunlich leistungsfähig (und überaus preisgünstig) bezeichnet werden. Einer Universitätsbibliothek, die in der Lage ist, ihren hilfeheischenden Benutzern ein solches Freeware-Paket anzubieten, kann zu diesem Engagement durchaus gratuliert werden.

Dr. Otto Oberhauser
