

371  
25. 8. 1965

## Ein Verfahren zur automatischen Klassifizierung für Veröffentlichungswesen, Bibliothekswesen und Dokumentation \*)

VON MARTIN SCHEELE, Schlitz/Hessen

### I. Einleitung

#### 1. Aufgabenstellung

Die Aufgabe der Dokumentation ist die Bewältigung des immer schneller anwachsenden Schrifttums.

\*) Dem Andenken meines am 6. März 1964 verstorbenen väterlichen Freundes, Dipl.-Ing. WILLI HEIMERDINGER, in Dankbarkeit gewidmet.

Dabei stehen die wissenschaftlichen Zeitschriften im Vordergrund des Interesses. Die Größe der Aufgabe und der Umfang des Literaturmaterials zwingen uns zu einer Rationalisierung durch den Einsatz von modernen Methoden. Das Dokumentationsproblem selbst ist vielschichtig und kann daher weder in einem Arbeitsgang noch auf einer Ebene gelöst werden. Wir

müssen vielmehr verschiedene Stufen der Dokumentation unterscheiden, die bei der reinen Titelbearbeitung der Veröffentlichungen beginnen und über Schlagwortgebung und Referat bis zur Auswertung der konkreten Daten und Werte reichen. Vor dem Einsatz moderner Verfahren hat man sich daher die Frage vorzulegen, auf welcher Stufe der Dokumentation der größte Rationalisierungseffekt zu erzielen ist und welche spezielle Methode das günstigste Verhältnis von Aufwand und Nutzen verspricht.

Die gründliche Untersuchung dieser Frage führte zu dem Ergebnis, daß man mit der Automatisierung der Dokumentation am besten auf der ersten, bibliographischen Stufe der reinen Titelbearbeitung ansetzt und sich dabei verfahrensmäßig am rationellsten der Lochstreifentechnik in Kombination mit elektronischen Rechenanlagen bedient.

Die Darstellung eines entsprechenden Verfahrens ist Gegenstand dieses Aufsatzes. Dabei kommt es mir darauf an, den Leser mit den grundsätzlichen Gedankengängen und Arbeitsschritten unserer Methode vertraut zu machen. Einzelheiten werden nur dann behandelt, wenn sie zum Verständnis des Ganzen unentbehrlich sind. Im übrigen habe ich mich bemüht, alle speziellen Dinge und alle Komplizierungen beiseite zu lassen, zumal dies schon aus Platzgründen notwendig ist.

Das geschilderte Verfahren wurde an der Dokumentationsstelle für Biologie in Schlitz/Hessen entwickelt. Die Planung und Durchführung der Programmierarbeiten für die elektronischen Rechenmaschinen lag in Händen von Herrn GERHARDT NATALIS. Wenn ich ihm dafür an dieser Stelle sehr herzlich danke, so muß ich noch betonen, daß seine Leistung wohl nur von einem Programmierfachmann in vollem Umfang gewürdigt werden kann. Die zur Erprobung des Verfahrens benötigten elektronischen Maschinen stellte die IBM Deutschland in außerordentlich großzügiger Weise zur Verfügung, wofür wir der Firma und allen beteiligten Herren unseren besonderen Dank aussprechen. Für die Durchführung der Lochkartenarbeiten danken wir den Damen und Herren der Hollerith-Abteilung der Max-Planck-Gesellschaft, deren verstorbenem Leiter, Dipl.-Ing. WILLI HEIMERDINGER, diese Arbeit gewidmet ist. Vor allem gilt aber unser Dank den beiden Persönlichkeiten und Institutionen, die unsere gesamte Arbeit überhaupt erst ermöglicht haben: Herrn Professor Dr. REINHOLD VON SENGBUSCH und dem Max-Planck-Institut für Kulturpflanzenzüchtung, an das unsere Dokumentationsstelle angeschlossen ist, sowie Herrn Dr. MARTIN CREMER und dem Institut für Dokumentationswesen, das unsere Arbeiten seit dem 1. Januar 1962 finanziert hat.

## 2. Definition der verwendeten Fachausdrücke<sup>1)</sup>

**Begriffsprinzip.** Im Rahmen der Methoden für das Wiederauffinden von Informationen (information retrieval) unterscheidet man zwei Prinzipien. Bei dem Dokument-Prinzip wird für jedes einzelne Dokument ein Datenträger (Karteikarte, Lochkarte, Magnetbandabschnitt) benutzt, auf dem man jeweils alle Begriffe fixiert, die in dem betreffenden Dokument vorkommen. Bei dem Begriffs-Prinzip ist es umgekehrt: Hier legt man für jeden einzelnen Begriff einen Datenträger an und fixiert darauf die Nummern aller Dokumente, in denen der betreffende Begriff enthalten ist.

**Current Contents.** Ein in den Vereinigten Staaten von Amerika herausgegebener Literaturdienst, der für bestimmte Gruppen von Zeitschriften bereits vor dem Erscheinen der laufenden Hefte nur deren Inhaltsverzeichnisse mit den dort aufgeführten Verfassern

<sup>1)</sup> Termini, die als allgemein bekannt gelten dürfen, sowie alle Begriffe, die im Textzusammenhang erläutert sind, werden hier nicht definiert.

und Aufsatztiteln sowie die Anschriften der Autoren bringt.

**KWIC-Index.** Ein von H. P. LUHN (USA) entwickeltes Verfahren, das in vollem Wortlaut Keyword-In-Context Indexing (oder Index) heißt und die Titel von Veröffentlichungen auf maschinell Wege nach allen in ihnen enthaltenen wesentlichen Wörtern (Keywords) alphabetisch sortiert und im Titelzusammenhang (Context) auflistet.

**Maschinenlochkarten.** Lochkarten, deren Bearbeitung maschinell erfolgt. Die Karten werden von den Maschinen einzeln (nacheinander) bearbeitet. Sie sind vor ihrer Verwendung ungelocht.

**Nadellochkarten.** Lochkarten, deren Bearbeitung vorwiegend manuell mit Hilfe von Nadeln oder Nadelgeräten erfolgt. Eine Vielzahl von Karten kann gleichzeitig bearbeitet werden. Sie tragen bereits vor ihrer Verwendung Löcher.

**Notation.** Jede Art von Bezeichnung oder Symbolisierung, bei der an die Stelle der ursprünglichen Wörter irgendwelche andere Schrift-Symbole treten.

**Schlagwort.** Der kürzestmögliche sprachliche Ausdruck für den Sachinhalt einer Veröffentlichung oder den Sachinhalt eines ihrer Teile. Das Schlagwort bezieht sich also grundsätzlich auf den Inhalt einer Arbeit. Dadurch unterscheidet es sich vom Stichwort, das aus dem Titel der betreffenden Veröffentlichung entnommen wird.

**Science Citation Index.** Ein in den Vereinigten Staaten von Amerika entwickeltes Verfahren, bei dem die Literaturangaben der Publikationen in den Mittelpunkt der Dokumentationstätigkeit gestellt werden. Von einer bestimmten Arbeit und ihrem Verfasser ausgehend, läßt sich damit vor allem feststellen, welche späteren Autoren diese Arbeit in ihren Veröffentlichungen zitiert haben.

**Thesaurus.** Jedes Wörterverzeichnis, in dem die Wörter nach den Verwandtschaftsbeziehungen ihrer Bedeutung zusammengestellt sind.

## 3. Begründung der Verfahrensweise

### Warum nur Titel?

Die Beschränkung unseres Verfahrens auf die Titel der Veröffentlichungen ist ein Hauptangriffspunkt der Kritik. Einer solchen Kritik kann man jedoch verschiedene wichtige Argumente entgegenhalten, die unser Vorgehen rechtfertigen.

Zunächst sind die Titel im Durchschnitt besser, als ein weit verbreitetes Vorurteil annimmt. In den Vereinigten Staaten von Amerika wurde eine genaue Untersuchung biologischer Aufsätze durchgeführt, wonach zwischen 50 und 70% der Titel den wesentlichsten Inhalt der Arbeiten in thematischer Form wiedergeben.

Daß auch den Benutzern mit den reinen Titeln bereits in erheblichem Maße gedient ist, beweist die schnelle Ausbreitung und Beliebtheit des amerikanischen KWIC-Indexing-Verfahrens und des „Current Contents“-Dienstes. Beide Projekte basieren ausschließlich auf der Auswertung von Titeln und haben bemerkenswerterweise bereits zu einer Verbesserung der Titelgebung durch die Autoren beigetragen.

Betrachten wir nun vergleichsweise den notwendigen Aufwand: Unter Berücksichtigung unserer weitgehenden Automatisierung wäre gegenüber der reinen

Titelbearbeitung für eine Inhaltsauswertung durch Schlagwörter oder Kurzreferate ein etwa 10facher Personalstab und ein etwa 20facher personeller Geldaufwand nötig. Dieser erhöhte Geldaufwand ist dadurch bedingt, daß man eine Inhaltsauswertung mittels Schlagwörtern oder Kurzreferaten weder automatisch vornehmen noch durch Hilfskräfte ausführen lassen kann. Hierfür sind Fachleute erforderlich, die entsprechend mehr Geld kosten.

Außerdem bringt der Einsatz von Fachkräften auch noch andere Probleme mit sich. Eines dieser Probleme ist der allgemein bekannte Personalmangel, der sich in der Dokumentation verstärkt auswirkt, weil die in Frage kommenden Wissenschaftler zunächst einmal in die Forschung und andere Berufszweige gehen. — Weiterhin ist es sehr schwierig, eine größere Zahl von Referenten so zu koordinieren, daß eine möglichst einheitliche Arbeitsweise erreicht wird.

Schließlich kommt ein technisch-organisatorischer Umstand hinzu, der die Beschränkung auf die Titel ganz besonders rationell erscheinen läßt: Bei jeder Dokumentationsmethode, ganz gleich ob es sich um einfache Zettelkarteien, um Nadellochkarten, Maschinenlochkarten oder elektronische Anlagen handelt, ist es notwendig, bestimmte Angaben über die einzelnen Veröffentlichungen durch Abschreiben aus der Originalliteratur zu übernehmen. Nur so ist der Aufbau eines Dokumentationsdienstes möglich. Dabei muß allen Dokumentationsstufen von der reinen Titelbearbeitung bis zur Auswertung der Daten und Werte eine gemeinsame Mindestauswahl solcher Angaben zugrunde gelegt werden. Zu diesen wichtigen Angaben der Zeitschriftenaufsätze, die hier in erster Linie zu berücksichtigen sind, rechnet man Verfasser, Titel, Zeitschrift, Band, Erscheinungsjahr und Seiten sowie gegebenenfalls einige zusätzliche Hinweise über die Anzahl der Literaturzitate, der Abbildungen und Tabellen sowie dergleichen mehr. Wenn man nun alle diese Angaben, zu denen also auch die Titel gehören, für jede Veröffentlichung sowieso abschreiben muß, bietet sich als günstigste Lösung ein Verfahren an, das ausschließlich mit diesen Angaben arbeitet und auf dem Vorgang des einfachen Abschreibens aus dem Originalmaterial weiter aufbaut. — Zugleich kann man jedoch dafür sorgen, daß der Weg zu den höheren Dokumentationsstufen offen bleibt und daß bei späteren zusätzlichen Wünschen keine Änderung der Methoden erforderlich wird.

#### *Warum Lochstreifenschreibmaschinen?*

Die Lochstreifentechnik liefert uns ein Verfahren, das es gestattet, einen sowieso zu schreibenden Text bei dem Schreibvorgang selbst gleich für beliebige Weiterverarbeitungsarten zu speichern. Dies geschieht mit Hilfe von Lochstreifenschreibmaschinen. Sie arbeiten wie eine normale elektrische Schreibmaschine und lochen gleichzeitig alle geschriebenen Zeichen nach einem besonderen Code in einen Papierstreifen ein. Dieser Lochstreifen stellt den Speicher dar. Alle in ihm gelochten Angaben sind jederzeit maschinell reproduzierbar und lassen sich entweder vermittels einer Leseeinrichtung der Lochstreifenschreibmaschine beliebig oft auf normale Karteikarten oder auf Nadellochkarten schreiben oder durch den Einsatz von besonderen Umwandlungsmaschinen auf

Maschinenlochkarten oder auf Magnetbänder übertragen. Jeder gewünschten weiterverarbeitenden Dokumentationsmethode kann also Rechnung getragen werden. Die Lochstreifenschreibmaschine gibt uns somit die Möglichkeit, den gesamten Aufnahmevorgang von Hilfskräften durchführen zu lassen, die alle oben genannten Angaben einfach nur abzuschreiben brauchen. Durch eine rationelle Arbeitsorganisation erzielt man dabei sehr gute Leistungen: Eine Gruppe von 4 Schreibkräften kann nach unseren Erfahrungen unter Berücksichtigung von Urlaub und sonstigen im Rahmen des Normalen bleibenden Ausfällen mit Hilfe einer Lochstreifenschreibmaschine bei mehrschichtigem Betrieb im Jahr durchschnittlich 100000 Literatureinheiten erfassen und korrigieren.

#### *Warum elektronische Rechenanlagen?*

Die elektronischen Rechenanlagen sind das am weitesten entwickelte technische Verfahren, das bisher für Dokumentationszwecke verwendet worden ist. Nur sie geben uns die Möglichkeit, die gesamte bibliographische Arbeit, soweit sie ausschließlich auf der Auswertung der aufgenommenen Titel und der übrigen genannten Angaben beruht, vollständig zu automatisieren und das gespeicherte Material in jeder Hinsicht gründlich auszunutzen.

Die schwierigste Teilaufgabe im Rahmen einer solchen vollständigen Automation ist die Entwicklung einer automatischen Klassifizierung der Titel der Veröffentlichungen. Nach Lösung dieses Problems sind die elektronischen Maschinen jedoch in der Lage, dem Menschen das Zusammenstellen von Titelbibliographien und das Durchsuchen von Titeln auf spezielle Themen weitgehend abzunehmen.

Welcher Aufwand dadurch gespart werden kann, ist wohl nur schwer abzuschätzen, denn bisher wird in dieser Hinsicht eine unüberblickbare Doppel- und Mehrfacharbeit geleistet: Unabhängig voneinander sehen ungezählte Bibliothekare, Dokumentare und Wissenschaftler dasselbe Literaturmaterial nach verschiedenen Themen durch, um entsprechende Bibliographien zu schaffen oder spezielle Arbeiten herauszufinden. Wir erfassen dagegen nur *einmal* die *gesamten* Originalveröffentlichungen aller einschlägigen Zeitschriften und lassen alle übrigen Arbeiten von den Maschinen automatisch erledigen. Dabei ist noch zu bedenken, daß auch das Herausschreiben der gewünschten Angaben selbsttätig, mit hoher Geschwindigkeit und völlig fehlerfrei erfolgt sowie beliebig oft wiederholbar ist, so daß sich auch in dieser Beziehung Einsparungen ergeben.

Außerdem ermöglichen die modernen Anlagen die Lösung einer Reihe von zusätzlichen Aufgaben, von denen hier nur einige genannt seien: Herstellung von Jahresverzeichnissen und Mehrjahresverzeichnissen der einzelnen Zeitschriften. Aufbau von alphabetischen und systematischen Registern. Herstellung von Katalogen für Bibliotheken, Institute und Einzelwissenschaftler. Zusammenstellung von Personalbibliographien. Aufbau von systematischen Wörterbüchern (Thesauren). Allgemeine Literaturforschung mit allen Möglichkeiten der Erfassung von Forschungslücken, der Charakterisierung von Zeitschriften und der Untersuchung der Entwicklung des wissenschaftlichen Veröffentlichungswesens überhaupt.

## II. Das Klassifikationssystem

Im Mittelpunkt unseres ganzen Verfahrens steht die automatische Klassifizierung und als Voraussetzung dafür ein geeignetes Klassifikationssystem. An ihm wurde bereits seit 1952 gearbeitet.

### 1. Allgemeine Systemtypen

Es gibt 2 Haupttypen von Ordnungssystemen: Hierarchische Systeme und Systeme aus frei kombinierbaren Grundbegriffen.

a) Hierarchische Systeme sind streng in eine Stufenfolge von Kategorien gegliedert. Sie beginnen mit einem sehr allgemeinen und enden mit einem sehr speziellen Begriff. Ihre logisch einwandfreie Anwendung liegt nur dann vor, wenn eine bestimmte Menge von Objekten nach einem durchgehenden und einheitlichen Gesichtspunkt (Einteilungsgrund) gegliedert wird. Ein Beispiel dafür ist das Phylogenetische System der Organismen. Als Einteilungsgrund dient dabei die natürliche Verwandtschaft.

b) Systeme aus frei kombinierbaren Grundbegriffen verzichten auf eine hierarchische Ordnung. Sie gründen sich auf die Erkenntnis, daß man bei der Gliederung besonders von Eigenschaftsbegriffen die Kategoriestufen vertauschen kann, so daß jeder Begriff beliebig als Oberbegriff oder Unterbegriff verwendbar ist. Als Beispiel seien die Begriffe „Blut“ und „Temperatur“ genannt. „Blut“ kann Oberbegriff für alle seine Eigenschaften, wie etwa „Temperatur“, „Farbe“ und „Viskosität“ sein. Ebenso gut kann aber „Temperatur“ als Oberbegriff für sein Vorkommen in Gestalt von „Bluttemperatur“, „Körpertemperatur“, „Wassertemperatur“, „Bodentemperatur“ und dergleichen dienen. — Systeme aus frei kombinierbaren Grundbegriffen haben neuerdings in der Dokumentation verschiedentlich Anwendung gefunden, um den zahlreichen Verknüpfungen zwischen den Begriffen möglichst weitgehend Rechnung zu tragen.

c) Beide Systeme weisen ihre speziellen Vor- und Nachteile auf. — Der Nachteil der hierarchischen Systeme besteht vor allem in ihrer Unbeweglichkeit und bei Anwendung auch auf Eigenschaften in dem Zwang, gleiche Begriffe an verschiedenen Stellen des Systems wiederkehren zu lassen. Beispielsweise kommt der Begriff „Temperatur“ in der internationalen Universal-Dezimalklassifikation (UDC) 44mal vor. — Der Nachteil der frei kombinierbaren Begriffe liegt vor allem in dem Fehlen von Oberbegriffen, da ein Mindestmaß an Hierarchie für die Bildung von Begriffsgruppen nicht entbehrt werden kann.

### 2. Eignes System

In unserem System der Naturwissenschaften, bei dem der Schwerpunkt auf dem Gebiete der Biologie liegt, haben wir folgenden Weg beschritten:

a) Objekte und Eigenschaften sind streng getrennt. Als Objekte sind die chemischen Verbindungen und die Organismen festgelegt. Alle anderen Begriffe gelten als Eigenschaften.

b) Für die Objekte werden rein hierarchische Systeme verwendet. Für die Eigenschaften entwickelten wir ein gemischtes System, das die Vorteile der beiden Ordnungsprinzipien (Hierarchie und freie Grundbegriffe) vereinigt und ihre Nachteile vermeidet.

c) Einteilungsgrund für die chemischen Verbindungen ist die Art und Anzahl der in ihnen vorkommenden Atome. Anders ausgedrückt, sind die Verbindungen also nach ihrer Bruttoformel festgelegt, die zugleich als Notation benutzt wird. Ihr Aufbau ist folgender: Am Anfang steht als einheitliches Symbol für alle chemischen Verbindungen ZO. Diese Zeichen wurden gewählt, weil Z sonst am Anfang keiner anderen Notation vorkommt und die Ziffer 0 diesen Buchstaben von den anschließenden Elementsymbolen abtrennt. Darauf folgt die Bruttoformel, bei der die Indizes als normale Ziffern erscheinen. Außerdem werden die nur einmal vorhandenen Atome mit der Ziffer 1 versehen. Beispielsweise bekommt das Natriumsulfat ( $\text{Na}_2\text{SO}_4$ ) die Notation ZONA2S1O4.

Eine solche Regelung ist notwendig, um Fehler bei der maschinellen Bearbeitung und Ausgabe zu vermeiden. Die Maschine erhält die Anweisung, jeweils ein oder zwei Buchstaben, die zwischen Ziffern stehen, als Elementsymbole anzusehen, während die Ziffern selbst die Anzahl der betreffenden Atome wiedergeben. Da bei diesen Notationen nur Großschreibung möglich ist, könnte die Maschine ohne die Verwendung der Ziffer 1 beispielsweise das Symbol für Zinn (Sn) nicht von den Symbolen für Schwefel (S) und Stickstoff (N) unterscheiden.

d) Einteilungsgrund für die Organismen ist ihre phylogenetische Verwandtschaft. Das auf diesem Einteilungsgrund basierende Natürliche oder Phylogenetische System der Organismen hat gegenüber dem System der chemischen Verbindungen zwei erhebliche Nachteile. — Erstens gibt es kein unveränderliches und allgemein anerkanntes System, weil die phylogenetische Verwandtschaft nicht direkt und eindeutig festgestellt werden kann, wie dies bei der Struktur der chemischen Verbindungen im allgemeinen der Fall ist. Die natürliche Verwandtschaft muß vielmehr indirekt durch die biologische Forschung erschlossen werden. Daraus resultieren zahlreiche unterschiedliche Meinungen über die richtige Einordnung der Organismen, und außerdem wandelt sich das System mit dem Fortschritt der Wissenschaft. — Zweitens gibt es für die Organismen kein der chemischen Formelsprache auch nur annähernd gleichwertiges Notationssystem, das die verwandtschaftlichen Verhältnisse durch alle Kategoriestufen hindurch widerspiegelt.

Beide Nachteile mußten wir beheben, denn für die Dokumentation und besonders für unser Verfahren sind sowohl ein unveränderliches Ordnungssystem als auch eine geeignete Notationsgebung erforderlich.

Deshalb wurde ein Ordnungssystem festgelegt, das sich zwar an die bekanntesten und ausführlichsten systematischen Werke und Übersichten anlehnt, aber nur die gebräuchlichsten Kategoriestufen berücksichtigt. Dazu entwickelten wir ein aus Ziffern und Buchstaben gemischtes Notationssystem, dessen Notationen den gesamten verwandtschaftlichen Zusammenhang der einzelnen Organismengruppen wiedergeben (S.5 oben).

e) Das Eigenschaftssystem enthält 4 Systemgruppen zu je 10 Einzelsystemen. Insgesamt sind also 40 Einzelsysteme vorhanden. Das zugehörige Notationssystem besteht aus 4stelligen Zahlennotationen. Die erste Ziffer gibt die Systemgruppe und die zweite Ziffer gibt das Einzelsystem an. Die dritte und vierte Ziffer sind für die einzelnen Grundbegriffe bestimmt.

Kategorie	Stellenwert der Notation	Zeichen
Abteilung	1. Stelle	Buchstabe
Klasse	2. Stelle	Ziffer
Ordnung	3. Stelle	Buchstabe
Familie	4. und 5. Stelle	Ziffern
Gattung	6. und 7. Stelle	Ziffern
Art	8. und 9. Stelle	Ziffern

Beispiel: Haushund (*Canis familiaris*)  
 Notation R9H010102  
 Chordata: R  
 Mammalia: R9  
 Carnivora: R9H  
 Canidae: R9H01  
 Canis: R9H0101  
 Haushund: R9H010102.

Es liegt somit eine 4stufige Gliederung vor. Dies ist das für ein brauchbares System bei uns erprobte und bewährte Mindestmaß an Hierarchie.

Jeder Grundbegriff kommt in dem System nur einmal vor. Alle Grundbegriffe lassen sich frei miteinander und mit den Begriffen der Objektsysteme kombinieren. Infolge der vierstufigen Hierarchie werden bei jeder Kombination der Grundbegriffe die zugehörigen Oberbegriffe automatisch mit verknüpft. Dadurch erreicht man ein Optimum an Informationsvermittlung.

Nachfolgend gebe ich einen Überblick über die Systemgruppen und Einzelsysteme des Gesamtsystems:

0 Hauptkategorien

- |                                  |                                    |
|----------------------------------|------------------------------------|
| 00 Semaphoronten                 | 06 Organisationsstufen             |
| 01 Gesamtwissenschaftsgliederung | 07 Zellbestandteile                |
| 02 Methoden                      | 08 Gliederung des Gesamtorganismus |
| 03 Zeitalter                     | 09 Verschiedene Leistungsbegriffe  |
| 04 Allgemeine Oberbegriffe       |                                    |
| 05 Wissenschaftseinzelbereiche   |                                    |

1 Grundwissenschaftskategorien

- |                             |                                   |
|-----------------------------|-----------------------------------|
| 10 Allgemeine Grundbegriffe | 17 Hauptgruppen im Periodensystem |
| 11 Größen                   | 18 Nebengruppen im Periodensystem |
| 12 Rhythmen                 | 19 Organisch-chemische Faktoren   |
| 13 Medium und Zustand       |                                   |
| 14 Mechanische Faktoren     |                                   |
| 15 Energetische Faktoren    |                                   |
| 16 Chemische Faktoren       |                                   |

2 Einzelkategorien der Idiobiologie

- |                               |                         |
|-------------------------------|-------------------------|
| 20 Nervensystem               | 25 Stützsystem          |
| 21 Verdauungssystem           | 26 Bewegungssystem      |
| 22 Urogenitalsystem           | 27 Sinnessystem         |
| 23 Angriffs- und Schutzsystem | 28 Atmungssystem        |
| 24 Kreislaufsystem            | 29 Verschiedene Systeme |

3 Einzelkategorien der Coenobiologie

- |                                     |   |
|-------------------------------------|---|
| 30 Biosen: Soziale Beziehungen      | 35 Geographie: Erdplanet als Ganzes       |
| 31 Biosen: Stoffliche Beziehungen   | 36 Geographie: Europa                     |
| 32 Coenosen: Land (Lithosphäre)     | 37 Geographie: Amerika, Antarktis, Arktis |
| 33 Coenosen: Luftraum (Atmosphäre)  | 38 Geographie: Asien, Afrika              |
| 34 Coenosen: Gewässer (Hydrosphäre) | 39 Substrat                               |

Das Gesamtsystem ist besonders auf die Biologie im weitesten Sinne und die Wissenschaften von der Erde zugeschnitten. Es enthält jedoch auch die Grundbegriffe aus Physik und Chemie und läßt sich im übrigen gemäß seinen allgemeinen Prinzipien beliebig erweitern und auf andere Gebiete ausdehnen.

In der *Systemgruppe 0* sind alle Grundbegriffe enthalten, die das Prinzip der 4stufigen Hierarchie sprengen müßten, wenn sie in einem der übrigen Einzel-

systeme erscheinen würden. Deshalb sind diese Begriffe zu gesonderten Systemen zusammengefaßt. Es handelt sich also um höhere Oberbegriffe oder um solche Grundbegriffe, die eine sehr allgemeine Bedeutung besitzen. Als Beispiele greifen wir die Einzelsysteme „Semaphoronten“, „Methoden“ und „Zellbestandteile“ heraus.

Unter dem Oberbegriff „Semaphoronten“ (Merkmalträger) (nach HENNIG 1950) sind die Entwicklungsstadien, Geschlechter und Anomalien sowie die systematischen Einheiten unterhalb der Art zusammengefaßt. Die entsprechenden Grundbegriffe dieses Systems kann man sowohl als Weitergliederung des Objektsystems der Organismen als auch im Sinne von Oberbegriffen für sämtliche biologischen Einzelsysteme der übrigen Systemgruppen auffassen. Das System der Semaphoronten nimmt also eine Zwischenstellung zwischen dem Objektsystem und dem Eigenschaftssystem ein.

Bei dem Einzelsystem der Methoden wird besonders deutlich, daß hier ein allgemeiner, eigener Gesichtspunkt vorliegt, der mit jedem anderen Grundbegriff der übrigen Einzelsysteme sinnvoll verknüpft werden kann.

Die Zellbestandteile bilden schließlich Beispiele von Grundbegriffen der Biologie, die wegen ihres allgemeinen Vorkommens bei allen Einzelsystemen der Gruppe 2 wiederkehren müßten, wenn dafür nicht im Rahmen der Hauptkategorien ein gesondertes Einzelsystem geschaffen worden wäre.

In der *Systemgruppe 1* sind vor allem die Grundbegriffe der Physik und Chemie zusammengefaßt.

In der *Systemgruppe 2* finden sich die einzelnen Organsysteme. An den zugehörigen Begriffen dieser Systemgruppe lassen sich die Vorteile des Prinzips der Kombination von Grundbegriffen besonders gut demonstrieren. Nehmen wir als Beispiel den Begriff „Lichtsinn“ (Notation 2730). Wie alle anderen Grundbegriffe ist dieser Begriff in seiner allgemeinsten Bedeutung, gewissermaßen frei im Raum der Begriffe stehend aufzufassen. Erst durch die Kombination mit Grundbegriffen vor allem der Systemgruppe 0 erhält er eine konkrete Bedeutung, wie die folgenden Beispiele veranschaulichen sollen:

- 2730 + 0601 (Körperausscheidungen): Tranenflussigkeit
- 2730 + 0610 (Zelle): Sehzelle
- 2730 + 0620 (Gewebe): Augengewebe
- 2730 + 0630 (Organ): Auge
- 2730 + 0640 (Gesamtorganismus): Sehleistung oder
- 2730 + 0630 + 0521 (Morphologie): Augenstruktur
- 2730 + 0630 + 0532 (Entwicklungsphysiologie): Augenwachstum
- 2730 + 0630 + 0533 (Vererbungslehre): Augenvererbung
- 2730 + 0630 + 0542 (Evolutionsphysiologie): Augenmutation

Wie diese hier aus Platzgründen unvollständig zusammengestellten Beispiele zeigen, ergibt die Kombination weniger, gut begründeter und gut definierter Grundbegriffe aus wenigen Einzelsystemen eine Vielzahl von konkreten Begriffen. Berücksichtigt man noch die zahlreichen Synonyme, dann wird verständlich, daß wir alle bisher bei der Bearbeitung von etwa 40000 Titeln vorkommenden Wörter und Begriffe des Eigenschaftsbereiches (etwa 50000) durch die Kombination von nur rund 2000 Grundbegriffen eindeutig festlegen konnten.

In der *Systemgruppe 3* haben wir die Einzelkategorien der Coenobiologie vor uns. Während sich die

Idiobiologie (Systemgruppe 2) mit den Einzellebewesen beschäftigt, sind unter der Coenobiologie alle Beziehungen zusammengefaßt, bei denen die Einzelorganismen als Teilmglieder einer höheren Einheit auftreten. Diese höheren Einheiten setzen sich entweder nur aus Einzellebewesen zusammen (beispielsweise bei Parasiten- oder Geschlechtsverhältnissen) oder aus organismischen und nichtorganismischen Teilmgliedern. Das letztere gilt für die Coenosen, die hier als Einheit aus Lebensraum (nichtorganismisch) und Lebensgemeinschaft (organismisch) definiert sind. In entsprechender Weise werden die geographischen Grundbegriffe benutzt. Daher lassen sich durch unser System zwangsläufig auch alle Begriffe aus den Wissenschaften von der Erde (Geographie, Geologie usw.) ausdrücken.

### III. Das Verfahren

Das Verfahren kann in 3 Abschnitte eingeteilt werden, die beim Arbeitsablauf zeitlich aufeinander folgen: Die Aufnahme der Literatureinheiten einschließlich ihrer Übertragung auf Magnetband, die automatische Klassifizierung mit allen dafür notwendigen Arbeitsvorgängen und die je nach Aufgabenstellung unterschiedliche Aufbereitung und Ausgabe des Materials.

#### 1. Aufnahme und Übertragung

a) Zeitschriftenweises Vorgehen: Aufnahme der Titel und sonstigen bibliographischen Angaben aller Originalveröffentlichungen möglichst vieler Zeitschriften mit Hilfe von Lochstreifenschreibmaschinen. Wie für alle Maschinenarbeiten ist dazu ein feststehendes Schema notwendig, das hier für jede Angabekategorie eine Kategoriennummer und für jede Schreibzeile eine Zeilennummer innerhalb jeder Kategorie enthält, so daß vor jeder Schreibzeile eine zweistellige Nummer steht. (Siehe Beispiele in Kapitel III/2/e.) Die einzelnen Kategorien sind folgendermaßen eingeteilt:

Kategorie 0: Veröffentlichungsnummer.

Sie enthält zunächst eine fünfstellige Zeitschriftennummer mit vorgestelltem Großbuchstaben. Der Buchstabe dient zur besonderen fachlichen Einordnung der Zeitschrift. Im übrigen werden alle Periodika ohne besondere Reihenfolge laufend durchnummeriert. — Als nächstes folgt eine dreistellige Bandangabe innerhalb der Zeitschrift. — Eine letzte dreistellige oder vierstellige Nummer kennzeichnet laufend jede einzelne Veröffentlichung innerhalb Band und Zeitschrift. — Durch diese Veröffentlichungsnummer wird also eine völlig eindeutige Signierung jeder Arbeit erreicht.

Kategorie 1: Für korporativen Verfasser (Institution) vorgesehen.

Kategorie 2: Verfasser.

Kategorie 3: Originalquelle.

Hier steht vor einer Klammer die Bandnummer, in der Klammer das Erscheinungsjahr und hinter der Klammer die Anfangs- und Endseite des Aufsatzes. Bei der *Aufnahme* der Literatur, um die es sich hier ja handelt, braucht in dieser Kategorie der Zeitschriftentitel nicht mitgeschrieben zu werden. Die Maschine speichert gesondert alle Zeitschriftentitel unter ihren Zeitschriftennummern (s. Kategorie 0) und schreibt sie bei der Ausgabe der Literaturstellen automatisch vor die Bandnummer.

Kategorie 4: Beigabenvermerke und sonstiges.

Zu dieser Kategorie gehören: Anzahl der Literaturzitate, der Abbildungen und der Tabellen sowie die Sprache des Aufsatzes und das Vorhandensein und die Sprache einer anderssprachigen Zusammenfassung. — Für die Ausgabe errechnet die Maschine außerdem aus Anfangs- und Endseite (s. Kategorie 3) den Umfang der Arbeit und schreibt diese Angabe als erste Zahl vor die übrigen Hinweise.

Kategorie 5: Für Sekundärquelle (Referatstelle in einem Referateorgan) vorgesehen.

Kategorie 6: Titel (mit Untertitel) und gegebenenfalls Titelübersetzung.

Kategorie 7: Schlagwörter oder Kurzreferat vorgesehen.

Kategorie 8: Für Science Citation Index-Verfahren vorgesehen.

Kategorie 9: Zur Verfügung für Ergänzungen.

Alle diese Angaben werden für jede Veröffentlichung auf eine Karteikarte DIN A6 geschrieben. Um das Ein- und Ausspannen auf der Schreibmaschine so weit wie möglich zu sparen, werden Karteikarten-Endlosformulare benutzt. Zugleich mit dem Beschreiben der Karten entsteht ein Lochstreifen. Die Karteikarten dienen später nur noch Kontrollzwecken.

b) Übertragung der in den Lochstreifen vorhandenen Informationen auf Magnetbänder mit der elektronischen Rechenanlage IBM 1401 und einem daran angeschlossenen Lochstreifenabfühler IBM 1011. Herausschreiben von Korrekturlisten des aufgenommenen Materials mit Hilfe des Druckers IBM 1403.

c) Korrektur der Aufnahme- und Übertragungsfehler.

#### 2. Automatische Klassifizierung

a) Heraussuchen und Zählen der in den Titeln (Angabekategorie 6) der Veröffentlichungen vorkommenden Wörter mit Hilfe der IBM 1401. Anschließend Verdichtung und Ausgabe der Wörter in Form von gelochten Wörterbuchkarten. Zugleich Herausschreiben einer Wörterliste mit den Häufigkeiten und Fundstellen (Wortfrequenzliste).

b) Klassifizierung der auf den Wörterbuchkarten stehenden Wörter und handschriftliche Eintragung der zugehörigen Notationen in diese Wörterbuchkarten.

Die an dieser Stelle des Arbeitsablaufes liegende Klassifizierung braucht durch den menschlichen Bearbeiter für jedes vorkommende Wort nur ein einziges Mal durchgeführt zu werden. Bei jedem weiteren Auftreten des gleichen Wortes in den Titeln der Veröffentlichungen klassifiziert dann die Maschine automatisch mit Hilfe der ihr mitgeteilten Notationen. Nur alle neuen und jeweils noch nicht vorgekommenen Wörter müssen den menschlichen Bearbeiter passieren. Je mehr Wörter und zugehörige Notationen die Maschine in einem internen Wörterbuch gespeichert hat, um so geringer wird die Zahl der hinzukommenden Wörter. Eines Tages dürfte dann ein Zustand erreicht werden, bei dem überhaupt nur noch neu in die Wissenschaft eingeführte Wörter hinzutreten.

Zunächst werden die Wörter in sieben verschiedene Wörterbuchkategorien eingestuft:

Füllwörter (1)	Artnamen (2)
Chem. Verbindungen (3)	Organismennamen (4)
Termini (5)	Nom./Term. (6)
	Vornamen (7)

Als Fullwörter werden alle Wörter geführt, die keinen Informationswert besitzen. (Beispiele: der, die das, und, oder, Untersuchungen, Ergebnisse.)

Die zur biologischen Nomenklatur gehörenden Artnamen sind unter einer besonderen Wörterbuchkategorie zusammengefaßt. Dies ist notwendig, weil die Artnamen einerseits isolierte Wörter sind, wie alle anderen Wörter auch, weil sie andererseits aber nur in Zusammenhang mit den zugehörigen Gattungsnamen die betreffende Organismenart eindeutig kennzeichnen. Die spezielle Lösung des sich daraus ergebenden Problems wird bei der Ausgabe beschrieben (Beispiele: palustris, Hustedti).

Die Wörterbuchkategorie 3 enthält alle chemischen Verbindungen, für die eine Bruttoformel angegeben werden kann (Beispiele: Natriumhydroxyd, Barbitursäure).

Als Organismennamen gelten alle wissenschaftlichen Bezeichnungen und alle Trivialnamen für Lebewesen innerhalb sämtlicher Kategorien vom Reich bis zur Gattung sowie die Art-Trivialnamen (Beispiele: Pflanzen, Vögel, Marchantiales, Drosophila, Tollkirsche).

Als Termini werden alle Wörter des Eigenschaftssystems eingruppiert (Beispiele: Albino, Immunität, Schwerkraftwirkung, Magennerven, Mittelamerika).

In der Wörterbuchkategorie 6 finden sich die aus Organismennamen und Eigenschaftsbegriffen zusammengesetzten Bezeichnungen und alle sonstigen Wörter, bei denen Organismen- und Eigenschaftsnotationen gemeinsam vorkommen (Beispiele: Bakterienzellwände, Amphibieneier, Blütenfarbe).

Für die menschlichen Vornamen wurde eine besondere Wörterbuchkategorie geschaffen, die in unserer weiteren Beschreibung nicht mehr berücksichtigt wird. Es hat sich nämlich gezeigt, daß beim Vorkommen dieser Wörter in den Titeln der Veröffentlichungen meistens Nachrufe oder sonstige biographische Aufsätze vorliegen (Beispiel für Titel: KARL LUEDERS 70 Jahre alt).

Die Füllwörter (WB 1) und Artnamen (WB 2) erhalten nur die Wörterbuchkategorie und keine Notationen zugeteilt. Die chemischen Verbindungen (WB 3) und Organismennamen (WB 4 und Teilwörter WB 6) werden nach den vorliegenden Objektsystemen klassifiziert und mit den entsprechenden Notationen versehen.

Die Klassifizierung der Termini (WB 5 und Teilwörter WB 6) geht nach dem Prinzip der Definition vor sich. Dabei wird die Bedeutung des Wortes durch Kombination der Notationen der im Eigenschaftssystem vorhandenen Grundbegriffe ausgedrückt, wobei gegebenenfalls auch noch Notationen aus den Objektsystemen hinzugezogen werden können. Das Wort „Pollen“ läßt sich beispielsweise als „Männliche (Notation 0092) Keimzellen (Notation 0000) von Gymnospermen (Notation G) und Angiospermen (Notation H)“ definieren und durch die betreffenden Notationen darstellen.

Alle synonymen Wörter der gleichen Sprache und die entsprechenden Wörter aus anderen Sprachen erhalten die gleichen Notationen.

Weiterhin wird jedem Wort ein Sprachmerkmal zugeteilt, so daß später auch alle Wörter der gleichen Sprache maschinell zusammengestellt werden können. Außerdem ergibt sich auf diese Weise ein mehrsprachiges Wörterbuch.

c) Lochen der Wörterbuchkategorie, des Sprachmerkmals und der Notationen in die Wörterbuchkarten. Anschließend werden die Wörterbuchkarten in die elektronische Maschine zurückgegeben. So entsteht ein Thesaurus auf Magnetband, der durch neue und jeweils noch nicht vorhandene Wörter laufend ergänzt wird.

d) Maschineninterne Kontroll- und Sortierarbeiten. Ausdrucken von alphabetischen und systematischen Wörterlisten (Thesauren). Weitere Prüf- und Korrekturvorgänge.

e) Nachdem alle Kontroll- und Korrekturarbeiten erledigt sind, liegen im grundsätzlichen zwei „saubere“

Magnetbänder vor: Ein Band, das die aufgenommenen Literatureinheiten mit allen zugehörigen Angaben enthält, und ein Band mit dem Thesaurus (Wörter und Notationen). Mit Hilfe dieser beiden Bänder und durch Verwendung einiger zusätzlicher interner Magnetbänder führt die Maschine die automatische Klassifizierung durch und baut zugleich zwei Register nach dem Begriffsprinzip auf, die der Beschleunigung der Abfrage dienen.

Als Beispiel bringen wir drei Literatureinheiten sowie einen Auszug aus dem alphabetisch geordneten Thesaurus und zeigen daran in vereinfachter, leicht verständlicher Form die Arbeitsgänge:

*Literatureinheiten*

Das Aufnahmeschema ist im Kapitel III/1/a ausführlich erklärt. Die hinter jedem Wort in Klammern stehenden Ziffern sind Wortnummern und dienen nur der Erläuterung.

- 00 A00011-046-021
- 20 Mutschler, Albert
- 30 46(1953) S. 364—418
- 40 0-3-1-Dt.-Engl., Russ.
- 60 Massenaufreten(1) der(2) Larven(3) von(4) Eristalis(5) tenax(6)
- 61 in(7) Abwasser(8)
  
- 00 B00104-085-006
- 20 Steiniger, Otto Hellberg, Karl-Heinz
- 30 85(1936) S. 12—29
- 40 15-0-2-Dt.-0
- 60 Die(1) Änderung(2) von(3) Blattgröße(4) und(5) Blütenfarbe(6)
- 61 häufig(7) angebauter(8) Blütenpflanzen(9) nach(10) Düngung(11)
- 62 mit(12) phenolhaltigem(13) Abwasser(14)
  
- 00 A00117-008-019
- 20 Koester-Uhlig, Heinrich, von
- 30 8(1959) S. 14—33
- 40 31-0-3-Dt.-0
- 60 Die(1) Größe(2) der(3) Chromosomen(4) von(5) Dipteren(6)

*Thesaurus-Auszug*

Wort	Wörterbuch-Kategorie	Sprach-merkmal	Notationen
Abwasser . . . . .	5	B	1311
Änderung . . . . .	5	B	0420 1019
Angebauter . . . . .	5	B	0220
Blattgröße . . . . .	5	B	1120 2950
Blütenfarbe . . . . .	6	B	H 1530 2270
Blütenpflanzen . . . . .	4	B	G H
Chromosomen . . . . .	5	A	0790
der . . . . .	1	B	—
die . . . . .	1	B	—
Dipteren . . . . .	4	A	09L
Düngung . . . . .	5	B	0226
Eristalis . . . . .	4	A	09L2109
Größe . . . . .	5	B	1120
häufig . . . . .	1	B	—
in . . . . .	1	B	—
Larven . . . . .	5	B	0030
Massenaufreten . . . . .	5	B	1115
mit . . . . .	1	B	—
nach . . . . .	1	B	—
phenolhaltigem . . . . .	3	X	Z0C6H6O1
tenax . . . . .	2	A	—
und . . . . .	1	B	—
von . . . . .	1	B	—

In einem ersten Arbeitsgang versieht die Maschine alle Wörter der Angabekategorie 6 (Titel) jeder Literatureinheit mit einer laufenden Nummer (bei unseren Beispielen hinter jedes Wort in Klammern gesetzt). Dann vergleicht sie diese Wörter mit dem Thesaurus und ordnet jedem Thesaurus-Wort die Veröffentlichungsnummern (Angabekategorie 0) aller Literatureinheiten zu, in deren Titel das Wort vorkommt. Diesen Veröffentlichungsnummern sind zugleich die laufenden Wortnummern beigelegt. Anschließend werden aus dieser maschineninternen Zwischenstufe heraus zwei Magnetband-Register aufgebaut: Ein Wortregister und ein Notationsregister.

In dem Wortregister sind die Füllwörter (Wörterbuchkategorie 1) nicht mehr enthalten, da sie vorher von der Maschine eliminiert wurden. Das Wortregister ist alphabetisch geordnet und umfaßt alle aussagefähigen Wörter mit den zugehörigen Veröffentlichungs- und Wortnummern.

Im Notationsregister ist auch die Wörterbuchkategorie 2 (Artnamen) nicht mehr geführt, weil dazu keine Notationen gehören. Das Notationsregister wird nach Notationen sortiert und erhält dadurch im Gegensatz zu dem alphabetischen Wortregister den Charakter eines systematischen Registers. Bei den einzelnen Notationen stehen ebenfalls die zugehörigen Veröffentlichungs- und Wortnummern. — Mit dem Aufbau des Notationsregisters ist zugleich die automatische Klassifizierung abgeschlossen.

Die Vorteile dieses Vorgehens liegen auf der Hand. Die automatische Klassifizierung erfolgt mehr als tausendmal schneller, als sie ein Mensch erledigen könnte, und arbeitet außerdem völlig einheitlich und fehlerfrei. Weiterhin stehen nun bereits alle einschlägigen Veröffentlichungsnummern beisammen (Begriffs-Prinzip!), so daß die Maschine im Rahmen der später zu schildernden Abfragevorgänge nicht mehr das gesamte gespeicherte Literaturmaterial zu durchsuchen braucht.

### 3. Aufbereitung und Ausgabe

Maschineninterne Aufbereitung des Literaturmaterials für die verschiedenen Auswerte- und Ausgabearten, von denen eingangs einige Möglichkeiten aufgeführt wurden. Anschließend daran maschinelles Herausschreiben von Übersichten, Listen oder beliebigen Karteikarten entsprechend den jeweils gewünschten Aufgabenstellungen.

Als Beispiel schildern wir abschließend den Auswerte- und Ausgabevorgang für Dokumentationsanfragen. Dabei betrachten wir unser ganzes Verfahren am besten vom Standpunkt des Benutzers aus. Die Verbindung zu den Benutzern ist überhaupt von grundsätzlicher Bedeutung, denn die Dokumentation hat eine dienende Aufgabe zu erfüllen. Nachfolgend führen wir einige Grundsätze auf, die für den Dokumentationsbenutzer wichtig sind, und bringen anschließend einige Abfrage-Beispiele.

a) Der Benutzer soll bei der Formulierung seiner Anfragen grundsätzlich frei sein und nicht zur Einhaltung irgendeines Schemas verpflichtet werden. Das Umsetzen dieser Benutzeranfragen für die Maschinen ist Sache der Dokumentationsstelle. Bei schwierigen oder unklaren Anfragen wird sich eine persönliche Zusammenarbeit zwischen Benutzer und Dokumentationsstelle empfehlen.

b) Die Maschine kann alle Fragen beantworten, die sich auf die im Kapitel III/1/a ausführlich geschilderten, für jede Literatureinheit gespeicherten Angaben und ihre beliebigen Kombinationen beziehen. Ein extremes Beispiel, das in der Praxis sicher kaum vorkommt, wäre die Frage: „Alle Veröffentlichungen von G. KUNTZE aus der Zeitschrift für Vererbungslehre während der Jahre 1950—1960 über die Vererbung der Augenfarbe beim Menschen, mit weniger als 20 Seiten Umfang, mehr als 10 Literaturangaben, 5 Abbildungen und 5 Tabellen, in deutscher Sprache und mit englischer Zusammenfassung“. — Für eine

rationelle Maschinenausnutzung ist es zweckmäßig, mehrere hundert Anfragen gleichzeitig zu bearbeiten.

c) Nach den Erfahrungen eines von uns durchgeführten und von mehr als 350 künftigen Benutzern beantworteten Anfragetestes wird unser Verfahren im Rahmen von Dokumentationsanfragen vorwiegend zwei Aufgaben zu erfüllen haben: Einmal die abonnementsmäßige, etwa monatliche Lieferung aller zu einem bestimmten Thema neu erschienenen Publikationen der erfaßten Zeitschriften und zum anderen das Beantworten von einmaligen Anfragen auf der Basis des gesamten gespeicherten Literaturmaterials.

d) Die Abonnementsanfragen sind meistens auf die besondere Arbeitsrichtung des anfragenden Instituts oder Wissenschaftlers zugeschnitten. Beispiele für solche Anfragen sind, ganz willkürlich aus dem Anfragetest herausgegriffen, etwa: „Biologie der Salmonellen“ oder „Tierische Zelldifferenzierung“ oder „Alles über die Gattung Centaurea“ oder „Menschliche Rassenkunde“ oder „Nukleolen bei Pflanzen und Tieren“ oder „Elektronenmikroskopie in Biologie und Medizin“. — In allen diesen Fällen möchte man also laufend über sein Arbeitsgebiet orientiert werden. Da schon die Bibliothekserfahrung gelehrt hat, daß auch Publikationen sehr wichtig sein können, die am Rande des eigentlichen Forschungsthemas liegen, ist unser Verfahren so ausgerichtet, daß bei diesen Abonnementsanfragen wahlweise engeres oder umfassenderes Material geliefert wird.

e) Die Möglichkeiten bei Einzelanfragen haben sich vor allem während der Entwicklung unserer Methode herauskristallisiert. Da ganze Zeitschriften, auch nach rückwärts, aufgenommen wurden, läßt sich durch eine Einzelanfrage an den Gesamtbestand unseres Literaturspeichers ermitteln, ob zu einer ganz speziellen Fragestellung bereits frühere Veröffentlichungen vorliegen. Soweit dies in den Titeln der Publikationen zum Ausdruck kommt, kann also den Benutzern auf diese Weise der gesamte von uns erfaßte Erkenntnisstand dienstbar gemacht werden. Derartige Fragen sind etwa: „Wo ist eine Methode zur Messung von Spaltöffnungen beschrieben?“ oder „Hat jemand die Speichelsekretion von Rebläusen in Abhängigkeit von täglichen Luftdruckänderungen bearbeitet?“ oder „Liegen schon Untersuchungen über artspezifische Unterschiede im Gehalt der Erythrozyten an Östradiol-Dehydrogenasen bei Säugetieren vor?“ oder „Ist bereits einmal das Vorkommen der Kohleule als Schädling an Mohn festgestellt worden?“ oder „Wurde schon ein Vergleich des Eiweißanteils vom Tabakmosaikvirus mit dem einer Temperaturmutante durchgeführt?“ — Diese Beispiele stammen aus unserem Literaturmaterial. Soweit überhaupt vorhanden, könnte keine normale Bibliographie und kein normales Register darauf Antwort geben. Die Maschine kann es aber. Hat man mit ihrer Hilfe dann erst einmal eine Arbeit gefunden, so ergeben sich meistens aus deren Literaturverzeichnis weitere Hinweise.

f) Abschließend einige Beispiele zur maschinellen Beantwortung von Anfragen. Wir greifen dafür auf das Kapitel III/2/e und die dortigen Beispiele zurück.

Angenommen, ein Benutzer stellt die Frage: „Alle Arbeiten über Abwasserprobleme.“ Dann wird über eine Lochkarte die einschlägige Notation 1311 in die elektronische Anlage eingegeben. Die Maschine sucht in ihrem Notationsregister die betreffende Notation auf und findet dort alle Veröffentlichungsnummern der einschlägigen Arbeiten. Mit diesen Veröffentlichungsnummern holt sie aus dem

Literatur-Magnetband alle entsprechenden Titel und sonstigen bibliographischen Angaben heraus und druckt sie in beliebiger Reihenfolge und Anordnung auf Listen oder Karteikarten.

Bei einer Anfrage „Änderung der Blütenfarbe“ werden die Notationen 0420, 1019 und H, 1530, 2270 eingegeben. Zugleich wird der Maschine über die betreffende Lochkarte durch besondere Symbole mitgeteilt, daß die beiden Notationsgruppen als logisches Produkt zu werten sind und daß diese Notationsgruppen jeweils eine Notationskette darstellen. Das logische Produkt besagt, daß die beteiligten Notationen alle gemeinsam vorkommen müssen. Eine Notationskette liegt vor, wenn zu einem Wort oder Begriff mehrere Notationen gehören, die in dem speziellen Fall zusammenbleiben sollen und sich nicht mit anderen Notationen anderer Wörter des gleichen Titels zu falschen Aussagen kombinieren dürfen. Dann macht die Maschine von den Wortnummern Gebrauch. Sie sucht in unserem Beispiel die Notationen für „Blütenfarbe“ (H, 1530, 2270) heraus und vergleicht alle unter diesen Notationen stehenden Veröffentlichungsnummern einschließlich ihrer Wortnummern miteinander. Dabei findet sie, daß nur die Nr. B00104-085-006 (6) bei allen drei Notationen identisch ist. Diese Nummer wird dann noch mit den Veröffentlichungsnummern unter den Notationen für das Wort „Änderung“ (0420, 1019) verglichen, wobei sich herausstellt, daß die gleiche Nummer auch dort vorkommt. Es liegt also eine Arbeit über „Änderung der Blütenfarbe“ mit der Veröffentlichungsnummer B00104-085-006 vor, die entsprechend herausgeschrieben wird.

Bei einer Anfrage nach allen Arbeiten über Dipteren sucht die Maschine nach der Notation O9L. Sie holt dann nicht nur alle Arbeiten heraus, die unter dieser Notation selbst stehen, sondern sämtliche Arbeiten zu allen Notationen, die mit O9L beginnen, wie beispielsweise die Notation für Eristalis (O9L2109). Hieran sieht man, daß durch unsere Notationsgebung zwangsläufig alle Unterbegriffe (in unserem Beispiel alle Familien und Gattungen der Dipteren) miterfaßt werden. In einem herkömmlichen alphabetischen Register wäre dazu das Aufsuchen von mehreren hundert oder gar tausend verschiedenen Namen erforderlich, was praktisch kein Mensch bewältigen kann.

Eine Anfrage nach einer bestimmten Organismenart, wie etwa *Eristalis tenax*, erfordert ein besonderes Vorgehen, weil der Artname (*tenax*) nur in Verbindung mit dem Gattungsnamen (*Eristalis*) den gewünschten Informationswert besitzt (s. auch die Beschreibung der Wörterbuchkategorien). In allen solchen Fällen sucht die Maschine die Gattung im Notationsregister, den Artnamen jedoch im Wortregister. Nur Trivialnamen, wie beispielsweise „Maikäfer“, die aus einem einzigen Wort bestehen, haben eine bis zur Art reichende Notation und sind daher auch im Notationsregister zu finden.

Die Lochkarten, in die unsere verschiedenen Anfragen eingelocht sind, werden aufbewahrt, so daß im Laufe der Zeit eine Anfragekartei entsteht. Wenn neue Anfragen eingehen, wird zunächst geprüft, ob diese Anfragen in der Kartei bereits vorhanden sind, wodurch unnötige Arbeit gespart werden kann.

#### IV. Schlußbemerkungen

##### 1. Grenzen des Verfahrens

Es wäre unwissenschaftlich, zu verschweigen, was unser Verfahren *nicht* leisten kann und welche Grenzen ihm gesetzt sind. Die Beschränkung auf die Titel der Veröffentlichungen wurde oben bereits ausführlich begründet. Damit hängt auch zusammen, daß sich keine Anfragen nach konkreten Untersuchungsergebnissen beantworten lassen („Wie viele Chromosomen hat die Sonnenblume?“ oder „Wie lang ist der Darm des Meerschweinchens?“ oder „Welche Pflanzen haben gelbe Blumenblätter und blühen in den ersten 4 Monaten des Jahres?“). Dazu ist der Aufbau der höchsten Dokumentationsstufe (Ergebnisdokumentation) erforderlich, für die ganz andere Voraussetzungen nötig sind. Weiterhin liegt es in der Natur der Sache, daß auch die automatische Klassifizierung eine Reihe von Problemen aufwirft, deren ausführliche Beschreibung hier nicht möglich ist. Vor allem handelt es sich dabei um sprachwissenschaftliche Fragen, wie sie in ähnlicher Form auch bei der maschinellen Übersetzung auftreten. Hierzu gehören beispielsweise die Probleme

der Homonymie sowie Fragen der Syntax. Soweit dies vorausschauend möglich war, haben wir diese Dinge bei der Entwicklung unseres Verfahrens berücksichtigt. Darüber hinaus werden uns die praktische Arbeit und die laufende Auskunftstätigkeit in die Lage versetzen, in Zusammenarbeit mit den Benutzern noch vorhandene Mängel auszumerzen. Parallel dazu wird die Arbeit an dem Klassifikationssystem weitergehen, weil das Verfahren die Möglichkeit bietet, das Material gegebenenfalls maschinell umzuklassifizieren. Trotz der verbleibenden Grenzen, die durch die Struktur der Sprache bedingt sind, glauben wir, den Benutzern mit dem Verfahren nicht unwesentlich weiterhelfen zu können, wobei auch zu bedenken ist, daß man den zweiten Schritt nicht vor dem ersten tun soll und daß neunzig- oder gar hundertprozentige Lösungen auf dem Gebiete der Dokumentation sowieso nicht möglich sind.

##### 2. Ausblick

Da das Verfahren seine Bewährungsprobe im Prinzip bestanden hat und jetzt dank den Bemühungen des Instituts für Dokumentationswesen bei der Max-Planck-Gesellschaft in Göttingen eine elektronische Anlage ausschließlich für Dokumentationszwecke zur Verfügung steht, hoffen wir, die Anwendung der Methode auf den Gesamtbereich der Wissenschaften ausdehnen und eine maschinell bearbeitbare Nationalbibliographie der deutschsprachigen Zeitschriftenaufsätze schaffen zu können.

Dann würden wir eine breite Grundlage der Automatisierung für Veröffentlichungswesen, Bibliothekswesen und Dokumentation besitzen, die allen zur Zeit überblickbaren Anwendungsmöglichkeiten gerecht wird. Zugleich ist wünschenswert, an diesem Material eine entsprechende Klassifikationsforschung zu treiben, ein Arbeitsbereich, der auf internationaler Ebene ständig an Bedeutung gewinnt und auch Eingang in die Universitäten finden sollte.

Die „Nationalbibliographie auf Magnetband“ dürfte keine Konkurrenz zu den vorhandenen Dokumentationsstellen sein und daher auch nicht direkt Auskünfte an Benutzer geben. Dies würde eine viel zu große Organisation erfordern, die in den schon arbeitenden und bewährten Dokumentationsstellen ja bereits vorliegt. Deshalb erscheint es als die beste Lösung, daß jede Dokumentationsstelle das für sie wichtige Material in duplizierter Form von der Nationalbibliographie erhalten und damit beliebig weiterarbeiten kann.

Auf weitere Sicht besteht vielleicht die Möglichkeit, die Nationalbibliographie im ganzen — als Magnetband-Duplikat — dem Ausland zum Tausch anzubieten, um dadurch entsprechende Dokumentationsvorhaben in anderen Ländern anzuregen. Eventuell läßt sich so erreichen, daß auf der ersten Dokumentationsstufe der Titel-Erfassung und -Auswertung in nicht allzu ferner Zeit eine internationale Regelung möglich wird, auf der dann weitere Arbeiten aufbauen können.

##### Literatur

Hier sind aus Platzgründen nur einige Überblicksarbeiten angeführt, die mit dem Aufbau der Dokumentation allgemein oder unserer Dokumentationsstelle im besonderen zusammenhängen. Wer spezielles Fachschrifttum sucht, sei auf die Bibliographie [4] und die einschlägigen Zeitschriften (in Deutschland besonders die „Nachrichten für Dokumentation“) verwiesen.

[1] CREMER, M.: Aufgabe und Funktion des Instituts für Dokumentationswesen. Mitt. Max-Planck-Ges. 1963, H. 1/2, 77—86. — [2] PIETSCH, E.: Struktur von Informationseinrichtungen. Nachr. Dok. 15, H. 1, 28—41 (1964). — [3] SCHEELE, M.: Die Lochkartenverfahren in Forschung und Dokumentation, mit besonderer Berücksichtigung der Biologie, 2. Aufl. Stuttgart: Schweizerbart 1959. — [4] SCHEELE, M.: Literatur über Lochkartenverfahren. Eine vermittels Sichtlochkarten aufgeschlüsselte Bibliographie über Lochkarten und die mit ihrer Anwendung zusammenhängenden

Fragen. Schlitz/Hessen: H. Guntrum II. K. G. 1959. — [5] SCHNEIDER, K.: Fünf Jahre KWIC-Indexing nach H. P. LUHN. Nachr. Dok. 14, H. 4, 200—205 (1963). — [6] SENGBUSCH, R. v.: Probleme der Auslese und ihre Beziehungen zu unserem kulturellen Leben. Jahrb. Max-Planck-Ges. 1959, 135—185.

*Max-Planck-Institut für Kulturpflanzenzüchtung,  
Hamburg*

Eingegangen am 19. Juli 1964