

Damit ein solches Maschinenprogramm in der mechanisierten Chemie-Dokumentation nutzbringend eingesetzt werden kann, muß es mit einem anderen Programm verbunden werden, welches das Auffinden der *gesuchten Begriffe selbst* zum Ziele hat, unabhängig von ihren wechselseitigen Verknüpfungen. Solange nicht sichergestellt ist, daß in einem gespeicherten Dokument die Begriffe der Fragestellung auch tatsächlich vorkommen, hat es nämlich keinen Sinn, daß man die Maschine auf die Suche nach der gewünschten Verknüpfung dieser Begriffe schickt. Gegenwärtig sind wir damit beschäftigt, den hintereinandergeschalteten Ablauf zweier solcher Maschinenprogramme mit unterschiedlicher Aufgabenstellung zu verwirklichen. Zur Vorselektion nach den Begriffen selbst dient uns das in der IDC^[3] bereits eingeführte Verfahren zum Recherchieren nach chemischen Fachbegriffen.

4. Ausblick

Betrachtet man die vertraute Strukturformel aus dieser Perspektive, so kann man auch sie als einen Graphen auffassen, wenn auch mit besonderen Eigenschaften. Nicht ein Stoffbegriff besetzt hier einen Punkt, sondern ein Atom. Die Bindungen zwischen den Atomen fungieren als Strecken in diesem Graphen. Tatsächlich ist die Strukturformel auch als ein immer noch hoch-

[3] S. Rössler u. A. Kolb, J. Chem. Doc. 10, Nr. 2, S. 128 (1970); E. Meyer, Angew. Chem. 82, 605 (1970); Angew. Chem. internat. Edit. 9, Heft 8 (1970).

gradig zusammengesetzter Begriff aufzufassen, aufgebaut aus vielen „Einfachbegriffen“, nämlich den Atomen selbst. Nun ist es eine mehr als hundert Jahre alte Erfahrung, daß die analytische Betrachtungsweise und Darstellungsweise für den Strukturformelbegriff ungleich anschaulicher und übersichtlicher ist, als die komplexe Nomenklatur oder gar eine willkürlich festgelegte Trivialbezeichnung für einen Stoff. Darüber hinaus ist diese Betrachtungsweise wissenschaftlich auch ungleich fruchtbarer, weil sie verwandtschaftliche Beziehungen zwischen den Stoffbegriffen unmittelbar ins Auge springen läßt.

Aus dieser Perspektive stellen die beschriebenen Graphen nichts anderes dar als die Übertragung der Strukturformelidee auf stärker ausgeweitete Zusammenhänge, nämlich auf die Zusammenhänge innerhalb einer Publikation. Der Gewinn an Übersichtlichkeit wird deutlich, wenn man sich die Mühe macht, den oben gezeichneten Graphen mit dem Originaltext der verfahrenstechnischen Patentschrift zu vergleichen. Wir verfolgen daher sehr aufmerksam, welchen Anklang solche Graphen bei unseren Kollegen in der Forschung und im Betrieb finden, als reines Mittel zur Schnellinformation, eventuell zusätzlich zu einem Referat oder zum Originaltext.

Diese Arbeit wurde durch Mittel des Bundesministeriums für Bildung und Wissenschaft gefördert. Für anregende Diskussionen danken wir dem Arbeitskreis für Makromolekularchemie-Dokumentation.

Eingegangen am 7. Januar 1970 [A 768]

Der Dokumentationsring der chemisch-pharmazeutischen Industrie; Ziele und Methoden

Von Wolfgang Nübling und Walter Steidle^[*]

Die im Dokumentationsring zusammenarbeitenden Firmen wählten für die Verschlüsselung der pharmazeutisch-chemischen Literatur das Lochbildverfahren. Chemische, biochemische und medizinische Fakten einer Publikation werden dabei in Grundbegriffe zerlegt, denen auf der Lochkarte jeweils eine Position zugeteilt ist. Dieses Verfahren hat sich seit zwölf Jahren bewährt.

1. Einleitung

Der 1958 gegründete Dokumentationsring^[1] der chemisch-pharmazeutischen Industrie hat sich die Aufgabe gestellt, chemische und verwandte Fachliteratur durch Verbesserung vorhandener und Entwicklung

neuer Methoden intensiver zu erfassen und damit besser nutzbar zu machen. Der Dokumentationsring versetzt seine Teilnehmer in die Lage, bei angemessenem eigenen Aufwand über das große Gesamtmaterial der Gemeinschaft verfügen zu können. Jedes Mitglied bearbeitet einen gleich großen Anteil der Quellen (Zeitschriftenliteratur, Patentschriften, andere Literatur). Diese Arbeitsteilung gestattet es auch kleineren

[*] Dr. W. Nübling
E. Merck
61 Darmstadt 2, Postfach 4119
Dr. W. Steidle
Knoll AG
67 Ludwigshafen

[1] Gründungsmitglieder: Farbenfabriken Bayer AG, Leverkusen; CIBA AG, Basel (Schweiz); Knoll AG, Ludwigshafen; E. Merck, Darmstadt; Dr. Karl Thomae GmbH, Biberach.

und mittleren Institutionen, eine umfangreiche Dokumentation mit den heute fast überall vorhandenen Lochkartenmaschinen zu betreiben.

Insbesondere der vom Dokumentationsring erfaßte pharmazeutisch-chemische, biochemische und medizinische Teil der Literatur ermöglicht es den Mitgliedern, Hochschulen und Kliniken mit Literatur zu versorgen, die diesen Stellen normalerweise in diesem Maße überhaupt nicht zugänglich ist. Als erwünschter Nebeneffekt ergab sich die Möglichkeit, die Forschungsergebnisse der Mitglieder nach den gleichen Methoden zu bearbeiten und so veröffentlichte und unveröffentlichte Daten zusammenzuführen.

2. Organisation des Dokumentationsringes

Zur Zeit haben sich 15 Mitglieder^[2] zusammengefunden, um die Dokumentationsprobleme wirtschaftlichen und schnell realisierbaren Lösungen zuzuführen. Die Entwicklung geeigneter Schlüssel für die Datenverarbeitung nimmt dabei einen hervorragenden Platz ein.

Die Erfassung und Verarbeitung der pharmazeutisch-chemischen Literatur macht besondere Schwierigkeiten. Die angesprochenen Fachgebiete sind weit gestreut; sie überdecken Medizin, Chemie und Pharmazie. Aus den Publikationen müssen alle Einzelheiten, die diese Fachbereiche betreffen, erfaßt und in geeigneten Ordnungssystemen festgehalten werden, denn nur dann besteht die Möglichkeit, bei späteren Anfragen die zur Bearbeitung eines bestimmten Themas notwendigen Unterlagen mit vertretbarem Aufwand herauszufinden. Dazu muß die Beziehung zwischen der chemischen Substanz und ihrer Anwendung oder Wirkung jederzeit hergestellt werden können.

Bei den Benutzern^[3] werden die verschiedenartigsten Computersysteme verwendet; für fast alle eignen sich jedoch Lochkarten als Eingabe-Medium. Auch ist der Austausch von Lochkarten in kurzen Zeitabständen, d.h. bei kleinen Kartenmengen, problemlos. So ist sichergestellt, daß alle Teilnehmer schnell mit aktuellem Material versorgt werden können. Unser Ziel war, die Benutzung des Systems mit einfachen Lochkartenmaschinen zu gewährleisten. Die meisten Benutzer gebrauchen inzwischen zwar Magnetbänder und Platteneinheiten als Datenspeicher, jedoch bewältigen auch heute noch einige Mitglieder das in der Zwi-

schenszeit gesammelte Material^[4] mit Hilfe von Lochkartenmaschinen.

Darüber hinaus erlauben Lochkarten das Erstellen von Spezialkarteien, die nur Teilbereiche umfassen oder nach bestimmten vorgegebenen Gesichtspunkten vorsortiert sind. Diese Spezialkarteien können, dezentralisiert aufgestellt, die Arbeit einzelner Betriebsstellen wesentlich erleichtern oder bestimmte Fragestellungen mit einfachsten Sortiermaschinen schnell und mit geringstem Aufwand lösen. Schließlich können andere Interessenten solche Teilkarteien erwerben und benutzen.

Die dezentrale Anfragenbeantwortung wird der zentralen vorgezogen. Die Verarbeitungswege im eigenen Haus sind kürzer. Erst durch den ständigen Dialog zwischen Fragesteller und Speicher wird ein optimales Ergebnis gewährleistet. Das setzt eine gute Zusammenarbeit zwischen Anfragendem und Dokumentationsstelle voraus, denn nur der Dokumentationsfachmann ist in der Lage, die Frage des Benutzers in eine der Speichermethode adäquate Form zu bringen.

Die stetig wachsenden Kosten für den Maschinenpark einer modernen Dokumentation werden allerdings früher oder später eine zentrale Speichereinheit erfordern, die den Dokumentationsstellen aller Mitglieder über entsprechende Terminals zur Verfügung steht. Darüber hinaus wird es nur eine solche Lösung gestatten, das gespeicherte Material durch Bereitstellung geeigneter Computer-Programme optimal zu nutzen. Zu dieser optimalen Nutzung gehören selbstverständlich nicht nur die gängigen Fundstellen-Nachweise, sondern auch Antworten auf ungewohnte Fragestellungen, wodurch z.B. Entwicklungstrends frühzeitig erkennbar werden oder automatisch vorteilhafte Kombinationen von Methoden, wie etwa in der analytischen oder präparativen Chemie, ermittelt werden können.

3. Verschlüsselungsmethoden des Dokumentationsringes

Bei der Wahl des Verschlüsselungssystems ließen wir uns vom Wunsch nach einem möglichst unkomplizierten Schlüssel leiten, der zudem den Gepflogenheiten und Gedankengängen der an der Verschlüsselung beteiligten Fachleute so weit wie möglich entgegenkommen sollte. Die Verschlüsselung muß schnell und fehlerfrei durchgeführt werden können. Die Fehler sind aber um so zahlreicher, je komplizierter und ungewohnter das Verschlüsselungssystem ist. Die Einspeicherung des Materials, d.h. die Verschlüsselung, darf auch nicht beliebig lange dauern, da jedes System nur dann sinnvoll ist, wenn das vorgegebene Volumen in der verfügbaren Zeit bewältigt werden kann. Von der Verschlüsselungsdauer ist auch der Zeitpunkt abhängig, zu dem das gespeicherte Material für Recherchen wieder zur Verfügung steht. Außerdem muß das System leicht erlernbar sein, damit es von

[2] Neben den Gründungsmitgliedern die Firmen Boehringer Mannheim GmbH, Mannheim; Laboratoires Diamant S.A., La Plaine-Saint-Denis (Frankreich); L'Equilibre Biologique, Boulogne-sur-Seine (Frankreich); J. R. Geigy AG, Basel (Schweiz); Chemie Grünenthal GmbH, Stolberg; N. V. Philips-Duphar, Weesp (Niederlande); Rhône-Poulenc, Vitry-sur-Seine (Frankreich); Roussel-Uclaf, Romainville (Frankreich); Schering AG, Berlin; Troponwerke Dinklage & Co., Köln-Mülheim.

[3] Nach dem vom Dokumentationsring entwickelten System zur Bearbeitung „pharmazeutischer“ Literatur im weitesten Sinne wird die Literatur seit 1964 von der Firma Derwent Publications Ltd., London, ausgewertet. Seither steht dieser Literaturdienst unter den Bezeichnungen Ringdoc (pharmazeutisch-chemische Literatur), Pestdoc (Pflanzenschutz-Literatur) und Vetdoc (veterinär-medizinische Literatur) in erweiterter Form mehr als 100 Abonnenten zur Verfügung.

[4] Seit Gründung des Dokumentationsringes wurden etwa 500 000 Arbeiten referiert. Ihr Inhalt ist in mehr als 1,5 Millionen Lochkarten festgehalten.

Steroid-Schlüssel									
1/1/2	2/1/2 	3/1/2 A ≤ 4	6/1/2 D ≤ 4	7/1/2 ≤ 5	9/1/2 KONDENSIERBAR m. O, S	11/1/2 MIT ≤ 5	15/1/2 	16/1/2 LAVEL 7 AM GERÜST	
1/1/1 SEID	2/1/1 NDR	3/1/1 A ≤ 5	6/1/1 D ≤ 6	7/1/1 SONST. ≤ 5	9/1/1 SP n. 7/12-10/12	11/1/1 ≤ 5	15/1/1 ISO	16/1/1 	
1/0/ KOHLENHYDRAT	2/0/ S, K 1C	3/0/ A ≥ 7	6/0/ D ≥ 7	7/0/ X	9/0/ X	11/0/ X	13/0/ 20	16/0/ 21	
1/1/ LEWISS	2/1/ X 2	3/1/ 6/1	7/1/ 7	9/1/ X	11/1/ 15	15/1/ 22-24	16/1/ ≥ 25		
1/2/ NUCLEINSAUREN	2/2/ S, K 3-4 C	3/2/ 3α 17-X	6/2/ α 17-X	7/2/ α X	9/2/ α X	11/2/ α	15/2/ a	16/2/ a	
1/3/ ANDERE STOFFE	2/3/ S, K ≥ 5 C	3/3/ 3β 17β	6/3/ 17β β	7/3/ β X	9/3/ β X	11/3/ β X	15/3/ b	16/3/ b	
1/4/ KLUSTEN- u. POLYM.	2/4/ ISOLIER. AUS	3/4/ 3-OH 3-Dok	6/4/ 3-OH 3-Dok	7/4/ 3-OH 3-Dok	9/4/ 3-OH 3-Dok	11/4/ 3-OH 3-Dok	15/4/ 3-OH 3-Dok	16/4/ 3-OH 3-Dok	
1/5/ 2/5/ NAKSWEIS	3/5/ 3-XZ 17-X	6/5/ 17-X XZ	7/5/ XZ XZ	9/5/ XZ XZ	11/5/ XZ XZ	15/5/ XZ XZ	16/5/ XZ XZ		
1/6/ VERFAHREN	2/6/ 3=Y 17-Y	3/6/ 17=Y =Y	7/6/ =Y =Y	9/6/ =Y =Y	11/6/ =Y =Y	15/6/ =Y =Y	16/6/ =Y =Y		
1/7/ KARTENTYP	2/7/ SYNTH. DAMEST.	3/7/ 3-C -C	6/7/ 17-C -C	7/7/ -C -C	9/7/ -C -C	11/7/ -C -C	15/7/ -C -C	16/7/ -C -C	
1/8/ PATENT	2/8/ STRUKT. UNSICHER	3/8/ 3 Δ 3-4	6/8/ 6 Δ 17-20	7/8/ 7 Δ 1-2 7-8	9/8/ 9 Δ M-12	11/8/ 11 Δ 11-12 11-18	15/8/ 20-27 22-24	16/8/ 20-21 ≥ 23-	
1/9/ AUS DER LITERATUR	2/9/ ALLG. FORMEL	3/9/ 3 ○ ○	6/9/ 17 ○ ○	7/9/ ○ ○	9/9/ ○ ○	11/9/ ○ ○	15/9/ ○ ○	16/9/ ○ ○	

A 770.1

49 Antiepileptica	67	68	69	70	71	72 Therapie ohne NW
49 Antipruriginosa	2	4	0			72 Therapie mit NW
49 Licht-Strahlenschutzmittel	2	4	1			72 Nebenwirkung
49 Desodorantien	2	4	3			72 Therapie der NW
49 Milatica	2	4	4			72 Dauertherapie
49 Mydratika	3					72 Prophylaxe
49 Cardiotonika	4					72 Kontraindikation
49 Gefäß-erweiternde Mittel	5					72 Kthologie
49 Gefäß-verengende Mittel	6					72 Prognose
49 blutdrucksenkende Mittel	7					72 Diagnose
49 Antihypertensiva	8					72 Kausalfaktor
49 blutdrucksteigernde Antihypertensiva	9					72 Statistik
49 Kreislaufmittel undefiniert	0	Krankheitschlüssel bzw. Galenik				72 Statistik

einer möglichst großen Zahl von Teilnehmern benutzt werden kann. Nur dann lassen sich die Kosten in erträglichen Grenzen halten.

Diese Überlegungen haben uns bewogen, für die Chemie statt topologischer Methoden einfachere und billigere zu wählen. Sie liefern allerdings kein eindeutiges Ergebnis, sondern bringen neben den richtigen Antworten immer einen – je nach Anfrage mehr oder weniger großen – Anteil an Ballast. Der Dokumentar ist also manchmal gezwungen, das von der Maschine gelieferte Material zu sichten, bevor er es dem Interessenten weitergibt. Da nur ein Bruchteil des gespeicherten Materials jemals wieder abgefragt wird, spielt die Wirtschaftlichkeit der Eingabe eine entscheidende Rolle. Allzu differenzierte Verschlüsselung führt häufig zum Ausschluß von Unterlagen, die für die Bearbeitung eines Themas sinnvoll und notwendig wären. Soll dieser Nachteil vermieden werden, so müssen erst in langwierigen Überlegungen alle notwendigen Begriffe ermittelt werden. In den meisten Fällen kommt aber der Dokumentationsfachmann, der solche differenzierte Fragen computergerecht formulieren muß, damit in beträchtliche Schwierigkeiten.

Ein sinnvoll aufgebautes Fragmentierungssystem vermeidet diese Schwierigkeiten von vornherein; schon durch die Auswahl der zu berücksichtigenden Fragmente ist es möglich, die Anzahl relevanter Antworten wunschgemäß zu vergrößern oder zu verkleinern. Durch Loslösung der Verknüpfungsmerkmale von den Fragmenten liefert die Recherche häufig Kombinationen, an die weder Fragesteller noch Dokumentar gedacht hatten, die aber für die Arbeit des Fragestellers von großem Wert sein können.

Als besonders günstig für die Aufnahme der chemischen, biochemischen und medizinischen Fakten der Publikation auf *einer* Lochkarte erwies sich das Lochbildverfahren. Jeder zu erfassende Tatbestand wird in

Kasuistik, Prophylaxe und Therapie

$$\overbrace{72/8} \quad \overbrace{72/3} \quad \overbrace{72/12.2}$$

des Heufiebers, des Bronchialasthmas

W.H.O.-Code 240	W.H.O.-Code 241
--------------------	--------------------

der Nesselsucht und des

W.H.O.-Code 243

allergischen Ekzems, sowie Bekämpfung

49/12

W.H.O.-Code 244

des Juckreizes mit dem Steroid

49/11 01/12.11

Decortilen®

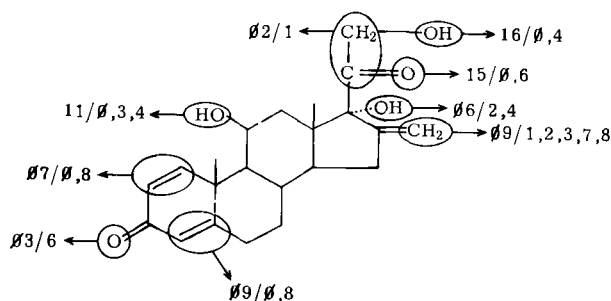


Abb. 1. Oben: Ausschnitte aus dem Lochbeleg mit medizinischer und partieller chemischer Verschlüsselung für die Behandlung des Heufiebers und Bronchialasthmas, der Nesselsucht und des allergischen Ekzems mit Decortilen®. Unten: Verschlüsseltes Referat und verschlüsselte Formel des Decortilens.

sinnvolle Grundbegriffe zerlegt, und jedem Grundbegriff wird eine bestimmte Position auf der Lochkarte zugeordnet. Auf diese Weise ist es möglich, jeden Grundbegriff allein oder in beliebiger Kombination mit anderen Grundbegriffen abzufragen.

Jeder der 960 Lochpositionen, also jedem bit, ist somit eine bestimmte Bedeutung zugeordnet. Dies gilt nicht nur für die Chemie – wo der Fragmentcode heute schon Allgemeingut geworden ist – sondern auch für Medizin, Biochemie und Biologie, wo Organismen, Organe, pharmazeutische Wirkgruppen, Stoffwechsel und Physiologie nach dem gleichen Prinzip erfaßt werden.

Die Verschlüsselung einer Publikation macht in der Regel die Anlage mehrerer Lochkarten notwendig. Überlagerungsregeln geben an, welche Verbindungen und Sachverhalte übereinandergeschlüsselt werden dürfen. Diese Regeln verhindern einerseits ein übermäßiges Anwachsen des Speichers, führen andererseits aber dazu, daß bei Recherchen Fehlkarten anfallen. Auf diese Weise wird die Verwendung von Schlagwörtern auch für die nichtchemischen Sachverhalte umgangen.

Wir sind der Ansicht, daß Schlagwörter wegen der Uneinheitlichkeit der Wortprägung und wegen ihrer Eindimensionalität kein ideales Mittel zur Dokumentation von Sachverhalten sind. Sie verbrauchen überdies relativ viel Platz auf dem Magnetband oder im Speicher, mit dem trotz der enormen Kapazität moderner Anlagen angesichts des großen und ständig wachsenden Literaturvolumens sparsam umgegangen werden sollte. Natürlich kann man diese Schwierigkeiten teilweise umgehen, indem man die Ausdrücke normiert und Thesauri zusammenstellt, in denen die erlaubten Wortprägungen gesammelt und die nicht erlaubten Ausdrücke erlaubten zugewiesen werden. Diese Tätigkeit ist aber mühsam, muß sich notwendigerweise über Jahre hinziehen und ist praktisch niemals beendet.

Dagegen erlaubt es die bitweise Klassifizierung, d.h. die Tatsache, daß für jeden Grundbegriff von vornherein eine Adresse festliegt, mit geringsten Hilfsmitteln eine eindeutige Speicherung vorzunehmen. Gleichzeitig ist ein außerordentlich schneller Zugriff möglich. Durch Kombination der Bausteine und durch Verwendung von Zahlen- und alpha-numerischen Schlüsseln in Teilbereichen ist es gelungen, alle gewünschten Sachverhalte befriedigend wiederzugeben.

Abbildung 1 zeigt ein Beispiel für die Verschlüsselung der Therapie von Krankheiten.

Ein weiterer großer Vorteil unserer Methode liegt in der Tatsache, daß sich wegen der Bindung an die Lochkarte Formblätter als Lochbelege verwenden lassen, auf denen die vorgegebenen Positionen für chemische und biologische Sachverhalte sowie Reaktionsschritte markiert sind, so daß es dem Bearbeiter leichtfällt, eine Arbeit korrekt zu verschlüsseln. Der Zwang zur Verschlüsselung auf dem vorgedruckten Lochbeleg bringt dem Verschlüsseler ständig die vorhandenen Möglichkeiten vor Augen und veranlaßt ihn, diese auch zu benutzen. Verschlüsselungsfehler und Auslassungen verschlüsselungsfähiger Begriffe werden dadurch auf ein Minimum beschränkt. Die Benutzung eines Beleglesers erspart heute die manuelle Eingabe mit ihren Fehlermöglichkeiten.

Wir kennen kein anderes System mit einer solch idealen Relation zwischen Eingabe- und Ausgabe-kosten, zumal der über einen längeren Zeitraum gemittelte Prozentsatz der bei einer Anfrage anfallenden Lochkarten nur 0.02% der befragten beträgt. Wegen der Vorsortierung der Karteien wird in der Regel nur ein Bruchteil des Bestandes durchgesucht. Da ferner mehrere Lochkarten zum gleichen Referat führen, ist die Zahl der als Ergebnis erhaltenen Literaturstellen trotz des oben beschriebenen Ballasts außerordentlich klein. Wir sind daher zur Ansicht gekommen, daß aufwendigere Systeme unwirtschaftlich wären.

Nach Fertigstellung der Grundlagen für die chemisch-pharmazeutische Dokumentation^[5] hat sich der Dokumentationsring anderen Problemen zugewendet. Es seien hier nur die Dokumentation chemischer Reaktionen^[6] sowie die Verschlüsselung von Kohlenhydraten, von pharmazeutischen Spezialitäten und von Markush-Formeln erwähnt.

Eingegangen am 19. Mai 1970 [A 770]

[5] Vgl. M. Scheublein, W. Steidle, W. Germann, W. Nübling u. R. Wilhelmi, *Naturwissenschaften* 55, 362 (1968).

[6] O. Schier, W. Nübling, W. Steidle u. J. Valls, *Angew. Chem.* 82, 622 (1970); *Angew. Chem. internat. Edit.* 9, Nr. 8 (1970).