

# Zur Mechanisierung der Dokumentation in der Chemie

VON DR. ERNST MEYER

BADISCHE ANILIN- & SODA-FABRIK AG., LUDWIGSHAFEN/RHEIN

*Veröffentlicht zum 100-jährigen Bestehen der Badischen Anilin- & Sodafabrik AG., am 6. April 1965*

*Der Grund dafür, daß sich mechanisierte Dokumentationsverfahren in der Chemie noch nicht in großem Umfange durchsetzen konnten, liegt vorwiegend in der geringen Flexibilität der verwendeten Systeme und in der Schwierigkeit, alle Fragestellungen zu präzisieren. Für die Dokumentation von chemischen Strukturformeln, Reaktionstypen und sonstigen Sachverhalten werden Methoden beschrieben, die anpassungsfähiger als die bisher gebräuchlichen sind. Um nicht auch für einfachere Anfragen elektronische Rechenanlagen einsetzen zu müssen, wird vorgeschlagen, von einem zentralen Speicher aus kleinere Karteien mit fertigem Material zu beliefern, das maschinell nach dem jeweils bestgeeigneten Code verschlüsselt ist.*

Stärker als alle anderen Wissenschaftszweige baut die Chemie auf zahllosen, jederzeit wiederholbaren experimentellen Erfahrungen auf, die in der Fachliteratur beschrieben sind. Diese Literatur zu erschließen ist deshalb besonders wichtig, und die Dokumentation in der Chemie ist entsprechend stark entwickelt.

Bei der Suche nach Literaturstellen helfen dem Chemiker zunächst die bekannten Handbücher (z. B. „Gmelin“, „Beilstein“, „Houben-Weyl“ usw.) und die Register der Referateorgane (Chemisches Zentralblatt, Chemical Abstracts). Aber je mehr der Umfang der Fachliteratur wächst, je genauer man deshalb die Fragestellung dem vorliegenden Problem anpassen muß, um so weniger genügen diese klassischen Hilfsmittel. Daher entstanden allenthalben umfangreiche Karteien. Sie wachsen beängstigend schnell, und der Arbeitsaufwand, neu veröffentlichtes Material zu klassifizieren und einzuordnen, wird sehr groß. Es liegt nahe, auch mechanische Hilfsmittel und Datenverarbeitungsanlagen einzusetzen.

Die besondere Schwierigkeit der Dokumentation besteht in der Menge des zu verarbeitenden Materials und in der Vielfalt der möglichen Fragestellungen; immerhin muß bei der Suche eine Intelligenzleistung simuliert werden. Ein gutes Dokumentationssystem sollte so flexibel sein, daß ganz neuartige Sachverhalte ohne Änderung des Schlüssels abrufbar gespeichert werden können. Darüber hinaus besteht der Wunsch, Fragen an den Speicher richten zu können, die bei der Einspeicherung noch nicht voraussehen waren. Fragen an den Speicher müssen so genau dem Problem gemäß formuliert werden können, daß möglichst wenig unzutreffendes Material mit herausfällt, andererseits aber keine wichtige Information verlorengeht. Schließlich dürfen die Kosten für Verschlüsselung, Einspeicherung und Recherche nicht zu hoch liegen.

Erst wenn alle diese Anforderungen weitgehend erfüllt sind, können mechanisierte Dokumentationssysteme auch bei sehr großen Informationsspeichern voll befriedigen. Die methodische und maschinelle Entwicklung auf diesem Gebiet befindet sich noch im Anfangsstadium. Erstes Ziel ist es deshalb, genügend anpassungsfähige Systeme zu schaffen. Im folgenden sollen zwei Verfahren beschrieben werden, deren praktische Be-

währung in größerem Rahmen noch aussteht. Eine Veröffentlichung in dieser Phase der Entwicklung soll dem Ideenaustausch und damit dem raschen Fortschritt dieser neuen Methodik dienen.

## 1. Dokumentation chemischer Strukturformeln

Ein Teilgebiet der Chemie-Dokumentation befaßt sich mit Sachverhalten, die durch Strukturformeln beschrieben werden können. Einerseits handelt es sich dabei um „harte“ Daten, deren Bedeutung eindeutig und scharf definiert ist; andererseits ist gerade hier die Zahl der Merkmale und ihrer möglichen Kombinationen besonders groß. Das rechtfertigt die Anwendung anderer Methoden als bei der Behandlung der nicht durch Formeln darstellbaren Sachverhalte. Beide Arten der Dokumentation müssen jedoch miteinander zu kombinieren sein, weil viele Fragen sich sowohl auf Strukturen als auch auf andere Merkmale beziehen.

Der Wunsch des Chemikers geht dahin, daß er jede beliebige Strukturformel oder Partialstruktur „anrufen“ kann. C. N. Mooers [1] wies bereits 1951 darauf hin, daß eine topologische Verschlüsselung [2] die Frage nach beliebigen Partialstrukturen erlaubt. Man numeriert bei dieser Verschlüsselung die Atome eines Moleküls in beliebiger Reihenfolge und gibt in der Chiffre für jedes Atom neben seiner Nummer und seinem Elementsymbol die Nummern der mit ihm verbundenen Nachbarn an (Abb. 1). Mehrfachbindungen und andere Besonderheiten können durch entsprechende Schlüsselzahlen in Verbindung mit den Nummern der betroffenen Atome verschlüsselt werden.

[1] C. N. Mooers, Zator Techn. Bull. 59, 1 (1951).

[2] Die topologische Darstellung eines aus Einzelementen bestehenden Systems (einer Strukturformel, eines Schaltbildes oder eines Fließschemas) soll im Gegensatz zur topographischen Wiedergabe, z. B. einer Konstruktionszeichnung, nicht die Lage aller Bausteine im zwei- oder dreidimensionalen Raum, sondern nur ihre gegenseitige funktionelle Verknüpfung in mathematisch-logischer Form festhalten.

Je nach der Reihenfolge der Numerierung entstehen äußerlich verschiedene Speichersätze; ihre Gleichwertigkeit kann aber durch zahlreiche Vergleiche von der Maschine festgestellt werden. Auch beliebige Partialstrukturen lassen sich auf diese Weise maschinell aufsuchen, indem man sie entsprechend verschlüsselt und vergleicht.

25344	007	019	+	017	028	018	+	018	017	019	031	+	019	007	018	031	020	+	020	019	+
25344	028	039	017	041	+	031	019	018	042	+	039	051	028	+	041	028	042	+	042	041	031
25344	+	051	039	063	+	063	051	076	+	065	076	066	+	066	065	079	+	076	063	065	089
25344	+	079	066	090	091	+	089	076	090	+	090	089	079	091	+	091	079	090	103	092	+
25344	092	091	+	103	091	+	-10	30	78	-37	51	-50	28	41	65	76					

Diese Verschlüsselung kann zwar von ungelernten Hilfskräften durchgeführt werden, ist aber so mühsam, daß sie für eine praktische Verwendung nicht in Betracht kommt. Erst seitdem es gelang, diese Arbeit weitgehend maschinell auszuführen, besteht die Aussicht, die Vorteile des topologischen Verfahrens für die Strukturformeldokumentation zu nutzen.

In der BASF wurde zunächst ein Verschlüsselungssystem entwickelt [3], das als Bausteine größere, oft vorkommende



1	O	2	—	—	—	1	N	2	—	—	—					
2	C	1	3	—	—	2	C	1	6	3	—					
3	C	2	4	7	—	3	C	2	4	—	—					
4	O	3	5	—	—	4	C	3	5	—	—					
5	C	4	6	—	—	5	O	4	6	—	—					
6	C	5	7	—	—	6	C	2	5	7	—					
7	C	6	3	8	—	7	C	6	8	—	—					
8	N	7	—	—	—	8	O	7	—	—	—					
9	R	—	—	—	—	9	R	—	—	—	—					
6	50	2	1	3	7	5	6	Doppelbindungen	6	50	2	6	4	3	7	8
5	70	9	3	4	5	6	7	Ringglieder	5	70	9	2	6	5	4	3

Abb. 1. Zwei gleichwertige topologische Speichersätze für die gleiche Strukturformel.

Gruppierungen (z. B. Anthrachinongerüste, Benzoylamino- und dgl.) verwendete; sie wurden einmal, sozusagen auf Vorrat, verschlüsselt und in einer Rechenanlage gespeichert. Wenn sie bei der Verschlüsselung verwendet wurden, so löste die Maschine sie in ihre Einzelatome auf und lieferte so den endgültigen Speichersatz. Dieses Verfahren befriedigte allerdings nicht ganz, vor allem, weil leicht Fehler unterlaufen können.

Deshalb wurde eine Maschine entwickelt, die von einer auf ein Rasterblatt gezeichneten Strukturformel (Abb. 2) die Bindestriche mit Photozellen abtastet und

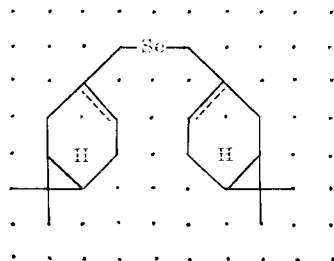


Abb. 2. Rasterblatt mit Strukturformel. Die roten Markierungen sind hier gestrichelt wiedergegeben.

[3] E. Meyer u. K. Wenke, Nachr. Dokumentat. 13, 13 (1962).

entsprechende Löcher in einen Lochstreifen stanzt. Heteroatome, Mehrfachbindungen und sonstige Besonderheiten sind mit roter Tinte gezeichnet, die von den Photozellen nicht registriert wird. Solche Strukturmerkmale werden in Form von Schlüsselzahlen mit den zugehörigen Rasterpunktnummern über eine Tastatur in den Lochstreifen gegeben.

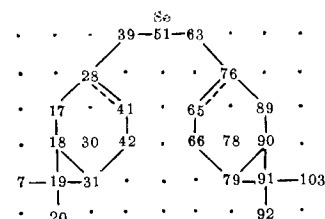


Abb. 3. Topologischer Speichersatz (oben, waagrecht von links nach rechts zu lesen) und zugehörige Formel mit Angabe der Rasterpunkt-Numerierung.

Aus diesem Lochstreifen erstellt eine elektronische Datenverarbeitungsanlage dann den topologischen Speichersatz (Abb. 3, oben) und druckt gleichzeitig ein Bild der Formel aus (Abb. 4), das eine Kontrolle der Verschlüsselung gestattet.

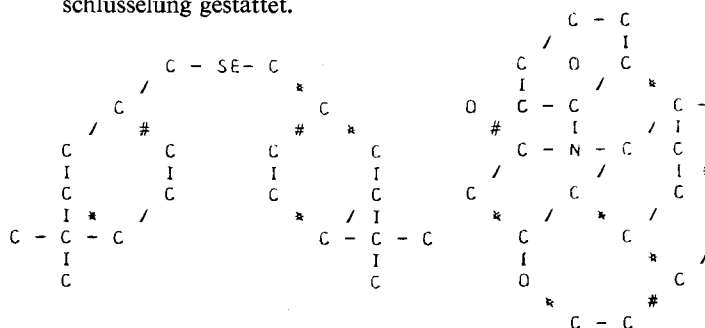
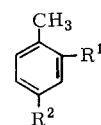


Abb. 4. Von der Rechenanlage aus den topologischen Speichersätzen rekonstruierte Strukturformeln (rechts: Strychnin).

Kompliziert wird das topologische Verfahren durch die Forderung der Chemiker, auch nicht genau definierte Strukturformeln (z. B. solche mit Substituenten an unbestimmter Stelle eines Ringes), Formeln mit wechselnden Substituenten (Markush-Formeln [4]) und sogar chemische Reaktionsschemata zu speichern und anzufordern. Aber auch diese Probleme können mit einigen Kunstgriffen grundsätzlich gelöst werden. Reaktionen z. B. werden verschlüsselt [4a], indem die transparenten Rasterblätter der Ausgangsstoffe und Endprodukte über-

[4] Markush setzte im US-Patentrecht die Zulässigkeit von Ansprüchen durch, die Stoffgruppen schützen. Nach ihm bezeichnet man daher als Markush-Formeln solche, die durch alternative Substituenten mehrere definierte Verbindungen darstellen, z.B.



$R^1 = \text{NH}_2, \text{NHCH}_3, \text{OCH}_3;$   
 $R^2 = \text{NO}_2, \text{CF}_3.$

[4a] E. Meyer, Nachr. Dokumentat. 13, 144 (1962); ADI Annual Meeting 1963, Short Papers, Part 2, S. 131.

einandergelegt, und die Strukturen so durchgepaust werden, daß identische Atome gleiche Rasterpunkt-nummern behalten.

In Abb. 5 ist als Beispiel die Reaktion einer Bromverbindung (Blatt A) mit Acetessigester (Blatt B) zum Keton (Blatt C) dargestellt. Die anderen Reaktionsprodukte (Äthanol und Kohlendioxyd) sind trivial und werden daher nicht aufgezeichnet. Eine spätere Anfrage kann dann beliebig variiert werden: Man kann nicht nur allgemein nach Synthesen mit Acetessigester fragen, sondern auch spezieller nach solchen mit Acetessigester und  $\alpha,\beta$ -ungesättigten Bromverbindungen oder, noch spezieller, mit Acetessigester und  $\alpha,\beta$ -ungesättigten, in  $\beta$ -Stellung verzweigten, eventuell sogar noch  $\varepsilon,\zeta$ -ungesättigten Bromverbindungen usw. Voraussetzung ist nur, daß mit den Bezeichnungen der einzelnen Rasterblätter (hier A, B und C) das Schema des Reaktionsablaufes eingetastet und im Speicher festgehalten wird. Auf diese Weise ist es später möglich, topologisch nicht nur nach allen denkbaren Partialstrukturen, sondern auch nach beliebigen Reaktionstypen zu fragen. Das ist wichtig, da die meisten, vom präparativ arbeitenden Chemiker gestellten Anfragen sich im Grunde auf Reaktionen beziehen.

Für das Problem der topologischen Verschlüsselung gibt es somit befriedigende Lösungen [5]. Es blieb nun die Frage offen, ob sich auch die Recherche mit wirt-

schaftlich tragbarem Aufwand durchführen läßt. Vorversuche auf einer IBM 704 zeigten, daß trotz der großen Geschwindigkeit dieser Maschinen die benötigte Zeit um etwa eine Größenordnung zu hoch liegt, wenn eine angefragte Strukturformel Atom für Atom mit *allen* gespeicherten Strukturformeln verglichen werden muß.

Wenn es jedoch gelänge, die Zahl der Strukturvergleiche für jede Anfrage auf etwa ein Prozent aller gespeicherten Strukturformeln zu beschränken, so dürfte auch die topologische Recherche durchaus mit annehmbarem Aufwand durchzuführen sein. Suchte man nach einzelnen Verbindungen, so böte das keine Schwierigkeit. Die Maschine prüfte zunächst Summenformel, Zahl der Ringe, Doppelbindungen usw. und müßte dann nur eine kleine Zahl von Isomeren auf topologische Übereinstimmung mit der angefragten Formel untersuchen. Aber derartige Recherchen lassen sich leichter nach alter Weise in einem Formelregister durchführen.

Wichtiger ist die Suche nach Verbindungsklassen, die durch Partialstrukturen beschrieben werden können. Bei ihnen kann man nur eine Mindestsummenformel angeben, und deshalb wird es notwendig, zur Vorprüfung der gespeicherten Strukturen weitere Kriterien heranzuziehen. Man kann z. B. dem topologischen Speichersatz Codezeichen voranstellen, die sich auf besonders charakteristische Gruppierungen in der Formel beziehen, etwa nach einem der gebräuchlichen Strukturformelcodes. Dies erforderte aber doppelte Verschlüsselungsarbeit und ist deshalb nur tragbar, wenn der Vorprüfungscode maschinell und automatisch aus dem topologischen Speichersatz der Formel erstellt werden kann. Maschinenprogramme für eine solche Umschlüsselung werden zur Zeit ausgearbeitet. Neue Methoden für den topologischen Vergleich sind auch an der Harvard-Universität entwickelt worden [9]. Sie erlauben in vielen Fällen einen wesentlich rascheren Strukturvergleich.

## 2. Dokumentation sonstiger Sachverhalte

Auch verfahrenstechnische, anwendungstechnische und andere Sachverhalte sollen möglichst vielseitig und maschinell abrufbar gespeichert werden. Dies geschieht bisher in der Regel nur durch reine Klassifizierung: für jedes Dokument (das heißt für jede Literaturstelle) werden aus Schlagworten oder Merkmalen, die in einem Fragebogen oder Schlüssel enthalten sind, die zutreffenden ausgewählt, gelocht und gespeichert. Dieses Verfahren hat den Nachteil, daß alle Begriffe, die nicht recht

[5] Inzwischen wurden auch in den USA zwei Verfahren entwickelt, die mit tragbarem Aufwand arbeiten. Das eine [6] scheint indessen keine Reaktionsverschlüsselung zu erlauben; das andere [7] erfordert beträchtlich mehr manuelle Arbeit, benötigt aber keine Spezialmaschine. Ferner ist neuerdings ein Verfahren [8] zur optischen Abtastung mit handelsüblichem Gerät (Fernsehkamera) in Entwicklung, bei dem allerdings der Datenverarbeitungsaufwand recht hoch ist.

[6] A. Feldman, D. B. Holland u. D. P. Jacobus, J. chem. Documentat. 3, 187 (1963).

[7] P. Horowitz u. E. M. Crane, Vortrag, ACS Meeting, 11. Sept. 1962.

[8] W. E. Cossum, M. E. Hardenbrook u. R. N. Wolfe, Proc. Amer. Documentat. Instit. 1964, 269.

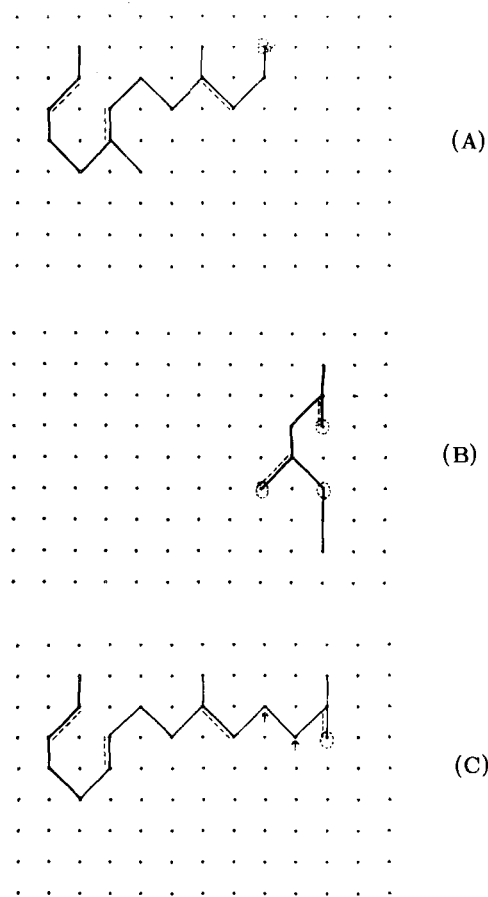


Abb. 5. Verschlüsselung von Reaktionen (vgl. Text). Man kann maschinell nach beliebig großen Partialstrukturen in Ausgangsmaterial und Endprodukt fragen. Bei topographischer Identität der einander entsprechenden Atome in den Rasterblättern erhält man als Antwort alle Reaktionen, die sich durch das vorgegebene Schema von Partialstrukturen beschreiben lassen.

in das Schema des Schlüssels passen – vor allem die ganz neuartigen und damit besonders interessanten – sich nicht befriedigend kennzeichnen lassen. In regelmäßigen Zeitabständen werden Änderungen und Ergänzungen des Schlüssels notwendig, dadurch wird der Speicher immer uneinheitlicher und die Suche immer schwieriger.

Für große Informationsspeicher muß die Analyse des Dokuments sehr tief gehen. Das heißt, nicht nur der Schlüssel muß sehr viele Klassifikationsmerkmale enthalten, sondern es müssen auch sehr viele Merkmale pro Dokument festgehalten werden. Neben der Erschwerung des Verschlüssels bringt dies vermehrten Ballast bei der Recherche, da sich mehr Möglichkeiten zu „fälschen“ Kombinationen der Schlüsselbegriffe ergeben und viele nicht im Dokument enthaltene Sachverhalte vor-täuschen.

Es liegt nun der Versuch nahe, in der mechanisierten Dokumentation von der reinen Klassifikation abzu-gehen und – ähnlich wie in der Sprache – maschinell abfragbare, syntaktische Beziehungen zwischen den verwendeten Schlüsselbegriffen herzustellen. Sie würden gleichzeitig die Bildung zusammengesetzter Begriffe und damit eine beträchtliche Verkleinerung des Wortschatzes oder Schlüssels ermöglichen. Als erste Ansätze in dieser Richtung sind die gelegentlich verwendeten „links“ und „role indicators“ [9a] zu werten. Da der Dokumenta-tionsspeicher nicht zur ausführlichen Darstellung, son-dern nur zum sicheren Wiederauffinden von Sachver-halten dienen soll, können Wortschatz und Syntax der „Maschinensprache“ viel kleiner gehalten werden als in natürlichen Sprachen. Weil andererseits aber auch der Aufwand für die Suche nicht zu groß werden darf, müs-sen Vokabular und syntaktische Beziehungen viel strenger normiert werden als im täglichen Sprachge-brauch. Weiter sollte sich die Codierung chemischer Verbindungen (auch die topologische!) in dieses System einfügen lassen, damit man maschinell auch nach Struk-turmerkmalen und Reaktionstypen im Rahmen kom-plexer Sachverhalte suchen kann. Die wichtigste For-derung aber bleibt, daß der Arbeitsaufwand für die Ver-schlüsselung nur unwesentlich wachsen darf. Im folgen-den soll ein Verfahren angegeben werden, das diesen Wünschen weitgehend Rechnung trägt.

Bei der Verschlüsselung werden zunächst die einzelnen Schlüsselbegriffe aus dem Dokument ausgewählt und in beliebiger Reihenfolge registriert, und zwar in Form ihrer Notationen [9b] oder – bei chemischen Verbin-dungen – ihrer (topologischen) Speichersätze. Die Notationen des Schlüssels sollten klassifikatorische Merkmale enthalten, aus denen zugehörige Oberbegriffe maschinell erkennbar sind; bei den Speichersätzen ist dies ohnehin der Fall.

[9] G. Salton u. E. W. Sussenguth jr., ADI Annual Meeting 1963, Short Papers, Part 2, S. 143; Scientific Report No. ISR-6 to The National Science Foundation, April 1964.

[9a] „Links“ sind gleichartige Codezeichen, die an zwei oder mehr Schlagwörter oder Begriffsnotationen angehängt werden und dadurch eine besonders enge Beziehung zwischen ihnen herstellen. „Role indicators“ geben an, in welchem Sinne oder Zusammenhang die zugehörigen Begriffe gebraucht sind, d. h. welche Rolle sie im Sachverhalt spielen.

[9b] Notation ist hier eine Folge von Druckzeichen als Symbol für einen Begriff.

Bei dieser Zusammenstellung wird, möglichst fort-laufend, jedem Begriff eine dokument-interne Ordnungs-zahl zugeteilt, die der Bearbeiter gleichzeitig mit Blei-oder Rotstift im Dokument an die zugehörigen Stich-wörter oder Formeln schreibt. Sodann werden „Aus-sagesätze“ oder „Aussagechiffren“ zusammengestellt, die die wesentlichen Sachverhalte des Dokuments ein-deutig rekonstruierbar beschreiben. Sie bestehen aus den entsprechenden Ordnungszahlen, die durch besondere Funktionszeichen miteinander syntaktisch verknüpft sind, ähnlich wie man in einem mathematischen Aus-druck die einzelnen Größen durch Operatoren verbindet. Als Funktionszeichen kommen im Fachgebiet Chemie vor allem die in Tabelle 1 genannten Buchstaben und Sonderzeichen in Betracht.

Tabelle 1. Syntaktische Operatoren.

Symbol und Umschreibung	Erläuterung
A „als“	(adjektivische Beziehung, physikalischer Zu-stand), z. B. „Oxydation <i>in der</i> Gasphase“, „azeotropes... Gemisch“. Die Nummer des Hauptbegriffs steht vor, die des erläuternden hinter dem Operator A.
C „mit“	(gleichwertige Stoffe oder Faktoren), z. B. Um-setzung von „Carbonsäuren <i>mit</i> Olefinen“.
E „mit oder ohne“	(Stoffe oder Faktoren, die vorhanden sein sollten oder dürfen, aber nicht müssen), z. B. Färben von „Polyesterfasern <i>mit oder ohne</i> Wolle-Beimischung“.
G „von“	(Beziehung zwischen Vorgang und Substrat), z. B. „Veresterung <i>von</i> Carbonsäuren“.
J „aus“	(Beziehung zwischen Endprodukt und Ausgangs-material), z. B. „Aromaten <i>aus</i> Cycloaliphaten“.
L „durch“	„in Gegenwart von“ (Hilfsstoffe, Apparate), z. B. „Hydrierung <i>mit</i> Raney-Nickel“, „Zerstäuben <i>durch</i> eine Düse“.
N „für“	(Anwendung, Verwendung, Zweck), z. B. „Phenole <i>als</i> Stabilisatoren“.
P „bei“	(quantitative Angabe von Reaktions-bedingungen), z. B. „Decarboxylierung <i>bei</i> 120 bis 150 °C“.
R „dann“, „bevor“	(zeitliche Reihenfolge), z. B. „Destillation <i>mit anschließender</i> Trocknung“.
v „oder“	(Disjunktion) verbindet Alternativen, z. B. „mit Phosgen <i>oder</i> Thionylchlorid“.
– „nicht“	„in Abwesenheit von“ (Negation); steht vor der entsprechenden Ordnungszahl und wird zusätzlich zu einem der Zeichen A...R benutzt.
( )	(Klammern) dienen – ähnlich wie in mathe-matischen Ausdrücken – zur engeren Verbin-dung oder Bildung von Untergruppen im Aussagesatz.
• „Punkt“	(Satzende); beendet einen Aussagesatz, mit dem ein Sachverhalt vollständig beschrieben wird.

Das Maschinenprogramm kann so gestaltet werden, daß bei der Recherche das Vorkommen von Begriffen mit bestimmten Operatoren (z. B. v, A, E, N, P und R) im Speicher unschädlich ist, auch wenn sie in der Frage nicht genannt werden, daß aber die Verknüpfung mit anderen Operatoren in der Frage ausdrücklich verlangt oder erlaubt werden muß. Auf die Weise erreicht man, daß in bestimmten Fällen automatisch auch Dokumente gefunden werden, in denen der Sachverhalt ausführ-licher beschrieben und verschlüsselt ist als in der Frage. Die Verwendung von Klammern und Disjunktions-

Operatoren („v“) in den Aussagechiffren erlaubt die Zusammenfassung von komplexen Beziehungen und gewährt dem Bearbeiter weitgehende Freiheit bei ihrer Darstellung, so daß er nicht zu sehr mit der Wahl der richtigen Formulierung belastet wird. Diese Zusammenfassungen werden bei der Recherche (unter Umständen auch schon bei der Einspeicherung) von der Maschine nach den Regeln der Booleschen Algebra [10] aufgelöst. Um zur Herstellung eindeutiger Beziehungen nicht zu viele Klammern verwenden zu müssen, wird eine Rangfolge der Operationszeichen festgelegt: Der im Alphabet vorangehende Buchstabe bindet jeweils enger als der folgende (ähnlich wie in der Algebra das Multiplikationszeichen dem Additionszeichen übergeordnet ist).

An einem Beispiel sei das Verfahren kurz erläutert. Zu verschlüsseln sei das Referat [10a]:

„Katalytische ⑥ Alkylierung ⑤ eines Isoparaffins ② mit einem Olefin ③. Man bringt das im Überschuß ① eingesetzte Isoparaffin, z. B. Isobutan ②, mit dem Olefin ③ in Anwesenheit eines sauren ⑧ Katalysators ④, wie Schwefelsäure ④, in einer Reaktionszone in Berührung. Das aus dem Katalysator, überschüssigem Isoparaffin und dem Reaktionsprodukt bestehende Gemisch wird aus der Reaktionszone abgezogen. Der Katalysator ④ wird praktisch vollständig aus dem Gemisch entfernt ⑨. Die zurückbleibende Kohlenwasserstoffphase bringt man in eine Fraktionierungszone. Die erfindungsgemäße Verbesserung des Verfahrens besteht darin, daß man die Kohlenwasserstoffphase, bevor man sie der Fraktionierungszone zuführt, wenigstens zum größten Teil einer ⑦ Blitzverdampfung ⑥ unterwirft. Die dabei freiwerdenden Isoparaffindämpfe ② werden kondensiert ⑩, und das Kondensat wird in die Reaktionszone zurückgeführt“ ⑩.

Die Begriffsliste für die Verschlüsselung sieht dann etwa folgendermaßen aus:

- 1 ..... (Notation für „Überschuß“)
- 2 ..... (Notation oder Speichersatz für „Isobutan“, Oberbegriff: „Isoparaffin“)
- 3 ..... (Notation oder Speichersatz für „Olefine“)
- 4 ..... (Notation oder Speichersatz für „Schwefelsäure“)
- 5 ..... (Notation für „Alkylierung“)
- 6 ..... (Notation für „Katalysator, Katalyse“)
- 7 ..... (Notation für „schnell“)
- 8 ..... (Notation für „Verdampfen, Verdampfung“)
- 9 ..... (Notation für „Kondensation“)
- 10 ..... (Notation für „Rückführung“)
- 11 ..... (Notation für „Fraktionierung“)
- 12 ..... (Notation für „Entfernung, Abtrennung“)
- 13 ..... (Notation für „Säure, sauer“)

Die zugehörigen Aussagesätze lauten:

5 A 6 G 2 C 3.

2 A 1 C 3 L (4 v 6 A 13) R 12 G (4 v 6) R 8 A 7 R 11 R 9 G 2 R 10 G 2.

Ähnlich, wie man aus der topologischen Verschlüsselung die Strukturformel rekonstruieren kann, läßt sich – wenn auch etwas weniger genau – aus diesen Chiffren und der Begriffsliste das Referat rekonstruieren. Es würde dann lauten:

„Katalytische Alkylierung von Isobutan mit Olefinen. Überschüssiges Isobutan mit Olefin (umgesetzt) in Gegenwart von Schwefelsäure oder einem (anderen) sauren Katalysator; dann Abtrennung der Schwefelsäure (bzw.) des Katalysators; dann schnelle Verdampfung; dann Fraktionierung; dann Kondensation von Isobutan; dann Rückführung von Isobutan“.

[10] Boolesche Algebra ist eine Methode, logische Zusammenhänge („und“, „oder“, „nicht“) durch mathematisch exakte Formeln auszudrücken. Nach ihren Regeln kann man diese Formeln umformen, ähnlich wie man in der normalen Algebra mit Gleichungen rechnet.

[10a] Die Ziffern neben den Stichworten sind nicht im Referat enthalten, sondern erst vom Bearbeiter hingeschrieben.

Wie schon erwähnt, sollen die Notationen eine begriffliche Hierarchie enthalten. Würde man z. B. dem Begriff „Säuren“ die Notation „47“ zuteilen, so könnte man „Mineralsäuren“ durch „47.1“ und „Schwefelsäure“ durch „47.1.1“ darstellen. Fragt man später nach 47, d. h. nach Säuren allgemein, so fällt die Literaturstelle, in der die Schwefelsäure genannt ist, automatisch mit heraus. Ähnliches gilt z. B. für die Reihe Isobutan – Isoparaffine – Paraffine. Gibt der Speicher auf eine Anfrage keine Antwort, so kann man die Fragestellung schrittweise verallgemeinern, bis entsprechende Literaturstellen gefunden werden.

Weil somit in der Chiffre alle wesentlichen Sachverhalte vollständig und eindeutig enthalten sind, lassen sich – ähnlich wie man aus dem topologischen Speichersatz beliebige Partialstrukturen abfragen kann – aus dem so verschlüsselten Referat auch alle implizit vorhandenen Teilsachverhalte maschinell abrufen. Freilich wird das dafür notwendige Maschinenprogramm beträchtlichen Umfang haben; die Rechenzeit aber dürfte nicht sonderlich lang sein, weil eine Prüfung der Begriffsliste auf die in der Frage geforderten Notationen vorausgeht und sich die Prüfung des Aussagesatzes bei der Mehrzahl der Referate erübrigt.

### 3. Belieferung anderer Karteien

Ein entscheidender Nachteil der beschriebenen Verfahren – wie überhaupt der mechanisierten Dokumentation mit sehr großen Informationsspeichern – liegt darin, daß man auf die Benutzung von Lochkartenmaschinen oder elektronischen Datenverarbeitungsanlagen angewiesen ist. Damit ist der Zugriff erschwert, denn auch einfache Anfragen muß man erst in die Maschinensprache übersetzen und möglicherweise lochen lassen. Die Antwort liegt dann frühestens nach dem nächsten Maschinenlauf vor. Wegen dieser Schwierigkeiten unterbleiben sicher viele, an sich interessante Anfragen, die mit einer kleinen, dem Benutzer besser zugänglichen Handkartei ohne weiteres beantwortet werden könnten.

Derartige Handkarteien werden freilich zu groß und unhandlich, wenn sie sich nicht auf ausgewählte Spezialgebiete beschränken. Sie sind nur dann wirklich brauchbar, wenn sie einen auf ihr jeweiliges Sachgebiet ausgerichteten eigenen Schlüssel verwenden. Daran scheiterte bisher ihre Koordinierung und zentrale Belieferung mit fertig verschlüsseltem Material. Die elektronische Datenverarbeitung kann auch hier einen Wandel schaffen. Wenn nämlich alle von einem Kunden in seinem Schlüssel benutzten Merkmale durch Einzelanfragen in einem elektronischen Großspeicher abgefragt werden können, so kann man die dabei als Antwort herausgegebenen Nummern der Dokumente auf eine Sichtlochkarte lochen. Nach einem entsprechenden Programm liefert die Rechenanlage zunächst Hollerithkarten mit beispielsweise 500 Lochstellen für Dokumentnummern. Jedes Schlüsselmerkmal erhält für je 500 Literaturstellen eine solche Karte, in der die Lochstellen derjenigen Dokumente ausgestanzt sind, die dieses Merkmal aufweisen: Für die nächsten 500 Dokumente wird ein neuer Satz Merkmalskarten angelegt und so fort.

Man benutzt bei der Recherche zunächst diese Hollerithkarten als Sichtlochkarten, d. h. man sucht aus

jedem Satz die Karten der gerade benötigten Merkmale heraus, legt sie übereinander und hält sie gegen das Licht: Die Lochstellen der gesuchten Dokumente scheinen dann durch. Wenn 5000 Dokumente bearbeitet, also zehn Sätze solcher Hollerithkarten beisammen sind, werden sie zur bequemen Handhabung bei späteren Anfragen auf einen einzigen Satz normaler Sichtlochkarten kumuliert, was mechanisch mit Spezialmaschinen in Lohnarbeit durchgeführt werden kann.

Auf diese Weise könnte man die hohen Kosten für die selbständige Bearbeitung und Verschlüsselung der Fachliteratur bei zahllosen kleineren Karteien entscheidend reduzieren. Bedingung dafür ist neben dem Vorliegen entsprechender Maschinenprogramme ein großer mechanisierter Informationsspeicher mit der Möglichkeit zu äußerst vielseitiger Fragestellung.

Eingegangen am 19. November 1964 [A 432]

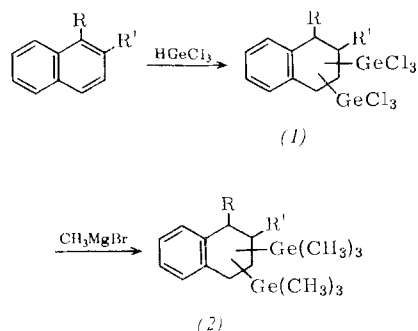
## ZUSCHRIFTEN

### Hydrogermylierung aromatischer Verbindungen

Von S. P. Kolesnikow und Dr. O. M. Nefedow

N.D.Zelinskij-Institut für Organische Chemie der Akademie der Wissenschaften der UdSSR, Moskau (UdSSR)

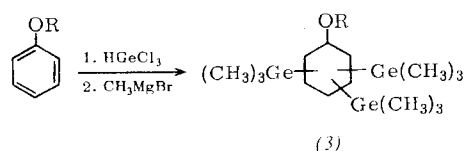
Erhitzt man 0,05 Mol Naphthalin mit 0,05 Mol  $\text{HGeCl}_3$  10 bis 15 min auf 110–130°C und methyliert das Gemisch anschließend mit überschüssigem  $\text{CH}_3\text{MgBr}$  in Äther, so entsteht mit 52,5-proz. Ausbeute 1.3-Bis-(trimethylgermyl)-1.2.3.4-tetrahydronaphthalin (2),  $\text{R} = \text{R}' = \text{H}$ ,  $\text{Kp} = 100$  bis 102°C/0,2 Torr.



Die Struktur dieses Produktes wird durch sein IR-Spektrum bewiesen. Das NMR-Spektrum (60 MHz, Tetramethylsilan als Bezugssubstanz) enthält Signale bei 0,06 und 0,12 ppm ( $\text{CH}_3$ -Gruppen am Ge), bei 2,75 und 2,49–2,60 ppm (nicht-aromatische CH-Gruppen), bei 6,86 und 6,94 ppm (aromatische Protonen) sowie zwei Multipletts bei 1,90 und 2,58 ppm ( $\text{CH}_2$ -Gruppen). Bei der Selen-Dehydrierung (290–300°C, 3 Std.) oder beim Kochen mit Chloranil in Xylol (10 Std.) bildet sich aus (2),  $\text{R} = \text{R}' = \text{H}$ , mit 65–80-proz. Ausbeute Naphthalin zurück.

Außer (1),  $\text{R} = \text{R}' = \text{H}$ , entsteht bei der Reaktion von Naphthalin mit  $\text{HGeCl}_3$  auch ein Pentakis-(trichlorgermyl)-decahydronaphthalin mit 10–20-proz. Ausbeute.

Die Hydrogermylierung von 1-Methylnaphthalin unter gleichen Bedingungen ergibt bereits nach 5 min ein Bis-(trichlorgermyl)-Derivat, das sich zur Bis-(trimethylgermyl)-Verbindung ( $\text{Kp} = 101$ –104°C/0,2 Torr) methylieren läßt. Gesamtausbeute: 62%. Ähnlich leicht verläuft die Anlagerung von  $\text{HGeCl}_3$  an Dimethylnaphthaline sowie an Anthracen, Phenanthren und ihre Derivate.



Mit Benzol, Toluol und Xylenen reagiert  $\text{HGeCl}_3$  unter diesen Bedingungen nicht; dagegen entstehen aus Alkoxybenzolen Addukte. So liefert 8,5-stündiges Erhitzen von Methoxybenzol mit äquimolaren Mengen  $\text{HGeCl}_3$  auf etwa 110°C und anschließende Methylierung ein Gemisch isomerer Methoxy-tris-(trimethylgermyl)-cyclohexane (3),  $\text{R} = \text{CH}_3$ . Mit 32-proz. Ausbeute isoliert man aus diesem Gemisch durch Destillation und Ausfrieren das 1-Methoxy-1.3.5-tris-(trimethylgermyl)-cyclohexan,  $\text{Fp} = 66^\circ\text{C}$ ,  $\text{Kp} = 105$ –107°C/0,6 Torr. Das IR-Spektrum dieser Verbindung enthält die für  $(\text{CH}_3)_3\text{Ge}$ - und  $\text{CH}_3\text{O}$ -Gruppen charakteristischen Banden. Im NMR-Spektrum findet man zwei Singulets bei 0,07 und 0,13 ppm (2:1,  $\text{CH}_3$ -Gruppen an Ge), ein Multiplett bei 1,0–2,0 ppm ( $\text{CH}_2$ - und CH-Gruppen) und ein Singulett bei 3,14 ppm ( $\text{CH}_3\text{O}$ -Gruppen).

Durch Erhitzen von Äthoxybenzol mit  $\text{HGeCl}_3$  (30 min, 110–130°C) und anschließende Methylierung entstehen die (3) entsprechenden Äthoxyderivate.

Es ist bemerkenswert, daß aus Alkoxybenzol und  $\text{HGeCl}_3$  kein Trichlorgermylbenzol entsteht, obwohl man beim Kochen von Dialkyläthern mit  $\text{HGeCl}_3$  Alkyltrichlorgermane mit Ausbeuten von 35–40% erhält [1].

Eingegangen am 11. Dezember 1964, ergänzt am 26. Januar 1965

[Z 905]

[1] O. M. Nefedow, S. P. Kolesnikow u. W. J. Schejtschenko, Angew. Chem. 76, 498 (1964); Angew. Chem. internat. Edit. 3, 508 (1964); Ber. Akad. Wiss. UdSSR, im Druck.

### 1.2.3.4-Tetrachlor-7.8-benzosquisfulvalen [1]

Von Prof. Dr. Y. Kitahara, Dr. I. Murata und Dr. S. Katagiri

Chemisches Institut der Tohoku Universität, Sendai (Japan)

Wir fanden, daß die Umsetzung von Tropon mit 1.2.3.4-Tetrachlorcyclopentadien in Methanol 1.2.3-Trichlor-6-(o-methoxyphenyl)-fulven und das p-Isomer liefert [2]. Die Bildung von Fulvenderivaten kann vermieden werden, wenn man 4.5-Benzotropen verwendet: Die Reaktion von 3,12 g (0,02 Mol) 4.5-Benzotropen und 4,08 g (0,02 Mol) 1.2.3.4-Tetrachlorcyclopentadien in 150 ml Methanol bei Raumtemperatur führt in ca. 44 Std. zu 1.2.3.4-Tetrachlor-7.8-benzosquisfulvalen (1) mit einer Maximalausbeute von 15% (0,51 g).

