ring index names is a longer range goal; it now appears that such development will require considerable simplification and standardization of the present international systems for naming ring structures.

Problems of ring names permeate almost the whole of chemical nomenclature and are compounded by the proliferation of ring-naming systems over the years. There are at present at least seven approved IUPAC systems for naming rings, and four more either are partially accepted or are being considered for acceptance. The use of these systems yields many approved names for the same ring or for combinations of one ring with other rings. An example is seen in Figure 4 which shows the variations in naming the pyridine ring as it occurs in combination with other rings. It is presently impossible to search *CA*

Nomenclature options for ring systems containing the following:

| | |
|---|---|
| Piperidine | Acridine |
| Pyridine | Phenanthridine |
| Indolizine | Quindoline |
| Pyrindine (2 isomers) | Quinindoline |
| Isoquinoline | Thebenidine |
| Naphthyridine (6 isomers) | Acrindoline |
| Quinoline | Numerous "indicated hydrogen" forms |
| Quinolizine | Various von Baeyer structures, e.g., |
| Quinuclidine | 1-Azabicyclo[4.1.0]heptane |

Figure 4. Ring nomenclature

| Number of Rings | Number of Basic Ring Systems | Percentage of Sample |
|---|---|---|
| 1 | 1,661 | 6.60 |
| 2 | 3,653 | 14.53 |
| 3 | 6,070 | 24.15 |
| 4 | 4,836 | 19.24 |
| 5 | 4,000 | 15.91 |
| 6 | 2,230 | 8.87 |
| 7 | 1,180 | 4.69 |
| more than 7 | 1,502 | 6.01 |

Figure 5. Ring system frequency analysis

Subject Indexes for all forms of such rings. They can be found by careful, time-consuming search of the Index of Ring Systems.

There are now ring frequency analyses which show the number of rings in the total collection of parent rings contained in CAS files; as shown in Figure 5, three-membered rings comprise the largest percentage of the known basic ring structures. Of the individual rings, the six-carbon ring has the highest frequency of occurrence. The pyridine ring, used as the example in Figure 4, is the heterocyclic ring that occurs most frequently.

Thus, the Chemical Registry System serves as a vocabulary control system which not only deals with chemical names and the corresponding structural diagrams interchangeably but also assures that a given substance is always represented in the *CA* Volume Subject Indexes by the same index name; as pointed out earlier, statistics such as those above on the content of the CAS indexes are available only because of the computer data base.

# A Multilingual Index Via the Multiterm System*

HERMAN SKOLNIK
Hercules Incorporated, Research Center, Wilmington, Del. 19899

Received February 16, 1972

**The Multiterm system, which was introduced by the author in 1970 as a new indexing method, is shown to be a unique concept for producing a multilingual index via computer processing. An inherent advantage of the multilingual Multiterm index is that on reaching a certain size, the translating of terms from a source language to others becomes essentially a computer operation. The new concept has significance for international cooperative programs.**

There has been considerable interest in the potential for worldwide information systems, particularly for various disciplines of science and areas of technology. Most recently, this interest was expressed in some detail in the Unesco publication "UNISIST," with many cogent arguments.[1]

Despite the apparent feasibility documented for a viable international information system, it is highly unlikely that real progress will be made towards its realization until solutions are found to two basic problems:

1. The linguistic barrier.
2. A universally accepted system for indexing.

The objective of the work described in this paper was to find solutions to these two problems. Input and output character limitation of our computer restricted the work to languages based on the Latin alphabet.

## THE MULTITERM INDEX

The Multiterm index is a system for communicating the informational content of documents by coordination of subject terms in defined directional orders.[2] It was conceived

*Presented before Middle Atlantic Regional Meeting, ACS, Philadelphia, Pa., Feb. 16, 1972. Hercules Research Center Contribution No. 1584.

and developed for processing with and communication by computer operations. In its most basic form, a Multiterm is a combination of associated subject terms, such as

C/R/P/A/ / (I)

in which C is a chemical prepared from reactant R by process P using catalyst A. The oblique stroke or virgule following each term is the designation in the computer program statements to initiate a wrap-around process whereby the following Multiterms are stored in addition to the input Multiterm:

R/P/A/ /C/ (Ia)

P/A/ /C/R/ (Ib)

A/ /C/R/P/ (Ic)

The double virgule marks the end of the input Multiterm, and indicates in each of the computer-processed Multiterms the logical order of the associated terms.

All documents are not amenable to the logical order expressed in Multiterm I, even in chemistry. They are, however, indexable within the logical order of most generic to most specific, of which Multiterm I is a subset.

For example, a document concerned with the study of the structure of graphite fibers using an X-ray method would be assigned the following Multiterm:

Fiber:Graphite-Q/Structure-D/Test Method-U/
X-Ray-U/ / (II)

A Multiterm in the logical order of most generic to most specific makes sense when read backwards—e.g., for Multiterm II: the use of X-rays as a test method for determining the structure of graphite fibers. The use of the letters, Q, D, and U, will be described later.

Just as the generic:specific relationship imparts meaning and consistency in a Multiterm, the generic:specific relationship also imparts meaning and consistency in a subject term, such as Fiber:Graphite. The generic:specific relationship of a subject term allows the searcher for information to browse among other generic:specific terms should he be interested in the generic field, such as all information on fiber-reinforced composite materials, which he would find under the term Fiber:X/.

Comparison of the Multiterm assigned to a document with the title of a document is quite revealing of the informational content of the Multiterm *vs.* the title.[3] For example, the title of the document of Multiterm II is: "Structural and Orientational Aspects of Graphitized Fibers." Another revealing exercise is to compare the informational content of the Multiterm with descriptors (usually Uniterms) assigned by indexers using a thesaurus and keywords assigned by the author.[3]

As originally conceived and implemented, a Multiterm was limited to 68 characters (for printout on 8½ × 11 paper).[3] This limitation has been eliminated by a new program and by a format change for both input and output. To illustrate, a paper entitled "The Oxychlorination of Ethylene at High Temperatures,"[5] was assigned the following Multiterm:

Vinyl Chloride-P/Ethylene-R/Oxychlorination/
Catalyst:Celite + MnX-U/Mechanism/D Exchange/ /
Ref: *J. Org. Chem.* 34, 271-3 (1969) (III)

which contains 88 characters (letters + spaces + characters), and which via the wrap-around program yields a total of six Multiterms in the printout in the same format as III. The amount of information in Multiterm III is equivalent to the following abstract:

> The mechanism of the oxychlorination of ethylene to vinyl chloride in the presence of Celite with a manganese compound as catalyst was studied by means of deuterium exchange.

There are 171 characters in this abstract, or about 75% more than in Multiterm III. There are nine keywords or Uniterms in the abstract, whereas there are six terms in Multiterm III, with the same informational content and retrieval power; on the other hand, there are three keywords in the title, but only two (oxychlorination and ethylene) are relevant for retrieval within a highly restrictive limitation relative to the total number of terms by which readers may want to retrieve the document.

As in any good indexing system, crosses, such as SEE and SEE ALSO, are used when and where necessary in the Multiterm system. For example, Multiterm III requires the following crosses:

Iron Oxide as Catalyst—See Catalyst:Iron Oxide
Celite—See Also Iron Oxide
Catalyst:Iron Oxide—See Also Catalyst:Celite
Deuterium Exchange—See D Exchange

Consistency of index terms within Multiterms is maintained preferably by principles and rules for categories of concepts rather than through reliance on a thesaurus or subject authority list. An important aid to consistency is the use of the generic:specific relationship with subject terms, as illustrated for catalysts, polymers, and stability in a previous paper.[2]

The Multiterm system is computer processable via operations which are common to essentially all computers. It is particularly amenable to the philosophy, principles, and rules of indexing which are important to any system designed to play a role in the understanding and communication of a broad spectrum of information within the world of science and technology.[4]

## MULTILINGUAL-MULTITERM INDEXING

Many journals display titles and abstracts of their articles in two, three, and more languages. A considerably smaller number of journals are produced in more than one language, such as *Angewandte Chemie* (German and English), and various Russian journals (via translation to English by others than the publisher).

Quite obviously, journals with multilingual titles and abstracts are candidates for keyword indexing. Although the indexing operation via computer would be economical, the need for thesaurus control increases computer costs appreciably and the retrieval power of the product leaves much to be desired. Furthermore, too few journals comprise the data base to make the operation feasible without a heavy investment in translations.

Whereas translating titles and abstracts requires a high level knowledge of grammar, translating single words or combinations of a few words that denote a subject is considerably less demanding of the translator. Consider a relatively simple title of a paper in English, "The Mechanism of Polyethylene Degradation," which was assigned the following Multiterm:

Polymer:Ethylene-Q/Degradation/Mechanism/ / (IV)

It is relatively easy to translate IV into French (IVF) and into German (IVG)

Polymer: Ethylene-Q/Degradation/Mecanisme/ / (IVF)

Polymerisat: Aethylen-Q/Abbau Mechanismus/ / (IVG)

Considerably more knowledge and skill, however, are required for translating the title:

"Mecanisme de degradation de polyethylene" (French)

"Abbaumechanismus von Polyaethylen" (German)

Because of the limitation of the print train for characters, such as accent and diacritical marks, we ignore such marks in French words and replace German umlauted letters with an e following the vowel

Another example is that of a paper[6] entitled "Actual Optimization Problems of the Reactors Used for Ammonia Synthesis" (English) and "Problèmes actuels dans l'optimalisation des réacteurs pour la synthèse de l'ammoniac" (French). We assigned the following Multiterm to this paper:

Ammonia-P/Reaction Conditions-E/Reactor-Design/ Optimization/Computer-U/ / (VE)

Ammoniac-P/Reaction, Conditions de-E/Reacteur-Plan/ Optimisation/Ordinateur-U/ / (VF)

Ammoniak-H/Reaktionsbedingungen-E/ Reaktionsgefaess-Entwurf/Optimierung/ Rechenanlage-V/ / (VG)

Ref: *Chim. Ind., Genie Chim.* (1971), 104(8), 1002–7

In each of the three languages, VE, VF, or VG, the Multiterm contains the following information: use of a computer for optimizing the design of a reactor for the synthesis of ammonia on the basis of the reaction conditions. This is, indeed, considerable information—far more than the author put into the title, and far more accurate. Furthermore, the translation of the terms in the Multiterm of the author's language (the source language) to a Multiterm in another language can be done reproducibly by anyone with a fair familiarity with the second language. Translation of titles, however, requires a relatively good knowledge of the two languages, and it is a considerably greater task than a word-for-word translation.

Multiterm III has the following input format for English, French, and German:

Vinyl Chloride-P/Ethylene-R/Oxychlorination/ Catalyst: Celite + MnX-U/Mechanism/D Exchange/ / (IIIE)

Vinyle, Chlorure de-P/Ethylene-R/Oxychloruration/ Catalyseur: Celite + MnX-U/Mecanisme/ D, Echange de / / (IIIF)

Vinylchlorid-H/Aethylen-R/Oxychlorierung/ Katalysator: Celite + MnX-V/Mechanismus/ D Austausch/ / (IIIG)

Ref: *J. Org. Chem.* 34, 271–3 (1969)

This input on computer processing of IIIE [wrap-around of IIIE to produce as many Multiterms as there are terms (six in IIIE), alphabetical sorting and merging with other Multiterms already in the system, and printout] yields an output that has the following format (the triple dots between Multiterms represent a multitude of Multiterms in the total printout of the index):

⋮

Catalyst: Celite + MnX-U/Mechanism/D Exchange/ / Vinyl Chloride-P/Ethylene-R/Oxychlorination/ Ref: *J. Org. Chem.* 34, 271–3 (1969)

⋮

D Exchange/ /Vinyl Chloride-P/Ethylene-R/ Oxychlorination/Catalyst: Celite + MnX-U/Mechanism/ Ref: *J. Org. Chem.* 34, 271–3 (1969)

⋮

Ethylene-R/Oxychlorination/Catalyst: Celite + MnX-U/ Mechanism/D Exchange/ /Vinyl Chloride-P/ Ref: *J. Org. Chem.* 34, 271–3 (1969)

⋮

Mechanism/D Exchange/ /Vinyl Chloride-P/Ethylene-R/ Oxychlorination/Catalyst: Celite + MnX-U/ Ref: *J. Org. Chem.* 34, 271–3 (1969)

⋮

Oxychlorination/Catalyst: Celite + MnX-U/Mechanism/ D Exchange/ /Vinyl Chloride-P/Ethylene-R/ Ref: *J. Org. Chem.* 34, 271–3 (1969)

⋮

Vinyl Chloride-P/Ethylene-R/Oxychlorination/ Catalyst: Celite + MnX-U/Mechanism/D Exchange/ / Ref: *J. Org. Chem.* 34, 271–3 (1969)

⋮

The index in French and in German has the same output format via computer processing of IIIF and IIIG, respectively.

## ROLE INDICATORS OR SUBJECT PROCESS TERMS

Computer processable systems need to be designed to keep the storage requirements to a minimum and to use the least read/writes in manipulating information. These design features allow the computer to operate at optimum speeds and at the least cost. Consequently, we use a series of letters following a subject to indicate actions on or by the subject. These actions are known generally as role indicators or process terms—such as analysis, determination, preparation, reaction, and use. Table I lists the mnemonic abbreviations of the several role indicators we have used in English and those we recommend for French and German.

In addition to making computer operations more economical, role indicators also impart a considerably higher degree of information content to Multiterms at almost no cost of space, and make reading of Multiterms essentially similar to reading expanded titles or compressed abstracts.

Table I. Role Indicator Symbols Used in Multiterms

| Role | Symbol | | |
|---|---|---|---|
| | English | French | German |
| Analysis | A | A (Analyse) | A (Analyse) |
| Composition | C | C (Composition) | Z (Zusammensetzung) |
| Determination | D | D (Determination) | B (Bestimmung) |
| Effect | E | E (Effet or Influence) | E (Einfluss) |
| Preparation | P | P (Preparation) | H (Herstellung) |
| Property or Quality | Q | Q (Qualite) | Q* (Eigenschaft) |
| Reaction | R | R (Reaction) | R (Reaktion) |
| Treatment | T | T (Traitement) | T* (Behandlung) |
| Use | U | U (Utilisation) | V (Verwendung) |

*Because the mnemonic letter B was used for Bestimmung and E for Einfluss, Eigenschaft was assigned Q and Behandlung T in harmony with the English and French symbols.

## A MULTILINGUAL-MULTITERM INDEX IS A MULTILINGUAL DICTIONARY

As Multiterm indexing proceeds in several languages, such as English, French, and German, in the format of Multiterms IIIE, IIIF, and IIIG, it soon becomes apparent that a multilingual dictionary is being produced as well as a multilingual index. Indeed, the wrap-around computer process followed by alphabetization yields a dictionary within the computer than can be used to translate the terms of new Multiterms from one language to the other two. The larger the Multiterm index in the three languages, the less we need to translate for input and the more the computer can take over the translating process. Thus, the multilingual Multiterm index can be programmed to produce via computer its own multilingual dictionary.

Multiterms, as we have described them, are subjects within context—they are basically coordinated subjects or terms denoting the informational content of documents. Consequently, the multilingual dictionary of Multiterms gives considerably more meaning to the relationship of words in the three languages than is possible in a word-for-word dictionary form. In this sense, the multilingual-Multiterm dictionary has instructional advantages. One of the instructional advantages is how strikingly the multilingual-Multiterm dictionary reveals the similarity of English, French, and German technical words, as illustrated by VE, VF, and VG, and by the following polymer terms:

Polymer: Acrylate-P/_____/ (VIE)
Polymere: Acrylate-P/_____/ (VIF)
Polymerisat: Akrylat-H/ _____/ (VIG)
Ref. _____

Polymer: Amide-P/_____/ (VIIE)
Polymere: Amide-P/_____/ (VIIF)
Polymerisat: Amid-H/_____/ (VIIG)
Ref. _____

Polymer: Ethylene-P/_____/ (VIIIE)
Polymere: Ethylene-P/_____/ (VIIIF)
Polymerisat: Aethylen-H/_____/ (VIIIG)
Ref. _____

Polymer: Vinyl Chloride-P/_____/ (IXE)
Polymere: Vinyle, Chlorure de-P/_____/ (IXF)
Polymerisat: Vinylchlorid-H/_____/ (IXG)
Ref. _____

## DISCUSSION

Whereas cooperative abstracting programs between nations have been initiated, with mixed results, indexing has yet to be tried on an international cooperative basis. There are many problems associated with international indexing operations, such as the lack of an acceptable index system, linguistics, and semantic differences between disciplines of science. Yet it is highly desirable that duplication of indexing and abstracting be avoided throughout the world and that a system be established that is equally accessible by different languages.

Although this paper explores the feasibility of the Multiterm system as a basis for a multilingual index for English, French, and German, it is quite obvious that the results are equally obtainable for any language based on the Latin alphabet. The Latin alphabet restriction, however, is imposed by the computers available to us. With greater computer facilities, similar results should be possible with languages based on other alphabets or even on ideographs.

The significance of the results and conclusions of the work described in this and the previous papers on the Multiterm system is that a new concept has been developed on which an international indexing enterprise can be based. Much work towards standardization remains to be done, even for the production of multilingual-Multiterm indexes in English, French, and German. The purpose of this paper was to report preliminary work in this direction, and no attempt was made to be definitive.

The preliminary work proves the feasibility of the Multiterm system for the production of multilingual indexes and of multilingual dictionaries, with the following advantages:

1. The index is multidimensional by correlating a subject to the role which it plays vis-à-vis related subjects.

2. The information content of a Multiterm is greater than that of titles of documents and essentially equal to a good abstract.

3. The Multiterm is readable text.

4. As a multilingual index, it is inherently translatable from one language to another via computer.

5. The index is self-sufficient and does not require the construction and maintenance of thesauri or subject authority lists.

6. The system is computer processable by software and hardware universally available.

7. Input into the system can be at the source (writer or publisher) in the language of the document.

## ACKNOWLEDGMENT

The author is indebted to Z. B. Blums and B. E. Clouser, Hercules Research Center, for assistance with the French and German terms, titles, and phrases.

## LITERATURE CITED

(1) "UNISIST—Study Report on the Feasibility of a World Science Information System," by Unesco (United Nations Educational, Scientific and Cultural Organization) and the International Council of Scientific Unions, 161 pp., Unesco, Paris, 1971.
(2) Skolnik, H., "The Multiterm Index: A New Concept in Information Storage and Retrieval," *J. Chem. Doc.* **10,** 81–4 (1970).
(3) Skolnik, H., and Clouser, B. E., "Designing an Information Awareness and Retrieval System for Chemical Propulsion Literature," *Ibid.*, **11,** 39–43 (1971).
(4) Skolnik, H., "Designing an Index," *The Percolator* **35,** 26–31 (1959).
(5) Magistro, A. J., Nicholas, P. P., and Carroll, R. T., "The Oxychlorination of Ethylene at High Temperatures," *J. Org. Chem.* **34,** 271–3 (1969).
(6) Fodor, L., "Actual Optimization Problems of the Reactors Used for Ammonia Synthesis," *Chim. Ind., Genie Chim.* **104**(8), 1002–7 (1971).

# Changes in IUPAC Nomenclature Rules for Organic Chemistry

STANLEY P. KLESNEY
The Dow Chemical Co., Midland, Mich. 48640

Received December 13, 1971

**Changes in the 1971 edition of the IUPAC book, "Nomenclature of Organic Chemistry," from the previous editions (Sections A and B, 1966, and Section C, 1965) are listed with details of the changes.**

"Nomenclature of Organic Chemistry, Sections A and B Third Edition and Section C Second Edition 1971," International Union of Pure and Applied Chemistry Commission on the Nomenclature of Organic Chemistry, Butterworths, London, 1971. xiii, 337 pp. $24.00.

This book combines under one cover "Nomenclature of Organic Chemistry, Sections A and B," Second Edition, 1966, $5.00, and Section C, First Edition, 1965, $9.50. Both available from Butterworths, London. In addition, publication of the first editions of Sections A and B was in *J. Amer. Chem. Soc.*, **82,** 5545 (1960), and of Section C in *Pure and Applied Chem. II,* Nos. 1–2 (1965) [IUPAC].

"Nomenclature of Organic Chemistry" comprises definitive rules for Sections A, Hydrocarbons; Section B, Fundamental Heterocyclic Systems; and Section C, Characteristic Groups Containing Carbon, Hydrogen, Oxygen, Nitrogen, Halogen, Sulfur, Selenium, and/or Tellurium.

The new edition of Sections A, B, and C contains considerable changes from the earlier editions, but these have been confined mostly to correction of errors, to clarifications, in a few cases to expansion of existing Rules, and to provision of better or additional examples.

However, major changes are the deletion of the Rules for order of complexity of side chains [Rule A-2.3(a) and parts of Rules A-2.4 and A-2.5], and of the Stelzner method of naming heterocyclic systems [Rule B-4] by replacement nomenclature. In both cases, the procedures of the deleted rules have been little used recently.

While use of the earlier editions is possible in conjunction with the following list of changes, most users will find it convenient and time-saving to have the new publication. The combined and expanded index is a particular useful feature.

The Commission invites suggestions for necessary extension and modification of its rules. Comments may be sent to S. P. Klesney, Central Report Index, 566 Building, The Dow Chemical Co., Midland, Mich. 48640 or to any member of the Commission.

Significant changes are given below. Page numbers of the current edition are listed first with page numbers for the earlier editions following in parenthesis.

## SECTIONS A and B

Page 7 (8). A-2.3 Part (a) dealing with the order of complexity for arranging radicals in a name has been deleted. Parts of other rules dealing with the order of complexity have also been deleted.

Page 8 (10). A-2.4 Delete A-24(a) dealing with the order of complexity.

Page 9 (11,12). A-2.5 Delete the (a) names based on the order of complexity.

Page 11 (13). A-3.1 Add to the rules: When, in cyclic compounds or their substitution products, the locants of a double bond differ by unity, only the lower locant is cited in the name; when they differ by more than unity, one locant is placed in parentheses after the other (see Rules A-31.3 and A-31.4).

Page 17 (19). A-11.3 Delete the reference to fulvene. Fulvene is now treated in A-61.6.

Page 20 (22). A-13.3 Add Benzhydryl (alternative to Diphenylmethyl)

$$(C_6H_5)_2\overset{\alpha}{C}H—$$

Page 29 (31). A-24.2 5,6,7,8-Tetrahydro-2-naphthyl and its structural formula have been transferred to the exceptions to this rule.