ton, 1961. Available from Superintendent of Documents, Washington, D. C. 20402, $1.25. Under revision.
(2) Hayward, H. W., Sneed, H. M. S., Turnipseed, J. H., Tauber, S. J., *J. Chem. Doc.* **5**, 183 (1965).
(3) Proposed Revised American Standard Code for Information Interchange," American Standards Association Committee X3.2, document 206, American Standards Association, New York, January 21, 1965.
(4) Garfield, E., ed., "Encyclopedia Chimicus Internationalis (Cumulative Index Chemicus)," Institute for Scientific Information, Philadelphia, Pa., 1962.
(5) "The Naming and Indexing of Chemical Compounds from Chemical Abstracts," American Chemical Society, Easton, Pa., 1962, p. 87N.
(6) Tauber, S. J., "Digital Handling of Chemical Structures and Associated Information" in "Association for Computing Machinery, Proceedings of the 20th National Conference," Association for Computing Machinery, New York, 1965, p. 206ff.; Marden, E. C., "HAYSTAQ—A Mechanized System for Searching Chemical Information," National Bureau of Standards Technical Note No. 264, National Bureau of Standards, U. S. Department of Commerce, Washington, D.C., 1965, p. 29ff. Available from Superintendent of Documents. Washington, D. C. 20402. 50¢.

Links and Roles in Coordinate Indexing and Searching: An Economic Study of Their Use, and An Evaluation of Their Effect on Relevance and Recall*

J. G. VAN OOT[a], J. L. SCHULTZ[a], R. E. McFARLANE[a], F. H. KVALNES[a]
Textile Fibers Department, E.I. du Pont de Nemours & Company, Inc., Wilmington, Delaware
and A. W. RIESTER
Film Department, E. I. du Pont de Nemours & Company, Inc., Buffalo, New York

Received March 24, 1966

In a two-phase evaluation of links, roles, and type of indexing vocabulary (prescribed terms *vs.* a freer vocabulary), it was found that certain roles, used only on selected terms, can increase relevance of an answer with only minor reduction of recall. Role definitions must be clearly mutually exclusive. Roles are not useful if, fairly frequently, any term is indexed in several roles in one document-link, or searched in several roles in one question. Links also increase relevance, with essentially no reduction in recall, but are economically used only when the indexing is such that repetition of terms from one link to another for the same document is not frequent.

This paper covers a three-year period of part-time work of the five authors plus assistance from other personnel: clerks, computer programmers, consultants, and experts in the various fields of chemistry and textile technology who judged the relevance of the documents retrieved in our test of links and roles discussed in this paper.

Of necessity then, this description must be highly condensed. We present, therefore, just an outline of the experimental procedures used, concentrating on the results; giving a philosophy of using links and roles which we developed from this study of the effectiveness of links and roles in reducing false retrieval, and of the economics of their use.

THE PROBLEM

The specific question which faced us when we began this experiment was whether or not our indexing system should be converted to the use of links and/or roles. By our indexing system, we mean only one of perhaps 10 information centers in the Du Pont Co. which use concept coordination techniques to index internal company reports of proprietary information.

Our system, serving the needs of the Textile Fibers and Film Departments, was started in 1950 (*1*). It was not until eight years later that our colleagues in the Engineering Department reported (*2*) the results of their studies of the use of links and roles in coordinate indexing. We followed with interest the successful establishment in the Plastics Department of a coordinate index to patents, using links and roles (*3*). Other departments started indexing centers during the period 1958–1962 using links and roles in concept coordination techniques.

* Presented before the Division of Chemical Literature, 148th National Meeting of the American Chemical Society, Chicago, Ill., Sept. 3, 1964.

[a] Present address: Secretary's Department, Information Systems Division, E. I. du Pont de Nemours & Co., 3211 Centre Road Building, Wilmington, Delaware 19898.

But our reports were not like patents, the depth of indexing was different, and our vocabulary more precoordinated as a result of its being developed for an edge-notched card system in 1950. In short, our information service was operating well, with about 20,000 reports already indexed without links and roles, and we felt that an evaluation of any change in the system, particularly a study of the economics involved, should be undertaken before any decision to change was made. This evaluation was divided into two phases. In the first phase, we compared our then operating system, used as the control, with an experimental one, using links, roles, and a freer vocabulary—*i.e.*, one with fewer bound terms than our control. The second phase involved the evaluation of a modified indexing system which was developed for our purposes based on the results of the first phase of the experiment.

PHASE I

In the first phase of our experiment, we re-indexed 550 reports that had already been indexed by our regular techniques. These reports were a representative cross section of the types of information we normally handle; the test sample contained a number of pioneering research, applied research, development, marketing, and manufacturing reports proportional to the number of each of these types we receive annually. The test indexing was done with a new vocabulary which we developed during the test; this vocabulary was edited, related terms and generic relationships were provided, and a thesaurus was printed with the use of an IBM 1401 computer. This vocabulary was freer than our control, having fewer precoordinations, and the indexing deeper (1.6 times more indexing terms per document than the control.)

Links were used to associate terms used to index one idea in the same report. False correlation was sometimes tolerated when a prohibitive number of links would have been required to prevent it. The effect of links was tested by searching both with and without links.

Two sets of role indicators were used. These are defined in Appendix I. Set I resembled closely what had become known outside Du Pont as the "Du Pont roles," while Set II was designed to test several modifications of Set I. Some of these modifications were later independently adopted by the Engineers Joint Council (EJC). Set I and Set II were combined in a single indexing by coding the role changes involved whenever necessary to permit retrieval with either set.

In the input stage, cumulative stop watch times were recorded for the indexing, the editing of the indexing of the other indexers, and, for the control system, the writing of a descriptive abstract of the information indexed.

In the searching stage we selected for the test every fifth question that had already been answered by the control system during several months in late 1961. Seventy-eight real questions were thus used. Only 56 of these resulted in answers from the test body of reports. These same questions were put to the test index, and searches were made manually from the experimental inverted file printout. Each question was searched on the test system in five different ways: (1) using neither links nor roles, (2) using links but no roles, (3) using links and Set I roles, (4) using links and Set II roles, and (5) using no links and Set I roles.

Answers from all six searches (five test and one control) were combined for each question into a single list of report numbers. These lists were distributed among about 30 experts in the various fields of technology who judged the relevance of each report to the question. (These experts were not told how the reports were selected; they judged only relevance of reports and not the effectiveness of each system).

For all 56 questions, a total of 225 relevant reports and 138 nonrelevant reports were listed. Total relevant reports here means the total number of reports turned up by any of the six systems and adjudged relevant by the experts; no attempt was made to determine if any of the other reports in the test, not retrieved by any system, were also relevant to a question.

In the discussion that follows, relevance ratio is expressed as the percentage of the total number of documents retrieved which were relevant to the questions; recall ratio is the percentage of the total *relevant* documents in the system that were retrieved.

$$\text{Relevance ratio} = \frac{\text{Relevant retrieved for subsystem tested} \times 100}{\text{Total retrieved for subsystem tested}}$$

$$\text{Recall ratio} = \frac{\text{Relevant retrieved for subsystem tested} \times 100}{\text{Total relevant for all systems}}$$

The effect of links can be seen in Table I by comparing System 1 with System 2 (when no roles were used) and also by comparing System 3 with System 4 (when Set I roles were used). In both cases recall of relevant documents was essentially unaffected by links, but the relevance of the answer was improved.

The effect of Set I roles is similarly shown by two comparisons: System 1 with System 3 (when links were not used) and System 2 with System 4 (when links were used). It is seen that roles reduced recall by about 7%, but increased the relevance ratio of the answer substantially. Stated another way, roles caused us to miss 15 documents, or 7%, of the total 225 relevant documents, but reduced false retrieval by about 30%.

The combined effect of links and roles is shown by comparing Systems 2 and 4. Again the figures show that the relevance of the answer was improved substantially at the price of failure to retrieve some of the relevant documents.

The values for the control system in Table I show that the use of a vocabulary that is more precoordinated than the test, and the screening of the answers, using abstracts of the documents cited, gives relatively high values for recall and relevance for the answers we actually sent back to our clients. Further comments about our screening procedure are given later.

The use of terms in the control vocabulary which were broader than those of the test vocabulary (the control vocabulary was originally developed for edge-notched cards) accounts for the fact that more documents were recalled by the control system than by System I.

Table I. Effect of Links and Roles on Recall and Relevance, Phase I

Total Relevant Documents 225
Total Nonrelevant Documents Retrieved 138
Total Number of Questions 56

System	Links	Roles	Recall, %	Relevance, %
1.	No	No	84	70
2.	Yes	No	84	82
3.	No	Yes	77	85
4.	Yes	Yes	77	90
Control	No	No	83	87

Why Roles Blocked Relevant Retrieval. We were, of course, curious as to why roles blocked relevant retrieval. An examination of the work sheets for questions where this occurred showed that some of the "misses" could be attributed to our lack of experience with roles. Even though we had a long list of ground rules that were drawn up in an attempt to ensure consistent role assignment to terms, there were many individual instances where these rules were inadequate or not applicable. We therefore conclude that roles do reduce false retrieval, but roles must be used consistently, both in indexing and searching, and only by personnel thoroughly experienced in their use for all types of information to be included.

More important than our lack of familiarity with roles (and this was borne out by the second phase of our experiment) was a lack of clearcut mutual exclusiveness in role definitions.

Roles Require Definitions That Are Clearly Mutually Exclusive. To illustrate this, let us use an example cited by Costello recently (*4*). Here the concept of cutting trees with saws is illustrated in three ways, which we shall identify as the words used in documents one, two, and three, respectively (Table II).

The term-role combinations for Document 1 are read "The operation. . .of saws used for cutting trees into veneers to be used for plywood." Those for Document 2 as "Saws to be used for cutting trees are discussed;" and for Document 3 as "The cutting of trees into veneers, using saws, is discussed." Note that in these examples, the term "saw" is in three different (EJC) roles: Role 9 (passively receiving action), Role 8 (primary subject of discussion), and Role 10 (means to accomplish the objective).

Our experiment showed quite clearly, when questions were asked like "what information do you have on the use of saws for cutting trees?", that the term "saw" had to be searched in at least all three of these roles in order to avoid losing information.

Document 1 discusses the "design, development, operation, maintenance, and repair of saws used for cutting trees. . ." etc. This, in our view, is clearly information on "saws"; in a lumberman's handbook, this concept would most probably be indexed under "Saws, operation of," etc. "Saws" is the principal subject of discussion just as much as "saws" is in Document 2, yet in one case Role 9 is used and in the other Role 8 is used. These role definitions are thus not clearly mutually exclusive. Hence, the indexer must use both sets of term-roles (resulting in an increase in the number of postings for the same idea), or the searcher must formulate two search questions, or form a union of both term-roles in one question in order to avoid losing information.

Table II. Term-Role Indexing

Document 1		Document 2		Document 3	
Operation	8	Saws	8	Cutting	8
Saws	9	Cutting	4	Trees	1
Cutting	4	Trees	1	Saws	10
Trees	1			Veneers	2
Veneers	2				
Plywood	4				

A similar case can be made for using roles to distinguish the thing which is the "primary topic of consideration" from that same thing which is discussed only as a "means to accomplish the primary topic of consideration." These opposing relationships can be distinguished by EJC Roles 8 and 10. The theoretical value of the ability to make this discrimination was, however, substantially diminished in our minds when our experts, in many cases like this, declared that information in both relationships was relevant and useful to them.

Finally, we found that the "cause" and "effect" roles were being used redundantly in indexing and searching along with the other roles for nonchemical terms, again resulting in the loss of the discrimination that these two roles are intended to provide. We therefore concluded that these roles were not very useful. At this point, it is appropriate to emphasize that the chemical roles—*e.g.*, reactant, product, solvent—caused little trouble. The problem exists mainly with the roles used on nonchemical terms.

We shall return to the problem of overlapping role definitions later. Suffice it to say at this point that the lack of discrimination caused by overlapping role definitions was largely responsible in our view for the blocking of relevant information (that is, a term would be indexed in one role, but lost on searching in another role). This led us to propose a *partial* set of roles which was tested in the second phase of our experiment.

PHASE II

In this phase, we began indexing new reports routinely, using links, but using only the roles we felt were clearly mutually exclusive in their meanings. Furthermore, we used roles only on those terms where the discriminating power of the roles was clearcut; in other words, not all terms were assigned role indicators. Stated very briefly, the set of roles tested in Phase II was similar to the EJC set of roles, except that no roles were assigned to the terms that would take the "primary topic of consideration," the "passively receiving an operation" and the "input" and "output"roles for physical changes. We used the chemical roles for reactant, product, solvent or medium, impurity, and special agent for only chemical—*i.e.*, not physical—changes.

In the input stage, our regular indexing vocabulary was used, roles and links were appended, abstracts written, and each document-link was identified with a term, "Report Index Test System," that would serve as a flag for the test documents in the search operation. Eleven hundred documents were so indexed, and incorporated

into our regular computer file. Cumulative stop watch times were kept for all technical and clerical operations.

Searches of all 126 actual questions we received during several weeks were made with and without roles, with and without links, and with and without both roles and links. Answers in the test body of reports were found for only 74 questions.

Results of Phase II. Analysis of the results (Table III) showed that roles eliminated 18% of the false retrieval while blocking only a little over 1% of the relevant reports, a definite improvement in recall over the results of Phase I. Links eliminated 11% of the false retrieval and did not prevent the retrieval of any relevant report. The use of both links and roles eliminated 27% of the false retrieval, but again blocked a little over 1% of the relevant reports. On the other hand, it was interesting to note that links affected retrieval—*e.g.*, screened out nonrelevant reports—in only 20% of the questions, while roles affected the answers to only 12% of the questions.

Table III. Effect of Links and Roles, Phase II

74 Questions

	False Retrieval Eliminated, %	Relevant Reports Blocked, %
Links	11	0
Roles	18	1
Roles & Links	27	1

Nevertheless, both links and roles, used only on some selected terms, served the purpose for which they were intended—namely, to reduce the number of nonrelevant reports in an answer. The question remains, however, as to how well they serve this purpose.

SCREENING OF OUTPUT

Part of the answer to this question can be found in the percentage of nonrelevant reports that turned up for all questions. When links and roles were not used, only 52% of the documents cited were relevant to the questions. This low percentage was attributed in part to some broad terms used to search some questions that required a more specific answer. Over half of the false drops occurred in only six "sticky" questions; one question alone out of the 74 accounted for 16% of the irrelevant reports retrieved. The use of links and roles increased relevance from 52% to only 59%. This in our view represents too much irrelevant material to send back to our clients. The answers thus still require screening before being sent to the client, some questions requiring more screening than others.

In our control system, the information specialist uses abstracts to screen the answers before copies of the abstracts of relevant documents are sent to client. The client then can further screen the answer using the abstracts, which themselves may contain enough information to answer his question. At any rate, the abstracts serve to reduce the number of the complete reports that the client has to order. For this purpose, the descriptive abstract is preferred to a listing of indexing terms with their roles, primarily because the client would be handicapped by unfamiliarity with the role use and meanings.

These abstracts are, incidentally, prepared by the indexers, and in the first phase of this test, we showed that an abstract could be dictated and the report indexed without links and roles in 26 minutes, whereas indexing only (no abstract) with links and roles required an average of 31 minutes per report.

The economics of the use of links and roles then revolves around the question as to whether the added input costs are offset by the savings in output screening time.

EQUATING COSTS TO SAVINGS

To answer this question, we developed a formula that equated the extra cost in input due to the assignment of links and roles to each term, to the cost savings resulting from not having to screen out as many nonpertinent reports from the answer. This equation takes into account not only the extra indexing time and the saving of screening time at output, but also the rate of document accession, the number of questions answered per unit time, the rate of change in document accession and question load, the size of the document file, and even document obsolescence in the file.

The actual times for each of these operations that we observed during the experiment were inserted into the equation in terms of actual average clerical salaries and actual average technical salaries plus overhead. Computer processing times per document-link-term-role combinations were used to calculate the effects of links and roles on computer costs on both updating and searching.

A generalized statement of the results of the solution to this equation is given in Figure 1. Here is plotted the cumulative input costs and output savings as a function of the age of the index. In our system the rate of document accession over the years has been fairly constant, and hence the input expense accumulates over the years in a linear fashion. The input expense line would curve upward or downward if the accession rate increased or decreased. As the volume of indexed documents increases, the cost savings increase along a rising curve. Eventually the two curves intersect. At this point the input costs have been repaid by the output savings. After this time the savings increase rapidly, while the input costs continue to accumulate linearly.

The critical factor, then, is the location of the breakeven

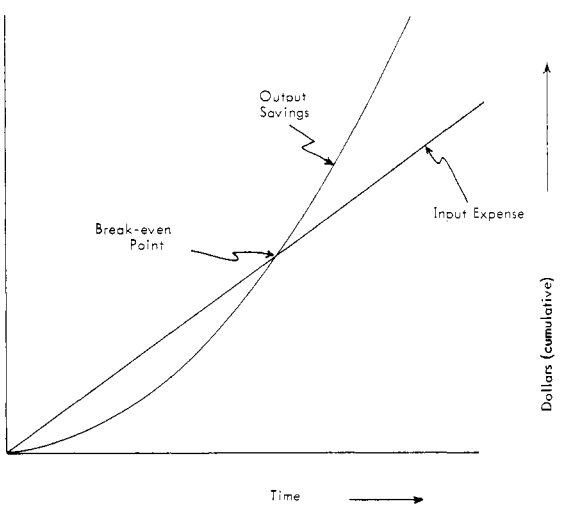

Figure 1. Comparison of input expense with output savings.

point in time: How long will it be before the extra input time is repaid by the saving of output screening time?

Some of the factors affecting the location of the breakeven point have long been recognized. If a system has a high rate of document accession, but relatively few questions, economics dictate that less time be spent with indexing and more with question analysis. The use of links and roles in this situation is probably not advisable.

On the other hand, if there are many questions directed to a file of only a few documents, economics indicate that indexing be done in depth, with links and roles employed to the fullest extent. This is nothing new. The real situations, in our case and many others, lie in between these extremes.

The generalized plot of input costs and output savings in Figure 1 applies to new systems as well as to major changes in older ones. A certain amount of input effort is required before the systems manager can realize a return on his investment. For a small system, the breakeven point may be reached after a few weeks or months; for larger systems it may be up to five years. In each case, the input dollars at this point would be commensurate with the size of the system. For our system (20,000 documents, increasing at a rate of 1600 documents a year and answering 1800 inquiries per year) we felt that a breakeven point any time before two years was a reasonable goal.

Our evaluation of the factors affecting the breakeven point in our system showed several things to avoid and several steps to be taken to make use of the discriminating power of links and roles economically.

Links and roles add extra postings to the indexing of a document. If we had to summarize our work in one sentence, it might be "Don't allow links or roles to cause the creation of any unnecessary document postings." Our system was particularly sensitive to computer costs, and probably this factor would be less prominent in other systems. (For example, two-thirds of the extra cost of using links in this test was attributed to computer updating; the computer also accounted for half of the extra cost for both links and roles.)

Nevertheless, the extra cost of using links and roles is almost directly dependent on the number of extra document-link-term-role combinations; the more postings one has for a document, the more it costs to process the index to that document into the system. If some of these postings are unnecessary, more money is being spent than is needed.

Earlier, in discussing overlapping role definitions, it was pointed out that where such overlap existed, the indexer resorted to redundant indexing—that is, putting one term in more than one role—in order to ensure that several viewpoints were indexed to avoid loss of information at search time. The consequence of putting one term in more than one role is the creation of extra postings, requiring extra cost to process them.

Our cost studies showed, moreover, that links were more expensive than roles for our system. This was due to the fact that links increased the number of postings per document by 62%, from 23 to 37. This is all the more striking in view of the fact that only 51% of the documents were indexed with more than one link, and that the average number of links was only 2.65 per document.

The 62% increase in postings per document suggests a high repetition of terms from one link to another. This would be the case for a document dealing generally with one main subject, where links are used to separate several subsidiary ideas from each other; for each subsidiary idea there is a link, but the terms for the main subject must be repeated in each of these links. Many documents seldom deal with several mutually exclusive subjects, but closely related aspects of one subject. For documentation purposes, we would make an exception for a document which dealt with several different chemical reactions.

To avoid these extra postings, we propose a change in the computer program that will provide for master and subsidiary links. Table IV shows how this would provide separation of the ideas in the indexing of a document without extra postings. Here terms linked "A" or "B" may be coordinated with terms linked "M," but "A" and "B" terms may not be coordinated together. This idea is an extension of Walker's singularity Sublink (5), but it makes provision for more than one term in the singularity aspect, and for more than one singularity aspect in a document. We hope to report on the utility of this technique after we have evaluated it more thoroughly.

Table IV. Proposed Use of Links

	Present Use		Master and Subsidiary
Weaving	A	B	M
Fabrics	A	B	M
Fiber Blends	A	B	M
"Dacron" Polyester	A	B	M
Cotton	A	B	M
Yarn Count	A		A
Drape	A		A
Composition		B	B
Dyeability		B	B
(Number of Postings		14	9)

There are, therefore, several factors which determine if links are economically feasible. If the documents in the system deal routinely with several entirely different subjects, so that terms are not excessively repeated from one link to another, links are worthwhile. For those documents dealing with one main subject and several subsidiary ideas, the extra duplicate term-link postings raise the cost of input as to make the use of links questionable, for our computer program anyway, unless the master and subsidiary link technique is adopted. The economics of links are also dependent on output savings—*i.e.*, on the time for screening reports in the answer.

SUMMARY

To summarize then, we have found that links and roles do substantially reduce the number of nonrelevant documents retrieved in a concept coordinate index. Roles can cause failure to retrieve a small percentage of relevant documents, if the role definitions are not clearly mutually exclusive. This is particularly true for roles on nonchemical terms. Overlapping role definitions also cause multiple term-role postings which are unnecessary, and which add to the cost of processing. For certain types of documents,

the use of links also adds extra postings, due to the repetition of terms from one link to the next. A modification of our present computer program to provide for master and subsidiary links where desirable, to avoid these extra term-link postings, is being evaluated.

CONCLUSIONS

After three years of evaluation, there is no one answer as to whether links and roles are economically justified. This answer must be determined for each index, taking into account the rate of document accession, the number of questions answered in any given time period, the rate of obsolescence of the indexed information, the nature of the technology involved, the depth of indexing, the type of indexing vocabulary—*e.g.*, free or precoordinated—the specificity of the indexing terms, the particular set of roles used, the nature of the information indexed as this relates to the division of the documents into discrete links, the mechanism of linking—*e.g.*, the use of master and subsidiary links—the efficiency of the computer program or other mechanism used to manipulate the index, and whether or not the index is searched by an information specialist or by the client directly.

The most important thing to watch out for in the evaluation of links and roles, as we see it, is the number of times more than one role is used for any term in either indexing or searching, or the number of times that terms are repeated from link to link in indexing a document. If either of these numbers is excessively high, a closer examination of the indexing and searching techniques and of system economics would be in order.

If we were designing a system from scratch today, we would use roles to distinguish the starting materials and final products of a chemical reaction. For our index to textile fiber and film technology, this is all we feel we need. For a more chemically oriented system, additional chemical roles to distinguish solvent, impurity, and special agent would be useful. For nonchemical terms, we would not recommend any of the roles we have studied so far. Some argument can be made for combining the "cause" and "effect" roles into a single role which might be useful, but in our experiment, the usefulness of the "cause" and "effect" roles was borderline.

We would use links to separate a document into discrete intellectual subdivisions, or into master or subsidiary links, as described above. If it can be done economically, we would have the computer create, at updating, the master and subsidiary links from the more conventionally linked terms on the tracing sheets. These recommendations are what we find expedient for our system under present conditions. For other systems, the results may be entirely different.

APPENDIX I. ROLE DEFINITIONS

Set I Roles (Phase I).

1. Using; by means of; with; classes of agents; devices; operations; methods; concepts; etc.
2. Cause; influence; independent (controllable) variable.
3. Reactant; input; raw material to chemical or physical change in which name of material changes.
4. Special agent: specific material that modifies material, process, operation, etc., generally without entering reaction.
5. Medium; vehicle; solvent; environment; mechanical support, as for catalyst; dispersion means.
6. By-product; waste; scrap; contaminant; impurity.
7. Product; output from chemical or physical change in which name of material changes.
8. Information, research, or development on; main subject of link.
9. Dependent (affected) variable.
10. Design information on—*e.g.*, blueprint, numerical data, flow sheet.
11. Physically processed, treated, or handled without change of name; of (as fabrics *of* nylon, tenacity *of* rayon); for (as antioxidants *for* nylon).
12. Patent or legal information on.
14. Sales and marketing information on; selling prices; customer service; market survey; marketing plans; mill processibility.
15. Manufacturing information on: repetitive manufacturing, not plant-scale research; operating instructions; production.
16. Construction information on& the actual building, installation, etc., of physical facilities.

– Modifiers; adjectives; proper nouns (people; places; companies).

Set II Roles (Phase I).

2. (a) Same as in Set I. (b) Secondary cause—*e.g.*, "drawing" in "effect of temperature of drawing" (coded as 42).
3. Reactant, input, raw material to chemical change only.
4. (a) Same as in Set I. (b) All functions of role 1 in Set I (coded as 1).
5. Same as in Set I.
6. Same as in Set I.
7. Product, output from chemical change only.
8. (a) Same as in Set I. (b) All functions of roles 10, 12, 14, 15, and 16—appropriate terms used in place of these roles (coded as 10, 12, 14, 15, and 16 respectively).
9. (a) Same as in Set I. (b) Secondary effect—*e.g.*, "fabrics" in "effect on luster of fabrics" (coded as 49).
11. (a) Physically processed, treated, or handled, with or without change of name (aspects included under 3 and 7 in Set I coded as 43 and 47 respectively). (b) *Of* as in Set I, but excluding aspects coded as 42 or 49. (c) *For* aspects excluded (role 13 used instead).
13. Used in; used or intended for; as antioxidants *for* nylon, yarn *for* carpets.

(90 + X). Absence of, where X is the appropriate positive role—*e.g.*, 94 = "not using," 97 = "expected product not obtained."

Modifiers and adjectives given the same code as the terms modified; proper noun terms coded like any other terms [coded as (20 + appropriate role) for adjectives].

Set III Roles (Phase II).

1. Intended (but not fabricated) end use.
2. Independent variable (cause).
3. Reactant (materials and classes of materials only).
4. Special agent; equipment, process, or method used.
5. Reaction medium; environment.
6. Impurity; by-product.
7. Reaction product (materials and classes of materials only).
9. Dependent variable (effect).

(20 + role number). Adjective; noun term used adjectivally (*e.g.*, "spinning, wet" in role 20 + means "wet-spun").

(90 + role number). Negation (used only when a negative term would otherwise have to be invented).

(0. Computer input signifying "no role assigned to term").

APPENDIX II. FORMULA EQUATING INPUT COSTS TO OUTPUT SAVINGS

$$s\left(\frac{F}{Q_{test}D_{test}}\right)\int_{o}^{tn} QD(P_t/P_o)\,dt = cD \qquad S = C$$

where

S = Total cumulative savings afforded by any test system by time t.

C = Total cumulative extra cost of indexing by any test system by time t.

s = Savings afforded by eliminating one false drop (cost of screening one document from a search answer).

F = False retrieval (number of documents) prevented in the test by any given system.

Q_{test} = Number of test questions.

D_{test} = Number of test documents.

Q = Searching load (rate of question receipt) at any time t.

D = Total number of documents in the index at any time t.

P_t = Average probability of retrieving a document added to the index t years ago.

P_o = Average probability of retrieving a document newly added to the index.

P_t/P_o = Document obsolescence factor; average relative probability of retrieving a document added to the index t years ago.

c = Extra cost of indexing one document by any test system (over the cost of indexing by the control system).

LITERATURE CITED

(1) Van Oot, J. G., *J. Chem. Doc.*, **2**, 220 (1962).
(2) Costello, J. C., Jr., Wall, Eugene, Chapter VIII, p. 141, in "Solutions for Mechanizing Coordinate Indexing," Vol. V, *Documentation, Incorporated*, 1959.
(3) Montague, B.A., *Am. Doc.*, **13** 104 (1962).
(4) Costello, J.C., Jr., *J. Chem. Doc.*, **4**, 116 (1964).
(5) Walker, J.F., *ibid.*, **4**, 45 (1964).

Foreign Patents Documentation*

M. HYAMS
Derwent Publications Ltd., Rochdale House, Theobalds Rd., London, W.C.I., England

Received November 29, 1965

This paper is concerned with rapid access to and maximum use of the patent literature published throughout the world.

I estimate that the 45 most active Patent Offices in the world publish some 6000 patent specifications weekly, about one half of which have already been published in another country. Altogether there are about 1000 completely new or first-issue inventions of chemical interest published per week. Only 25% of these basic chemical inventions ever get published in the United States, and half of these are published earlier in some other country. It is clear, therefore, why United States firms are well advised to gather their patent information from an international source, and not merely rely upon specifications published by the United States Patent Office.

I have grouped together different countries on a patent information basis: The first section deals with the five major issuing countries. The second section deals with three minor but quick-issuing countries of particular importance, and the third section with nine further countries of special interest.

Statistics are presented in Section D for the total number of specifications filed in the eight countries dealt with under Sections A and B. For each of these countries, figures are provided showing the number of inventions currently protected in each of the seven broad fields of chemical interest, and also the country of origin or first priority. These figures enable you to assess the volume of work to be carried out in following foreign patents in your particular sphere of interest, and the type of information to be expected from the individual countries.

* Presented in part before the Division of Chemical Literature, 150th National Meeting of the American Chemical Society, Atlantic City, N. J., September 15, 1965.