Evaluation of Coordinate Index Systems During File Development

Part II: An Application*

DONALD W. KING†
Westat Research, Inc., Bethesda, Maryland
PATRICIA M. McDONNELL
U. S. Patent Office, Washington, D. C.

Received October 19, 1966

This paper describes the application of an evaluation program to a combined coordinate index–conventional classification search system currently being developed in the Office of Patent Classification of the U. S. Patent Office. The nature of the file and of the evaluation program are reviewed; the experiments are described and their results presented; and conclusions based at least in part on the results of the evaluation program are discussed.

In 1964 work was begun on the reclassification of patents in Class 167, Medicines, Poisons, and Cosmetics. Patent claims in this class were found to fall into three categories:

(1) methods of utilizing chemical compounds or compositions (mixtures) for medicinal, biocidal, or cosmetic purposes;
(2) chemical compositions disclosed as having utility for one of these purposes;
(3) methods of preparing these compositions.

Examination of patent applications was observed to require searches based upon any or all of these categories.

Further analysis showed that the utilities for medicines and poisons lay in several well-defined areas:

(1) elimination or repulsion of life forms which are considered external pests;
(2) elimination or neutralization of life forms which are uninvited guests in an animal host;
(3) treatment of life forms to alleviate or correct certain disease conditions.

A decision was made to build a conventional classification schedule based primarily on the active ingredients of the claimed compositions and also to construct a supplementary coordinate index search system relating to pharmacological and biocidal utility. Cosmetic utility is to be excluded from the coordinate index search system, but is to be included as a "special" area, with a high priority in the conventional classification schedule.

The basic indexing tool is a thesaurus of disease conditions of animals and pharmacological effects of chemical agents. This thesaurus contains approximately 4000 entries and is intended to direct the indexer or searcher to select one or more descriptors from the several descriptor lists which were also prepared. These descriptor lists or facets are as follows:

1. Disease conditions and pharmacological effects.
2. Organisms attacked by chemical agents.
3. Body parts pathologically affected.
4. Subject treated.
5. Route of administration.
6. Form of administration

The subclass schedule serves as a seventh facet or descriptor list. Lists 1 through 6 contain a total of approximately 600 terms, and the classification schedule approximately 350 subclasses. The entire file contains about 22,000 patents. A more detailed description of the system may be found elsewhere (*2*, *3*).

THE EVALUATION PROGRAM

The evaluation program described previously (*1*) was carried out during the initial stages of the developmental phase and included an indexing experiment and a preliminary search experiment.

The principal objectives of this program were:

1. to predict the performance of the system once it is made operational;
2. to reveal specific needs, if any, for modifying and correcting the system before changes become too costly;
3. to investigate indexing accuracy of Examiners, Classifiers, and Analysts as potential indexers;
4. to ensure that input errors will not be a significant factor contributing to unsatisfactory operational searches;
5. to perform a preliminary investigation of operational search strategies.

These objectives were accomplished by investigating three factors: (a) the *cost* associated with input, equipment, and operation; (b) the predicted *search effectiveness* once

* Presented before the Division of Chemical Literature, 152nd National Meeting of the American Chemical Society, New York, N. Y., Sept. 15, 1966.

† Work accomplished under contract to the U. S. Patent Office.

the system has become operational; and(c) the *time* involved in development, implementation, and operation of the system. These factors were investigated by experimentation, and search effectiveness was measured in part by mathematical models described previously (*1*).

The indexing experiment was used to estimate the cost of indexing the entire file. The observed average number of terms per document was combined with information obtained in the systems analysis study to estimate equipment, operating, and updating costs. Searching costs were estimated from the preliminary search experiment.

Search effectiveness is a measure in part of the ability of the user to retrieve only the documents desired. Failure to retrieve desired documents and the retrieval of unwanted documents may nearly always be attributed to indexing and searching errors. The evaluation program provides a means for identifying sources of errors, measuring them, and relating them to search effectiveness. Indexing accuracy is measured in the indexing experiment; the effect of indexing errors on search effectiveness is estimated by mathematical models; and the results are tested in the preliminary search experiment. The preliminary search experiment also serves to determine the effect of alternative search strategies on search effectiveness.

Finally, the evaluation program provides information as to the time required for system development and implementation, as well as for operation and updating. Time information is particularly important for planning and programming system development, since two nonconventional search modes—*i.e.*, multicolumn sorter and optical coincidence equipment—are planned to be implemented with the conventional classification schedule.

THE INDEXING EXPERIMENT

An indexing experiment was performed on the coordinate index portion of the system with the following objectives:

1. to provide a basis for choosing among potential indexers;
2. to estimate parameters of retrieval models for measuring the effect of input errors on searching;
3. to ensure that the level of input accuracy is satisfactory;
4. to estimate total time required to index the entire file.

Four Patent Classifiers, two Patent Examiners, and two Analysts (Aids) were selected for the indexing experiment. The first six were highly trained technically with varying degrees of familiarity with the specific subject matter to be indexed. One of the Aids had some technical background and familiarity with the documents, but the other was lacking in both. Of the eight persons involved, only one Classifier was familiar with the term lists and the nature of the system. None in the group had experience in coordinate indexing.

A seminar was held to acquaint indexers with the coordinate index term lists, the thesaurus, and their relation to the conventional classification schedule. Instruction was also given for analyzing and coding the documents. Each indexer was given 15 to 20 patents on which to practice. These indexed patents were then reviewed, and a number of errors due to inconsistencies and ambiguities in these indexing aids and in the instructions were found. Therefore, the actual indexing experiment was delayed for several weeks to revise the thesaurus, term lists, and indexing techniques. After the modifications were completed, 60 sampled patents were indexed twice by combinations of two indexers selected from the three groups. The coding sheets were compared by the system designers, and a "correct" indexing was determined.

The first objective of the indexing experiment was to provide a basis for choosing among potential indexers. Table I gives the proportion of correct terms selected as well as the proportion of terms chosen incorrectly. Indexers were selected from among the Classifiers (C_2, C_3, C_4) and Analysts (A_2). Examiners were not chosen, since they were not significantly better indexers and are not generally available to the Office of Patent Classification.

The second objective of the indexing experiment was to estimate parameters of retrieval models which, in turn, are used to estimate the effect of input errors on searching. The retrieval models provide estimates of (1) total retrieval, (2) the number of missed documents resulting from indexing errors, and (3) the number of false drops resulting from indexing errors. The model parameters for making these estimates are designated as p_3, the proportion of terms which should be indexed and were; p_2, the proportion of terms which should not be indexed but were; and Q_k, the proportion of the file which should contain the k terms used in the search query. The first two parameters are estimated directly in the indexing experiment and their values are in Table I. The third parameter is estimated subsequently by applying search queries to the indexing judged to be "correct" in the indexing experiment.

The next objective of the indexing experiment was to establish that the level of input accuracy is satisfactory.

Table I. Estimated Proportion of Correct and Incorrect Terms Indexed for Indexing Experiment[a]

	Proportion of Correct Terms Indexed (p_3)	Proportion of Incorrect Terms Indexed (p_2)
Classifier		
C_1	0.81	0.0002
C_2	0.94	0.0003
C_3	0.93	0.0001
C_4	0.95	0.0000
	Av. 0.92 (0.82, 0.98)	0.0001 (<0.0002)
Analyst		
A_1	0.64	0.0003
A_2	0.95	0.0004
	Av. 0.82 (0.69, 0.92)	0.0003 (0.0002, 0.0005)
Examiner		
E_1	0.99	0.0001
E_2	0.93	0.0002
	Av. 0.96 (0.88, 1.00)	0.0001 (<0.0002)

[a] 95% confidence limits.

The number of documents missed and the number of false drops are a function of p_2 and p_3. The number of false drops resulting from input errors as estimated by model was found to be negligible. The proportion of documents missed because of indexing errors is given in Table II by number of terms used in search queries.

Indexing accuracy of $p_3 = 0.94$ is the average for indexers chosen for preparation of this file. One may expect to retrieve, therefore, approximately 83% of documents actually containing the three terms asked in a three-term query; the remaining 17% are missed because of indexing errors. The other two columns yield estimates for two indexers not chosen for the file preparation; values of p_3 at 0.81 and 0.64 are clearly not satisfactory from the standpoint of missed documents. (The input accuracy for the complete file should be even higher, since all coordinate indexing is currently being checked for the purpose of establishing the adequacy of the thesaurus and term lists.)

Table II. Model Estimates of Proportion of Documents Missed Because of Indexing Errors[a]

No. of terms in search query	$p = 0.94$ (Average of indexers chosen)	$p = 0.81$ (Classifier not chosen)	$p = 0.64$ (Analyst not chosen)
1	0.06	0.19	0.36
2	0.12	0.35	0.59
3	0.17	0.47	0.74
4	0.22	0.57	0.83
5	0.27	0.65	0.89
6	0.31	0.72	0.93

[a] Indexer accuracy(p_3) estimated for three indexers.

One can see that the proportion of documents missed because of indexing errors is quite sensitive to values of p_3 and the number of terms used in the search query. The estimates in Table II also indicate the need to maintain high standards of input accuracy. Detailed results of model estimates are given in the next section.

Three basic assumptions regarding the estimates of proportion of documents missed because of indexing errors given in Table II are that (1) input errors occur with the same frequency in terms asked in search queries and in all terms indexed; (2) errors of omission are independent among terms; and (3) input errors occur with the same frequency in relevant documents as in nonrelevant documents (where relevant refers to usefulness of documents to Examiners). These assumptions were investigated in the preliminary search experiment, and all evidence thus far supports their validity.

The fourth objective of the indexing experiment was to estimate the total time required to index the entire file, and thereby gain some insight into the cost of preparing the coordinate index search system. Among the indexers chosen to index the file, the average time for entering one patent (six to eight terms) into the coordinate index system alone was 12 minutes; and for the Classifier who was familiar with the term lists, 6½ minutes. A decision was made to encode documents concurrently for the coordinate index and conventional classification systems, since it was found that most of the encoding time for each is spent in reading the document.

SEARCH EXPERIMENTS

A preliminary search experiment was conducted in conjunction with the indexing experiment as part of evaluation during the developmental phase (*1*). The two experiments, together with mathematical models, yield information (1) to determine whether one can expect the total retrieval to be within acceptable limits, (2) to estimate the effect of input errors on search results, (3) to indicate whether it is necessary to modify the term lists and structure, (4) to assist in final selection of search and retrieval devices, and (5) to gain an insight concerning operational search strategies.

Two preliminary search experiments were actually performed. The first, conducted on a sample of indexed documents, showed that input accuracy was satisfactory and that the total retrieval would not be excessively large. However, a large portion of relevant documents were missed because of search strategy. A second search experiment with improved strategy was performed, and the proportion of missed documents and total retrieval were both reduced substantially.

The initial search experiment involved search queries based on claims of recently issued patent documents. These documents are referred to as source documents. Queries were formulated from each source document and included both specific and broad coordinate index terms and subclasses. A total of 97 specific and broad search queries from 30 source documents was posed to three sets of indexed documents as follows:

(1) A sample of approximately 500 documents was selected from the entire file of nearly 22,000 documents, and these were indexed and reclassified in the normal manner. The average number of documents retrieved from the entire file was estimated from these searches.

(2) Patents cited by the Examiner in the source patent documents were also indexed in the normal manner. These cited patents were termed relevant documents (*4*). The search queries formulated from the source patent documents were posed to the corresponding indexed relevant documents to determine if they were retrieved. If relevant documents were missed, it was possible to determine *why* they were not retrieved. The proportion of documents missed because of input errors was compared to estimates of the same proportion made with the retrieval models.

(3) Searches were also made against the "correctly indexed" file of 60 documents prepared in the indexing experiment to provide estimates of the retrieval model parameters (Q_k).

The general results were revealing concerning expected system performance. The average total retrieval of the entire file was estimated from searches of the first set of documents to be 41.0 ± 17.3 (90% confidence limits) documents per search query. Queries posed to the second file showed that 17 % of the relevant documents were not retrieved because of indexing errors and an additional 44% were missed because of poorly conceived search strategies. A thorough study of the missed relevant documents and models of total retrieval provided considerable insight into alternative methods of searching the file. A second

set of queries was formulated and posed; this experiment will be described later.

Values of the proportion of documents retrieved and the number of documents retrieved with various numbers of subclasses (in logical sum) are dependent on the relationship of coordinate index terms and subclasses used in search queries. For instance, if one were to pick five coordinate index terms randomly and apply them in a search query, a much smaller retrieval would be expected than five terms actually used operationally in a search query. The reason for this is that terms used in a search query have an inherent relationship to one another that also occurs in the text of documents. For instance, the term "Inflammation" is commonly used in combination with a specific body part—*e.g.*, Stomach—and rarely with an organism attacked by a chemical agent—*e.g.*, Protozoan. The correlation of terms is taken into account in the retrieval models by parameters Q_k.

The general search strategy in the first experiment involved searching various conventional subclasses in logical sum and coordinate index terms in logical product. A typical search query had as many as four subclasses in logical sum and anywhere from one to five coordinate index terms in logical product. The first step in analyzing this method of formulating search queries was to generate a table of average total retrieval of search queries having all combinations of one to four subclasses and zero to five index terms. A mathematical model was used to estimate the 24 table entries, since there were not sufficient observations for each. These averages are given in Table III by the number of subclasses and number of terms incorporated in search queries.

The estimated average number of documents retrieved without using coordinate index terms is given in the first column of Table III. On the average, one would expect 105 documents in searches utilizing one subclass, 250 for searches involving two subclasses and so on. The estimated average number of documents per subclass for the entire class is 95.2 documents. The proportion of documents retrieved is given in the first row. This portion indicates the discrimination one can expect as a result of applying coordinate index terms. For example, if one coordinate index term is used in a search query, one can expect to retrieve approximately 20% of the documents in the subclasses which are searched. When five terms are asked, one can expect less than 1% of the documents in the designated subclasses to be retrieved. Thus, average total retrieval is quite sensitive to the number of coordinate index terms used in a query.

The results of the search experiment may appear contraintuitive, since the number of documents retrieved increases by a greater amount than the average subclass size when additional subclasses are used in search queries. As indicated in the table, the average total retrieval for one subclass is 105 documents per search query; and for two subclasses, 250 documents per search query (which is greater than twice 105). The reason for this anomaly is not known for certain. However, one might speculate that subclasses added to search queries are inherently broader and contain more documents. The fact that this phenomenon occurs appears significant and will be investigated further.

The entries in Table III further indicate what one can expect with various numbers of coordinate index terms and subclasses used in search queries. For instance, if two coordinate index terms and three subclasses are used in a combined search query, one would expect a retrieval of 55 documents per search query on the average. If five coordinate index terms and two subclasses are used, the average total retrieval is reduced to 1.5 documents per search query.

It is clear from Table III that one may either use fewer subclasses or more coordinate index terms in his search queries in order to maintain a desirable total retrieval. However, results in Table II indicate that an increased number of coordinate index terms will yield a substantial increase in documents missed because of input errors. Therefore, the number of documents missed because of indexing errors with various numbers of coordinate index terms and subclasses used in search queries is also important.

Estimates of the number of documents missed because of input errors are given in Table IV. The proportion of documents missed because of input errors is found for each table entry by dividing the number of missed documents by the corresponding total retrieval. For this information to be useful, one assumes that the model estimate of the proportion of relevant documents missed because of indexing errors is the same as the proportion

Table III. Average Total Retrieval Estimated by Model for Queries Having 1 Through 4 Subclasses and 0 Through 5 Coordinate Index Terms

Number of subclasses (logical sum)	Number of coordinate index terms (logical products) 0	1	2	3	4	5
	Proportion retrieval					
	1.00	0.203	0.124	0.027	0.011	0.006
All	22,000	4470	2370	954	242	132
4	1,160	235	140	31	13	7.0
3	440	89	55	12	4.8	2.6
2	250	51	31	6.8	2.8	1.5
1	105	21	13	2.8	1.2	0.6

Table IV. Average Number of Missed Documents Estimated by Model for Queries Having 1 Through 4 Subclasses and 1 Through 5 Coordinate Index Terms

Number of subclasses (logical sum)	Number of coordinate index terms (logical product) 1	2	3	4	5
All	285	372	123	70.8	55.2
4	15.0	19.6	6.5	3.7	2.9
3	5.7	7.4	2.5	1.4	1.1
2	3.2	4.2	1.4	0.8	0.6
1	1.4	1.8	0.6	0.3	0.3

of nonrelevant documents missed. That is, if 10% of all documents are missed because of indexing errors, the same proportion of relevant documents will also be missed; otherwise, one must adjust the model estimates.

The assumption mentioned above was tested, in part, by searches posed to the set of relevant documents cited in the source patent documents which were also indexed in the normal manner. Documents not retrieved from this set were studied to establish reasons for their not being retrieved. It was found that 17% of these relevant cited documents were not retrieved because of indexing errors. The model estimate of the same proportion was 14%. (The model estimate of the proportion of documents missed was the sum of the number of missed documents [from Table IV] divided by the sum of total retrieval [from Table III] for the corresponding number of subclasses and terms used in all search queries.) Thus, the basic assumptions of the model appear to hold.

A second search experiment involving a new strategy was conducted on the sample file. Sequence of search queries were formulated from 12 new source patents. The search queries from each source document were ordered from what was considered to be the most specific to the broadest reasonable sets of terms. One of the principal differences in the new sequence of search queries is that a query rarely had more than one subclass and, if so, subclasses were searched in logical product rather than logical sum. This strategy more or less ordered the search responses. Furthermore, in the initial search experiment, queries tended to be either quite specific or quite general. The objective of the initial searches was not to attempt to optimize a search strategy, but to determine within rough bounds what total retrieval might be expected with specific queries as well as with the broadest reasonable queries. In the second experiment, queries were intentionally sequenced from very specific to very general. However, these queries still tended to lie somewhere between the two classes of queries in the initial experiment. The intent of the second experiment was to see whether a particular search strategy (sequencing queries from specific to general) would yield a reasonable total retrieval per query and a reasonable total retrieval for a sequence of queries.

In the second search experiment, 277 queries were applied to the sample file of 500 documents. The average number of documents retrieved from the entire file per search query was estimated to be 13.6 ± 4.2 (95% confidence limits) compared to 41.0 ± 17.3 on the initial experiment. Estimate of total retrieval per search query made by model is 10.9, which is sufficiently close to 13.6 ± 4.2.

One assumes that an Examiner will pursue his sequence of search queries until a set of desired documents is retrieved. The point at which all relevant documents corresponding to their source documents were retrieved was recorded in the sequence of queries. The average number of documents examined per search sequence necessary to obtain the corresponding relevant documents is estimated to be 44.5. One must recognize that these preliminary search experiments were simulated rather than "real life" since (1) the relevant documents were prespecified; (2) the search queries were prepared and put into search sequences in advance of any actual searching—that is, they were not modified as searching progressed, but were organized conceptually in advance; and, (3) the "searching" amounted simply to applying the search query (in the sequence specified by the searcher) to the file and checking to see whether prespecified relevant documents had been retrieved. However, despite these caveats, the search results yielded considerable understanding of the system and aided in making key system decisions.

The estimated distribution of documents retrieved per search query is given in Figure 1. Total retrieval (x)

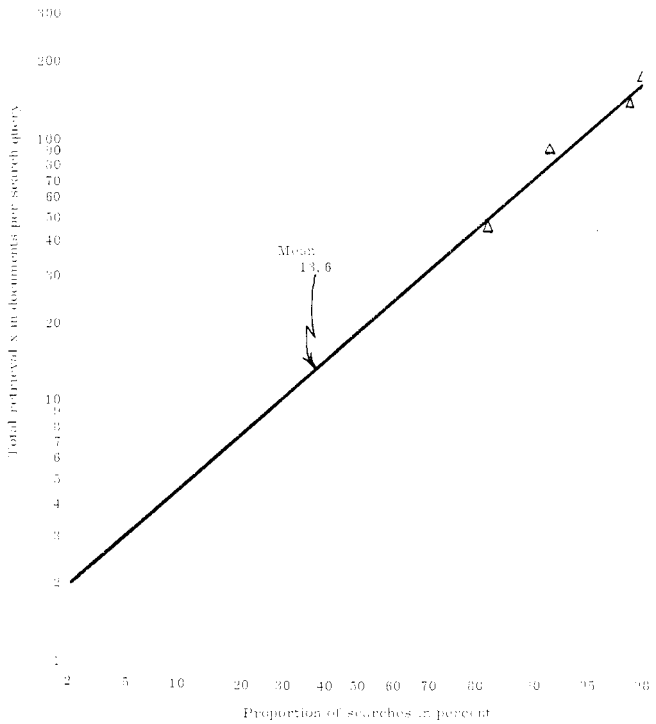

Figure 1. Cumulative proportion of searches having less than x number of documents retrieved.

for the entire file per search query is given on the ordinate and the proportion of searches having that number or fewer is given on the abscissa. Thus, approximately 50% of the search queries will retrieve fewer than 18 documents, 25% of the search queries fewer than 9 documents, and 75% of the search queries fewer than 38 documents. This information is important because one of the major decisions concerning the system involves choosing among various search and retrieval systems. A key consideration in this decision is the distribution of the number of documents which must be physically retrieved once they are identified as having the subclasses and coordinate index terms specified in the search queries.

CONCLUSIONS

The first objective of the evaluation was to predict the performance of the system once it is made operational. The average number of documents retrieved for the entire file was estimated to be 13.6 documents per individual search query and 44.5 per search sequence, well within acceptable limits. A table is provided which gives the average total retrieval per search query for search queries

having various numbers of coordinate index terms and subclasses and for search sequences. Preliminary results indicate that coordinate index terms will considerably reduce retrieval by the classification schedule alone and will still provide representative desired documents.

The next objective was to reveal specific needs, if any, for modifying and correcting the system before changes become too costly. Some modifications of the original term lists were made as an indirect result of the evaluation program in the preparation of the sample files. The expected number and distribution of documents retrieved suggesta a need to investigate further methods for physically retrieving documents once they are identified.

Another objective of the evaluation was to investigate indexing accuracy of Examiners, Classifiers, and Analysts as potential indexers. The indexing experiment showed that input errors involving coordinate index terms should not be an appreciable source of search failures if satisfactory input standards are maintained. The four indexers chosen for preparation of the file all had satisfactory indexing accuracy.

The fourth objective was to ensure that the system will perform satisfactorily from the standpoint of input accuracy. The system appears to be satisfactory in that few relevant documents should be missed either from indexing errors or errors in searching if the proper search strategy is used.

The final objective was to provide a preliminary investigation of search strategies. The study yielded good information concerning potential alternative search strategies. However, further studies of search strategies are anticipated in an operational search experiment to further improve search results and to train Examiners in the use of the system.

ACKNOWLEDGMENT

The authors wish to acknowledge the assistance and cooperation of H. S. Vincent, project coordinator for the design and development of the combined search system, and to thank especially A. J. Cook, F. J. Cullen, D. J. Funderburk, R. L. Huff, A. Mazel, D. B. Moyer, and S. Rosen who participated in the indexing experiment.

LITERATURE CITED

(1) King, D. W., *J. Chem. Doc.* **5**, 96 (1965).
(2) Vincent, H. S., "A Combined Manual and Mechanized Search System," Proceedings of the Fifth Annual ICIREPAT Meeting, in press.
(3) McDonnell, P. M., and Vincent, H. S., "A Combined Manual-Mechanical Search System for Patents Relating to Medicines, Poisons and Cosmetics–Present Status," Paper presented at the Sixth Annual ICIREPAT Meeting, The Hague, October, 1966.
(4) The term "relevant" is used in the sense of being useful to the Examiner. Relevant documents are those cited in source patents. Mere citation, however, cannot be construed as an indication of what one commonly considers relevancy. An issued patent contains a list of all references cited during prosecution of the case for any reason, such as requirement for restriction, rejection of claims later withdrawn, etc. Furthermore, no cited reference fully anticipates the claimed subject matter of the source documents.

Designing an Author-Based Correspondence Information System*

HERMAN SKOLNIK and MARION H. PAYSON
Research Center, Hercules Incorporated, Wilmington, Delaware

Received May 16, 1966

The problems associated with a centralized correspondence information system in a research and development environment are examined. These problems relate to receipt and control, retention periods, circulation, filing, and retrieval. Various mechanisms for solving these problems are considered, and a correspondence information system based on filing alphabetically by author is described.

In many research and development environments, technical correspondence is a storehouse of information almost equal in importance to the report literature and to the library. Indeed, these three categories of document holdings are considered so important in some research and development centers that they are maintained as three separate and distinct document areas. The needs of scientists for, and the nature of, these three categories are such that their separation is often well advised. From the viewpoint of protection and control, the philosophy and policies invoked for reports and correspondence are completely different from those for a library.

*Presented before the Division of Chemical Literature, 152nd National Meeting of the American Chemical Society, New York, N. Y., Sept. 13, 1966.