

## Maurer School of Law: Indiana University Digital Repository @ Maurer Law

---

Articles by Maurer Faculty

Faculty Scholarship

---

1965

# A Legal Document Retrieval System for the Federal Aviation Agency

F. Reed Dickerson

*Indiana University School of Law*

Follow this and additional works at: <http://www.repository.law.indiana.edu/facpub>

 Part of the [Air and Space Law Commons](#), [Cataloging and Metadata Commons](#), and the [Legal Writing and Research Commons](#)

---

### Recommended Citation

Dickerson, F. Reed, "A Legal Document Retrieval System for the Federal Aviation Agency" (1965). *Articles by Maurer Faculty*. Paper 1512.

<http://www.repository.law.indiana.edu/facpub/1512>

This Article is brought to you for free and open access by the Faculty Scholarship at Digital Repository @ Maurer Law. It has been accepted for inclusion in Articles by Maurer Faculty by an authorized administrator of Digital Repository @ Maurer Law. For more information, please contact [wattn@indiana.edu](mailto:wattn@indiana.edu).

# A LEGAL DOCUMENT RETRIEVAL SYSTEM FOR THE FEDERAL AVIATION AGENCY

*F. Reed Dickerson  
Indiana University, Consultant*

## A. Mission

The purpose of this study is (1) to analyze the legal research needs of the lawyers in the General Counsel's Office with particular reference to the finding of pertinent legal opinions and pertinent provisions of the Federal Aviation Regulations and other authoritative sources, and (2) to make appropriate recommendations for meeting those needs.

## B. General Approach

This study begins with an analysis of the legal research needs of the General Counsel's Office. Preferably, solutions should be tailored to problems and not the reverse. The attractions of modern technology are considerable and there is a strong temptation to start with an elaborate computer such as FAA's IBM 1401 and ask, "What legal problems can we use it to solve?" Such an approach, aptly described as "a solution in search of a problem," risks the adoption of overly elaborate techniques, making research not only unnecessarily difficult but unnecessarily expensive.

Between the status quo and automatic data processing represented by electronic computers lies a wide range of data systems that employ no equipment, only modest mechanical equipment, or non-electric business machines. In each instance the objective should be to match the legal researcher's need with search devices that are no more complicated and expensive than necessary.

Even this word of caution may not dispel the widespread feeling typified in Arnold Dumey's experience with a government client for whom, by a mere rearrangement of data, he had reduced operating identification time from 20 minutes to 1 minute. When Dumey asked him how the system was going, the client replied, "Fine, fine. But can't you do it by machine?"

Another, and more serious, disadvantage of the computer-minded approach is that it tends to distract attention from the fundamental problems of classification and indexing. These are the heart of any system of searching and, for the most part, they exist independently of specific hardware.

Even with the limited accomplishments of machine indexing, the problems of classification and indexing must ultimately be solved by human analysis. At present there is no known operational system in which they can be turned over wholly to a computer. This includes even computer-assisted attempts to index every word of search value that appears in the stored text.

On the other hand, it is desirable not to stick wholly to traditional methods of classification and indexing. Modern methods have freed these functions of limitations that until recently were believed to be inevitable. So far as economic considerations permit, it is desirable to exploit the new capabilities for legal research that these methods now make possible.

This study will try first to define a minimum-requirements system and then show how it can be supplemented, step-by-step, as experience demonstrates the specific need. As for the equipment, it will suggest the simple until experience justifies the more complex.

### C. GC Search Needs

Almost every branch in the General Counsel's Office is engaged, in addition to the performance of its other duties, in writing legal interpretations. Some of them relate to the Federal Aviation Act of 1958. Others relate to matters of general law such as government procurement, personnel, or evidence. The great bulk of them relate to FAA's own regulations, now called "FAR's," or to their predecessors, the "CAR's" (Civil Air Regulations) and the Regulations of the Administrator.

In the writing of these interpretations it is often necessary to find and refer to earlier ones. The existing files of interpretations of current interest for precedent purposes go back to 1939. It is also often necessary to find and refer to the relevant provisions of the regulations, statutes, executive orders, and agency policy statements. Lawyers preparing new FAR's often face the same search problems.

At present the most serious search problems relate to the findings of pertinent legal interpretations. A legal interpretation that is considered important enough that copies of it should be sent to regional offices is called a "legal opinion." It is digested and put in a separate file. If it is not so important, it is simply filed, undigested and unindexed, where it remains a "legal interpretation." For the purposes of this study, only interpretations important enough to be classed as "legal opinions" will be considered, because only they are relevant to current legal research.

The current method of handling GC legal opinions is to include in the master file only those legal interpretations for which the author has made a headnote card, and to file the opinions according to a general subject-matter arrangement or according to the parts of the FAR's that are respectively affected. This means that from 1 to 7 copies of the opinion and its headnote need to be made. (The average falls somewhere between 1 and 2.)

The current storage system has been defective in at least three important respects. First, the file of legal opinions has been incomplete in that it has omitted many legal interpretations that should have been digested and treated as legal opinions. Directive 1-6 of the General Counsel, dated December 15, 1959, now numbered GC 2000.6, provided for digesting and indexing at the source (by the lawyer writing the opinion), leaving only ministerial functions for the Files Supervisor. Unfortunately, in the rush to meet day-to-day deadlines, the great bulk of interpretations that belong in the opinion system have not been digested or indexed by their authors and thus not stored as "legal opinions." Accordingly, they have become for all practical purposes inaccessible.

The total number of legal opinions now stored and indexed in this system is about 335. This aggregate represents a 4-3/4 year period in which many times that number of legal interpretations have been prepared. Of the many that should have been digested, indexed, and stored as legal opinions, about 90 percent have been omitted.

Fortunately, a recent project has brought most of these into the opinion system. Jonker Business Machines, Inc., under a contract with FAA, has assembled a much more adequate set of GC legal opinions. The effort has yielded an aggregate body of opinions numbering about 3500, covering on the average about two legal issues, and consisting on the average of about 2.14 pages.

The second deficiency in the current storage system has been that some opinions have lacked the information necessary to make them adequately indexable or useful. Many, for example, have failed to include an adequate recital of the fact situations to which they are addressed. Mere reference to the inquiry made in another document is inadequate unless the document referred to is attached to and made a part of the opinion. Others have failed to recite all the vital statistics relating to the opinion that are necessary to adequate selection and identification: the statutory or regulatory authority on which it is based, the person for whom it was written, the person by whom it was written, the date, a short descriptive subject heading, and its approval for inclusion as a "legal opinion."

Third, the opinions have been stored according to a subject-matter arrangement, which requires as many copies of each opinion as there are subject headings under which it is filed. A storage system so arranged is unnecessarily bulky, unwieldy, and expensive, even with shallow indexing.

In every research system, there is in addition to the storage element the element of retrieval. Although closely related, these two elements need to be evaluated, and often treated, separately. One of the serious limitations in traditional systems is that they have often been combined when they should have been kept separate.

GC's most troublesome research problems seem to lie in the area of retrieval rather than storage. Although most of the basic materials are already on hand in usable form (at least in the Washington office), they remain largely inaccessible. The present retrieval system is built around a central card file of opinion digests arranged according to the relatively crude system of subject headings by which the opinion themselves are arranged.

This classification system was first outlined in the Directive of December 15, 1959, mentioned above, which appended a hierarchical subject-matter classification with eight main subject headings. ("AIRPORTS," "AIRSPACE RULES," "CONTRACTS," "ENFORCEMENT," "GENERAL LAW," "LEGISLATION," "LITIGATION," and "SAFETY RULES"), with subheadings ranging in number from about 10 (for "LITIGATION") to about 73 (for "CONTRACTS"). Under the main headings, "AIRSPACE RULES," "ENFORCEMENT," and "SAFETY RULES," classification is according to relevant section number of the FAR's. This system has since been expanded and refined to increase the number of main subject headings to 12 (yellow tabs) and to insert about 48 subheadings (white tabs). In general, it adopts EAA's classification system for correspondence (see Appendix 2, §1, OA P 1320.1 (3/9/62)), FAA Directive Systems Handbook). Unfortunately, the latter was designed mainly for storage purposes, rather than for retrieving opinions that need to be constantly referred to.

Besides having an inadequate conceptual matrix, the current indexing system is insufficiently uniform in depth, format, and operation.

The inadequacies of this system have not been compensated for by the fact that several GC divisions have their own files or indexes of opinions of special interest to them, because the reach of the searcher has been limited accordingly. The Office needs a consolidated, deeper, and more uniform index that covers all its legal opinions. The same may also be true of the FAR's and relevant statutes, executive orders, and Agency policy statements.

#### **D. A Minimum Requirements System For GC Opinions**

##### *1. Improving GC's storage system*

The first steps necessary to improving GC's storage and retrieval system are (1) to require that each opinion show on its face the elements necessary to selection and identification: the facts, relevant statutory or regulatory authority, addressee, author, date, descriptive subject heading, and approval for inclusion, and (2) to centralize the administrative responsibility for making sure that the interpretations that belong in the legal opinion file are actually so included. If these things are done, future Jonker-type surveys should become unnecessary.

Second, it is desirable not to complicate the storage of legal opinions with retrieval aids that require opinions to be arranged according to subject matter. The fact that the same opinion fits logically under several headings requires it to be duplicated to provide as many copies as there are headings. In such a system, the deeper the indexing the more unnecessarily complicated and unwieldy it becomes. This is not merely a problem of bulk.

Instead of being arranged according to subject matter, a large mass of materials that is being searched from many points of view should be filed in simple order of date of preparation or date of absorption into the system (no particular order is required). A numerical sequence is established by assigning to each item a document, or "accession," number or, if regional duplicates are to be on microfilm, a microfilm number. Such a system requires only that the storage file include the original of each document. It has maximum flexibility, because it is freely open-ended: new, and even old, legal opinions can be absorbed simply by adding them at the end of the file and numbering them serially. Storage is unencumbered by the many difficulties of classification.

##### *2. Improving GC's Retrieval System*

As for retrieval, the key to any storage facility, whether operated electronically, mechanically, or manually, is an adequate index. Because the index is the almost universal entry point for every system, it should get the fullest attention. GC's main legal research problem, therefore, can be solved by creating an adequate master index, one that gives ready access to the document numbers of the relevant GC legal opinions. To this result, unfortunately, there are no easy short cuts.

The limitations inherent in traditional methods of indexing have created two serious deficiencies: (1) lack of depth, and (2) limitation of full effectiveness to searches that reflect the same point of view as that reflected in the index.

Methods now available free indexing from both limitations. It is now possible to index as deeply as the objectives of the user require. Thus, the

indexer can use as many index terms as he thinks the searcher will need for flexible access to the contents of the particular document. It is also possible to index so that the system can be entered with the same facility regardless of the point of view of the user and regardless of the level of generality on which he searches. This is made practicable by avoiding hierarchical indexing in favor of linear (non-hierarchical) indexing. With the latter, each term, regardless of whether it is a general legal term or a specific factual one, is given equal dignity with every other term. This is done simply by placing all index terms, including those designating subclasses, in alphabetical order. It provides maximum flexibility, because relative importance, which may vary according to the searcher's viewpoint, is not reflected.

For example, under hierarchical indexing the following concepts would ordinarily appear in only one place in the index and could be located only through the parent term "aircraft":

- Aircraft
  - Fixed-wing aircraft
  - Rotocraft
    - Normal rotocraft
    - Transport rotocraft

Under linear (non-hierarchical) indexing, on the other hand, the same concepts would take their alphabetical places in a single list and each could be located through its own term:

- Aircraft
  - \* \* \*
- Fixed-wing aircraft
  - \* \* \*
- Normal rotocraft
  - \* \* \*
- Rotocraft
  - \* \* \*
- Transport rotocraft

The significant difference here is not so much one of form as one of substance. For example, it is sometimes typographically desirable in an index to take index terms that have a common initial element and combine them, using that element as a common heading and letting the remainder of each term stand as a subheading. Although this gives the result a hierarchical appearance, substantive linearity exists so long as each such subheading either has no search value or has also been carried in its appropriate alphabetical location in the primary list as an index term or as the initial element of an index term.

Conversely, an apparently linear list in which such groupings have not been attempted is, nevertheless, substantively hierarchical wherever an index term consists of several elements each of which has significant search value but not all of which also appear in their respective alphabetical positions in the primary list. Here, again, substantive linearity can be achieved by making sure that each such component does so appear.

The most flexible modern indexing is not only non-hierarchical but "coordinate" in the sense that, except as explained below, every item in the index, whether a word or a phrase, should represent what, in the light of FAA's needs, is considered to be a significant single concept (e.g., "air-

worthiness," "Aircraft," "fixed-wing aircraft"). Search problems involving combinations of concepts (e.g., "airworthiness of fixed-wing aircraft") can be handled in most cases by matching the perspective document entries for the concepts involved to find those common to each. When volume and depth of indexing are not too great, this can be done manually (a device for more sophisticated systems is described in paragraph I, below). Coordination on a word-by-word basis (e.g., "fixed-wing" and "aircraft") rather than a concept-by-concept basis (e.g., "fixed-wing aircraft") is undesirable, except where the single word is an accepted designation for a relevant concept.

A weakness of coordinate indexing is that it tends also to produce "false drops" (unwanted documents). This happens wherever the coordinate terms in a search question happen to appear in a document but are unrelated to each other. For example, if the searcher is interested in the airworthiness of fixed-wing aircraft and his search question is phrased in terms of the logical product of "airworthiness" and "fixed-wing aircraft," the search will yield not only documents that deal with the airworthiness of fixed-wing aircraft but any document that, for example, happens to deal with both (1) the airworthiness of rotocraft, and (2) the cost of fixed-wing aircraft. This tendency to include unwanted documents can be reduced by adding to the list of coordinate concepts the composite index terms for which there is a recurrent search need (e.g., "air-worthiness of fixed-wing aircraft").

One approach to building a subject index is to start from scratch to analyze the text of the opinions and select the terms that appear to be the most helpfully descriptive of their contents. On the other hand, because there are existing indexes of opinions and other FAA materials of likely relevance (such as the current GC subject index and the one Alan Barr developed in Oklahoma City), it is probably preferable to begin with these, dumping them into a mulligan stew of potential index terms. To these can be added whatever additional terms GC lawyers, after an examination of the raw documents being stored and with the help of a skilled indexer, select as helpful for search purposes.

An alphabetized list of such terms without documentary references would constitute a tentative "thesaurus" (or "concordance" or "word list"). As such, it would provide the basic fund from which the final product would emerge. Because it contained all doubtful and marginal terms, the initial list would undoubtedly be longer than ultimately desirable.

Such a preliminary thesaurus should be cleaned up to remove surplusage and the potential duplications and overlappings created by synonyms and near-synonyms. In each case the most appropriate index term (the one most likely to occur to the typical searcher) should be selected as the primary link with the relevant documentation. So far as useful for search purposes, synonymous terms should be retained, but only for the purpose of cross referencing to the primary index term. A thesaurus should also include helpful cross references between terms referring to closely related, though different, concepts.

In making such a list, it is important to nail down the primary index terms during the initial indexing of each document; later it is hard to add new primary index terms for those previously indexed. (The same is not true

of new terms appearing in documents not already indexed.) New cross-references, on the other hand, can easily be added at any time.

When the thesaurus (basic word list) has been cleaned up, it can be converted into a thesaurus-index by adding after each primary index term the document numbers of the relevant documents that have been stored. The alternative approach is to develop a separate index. A separate thesaurus would consist of the primary and secondary index terms and cross references (but without document references). A separate index would consist of the primary index terms plus document references (but without secondary index terms and cross references).

Whether an index can be satisfactorily combined with a thesaurus depends on the respective volumes of legal opinions indexed, index terms used, and cross references. Although the greater flexibility required by more massive systems has prompted their separation in the Department of Justice's "LEX" system for searching anti-trust materials (see Appendix C1) and in the system now being developed by C.I.A. (see Appendix C2), it is recommended that FAA use a combined thesaurus-index until the need for separate search tools becomes plain.

In exploiting the more flexible resources of modern indexing, the indexer should be careful not to over-index, (i.e., index overly narrow concepts or those with little search value or logical relevance). Over-indexing increases the risks that the indexing will be done inconsistently and that by searching in overly specific terms the searcher will overlook important, relevant documents. Despite this, it is desirable to index on several levels of generality. Here, no advice is possible beyond the general observation that good indexing depends on the practical and informed judgment of lawyer-indexers sufficiently acquainted with the material indexed and the needs of FAA. (Most traditional legal indexes suffer from over-generality.)

To get the feel of these matters, and especially to get a more accurate estimate of how deep the index needs to go, it would be desirable to have several GC lawyers who are thoroughly conversant with current legal search problems index a selected sampling of recent opinions. Needless to say, the present average depth of between one and two terms per legal issue is seriously inadequate.

One of the most useful entry points in the present system of digesting and indexing GC opinions is through the relevant provisions of the Federal Aviation Act or applicable Agency regulation. It is desirable to preserve this feature in any new indexing system. An index of opinions by applicable section of the Act and by applicable section of the FAR's should therefore be created and appended to the thesaurus-index (or index).

So far as it relates to Agency regulations, this part of the index can be designated and arranged according to the existing Act and FAR systems. Opinions interpreting earlier regulations, whether CAR's or Regulations of the Administrator, should be listed under their respective counterparts in the FAR's. This can be done with the help of the codification distribution table in Part 1 of Title 14, Code of Federal Regulations. Opinions interpreting earlier regulations but having no counterparts in the FAR's presumably have no current force or significance and can safely be ignored (this assumes that the current codification project is completed before the new



index). So can earlier opinions that were codified in the preparation of current FAR's. Only citations of the Act and the FAR's would appear as index entries. GC opinions relating to the Federal Aviation Act can be handled similarly. Presumably, these entries would precede those relating to the FAR's.

One benefit of this part of the index would be to provide the basic materials, properly arranged, for a full annotation of the Act and the FAR's. Such an annotation will be discussed in a later paragraph.

Whatever approach is adopted, individual index terms and other entries can be key (or paper) punched and machine sorted alphabetically. What particular devices, electronic or mechanical, will turn out to be the most appropriate for sorting and print-out, taking into account cost and other economic considerations, can be determined only when the specific operation can be more precisely defined and analyzed with the help of FAA's systems experts. At this stage, such an analysis is not feasible.

For purposes of machine processing and cross checking, each index term (e.g., "Damages") should be given a code designation (e.g., "D-116"). With this aid, visual cross checking can be used to guard against human error. For example, misspellings can be intercepted and kept from producing separate index terms during machine sorting. Once these designations have served their purpose, they should be dropped so as not to encumber the final index. If found to be economically acceptable, machine sorting can be done on FAA's IBM 1401.

In the assignment of code designations, adequate "air space" should be left between specific items so that other items can later be inserted at their appropriate places and designated without disturbing the numerical order of the existing code designations.

After machine sorting, the resulting print-out can be photographically reduced for publication and distribution as a practical search tool. Machine sorting is desirable, if economically feasible, because it is faster and more accurate than manual sorting. Moreover, the print-out capabilities of a computer combined with photographic duplication and reduction avoid the expense and inaccuracies of separate printing and proofreading operations.

With respect to the final form of the thesaurus-index, hard copy provides the greatest possible accessibility. It makes it possible to provide each lawyer, whether in Washington or a regional office, with a copy of the primary search tool. With this at hand, showing as a minimum both the document numbers and the descriptive titles, the lawyer can do much of his preliminary searching at his desk, thus reducing the need for pulling specific legal opinions from the storage file. In addition, the familiarity of its search format minimizes any problems of user acceptance.

A thesaurus-index, like a separate index, must include as a bare minimum the respective document numbers. To facilitate search, it is also desirable to include with each document number a short title or heading describing the specific subject of the opinion, including a reference to the pertinent statute or regulation. It might also be useful to include the date of the opinion. Without cluttering the thesaurus-index, this information makes it unnecessary for the searcher to consult the document storage file to eliminate documents that are obviously of no interest.

The up-dating of the storage file, as we have seen, can be handled by

numbering the opinions in the order in which they are adopted and sending copies to the regional offices. These should be distributed as soon as they are issued. The printed thesaurus-index, on the other hand, should be revised and distributed only periodically. This can be done either by reissuing the thesaurus-index after a substantial time has elapsed during which a sizeable number of new opinions have accumulated (e.g., quarterly or semi-annually), or by maintaining the thesaurus-index on a loose-leaf basis with perhaps bi-weekly or monthly issues of new replacement pages.

Under the circumstances now anticipated, it is believed that the former approach would be both easier and cheaper. It is anticipated that with proper methods of selection and administration about 25 new legal opinions will be added to the storage file each month. Assuming an average depth of indexing of 8 terms an opinion, there is strong likelihood that upwards of 400 changes in the thesaurus-index would be involved each month. These would probably affect a large number of pages and, accumulated over the same period as would be used for a complete reprinting, the total number would tend to approach the total number of pages in the thesaurus-index. At best, therefore, paper saving would be a minimal advantage to a loose-leaf system. Even this would be reduced insofar as successive batches of changes affected the same pages of the thesaurus-index, because those pages might have to be replaced as many as several times during the period which, under the method of complete replacement, they would have to be replaced only once.

Finally, the small net saving in paper, if any, resulting from using the loose-leaf approach would probably be more than offset by the man hours needed for filing replacement pages.

Assuming an average increase of 25 legal opinions a month, quarterly replacement would seem to be adequate. For the growing accumulation of new indexed opinions, which would probably not exceed 75, no formal supplementary index would appear to be necessary or desirable. A current reading file would probably serve the current search need.

The system recommended contemplates that the master storage file would be kept at the headquarters office (presumably in the law area of the FAA library) and one set of copies would be kept in each regional office. Xerox copies would be more usable but, if expense became an important factor, the regional files could be put on microfilm. (Before doing the latter, it would probably be desirable to poll the regional attorneys as to their preferences.) Although only 10 sets of basic legal opinions would be needed, thesaurus-indexes should be distributed to all lawyers in the headquarters office (62) and to those in the field (37), a total of 99. The indexing function, on the other hand, should be located, at least initially, in the Office of General Council (see paragraph O, below).

The system just described is a rock-bottom, no-frills system, one that can be supplemented, as experience shows the need, with refinements and additional search aids. (Some of them will be described in later paragraphs.) If it looks complicated, it is only because the preparation of an adequate index is never simple.

A consistent, unified indexing system is desirable not only on its own account but in case the system develops in size and complexity to the point where searching by machine or computer becomes desirable.

## **E. Separate Thesaurus and Index**

If the volume of stored materials increases greatly, or if the thesaurus-index recommended in paragraph D otherwise proves unwieldy for searching the existing materials, it may become desirable to separate the thesaurus from the index. The advantages of separation for any large, sophisticated system are that (1) the thesaurus can be used more flexibly as a source of browsing for helpful search terms, because it is unencumbered with a vast number of often repeated document numbers and other information, and (2) the index can be used for more quickly tracing the relevant document numbers for the relevant primary search terms, because it is unencumbered by cross references and synonymous and related terms and their accompanying document numbers.

The later creation of a separate thesaurus would require the key (or paper) repunching of all items carrying document entries in the thesaurus-index.

## **F. Opinion Digest**

Although the descriptive document titles appearing in the indexes can serve as rudimentary digests, experience may show that the lawyer's time could be further saved, and undesirable congestion at the storage file reduced, by creating a separate list of opinion digests, showing in greater detail the critical aspects of each opinion and arranged according to document number. (This, too, has been done in the Department of Justice's "LEX" storage and retrieval system.) Such a device would enable the lawyer to screen materials of potential interest more thoroughly without leaving his desk.

Unlike the descriptive document titles in the subject index, each digest would appear only once in the digest list: at the location determined by its document number. For each document, the digest should be prepared in the first instance by the author of the original document; in default of him by the central indexer. The rudiments of such a digest would, of course, be suggested by the list of primary search terms developed for that document by the indexer.

Here, again, machine processing would be useful. Each digest entry could be key (or paper) punched, and the aggregate machine sorted. The printout could be photographically reduced and published similarly to the thesaurus and index.

As with other search tools not considered to be minimum essentials, preparation of a list of opinion digests should be deferred until experience shows the specific need.

## **G. Search Strategy**

Under the recommended minimum requirements system, the entry point would normally be the thesaurus (if combined, the thesaurus-index). Examination of the thesaurus would disclose the primary search words through which the primary materials would become available. Through its cross references, the thesaurus would allow the searcher, by browsing, to orient his search inquiries. Then, having tentatively formulated his inquiry in

terms of the relevant primary search words, the searcher would examine the appended document entries.

When the question involved the logical product of several search terms (i.e., searching in combination), he would match the relevant document lists and look for the document entries common to those lists. An examination of their descriptive titles would enable him to eliminate obviously inapplicable documents. Use of the list of opinion digests, if any, would reduce the number of potentially interesting documents still further. The remainder he would evaluate by going to the files and examining the documents themselves. All of this is, of course, conventional search strategy.

Searchers armed with pertinent section numbers of the Federal Aviation Act or the FAR's would enter the system through the index of opinions according to the respective locations of those numbers in the Act or the FAR's.

## **H. Alternative Approaches: "Conventional" Card Systems**

If printing a thesaurus or an index would be too expensive, the volume of documents or their search terms is not too large, and the traffic in opinion searching is not concentrated too heavily on the common search tools, a centralized card system might be adequate. However, unless the system can be handled manually, it is likely to be more expensive and unwieldy than the recommended hard-copy thesaurus and index. The information now available suggests that for FAA this would be the case.

The simplest method of classifying and sorting documents is to code on a card representing the source document each of the relevant search terms needed for search purposes. The result would be a card file arranged by document number. For quick sorting, a specific location on the card form must be assigned to each search term used in the system (whether or not it is relevant to the particular document). This system of "dedicated spaces" makes it possible to sort (1) in very simple systems, by scanning the card file, search term by search term, to see which cards have been checked at that location, (2) in more complicated systems, by successively inserting a rod in the pre-punched holes respectively assigned to the search terms used in the system, to see which cards drop out as a result of having been notched at the relevant locations, or (3) in still more complicated systems, by the machine sorting of cards that have been key-punched or otherwise coded on the same basis. In any non-electronic system involving thousands of documents, not to use dedicated spaces greatly impairs the usefulness of the system, because in effect it deprives it of an adequate subject index.

This kind of grouping, in which each card represents a document, is considered "conventional." Unfortunately, because of the use of dedicated spaces, the limitations of each specific system are defined by the total volume of search terms used in it. In general, its utility is limited to the number of search terms that can be expressed or coded on a single card or sheet. For a Zatoperk edge-notched card, the maximum is 65. For a Hollerith card, the practical maximum is 840. For a Recordak Miracode code sheet (film medium), the maximum is 1000. Thus, such a device is suitable only for a system in which the total number of search terms in the index does not exceed the capacity of the particular kind of card or sheet.

When the capacity is so exceeded, it is necessary to go to a device with a higher descriptor capacity. Theoretically, through the device of super-

imposed coding, higher capacities are possible. However, for practical purposes, the need for higher descriptor capacity is more appropriately handled by a computer.

How great a descriptor capacity should GC anticipate? In gaining access to anti-trust materials, the Department of Justice's "LEX" system uses about 3500 primary index terms. The Chief Counsel's Office of the Internal Revenue Service has already developed about 5000 index terms and expects ultimately to quadruple that number. At this time it is possible only to guess the total number of index terms needed to gain access to GC's legal opinions. However, it seems likely that the volume of opinions and depth of indexing will produce a number requiring, for any conventional card system, the use of facilities and equipment costing as much as or more than a printed thesaurus and index, but without their practical advantages.

Another serious limitation on this kind of card index is that, although it may substitute in some situations for a printed index, it is an inadequate substitute for a printed thesaurus, because the limitations imposed by the size of the card on the number of search terms that can be included make it necessary to eliminate most if not all synonyms that might otherwise be useful for search purposes.

Also, every card system has the inherent limitation that, being a single device, it must be shared by all the lawyers interested in GC legal opinions. Not being immediately available at each desk, it is not only inconvenient but subject to bottlenecking wherever simultaneous use becomes significant.

If a conventional card system is used at any point, whether as the principal search tool or as a step in developing that tool, it is highly desirable, if the cards themselves are not in machine-readable form, that they be typed on a Flexowriter, SCM Typetronic, or similar device for producing a punched tape or other machine-readable byproduct. This takes little or no extra labor and improves the compatibility of the card system with more sophisticated systems to which experience may make it desirable to convert. Otherwise conversion may be unnecessarily expensive.

## **I. Alternative Approaches: Inverted Card Systems**

When the number of primary search terms becomes very large and especially when indexing has been done in depth and the searcher often searches in combinations of search terms (as when the searcher wants every GC opinion that deals with "pilots," "VFR Rules," "Chartered flights," and "Alaska"), a card system using so-called "inverted grouping" should be considered. Here, the critical unit is not a card that represents a document and lists the relevant search terms, but a card that represents a search term and lists the documents to which the search term applies. (Such a card corresponds to a primary index word as it appears with its documentary references in a conventional index.)

With inverted grouping, which can be used also by digital computers, the number of recordable search terms is limited only by the number of cards that it is feasible to handle. Theoretically, the number of usable search terms is unlimited, a capability that helps to make possible deeper indexing. Moreover, unlike "conventional" grouping, inverted grouping permits secondary search terms (e.g., synonymous or related terms likely to be

useful for search purposes) to be interspersed at their appropriate alphabetical locations in the card file.

The significant limitation on inverted grouping lies, instead, in the number of documents that can be searched. The manually usable Port-a-Punch card, for example, although unlimited as to search terms, can accommodate only 480 documents. The standard Termatrix descriptor card, requiring modest mechanical equipment, can accommodate 10,000 documents. The card used by the Bureau of Standards can accommodate 20,000. A microfilm variant of the Termatrix card can accommodate 100,000.

Digital computers have, for present purposes, a capacity that is practically unlimited. FAA's IBM 1401 could undoubtedly handle the Agency's legal retrieval problems with more than enough capacity to spare.

The capacity of a Uniterm card, which does not dedicate spaces for particular documents, is taxed only by the demands of the descriptor card having the highest incidence of applicability to the documents indexed. Although non-use of dedicated spaces gives a Uniterm card a higher potential document capacity than a card of equal size that uses dedicated spaces, the absence of dedicated spaces impairs its utility for quick searches by combination of search terms.

Facility and speed in such a search depends on the searcher's being able to match the corresponding document references for each term that define the search question. This disability is easily seen in the conventional printed coordinate index, where not only are the search terms and their entries widely separated but the documents common to each of the relevant search terms are likely to appear in different relative locations in the document entries for the respective search terms, thus making comparison difficult. Together with the use of the hierarchical form, these two factors create probably the most serious limitations in the traditional indexing of law materials. Moreover, they are formidable reasons why traditional indexing is comparatively shallow.

Under the Uniterm system, as with conventional indexes, it is necessary to compare the respective lists of document addresses one-by-one against the results of the immediately preceding comparison. Although this relatively inefficient method of matching document entries is used also by digital computers, their great speed makes this limitation minimal for legal searches.

Where the general capabilities of such a computer are not needed, on the other hand, the use of dedicated spaces becomes highly desirable, because they greatly facilitate the matching of corresponding document entries. The best known system for simultaneous matching uses optical coincidence (commonly called "peek-a-boo"). Here, each document is assigned a specific location on every search card in the system, whether or not the particular search term applies to that document. The applicability of a particular search term to a particular document is then shown by punching or drilling a hole in the search term card at the location dedicated to that document. Under this system, when all the search term cards that define the search question are stacked, light shines through only at the locations that represent documents to which every selected search term applies. Commercial versions of this system include Port-a-Punch, Termatrix, and Keydex.

Despite its impressive advantages, a system that uses optical coinci-

dence of this kind shares some of the general limitations of conventional card systems: (1) the inconvenience of having to use a single, centralized facility, (2) the possible delays and traffic problems caused by simultaneous use of that facility, and (3) the difficulties of updating, especially when parallel research facilities are being maintained at remote locations.

It is difficult to update such a system because new materials cannot be absorbed simply by adding new cards at the end of the file. Instead, all affected descriptor cards must be (1) replaced or (2) pulled and stacked, separately for each document, for the purposes of punching or drilling. The latter approach can become a serious mechanical problem when duplicate systems are being maintained for FAA's regional offices. Accuracy and economy require that all systems be up-dated simultaneously. Serious delays could develop if as much as one regional office did not send in its cards for up-dating. The problem can be reduced, of course, by incurring the expense of issuing new cards with each updating.

It is recommended, therefore, that unless the difficulties of searching in combination of terms become serious the optical coincidence ("peek-a-boo") card system not be adopted.

If, on the other hand, the need for such a system arises, it can be developed readily from the thesaurus-index developed under the minimum requirements system already described. Of course, if deeper indexing is attempted at this time, the added terms would simply be reflected in additional descriptor cards.

The recommended minimum requirements system is, therefore, fully compatible with an inverted card system.

## **J. Alternative Approaches: "KWIC"**

Keywords-in-context indexing ("KWIC") is a means of indexing documents by the words used in their titles. The index is generated by a computer by following a dictionary of key words or, more simply, by selecting every word in each title except those appearing in a list of common words with no search value (e.g., "a," "the," "from," etc.). For each selected word the machine makes an index entry in which the key word appears in the middle, flanked by the words of the original title most closely preceding it and those most closely following it. For example, an entry in a KWIC index looks like this (the key word has been underlined to identify it):

of Titles with a                      Subject Heading Classification

Breaking each entry just before the key word makes the word conspicuous enough that it can easily be seen and the entries machine-alphabetized accordingly.

The main advantage of this system is that the indexing is done automatically, quickly, cheaply, and with a small use of manpower. Its main weakness is that, unless supplemented by humanly-selected index terms, the result is limited in depth and quality by the depth and quality of the existing document titles. Indexing limited to conventional titles naturally tends to be shallow. An analysis of the KWIC index for the American Bar Foundation's *Index to Legal Theses and Research Projects—1960-1961* showed that apart from depth, about 65 percent of the time the computer

produced the same results as a human indexer (see 11 American Documentation 48, 51 (1964) ).

If sufficiently informative titles already exist, the KWIC system will produce very satisfactory results. However, if the titles are inadequate or non-existent, adequate titles must first be prepared. Because the production of such titles requires the judgments necessary for initial indexing, much of the labor-saving value of computer-generated indexing is lost. Because the titles of GC opinions appear to be insufficiently informative, it is doubtful that KWIC indexing could offer much in alleviating GC's search problem in this area. Because human indexing appears to be unavoidable, KWIC indexing offers only the advantages of machine sorting, which, of course, are not peculiar to it.

## **K. Searching by Computer**

Despite the many efforts in the field, few systems for searching legal materials by electronic computer (as distinct from using one to help build a traditional hard-copy index) are actually operational. Perhaps the most important of these is the system of full-text searching, discussed in Appendix C(3), that is used by John F. Harty of the University of Pittsburgh in dealing with the statutes of Pennsylvania. Some difficulties are technical. The most important, however, are over-elaborateness for most search purposes, inconvenience, and high cost.

Inconvenience inheres in the fact that interposed between the lawyer and the computer are technical aspects that require him to formulate his question very precisely. For a system such as proposed for FAA, the inconvenience and time involved would appear to be out of proportion to the difficulty and significance of most search problems of the typical GC lawyer. These difficulties, compounded by any shortage of machine time, seriously limit his capacity to browse. Ordinarily, he needs several shots at this material.

Searching by computer, in addition to being over-elaborate and inconvenient for most legal searching, is expensive. These factors restrict the kinds of legal searches that warrant the use of a computer to those involving an appropriate combination of size, complexity, importance, and need for speed.

If, because of unforeseen conditions, an appropriate combination of these circumstances should later exist in GC, the system recommended for immediate adoption is, fortunately, fully compatible with machine searching (other than of full text). A hard-copy thesaurus, such as the one described, is necessary in every search system, and would be retained. The printed index, on the other hand, would be replaced by putting each index term and its family of document numbers into machine-readable form and transferring them to tape, disk, or other electronic memory. The same could be done for the list of opinion digests. It would seem to be unnecessary to store full text in electronic form (as is done under the Harty system).

On the basis of present information, however, it seems unlikely that in the foreseeable future GC will need to search its opinion files by computer. A computer, of course, is useful in helping to build more adequate traditional search tools such as a hard-copy index. When that is done, the input should



be formatted to permit machine recall on a coordinate basis, in case machine searching later becomes desirable.

## **L. A Regulations Index**

Discussions with GC attorneys indicate the strong advisability of preparing an index to the Federal Aviation Regulations comparable to that needed for GC legal opinions.

Here the document unit carrying an index identification would be the section or subsection, and its number would serve as an appropriate reference in the index.

The determining considerations are the same as those that apply to GC opinions. Again, volume and depth of indexing should be decisive factors in determining the nature and form of the index. Although the codification program is not yet finished, it is estimated that the completed FAR's will consist of about 2723 sections averaging in length about 3/16ths of a page.

Although the total text will probably be substantially shorter than that of the total body of GC legal opinions, the greater compactness of drafted documents, as compared with legal opinions, is likely to call for correspondingly deeper indexing (presently estimated at about 10-12 terms per section).

Although both factors suggest that the same kind of minimum requirements indexing system as recommended above for GC opinions would be appropriate here, the greater depth of indexing and the lower bulk of stored materials increase the probability that experience with such a system will show the need for searching in combinations of terms (as through a system using optical coincidence) or even searching full text with the help of a computer. However, current knowledge suggests that the differences will not be significant enough to warrant a materially different system of indexing. Fortunately, the minimum requirements system is compatible with such systems and the decision to try them can be deferred until the capabilities of the more modest systems have been adequately tested.

Despite the large overlap in index terms that is likely to occur, the FAR subject index should be developed separately from the GC opinions subject index. After both have been prepared, the extent of the overlap can be determined and, if the systems are compatible, a decision made whether greater usability is to be gained by consolidating them or by keeping them separate. Elements affecting the desirability of consolidation will be (1) the total number of index terms, (2) the percentage of non-overlapping terms, and (3) the average length of the document references under the primary search terms. The larger these elements are, the less desirable consolidation will be.

## **M. Other Indexes**

Although the need is not so critical, it would seem desirable also to develop (1) an index to the Federal Aviation Act of 1958 and other statutes of interest to FAA, (2) an index to the pertinent executive orders, and (3) an index to FAA's policy statements.

Like the index to the GC legal opinions and that to the FAR's each index should be developed separately. Which indexes, if any, should then

be consolidated will depend on what experience reveals with respect to the factors, affecting consolidation, that were discussed in the preceding section.

## **N. Annotated Act and Regulations**

The parts of the recommended legal opinion index that follow the subject matter entries (those that index by reference to the pertinent sections of the Federal Aviation Act and the Federal Aviation Regulations) will provide the basic elements for a fully annotated Act and regulations system. Although there is little doubt that annotations of the Act and the FAR's would be useful to GC attorneys, it is possible that the need would be adequately served by the document headings listed according to section in the basic index, without a full-blown annotation. This possibility should be tested.

If later experience shows the need, a full-blown annotation can be built by (1) inserting the full text of the section of the statute or regulation after each corresponding section-number heading, and (2) adding, after the references to the GC legal opinions, the citations and brief synopses of the relevant executive orders, FAA policy statements, and decisions of the courts, Attorney General, Comptroller General, and Civil Aeronautics Board.

## **O. Executing the Program**

The development and maintenance of a subject index of voluminous legal materials to be explored in adequate depth requires centralizing the function in persons who are expert in indexing, who are conversant with the improvements in indexing made possible by modern methods and devices, and who have enough legal background to know what elements and categories are significant for GC's legal purposes. Subject indexing cannot be done adequately either by the typical lawyer or by an expert indexer operating in a legal vacuum.

Specifically, the program should be set up under the supervision of an experienced and imaginative legal indexer, solidly grounded in the capabilities of modern storage and retrieval methods. Preferably he should be a lawyer (FAA experience would be desirable, but not necessary). He should be assisted in the subject indexing function by two lawyers experienced in FAA matters. Organizationally, the unit should be located in the Office of General Counsel, rather than in the FAA Library, and the chief of the unit should be under the General Counsel or other official with the prestige and drive necessary to getting the job done. Because an operation of this kind cannot itself carry prestige, it should have prestige sponsorship and support.

The two legal indexers should not only have broad FAA experience, but be painstaking lawyers whose age or workload would minimize any disruptive effort on current GC matters that their transfer to this work might tend to have.

Although the original authors of legal opinions tend to be inexpert and inconsistent at indexing and to operate sporadically, it is still worthwhile to encourage them to index at the source, so far as circumstances permit. A document analysis worksheet would be helpful for this purpose (see Ap-

pendix B). Even if the indexing of an opinion is initially done by a specialist in indexing, the result should, so far as possible, be checked by the original author. This is helpful for quality control.

Once the program is well established, it should take only one legal indexer to maintain it. The other two lawyers will probably not be needed. If its operations become routine, solidly supported by GC lawyers, consideration should then be given to transferring it to the FAA Library.

## **P. Summary of Recommendations**

On the basis of this analysis the following recommendations are made:

(1) As a minimum, a central storage file of GC opinions, numbered and arranged serially, should be maintained in the Washington office and in each of the regional offices. Each opinion should recite the relevant facts, carry a descriptive heading, and show the date, author, addressee, document number, and approval for inclusion. The Washington files should contain the original opinions and should be located in the law division of the FAA library. Regional storage files should be in Xerox copy, or, if that is too expensive, on microfilm.

(2) As a minimum, a thesaurus-index of the GC opinions should be developed. It should be non-hierarchical and should index to a depth of about 6-8 subject terms per legal issue in addition to about 4-5 informational terms (e.g., date, author, addressee, company, etc.) per legal document. The thesaurus-index should include all secondary search terms likely to be used by searchers, cross referenced to the respective parent search terms. Each document reference should show both the document number and the descriptive document heading. If the thesaurus-index becomes unwieldy, it should be split into (A) a thesaurus, consisting of all search terms and cross references, but no document references, and (B) an index, consisting of all parent search terms and their document references, but no secondary search terms. These documents should be developed by key (or paper) punching, machine sorting, and print-out. The print-out should be photographically reduced and hard copy distributed to every GC lawyer, whether in Washington or in a regional office. The thesaurus and index should be updated and replacement copies distributed on a quarterly or semi-annual basis.

(3) As a minimum, an index to the GC legal opinions arranged by applicable section of the Federal Aviation Act or the FAR's should be developed and distributed on the same basis as, and as an adjunct to, the subject index.

(4) As a minimum, an index to the FAR's should be developed and distributed on the same basis as the subject index to the GC legal opinions.

(5) If experience shows that the descriptive document headings in the index are inadequate for preliminary sifting purposes, more informative digests of the opinions should be developed, arranged according to document number, and prepared and published similarly to the subject index to the GC legal opinions.

(6) Input materials for machine sorting should be formatted to permit machine recall on a coordinate basis by FAA's IBM 1401.

(7) The following should be undertaken only if the specific need develops:

(a) Development of an index to the Federal Aviation Act of 1958 and other statutes of interest to FAA.

(b) Development of an index to the pertinent executive orders.

(c) Development of an index to FAA's policy statements of interest to GC.

(d) Consolidation of two or more indexes.

(e) Searching by computer (as distinct from the sorting of items and printing out of indexes).

(8) An indexing unit, with prestige sponsorship and strong backing, should be established in GC for the purpose of carrying out these recommendations. To set up the program will require the services of an expert legal indexer with an understanding of computer capabilities and two lawyer-indexers, preferably with long FAA experience. After the program is well established, one of the latter should be retained to maintain the program.

## APPENDIX A

### GLOSSARY

**COORDINATE INDEX** means an index in which there is a separate index term for each separate concept (e.g., "airman," "certification," "eligibility"), whether it represents a main classification or a subclassification. The terms defined, being of equal dignity for search purposes, are arranged in a single alphabetical sequence.

**DEPTH OF INDEXING** refers to the number of index terms under which a particular document is indexed (e.g., for one GC legal opinion, the terms "lease," "landing area," "exclusive rights," "Anthony Municipal Airport," "Federal funds," "lighting system," "CAA §303," "General Counsel," "R. E. Elwell," "1/27/5/" represent indexing to a depth of 10 terms).

**DESCRIPTOR** means index term.

**HARD COPY** means a printed or typed record that can be read in the conventional way.

**HIERARCHICAL INDEX** means an index in which some index terms appear as subheadings under primary index terms without also appearing in their appropriate respective alphabetical locations in the primary alphabetical sequence.

**INPUT** means material in machine-readable form for introduction into a machine or computer.

**KEY WORD** means index term.

**LINEAR (NON-HIERARCHICAL) INDEX** means an index in which there are no index terms that appear as subheadings under primary index terms without also appearing in their appropriate respective alphabetical locations in the primary alphabetical sequence.

**PRINT-OUT** means the hard copy produced by a machine or computer.

**THESAURUS** means a list of words or phrases useful for searching the documents to which it refers.

APPENDIX B  
DOCUMENT ANALYSIS WORKSHEET

FEDERAL AVIATION AGENCY  
OGC DOCUMENT INDEX  
PUNCH CARD INPUT FORM

DOCUMENT NUMBER	DATE	AUTHOR	DESCRIPTIVE HEADING

SUBJECT MATTER INDEX TERMS (use terms in THESAURUS; if a new term is proposed, underline it):

CODE DESIG.

- PERTINENT SECTIONS OF FEDERAL AVIATION REGULATIONS:
- PERTINENT SECTIONS OF FEDERAL AVIATION ACT OF 1958:
- PERTINENT SECTIONS OF OTHER ACTS, EXECUTIVE ORDERS, OR REGULATIONS:

## APPENDIX C

### OTHER GOVERNMENT SYSTEMS

#### *(1) Department of Justice*

The Anti-Trust Division of the Department of Justice has adopted a manually indexed, machine processed legal research system called "LEX," designed to locate legal materials relevant to anti-trust law (an earlier version is described in 64M M.U.L.L. 41 (1964)). It is intended ultimately to cover (1) court decisions, (2) materials originating in the Anti-Trust Division, (3) legislative history, and (4) materials from legal periodicals. So far, only the second kind of materials and some of the fourth have been fully processed. Materials are stored on microfilm, with provision for needed print-outs.

Basic to the system is a thesaurus of search terms, with appropriate cross references to the parent search terms, which serve as index terms. The current thesaurus lists about 3500 search terms. Individual terms and their cross references are key punched and machine sorted by an IBM 1401. The hard copy thesaurus is prepared by photographing print out.

The principal index is called a "descriptor" index. It lists for each index term (e.g., "JEOPARDY, DOUBLE") the date of each pertinent document (e.g., "1-59"), its microfilm number (e.g., "AG 0382"), and a descriptive "document title" (e.g., "MEMO. IDENTITY OF OFFENSE & DOUBLE JEOPARDY UNDER SEC. 2 OF THE SHERMAN ACT"). It reflects indexing to an average depth of about 15-17 subject matter terms and about 8 additional search terms covering such matters as name of author, kind of company, and name of commodity.

The descriptor index, which is a comprehensive subject index, provides the searcher with not only the relevant document numbers that are of potential interest to him but, through the descriptive headings, means for culling out the more obviously inapplicable documents. The system also includes a screen of finer mesh for additional sifting at the lawyer's desk: document digests arranged by microfilm number of the stored documents and showing their respective dates. Both this list and the subject index are distributed in hard copy to each lawyer. Both are results of key punching and machine sorting.

The Division's "subject" index is actually a second subject index. It differs from the descriptor index in that it (1) includes only the less important anti-trust documents and (2) classified them only under very general subject classifications.

Also published and distributed is a "statute and rule" index. This is not, as its name seems to suggest, an index of the statute and regulations but a listing of documents according to the sections of the statutes and rules respectively involved.

The Department has also developed separate indexes by author, by company, and by commodity. These, however, have not been published and distributed within the Department.

Although the system, taken as a whole, seems to be unnecessarily elaborate for FAA, the following specific search tools could be useful to it: (1) thesaurus, (2) descriptor index (i.e., comprehensive subject index), (3) list of document abstracts (digests), and (4) list of documents by applicable

statute and rule. Indeed, items (1) and (2) constitute the essential elements of the minimum requirements system recommended above.

## *(2) Central Intelligence Agency*

The Office of General Counsel, C.I.A., through its Legislative Counsel, has developed and is perfecting a system for storing and retrieving the vast body of C.I.A. legal memoranda, legal opinions, and other documents involved in any of its specific problems or situations with legal aspects (each is called a "case"). Present estimates place the total number of documents in the system at 97,000, with an average of 2.4 pages a document.

The larger unit of storage is the "folder," which includes all documents and papers relating to a particular situation or problem. The smaller unit of storage is the document. Because all documents relating to the same case are located in the same folder, a document can be located either through its document number or through the number of the folder of which it is a part. Thus the system uses two series of accession numbers, or storage addresses. Folders are arranged according to serial number rather than subject matter.

Access to the pertinent file or document is through a printed list of search terms, commonly known as a "thesaurus," which gives the searcher the relevant terms that he can expect to find in the pertinent indexes. Of these, C.I.A. has developed four:

(1) The "subject" index is a hierarchical index about 1605 pages long that lists under each index term the author ("originator"), date, folder number, document number, and other information not relevant here. If the particular entry involves a "key" (important) document, it is designated according to the kind of source (e.g., "A JUST," where "A" means an agreement and "JUST" means the Department of Justice).

(2) The "proper name" index is an index about 1068 pages long that shows the same information for each document as the subject index but groups it alphabetically according to the names of the significant persons involved (e.g., "Dulles, Allan") or of special projects (e.g., "Q. R. Rabbit"). This is helpful to searchers interested in accumulating the available information on a particular person or project.

(3) The "folder-document number" index is an index about 1605 pages long that lists each folder and each document serially by assigned number, showing for each unit its date and a list of the index terms assigned to it in the subject index. As a list of folder and document digests, it is useful for a preliminary screening of the folder or document entries that an examination of the subject index makes potentially interesting.

(4) The "key document" index is a list about 534 pages long that shows for key documents the same information that appears in the "folder-document number" index but arranges it according to kind of document (e.g., "A," an agreement) and source (e.g., "JUST," the Department of Justice). Thus, a searcher whose only lead is the fact that what he is looking for is an agreement originating in the Department of Justice will enter the system through this index by scanning the entries marked "A JUST."

The present system prints and distributes a hard-copy thesaurus and four hard-copy indexes for each lawyer. Updating will probably be on a monthly basis. The materials are key punched and machine sorted by an IBM 1401, the print-out serving as the basis for hard copy. However, no computer is used for searching.

Although it parallels the minimum requirements system recommended above, this system appears to be overelaborate for FAA purposes. A separate key document index would be unnecessary for FAA, because all GC legal opinions are "key documents" in the sense used by C.I.A.

On the other hand, its "proper name" index represents the counterpart of the proposed index that would list documentation by sections of the Federal Aviation Act or the FAR's.

The most serious objection to the C.I.A. system is its use of an hierarchical arrangement of terms in the subject index. For the reasons given above, a non-hierarchical arrangement of index terms would seem to be preferable for FAA. Otherwise, the system provides many parallels to the minimum requirements system recommended in this report.

### *(3) Department of the Air Force*

Through a contract with John F. Harty of the University of Pittsburgh, the Department of the Air Force is now testing and refining a system of electronic storage and retrieval of legal materials called "LITE." Operating at the Air Force Finance Center, Denver, it has stored on electronic tape or disk complete text of all of the United States Code and a large number of opinions of the Comptroller General, a total of 18 million words. IBM, which is serving as sub-contractor, is providing the use of both a 1410 and a 1310 random access computer. Testing, which is being done at every organizational level, has consisted so far of about 150 searches based on actual legal problems. Evaluation includes the documentation of machine costs. Heretofore, the Harty system has been used primarily to search the statutes of Pennsylvania and several other states.

If successful (the officials in charge appear optimistic), the system will centralize research and Defense communication. Because all machine searches will be done at one location, lawyers located at other installations will send their search requests to the legal research center.

The system will attempt to blanket all materials of interest to the Air Force lawyer. In addition to the United States Code and the Comptroller General Opinions, it will ultimately include relevant uncodified statutes, selected court cases, JAG opinions, General Counsel opinions, other interpretations, and internal directives. The aggregate of stored materials is expected to total 100 million words.

In general, the Harty system builds a word-for-word index from the natural language of the documents stored. Full text is first stored in the computer. From this, the machine builds a thesaurus that includes every word with any possible search value (this excludes such words as "the," "for," "from," "unless," etc.), providing practically 100 percent depth. The thesaurus is then refined to correct misspellings, consolidate word variants (e.g., "fly," "flying," "flew," etc.), and related synonyms. This is necessary to make it more efficient and easier to use. Depending on the specific request, the machine provides the searcher with the total number of



responsive documents, their citations, digest lists of all applicable index terms, or full text.

This system gives the searcher the widest and most flexible possible access to original text, because there is almost no advance selection by the indexer. The disadvantage is that it imposes a correspondingly higher burden on the searcher to frame his question in the most significant categories. For this reason, it appears to be unwieldy for routine searching, especially for lawyers located at other installations.

Such a system appears to be unnecessarily elaborate and expensive for GC's current needs. Also, it is not compatible with the recommended minimum requirements system, because, if the need for a machine-generated natural-language index later appears, it will be necessary to start again from scratch, beginning with the punching of original text on machine readable cards or tape.

At present, it seems unlikely that so elaborate and expensive a system could be justified for storing and searching GC's legal opinions. The same seems to be true for the FAR's and other materials of interest to GC lawyers.

#### *(4) Internal Revenue Service*

The Office of Chief Counsel, Internal Revenue Service, is developing a storage and retrieval system for (1) its abstracts of pending cases, and (2) the documents relating to closed cases. Because these two categories involve significantly different kinds of problems, they have been dealt with differently. Of the roughly 150,000 documents represented, about 25,000 (averaging about 2 pages) relate to pending cases and about 125,000 (averaging 4-5 pages) relate to closed cases. The latter fund is expected to grow steadily.

For pending cases, the main search tool is an index that has some of the characteristics of an annotation of the Internal Revenue Code and some of the characteristics of a hierarchical subject index. The index has a three-level breakdown. The first-level breakdown is by sections of the Internal Revenue Code (e.g., "§6653. Failure to Pay Tax"). The second-level breakdown is by "tax problem concept," that is, by subject matter (e.g., "Penalty for Fraud"). The third-level breakdown is made up of subordinate classifications of the subject matter breakdown (e.g., "Net Worth Material," "Intent to Reduce Taxes," "Doctor," "Failure to Report Alleged Income").

The use of a hierarchical index is believed by its authors to be justified by the needs of lawyers interested in pending cases. Because the great bulk of searches are code-oriented (99.5% cases are tried under the I.R.C.) and in terms of the selected categories, the typical searcher benefits from a hierarchical arrangement oriented by his needs. Only a few lawyers search sufficiently from other points of view to be seriously disadvantaged. Even so, the disadvantage has been serious enough to prompt the development of a KWIC index (non-hierarchical) covering the key words in the titles used in the master index.

Although this kind of indexing appears to be adequate for the specialized and largely uniform needs of search pending cases, its authors consider it inadequate for its second category of document: those relating to closed cases, which are consulted for their precedent value. It is this category that provides the closest parallel to GC opinions in FAA.

Because the indexing used for the first category is considered too shallow for the second, I.R.S. proposes to index the latter to a depth of from 15 to 20 terms. This has already produced about 5,000 search terms and is expected ultimately to produce a much larger total. Indexing to this depth is expected to fragment the relevant concepts to the point where searching in combinations of terms (to an anticipated maximum of 5) is likely. The system, therefore, must coordinate to at least 5 terms.

Consequently, it is expected that some form of inverted grouping system will be necessary to facilitate the matching of document entries. This will either be descriptor cards using optical coincidence ("peek-a-boo") or some variant of the Uniterm card. The former will provide immediate matching but will be difficult to update. The latter will require matching by inspection, but will be easy to update because it involves only the creation of replacement descriptor cards.

Under both systems, storage is on microfilm (in the first category only case abstracts are involved). Accordingly, document numbers refer to microfilm records.

Because of the desire to provide for a decentralizing searching operation and to avoid expensive hardware, it is planned to provide each regional office with a set of the stored materials in microfilm and a set of the applicable search tools. For the 250 attorneys in the headquarters office it is planned to maintain 10 sets of the materials to avoid traffic jams at the search facilities.

I.R.S.'s approach to storing and retrieving agency precedent appears to be over-indexed for FAA purposes, where an average depth of 8 to 10 subject terms per document would seem to be adequate. If so, the use of optical coincidence or Uniterm cards seems less likely to be appropriate.

Although I.R.S. plans to use computers largely for generating search tools rather than for searching, it will do inventory searching on pending cases by computer and ultimately machine searching of closed cases. When desirable, this can be done readily because it is using a formatted system that will permit machine recall on a coordinate basis. Its present system is the one developed for Navy Intelligence. Although written for an IBM 1410, it can easily be written for FAA's 1401. Fortunately, no extra cost is involved in using a format of this kind.

#### *(5) Federal Trade Commission*

The Federal Trade Commission is engaged in indexing (1) its published decisions (1916-1962, 60 volumes), (2) the relevant decisions of the Supreme Court and Circuit Courts of Appeals (1918-1962, 7 volumes), and (3) the relevant non-case materials: law reviews, legislative history, reporting services, Law Week, Federal Second, treatises, monographs, and even the names of useful experts.

In general, the system follows the "point of law" approach, developed by Professor Robert T. Morgan at Oklahoma State University. A case is classified by one or more short, descriptive sentences (called "concepts"), each of which begins with a key term for the purpose of alphabetizing. It is also classified by commodity and by section of the applicable statute. For identification, each concept or commodity is assigned a number.

The resulting materials are stored on cards according to the inverted system of arrangement (i.e., a card represents an idea rather than a docu-

ment). Because there is a separate thesaurus and a separate index, there are two basic card files. In the thesaurus card file, each card shows a concept or commodity plus its assigned number, as in the following entry:

DOMESTIC PRODUCT COMM HOLDS A PRODUCT WITH 135  
SOME FOREIGN INGREDIENTS TO BE A, IF FINAL  
FABRICATION OR BLENDING IS IN U S

(concept)

(concept number)

In the index card file, each card shows the number assigned to the concept or commodity that it represents plus the citations to the cases, documents, and other items of information that relate to that concept or commodity, as in the following entry:

120 5FM DISCLOSURE CONTENTS 54F1026 54F1292 06S533  
120T01 59 MICH821 SEE JIM FERGUSON

(concept number)

(relevant indexed materials)

The thesaurus card file is used to create hard-copy printouts called "code books," which are in the form of lists of concepts and commodities and their respective identification numbers. The index card file, on the other hand, is not used to create a similar hard-copy index, and therefore must be searched by machine or computer. In each instance the searcher uses the appropriate thesaurus code book manually to find the relevant concept or commodity identification numbers. Armed with these, he then conducts a machine search in which he asks for the citations corresponding to those numbers. A 407 accounting machine is used both for sorting index cards and for printout. (These operations could be performed with an IBM 1401 or other computer of similar capabilities.)

Instead of using a single printed master thesaurus code book, the system uses four. Concepts and commodity entries are distributed among these according to general kinds of cases indexed and the corresponding source laws to which they pertain, as follows:

I Deceptive Practices	Sections 5 and 12 of the FTC Act
II Restraint of Trade	Sections 2 and 3 of the Clayton Act
III Mergers	Section 7 of the Clayton Act
IV Fur and Fabrics	Miscellaneous acts

Within each thesaurus code book, concept entries are grouped according to specific section or part of the source law involved and, within each such category, listed alphabetically. As a result, the total classification is highly hierarchical. The limitations inherent in such a system are lessened somewhat by the use of cross references from other helpful search terms. No use is made of the matching capability available in the use of dedicated spaces, presumably because relatively shallow indexing does not call for searching in combination of terms. Each code book includes its own commodity list.

However well adapted this specific approach may be to the needs of the Federal Trade Commission, it does not appear to be sufficiently flexible for FAA's needs, which call for deeper indexing and indexing not so subject to the limitations of hierarchical arrangement.