

**U.S. DEPARTMENT OF COMMERCE
National Technical Information Service**

AD-A009 237

**A COMPARATIVE EVALUATION OF MACHINE
AIDED INDEXING AT THE DEFENSE DOCU-
MENTATION CENTER VS THE MANUAL IN-
DEXING DONE BY THE NSWC/DAHLGREN
LABORATORY TECHNICAL LIBRARY**

C. C. Lyon

**Naval Surface Weapons Center
Dahlgren Laboratory, Virginia**

March 1975

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR-3317	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD-A009 237
4. TITLE (and Subtitle) A COMPARATIVE EVALUATION OF MACHINE AIDED INDEXING AT THE DEFENSE DOCUMENTATION CENTER VS THE MANUAL INDEXING DONE BY THE NSWC/DAHLGREN LABORATORY TECHNICAL LIBRARY		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Cathryn C. Lyon.		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Surface Weapons Center Dahlgren Laboratory Dahlgren, Virginia 22448		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE March 1975
		13. NUMBER OF PAGES 87
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release; Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
REPRODUCED BY NATIONAL TECHNICAL INFORMATION SERVICE U. S. DEPARTMENT OF COMMERCE SPRINGFIELD, VA. 22161		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A study was made of the DDC Machine indexing system (MAI) using 551,1498s (WIS) published from 1973/74. The 1498s are all on electronic sensors. The author indexed the 551 documents manually using the Uniterm System of Coordinate Indexing. Technical Information Specialist assisted with the indexing and final acceptance of answers. Questions submitted by users at NSWC/DL were run by DDC and the resulting print-outs were compared with the retrieved material from the author's files.		

A COMPARATIVE EVALUATION OF MACHINE AIDED
INDEXING AT THE DEFENSE DOCUMENTATION CENTER
VS THE MANUAL INDEXING DONE BY
THE NSWC/DAHLGREN LABORATORY TECHNICAL LIBRARY

Cathryn C. Lyon
Head, Technical Information Division
Naval Surface Weapons Center/Dahlgren Laboratory

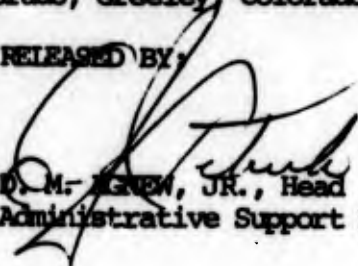


Approved for Public Release; Distribution Unlimited

FOREWORD

This paper is an attempt to evaluate the machine aided indexing developed by Paul Klingbiel and his staff at the Defense Documentation Center vs the manual indexing which is applied to the Computer Storage and Retrieval Program in the Technical Library at NSWC/Dahlgren Laboratory. It is the result of Independent Study for four hours in the Masters in Public Administration Program sponsored by the University of Northern Colorado, Greeley, Colorado.

RELEASED BY:


D. M. GREEN, JR., Head
Administrative Support Department

ABSTRACT

A study was made of the DDC Machine indexing system (MAI) using 551,1498s (WIS) published from 1973/74. The 1498s are all on electronic sensors. The author indexed the 551 documents manually using the Uniterm System of Coordinate Indexing. Technical Information Specialist assisted with the indexing and final acceptance of answers. Questions submitted by users at NSWC/DL were run by DDC and the resulting print-outs were compared with the retrieved material from the author's files.

TABLE OF CONTENTS

Chapter		PAGE
I	STUDY BACKGROUND	1
II	EVALUATION METHODS FOR COMPARING INDEXING TECHNIQUES	6
III	DESCRIPTION OF THE DDC MACHINE INDEXING PROGRAM (MAI)	14
IV	UNITERM METHOD FOR INDEXING 1498's	26
V	CONCLUSIONS	31

APPENDIX A - Figures 1 through 4

APPENDIX B - Questions and Results

APPENDIX C - Distribution

CHAPTER I STUDY BACKGROUND

These four hours of independent study in the Masters in Public Administration are devoted to measuring the retrieval effectiveness of the machine indexing system of the Defense Documentation Center versus indexing, to be accomplished by the author and her staff.

The Defense Documentation Center is responsible for collection and dissemination of reports pertaining to the Research and Development contracted for the Department of Defense.

The DDC document retrieval system has been designed and implemented to satisfy the needs of users of DOD information. In its early days, DDC was concerned with the acquisition and bibliographic identification of documents required by its users. In the middle fifties a UNIVAC computer was installed and efforts were concentrated on storing for retrieval not only by bibliographic addresses but by subject approach to the collection. Those of us who used their services in the 1950's can recall frustration when the new computer could not respond because it was down for repair. This was largely due to the newness of the animal.

Since the 50's, DDC has been a leader in the processing, i.e. indexing, storing and retrieving of information, using sophisticated computers. The data bank with which I am concerned is called the 1498 or Work Unit Information System (WUIS). The 1498s a one page document, (Figure 1) reflect the on-going R&D of DOD that has not reached formal reporting stages.

During the 60's, Paul Klingbiel, in the Development Directorate of DDC, began work on machine indexing of documents using the 1498s for his investigation. It is the author's purpose to evaluate Mr. Klingbiel's machine indexing of the 1498s by indexing manually a sample group of 551 documents from this data bank. Technical Information Specialists in my Technical Information Division will assist with the indexing in order to be sure I am obtaining the correct depth of terminology. I hope that the evaluation will prove the effectiveness of machine indexing because manual indexing is costly and time consuming.

The references at the end of this report reflect many authors' discussion of the "right way" to evaluate a retrieval system. There is no single agreement among authors as to this "right way" unless it happens

to be a co-author of the evaluation. The discussion of evaluation methods will come in Chapter II of this report.

Intellectual indexing of text occurs in three steps: selection of the concepts in the document; fitting the concepts to the thesaurus of generic descriptors; and posting the descriptors to the storage medium.

Indexing involves both explicit and implicit concepts. The broad, general category of a document may not be mentioned in the text but the indexer's knowledge leads him to use the general concept in addition to those explicitly laid out.

A controlled Thesaurus, whether of single descriptors or hierarchical associations, is essential to any indexing. The manual intellectual selection of a descriptor is weighted against the thesaurus for usage. Usually, in our established system at Dahlgren Laboratory, if the concept has not been recorded before, there will be a period of consultation before the descriptor is then made a part of the thesaurus. It will be noted that the DDC Machine Indexing Program has the capability built in to check against existing descriptors and when there is none, the MAI will reject the descriptor

to a part in the program from which it is considered by humans for inclusion in the storage thesaurus.

Human analysis and indexing has the advantage of making decisions when the unusual or new concept appears. If one has indexed documents it is worth checking yourself for consistency in six months. I have found that the first 6-10 explicit descriptors will still be the same, but additional concepts will be raised by the experiences of literature searching over the 6 months. I think this is the problem that can not be effectively addressed by the machine analysis, at least not until the new concepts have entered the literature as accepted terminology. However, human indexers are inconsistent by virtue of being intellectual individuals and there will be a certain arbitrariness in assigning descriptors. Machine indexing has the advantage of being consistent as long as the author has intended the same concept as that established by the automatic thesaurus.

Much research has been done on the storage of complete text by optics or other methods. However, the costs of input and output at the present time makes it uneconomical to store the complete document. In addition, scanning full length documents is impractical because of the processing

time required. Automatic indexing may lose those concepts implied by the text and the concepts may only be capable of pick-up by manual indexing.

The computer does not criticize grammar, therefore descriptors can ignore such things as hyphenation, singles posted to plural terms, spelling, or grammatical usage. While a request for a search usually does not give it exactly as in-putted to the computer, the descriptors used for retrieval must identify the problem clearly enough that correct answers will result. Whether to use a single word or a group of words depends upon the loss of relevance if such as "air-to-air" or "surface-to-surface" as a concept is used. Single terms in this case would create unnecessary coordinating.

As the chapter on the DDC program indicates, the identification of significant descriptors should reduce the number of entries in the thesaurus. However, the information which may later become important should be made available when its time is ripe. For this reason the program does automatically scan for new terminology in the field of military research. As I see it, the work at DDC is toward creating and maintaining a self-adaptive data base of information.

CHAPTER II EVALUATION METHODS FOR COMPARING INDEXING TECHNIQUES

An Information Retrieval System implies the complete organization for acquisition, indexing, storing and retrieving information required by a group of users. There is expected to be a subject index to the items in the store, also a method for recording and retrieving the identifying bibliographic items. The arrangements in the index or thesaurus will depend upon its format, i.e. hierarchal arrangements or completely single descriptors.

A good system should be developed from the point of the user if that is the reason the organization has established the data bank¹. One must be careful that it does not wholly reflect the point of view of the manager, the designer, or other than the user. Mr. Klingbiel and Mr. Charles Jacobs have developed the program for Machine Aided Indexing with the idea that the users' needs are the important demand placed upon their work. They must work within a time frame and economical base with service to a large population of users in every imaginable discipline in the world.

¹Data bank and document collection are interchangeable in this report.

It is recognized that as much similarity of technology terms as possible must be established between the MAI population of users and those being used by the author. This has been established by my questions because scientists and engineers who have submitted questions for my use in the study are a micro of the same DOD population served by DDC. Thus, we have as similar as is practical, the same limitations or advantages enjoyed by DDC and the smaller group of users at Dahlgren Laboratory.

The particular population approached at Dahlgren consists of the researchers of electronic sensors. The users in the normal course of events require literature searches of electronic sensor 1498s so that the questions submitted by me from them to DDC are typical of current requirements.

The vocabulary of the indexing language I use should be a complete collection of sought terms in the Natural Language Data Base for the MAI system. Likewise, the indexing of the request of 1498s should be as complete as experience with the "micro" users has indicated necessary. The DDC researchers have set up a system that does not overlook new descriptors being received, "add-ons" to the Index terms are a part

of every in-put to the computer. The "strings" of words developed by DLC as the basis for the computer Index hopefully, will not limit the kinds of requests that the 1498s will be subjected to. The Uniterm System of Coordinate Indexing which I have used can be seen to allow greater use of "colloquial" or "one of a kind" descriptors without adding too much cost if they are used discreetly.

Any retrieval system should have as its primary function that of saving its users to as great an extent as possible, from perusing and discarding irrelevant documents in their search for relevant ones.

William S. Cooper² lists the following to describe the user-requester operations:

1. He may want only a specific answer, for which one document will suffice.
2. He may not have time to read all the literature but it would be useful to randomly select a sample of the literature; this may cover the whole area but he may only want a few reports.

²Cooper, William S., American Documentation, January 1968.

3. A user may want to make a thorough search to see all relevant documents. If only one is available that may be all.

4. A user may want to sample an area and just specify the ideal size like 10% of retrieval.

Let me define two terms which will be used for judging the output answers which result from the questions put to DDC's MAI and the manual UNITERM SYSTEM which I have used for this study.

The words are:

1. Precision - this has to do with the relevance ratio, accuracy, distillation and acceptance rate of answers.

2. Recall - has to do with the proportion of relevant documents retrieved.

High recall means nothing if the requester does not want more than a small fraction of the relevant documents. In some cases, the elimination of excess retrieval can be programmed in the request. For instance, a question can ask for documents relevant to "high temperature mica capacitors" and want only those of a particular date or application.

In indexing the 1498s , it occurred to me that ranking these particular documents by hierarchy may be too limiting for the kinds of request that 1498s are subjected to.

The primary function of a retrieval system is conceived to be that of saving its users to as great an extent as possible, the labor of reading and discarding irrelevant documents in their search for relevant ones as will be evident in the comparison that will follow - overall performance of a system can be affected by changing a single variable such as the coordination level.

It would be rash to suggest that the results which I obtain will necessarily be valid for every other subject field. The correct approach of course, is to find what system lends itself best to information retrieval under a given set of circumstances. The audience of users at Dahlgren Laboratory constitute a representative group of DDC's users. By use of the Uniterm System, I feel this investigation of the hierarchy strings of DDC in comparison with the Uniterm looser indexing will show whether or not specificity is served in both systems.

At this point, I need to make myself "clear" about the point-of-view taken for evaluating results. For instance, there could be several points of view such as the systems' users, the manager of the program, the designer of the program or the indexer. In the case of this evaluation - the decision of relevancy will come from the users who were kind enough to originate the questions for me.

Each system will be checked for:

1. References which are the answer to the question (precision)
2. References which were recalled but lacked relevance

As in all computer programs, there is a problem of noise. Marie Konigova³ defines noise in the following manner:

1. First order noise occurs when the system doesn't deliver a relevant document. This indicates in most cases an indexing error.

This will generally happen if the indexer mistakenly omits one of the appropriate descriptors. From the users point of view this kind of noise

³Konigova, Marie, Information Storage and Retrieval, V.6, #6, pp. 437-444, May 1971.

represents a more bothersome error than second order noise.

2. Second order noise occurs when the system delivers a document which is not really relevant. It indicates how relevance is caused by an error in the selection language.

William Cooper writing On Selecting a Measure of Retrieval Effectiveness

says (and my reading supports) there is no substantial agreement in the field as to what good retrieval performance is or how to measure it. He goes on to say that it is too much to hope that a measure of this ideal nature could ever be described in such a way as to appear totally convincing to everyone - one obvious difficulty being the problem of describing an idealized evaluation procedure in acceptable operational term when this procedure may in fact be operationally unpracticable. It is based on the assumption that an ideal evaluation methodology must somehow measure the ultimate worth of a retrieval system to its users in terms of an appropriate utility. One procedure in report writing will obviously have to improve with the advent of Machine Aided Indexing - that is the development of the title. With the use of KWIK Indexes and operations such as MAI using key

words from the title, authors should be concise but as inclusive as possible of the concepts within his report. Improvement in this area may be just the variable that will make machine indexing economical and specific enough to eclipse all manual indexing.

I will take a rather simplistic approach for establishing ratios. The 40 questions will be judged by an adaptation of Cyril Cleverdon's plan in the Cranfield Tests. In addition to judging each report for relevancy according to the two outlined already referenced in this chapter, I'll apply the 40 questions to each of the two indexing systems and use the following method for establishing recall and precision ratios.

	Relevant	Non-Relevant
Retrieved	a	b
Not Retrieved	c	
	a + c	

Recall Ratio $\frac{a}{a + c}$

Precision Ratio $\frac{a}{a + b}$

I must remind the reader that I am not evaluating cost effectiveness of either system.

CHAPTER III DESCRIPTION OF THE DDC MACHINE AIDED INDEXING PROGRAM (MAI)

In a report published in June 1969 called "Machine-Aided Indexing", the author, Paul Klingbiel, touches on two approaches to automatic indexing which he considered while developing the MAI program. He says that there is (1) statistical analysis, or (2) syntactic analysis. The statistical technique requires fairly extensive stretches of text. DDC is constrained to work with title and abstracts. For the 1498 collection this usually means less than 150 words per document. In addition, Klingbiel says, any machine indexing technique must compete with manual indexing cost to be of serious interest to DDC. Mr. Klingbiel has developed a system that does not depend upon statistical analysis. He uses linguistic techniques but he feels complete syntactic analysis of sentences by the computer are not required.

In the DDC MAI program, individual words are read into a computer and are either held for further consideration or eliminated from further processing. The output is a list of candidate index terms and a screened exception list of terms and phrases for human review. Eventually the list of candidate

terms will be entered into the Natural Language Data Base that will have the capability of posting terms directly to the data base, switching synonyms to postable terms, or out-putting recognized terms for technical consideration.

The computer programs for the MAI System were written primarily in SLWETH/I, the assembly language for the UNIVAC 1107. Eventually all related programs will be written for the UNIVAC 1108 running under EXEC8.

Manual indexing for storage and retrieval required a thesaurus. However, because of new developments in computer technology, particularly developments in time-sharing and remote access devices, emphasis is being placed on automated solutions for indexing, storage and retrieval for large collections of documents.

Basically, the language problem asks how to index and store information so that it can be efficiently retrieved. DDC serves a very large audience and must be able to retrieve with fairly accurate specificity, any documents that pertain to a users request for information. For instance, a term such as "electronic sensors" is far too broad for the narrow boundaries of research on laser detection. However, economics dictates that DDC can

not create a language base of such narrow specificity that the computer is filled up with too many descriptors. The MAI program, as it will be described, is developed to have as narrow specificity as possible while maintaining control of the vocabulary in its hierarchical growth. I am using a Modified Uniterm system (Figure 2) to manually index and post the sample 1498 documents. The sample 1498s consist of 551 documents pulled from 1973/74 inputs on electronic sensors. The manual indexing will be described in Chapter IV. The Uniterm System allows for some what more specificity and this will be the crux of the evaluation. In Technical Libraries such as at Dahlgren Laboratory, the user has two general types of requests. First, when starting in a new area or attempting to bring a current research project up-to-date, those concerned will ask for a state-of-the-art survey. The other need is usually quite specific to a hardware or component problem and to answer this type of request, depth of indexing is a requirement.

The MAI System uses natural language which then requires a mechanism to correlate singulars, plurals, synonyms and related concepts.

The Natural Language Data Base is required for two things: It is a final screen for index terms selected by the MAI indexing and as a translation for English language retrieval. The in-puts to the data base are of two forms. The first is either a single word or a phrase. The other major type of in-put is the lead-in or reference term. This can be either of two forms:

A use B

or

A use B and C

Where "A" is the lead-in term and "B" and "C" are both posting points, an incoming index term is matched against the file and takes one of these paths.

1. If the term matches a posting point, the accession number associated with the index term is posted (added) to the list under the matched term.
2. If the term matches a reference item, the accession number is posted to the posting point(s) referred to by the entry, i.e. term B (and C) in the example above.

If a request for information uses terms that are not posted in this manner there would, of course, be no output for the researcher.

As indicated before, users ask by general topics as well as very specific items or concepts. Therefore, Mr. Klingbiel had to consider hierarchal concepts in order to relate items most often in their generic sense.

For instance, (1) light cruisers, (2) cruisers, (3) destroyers and (4) naval ships. Number (1) is a sub-set of number (2) which is a sub-set of number (4). Number (3) is another sub-set of number (4). Because the transactions for this program need a free-form input, DDC decided to use the Text Processing System (TPS), a remote keyboard-to-computer system. This system enables one to enter a flexible record in free form. The programs that create the data base include not only the actual formatting programs but sorts programs to print the data base in various formats and a program to load the data base on mass storage as a part of a data management system, FMS-8.

Each main term is assigned a unique identifying number and all references are tied together to main terms by that code. The creation program

recognized four types of input records:

1. Main terms
2. Synonyms
3. Multiple-posting terms
4. Narrower terms

Taking Figure 1 as a sample, it is pointed out that the MAI uses those items indicated for automatic indexing. There is a program for the indexing staff to review the key words, etc. furnished by the author of the document.

Mr. Klingbiel describes his terminology for syntactic types. He uses six syntactic types which are parts of speech that constitute index terms or elements of index terms. The six possible components of index are:

- (1) N - class of nouns each of whose members is acceptable as a free form
- (2) A - class of adjectives that can function only in the role of modifier
- (3) Z - class of nouns each of whose members is acceptable only in combination with an N or an A or another Z, such as NZ, ZN, AZ or ZZ and of course strings of three or more such as AXN
- (4) + - the word "and"
- (5) P - the word "of"
- (6) C - the word "or"

The fact that "temperature" tends to occur with high frequency and tends to be nondiscriminating in isolation suggests that "temperature" be considered as a "Z". The recognition that "temperature" can function adjectively does not preclude assigning a "Z" to the term, quite the reverse, the fact that "temperature" is desired only in combination strengthens the argument. From an indexing standpoint, "low temperature alloys" is as logically represented by the syntactic formula AZZ as by AAZ. Moreover, "body temperature" requires either an NZ or a .ZZ combination since "temperature" does occur in a noun head position.

The approach the DDC people have taken is as follows

1. Only a subset of English is pertinent to indexing. For instance, verbs are never used as index points. Adverbs ending in -ly or -lly are rarely used. In the rare case the word can be designated as an adjective.

2. Nouns and adjectives can be distinguished as follows: An adjective is a word which never appears in isolation (or as a free form) as an index term. An adjective is always in a modifying, never a head position. This condition can be considered completely unambiguous since in natural scientific English the modifier precedes the head noun rather than following it.

Nouns can appear either as heads of structures or modifications or modifiers. Plurals which are a standard form of index terms, always appear either in isolation or in a head position. Whether such nouns are typed as N or as Z is a decision based largely on the noun's utility as a discriminating element in the data base for which the system is built. Singular nouns may be categorized as "Z" the decision is based on a study of occurrence through some such means as a permuted list.

A basic contingency factor is incorporated in the part-of-speech mechanism of stand-alone nouns vs. nouns requiring modification. However, there are three screens before they are accepted:

1. Words read in are either accepted for further analysis or rejected.
2. Words passed on for analysis are stored and a syntactic formula is built up until the indexing process is halted by either a word reject, a conditional word such as "and", "or", "of" or by punctuation.
3. The Natural Language Data Base is the final screen before posting.

A match with a plural stand-alone noun is passed for posting; singulars of the same noun (these must be N's not Z's) are detected and posted on the plural form. .. wide range of "use" references not involving plurals are

also detected and posted on the preferred term. Long sequences of words of appropriate syntactic type will probably not match and will be displayed for technical review.

Here are some samples of format matches:

AA + AN Biological, Chemical and Physical Oceanography

AZ - Biomechanical forces

A&AZ - Ballistic and Orbital Vehicles

ZZZ - Air breathing engines

The main source of input to the Natural Language Data Base is the terminology of the users as expressed in incoming DDG 1498s (WUIS). A word-by-word comparison was made of the existing vocabularies by Mr. Klingbiel. Each term was broken into its component words, and a computer file of unique words was developed for each vocabulary. This DDC list contains 14,325 unique words. The formal vocabulary and list were compared with the list of 21,463 unique words contained in a corpus of one million running words of text used in the MAI experiment. 7,727 words occurred in both lists. Other matches were made which led to insights into the relationship of formal vocabularies to natural language as it

occurs in scientific and technical literature. The major effort in the NLDB creation has been in the area of natural language synonyms. One of the uses of the data base is as a final screen for terms generated by the MAI programs. This screen works a filter for syntactically correct but semantically worthless terms such as "two areas" and as a normalizer.

The MAI system now operates with nine word classes;

1. two adjective classes, A and B
2. three noun classes, N, Z, and R
3. four classes consisting of one word each T, P, X and Y which indicate respectively, "and", "of", "or", and "other"

In The Recognition Dictionary a single subroutine (MACRO) is associated with each English word held in the dictionary. Such a unique assignment within a context-sensitive environment is possible only within a restricted subset of English. The fact that the indexable subset of English does not contain verbs, the acceptance of noun-noun combinations and the careful assignment of words which could act as either adjectives or nouns have made unique part-of-speech assignment a practical approach.

The next step in the DDC program is the temporary storage. As words are read they are matched against the Recognition Dictionary. All

words, with some exceptions are copied into the temporary storage area until the indexing cycle is halted. When the indexing cycle is halted, the words held in temporary storage are either printed out as candidate index terms, printed out on an exception list for human analysis or destroyed. As a word is read and recognized, its syntactic type is noted and placed in the format register, a record temporary storage area. Syntactic symbols in the format register are matched against a dictionary of permissible syntactic formats whenever the cycle is halted.

The format dictionary contains 77 permissible syntactic formats ranging from a single N (a noun judged to be meaningful in isolation) to a string such as ZZZZ (a string of four contiguous nouns, none of which is meaningful enough to stand alone, i.e. San Bernardino counties planning).

At the indexing cycle, assume that all temporary storage locations are empty and that the recognition and format dictionaries are available. The first English word is read in and matched against the Recognition Dictionary. If the word is unrecognized, it is printed out for human review, and the second word is read in. If the word is recognized, its

MACRO⁴ will move it to temporary storage and place its syntactic type in the format register provided it is not one of several words which may not appear initially. When the cycle is interrupted, the syntactic format held in the format register is matched against the permissible formats of the format dictionary. A match is considered a candidate index term; a mismatch results in a readout of the string for manual review.

⁴ MACRO refers to a list of Housekeeping and Index Selecting steps built into the program.

CHAPTER IV UNITERM METHOD FOR INDEXING 1498s

I chose The Uniterm System of Coordinate Indexing as the medium for indexing the 1498s. It was developed as a manual documentation method. I have modified it to suit purposes in a one shot application such as this study. The Uniterm System was developed by the late Dr. Mortimer Taube in his research for a method to organize military documents for storage and retrieval during World War II. The Uniterm approach has been used by Dr. Cyril W. Cleverdon in the Cranfield Tests he conducted during the 60's in his work for ASLIB in Great Britain. See Figure 2 for the kind of record used and the manner for coordinating terms. In the early days of the system, individual words were used almost exclusively. As experience with the system grew, it was recognized that "bound terms" made sense. Involved is the analysis of the contents of documents in terms of the key words that appear in the title, abstract and body of the report. These key words or descriptors include not only single English words but abbreviations (defined) see Figure 3 . Identifiers are, for instance, "Mark 48" the nomenclature assigned by the services for a specific torpedo;

numbers are used; in the case of this work plurals are used and it is recognized that the system does not have to recognize grammatical usage for storage. In this study some arbitrary assignments have been made i.e. recording visual and visible on the same card for sake of economy of cards. The 1498s are all concerned with "electronic sensors" so there is no card for them because a question can assume that any one of the cards will reveal a relationship to "electronic sensors". It will be noted that the cards for guidance (control) and navigation are very full. The use of these terms approach the same general size since electronic sensors in the military have their greatest application in missile guidance and navigation.

I applied this system of indexing in the 1950's to 6000 items of manufacturer's technical information and later to a report collection in a first effort at manual documentation. It is preferable that the indexer have some familiarity with the subject involved, but the fact that he or she did not have to use absolute relationships such as knowing that "electron spin-lattice interactions" should include indexing up to "electron paramagnetic resonance" one could use a bright person with a

general liberal arts background. We were able to organize the literature required by an electronics industry to obtain faster retrieval than had been possible previously. By the 60's industry and the military had reached a point that required quicker response to the progress brought about by competition. Information had to be more available and exactly pertinent. Military libraries recognized the need for discipline oriented indexers. In our case at the Dahlgren Laboratory, we combined indexing and literature in one person. This effected, in my opinion, a more comprehensive look at the literature searching requirements of our laboratory. At present, there are six technical information specialists in the fields of physics/chemistry, mathematics, ordnance, electronics and engineering. The Thesaurus used for our computer storage and retrieval of about 100,000 documents is a result of indexing experience in the Uniterm Mode. We serve an audience which is largely concerned with specific weapon systems and their parameter equipment. We have to be able to retrieve very specifically for items such as "deflection of the vertical" or "at sea". In a large collection such as DDC is concerned with, this term might be included in "vertical orientation". Economics

and preference require that they be able to relate this terminology automatically in machine indexing.

The collection of 1498s selected from the total input for 1 year, 1973/74, as used for this study are concerned with some form of electronic sensors. Ideas that are related i.e. low-cost, etc. have all been included on the "cost effectiveness" card. "Ground" has been used to record items pertaining to "surface" as well as "ground". However, "terrain" has its own card. General terms such as "systems", "devices" etc. have been omitted because of the brevity and purpose of this study. In the original Uniterm System, descriptors were very narrowly used. "Phased arrays" would have been separate items as would "charge coupled devices". These are bound terms in the Thesaurus devised at Dahlgren; these combinations are recognized and are required for accurate input to the Computer Program. For the purpose of doing this study, the Uniterm System has been useful in point of time and depth of indexing. There are approximately 1,500 descriptor cards as a result of indexing less than 600 documents. In any new program, index terms will rise exponentially in the early development of a program. In the case of the 1498s concerned with electronic

sensors, the applications vary so much that terms or descriptors will
add up very quickly.

Figure 4 is a copy of the work sheet for the 1498 displayed in

Figure 1.

CHAPTER V CONCLUSION

The answers to the 40 questions resulted in these statistics:

MAI Recall 57% Precision 85%

DL Recall 92% Precision 90%

Let me review the program as far as type of documents and number of them. The indexing and retrieval has been done with 551 1498s (WUIS) documents from DDC. They are all on electronic sensors for the year 1973/74. The documents are largely application oriented R&D so that the items discussed are relatively apparent. The total results of the study should be considered in this respect, but it could perhaps represent the whole electronics collection as far as indexing and retrieval requirements are concerned. Other fields such as chemistry etc. may be more prone to concepts and would come up with different results.

After working with the indexing, retrieval and totals, I find I've changed my mind about the reason for MAI's lower recall rate. At first it seemed that the difference between the indexing systems might be the

questions. Some questions are apt to be for a concept i.e. "Signal processing techniques for obtaining radiometric target signatures" as different from a hardware question such as "ground laser locators". The concept might be open to individual judgment and therefore, I thought that this might be the grounds for some disagreement in the results.

With the recall for DL at 92%, I believe the problem with MAI at 52% must lie with the key puncher. In two cases, question 23 "ground laser locators" and 36 "Helmet mounted sighting devices" the descriptors are completely apparent in the documents indexed. Where DL missed on Recall, it appears to be mostly a carelessness. Except, in two cases MAI had added a term manually which was not in the documents that were indexed. A look at question 6 will illustrate this point. MAI had an 85% precision result of those recalled. DL had 90% precision. It is likely that this result may also be an input factor.

Questions 6, 8, 11, 26, and 29 appear to have had some manual enrichment because the terminology itself did not appear in the documents as I indexed them. Perhaps these descriptors will or have even by now been added to the MAI Natural Language Data Base.

It now seems that the human element between the document and computer is the most vital element. In questions such as #28 on Fluidic sensors, etc. and #37 Improvements in air traffic control etc., the fact that both questions are rather lengthy and more of the concept type - but answered currently by MAI and DL, indicates that the human input or indexing was up to snuff. Until such time as economics allows for complete optical indexing it appears that we must make an effort for excellence in the human interface.

BIBLIOGRAPHY

- Cleverdon, Cyril W., ASLIB Cranfield Research Project. Report on the first stage of an investigation into the comparative efficiency of indexing system. College of Astronautics, Cranfield, England, Sept. 1960.
- Cleverdon, Cyril W., ASLIB Cranfield Research Project. Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems. Cranfield, England, October, 1962.
- Cleverdon, Cyril W., Description and Evaluation of information systems. Annual Review of Information Science and Technology, Vol. 6, 1971, pp. 41-73.
- Cleverdon, Cyril W. The Cranfield Tests on Index Language Devices. ASLIB Proceedings. Vol 19, No. 6, pp. 157-209, London, 1967.
- Cleverdon, Cyril W. and J. Mills. The testing of index language devices. n.d.
- Cooper, William S. Expected search lengths. A single measure of retrieval effectiveness based on the weak ordering action of retrieval systems. American Documentation. January 1968, Vol. 19, No. 1, pp. 30-41.
- Cooper, William S. On Selecting a Measure of Retrieval Effectiveness. ASIS, M/A 1973, Vol. 24, No. 2, p. 87.
- Cuadra, Carlos A., ed. Annual Review of Information Science and Technology, Vol. 2, Interscience Publishers, New York, 1967.
- Cuadra, Carlos A., ed. Annual Review of Information Science and Technology, Vol. 4, Encyclopedia Britannica, Chicago, Ill., 1969.
- Cuadra, Carlos A., ed. Annual Review of Information Science and Technology, Vol. 5, Encyclopedia Britannica, Chicago, Ill., 1970.
- Jackson, David M. Basis for an improvability measure for retrieval performance. (OSU-CISRC-TR-70-25).
- Jacobs, Charles R. Machine-Aided Indexing. DDC. 1972, AD754400.
- Jardine, N. and C. J. Van Rusbergen. The use of hierarchic clustering in information retrieval. Information Storage and Retrieval. Vol. 7, pp. 217-249. Pergamon Press, 1971.

- Johnson, Samuel. The role of automatic indexing in access control: A modular view. ASIS Journal. May-June 1974, pp. 162-171.
- Keen, Michael. Search strategy evaluation in manual and automated systems. ASLIB Proceedings. Vol. 20, January 1968.
- Kent, Allen, et al., ed. Electronic handling of information: testing and evaluation. Academic Press, London, 1967.
- Kent, Allen. Textbook on mechanized information retrieval. 2nd ed., John Wiley, New York, 1966.
- Klingbiel, Paul H. Machine-Aided Indexing. DDC. 1971, AD721875.
- Klingbiel, Paul H. Machine-Aided Indexing. DDC. 1969, AD696200.
- Klingbiel, Paul H. A technique for machine-aided indexing, Information Storage and Retrieval. September, 1973.
- Lancaster, F. Wilfred. Information Retrieval Systems. John Wiley and Sons, New York, 1968.
- Pollack, Stephen M. Measures for the comparison of Information Retrieval System. American Documentation, October 1968, pp. 387-397.
- Selton, Gerard. A comparison between manual and automatic indexing methods. American Documentation, January 1969, p. 61.
- Sharp, John R. Some fundamentals of information retrieval. London House, 1965.
- Stevens, Mary Elizabeth. Automatic indexing: a state of the art report. NBS Mono. 91. 1965.
- Studies in indexing and cataloging. Management Information Services. Detroit, 1972.
- Tonik, Albert B., ed. Information retrieval. The user's viewpoint. An aid to design. 4th Annual Colloquium on Information Retrieval. International Information Inc., 1967.
- Vickery, B. C. On retrieval system theory. Butterworth, London, 1961.

APPENDIX A

UNCLASSIFIED
DDC REPORT NUMBER KEMPSM, SEP 24, 1974
DDC FOR AT B0001

ACCESSION NUMBER: DA0E1173

SUMMARY DATE: 15 MAY 74

PRIMARY NUMBER CODE:
63306A 1S263306D07000

001 F04D345

START DATE: APR 72

COMPLETION DATE: APR 74

RESPONSIBLE GOV'T ORGANIZATION
NAME
MICOM ARMY MISSILE COMMAND

PERFORMING ORGANIZATION
NAME
INTERNATIONAL LASER SYSTEMS
INC

ADDRESS
REDSTONE AR AL 35809

ADDRESS
ORLANDO FL 32804

RESPONSIBLE INDIVIDUAL
EMMONS, G

PRINCIPAL INVESTIGATOR
CROW, T

TELEPHONE NUMBER
2058761475

TELEPHONE NUMBER
3052954010

TITLE: (U) GROUND LASER LOCATOR DESIGNATOR

KEYWORDS: (U) DESIGNATOR ; (U) GUIDANCE LASER ; (U) GUIDED MISSILES
; (U) LASER ;

DESCRIPTORS: (U) AIRBORNE ; (U) GROUND SUPPORT EQUIPMENT ; (U) GUIDED
MISSILES ; (U) *LASERS ; (U) LIGHT HOMING ; (U) *OPTICAL TARGET
DESIGNATORS ; (U) PROJECTILES ;

TECHNICAL OBJECTIVE: (U) THE GROUND LASER LOCATOR DESIGNATOR (GLLD)
IS CONCEIVED AS A SYSTEM IN ITSELF, CONSISTING OF A LASER,
RANGEFINDER, TRACKING UNIT AND MOUNT, DAYSIGHT, AND NIGHT SIGHT
INTERFACE. THE GLLD WILL SATISFY ALL OF THE GROUND MOUNTED,
PRECISION DESIGNATOR REQUIREMENTS OF THE HELLFIRE AND CLGP WEAPON
SYSTEMS- IT WILL PROVIDE DATA INPUTS TO ARTILLERY FIRE CONTROL
SYSTEMS- AND IT WILL BE COMPATIBLE WITH THE TERMINAL HOMING
SYSTEMS OF ALL SERVICES. THE OBJECTIVE OF THIS PROGRAM IS TO
PROVIDE ADVANCED DEVELOPMENT MODELS OF THE GLLD AND TO CRITICALLY
EVALUATE THEIR CAPABILITIES.

APPROACH: (U) THE CONTRACTOR SHALL DESIGN, FABRICATE, AND ASSEMBLE
ADVANCED DEVELOPMENT MODELS OF THE GLLD- AND THE GOVERNMENT WILL
TEST THESE EQUIPMENTS FOR CONFORMANCE WITH WEAPON SYSTEM
REQUIREMENTS. NOTE- THIS IS ONE OF TWO CONTRACTS AWARDED- THE

Figure 1

A sample 1498 (WUIS) document

- 37 -

UNCLASSIFIED

PAGE 34

LOCATION, LOCATORS

0	1	2	3	4	5	6	7	8	9
			133	134 474	135 475		447		

LASERS

0	1	2	3	4	5	6	7	8	9
30 50 470	131 241 431	103 452 482	103 134 215 153 473	134 134 454	135 135 215	136	107 137 167 431 467	128 138 208 458	109 149

GROUND

0	1	2	3	4	5	6	7	8	9
200	191 221		133	134 224	135 185 485	136 206 216 236	107 187	148 228	149 169

FLIR (Forward Looking Infrared)

0	1	2	3	4	5	6	7	8	9
160	181	212	193	434	115	256	437	218	119
340		282	223		465				
			453						

Figure 3

An abbreviation card for UNITERM posting

134 ground laser
locator designator (GLLD)
(formerly GLD station)
masers
lasers
optics
control/guidance
guidance laser
guided missiles
rangefinder
tracking mount
daysight
nightsight
interface
Hellfire
CRGP weapon system
artillery fire control systems
terminal homing system

Figure 4

Sample of a worksheet for Figure 1

#-i

APPENDIX B

41

1 Fluxgate Magnetometer

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (3)	

Recall: $\frac{a}{a+c} = \frac{1}{1+3} = \frac{1}{4}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (4)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{4}{4+0} = 1$

Precision: $\frac{a}{a+b} = \frac{4}{4+0} = 1$

CONCLUSIONS:

The MAI program does not seem to include gradiometers in the fluxgate magnetometer category.

Reproduced from
best available copy.

2 Techniques for Use of Explosive Ordnance

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

CONCLUSIONS:

No disagreement.

3 Utilization of Extremely Low Frequency and Ultra Low Frequency Radiowaves for Communications

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (1)	

Recall: $\frac{a}{a+c} = \frac{1}{1+1} = \frac{1}{2}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{2}{2+0} = 1$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

CONCLUSIONS:

MAI did not pick up DNO20160.

44

4 Signal Processing Techniques for Obtaining Radiometric Target Signatures

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (2)	

Recall: $\frac{a}{a+c} = \frac{1}{1+2} = \frac{1}{3}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (3)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{3}{3+0} = 1$

Precision: $\frac{a}{a+b} = \frac{3}{3+0} = 1$

CONCLUSIONS:

MAI did not pick up DF 083610 and DF 063410.

45

5 Target surveillance by use of sonobuoys for AS^h

MAI

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (1)	

Recall: $\frac{a}{a+c} = \frac{2}{2+1} = \frac{2}{3}$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (3)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{3}{3+0} = 1$

Precision: $\frac{a}{a+b} = \frac{3}{3+0} = 1$

CONCLUSIONS:

MAI did not pick up DN475051.

46

6 Utilization of satellite link for a system of locating ships, liferafts, or lifeboats at sea with a display of location ashore

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (2)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = \frac{1}{1}$

Precision: $\frac{a}{a+b} = \frac{1}{1+2} = \frac{1}{3}$

DL

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

CONCLUSIONS:

MAI picked up DN326040 but it appears to be from a manual addition because the document does not contain this language in the portions used for machine indexing. MAI did not pick up DN214742 which is the only one called relevant by Dahlgren Laboratory.

47

7 Use of charge coupled devices as image sensors and trackers.
 Application required for Modular Missile Seeker (Single Solid
 State Clip desired)

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (1)	b ()
Not retrieved	c ()	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

CONCLUSIONS:

Agreement between systems.

48

8 Applications of Television for Air to Ground Reconnaissance

MAI

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (2)	

Recall: $\frac{a}{a+c} = \frac{2}{2+2} = \frac{2}{4}$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (3)	b (0)
Not retrieved	c (1)	

Recall: $\frac{a}{a+c} = \frac{3}{3+1} = \frac{3}{4}$

Precision: $\frac{a}{a+b} = \frac{3}{3+0} = 1$

CONCLUSIONS:

MAI picked up DAOD2179 by manual review - it couldn't have been picked up by machine indexing. Dahlgren Laboratory missed DAOD2179. MAI missed two others - DN482006 and DN237215.

49

9 Spark Gap Oscillators used as Signal and Noise Generators

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

CONCLUSIONS:

No disagreement.

50

10 Brown Magnetometer as Magnetic Surveillance Sensor

MAI

	Relevant	Non-Relevant
Retrieved	a (0)	b (4)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{0}{0+0} = 0$

Precision: $\frac{a}{a+b} = \frac{0}{0+4} = \frac{0}{4} = 0$

DL

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{2}{2+0} = 1$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

CONCLUSIONS:

The question asked for "Brown" magnetometers. MAI did not pick up either one. DN234455 or DN3344261. MAI picked up 4 irrelevant documents.

11 Lighting for Night Photographic Reconnaissance

MAI

	Relevant	Non-Relevant
Retrieved	a (3)	b (1)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{3}{3+0} = 1$

Precision: $\frac{a}{a+b} = \frac{3}{3+1} = \frac{3}{4}$

DL

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (1)	

Recall: $\frac{a}{a+c} = \frac{2}{2+1} = \frac{2}{3}$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

CONCLUSIONS:

DL should have included DAOD2538 which was picked up by MAI. DAOD2540 picked up by MAI appears to have had some manual editing since the term night reconnaissance does not appear in MAI portions.

52

12 Inertial Positioning System used for Mapping

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (6)	

Recall: $\frac{a}{a+c} = \frac{1}{1+6} = \frac{1}{7}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (7)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{7}{7+0} = 1$

Precision: $\frac{a}{a+b} = \frac{7}{7+0} = 1$

CONCLUSIONS:

DL picked up 7 relevant answers. MAI picked up only 1 of the 7.

53

13 Simulation of Low Light Level Television and Forward Looking Infrared Displays in Non Real Time

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (2)	

Recall: $\frac{a}{a+c} = \frac{1}{1+2} = \frac{1}{3}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (3)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{3}{3+0} = 1$

Precision: $\frac{a}{a+b} = \frac{3}{3+0} = 1$

CONCLUSIONS:

DL picked up 3 relevant documents and MAI only one of these.

14 Electrooptical Sensor Simulation for Use in Training

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (2)	

Recall: $\frac{a}{a+c} = \frac{1}{1+2} = \frac{1}{3}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (3)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{3}{3+0} = 1$

Precision: $\frac{a}{a+b} = \frac{3}{3+0} = 1$

CONCLUSIONS:

DL picked up three relevant documents. MAI picked up only one of these.

55

15 A System to Automatically Maintain Alignment of Airborne Sensing and Designating Equipment in a Flight Environment

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (8)	

Recall: $\frac{a}{a+c} = \frac{1}{1+8} = \frac{1}{9}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (9)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{9}{9+0} = \frac{9}{9} = 1$

Precision: $\frac{a}{a+b} = \frac{9}{9+0} = \frac{9}{9} = 1$

CONCLUSIONS:

DL picked up 9 relevant documents and MAI picked up one of these.

56

16 Development of Instrumentation for Electromagnetic Pulse Sensors, the Measurement Techniques, and Calibration Standards

MAI

	Relevant	Non-Relevant
Retrieved	a (4)	b (0)
Not retrieved	c (2)	

Recall: $\frac{a}{a+c} = \frac{4}{4+2} = \frac{4}{6}$

Precision: $\frac{a}{a+b} = \frac{4}{4+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (4)	b (2)
Not retrieved	c (2)	

Recall: $\frac{a}{a+c} = \frac{4}{4+2} = \frac{4}{6}$

Precision: $\frac{a}{a+b} = \frac{4}{4+2} = \frac{4}{6}$

320 - manual addition - not on 1498.

CONCLUSIONS:

MAI picked up 4 relevant documents. DL picked up only 2 plus 2 that MAI missed. DL also picked up 2 that were irrelevant.

57

17 Low Cost, Light Weight, Doppler Navigation System

MAI

	Relevant	Non-Relevant
Retrieved	a (3)	b (0)
Not retrieved	c (8)	

Recall: $\frac{a}{a+c} = \frac{3}{3+8} = \frac{3}{11}$

Precision: $\frac{a}{a+b} = \frac{3}{3+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (11)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{11}{11+0} = 1$

Precision: $\frac{a}{a+b} = \frac{11}{11+0} = 1$

CONCLUSIONS:

MAI picked up 3 relevant documents. DL picked up 11 documents which included the MAI 3 documents.

58

18 Monopulse Radars and Processors to Improve Low Angle Tracking

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (3)	

Recall: $\frac{a}{a+c} = \frac{1}{1+3} = \frac{1}{4}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (4)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{4}{4+0} = 1$

Precision: $\frac{a}{a+b} = \frac{4}{4+0} = 1$

CONCLUSIONS:

DL picked up 4 relevant documents. MAI picked up only 1 of the 4.

19 Forward Looking Infrared (FLIR) Sensor to be Used in Navy Aircraft

MAI

	Relevant	Non-Relevant
Retrieved	a (13)	b (0)
Not retrieved	c (4)	

Recall: $\frac{a}{a+c} = \frac{13}{13+4} = \frac{13}{17}$

Precision: $\frac{a}{a+b} = \frac{13}{13+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (15)	b (0)
Not retrieved	c (2)	

Recall: $\frac{a}{a+c} = \frac{15}{15+2} = \frac{15}{17}$

Precision: $\frac{a}{a+b} = \frac{15}{15+0} = 1$

CONCLUSIONS:

There were 4 documents not picked up by MAI that DL retrieved. DL did not pick up 2 that were on the MAI list. The errors must be in machine indexing for MAI. On DL side it must be improper indexing not to pick up the 2 documents listed by MAI.

60

20 Use of FM/CW, Millimeter-wave Radiometric Tracking for Terminal Guidance of Air-to-surface Conventional Munitions

MAI

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (1)	

Recall: $\frac{a}{a + c} = \frac{2}{2+1} = \frac{2}{3}$

Precision: $\frac{a}{a + b} = \frac{2}{2+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (1)	

Recall: $\frac{a}{a + c} = \frac{2}{2+1} = \frac{2}{3}$

Precision: $\frac{a}{a + b} = \frac{2}{2+0} = 1$

CONCLUSIONS:

MAI and DL each picked up DF628020 but MAI listed DF090570 which DL did not pick up and MAI did not pick up DN230052 listed by DL.

61

21 Use of Laser Designators and Laser Guidance to Increase Ordnance Delivery Capability

MAI

	Relevant	Non-Relevant
Retrieved	a (10)	b (0)
Not retrieved	c (2)	

Recall: $\frac{a}{a + c} = \frac{10}{10+2} = \frac{10}{12} = \frac{5}{6}$

Precision: $\frac{a}{a + b} = \frac{10}{10+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (12)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a + c} = \frac{12}{12+0} = 1$

Precision: $\frac{a}{a + b} = \frac{12}{12+0} = 1$

CONCLUSIONS:

MAI did not pick up DF626990 and DF627880.

62

22 Angular Motion Sensors to Maintain Alignment of Sensing Equipments in Space Craft

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (2)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+2} = \frac{1}{3}$

DL

	Relevant	Non-Relevant
Retrieved	a (0)	b (1)
Not retrieved	c (1)	

Recall: $\frac{a}{a+c} = \frac{0}{0+1} = \frac{0}{1} = 0$

Precision: $\frac{a}{a+b} = \frac{0}{0+1} = \frac{0}{1} = 0$

CONCLUSIONS:

DL picked up DF36848 which was irrelevant and did not pick up DF368409 which MAI did as relevant. MAI picked up DF368170 as irrelevant.

63

23 Ground Laser Locator Designator to be Used for Range Finding,
Tracking in Either Day or Night Mode of Operation

MAI

	Relevant	Non-Relevant
Retrieved	a (0)	b (1)
Not retrieved	c (4)	

Recall: $\frac{a}{a+c} = \frac{0}{0+4} = \frac{0}{4} = 0$

Precision: $\frac{a}{a+b} = \frac{0}{0+1} = \frac{0}{1} = 0$

DL

	Relevant	Non-Relevant
Retrieved	a (4)	b (1)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{4}{4+0} = 1$

Precision: $\frac{a}{a+b} = \frac{4}{4+1} = \frac{4}{5}$

CONCLUSIONS:

MAI did not pick up 4 relevant documents which had the actual words ground laser locators... in the title. They had 1 irrelevant title. DL picked up one irrelevant document but had the 4 relevant documents.

64

24 Simple High Accuracy Guidance Air to Air Missile.
Pilot Guided, Man in the Loop, Concept

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (1)	

Recall: $\frac{a}{a+c} = \frac{1}{1+1} = \frac{1}{2}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{2}{2+0} = 1$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

CONCLUSIONS:

MAI did not pick up DN230077.

65

25 Infrared Signature Measurement

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (8)
Not retrieved	c (2)	d (-)

Recall: $\frac{a}{a + c} = \frac{1}{1+2} = \frac{1}{3}$

Precision: $\frac{a}{a + b} = \frac{1}{1+8} = \frac{1}{9}$

DL

	Relevant	Non-Relevant
Retrieved	a (3)	b (1)
Not retrieved	c (0)	d (0)

Recall: $\frac{a}{a + c} = \frac{3}{3+0} = 1$

Precision: $\frac{a}{a + b} = \frac{3}{3+1} = \frac{3}{4}$

CONCLUSIONS:

MAI picked up 7 irrelevant documents. One of the two which are relevant had to have had manual indexing since the term infrared did not appear in the MAI portion for machine indexing. DL included a non-relevant document which had also been included in MAI.

66

26 Low Cost Solid State Gyro Capable of Withstanding High Acceleration
to be Used in Guidance System in Artillery Round

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (1)	d (0)

Recall: $\frac{a}{a+c} = \frac{1}{1+1} = \frac{1}{2}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (1)	d (0)

Recall: $\frac{a}{a+c} = \frac{1}{1+1} = \frac{1}{2}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

CONCLUSIONS:

MAI retrieved DF083130 which could not have been retrieved
by machine indexing. They did not retrieve DA060143 which DL did.
DL of course did not pull DF083130.

67

27 High Power, High Efficiency Radar Traveling Wave Tube in the X-Band

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (2)	d (0)

Recall: $\frac{a}{a+c} = \frac{1}{1+2} = \frac{1}{3}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (3)	b (0)
Not retrieved	c (0)	d (0)

Recall: $\frac{a}{a+c} = \frac{3}{3+0} = 1$

Precision: $\frac{a}{a+b} = \frac{3}{3+0} = 1$

CONCLUSIONS:

MAI did not pick up DF092400 and DN484050 which are relevant and terms were obvious in the portion indexed by MAI.

68

28 Fluidic Sensors and Transducers for Sensing Flight Vehicles
Environment to be Used for Safeing, Arming and Fuzing Munitions

MAI

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{2}{2+0} = 1$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{2}{2+0} = 1$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

CONCLUSIONS:

No disagreement.

69

29 Radio Frequency Interference (RFI) Reductions

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (1)
Not retrieved	c (1)	

Recall: $\frac{a}{a+c} = \frac{1}{1+1} = \frac{1}{2}$

Precision: $\frac{a}{a+b} = \frac{1}{1+1} = \frac{1}{2}$

DL

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

CONCLUSIONS:

MAI had one relevant DF091680 which would only have been retrieved by manual addition of terms. They also had one irrelevant document and did not pick up at all which DL had retrieved.

70

30 Automatic Passive Tracker High Sensitivity and Accuracy are Required to be Used for Fire Control Sight in Airborne Application

MAI

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{2}{2+0} = 1$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{2}{2+0} = 1$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

CONCLUSIONS:

No disagreement.

71

31 Ferrite Materials for Use in Phased Array Radar Systems

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

CONCLUSIONS:

No disagreement.

72

32 Fuzing for Mid Course Intercept Applicable to an Exoatmospheric Intercept to be Used for Ballistic Missile Defense

MAI

	Relevant	Non-Relevant
Retrieved	a (4)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{4}{4+0} = 1$

Precision: $\frac{a}{a+b} = \frac{4}{4+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (4)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{4}{4+0} = 1$

Precision: $\frac{a}{a+b} = \frac{4}{4+0} = 1$

CONCLUSIONS:

No disagreement.

73

33 Image Intensifier Tube to be Used for Night Vision Incorporated
in Tank Drivers Periscopes and Gun Sights

MAI

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (3)	

Recall: $\frac{a}{a+c} = \frac{2}{2+3} = \frac{2}{5}$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (5)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{5}{5+0} = 1$

Precision: $\frac{a}{a+b} = \frac{5}{5+0} = 1$

CONCLUSIONS:

MAI did not pick up three documents that are relevant and terms appear in the portions indexed by MAI so should have been included.

74

34

Radio Frequency Device for Air Craft Use to Prevent Collision,
Provide Proximity Warning in Formation Flying

MAI

	Relevant	Non-Relevant
Retrieved	a (3)	b (0)
Not retrieved	c (1)	

Recall: $\frac{a}{a+c} = \frac{3}{3+1} = \frac{3}{4}$

Precision: $\frac{a}{a+b} = \frac{3}{3+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (4)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{4}{4+0} = 1$

Precision: $\frac{a}{a+b} = \frac{4}{4+0} = 1$

CONCLUSIONS:

MAI did not pick up DN333035.

75

35 High Power Transistors Development

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (1)
Not retrieved	c (1)	

Recall: $\frac{a}{a+c} = \frac{1}{1+1} = \frac{1}{2}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{2}{2+0} = 1$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

CONCLUSIONS:

MAI did not pick up DF 095400.

Fl

36 **Helmet Mounted Sighting and Pointing Devices for Target Acquisition and Weapon Pointing**

MAI

	Relevant	Non-Relevant
Retrieved	a (0)	b (0)
Not retrieved	c (4)	

Recall: $\frac{a}{a+c} = \frac{0}{0+4} = 0$

Precision: $\frac{a}{a+b} = \frac{0}{0+4} = 0$

DL

	Relevant	Non-Relevant
Retrieved	a (4)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{4}{4+0} = 1$

Precision: $\frac{a}{a+b} = \frac{4}{4+0} = 1$

CONCLUSIONS:

MAI key puncher must have missed the four documents because the documents all used the words "helmet mounted" in four instances.

77

37 Improvements in Air Traffic Control Equipment in the Area of
Sensor and Data Link Development

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

CONCLUSIONS:

No disagreement.

78

38 Airborne Devices for Computing Time of Flight for Laser Guided Weapons for Fuzing Functions

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{1}{1+0} = 1$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

CONCLUSIONS:

No disagreement.

79

39 Microelectronic Circuitry for U. S. Navy Electronic Equipment to Provide Improved Performance, Reliability, Maintainability, Serviceability and Commonability.

MAI

	Relevant	Non-Relevant
Retrieved	a (1)	b (0)
Not retrieved	c (2)	

Recall: $\frac{a}{a+c} = \frac{1}{1+2} = \frac{1}{3}$

Precision: $\frac{a}{a+b} = \frac{1}{1+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (3)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{3}{3+0} = 1$

Precision: $\frac{a}{a+b} = \frac{3}{3+0} = 1$

CONCLUSIONS:

MAI did not pick up DN230052 and DN234580.

80

40 Missile Scoring Systems for Subsonic Drone Engagements in
 Combat Fighter Training

MAI

	Relevant	Non-Relevant
Retrieved	a (2)	b (0)
Not retrieved	c (1)	

Recall: $\frac{a}{a+c} = \frac{2}{2+1} = \frac{2}{3}$

Precision: $\frac{a}{a+b} = \frac{2}{2+0} = 1$

DL

	Relevant	Non-Relevant
Retrieved	a (3)	b (0)
Not retrieved	c (0)	

Recall: $\frac{a}{a+c} = \frac{3}{3+0} = 1$

Precision: $\frac{a}{a+b} = \frac{3}{3+0} = 1$

CONCLUSIONS:

MAI did not pick up DN320164.