UNCLASSIFIED

AD **409 276**

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA

Mechanized Indexing Methods and Their Testing[*]

John O'Connor

Institute for Scientific Information

ABSTRACT

Methods of mechanized indexing (subject indexing by computer) which have been proposed are systematically summarized. Every suggested method consists of some document preparation process (mostly or wholly mechanical) followed by the application of indexing rules to the prepared document. A comprehensive document preparation is described from which proposed methods can be derived by selection. It includes full text input, "document place" (title, abstract, etc.) marking, sentence and paragraph marking, pronoun replacement and other syntactic marking. It also includes addition of "thesaurus" headings, position numbers, weighted frequencies, "closely associated" expressions, importance measures, and reference information. (Some questions are raised about some of these preparation procedures). Three kinds of indexing rules are then distinguished and illustrated.

Several general comments on mechanized indexing include remarks on the argument that good mechanized indexing is not feasible, and the argument that mechanized indexing has the advantage, compared to human indexing, of consistency.

Some problems of testing mechanized (or any other) indexing quality by the quality of the retrieval it permits are described. "Index duplication" studies are suggested as an alternative kind of empirical investigation of mechanized indexing methods.

A Postscript raises a question about mechanized indexing which is of broader social significance.

1. Introduction

Most or all retrieval systems for scientific documents require subject indexing; the exceptions are text searching systems, which are still experimental. (Swanson,a;Kehl). "Subject indexing"here means the assignment to a document of words or phrases, indicating its content, which can be used later to aid searching. Familiar illustrations are the listing of a document under subject headings or class names in a card catalogue or book-form subject index. Mechanized retrieval systems also usually use subject indexing. For example, in a pharmaceutical retrieval system using a punched card sorter for searching, one of the documents is represented by the index term set: skin, mycobacteria, bone, vaccination, therapy, humans, tuberculosis, bacteria, wounds and injuries, infection, toxicity, children.

In many or most such systems the indexing is done by subject specialists. Such people are in short supply and are relatively expensive. One estimate is that subject indexing accounts for about three quarters of the cost of operating a retrieval system.

Accordingly in recent years a number of methods have been proposed for subject indexing by computer (mechanized indexing). Most of these techniques require that the full texts of documents be in machine-readable form. At present this usually requires keypunching which is more expensive than a specialist's indexing effort. But the study of mechanized indexing methods presupposes the development of print readers which will machine text economically, and/or the increasing

use of printing processes which produce a machine-readable text
as a by-product.

This paper summarizes the proposed methods of mechanized
indexing with some comments, and then discusses the question of
how the effectiveness of such methods can be empirically studied.

2. Document Preparation

Each mechanized indexing method envisions preparing the
document text in various ways, mostly by computer, and then using
the prepared text as the input to some kind of indexing rules.
This section describes a comprehensive text preparation from
which any preparation that has been actually proposed can be
derived by selection. In the course of the exposition some presently
unsolved problems of mechanized preparation are indicated.

We begin with the input of full text, including distinctive
codes for special symbols such as integral signs, chemical reaction
arrows, etc.. Codes are also used to represent upper case, italics,
boldface, different print-sizes (e.g. smaller print is sometimes
used by an author for material he regards as subsidiary), etc..
Subscripts, superscripts, and any other such devices are recorded.
New line beginnings and indentations (as for new paragraphs) are
represented, as are other spaces (such as those usually separating
a heading from surrounding text). Material which does not occur
in an obvious sequence with the rest of the document, such as
textual material accompanying figures and footnotes, is distinctively
coded and placed in some standard position, e.g. after the rest
of the text.

Each word and non-verbal text expression (e.g. an integral
sign) is marked with its "document place". This might be title,
abstract or summary, heading, main text, footnote, expressions
in a figure, references etc. (Oswald et al)(2).The "document places"
might not be specific enough. For instance, should a last para-
graph beginning "In conclusion", a section labelled "abstract",
and a brief summary above the title all be labelled "abstract"?
Should distinctions be made between headings and sub-headings,
etc., between expressions in a figure (e.g. a table) and expres-
sions labelling the figure, and so on.

These questions shade into the problems of how to identify
document places mechanically. These problems have not been
much discussed. To avoid such complications, document places
can be marked by human editing before machine input. To the
extent that they are not given unambiguous rules, the editors
will also be uncertain.

The text is marked with paragraph and sentence divisions.
Some document places, such as title, headings, and references,
might be exempted from such marking. The paragraph and sentence
marking might be done mechanically. But ".... even with the
aid of a dozen different tests performed by the machine, the
true end of a sentence cannot be determined with certainty",
(Luhn,b). Alternatively human pre-editing might add paragraph
and sentence marks.

To each pronoun its antecedent is added. This includes
not only pronouns such as "it" and "they", but also expressions
such as "this condition" and "these results", and abbreviations

such as "the S group". Pronoun replacement might also be
extended to cases such as the following: the expression "the
gland" is short for"the adrenal gland" in a paper on that
organ (Harris). Decisions have to be made about just how far
to carry such expansions of compressed reference. For example,
suppose that a complicated therapeutic treatment is described
in a lengthy passage, and then later in the paper the expres-
sion" the treated patients" occurs?

Even if clear-cut general decisions have been made about
what replacements are wanted, the problems of how to accomplish
them mechanically have not yet been solved. For instance, inter-
sentence pronoun-antecedent relations still present difficulties
even for simpler abbreviations such as "it" and "this condition".

Syntactic information is added to the text words. Each
word is marked with its part of speech, and more generally each
sentence is marked with its syntactic structure. These processes
have been mechanized with great,but not yet total,success (e.g.
Kuno and Oettinger).

Each sentence is next "kernelized"(Harris). "Kernelization"
needs some explanation. There are a few kernel sentence types
(in English at least), mostly of forms NV,NVN, NVPN, NVNPN,
N is N, N is A, and N is PN. (Here N,V,P, and A mean noun,verb,
preposition, and adjective). Several of these can be combined
by transformations. For example, N is A can be transformed to
the noun phrase AN and substituted in NVPN to give ANVPN. Ker-
nelization is the unraveling of a sentence into its constituent

kernels and transformations. For example, the sentence "The
optical rotatory power of proteins is very sensitive to the
experimental conditions under which it is measured, particularly
the wavelength of light which is used" is analysed into the
following kernels: the power is rotatory/the rotation is optical/
the power is of proteins/the power is very sensitive to the
conditions/ the conditions are experimental/ the power is
measured under the conditions ("which" is associated with this
kernel as a "connector")/ the power is very sensitive to the
wavelength ("particularly" is a "connector" of this kernel)/
the wavelength is of light/ light is used. (The similarity
in style to a first grade reader is not completely accidental.)

Harris has conjectured that a less ultimate kernelization
would be preferable for various retrieval purposes, including
mechanized indexing. However where it is best to halt kernel-
ization for this purpose (or others) is still a question.

Mechanized kernelization, without regard to the problem
of where it is best to halt it, is still a problem for research.

Variations in linguistic expression of the same meanings
("same" at least for purposes of the retrieval system) are
minimized. The set of rules used for this purpose is often
called a "thesaurus". An input to a thesaurus rule might be
called an "entry", and is sometimes called a "keyword".
The output of a thesaurus rule is usually called a "heading"
or "head". The thesauric phase of document preparation consists
of marking all words and phrases which are entries with the
appropriate thesaurus headings.

A variety of kinds of rules for headings are used. Phrases which are to be treated as units, e.g. "side action" are so marked. Phrases so marked might include all two and three word sequences unbroken by punctuation and not containing "empty" words,("the","and","of", etc..) (Luhn, a).

An entry expression (word phrase,or non-verbal expression) may have as corresponding head a preferred synonym; e.g. "side action" as entry may lead to the head "side effect". Or an entry may be under a head which is a synonym in some contexts, (partial synonym), e.g. "reaction " and "biological response". If an entry expression is too specific for purposes of the retrieval system, its corresponding thesauric heading will be more general; e.g. "butterfly" and "beetle" both may lead to the heading "insect" in an electronics collection (Oswald et al). An entry expression under a "more general" heading need not even represent a species of the genus represented by that heading, but simply have the heading occur in a definition of the entry; e.g. the entry "hypercalcemia" (an excess of calcium in the blood) might be placed under the heading "calcium" (Montgomery and Swanson), even though hypercalcemia is not a species of calcium. Inflectional variants of a word may be standardized to the same stem, for example, "toxic", "toxicity","intoxicate", etc. may all be standardized to "toxic". However if such stem selection is attempted not by rules specifying particular entry words but by general truncation rules or affix removing rules (a list of affixes being provided), difficulties can arise. For example,

truncation may equate "differentiate" and "difference", when they should be kept distinct (e.g. in a mathematical content). (Luhn c, Bar-Hillel,b).

An entry may have more than one partial synonym as a thesaurus head; for example "reaction" might have both "biological response" and "chemical transformation" as heads. All heads might be added in such a case. Alternatively, if thesaurus heads have some relationships among themselve these might be used to help select the appropriate head. For example, if "most" of the heads for "nearby" entries are characterized in the thesaurus as "biologic", the "biologic response" is the head selected for "reaction". How reliable such procedures can be is not known.

The syntactic analysis can help some in reducing the range of possible heads for an entry. For example, if "trains" has been identified as a noun, then it does not mean "trains" in the sense of "instructs". But syntactic information by itself seems insufficient for resolving mechanized indexing's multiple problems. The experience of machine translation suggests this, and the example of "reaction" illustrates it.

Each word and non-verbal text expression is assigned a paragraph number, sentence number (in the paragraph), and word number (in the sentence). In numbering paragraphs, sentences, and words, decisions must be made about how to count non-verbal expressions such as formulas and equations. The numbering of expressions in such places as figures and footnotes

also requires special decisions. Further interesting questions
are whether the word counts and position numbering should take
account of pronoun replacements, phrases treated as units,
and syntactic rearrangements.

First and last paragraphs and first and last sentences
in paragraphs may be specially marked (Baxendale).

The first occurrences of non-empty words may also be
marked (Storm).

Each expression (word, unit phrase, thesaurus head, or
non-verbal expression) is marked with its frequencies in the
document. These frequencies include at least its absolute
frequencies in the whole document (Luhn a) and in its
document place. Frequencies are weighted by the total number
of words (or words and other expressions) in the document,
and↑the total number in the expression's document place.
   perhaps
Frequencies are also weighted by the average frequency of
expressions in the document, and in the document places.

A frequency of an expression is also weighted by its
frequencies in various kinds of literature, "general"
literature and various special kinds of literature of interest
to the retrieval system. For example, a document frequency of
the word "emotion" which is only average for "general"
literature may be quite significant for "electronics" liter-
ature in which the word is rare (Oswald et al, Bar-Hillel,a).

Expressions which may not occur in the document or as thesaurus heads marking it are added if they are "associated closely" with expressions of the document, by any of a number of association measures (Maron and Kuhns, Stiles, Needham, Giuliano and Jones). The inputs for every proposed association measure are: for each expression the number of documents in which it occurs, and for each pair of expressions the number of documents in which they co-occur. Weighting for document places, (e.g. the expressions co-occur in the abstract), syntactic roles, position, etc. have not been considered, primarily because of computational difficulties. The literature within which occurrences and co-occurrences are counted may be the retrieval system's document collection or the literature of some specified field.

The addition of closely associated expression carried on through several "generations" of associations can add expressions helpful for retrieval and otherwise absent from the document. For example, a document may contain "fungicidal" and be relevant to a search for documents on "weatherproofing". If "fungicidal" occurs relatively frequently with "fungus" in documents in the retrieval system (and this is a measure of "close association") and "fungus" in turn is closely enough associated with "weatherproofing", then "weatherproofing" will be an expression added to the document (Stiles). In general, association over several "generations" may add to a document containing expression X and expression y which is "similar in meaning" to X and therfore may have a "similar environment" (similar close associations)(Giuliano and Jones).

However, except in the vaguest sense of the expressions involved, it is not known how true it is that "similar meanings" imply "similar environments". And even if this is true in some precise sense, it is still a question whether such "similarity of environment" is reflected in occurrences and co-occurrences of words and phrases (the only expressions so far considered).

Decisions have to be made about whether expressions added because they are closely associated are to be marked with any document place, syntactic, thesauric, position number, or frequency information (or analogues of these), perhaps on the basis of the markings of some document expressions with which they closely associate.

For references in the document which have themselves been indexed in the retrieval system, the indexing terms (not necessarily as definite index terms for the document being prepared) are added.(2) The index terms of documents in the retrieval system whose bibliographies "closely resemble" the bibliography of the document being processed are also added (Salton, Kessler). In general, material is added from documents "closely connected" to the processed document by citation relations (e.g. the processed document and some others form a small set within which citation is unusually frequent) (Salton, Kessler) (2,3). And the material added need not be index terms but may be title words, frequent words, etc., from the bibliographically related documents.

Decisions have to be made about whether expressions added from bibliographically related documents are to be marked with any analogues of place, syntactic, etc. information (e.g. more weight for a reference in the first paragraph).

As the last step in preparation for mechanized indexing, each expression in the document is marked with an importance measure, obtained by a dictionary look-up, appropriate for that retrieval system. For example, in a physics retrieval system, the words "how", "measure", and "protons" have respectively weights of 0, 1, and 9 (Swanson b).

3. Indexing Rules

We now have a document in which there has been pronoun replacement and each expression (word, unit phrase, non-verbal expression, thesaurus head) is accompanied by information about its document place, syntactic roles, position, frequencies and importance measure. In addition, original document expressions are accompanied by thesaurus heads. And the whole document or expressions in it are accompanied by "closely associated" expressions, and bibliographically related expressions,which in turn may be marked with some analogues of place, syntactic, position, frequencies, and importance measures.

This whole marked document is input to indexing rules. These rules may select expressions which occurred in the original text (unprepared document),select expressions added during the document preparation, or assign terms on the basis of the prepared document.

Selection rules of either kind assign a score to each expression on the basis of some function of its marks and

perhaps those of some other expressions in the document too

(e.g. other expressions near in position). The expressions with

the highest scores (e.g. highest 1%, first twenty, above some

absolute figure) are selected to be the indexing terms.

Some examples of original text selections are the following.
Most frequent words, (omitting "empty" words),↑perhaps taking
(Luhn a,b)

account of some weighting of the frequencies (Oswald et al). Most

frequent words in first and last sentences of paragraphs (Baxendale).

Most frequent word pairs (omitting empty words) (Oswald et al).

Frequent non-"empty" words in kernels, and centers (gramatically

essential words) in kernels connected to frequent word kernels

(e.g. by if... then...) (Harris). Certain words whenever they

occur (Luhn a, Harris).

Some examples of prepared text selections are a

thesaurus heading which occurs at least twice within two para-

graphs (Luhn a), any thesaurus heading which occurs (Swanson b),

closely associated expressions (Stiles).

Assignment rules consists of a standard vocabulary of

indexing terms, and rules for determining a weight W_j for

each indexing term T_j from any marked document which might

be an input to the assignment rules. Those terms with the

largest weights (by some measure) for a document are the

indexing terms for that document.

Some examples of proposed assignment rules are the

following. Each word and index term are associated by a
weight, determined by the frequency of occurrence of that
word in documents indexed (by humans) with that term; the
words of a document to be mechanically indexed thus con-
tribute a total weight to each index term, and the terms
with highest weights are assigned (Maron). Index terms
are class names in a classification obtained by factor
analysis (each factor determining a class); and each word
in a document (of those used in the factor analysis)
contributes as its weight to a factor its loading on that
factor; the heaviest weighted factors for a document are its
assigned terms (Borko and Bernick).

4. Several General Remarks on Mechanized Indexing Methods.

In existing retrieval systems, many of the documents
examined are not indexed because they are not important
enough to process, and/or indexing them would "clutter up"
the system. If indexing can be mechanized, either quick
human selection must be feasible, or mechanical selection
(before document preparation or as a result of indexing rules).
Otherwise the cost of processing everything received must be
acceptable and this must not unduly "clutter up" the system
(e.g. too many trivially relevant documents in response to

to searches). This question of document selection for indexing has received almost no explicit discussion in mechanized indexing literature. (An exception is Swanson, c). Perhaps it would be usually satisfactory to have cerebral (human) selection of papers for machine indexing. For example, people familiar with the pharmaceutical indexing system at the Merck Sharp & Dohme Research Center (Lansdale, Pennsylvania) estimate that a specialist who now indexes could decide in about a minute whether or not a paper should be indexed, while it takes an average of fifteen minutes actually to index a paper.

Two different retrieval systems may index the same document in quite different ways, because of different user interests. If mechanized indexing is feasible, where can such differences of indexing enter? The two indexing systems can differ in thesaurus rules, importance measures, the choice of literature from which to derive weights for document expression frequencies, or the choice of literature from which to derive associations among expressions. And even if there are no differences between the indexing pro-cedures of the two retrieval systems in any of these document preparation phases, if they use assignment indexing rules, differences can enter there (even for the same indexing vocabulary).

Bar-Hillel has argued that good quality mechanized indexing will not be feasible, at least for several decades (Bar-Hillel a,b). He emphasized the frequent differences between a literature searcher's vocabulary and a "relevant" document's language, and expresses the opinion that such differences cannot be successfully overcome by machine methods.

The arguments are not conclusive, but are useful reminders of difficulties. A "thesaurus", "closely associated" terms, and "bibliographically related" material are each intended to help meet this problem. Determining whether they can do so satisfactorily presumably requires a great deal of careful empirical study.

A specific remark of Bar-Hillel on digrams (two word sequences) needs some comment here. Suppose that digram frequencies for some field of literature are needed, for example to weight digram frequencies in a document. For the 10^8 digrams in English he argues that "no practical method is in view how to arrive at their relative frequency list" (Bar-Hillel, a). However it might not be necessary to consider 10^8 possible digrams separately. A sample of N running words of text generates about N different digrams (or trigrams or n-grams). The frequency in the literature sampled of any digram absent from the sample has an upper limit set by the sampling error.

It has been argued that mechanized indexing has the
advantage of consistency. The same program operating on the
same document will always produce the same result; while
different human indexers, and sometimes even the same
indexer at different times, often produce varying indexing
of the same document (Baxendale).

However this argument by itself says very little in
favor of mechanized indexing. For two humanly produced
index sets for a document which differ somewhat may both
be quite useful, though imperfect, while the index set
which the same program will always reproduce for the same
document may be worthless. Of what value is consistency
then?

A more recent illustration of a similar confusion
is to be found in the following passage. "Are human indexers
both self-consistent and consistent with one another? If
so, are they making choices consistent with effective
retrieval of the indexed information? If the answer to
both questions is 'yes', then clearly the intellectual
aspects of indexing are of much interest for further
analysis. If the answer turns out to be 'no', we might
reasonably conclude that the only reliable and effective
kind of human indexing is that which is already machine-
like in nature."(Montgomery and Swanson) . This assumes
that the answers to the two questions may not be "no and
yes".

5. Some Problems in Retrieval Testing of

      Mechanized Indexing Methods.

    The principal question about any proposed method of
mechanized indexing is a simple one: how good is the
indexing produced by the method? Unfortunately the sim-
plicity of the question is deceptive.

    Indexing is done to aid retrieval. Therefore
indexing quality can seemingly be determined by measuring
the quality of the retrieval it permits (supposing such
measurement possible). "Classification and indexing are
necessarily the proper points of beginning [attempts to
improve retrieval] since, once these operations have been
completed ... the die is already cast with respect to the
effectiveness of the library as an instrument for infor-
mation retrieval" (Montgomery and Swanson, p. 266). " [to
determine the quality of mechanized indexing, supposing
retrieval quality can be measured] take a collection
which has been mechanically indexed, perform retrievals on
the basis of the mechanized indexing, and see if the retrievals
are good enough" (O'Connor, p. 273).

    This approach to measuring indexing quality assumes
that retrieval quality is the result only of indexing. But
retrieval quality may also be seriously influenced by
many other factors, which might be called "search aids".

These include such factors as how the indexing vocabulary
is arranged for consultation by searchers, what kind of
cross-references are provided, when searchers are distinct
from users,what the searchers' backgrounds are and the
nature of searcher-user communication (e.g. written or
in person), the delay between a search question formulation
and first search results (determining how many search
cycles are feasible), and what kind of "intermediate
information" about selected documents is provided as output
(e.g. authors, titles, abstract, "tailored" output of
various kinds such as the contexts of certain words spe-
cified by the searcher, etc.)

These factors may have only a secondary effect on
retrieval, compared to the effect of indexing, but this is
not known. The Ramo-Wooldridge study has included some
varying of search aids (Swanson a). But the whole subject,
both particular cases and in general, needs much further
study.

For the sake of further discussion, let us assume
here that the quality of indexing can be determined by
measuring the quality of retrieval.

How can the quality of retrieval be measured?

One begins with a user of the retrieval system who
asks for documents of kind S, and who is able to determine,
when presented with a document, whether it is or is not

of kind S. But then many distinctions have to be made.

Does the user want any one S document (to answer a question), a few (to start on a subject), most in the collection (for a good grasp of the subject), or all in the collection (an exhaustiveness needed for scientific, military, safety, or legal purposes)?

Does the user want, in addition to S documents, "related" documents? These are not S documents but are likely to be of interest to someone looking for S documents (especially if there are none of the latter). They may be produced by a search as a matter of course, through cross -references, the relatively general character of some search terms, etc.. For example, searching the modifications (brief descriptions) under a heading in the Chemical Abstracts subject index leads to brief descriptions of sought for documents if any appear there, and to modifications representing other documents which share a heading with the kind of document sought (Bernier). If the user does want "related" documents, how can the quality of retrieved "related" documents be judged? Presumably the user judges; it was good to retrieve those "related" documents he finds he is glad to have.

A user cannot always judge immediately whether or not a document is of kind S. For example, in the Ramo-

Wooldridge text searching study physicists judged the
relevance of documents to various questions (and even
checked one another's work). Nonetheless it sometimes
happened that a document judged irrelevant for a question
but retrieved for it, was found upon re-examination to
be relevant after all. In any such case the emphasis on
certain expressions and combinations of expressions in a
document, because they were used in searching, indicated
to a re-examining physicist a relevance previously
unnoticed (Swanson, a). A similar complication might be
involved when a user judges the value of "related" docu-
ments. The question concerning retrieval quality measure-
ment which this point raises is the following. How much
reflection and discussion should precede a user's final
judgment about the value of a retrieved document? For
the sake of further discussion, we shall assume we have
satisfactorily answered this question in some way.

In measuring the quality of output from some
retrieval system, do we want to measure it against some
absolute standard, or against the output of another
retrieval system for the same search questions?

Under some circumstances, comparative measurements
of two retrieval systems can be made by confining atten-
tion to the retrieved sets produced by the two systems.
For example, if retrieval questions are always for any

one S document, and system A always produces at least one
while system B often does not, then enough information is
available.

However even if attention can be confined to retrieved
sets, the results may not always favor A. In that case a
satisfactory general measure of the value of a retrieved set
is needed, and this raises some problems. For example, should
all search questions be rated equally important, all documents
of kind S be valued equally, all "related" documents welcomed
by the user be graded equally, how much penalty should be
assigned for irrelevant documents retrieved, etc.? In simple
general terms, how shall the retrieval of a document set for
a question be graded? (4)

Problems similar to those described in the preceding
paragraph arise for tests of a retrieval system against an
absolute standard rather than by comparison with another
retrieval system. They can also be summarized as follows: how
shall the retrieval of a document set for a question be
graded?

For some kinds of retrieval question it is necessary
(in either absolute or comparison tests) to know how many
unretrieved documents of kind S there are. This problem has
been dealt with in several retrieval tests by preparing
question-relevant document sets, "salting" the collection

with the, and then using the questions for searches (Cleverdon, Mooers, Swanson a, Fels). This procedure has the disadvantage that the retrieval questions are artificial, and it is uncertain how important the differences might be between questions arising from users' genuine information needs and questions formulated for an experiment. Thus while the results of such experiments are of interest, it is not clear how much we can generalize from them.

The problem of determining which relevant documents were not retrieved by a retrieval system under test might be dealt with satisfactorily by an approach which would permit real search questions from real users to be employed in testing. A group of subject specialists cooperate to cover a collection better than does any usual retrieval system. This coverage might be a relatively slight extension of their usual work. On a very part-time basis they then indicate documents relevant to real retrieval questions which are used for the test. (Apparently this method has not yet been tried.)

A retrieval test which uses the method of either of the preceding two paragraphs encounters the measure definition problems described earlier for tests which only use retrieved

sets,and other measure problems of a similar kind. How should
the document set retrieved for a question be graded, and
how should the non-retrieval of the rest of the collection
be scored?

In summary, attempts to determine the quality of
indexing by measuring retrieval quality encounter problems
presented by the role of search aids, evaluation of documents,
and evaluation of sets of documents, and may also face the
problem of identifying relevant documents not retrieved by
the system under test (5).

6. Indexing Duplication Studies

One alternative way of investigating mechanized
indexing methods empirically might be called an "indexing
duplication" study (O'Connor, Maron). Such a study is
probably much less expensive than retrieval testing.

An indexing duplication investigation is done in
the following way. Select a well-reputed retrieval system

(admittedly a crude judgment), for which the subject
indexing has been humanly done. For each indexing term
T, try to find a computer rule which will assign T to
just those documents assigned T by the human indexers.
(Since a few "false drops" are usually regarded as less
undesirable than a few relevant documents not retrieved,
a bit of overassigning of T by a computer rule might be
permitted). Since indexers occasionally err, important
cases should be double-checked with someone familiar with
the system's indexing.

An indexing duplication investigation should not
be called a test of mechanized indexing methods. For
one can always ask: how good, really, is the human
indexing being used as a standard? Such a study should
rather be thought of as a method of investigation. It
can provide specific empirical material suggesting
hypotheses about the interrelationships of subjects
occurring in a document (not yet a completely well-
defined concept), human indexing of the document, and
mechanized indexing of the document by various methods.
Results of such a study will be reported in other papers
(for some preliminary results, see O'Connor).

7. Postscript

A detached observer of our culture might be puzzled by our attempts to use machines for intellectual work, such as subject indexing or language translation. He might suggest that perhaps it would be more reliable, quicker, and less expensive to educate many more people to do intellectual work, and develop more machines for menial jobs. Being detached, he wouldn't suggest that this would also be more humane. He might elaborate what he did say by remarking that we appear to have forgotten that we know a great deal more about "programming" people successfully to do intellectual work than we know about programming machines for such purposes.

It is not inconsistent with the viewpoint of the preceding paragraph (which is mine except that I'm not detached) to be interested in mechanized indexing. The subject is of pure scientific interest. Further, if we continue to waste human intellectual potential, any successful methods of computer indexing will be quite valuable. And in a longer view, subject indexing is relatively onerous compared to some other intellectual work, therefore its successful mechanization would be another transfer of drudgery from people to machines.

NOTES

(1)   The references to the bibliography and the bibliography itself are intended to be selective rather than exhaustive.

(2)   Garfield, Eugene, private communication.

(3)   Sher, Irving, private communication.

(4)   This paragraph has been much helped by discussion with Russel Kirsch and Phyllis Baxendale.

(5)   This whole section has been much helped by discussions with Lea Bohnert.

BIBLIOGRAPHY

BAR-HILLEL, Y.

   (a)   The Mechanization of Literature Searching , paper 4-8
in The Mechanization of Thought Processes, Natl. Physical Lab.,
Symp. No. 10, Her Majesty's Stationery Office, 1959.

   (b)   Some Theoretical Aspects of the Mechanization of
Literature Searching , Hebrew Univ., Jerusalem, 1960. Astia
No. AD 236772.

BAXENDALE, PHYLLIS.   Machine Made Index for Technical
      Literature - an Experiment , IBM J. of Res. and Dev.,
      Vol. 2, No. 4. Oct., 1958, p. 354.

BERNIER, CHARLES. Correlative Indexes VI: Serendipity,
      Suggestiveness, and Display. American Documentation,
Vol. 11, No. 4, Oct., 1960, pp. 277-287.

BORKO, H. andM.D. BERNICK. Automatic Document Classification.
      J. Assoc. for Computing Machinery, Vol. 10, No. 2,
      April, 1963, pp. 151-162.

CLEVERDON, CYRIL. Report on the Testing and Analysis of an
      Investigation into the Comparative Efficiency of Indexing
Systems, Aslib. Cranfield Research Project, Cranfield,
England, 1962.

FELS, E.M. Evaluation of the Performance of an Information
      Retrieval System by Modified Mooers Plan. American Documentation,
Vol. 14, No. 1, Ja., 1963, pp. 28-34.

GIULIANO, V.E. & P.E. JONES. Linear Associative Information
      Retrieval, Arthur D. Little, Cambridge, Mass., 1962.
(This report also appears in the book Vistas in Information
Handling.)

HARRIS, Z.S.   Linguistic Transformations for Information
      Retrieval, International Conference on Scientific Information,
National Academy of Sciences, Washington, 1959, Vol. 2, p. 937.

KEHL, W.B. & J.F. HORTY, et al. An Information Retrieval
      Language for Legal Studies, Comm ACM, Vol. 4, No. 9,
Sept., 1961, p. 380.

KESSLER, MM. Bibliographic Coupling between Scientific Papers.
      Reps. R1 ff. M.I.T. Libraries, Nov., 1961 ff.

KUNO, SUSUMU & ANTHONY G. OETTINGER. Multiple-Path Syntactic
    Analyzer, in pre-prints (and Proceedings) of the Munich
1962 Conference of the International Federation of Information
Processing Societies.

LUHN, H.P.

   (a) A Statistical Approach to Mechanized Encoding and
Searching of Literary Information, IBM Journal of Research
and Development, Vol. 1, No. 4, Oct., 1957.

   (b) Potentialities of Auto-Encoding of Scientific
Literature. IBM Research Center, Yorktown Heights, N.Y. 1959.

   (c) The Automatic Creation of Literature Abstracts.
IBM J. of Research and Development, Vol. 2, No. 2, April 1958,
pp. 159-165.

MARON, M.E. & J.L. KUHNS. On Relevance, Probabilistic
    Indexing and Information Retrieval. J. of ACM, Vol. 7, No. 3,
July, 1960, pp. 216-244.

MARON, M.E. Automatic Indexing: and Experimental Inquiry,
    Journ. ACM, Vol. 8, No. 3, July, 1961.

MONTGOMERY, CHRISTINE & DON R. SWANSON. Machinelike Indexing
    by People. American Documentation, Vol. 13, No. 4, Oct., 1962,
pp. 359-366.

NEEDHAM, ROGER. A Method for Using Computers in Information
    Classification. Pre-prints (and Proceedings) of the
Munich 1962 Conference of the International Federation of
Information Processing Societies.

O'CONNOR , J. Some Remarks on Mechanized Indexing and Some
    Small-Scale Empirical Results, Machine Indexing, Amer. U.,
1962, pp. 2662-279. A longer report, "Mechanized Indexing"
is Astia No. AD250209, and is available from the author.

OSWALD, V.A. Jr., et al. Automatic Indexing and Abstracting
    of the Contents of Documents RADC-TR-59-208, 31 Oct., 1959.
Summarized in Edmundson, H.P. and R.E. Wyllys, "Automatic
Abstracting and Indexing", Comm ACM, Vol. 4, No. 5, May, 1961.

SALTON, GERARD. Some Experiments in the Generation of Word
    and Document Associations. 1962 Fall Joint Computer
Conference, Spartan Books, Washington , D.C. 1962, pp. 234-250.

STILES, H.E. The Association Factor in Information Retrieval
    ACM, Vol. 8, No. 2, April, 1961, pp 271-279.

STORM, E.   Some Experimental Procedures for the Identification
    of Information Content, Information Storage and Retrieval,
Sci. Rep. No. 1 SR-1, Computation Laboratory, Harvard University,
Nov., 1961, pp. I-1 to I-34.

SWANSON, DON R.

   (a) Searching Natural Language Text by Computer, Science,
Vol. 132, No. 34, Oct. 21, 1960, p. 1099.

   (b) Interrogating a Computer in Natural Language.
Pre-prints (and Proceedings) of the Munich 1962 Conference
of the International Federation of Information Processing
Societies.

   (c) Research Procedures for Automatic Indexing. Machine
Indexing, American University, 1962, pp. 281-304.

ASSISTANT SEC. OF DEF. FOR RES. AND ENG.
INFORMATION OFFICE LIBRARY BRANCH
PENTAGON BUILDING
WASHINGTON 25, D. C.   (2 COPIES)

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA   (10)

CHIEF OF NAVAL RESEARCH
DEPARTMENT OF THE NAVY
WASHINGTON 25, D. C.
ATTN. CODE 437 INFORMATION SYSTEMS BRANCH
(2 COPIES)

CHIEF OF NAVAL OPERATIONS
OP-07T-12
NAVY DEPARTMENT"
WASHINGTON 25, D. C.

DIRECTOR, NAVAL RESEARCH LABORATORY
TECHNICAL INFORMATION OFFICER /CODE 2000/
WASHINGTON 25, D. C. (6 copies)

COMMANDING OFFICER, OFFICE OF NAVAL RESEARCH
NAVY #100, FLEET POST OFFICE
NEW YORK 13, NEW YORK   (10 COPIES)

COMMANDING OFFICER, ONR BRANCH OFFICE
346 BROADWAY
NEW YORK 13, NEW YORK

COMMANDING OFFICER, ONR BRANCH OFFICE
495 SUMMER STREET
BOSTON 10, MASSACHUSETTS

BUREAU OF SHIPS
DEPARTMENT OF THE NAVY
WASHINGTON 25,DC
ATTN CODE 607A NTDS

BUREAU OF NAVAL WEAPONS
DEPARTMENT OF THE NAVY
WASHINGTON 25, D. C.
ATTN RAAV AVIONICS DIVISION

BUREAU OF NAVAL WEAPONS
DEPARTMENT OF THE NAVY
WASHINGTON 25, D. C.
ATTN RMWC MISSILE WEAPONS CONTROL DIV.

BUREAU OF NAVAL WEAPONS
DEPARTMENT OF THE NAVY
WASHINGTON 25, D. C.
ATTN. RUDC ASW DETECTION
 AND CONTROL DIVISION

BUREAU OF SHIPS
DEPARTMENT OF THE NAVY
WASHINGTON 25, D. C.
ATTN COMMUNICATIONS BRANCH CODE 686

NAVAL ORDNANCE LABORATORY
WHITE OAKS
SILVER SPRING 19, MARYLAND
ATTN TECHNICAL LIBRARY

DAVID TAYLOR MODEL BASIN
WASHINGTON 7, D. C.
ATTN TECHNICAL LIBRARY

NAVAL ELECTRONICS LABORATORY
SAN DIEGO 52, CALIFORNIA
ATTN TECHNICAL LIBRARY

UNIVERSITY OF ILLINOIS
CONTROL SYSTEMS LABORATORY
URBANA, ILLINOIS
ATTN. D. ALPERT

AIR FORCE CAMBRIDGE RES. LABORATORIES
LAURENCE C. HANSCOM FIELD
BEDFORD, MASSACHUSETTS
ATTN RESEARCH LIBRARY, CRX2-R

TECHNICAL INFORMATION OFFICER
US ARMY SIGNAL RESEARCH & DEV. LAB.
FORT MONMOUTH, NEW JERSEY
ATTN DATA EQUIPMENT BRANCH

COMMANDING OFFICER
U.S. ARMY RESEARCH OFFICE/DURHAM/
BOX CM, DUKE STATION
DURHAM, NORTH CAROLINA
ATTN CRD-AA-IP

NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE, MARYLAND
ATTN R-4, HOWARD CAMPAIGNE

U. S. NAVAL WEAPONS LABORATORY
DAHLGREN, VIRGINIA
ATTN. HEAD, COMPUTION DIV.,G.H.GLEISSNER

NATIONAL BUREAU OF STANDARDS
DATA PROCESSING SYSTEMS DIVISION
ROOM 239 BLDG. 10
ATTN A.K.SMILOW
WASHINGTON 25, dD.C.

ABERDEEN PROVING GROUND, BRL
ABERDEEN PROVING GROUND, MARYLAND
ATTN. J. H. GIESE, CHIEF COMPUTION
          LABORATORY

OFFICE OF NAVAL RESEARCH
RESIDENT REPRESENTATIVE
UNIVERSITY OF PENNSYLVANIA
3438 WALNUT STREET
PHILADELPHIA 4, Penna.

COMMANDING OFFICER
ONR BRANCH OFFICE
JOHN CRERAR LIBRARY BLDG.
86 EAST RANDOLPH STREET
CHICAGO L, ILLINOIS

COMMANDING OFFICER
ONR BRANCH OFFICE
1030 E. GREEN STREET
PASADENA, CALIFORNIA

COMMANDING OFFICER
ONR BRANCH OFFICE
1000 GEARY STREET
SAN FRANCISCO 9, CALIFORNIA

NATIONAL BUREAU OF STANDARDS
WASHINGTON 25, D. C.
ATTN MR. R.D. ELBOURN

DYNAMIC ANALYSIS AND CONTROL LAB.
MASSACHUSETTS INST. OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS
ATTN. D. W. BAUMANN

SYRACUSE UNIVERSITY
ELECTRICAL ENG. DPT.
SYRACUSE 10, NEW YORK
ATTN. DR. STANFORD GOLDMAN

CORNELL UNIVERSITY
COGNITIVE SYSTEMS RES. PROGRAM
HOLLISTER HALL
ITHACA, NEW YORK
ATTN. DR. FRANK ROSENBLATT

NASA
GODDARD SPACE FLIGHT CENTER
WASHINGTON 25, D. C.
ATTN ARTHUR SHAPIRO

NATIONAL PHYSICAL LABORATORY
TEDDINGTON, MIDDLESEX
ENGLAND
ATTN DR. A.M. UTTLEY, SUPERINTENDANT
AUTONOMICS DIVISION

UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN
ATTN DEPT OF PSYCHOLOGY
DR. JAMES OLDS

DIAMOND ORDNANCE FUZE LABORATORY
CONNECTICUT AVE. & VAN NESS ST.
WASHINGTON 25, D. C.
ORDTL-o12, E. W. Channel

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
ATTN SCHOOL OF APPLIED SCIENCE.
DEAN HARVEY BROOK

COMMANDING OFFICER AND DIRECTOR
U. S. NAVAL TRAINING DEVICE CENTER
PORT WASHINGTON
LONG ISLAND, NEW YORK
ATTN TECHNICAL LIBRARY

U.S. ARMY BIOLOGICAL WARFARE LABS.
BLD. 29, ROOM 516
BETHESDA 14, MD.
ATTN CLIFFORD J.MALONEY
DIVISION OF BIOLOGICS STDS

LABORATORY FOR ELECTRONICS, INC.
1079 COMMONWEALTH AVE.
BOSTON 15, MASSACHUSETTS
ATTN DR. H. FULLER

STANFORD RESEARCH INSTITUTE
COMPUTER LABORATORY
MENLO PARK, CALIFORNIA
ATTN. H. D. CRANE

THE RAND CORP.
1700 MAIN STREET
SANTA MONICA, CALIFORNIA
ATTN NUMERICAL ANALYSIS DEPT.
WILLIS H. WARE

COMMUNICATIONS SCIENCES LAB
UNIVERSITY OF MICHIGAN
180 FRIEZE BUILDING
ANN ARBOR, MICHIGAN
ATTN GORDON E. PETERSON

NATIONAL SCIENCE FOUNDATION
PROGRAM DIR. FOR DOCUMENTATION RES
WASHINGTON 25, D. C.
ATTN. HELEN L. BROWNSON

WAYNE STATE UNIVERSITY
DETROIT, MICHIGAN
ATTN. DEPT. OF SLAVIC LANGUAGES
PROF. HARRY H. JOSSELSON

UNIV. OF CALIFORNIA -LA
LOS ANGELES 24, CALIFORNIA
ATTN. DEPT. OF ENGINEERING
PROF. GERALD ESTRIN

HEBREW UNIVERSITY
JERUSALEM ISRAEL
ATTN PROF. Y. BAR-HILLEL

MASS. INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS
ATTN. PROF. N. CHOMSKI.

UNIVERSITY OF ILLINOIS
ATTN JOHN R. PASTA
CHAMPAIGN, URBANA ILLINOIS

NAVAL RESEARCH LABORATORY
WASHINGTON 25, D. C.
ATTN. SECURITY SYSTEMS
CODE 5266, MR. G. ABRAHAM

ZATOR COMPANY
140 1/2 MT. AUBURN
CAMBRIDGE 38, MASSACHUSETTS
ATTN. R. J. SOLOMONOFF

TELECOMPUTING CORPORATION
12838 SATICOY STREET
NORTH HOLLYWOOD, CALIFORNIA
DATA INSTRUMENTS DIVISION
FIELD ENGINEERING DEPT.

MASSACHUSETTS INST. OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS
ATTN. PROF. V. H. YNGVE

BELL TELEPHONE LABORATORIES
MURRAY HILL LABORATORY
MURRAY HILL, NEW JERSEY
ATTN DR. EDWARD F. MOORE

BIOLOGICAL ABSTRACTS
3815 WALNUT STREET
PHILADELPHIA 4, Penna.
ATTN. G. MILES CONRAD

UNIVERSITY OF SASKATCHEWAN
DEPT OF ELECTRICAL ENG.
SASKATOON, SASK. CANADA
ATTN DR. ANDREW D. BOOTH

.CAMBRIDGE LANGUAGE RES. UNIT
20 MILLINGTON ROAD
CAMBRIDGE, ENGLAND
ATTN. MRS. MARGARET M. BRAITHWAITE

COLLEGE OF AERONAUTICS
CRANFIELD, BLETCHLEY
BUCKS, ENGLAND
ATTN. DR. CYRIL W. CLEVERDON
LIBRARIAN

COLUMBIA UNIVERSITY
NEW YORK, NEW YORK
ATTN. DR. MAURICE F. TAUBER
PROF. OF LIBRARY SERVICE

COMPUTER CONTROL COMPANY, INC.
983 CONCORD STREET
FRAMINGHAM, MASSACHUSETTS
ATTN BEN KESSEL, PRESIDENT

COMPUTER USAGE COMPANY
655 MADISON AVENUE
NEW YORK 17, NEW YORK
ATTN. DR. ASCHER OPLER

COUNCIL ON LIBRARY RESOURCES
1025 CONNECTICUT AVE. N. W.
WASHINGTON 8, D.C.
ATTN. DR. LAURENCE B. HEILPRIN

DOCUMENTATION INCORPORATED
7900 NORFOLK AVENUE
BETHESDA 14, MARYLAND
ATTN DR. MORTIMER TAUBE

ZATOR COMPANY
140 1/2 AUBURN STREET
CAMBRIDGE 38 MASS.
ATTN CALVIN N. MOOERS

CARNEGIE INSTITUTE OF TECHNOLOGY
PITTSBURGH, PENNSYLVANIA
ATTN. DIRECTOR, COMPUTATION CENTER
ALAN J. PERLIS

ROME AIR DEVELOPMENT CENTER, RCOR
DCS/OPERATIONS, USAF
GRIFFISS AIR FORCE BASE, NEW YORK
ATTN IRVING J. GABELMAN

AIR FORCE OFFICE OF SCIENTIFIC RES.
DIRECTORATE OF INFORMATION SCIENCES
WASHINGTON 25, D. C.
ATTN. DR. HAROLD WOOSTER

HUNTER COLLEGE
NEW YORK 21, New York
ATTN DEAN MINA REES

RADIO CORPORATION OF AMERICA
306/2
DATA SYSTEMS DIVISION
8500 BALBOA BLVD.
VAN NUYS, CALIFORNIA
ATTN. JOSEPH E. KARROLL

UNITED STATES AIR FORCE
AIR UNIVERSITY LIBRARY
MAXWELL AIR FORCE BASE, ALABAMA
ATTN AUL -9539

MR. SIDNEY KAPLAN
1814 GLEN PARK AVENUE
SILVER SPRING, MD.

NATIONAL BUREAU OF STANDARDS
WASHINGTON 25, D.C.
ATTN MISS IDA RHODES, 229 STUCCO BLDG.

DEPARTMENT OF THE ARMY
OFF. OF THE CHIEF OF RES. & DEVEL.
PENTAGON, ROOM 3D442
WASHINGTON 25, D. C.
ATTN MR. L. H. GEIGER

COUNCIL ON LIBRARY RESOURCES
SUITE 704, 1025 CONNECTICUT AVE. N.W
WASHINGTON 6, D. C.
ATTN. MR. VERNER CLAPP, PRESIDENT

DOCUMENTATION INCORPORATED
4827 RUGBY AVENUE
BETHESDA, MARYLAND
ATTN. MR. EUGENE WALL

DR. GEORGE MALCOLM DYSON
CHEMICAL ABSTRACTS
OHIO STATE UNIVERSITY
COLUMBUS 10, OHIO

ESSO RESEARCH & ENGINEERING CO.
TECHNICAL INFORMATION DIVISION
P. O. BOX 51
LINDEN, NEW JERSEY
ATTN MR. B. H. WEIL

ROYAL AIRCRAFT ESTABLISHMENT
MATHEMATICS DEPT.
FARNBOROUGH, HAMPSHIRE, ENGLAND
ATTN MR. R. A. FAIRTHORNE
MINISTER OF AVIATION

INSTITUTE FOR SCIENTIFIC INFORMATION
33 SOUTH SEVENTEENTH STREET
PHILADELPHIA 3, PENNSYLVANIA
ATTN DR. EUGENE GARFIELD

GMELIN INSTITUTE
VARRENTRAPPSTRASSE 40/42
FRANKFURT/MAIN, GERMANY
ATTN G. MULERT

HARVARD COMPUTATION LAB.
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
ATTN. DR. ANTHONY OETTINGER

HERCULES POWDER COMPANY
HERCULES RESEARCH CENTER
WILMINGTON 99, DELAWARE
ATTN MR. HERMAN SKOLNIK

ESSO RESEARCH AND ENGINEERING CO.
P. O. BOX 51
LINDEN,NEW JERSEY
ATTN MR. GERALD JAHODA

HERNER AND COMPANY
906 NEW HAMPSHIRE AVENUE
WASHINGTON 7, D. C.
ATTN MR. SAUL HERNER, PRESIDENT

IBM RESEARCH
YORKTOWN HEIGHTS, NEW YORK
ATTN HANS PETER LUHN

IBM RESEARCH CENTER
BOX 218
YORKTOWN HEIGHTS, NEW YORK"
ATTN DR. GILBERT KING

INSTITUTO CARLO ERBA
PER RICERCHE TERAPEUTICHE
VIA IMBONATI, 24-TELEGONO 60,05,23
MILANO, ITALIA
ATTN DR. F. WEISS
DOCUMENTATION SECTION

ITEK CORPORATION
45 THOREAU STREET
ATTN MISS THYLLIS M. WILLIAMS
CONCORD, MASS.

ITEK CORPORATION
INFORMATION SCIENCES LAB.
LEXINGTON 73, MASS.
ATTN JOHN W. KUIPERS

LIBRARY OF CONGRESS
SUBJECT CATALOGING DIVISION
WASHINGTON, D. C.
ATTN RICHARD S. ANGELL

LIBRARY OF CONGRESS
WASHINGTON 25, D. C.
ATTN JOHN SHERROD

LINDE COMPANY
P. O. BOX 44
TONAWANDA, NEW YORK
ATTN MR. S. M. COHEN

MASSACHUSETTS INST. OF TECHNOLOGY
CAMBRIDGE 39, MASSACHUSETTS
ATTN LIBRARY

U.S. PATENT OFFICE
RESEARCH AND DEVELOPMENT
WASHINGTON, D. C.
ATTN. R. A. SPENCER

PURDUE UNIVERSITY
SCHOOL OF MECHANICAL ENGINEERING
LAFAYETTE, INDIANA
ATTN DR. Y. S. TOULOUKIAN

THOMPSON RAMO WOOLRIDGE INC.
8433 FALLBROOK AVENUE
CANOGA PARK, CALIFORNIA
ATTN D. R. SWANSON

RAND CORPORATION
1700 MAIN STREET
SANTA MONICA, CALIFORNIA
ATTN LIBRARY

MRS. LEA M. BOHNERT
1500 MASS. AVE. N. W.
WASHINGTON 5, D. C.

UNIVAC DIV. OF SPERRY RAND
P. O. BOX 500
BLUE BELL, PENNSYLVANIA
ATTN. T. H. BONN
CHIEF ENG. RES. & DIV.

RUTGERS UNIVERSITY
5 HUNTINGTON STREET
NEW BRUNSWICK, NEW JERSEY
ATTN DR. RALPH R. SHAW

NATIONAL SECURITY AGENCY
DEPUTY CHIEF, OFFICE OF CENTRAL REF.
FORT MEADE, MARYLAND
ATTN C-3, DR. J. ALBERT SANFORD

STANFORD RESEARCH INSTITUTE
MENLO PARK, CALIFORNIA
ATTN MR. CHARLES BOURNE

UNIVERSITY OF PENNSYLVANIA
DEPT. OF LINGUISTICS
PHILADELPHIA 4, PENNSYLVANIA
ATTN. DR. ZELLIG HARRIS

UNIVERSITY OF PENNSYLVANIA
MOORE SCHOOL OF ELEC. ENG.
200 SOUTH 33rd STREET
PHILADELPHIA 4, PENNSYLVANIA
ATTN. MISS ANNA LOUISE CAMPION

MAUCHLY ASSOCIATES, INC.
INDIANA AVENUE
FORT WASH INDUSTRIAL PARK
ATTN DR. JOHN MAUCHLY
FORT WASHINGTON, PENNA.

NATIONAL BIOMEDICAL RES. FOUND. INC.
8600 16th STREET SUITE 310
SILVER SPRING, MARYLAND
ATTN DR. R. S. LEDLEY

NATIONAL BUREAU OF STANDARDS
WASHINGTON 25, D. C.
ATTN MRS. ETHEL MARDEN

NATIONAL BUREAU OF STANDARDS
OFF. OF BASIC INSTRUMENTATION
WASHINGTON 25, D. C.
ATTN DR. JOSHUA STERN

NATIONAL BUREAU OF STANDARDS
APPLICATIONS ENGINEERING SECTION
WASHINGTON 25, D. C.
ATTN MISS MARY E. STEVENS

NATIONAL LIBRARY OF MEDICINE
ROOM M. 107
8600 WISCONSIN AVENUE
BETHESDA 14, MARYLAND
ATTN MR. SEYMOUR I. TAINE
CHIEF BIBLOGRAPHIC SERVICES DIV.

NATIONAL SCIENCE FOUNDATION
1951 CONSTITUTION AVENUE, N. W.
WASHINGTON 25, D. C.
ATTN RICHARD SEE

LABORATOIRE PSYCHOLOGIE SOCIALE
2 RUE CHAPON
PARIS 3C FRANCE
ATTN ROBERT PAGES
PARIS 4, FRANCE

DEPARTMENT OF COMMERCE
U. S. PATENT OFFICE"
WASHINGTON 25, D. C.
ATTN MR. HERBERT R. KOLLER

DEPARTMENT OF COMMERCE
U, S. PATENT OFFICE
WASHINGTON 25, D. C.
ATTN MR. HAROLD PFEFFER

MONSANTO CHEMICAL COMPANY
800 N. LINDBERG
ST. LOUIS 66 ,MISSOURI
ATTN DR. W. H. WALDO

WESTERN RESERVE UNIVERSITY
CENTER FOR DOCUMENTATION
AND COMMUNICATION RESEARCH
CLEVELAND 6, OHIO
ATTN ALLEN KENT

CHIEF OF NAVAL OPERATIONS-OP-923 M4
DEPARTMENT OF THE NAVY
WASHINGTON 25, D. C.
ATTN KOINES/MUROS

DEPARTMENT OF THE ARMY
OFFICE OF THE ASST. COFD FOR INTELLIGENCE
ROOM 2B529, PENTAGON
WASHINGTON, D. C.
ATTN JOHN F. KULLGREN

NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE, MARYLAND
ATTN MR. IGNATIUS G. MATTINGLY, R44

DIVISION OF AUTOMATIAL DATA PROC. AOP
DEPARTMENT OF STATE
WASHINGTON 25, D. C.
ATTN F. P. DIBLAST
19A16

MR. PAUL W. HOWERTON
ROOM 1053 M. BLDG.
CODE 163
CIA
WASHINGTON, D. C.

DR. HENRY HIZ
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA 4, PENNSYLVANIA

UNIVERSITY OF PENNSYLVANIA
MECHANICAL LANGUAGES PROJECTS
MOORE SCHOOL OF ELECT. ENG.
PHILADELPHIA 4, PENNA.
ATTN DR. SAUL GORN, DIRECTOR

DR. VINCENT GIULIANO
ARTHUR D. LITTLE, INC.
ACORN PARK
CAMBRIDGE 40, MASS.

PROF. WESLEY SIMONTON
LIBRARY SCHOOL
UNIVERSITY OF MINNESOTA
MINEAPOLIS 14, MINNESOTA

MR. BERNARD M. FRY, DEPUTY HEAD
OFFICE OF SCIENCE INFO. SERVICE
NATIONAL SCIENCE FOUNDATION
1951 CONSTITUTION AVENUE, N.W.
WASHINGTON 25, D. C.

INSTITUTE FOR SCIENTIFIC INFO
33 SOUTH SEVENTEENTH STREET
PHILADELPHIA 3, PENNA.
ATTN DR. IRVING H. SHER

INTERNATIONAL BUSINESS MACHINES
ADVANCED SYSTEMS DEVELOPMENT DIV
SAN JOSE 14, CALIF.
ATTN I. A. WARHEIT

DR. KARL F. HEUMANN, DIRECTOR
OFFICE OF DOCUMENTATION
NATIONAL ACADEMY OF SCIENCES
WASHINGTON 25, D. C.

APPLIED PHYSICS LABORATORY
JOHNS HOPKINS UNIVERSITY
8261 GEORGIA AVENUE
SILVER SPRING, MARYLAND
ATTN DOCUMENT LIBRARY

BUREAU OF SUPPLIES AND ACCOUNTS
CHIEF, NAVY DEPARTMENT
WASHINGTON, D. C.
ATTN CODE W3

DR. NOAH S. PRYWES
MOORE SCHOOL OF ENGINEERING
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA 4, PENNA.

NATIONAL AERONAUTICS & SPACE ADM
GODDARD SPACE FLIGHT CENTER"
GREENBELT, MARYLAND
ATTN CHIEF, DATA SYSTEMS DIV.
C.V.L. SMITH

RABINOW ENGINEERING CO. INC."
7712 NEW HAMPSHIRE AVENUE
WASHINGTON 12, D. C.

PARMELY C. DANIELS, RES. SUPPORT
ARMY RESEARCH OFFICE _ OCRD
DEPT. ARMY PENTAGON
WASHINGTON 25, D. C.

MR. M. E. MARON
RAND CORPORATION
SANTA MONICA, CALIF.

INFORMATION UNIT HEAD
SCIENCE INFORMATION DEPT.
SMITH, KLINE AND FRENCH
1500 SPRING GARDEN ST.
PHILADELPHIA, PENNA.

ROBERT M. MASON
CODE 4558
NAVAL RESEARCH LABORATORY
WASHINGTON 25, D. C.

STANLEY ROSEN
GENERAL ELECTRIC
INFORMATION SYSTEMS SECTION
4901 FAIRMONT AVENUE
BETHESDA, MARYLAND (THREE COPIES)

A.S. LOE
SYSTEMS RESEARCH DIV., CODE W 4
BUREAU OF SUPPLIES AND ACCOUNTS
WASHINGTON 25, D. C.

MR. A. BODIAN
NAVY DEPARTMENT
OPNAV -OP-922
THE PENTAGON
WASHINGTON 25, D. C.

GEORGE C. FRANCIS
COMPUTING LAB., BRL
ABERDEEN PROVING GROUND
MARYLAND

MR. L. L. KOFFMAN
ARMY MAP SERVICE
5500 BROOKS LANE
WASHINGTON 25, D. C.

DR. PAUL GARVIN, THOMPSON
RAMO-WOOLDRIDGE-INC
8433 FALLBROOK AVE.
CANOGA PARK, CALIF.

NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE,
MARYLAND
ATTN R.-42, R. WIGGINGTON

FEDERAL AVIATION AGENCY
BUREAU OF RES. AND DEVELOPMENT
WASHINGTON 25, D.C.
ATTN RD-375 Mr. Harry Hayman

MCDONNEL AIRCRAFT CORP.
MISSILE ENGINEERING DIVISION
MUNICIPAL AIRPORT
ST. LOUIS, MISSOURI
ATTN MANAGER ADV. DIGITAL TECH.
R. E. ACKER

INSTITUTE FOR SPACE STUDIES
475 RIVERSIDE DRIVE
NEW YORK 27, NEW YORK
ATTN MR. ALBERT ARKING

DAVID TAYLOR MODEL BASIN
WASHINGTON 7, D. C.
ATTN AERODYNAMICS LAB. CODE 628
MISS CRAVENS

LEXICAL DATA HANDLING SECT. RCWIP
HQ ROME AIR DEVELOPMENT CENTER
UNITED STATES AIR FORCE
GIRFFISS AIR FORCE BASE, N.Y.
ATTN OSKAR REINSON
STAFF LINGUIST

LINCOLN LABORATORY
MASSACHUSETTES INSTITUTE OF TECH
LEXINGTON 73, MASSACHUSETTS
ATTN LIBRARY

AMERICAN UNIVERSITY
SCHOOL OF GOVERNMENT
1901 F. STREET N.W.
WASHINGTON 6, D.C.
ATTN LOWELL H. HATTERY

PROF. J. D. BERNAL
DEAN, BIRKBECK COLLEGE
MALET STREET
LONDON WC 1, ENGLAND

DR. HERBERT SIMON
CARNEGIE INST. OF TECHNOLOGY
PITTSBURGH, PENNA.

MR. TERRY SAVAGE
IBM ASDD P, O. BOX 344
YORKTOWN HEIGHTS, N. Y.

MRS. PAULINE GREEN, LIBRARIAN
GENERAL ELECTRIC COMPANY
DEFENSE SYSTEMS DEPART.
INFORMATION SYSTEMS SECTION
4901 FAIRMONT AVENUE
WASHINGTON 14, D. C.

SCIENTIFIC ATTACHE
SWEDISH EMBASSY
2242 R STREET, N.W.
WASHINGTON 8, D. C.

MR. JOHN ROBB
BUILDING 82
R.C.A. 6901 N. CRESCENT BLVD.
PENNSAUKEN, N. J.

COLUMBIA UNIVERSITY, LIBRARY
ACQUISITIONS DEPT.
535 WEST 114th STREET
NEW YORK 27, N.Y.

MR. JAMES DECKER MACK
LIBRARIAN
LEHIGH UNIVERSITY
BETHLEHEM, PENNA.

DR. ERIC DE CROLLIER
CENTRE FRANCAIS DECHANGES ET DE
DOCUMENTATION TECHNIQUES
32, CORSO MAGNETA
MILAN, ITALY

DR. C. J. DE HANN
NETHERLANDS PATENT OFFICE
S GRAVENHAGE
WILLEN WITSENPLEIN
THE HAGUE, NETHERLANDS

M. J. E. L. FARRADINE
TORRIN
CROFTON ROAD
ORPINGTON, KENT, ENGLAND

HENRY J. DUBESTER
6531 ELGIN LANE
BETHESDA, MARYLAND

DR. R. M. HAYES
ADVANCED INFORMATION SYSTEMS INC.
3002 MIDVALE AVENUE
LOS ANGELES 34, CALIF.

MR. R. N. NEEDHAM
CAMBRIDGE LANG. RESEARCH UNIT
20 MILLINGTON ROAD
CAMBRIDGE, ENGLAND

MRS. MADELINE M. HENDERSON
NATIONAL SCIENCE FOUNDATION
1500 MASS. AVENUE N. W.
WASHINGTON 25, D. C.

DR. B. C. VICKERY
DEPT. OF SCIENTIFIC AND IND. RES.
NATIONAL LENDING LIBRARY FOR
SCIENCE AND TECHNOLOGY
BOSTON SPA.
YORKSHIRE, ENGLAND

DR. S. R. RANGANATHAN
774 SEVENTEENTH CROSS ROAD
BANGALORE 3, INDIA

DR. JACQUES SAMAIN
CHEF DE SERVICE CENTRE NATIONAL
DE IS RECHERCHE SCIENTIFIQUE
74 RUE DES SAINTS-PERES
PARIS 7, FRANCE

MRS. CLAIRE K. SCHULTZ
LINE LEXINGTON
PENNA

MR. JAMES W, PERRY
UNIVERSITY OF ARIZONA
TUCSON, ARIZONA

DR. CHARLES L. BERNIER
ARMED SERV. TECH. INFO. AGENCY
B. BLDG.
ARLINGTON HALL STATION
ARLINGTON 12, VA.

INSTITUTE FOR DEFENSE ANALYSIS
COMMUNICATIONS RES. DIVISION
VON NEUMANN HALL
PRINCETON, NEW JERSEY

SYLVANIA ELECTRIC PRODUCTS, INC.
40 BYLVAN ROAD
WALTHAM 54, MASS.
ATTN MR. GERALD COHEN

MR. KENNETH WEBB
FEDERAL SYSTEMS DIVISION
IBM
7220 WISCONSIN AVENUE
BETHESDA, MARYLAND

MR. A. J. NEUMANN
STAFF ENGINEER, INFO SERVICES
RAYTHEON COMPANY
LEXINGTON 73, MASSACHUSETTS.

ALLAN KIRON
OFFICE OF RESEARCH AND DEVEL.
U. S. DEPARTMENT OF COMMERCE
PATENT OFFICE
WASHINGTON 25,D.C.

AIR FORCE OFFICE OF SCIENTIFIC RES.
INFORMATION S RESEARCH DIV.
WASHINGTON 25, D. C.
ATTN R. W. SWANSON