STATISTICAL ASSOCIATION PROCEDURES FOR MESSAGE CONTENT ANALYSIS

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-63-159

April 1963

J. Spiegel  E. Bennett  E. Haines

Prepared for
OPERATIONAL APPLICATIONS LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Prepared by
THE MITRE CORPORATION
Bedford, Massachusetts
Contract AF33(600)-39852  Project 702
NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.
When US Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Do not return this copy. Retain or destroy.
STATISTICAL ASSOCIATION PROCEDURES
FOR MESSAGE CONTENT ANALYSIS

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-63-159

April 1963

J. Spiegel  E. Bennett  E. Haines

Prepared for
OPERATIONAL APPLICATIONS LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Prepared by
THE MITRE CORPORATION
Bedford, Massachusetts
Contract AF33(600)-39852  Project 702
STATISTICAL ASSOCIATION PROCEDURES FOR MESSAGE CONTENT ANALYSIS
MITRE SR-79 Information System Language Studies, Number 1

by


FOREWORD

This paper is substantially the same as "Document and Message Routing through Communication Content Analysis," presented at the International Federation for Information Processing Congress, Symposium on Optimum Routing in Large Networks, in Munich, 31 August 1962.
This is an introductory report of an investigation concerned with developing procedures for utilizing certain statistical properties of messages or documents; these properties to be used for message routing or retrieval. This approach applies the most elementary relation among the words making up a message that of word-word co-occurrence probability patterns. It is shown that any message material, be it natural language, code, or index terms, can be processed provided that the input is compatible with the input requirements of the computer.
# TABLE OF CONTENTS

## THE DEVELOPMENT OF A METHODOLOGY FOR STATISTICAL ASSOCIATION

- Approach to the Problem ............................. 1
- Previous Research on Statistical Association Techniques ...................... 4
- Preliminary Considerations for a Statistical Association Methodology ...... 7
- The Development of a Statistical Association Technique ..................... 12

## PROGRAMMING THE STATISTICAL ASSOCIATION TECHNIQUE

- Input Program ........................................ 19
- Frequency Matrix Formulation Program .................................. 23
- Statistical Language Information .......................................... 25
- Association Matrix Program ................................................ 25
- An Algorithm for Routing or Retrieval .................................... 27
- Conclusion .............................................. 34

## REFERENCES

- REFERENCES ........................................... 55
THE DEVELOPMENT OF A METHODOLOGY FOR STATISTICAL ASSOCIATION

Approach to the Problem

The work of this investigation is concerned with developing procedures for coding certain statistical properties of messages (or documents) contained within an information system, and then using these codes for message routing or retrieval. The statistical approach applies the most elementary relation among message units, that of co-occurrence probability patterns. The basic strategy is to proceed as far as possible with a minimum of assumptions about the linguistic or semantic organization of the information within the message structure.

This strategy implies a rather mechanistic approach to language processing and that is indeed the case. We assume the information contained in a message is carried by the words that make it up, and by the manner in which they are strung together. Further, we assume a person generating a message or document chooses words in a nonrandom fashion and combines them according to semantic and syntactic rules that are regular and, at least in our culture, to some extent predictable. That is, both the selection of elements and their co-occurrence with other elements are subject to restrictions by the contexts in which they occur. Evidence from information redundancy experiments indicates such an assumption is reasonable. We intend to exploit the regularities of these
sociations among words, ignoring the specific nature of the rules which produce such regularity and thereby restricting ourselves to coding and use of the resulting statistical features alone.

Before discussing specific coding and routing or retrieval techniques, it is clear certain ideals or requirements should be taken into account. First, any machine technique for coding messages should be able to accept and analyze large amounts of natural message content relating to a wide range of topics. In responding to retrieval search or routing demands, a routing technique should be able to draw upon its total resource of stored content coded information, not only to select an appropriate response, but more important, to improve its program for interpreting such demands and responding to them. Both techniques should be able to improve with experience, and the combined system of the two techniques should be able to accept and use program or procedural information in the same form as any other input, including data. The system should be able to code the content from messages in a fully mechanical manner. It also should be able to relate new content to other statistically relevant content already in memory. From its reservoir of information, it should be able to elicit the necessary clues as to which messages are relevant to each other, especially in response to a message which is also a query. For such a system to be reasonably adaptable, it also should be able to perform these functions without an index, grammar book, dictionary, thesaurus or other formal constraint.
In reviewing these requirements it is evident that we are defining an approach similar in many ways to the way humans appear to retrieve information from their own memories. Typically, humans seem to start with the query words and then associate these with other words until the information they seek is brought to their conscious attention. This process of association of elements is so basic and obvious that Aristotle reasoned that to learn was to associate. However, although association theory has been known for many years, little use has been made of it as a methodology for information processing. In fact, literature on the use of statistical associations for information processing is quite limited, although at least three significant contributions of a methodological nature appear to be of direct relevance. All are concerned with the use of index terms, from a specified library of index terms, to retrieve documents from a specified library of documents. All involve obtaining descriptive statistics to indicate the extent to which specific index terms occur together in tagging the various documents of the library. Such descriptive statistics then are used to expand from one or more index terms used in a query to a set of associated terms, based upon evidence of the co-occurrence tendencies of the various terms.

¹For an interesting example of one way humans retrieve, see Freud (1938, p. 57A). References are listed by name and date at the end of the report.
Previous Research on Statistical Association Techniques

Maron and Kuhns (1960) investigated statistical association techniques as part of a more general methodological attack on the problem of document retrieval. Starting with a catalog of index terms, and a library of documents, a statistical matrix of association frequencies is developed.

\[
\begin{array}{ccc}
& T_k & \bar{T}_k \\
T_j & x = N(T_j, T_k) & u = N(T_j, \bar{T}_k) & N(T_j) \\
\bar{T}_j & v = N(\bar{T}_j, T_k) & y = N(\bar{T}_j, \bar{T}_k) & N(\bar{T}_j) \\
N(T_k) & N(\bar{T}_k) & n
\end{array}
\]

where

- \( T_j \) is a tag in the original request.
- \( T_k \) is a tag not in the original request.
- \( N(T_j, T_k) \) = the number of documents in the library tagged jointly with both \( T_j \) and \( T_k \).
- \( N(T_j, \bar{T}_k) \) = the number of documents tagged with \( T_j \) and not with \( T_k \).
- \( N(T_j) \) = the total number of documents tagged with \( T_j \).
- \( N(\bar{T}_j) \) = the total number of documents not tagged with \( T_j \).
- \( n \) = the total number of documents, and so forth.
From these descriptive statistics, Maron and Kuhns develop three different measures of closeness of association for index terms. One is the conditional probability that if a term in the original request \( T \) is assigned to a document, then the additional term \( T_j \) also will be assigned:

\[
P(T_k | T_j) = \frac{N(T_j, T_k)}{N(T_j)}
\]  

(1)

The second measure is the inverse conditional probability; that is, the probability that if the additional term \( T_k \) is assigned to a document, then the original request term \( T_j \) also would be:

\[
P(T_j | T_k) = \frac{N(T_j, T_k)}{N(T_k)}
\]  

(2)

Finally they use the contingency estimate, or estimate of the frequency of co-occurrence, independent of the individual and separate influences of the two terms which form the co-occurrence in question. They remove the magnitude to be expected on the basis of chance from the actual cell magnitude, taking into account the number of times the individual tags are used.

\[
\delta(T_j, T_k) = N(T_jT_k) - \frac{N(T_j)N(T_k)}{n}
\]  

(3)

Maron and Kuhns then introduce an arbitrary coefficient of association, based upon \( \delta(T_j, T_k) \), which ranges conveniently from -1 to +1 with a magnitude of zero for the condition were \( \delta(T_j, T_k) \) is zero. This co-
efficient is of the form:

\[ Q(T_j, T_k) = \frac{n^2}{(xy + uv)} \] (4)

Stiles (1961) also starts with a contingency table of the form previously noted. However, he introduces a different coefficient of association:

\[ \log_{10} \frac{n \left( |n^2| - \frac{p^2}{2} \right)^2}{N(T_j) N(T_k) N(T_{jk}) N(T_{jk})} \] (5)

Doyle (1960) uses still a third measure also drawn from a contingency table, to indicate strength of association:

\[ \frac{N(T_j, T_k)n}{N(T_j) N(T_k)} \] (6)

In each of the three approaches cited, the investigators tend to adopt the same basic data structure from which to develop their analyses. They pass over the question of how many terms are used to index any particular document and start with the total population of indexed documents as a base. They divide this population of documents into those that exhibit the common property of having been indexed by \( T_j \), with and without \( T_k \), and those not indexed by \( T_j \), with and without \( T_k \). Using various normalizing procedures, they adjust the sizes of these various groups, especially the group \( T_j, T_k \), to remove any effect that might result from the tendencies of \( T_j \) and \( T_k \), separately, to occur frequently in general. Some kind of normalization is required, because the more frequently an index word occurs, the more likely it will co-occur with some other term, simply on the basis of chance. The techniques used by Maron and Kuhns,
Stiles and Doyle, however, do not treat the fact that the more lengthy the string of index words used to index a document, the more likely that co-occurrences involving the terms in the string are due to chance.

For a library retrieval problem this might be little more than a minor omission, if, for example, the number of terms used to index all documents is a constant. However, if data on statistical co-occurrence are drawn from the actual strings of words in natural language that comprise the body of a document or message, then such factors as string length, word position in the string, and vocabulary size might significantly influence the tendency of words to co-occur. Accordingly, we would like to argue that a statistical association technique should take into account such factors and, further, that it should not be dependent upon the particular level of message aggregation being considered.

Preliminary Considerations for a Statistical Association Methodology

Before discussing a method for accounting for these effects, it would be useful to define our terms and examine their implications. As previously stated, the message is a carrier of information or content. The smallest message carrier of content is probably the alphabetical letter, number, or arbitrary punctuation mark. This is a message of minimum size. A continuous string of such marks, commonly a word, may be thought of as a somewhat larger message. At a still larger level of aggregation, a string of words, perhaps a sentence or a paragraph, is also a message. Similarly, documents, books, clusters of books, and so
forth, are messages of increasing levels of aggregation.

Analytical techniques for determining message or document content do not necessarily have to change radically because of the magnitude of message aggregation being considered. The procedures one uses to examine the subject matter index of a library card file may be similar to the procedures for understanding and searching the individual book cards, which in turn may parallel the procedures used with a book's table of chapter contents, its page index, or the paragraphs and sentences of an individual page itself.

Therefore, to maintain stress upon the common denominator, we will consider all of the strings that constitute messages as a class, becoming specific, when necessary, by indicating the size or level of aggregation for any string. Alphabetical, numerical, or punctuation mark messages are one level of aggregation smaller than those considered in detail at this point. The units of immediate concern are words, strings consisting of a few words, and strings of such strings, including those larger strings that range from sentences or titles, to paragraphs or abstracts, to articles, and so forth.

We establish the following working definition: a word type is the smallest unit of analysis and always has the identical configuration of alphabetical, numerical, and conventional marks. Thus, the word type man is different from men or man's. Similarly is, are, and am are dif-
ferent types. Types may vary in size from one symbol to many. The only requirement is that the symbol arrangement remains the same for the same type.

The ability of a person to react differently to the string of letters man in contrast to the string men, man, or manx reflects the influence of differing structural arrangements of identifiable elements. The string man is a unique system that might be represented by the simple flowgraph below in which the numbers give the distance between the elements of the string

```
    m   2   n
     \\
  --\  !  /--
    1   1   1
     /   \  /
  a   m

```

or, by the somewhat more redundant association list

```
    m   1   a
    a   1   n
    m   2   n
```

The arrangement or association of words can be represented in the same way to identify a sentence, or the association of sentences can identify a paragraph. This also applies to messages of larger aggregation. For example, the string Mary would like John has an identity characterized by the co-occurrence of the four words, the specific sequence of the words, and the distance among them:
In association list form the string would have the representation:

```
Mary 1 would
Mary 2 like
Mary 3 John
would 1 like
would 2 John
like 1 John
```

In this way a message at any level of aggregation can be represented structurally by its co-occurring units at the next lower level by merely specifying the directions and distances among them.

As further illustration consider the following title, descriptors, and abstract as one message:

```
(title) Psychophysical relations in the visual perception of length, area and volume.
```

---

Subjective length, area and volume as functions of the corresponding stimulus variables were studied in three experiments. The exponents of the psychophysical power functions scattered around 1 for perception of real space. For perspective drawings of cubes and spheres, however, the exponents were about 0.75. It was tentatively concluded that perspective is an insufficient cue to visual volume. The results are discussed with special reference to certain cartographic symbols representing population magnitude.

Just for this example, we will establish the following convention. A word type is any unique sequence of exclusively alphabetical symbols with one or more blank spaces preceding and following it, but without blank spaces in the sequence itself. Capital and lower case letters are to be considered identical, and all numbers and punctuation are ignored in identifying types. A primary string is specified as terminating with the presence of a punctuation mark directly followed by two or more spaces. This specification results in choosing as primary strings those sequences of words that correspond to what we ordinarily identify as sentences. Accepting these conventions we can represent the message as a secondary string composed of sentence length primary strings:

Psychophysical relations in the visual perception of length area and volume. Visual perception stimulation tests measurement. Subjective length area and volume as functions of the corresponding stimulus variables were studied in three experiments. The exponents of the psychophysical power functions scattered around for perception of real space. For perspective drawings of cubes and spheres however the exponents were about. It was tentatively concluded that perspective is an insufficient cue to visual volume. The results are discussed with special reference to certain cartographic symbols representing population magnitude.
This message, or any part of it, also can be represented by an association matrix, where the columns represent the first word in a pair, the rows represent the second word, and the cell entries indicate the frequency for each of the co-occurrences. This matrix is, in effect, a simple coded representation of part of the structural content of this one message. With the addition of other messages from the same corpus, the matrix could gradually grow to reflect the co-occurrences of types in all the messages of the corpus in question. This matrix would reflect the statistical structure of the corpus, showing which types were associated and to what extent.

The Development of a Statistical Association Technique

The actual frequency of occurrence of any pair of word types is partially a function of the relevant tendency for the two word types to co-occur because they are associated in some meaningful manner. However, it is also a function of the separate tendencies, irrelevant for this purpose, of either of the word types to occur with all other word types in general. For example, a specific word type will be the first type in as many pairs as there are other types following it in a string. Similarly it will be the second type in as many pairs as there are other types preceding it in a string. A word type will also form pairs as a function of how frequently it occurs as a type in the set of strings under consideration.

It is desirable to normalize to eliminate these extraneous influences: frequency of word occurrence, relative word position, and string length. This can be accomplished by subtracting from the actual
frequency of pair occurrence an estimate of the frequency expected on
the basis of chance and position of occurrences as well as sentence
length for each of the two words that comprise the pair in question,
as follows. We start with a matrix of frequencies of co-occurrences:

<table>
<thead>
<tr>
<th></th>
<th>$x_j$</th>
<th>$x_k$</th>
<th>$(x_j, x_k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_j$</td>
<td>$N(x_j, y_j)$</td>
<td>$N(x_k, y_j)$</td>
<td>$N((x_j, x_k), y_j)$</td>
</tr>
<tr>
<td>$y_k$</td>
<td>$N(x_j, y_k)$</td>
<td>$N(x_k, y_k)$</td>
<td>$N((x_j, x_k), y_k)$</td>
</tr>
<tr>
<td>$(y_j, y_k)$</td>
<td>$N(x_j, (y_j, y_k))$</td>
<td>$N(x_k, (y_j, y_k))$</td>
<td>$N((x_j, x_k), (y_j, y_k))$</td>
</tr>
<tr>
<td></td>
<td>$N(x_j)$</td>
<td>$N(x_k)$</td>
<td>$N(x_j, x_k)$</td>
</tr>
</tbody>
</table>
where

\[ N(x_j, y_j) = \text{the frequency of co-occurrences with word type } j \text{ preceding word type } j. \]

\[ N(x_j, y_k) = \text{the frequency of co-occurrences with word type } j \text{ preceding word type } k. \]

\[ N(x_j, (J, J)) = \text{the frequency of co-occurrences with word type } j \text{ preceding token which are not of word type } j \text{ and not of word type } k. \]

\[ N(x_j) = \text{the sum of the frequencies of all co-occurrences with word type } j \text{ in the first position.} \]

\[ N(y_j) = \text{the sum of the frequencies of all co-occurrences with word type } k \text{ in the second position.} \]

\[ N_0 = \text{the grand total frequency of co-occurrences.} \]

The total frequency of pairs that includes the word type \( j \) in the first position, \( N(x_j) \), is equal to the portion of the length of the string that follows the type \( j \), summed over the total number of occurrences of the type. Similarly, the total frequency of pairs that includes the type \( k \) in the second position, \( N(y_k) \), is equal to the length of the string that precedes the type \( k \), summed over the total number of occurrences of the type.

The row and column totals \( N(x_j), N(x_k), N(y_j), N(y_k) \) and so forth, supply a statistical estimate of the cell magnitude that could be expected because of the extraneous factors of frequency, position, and
string length. Subtracting the customary contingency table correction from the actual cell magnitudes, this estimate of cell magnitude can serve as a first level normalization.

Even with this correction, the cell frequencies are still a function of the actual magnitude of the total corpus of pairs and the total number of word types included in the entire matrix. Thus the greater the total number of pairs, the greater the number to be expected in any cell. Similarly, the fewer the number of word types, the fewer the number of matrix cells, and, therefore, the greater the number of pairs to be expected in any one cell. Consequently, correction of cell frequencies proportional to the total frequency of pairs and inversely proportional to the number of matrix cells results in a set of weights which is normalized for extraneous factors. The resultant cell weights, \( Z_s \), serve as one estimate of the influence of association forces independent of individual frequencies, sentence lengths, number of different types and total number of pairs within the corpus under consideration:

\[
Z(x_j, y_k) = n^2 \left[ \frac{N(x_j, y_k)}{N_o} - \frac{N(x_j)N(y_k)}{N_o^2} \right]
\]

\( ^3 \)Note that this initial correction is identical to the contingency table correction made by Maron and Kuhns, and Stiles on their matrix tabular data, although these investigators use row and column totals based upon frequency of type occurrence, ignoring the variable of how many types are used to identify a document (our notion of string length).
where

\[ N(x_j, y_k) \] is the frequency of co-occurrence of types \( i \) and \( j \).

\[ N(x_j) \] is the total frequency of co-occurrences with token \( i \) as first type.

\[ N(y_k) \] is the total frequency of co-occurrences with type \( j \) as second type.

\[ N_o \] is the total frequency of co-occurrence of all types.

\[ n \] is the number of different types.

When the direction of co-occurrence is not considered, the matrix can be collapsed into triangular form which reflects joint occurrence, where pairs with the words reversed in direction are combined. Each matrix cell of such a triangular matrix, except the cell where \( j \) equals \( k \), is, in effect, the sum of two cells

\[ N(x_j, y_k) + N(x_k, y_j) \]

In this case, the correction for extraneous factors would be:

\[ Z'(x_j, y_k) = \frac{n(n+1)}{2} \left[ \frac{N(x_j, y_k) + N(x_k, y_j)}{N_o} - \frac{N(x_j + y_j) N(x_k + y_k)}{2N_o^2} \right] \]

where \( N(x_j + y_j) \) is the total frequency of pairs containing type \( j \) in either position. Therefore, \( N(x_j, y_j) \) is counted twice.

If the matter of distance of displacement of the words in the pairs is ignored for the moment, a matrix of co-occurrences based upon the statistic \( Z'(x_j, y_k) \) would appear to reflect one statistical tendency of pairs of types to associate. The matrix is adaptive in that it starts
with no cell weights if there has been no input of strings. Then as the inputs begin and continue, the matrix continues to grow and change as it digests ever-increasing quantities of pairs. Each normalized cell weight, $Z'$, rises and falls with time as each specific association increases or decreases in relative frequency. In this way, the matrix memory of associations changes with time, maintaining a cumulative pattern of associations reflecting one statistical characteristic of messages fed into it in the past.

In addition to this adaptive characteristic of changing memory with time and with changes in inputs, the matrix is also readily subject to formal education. Any specific cell weight can be strengthened by repeatedly reading into the matrix memory the specific strings that contain the desired association. For example, by introducing the strings is am, is are, am is, am are, are is, and are am, we can increase the statistical tendency of the tokens is, am, and are to be associated.

More complex learning can be accomplished by the introduction of strings such as man men, men man, singular plural, plural singular, man singular, men plural. In a similar way, we can build chains, lists, trees, and circles of associations. A chain would be formed through the repetitive input of the strings of types such as $a b$, $b c$, $c d$, and so forth. A list would involve input strings of the form $a b$, $a c$, $a d$, $a e$, $a f$, where the word $a$ is the list heading, and the other words are subordinate entries in the list. A tree would involve introducing the
strings a b, b c, b d, c e, c f, d g, d h. Circular associations of the form a b, b c, c d, d a could also be formed. In fact, any particular configuration of links is possible through the development of an appropriate set of input strings.

We have outlined one method for building a matrix of statistical associations which has the general properties of being able to accept and analyze unlimited amounts of natural language text dealing with a wide range of topics. The potential uses of such a technique for forming content association matrices are varied, each potential use depending on a number of further considerations. First, one must evaluate the suitability of an association model to the problem area. Second, there is the question of whether a statistical approach, rather than a deterministic solution, is acceptable. Finally, there is the empirical consideration of whether a matrix of content associations of this type is powerful enough by itself to satisfy the needs implicit in any problem at hand. The specific applications that we intend to explore are: (1) message or document retrieval from a library of such messages or documents, and (2) message or document dissemination or routing. We are currently in the process of implementing a computer-based experimental system for establishing and using the statistical association matrix for message routing and retrieval along the lines suggested above.
PROGRAMMING THE STATISTICAL ASSOCIATION TECHNIQUE

Input Program

As indicated previously, any natural language message material can serve as an input as long as it is in a form compatible with the input requirements of the computer. The input may consist of a complete message or document, message abstracts, titles, or key words and may be placed in the computer by any acceptable input device. At present, very few extensive documents or message files are in a form suitable for use as direct input; to make up a deck of punched cards for a book, for example, is a major effort. However, there are two types of inputs currently available that show promise: the teletypesetter paper tape used by the national magazine and newspapers for regional printing control and the paper tape output used by some document producing agencies for the same or similar purposes. 4

For our initial efforts, a search was undertaken to locate suitable natural language corpora already in a computer-compatible form. Certain criteria of adequacy were: (1) representative of a heterogeneous message or document file; (2) pre-indexed so that criteria of retrieval success could be simply developed; (3) relatively recent; and (4) in a form convenient for input.

4The reader is referred to Stevens (1961) for an excellent discussion of the types of texts available for use in experimental efforts. Cornelius (1961) discusses input problems in general and appears rather reluctant to place any hope on an input device short of an optical reading device.
We found that the Armed Services Technical Information Agency Technical Abstract Bulletin (ASTIA TAB) met these criteria. In addition, the TABs were already being printed from Remington Rand punched paper tapes. Arrangements were made through the Chief of the ASTIA Data Processing Branch and the Director of the Office of Technical Services, U. S. Department of Commerce to borrow the punched paper tapes for two TAB issues, 15 March and 1 April 1962. With the use of an IBM Paper tape reader, the TABs were transferred directly onto magnetic tape in a form compatible with the 1410 computer. Due to errors in the paper tape and certain incompatible character sets, the entire corpus had to be both hand and machine edited.

For those unfamiliar with ASTIA TABs, a typical abstract is given in Fig. 1. It should be noted that a great deal of information is provided for the reader and that many different types of system inputs therefore are available: author names, titles, descriptors, as well as an abstract.

For simplicity of present discussion, in the following material we will restrict the system input to the descriptor list, which we will call the descriptor string. All of the descriptors for a single document are treated as one long sentence of separate words in sequence. We do not maintain the actual descriptors separated by punctuation but use all of the individual words in the entire set of descriptor words as individual types in one long string. All punctuation and capitalization are ignored. (See Fig. 2.)

DESCRIPTORS: (Spaceships, Satellite vehicles, Radiators, Radiant heating panels, Metal plates, Refrigerant condensers, Heat, Temperature control, Heat transfer, Heat exchangers, Convection, Thermal radiation, Design, Configuration, Effectiveness, Mathematical analysis, Military requirements.)

The thermal analysis of component elements of space radiators is described. Elements include rectangular and circular plates of uniform thickness, triangular and trapezoidal fins, and constant temperature-gradient fins. A complete condenser and a radiator are analyzed and illustrative examples given. The thermal analyses produced relationships between the physical properties and dimensions, element and environmental temperatures, and rates of heat transfer. These are shown graphically for all types of elements. The optimum proportions of space radiator elements having the greatest ratio of heat radiation rate per pound of weight are also indicated graphically, and procedures for their calculation are shown. The discussions on condensers and radiators include dimensional-thermal relationships and weight-optimizing procedures for complete units. (Author)

Taken from ASTIA TAB, 15 March 1962.

Fig. 1 Typical ASTIA Abstract
Fig. 2 A Sample of the Descriptor Strings. (The First Two Groups of Numbers Are the AD Number)
**Frequency Matrix Formulation Program**

The frequency matrix formulation program converts the input described above into a matrix of word pair co-occurrences suitable for use in the association matrix program. The program actually is made up of several distinct subroutines. The subroutines and the machines for which they are written are listed:

1. Packing subroutine - IBM 1410 computer
2. Concordance subroutine - IBM 7090 computer
3. Pairing subroutine - IBM 7090 computer
4. Sorting subroutine - IBM 7090 computer
5. Counting subroutine - IBM 7090 computer
6. Printing subroutine (optional) - IBM 1410 computer.

Although the specific functions of these subroutines are suggested by their names, a quick summary of what they do may be informative.

(a) **Packing.** To edit input material expeditiously, an editing program was used. However, the edited output to the concordance subroutine was not adequate. That is, the tape record lengths are small (72 characters) while the concordance can handle up to 1000-character record-lengths. Thus, to provide the input in a more efficient format, a packing subroutine was written whose output results in 1000-character record-lengths.

(b) **Concordance.** This subroutine accepts the output of the packing subroutine and writes a tape which adds for each word a series of identifying numbers. Each word is thus identified as having come from a particular message, a particular sentence, and a particular position in that sentence. The concordance tape is never erased since it contains the entire corpus. The tape is used for several purposes. It is used

---

All of the subroutines in this program were written by members of the Programming Research Subdepartment of the Computer Applications Department at MITRE. In addition, members of this group aided in the conceptualization of much of this work.
during the retrieval program in order to assign document relevancy numbers to each document sentence, and it can also be used at a later time with different programs to generate statistical information about various linguistic parameters such as word length, sentence length, word distances, and the like, as desired.

(c) Pairing. This subroutine works interwoven with the concordance subroutine although they are independent of each other. It is initiated by the discovery of an end-of-sentence mark. The function of this routine is to pair words and prepare an output in the form: "Word - Word" for every word in the string. For example, if the input sentence were "270 563 Algebraic topology groups mathematics." the subroutine would write:

- algebraic - topology
- topology - algebraic
- algebraic - groups
- groups - algebraic
- algebraic - mathematics
- mathematics - algebraic
- topology - groups
- groups - topology
- topology - mathematics
- mathematics - topology
- groups - mathematics
- mathematics - groups

(d) Sorting. In order to count the number of common word pairs, all the word-pairs must be put into some cogent order. This is the most time-consuming portion of the entire program and perhaps most basic to it. The alphabetizing of the word pairs is done by the IBM - 9 SORT program.

(e) Counting. The list of alphabetized word pairs is then reviewed and identical word pairs counted and combined.

---

6 *ASTIA AD 270 563, 1 April 1962*
(f) **Printing.** The printing subroutine is optional and when the system is checked out and operable it will rarely be used. At present, however, for checkout purposes, printouts of the frequency matrix formulation program are being made (see Fig. 3).

**Statistical Language Information**

At present, we have not prepared a detailed program to provide statistical language information. However, certain parameters essential to the operation of the present program are computed and can be printed out at relatively small time and computer cost. These are:

- (a) Number of Tokens
- (b) Number of Types
- (c) Number of Strings
- (d) Frequency Distribution of string lengths in steps of 1 from 1 to 50.

Other linguistic information can, of course, be obtained from the concordance tape.

**Association Matrix Program**

The word-word association matrix uses the output of the frequency matrix generation program. The cell values resulting from the normalization subroutines of the association matrix represent the degree of association between two words after certain corpus artifacts are removed. As we noted in our previous discussion, a cell value in the original frequency matrix is affected by three artifacts; the number of times a word appears in the corpus, the length of the corpus; and, the average string lengths. Other variables probably tend to affect a frequency
Fig. 3  A Sample of the Frequency Matrix for the Query Term "CONVECTION"

Fig. 4  A Sample of the Association Matrix for the Query Term "CONVECTION"
matrix cell value but these cannot, at this time, be fully accounted for and thus subtracted from the cell value.

The normalization program includes an optional printout routine for the 1410 computer (see Fig. 4). As with other optional printout routines, we do not expect it to be used after checkout is completed. However, this routine is useful when we deliberately alter cell values by adding special word strings to strengthen specific word associations.

An Algorithm for Routing or Retrieval

We start with any set of query or routing types, all of which must appear at least once in the corpus. These types serve as the means by which we expand to another set of types, all of which are highly associated through the normalized matrix to the original query types taken as a whole. This set of associated types, combined with the original set of query types, then serves as the basis for deriving another set of types. The procedure can, of course, be repeated any number of times.

This informal characterization of the algorithm can be stated more precisely: Given a set of query types, the matrix is searched to locate all types which have been associated with each and every one of the query types in the set. From this group of words, those (equal in number to the number of query types) that have the highest sum of normalized matrix weights (when summed over all of the query types) are selected to form a set of first order types.
Having obtained this set of first order associates, we form a new set combining these first order types with the original query types. With this larger set of joint first order and query types, the matrix again is searched to locate all types that have been associated with each and every one of the types in this expanded set. From this newly located group of types, those (equal in number to the number of joint first order and query types) that have the highest sum of normalized matrix weights (when summed over all of the first order plus query types) now are selected to form a set of second order types.

The procedure for determining first order associates can be presented in a symbolic form as follows:

Let \( a_{jk} \) = the \( Z'_{jk} \) for \( \tau_j \) with respect to \( q_k \) where, \( q \in Q \)

\[ Q = \{ \text{query terms} \} \]

\( \tau_j \) is any term in the normalized matrix but \( \notin Q \)

\( j = \) any row of the normalized matrix

\( k = \) any column of the normalized matrix

then \( \tau_j \in A = (k) a_{jk} \) & \( s_j \) is among the \( n_q \) highest sum

where, \( A = \{ \text{first order associates} \} \)

\[ s_j = \sum_{k=1}^{n_q} a_{jk} \]

\( n_q = \) the number of terms in the class \( Q \)
The second order associates are derived in a similar fashion as follows:

Let $\beta_{jk}$ be the $Z'_{jk}$ for $\tau_j$ with respect to $\alpha_k$

where, $a \in A$

$\tau_j$ = any term in the normalized matrix but $\not\in Q/A$,

then $\tau_j \in B = (k) \alpha_{jk} \beta_{jk} & \sigma_j \in \text{the } 2n_q \text{ highest sums}$

where, $B = \{\text{second order associates}\}$

$$s'_j = \sum_{k=1}^{n_q} \alpha_{jk} + \sum_{k=1}^{n_a} \beta_{jk}$$

$n_a$ = the number of terms in the class A.

From the above it follows that $Q, Z, B$ are mutually exclusive.

Having derived the first and second order association terms we can then note for each document the occurrence of each query term, each first order term, and each second order term. The documents then are ordered according to the following rules and definitions:

Let $n_b$ = the number of terms in the class B (2nd order associates)

$$n_q = n_a = n_b/2$$

$$j = n_q + n_a + n_b$$

$$k = 100n_q + 10n_a + n_b$$

$d_{j,k}$ = a message or document with j and k indices as defined above.

$d_1 \rightarrow d_2$ means that $d_1$ is more relevant than $d_2$. 
The ordering of messages or documents on the basis of relevance is then:

\[ D_j^r > D_{j-1} \]

and within the \( j \) set of messages

\[ D_{j,k}^r > D_{j,k-1} \]

In such an ordering each cut "\( j \)" is further subdivided by "\( k \)."

This procedure, of course, presumes that messages containing the query types are more relevant than those that do not, those that contain first order associates are more relevant than those that do not, and so forth.

Let us now examine a specific example. We derive the association matrix in the manner described above for 500 documents as represented by their ASTIA descriptors, and a printout is obtained.

The query types chosen \((Q)\) are \textit{thermal, radiation, convection}. We then search for those other types that are associated, either negatively or positively, with all three of them. Fig. 5 gives the more highly associated types and their values as determined from the association matrix, as well as their algebraic sum. Based upon the rules listed above, we select the three first-order associate terms \((\text{Class A})\) that have the highest algebraic sum but that also satisfy the requirement of mutual exclusion. \textit{Radiation} and \textit{thermal} are highest with values of 365 and 157, respectively. However, they already have been chosen as query terms and therefore are rejected. The next three highest are \textit{heat} with 133, \textit{temperature} with 115, and \textit{transfer} with 107.
<table>
<thead>
<tr>
<th>Associated Terms</th>
<th>Radiation</th>
<th>Thermal</th>
<th>Convection</th>
<th>Algebraic Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>radiation</td>
<td>245</td>
<td>116</td>
<td>4</td>
<td>365</td>
</tr>
<tr>
<td>thermal</td>
<td>116</td>
<td>36</td>
<td>5</td>
<td>157</td>
</tr>
<tr>
<td>heat</td>
<td>8</td>
<td>100</td>
<td>25</td>
<td>133</td>
</tr>
<tr>
<td>temperature</td>
<td>87</td>
<td>25</td>
<td>3</td>
<td>115</td>
</tr>
<tr>
<td>transfer</td>
<td>28</td>
<td>66</td>
<td>13</td>
<td>107</td>
</tr>
<tr>
<td>combustion</td>
<td>39</td>
<td>20</td>
<td>6</td>
<td>65</td>
</tr>
<tr>
<td>spaceships</td>
<td>39</td>
<td>7</td>
<td>6</td>
<td>52</td>
</tr>
<tr>
<td>mathematical</td>
<td>15</td>
<td>14</td>
<td>4</td>
<td>33</td>
</tr>
<tr>
<td>panels</td>
<td>16</td>
<td>3</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>control</td>
<td>17</td>
<td>2</td>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td>heating</td>
<td>5</td>
<td>11</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>radiant</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>radiators</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>refrigerant</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>19</td>
</tr>
</tbody>
</table>

(Summed over the three query terms; the cell weight is $Z'$ as previously defined in the paper; each $Z'$ has been multiplied by $1 \times 10^8$.)

Fig. 5 High First-Order Associations Ordered by Algebraic Sum
Now having six terms, we then search for all of the words that are associates of all the six taken as a group, and obtain their algebraic sums. Fig. 6 gives the types most highly associated with all six terms. The six second-order associates with the highest sum are chosen by a process similar to the one for the first order associates, terms previously chosen being eliminated. In the order of their algebraic sum, these are *heating* (195), *gas* (148), *boundary* (140), *layer* (140) *exchangers* (99), and in this case because of the tie in value, *hydrogen* and *spaceships* both with (67).

Having obtained the first and second order association types, we then determine the number of Q, A, and B terms each document contains. In other words, we develop the j index value for each document. If any term is repeated within the document, it is recorded only once. Those documents that have a higher j should be more relevant than those having a lower j. However, for some js the number of documents is quite large, and it is necessary to use the k index within each j. Fig. 7 shows in rank order the documents retrieved by this procedure for the query: *thermal, radiation, convection* in terms of j, and k within j, with a cut-off point at j=3.

The selection and ordering of the documents shown in Fig. 7 have a high degree of face validity; however there is another side to this coin: the number of relevant documents that were not retrieved by the algorithm. The authors and their staff rated all 500 documents on the
<table>
<thead>
<tr>
<th>word</th>
<th>radiation</th>
<th>thermal</th>
<th>convection</th>
<th>heat</th>
<th>temperature</th>
<th>transfer</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>heat</td>
<td>8</td>
<td>100</td>
<td>25</td>
<td>110</td>
<td>131</td>
<td>214</td>
<td>588</td>
</tr>
<tr>
<td>radiation</td>
<td>245</td>
<td>116</td>
<td>4</td>
<td>8</td>
<td>87</td>
<td>28</td>
<td>488</td>
</tr>
<tr>
<td>transfer</td>
<td>28</td>
<td>66</td>
<td>13</td>
<td>214</td>
<td>12</td>
<td>20</td>
<td>353</td>
</tr>
<tr>
<td>thermal</td>
<td>116</td>
<td>36</td>
<td>5</td>
<td>100</td>
<td>25</td>
<td>66</td>
<td>348</td>
</tr>
<tr>
<td>temperature</td>
<td>87</td>
<td>25</td>
<td>3</td>
<td>131</td>
<td>-41</td>
<td>12</td>
<td>217</td>
</tr>
<tr>
<td>heating</td>
<td>5</td>
<td>11</td>
<td>6</td>
<td>78</td>
<td>44</td>
<td>51</td>
<td>195</td>
</tr>
<tr>
<td>gas</td>
<td>-27</td>
<td>22</td>
<td>13</td>
<td>103</td>
<td>-3</td>
<td>37</td>
<td>148</td>
</tr>
<tr>
<td>boundary</td>
<td>-7</td>
<td>5</td>
<td>13</td>
<td>73</td>
<td>-3</td>
<td>59</td>
<td>140</td>
</tr>
<tr>
<td>layer</td>
<td>-7</td>
<td>5</td>
<td>13</td>
<td>73</td>
<td>-3</td>
<td>59</td>
<td>140</td>
</tr>
<tr>
<td>exchangers</td>
<td>0</td>
<td>10</td>
<td>6</td>
<td>62</td>
<td>2</td>
<td>19</td>
<td>99</td>
</tr>
<tr>
<td>hydrogen</td>
<td>-10</td>
<td>11</td>
<td>6</td>
<td>32</td>
<td>13</td>
<td>15</td>
<td>67</td>
</tr>
<tr>
<td>spaceships</td>
<td>39</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>9</td>
<td>2</td>
<td>67</td>
</tr>
<tr>
<td>panels</td>
<td>16</td>
<td>3</td>
<td>7</td>
<td>21</td>
<td>5</td>
<td>5</td>
<td>57</td>
</tr>
<tr>
<td>radiant</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>20</td>
<td>5</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>radiators</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>20</td>
<td>5</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>refrigerant</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>20</td>
<td>5</td>
<td>6</td>
<td>50</td>
</tr>
</tbody>
</table>

(Summed over all query and first order terms; the cell weight is $Z'$ as previously defined; each $Z'$ has been multiplied by $1 \times 10^8$.)

Fig. 6 High Second-Order Associations Ordered by Algebraic Sum
basis of their relevance to the query thermal, radiation, convection. The judgment of relevance was made on the basis of the more complete document abstract rather than on the descriptor string. Seven additional relevant documents were located. Of these, two were selected by the algorithm with a cut-off point set at \( j=2 \) and one was picked up with \( j=1 \). The remaining four were poorly indexed and would have been judged not relevant by a human who depended upon the descriptor string only, as the matrix did, rather than upon review of the abstracts.

**Conclusion**

The work done to date, and reported at this time is introductory. Continuing analysis of statistical association procedures is now in progress, and additional results will be presented in subsequent reports.
AD-269 587 Div. 12
(TISTA/SEC) OTS price $4.00

SPACE RADIATOR ANALYSIS AND DESIGN. PART I,
by D. B. Mackey and C. F. Bach, Oct 61, 287p. (incl. illus. tables, refs. (Rept. no. ASD 61-66)
(Contract AF Project 60-6163, Proj. 6146)
(Rept. no. 6130, pt. 1) Unclassified report


The thermal analysis of component elements of space radiators is described. Elements include rectangular and circular plates of uniform thickness, triangular and trapezoidal fins, and constant temperature-gradient fins. A complete condenser and a radiator are analyzed and illustrative examples given. The thermal analysis produced relationships between the physical properties and dimensions, element and environmental temperatures, and rates of heat transfer. These are shown graphically for all types of elements. The optimum proportions of space radiator elements having the greatest rate of heat radiation rate per pound of weight are also indicated graphically, and procedures for their calculation are shown. The discussions on condensers and radiators include dimensional-thermal relationships and weight-optimizing procedures for complete units. (Author)

\[ J = 8 \]
\[ k = 233 \]

\[ J = 9 \]
\[ k = 333 \]

Fig. 7 Rank Order of Documents Retrieved
CONVECTIVE HEAT TRANSFER WITH CHEMICAL REACTION,
THEORETICAL DEVELOPMENT OF CORRELATION FORMULAE FOR THE PREDICTION OF HEAT FLUXES IN HIGH PERFORMANCE ROCKETS MOTORS AND RELATED SYSTEMS.
Internal technical rep. on Research on Combustion Kinetics,
by Daniel E. Rother. Aug 61, 123p. incl. illus., tables, 129 refs.
(ARL 99, pt. 1)

Unclassified report

Descriptors: (Heat transfer, Convection, Boundary layer, Gas flow, Chemical reactions, Thermochemistry, Thermodynamics, Rocket motors.) (Transport properties, Recombination reactions, Boundary layer, Gas flow, Hypersonic flow, Hypersonic nozzles.) (Dissociation, Combustion, Hydrogen, Oxygen, Diffusion.) (Experimental data, Tables.)

Energy transfer in chemically reacting boundary layer flows is discussed from the point of view of the investigator, who is seeking to extend existing correlation formulae to cases in which thermochanical effects influence heat transfer rates. Emphasis is placed on the prediction of convective heat fluxes in high performance rocket motors; however, examples are also taken from the field of aerospace, to follow: the appropriate driving force for heat transfer with chemical reaction, effects of the enhanced efficiency of energy transport by diffusion as compared to ordinary conduction, calculation of the turbulent film conductance in axisymmetric nozzles, thermodynamic calculation of enthalpy/mixture-ratio charts for combustion gas mixtures, effects of chemical non-equilibrium in the gas phase, effects of surface catalyzed exothermic recombination reactions, estimation of transport properties in partially dissociated gas mixtures with emphasis on the binary diffusion coefficients pertaining to molecular fragments. (Author)

j = 7
k = 034

Fig. 7 (Continued)
Blast Tanks.
Quarterly progress rept. no. A for period ending 31 July 61,
(Contract AF 33(616)3154, Proj. J086)
Unclassified report

DESCRIPTIONS: (Liquid rocket propellants, Cryogenics, Propellant tanks, Fuel tanks,
Propellant tank liners, Insulating materials, Design, Processing, Heat transfer, Aerodynamic
heating, Test methods, Test equipment.) (Liquidified gases, Hydrogen.) Thermal insulation.

The design and fabrication of the 7000 gal.
liquid hydrogen fuel tanks of TI and stainless
steel are described. Insulating materials and
methods of insulation are also described. Tests
were successfully conducted to check out the
tank drain assembly. Further tests will be con-
ducted to determine the heat transfer of the
tanks during simulated rocket flight conditions.

\[ j = 5 \]
\[ k = 122 \]

Fig. 7 (Continued)
Near the leading edge of a sharp flat plate in high speed, low density flow of high temperature gases the induced shock wave is almost straight and the pressure and velocity downstream are therefore approximately constant. Assuming that the outer edge of the viscous layer coincides with the shock wave in this region of the flow, it is possible to integrate the boundary layer equations with first order slip in the boundary layer. Of the first two terms in this series have been found in closed, analytical form. To terms in $x$ to the first power the result is applied to the flat plate flying at Mach number 20 at 255,000 feet altitude, where the effects of slip are significant for a considerable distance downstream of the leading edge. (Author)

$$j = 5$$

$$k = 023$$

Fig. 7 (Continued)
The effects of conduction and viscosity on the stability of laminar flame are examined. If \( j \) denotes the ratio of the wave length of a disturbance to the flame width and \( \alpha \) is the ratio of the ultimate temperature of the burned gas to the initial temperature of the fresh mixture, the flame is found to be stable if \( j < \alpha \). Experiments were performed to determine the Prandtl number of the mixture. It is further shown that the stabilization is achieved primarily through the effect of heat conduction on the flame speed rather than the influence of viscosity. (Author)

\[
j = 5 \\
k = 023
\]

It further shown that the stabilization is achieved primarily through the effect of heat conduction on the flame speed rather than the influence of viscosity. (Author)
A compilation is presented of original test data on emissivity, reflectance, and adsorption of Cr, Nb, Mo, Ta, and W. The data were taken from the literature published during the period 1940-1959 inclusive, and as much of the 1960 literature as could be obtained. The following sources were searched: Chemical Abstracts, Ceramic Abstracts, Metallurgical Abstracts, Nuclear Science Abstracts, and the files of the Defense Metals Information Center (DMIC). An attempt was made to evaluate these sources of data according to the apparent thoroughness of methods and techniques as described by the various investigators. In many cases the descriptions in the literature are a summary of methods and results, and a complete evaluation is impossible. Curves are presented which appear to indicate the most probable values for the various conditions and materials. (Author)

\[ j = 4 \]
\[ k = 220 \]

Fig. 7 (Continued)
EXPERIMENTAL DETERMINATION OF THE SLOW NO DECOMPOSITION REGIME AROUND 3000 DEGREES K BEHIND SHOCK WAVES.

DESCRIPTORS: (Shock waves, Nitrogen compounds, Oxides, Infrared radiation, Velocity, Measurement, Temperature, Density, Optical systems.) (Heat transfer, Gases, Load distribution, Mach number, Vacuum systems, Photographic analysis.) (Electronic equipment, Shock tubes, Oscillograms, Oscilloscopes, Miniaturized electronic equipment, Piezoelectric gages.)

The decomposition of nitric oxide between 2250 and 3600 K has been studied in shock tube experiments. The emitted infrared radiation of the fundamental band system of NO has been used to determine the time duration of the slow decomposition regime of NO behind shock waves in pure NO. It is found that the temperature dependency of the duration is in good agreement with the theoretical model we have described previously. (Author)

\[ J = 4 \]
\[ k = 130 \]

Fig. 7 (Continued)
A thermoelectric heating and cooling module was constructed for installation in a water to water air conditioning system aboard a submarine. This module has a cooling rating of 2550 Btu/hr, at a coefficient of performance of 0.75 and an operating current of 35 amperes dc. This rating was based on a 65 F sink water temperature and a chill water temperature of 55 F. The unit was designed to withstand submergence pressures and the corrosive effect of sea water in all water passages. It occupies a volume 1 ft. by 1 ft. by 3 inches and was designed for ease in stacking into larger capacity units without additional space being required for coupling between units. It has a weight of 55 pounds. (Author)

j = 4
k = 121
An engineering model of a system for reducing CO₂ at a rate of 500 cc/min by reaction with H₂ over heated catalysts was developed. The primary products of the reaction were solid C and water vapor. To recover breathable O₂, condensed water vapor would be fed to an electrolysis cell; the H₂ by-product of the electrolysis would then be used to reduce more CO₂. Solid carbon is removed periodically from the apparatus and discarded. In the final test, the apparatus was operated continuously for a period of 11 hr. at about 10% above the target conversion rate. We estimate that the C deposits could be accumulated for at least 2 days in the present reactor before interrupting the process for removal of C and renewal of catalyst. With appropriate maintenance procedures, the apparatus should operate for the specified maximum of 3 yr without difficulty.

Fig. 7 (Continued)
Various problems associated with the macroscopic magnetohydrodynamic effects on exothermal flow considerations. A steady exothermal wave traveling in an ionized medium under the influence of a transverse magnetic field was found to display properties similar to those of a classical detonation or deflagration wave. For a hydromagnetic exothermal wave, it was found that a discontinuity either in thermodynamic quantities or in magnetic field strength appears as soon as the flow reaches the transition region.

A solution to the problem of strong interaction between shock waves and the boundary layer has been obtained for the case where velocity slip and temperature jump boundary conditions are consistent at the wall. It is shown that the addition of slip boundary conditions yields a correction of order (boundary layer thickness/δ) to the no-slip solution. Estimates are made of the effect of slip on induced pressures and skin friction for the case of the adiabatic wall. In addition, it is shown that the inclusion of slip boundary conditions does not change the energy transfer to the wall from the no-slip values.
Design information on liquid propellant tank pressurization systems is presented. The area covered are pressurization gas requirements, including hand calculation procedures and nomographs; tankage, including material properties and volume and wall area curves; and components, including stored gas system weight curves and a simple but accurate heat exchanger design method. (Author)

\[ j = 4 \]
\[ k = 013 \]
The total and the spectral intensity of the radiation emitted by rocket combustion products at high pressure was studied. Spectral and total radiation data are presented for 1 rocket propellant systems: \( \text{HNO}_3 \) and \( \text{HNO}_4 \); \( \text{N}_2\text{O}_4 \); \( \text{NO}_2 \) and \( \text{N}_2\text{O}_4 \). Each of the combinations contained \( \text{NO}_2 \) as the principal radiating gas. The measurements were made at a nominal combustion pressure of 7000 psi. Emissivity correlations and methods of estimating the temperature of the combustion gases are developed. A technique is given for obtaining direct observation of high temperature gases inside the combustion chamber of a small rocket motor.

The method consists of minimizing a positive functional which vanishes only when the correct stationary space distribution is assumed. (Author)

\[
\begin{align*}
\text{j} &= 3 \\
\text{k} &= 210
\end{align*}
\]
AD-514 Div. 25
(TSTP/MFA) OTS price $2.60

Aerospace Corp., El Segundo, Calif.

Heat Dissipation through Diode Lead Wires Under Steady-State Conditions,
by W. D. Buckman, Oct 61, 20p. incl. illus. tables (Rept. no. TDR-61/3)
(Contract AF 04(627)-1)

Unclassified report

DESCRIPTORS: (+Heat transfer, Diodes, Wire, Conductors, Nickel, Copper, Thermal conductivity, Electric connectors, Tests, Thermodynamics.) (Experimental data, Tables, Functions, Equations.)

Results of an investigation of the capabilities of wire leads to function as heat dissipating media are presented. Experimental work related to this project has confirmed that leads may serve as heat sinks to an extent greater than has generally been recognized, and has also served as a basis for derivation of the theoretical relationships which define the significant parameters involved. The experiments were conducted to generally determine the effects of varying wire lead materials, lengths, and diameters under both radiative and convective ambient conditions. The mathematical relationships which have been obtained provide quantitative methods for predicting the effect and behavior of component-generated heat on performance, and will permit better correlation between component wattage ratings as stated by the manufacturer and noted by the user. (Author)

Fig. 7 (Continued)


Contents:
High temperature - high friction materials
Lining compositions
Procedures
Data
Rocket motor material evaluation
Test facilities
Jet piercing torch data
Final report on 2 firings on 3.75 in. Rocket Data from 10 firings on 3.75 in. Rocket Final report on 15 firings on 5.75 in. Rocket Data from 7 firings on 5.75 in. Rocket

\[
\begin{align*}
\text{j} &= 3 \\
\text{k} &= 120
\end{align*}
\]

\[
\begin{align*}
\text{j} &= 3 \\
\text{k} &= 120
\end{align*}
\]

Fig. 7 (Continued)
DEVELOPMENT AND EVALUATION OF HIGH TEMPERATURE PROTECTIVE COATINGS FOR COLUMBIUM ALLOYS. PART II. COATING EVALUATION.

(Contract AF 33(616)72415, Proj. 7351)
(ABD TN 61-66, pt. 2) Unclassified report

DESCRIPTIONS: (Refractory coatings, Niobium alloys, Titanium alloys, Molybdenum alloys, Tungsten alloys, Directional alloys, Niobium, High temperature research, Heat resistant alloys.) (Oxidation, Creep, Thermal stresses, Mechanical properties.) *Oxidation inhibitors.

A comparative evaluation was made of 18 coating-base metal systems, six different coatings applied to 3 Nb base materials (B-51, alloy and unalloyed Nb). The 18 coating-base metal systems were tested under the same conditions in cyclic oxidation (2300 and 2500 F), thermal shock (2300 to 2500 F), bend-oxidation (2500 F) and stress-oxidation (2500 F) tests. The tests produced directly comparable data between the coating-base metal systems relating to the protective nature of such coating and the effect of the coating and the coating treatment on the mechanical properties of the substrate. (Author)

\[ J = 3 \]
\[ k = 120 \]

Fig. 7 (Continued)
**UNCLASSIFIED REPORT**

**DESCRIPTORS:** *Handbooks, Textiles, Synthetic fibers, Fibers, Threads, Nylon, Dacron, Deceleration, Parachute fabrics, Coatings.*

(Physical properties, Mechanical properties, Porosity, Semipermeability, Climatic factors, Radiation effects, Temperature, Aerodynamic heating, Cooling, Aging, Friction.)

Bibliography.

Content: Aging properties; Design data, basic; Friction, abrasion, wear; Impact loading; Porosity and air permeability; Sensitivity; Sunlight and weather resistance; Temperature properties; Chemical resistance; Radiation properties; Aerodynamic heating.

---

AD-269 384 Div. 14. 1
(T1STN/GE1) OTS price $1.50


**HANDBOOK OF FIBROUS MATERIALS**


(Contract AF 33(656)7504, Proj. 7301) In cooperation with Fabric Research Labs. Inc.

WADD TR 60-384, pt. 2

---

AD-269 961 Div. 9
(T1STP/TL) OTS price $1.10

Durham U. (Gt. Brit.).

**RESEARCH ON THERMAL CONVECTION IN ROTATING FLUIDS.**


(Contract AF 61(052)216)

(AFCHL-601) Unclassified report

**DESCRIPTORS:** (Cylindrical bodies, Fluids, Fluids flow, Rotation, Thermodynamics, Heat transfer, Convection.) (Liquids, Heat transfer, Hydrodynamics.) (Water, Glycols, Density, Acceleration, Gravity, Temperature, Oscillation, Viscosity.)

This report summarizes progress made with two investigations of thermal convection in a rotating liquid contained between concentric cylinders. In the first, careful measurements of the heat transfer coefficients at different rates of rotation were made and striking results obtained. In the second, an extension of earlier work on instabilities and other properties of the wave regime, is being carried out. (Author)

---

\[ j = 3 \]

\[ k = 030 \]

*Those terms were not included on the magnetic tape.*

---

Fig. 7 (Continued)
AD-269 553 Div. 25, 30
(TISTP/WH) OTS price $10.50

Lockheed Aircraft Corp., Marietta, Ga.
MAIN PROPELLANT TANK PRECIPITIZATION SYSTEM
STUDY AND TEST PROGRAM, VOLUME IV, COMPUTER
PROGRAM.
Final rept., 1 July 60-31 Oct 61.
Dec 61, 176p. incl. illus. tables. (Rept. no.
ER-5206)
(Contracts AF 04(611)6087 and AF 14(611)7032,
Proj. 6753)
(SSD TR 61-21, vol. 4) Unclassified report

DESCRIPTORS: (Guided missiles, fuel tanks, propellant tanks, heat transfer, aerodynamic heating.) (Thermodynamics, Equations, digital computers.) (Programming.) (Pressure tanks, gases, heat transfer.)

A computer program which can be used to determine the pressurizing gas requirements for a missile propellant tank pressurization system is described. The program is applicable to both cryogenic and storable propellants when pressurized with stored gas, evaporated propellant, and main tank injection methods of pressurization. (Author)

\[ j = 3 \]
\[ k = 0.021 \]

Fig. 7 (Continued)
The results of studies related to the oxidation of W and its alloys are studied. The pressure of WO3 polymer over WO2 was measured in a N gas flow integrator with aperiodic unit throttle. Literature data for WO2 were combined with vapor pressures determined in the project to give thermodynamic values for WBO2 and W2BO5. W oxidation rates were measured from 800 to 1700°C and in O pressures between 0.2 and 0.02 atmospheres. The effects of O pressure indicate that the rate may be governed by O dissociating to atoms at the reacting surface. The oxidation rate is demonstrated to be independent of the oxide evaporation rate. All of the evidence indicates that if an oxide barrier layer is present at temperatures above 800°C it must be very thin. Studies on the oxidation of Ta - W alloys between 800 and 1200°C indicate that the 50-50 alloy has the greatest oxidation resistance, oxidizing at a rate only 10 times slower than W alone. (Author)

\[ j = 3 \]
\[ k = 0.021 \]

Fig. 7 (Continued)
AD-269 628  Div. 12, 9
(TISTA/646) OTS price $1.60


HEAT EXCHANGE AT THE FRONTAL POINTS OF BLUNT BODIES WASHED BY A SUPERSONIC GAS FLOW.

by V. P. Hnatovich. 7 Aug 61, 16p. Incl. Illus. (Trans. no. WCL-1108/1 of Konvektivny V Lezhny Toplovochmen, Moscow, pp. 16-26, 1960) Unclassified report

DESCRIPTIONS: (Blunt bodies, Aerodynamics, Supersonic, Aerodynamic heating, Heat transfer, Mathematical analysis, USSR.)

\[ j=3 \]
\[ k=021 \]

AD-269 148  Div. 22, 14, 4, 25, 9
(TISTA/558) OTS price $0.10

Army Rocket and Guided Missile Agency, Huntsville, Ala.

QUARTERLY RESEARCH REVIEW NO. 30. 1 MAY-1 JULY 61.


DESCRIPTIONS: (Reinforcing materials, Glass textiles, Phenolic resins, Mechanical properties.) (Boron, Preparation, Purification.) (Nuclear physics, Radioactive decay.)(Ethylene, Nitrogen compounds, Fluorides, Combustion.) (Electronics, Atmospherics, Noise analyzers, Theory.) (Gases, Magnetohydrodynamics, Plasma physics, Plasma jets, High temperature research.) (Hypersonic wind tunnels, Hypersonics, Re-entry aerodynamics, Aerodynamic heating. Simulation.) (Solid state physics, Electronics, Transport properties.)

Contents: MATERIALS (Mechanics of materials and structures and preparation of high purity boron); NUCLEAR PHYSICS (Isomeric yields from \( (n, 2n) \) reactions); FUELS AND COMBUSTION (Burning rates of ethylene-NF3); PHYSICAL ELECTRONICS (Spectrum of amplitude-modulated noise after square-law detection III); HIGH TEMPERATURE PHYSICS (Simplified probe theory III, Plasma jet in probe measurements, Analysis of probe data by simplified double probe theory, Graphs for the truncation of partition functions of selected elements, Control system for the hyperthermal test facility, and Electron capture in a magnetic field; and SOLID STATE PHYSICS (Transport of fast electrons).

\[ j=3 \]
\[ k=021 \]

Fig. 7 (Continued)
AD-269 160  Div. 17, 27, 1, 20
(TISTM/GEC) OTS price $7.60

Foreign Tech. Div., Air Force Systems Command,
Wright-Patterson Air Force Base, Ohio.
INVESTIGATIONS OF HEAT RESISTANT ALLOYS (SELECTED
ARTICLES). (Issledovaniya Chornevychy Splavov
(Trudy 125)).
4 Dec 61, 69p. 11 refs. (Trans. in Full Texts-TT-61-31
of Gesundstromanzy Nauchno-tekhnicheskaya
Izdatel'stvo Ogoron, Moskva, pp. 17-34,
45-52, and 65-68, 1965)
Unclassified report

DESCRIPTIONS: 1. Heat resistant alloys,
2. Corrosion-resistant alloys, High temperature
research, Stainless steel, Austenite, Steel,
Titanium alloys, Aluminum alloys, Tungsten
alloys, Silicon alloys, Chromium alloys,
Boron alloys.) (Tests, Crystal structure,
Grains [Metallurgy], Mechanical properties,
Physical properties, Surface properties.)
(Aircraft, Hypersonics, Supersonic planes,
Gas turbine blades for Jet engines, Nuclear
power plants.) USSR.

Contents:
Selective alloying as a method for improving
heat resistance, by S. N. Vinarov
Investigation of the dependence of the structure
and properties of gas turbine blades on the
duration of operations, by S. T. Kishokha,
A. A. Kiypa, N. V. Karpykha
Strength of alloys in contact with Na, by S. T.
Kishokha, and G. F. Benadikova
High temperature corrosion resistance, by G. N.
Kashina

j=3
k=021

AD-271 181  Div. 12
(TISTA/SEG) OTS price $1.01

National Aeronautics and Space Administration,
Washington, D. C.
DETERMINATION OF NUCLEAR-ROCKET POWER LEVELS FOR
UNMANNED MARS VEHICLES STARTING FROM ORBIT AROUND
EARTH.
by Richard H. Cavicchi and James W. Misser.
JPL (2, Cwr. incl. Illus. 8 refs. (NASA Technical
Note D-474)
Unclassified report

Also available from NASA, Wash. 25, D. C. as
NASA Technical note D-474.

DESCRIPTIONS: 1. Space flight, 2. Space probes,
Mars, Space ships, 3. Nuclear propulsion,
Hydrogen, Temperature, Thrust, Specific
Impulse, Design, Feasibility studies, Military
requirements, Theory, Mathematical analysis.)

Nuclear-powered Earth-orbital-launch probes can
place greater payloads in orbit around Mars than
chemical vehicles if reactor power exceeds
50 mw. Suitable reactor powers for this mission
are about 150, 400, and 1000 mw for 33,000-
81,000-, and 200,000-lb vehicles, respectively.
Whereas a 33,000-lb vehicle requires greater
than a 200-day coast, a 165-day coast is fea-
sible for an 81,000-lb vehicle. A hydrogen
temperature of 4000 F in the nozzle appears to
be a good compromise. Use of a solid-propellant
rocket to achieve an orbit about Mars from
does not yield an payload advantage. Using
optimum firing dates, a 33,000-lb vehicle
could orbit an acceptable payload about Mars,
and an 81,000-lb vehicle could load freight on
Mars. (Author)

j=3
k=012

Fig. 7 (Continued)
REFERENCES


Hq. ESD, L.G. Hanscom Field, Bedford, Mass.

Unclassified Report
This is an introductory report of an investigation concerned with developing procedures for utilizing certain statistical properties of messages or

1. Coding
2. Data processing systems
3. Library science
I. Project No. 702
II. AF33(600)-39852
III. The MITRE Corporation Bedford, Mass.
IV. Spiegel, J.
V. Bennett, E.

Hq. ESD, L.G. Hanscom Field, Bedford, Mass.

Unclassified Report
This is an introductory report of an investigation concerned with developing procedures for utilizing certain statistical properties of messages or

1. Coding
2. Data processing systems
3. Library science
I. Project No. 702
II. AF33(600)-39852
III. The MITRE Corporation Bedford, Mass.
IV. Spiegel, J.
V. Bennett, E.
documents; these properties to be used for message routing or retrieval. This approach applies the most elementary relation among the words making up a message, that of word-word co-occurrence probability patterns. It is shown that any message material, be it natural language, code, or index terms, can be processed provided that the input is compatible with the input requirements of the computer. Unclassified Abstract