

UNCLASSIFIED

AD 4 3 6 8 6 7

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA



UNCLASSIFIED

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.

64-12

436867

(17) next no
TR-7

(1)

(9)

MONTHLY TECHNICAL REPORT.

How Nov-63.

6

FORT DETRICK THESAURUS PROJECT.

(11)

6 Dec 1963

(15)

Contract DA 19-054-AMC-132(A)

AD NO. _____
DDC FILE COPY

436867

Prepared for:

U. S. Army Biological Laboratories
Fort Detrick, Maryland

Prepared by:

General Electric Co.
Information Systems Operation
1901 Fairmont Avenue
Washington, D. C. 20014

DDC
RECORDED
APR 28 1964
RESERVED
TISA B

(5) 354650

2

FORT DETRICK THESAURUS PROJECT

INTRODUCTION

This informal report presents a summary of the technical progress for the month of November 1963.

PROGRESS IN NOVEMBER

Work during November fell into the following areas:

- o Completion of Phase I report
- o Planning for Phase II effort

Completion of Phase I report

The remaining few percent of the descriptor grouping structuring was completed, and the report, Descriptor Grouping Recommendations, was delivered to Fort Detrick on 6 November 1963. The approximate number of descriptors associated with each of the 16 top-level descriptors in Display I is as follows:

<u>Top-level descriptors</u>	<u>Approx. number of descriptors</u>
Anatomy	1,000
Automatic Data Processing	140
Biological and Chemical Warfare	300
Chemistry and Biochemistry	490
Ecology	40
Epidemiology	110
General	50
Genetics	270
Immunity	370
Laboratory	1,570
Material Medica	700
Mathematics and Statistics	80
Pathology	860
Physics and Biophysics	300
Physiology	670
Taxonomy	3,070

The figures above reveal that approximately 10,000 descriptors are tabulated in Display I. The number of unique descriptors is less than this figure since many of the descriptors appear two or more times. In Display II, Candidate List of Frequently Used Descriptors, approximately 1,800 descriptors are tabulated. Here again the number of unique descriptors is less than this figure since some descriptors appear more than once.

Planning for Phase II effort

During November some plans for the Phase II effort were formulated. In some areas Fort Detrick and General Electric must mutually agree on certain items related to Phase II work. These plans and agreements concerning Phase II fall into the following areas:

- o Approach for indexing Fort Detrick documents.
- o Revision of Display I and Display II.
- o Generation of Display III.
- o Statistical data on document-descriptor relationships.
- o Format of machine readable medium for document number-descriptor records.
- o Format of machine readable medium for record of Thesaurus information.
- o Type of machine readable medium to be furnished to Fort Detrick.

Approach for indexing - A separate worksheet will be used for each document being indexed. Recorded information will include document accession number, assigned descriptors each identified by a four or five character numeric code, number of descriptors assigned to each document, the main group heading from which each descriptor was selected, security classification of document, indexer's identification, date of indexing, and any other information mutually agreed upon by Fort Detrick and General Electric. Present estimates call for the selection of approximately 20 descriptors per document. About 650 documents for indexing have been ordered from the Defense Documentation Center. A portion of these have been received, and some order forms have been returned together with information about some of the documents not being available from DDC. Over 90% of the documents to be indexed must be supplied by Fort Detrick.

Revision of Display I and Display II - Display I, Total Groupings of Descriptors, and Display II, List of Frequently Used Descriptors, will be revised based upon knowledge acquired during the indexing period and upon guidance provided by Fort Detrick personnel. For example, it is evident that the munitions documents indexed during Phase II will reveal information to revise the descriptor displays. It is expected that greater revision will be made in Display II.

Generation of Display III - Display III, Alphabetic Listing of Descriptors, will be an alphabetic tabulation of individual and combined descriptors together with associated designations when appropriate. Designations presently being considered are the following:

"Use" designates an individual descriptor or a combination of descriptors being equivalent, nearly equivalent, or similar to and to be used instead of the entry.

"Also see" suggests descriptor(s), which in addition to the entry, might also be of interest although there is not necessarily a structural relationship between these descriptors and the entry.

"Included in" designates an individual or a combination of descriptors which encompasses the concept of the entry and must be used in handling the entry concept.

"Indented to" designates the individual descriptor(s) or combination(s) of descriptors to which the entry is indented.

In addition to the above designations, it is planned that each entry will include a listing of the Display I main group heading(s) under which the entry appears and will also include some designation to appropriately identify the entry as a descriptor or a combination of descriptors appearing in Display II.

Statistical data on document-descriptor relationships - Present plans are to present data to show the number of times each descriptor was used for indexing the Fort Detrick documents. Other data will show variations in the number of descriptors per document used for indexing. Also, the data will indicate the variations in the number of descriptors selected for indexing as a function of the Display I main headings under which the various descriptors appear.

Format of machine readable medium for document number-descriptor records - Plans are to use a five character numerical accession number for each document. It may be advisable to format the record with a character to indicate the security classification of each document. The use of serial records or an inverted file must be mutually agreed upon as well as the detailed format to be used within each record.

Format of machine readable medium for record of Thesaurus information - Fort Detrick and General Electric must mutually agree on the type of and format of the Thesaurus record to be produced.

Type of machine readable medium to be furnished to Fort Detrick - Unless mutually agreed upon plans are to use a Friden 8-channel punched paper tape as the machine readable medium.

"Included in" designates an individual or a combination of descriptors which encompasses the concept of the entry and must be used in handling the entry concept.

"Indented to" designates the individual descriptor(s) or combination(s) of descriptors to which the entry is indented.

In addition to the above designations, it is planned that each entry will include a listing of the Display I main group heading(s) under which the entry appears and will also include some designation to appropriately identify the entry as a descriptor or a combination of descriptors appearing in Display II.

Statistical data on document-descriptor relationships - Present plans are to present data to show the number of times each descriptor was used for indexing the Fort Detrick documents. Other data will show variations in the number of descriptors per document used for indexing. Also, the data will indicate the variations in the number of descriptors selected for indexing as a function of the Display I main headings under which the various descriptors appear.

Format of machine readable medium for document number-descriptor records - Plans are to use a five character numerical accession number for each document. It may be advisable to format the record with a character to indicate the security classification of each document. The use of serial records or an inverted file must be mutually agreed upon as well as the detailed format to be used within each record.

Format of machine readable medium for record of Thesaurus information - Fort Detrick and General Electric must mutually agree on the type of and format of the Thesaurus record to be produced.

Type of machine readable medium to be furnished to Fort Detrick - Unless mutually agreed upon plans are to use a Friden 8-channel punched paper tape as the machine readable medium.