

25X1A5a1



T&E/365

12 April 1963

A Proposal for the
DEVELOPMENT OF AUTO-INSTRUCTIONAL MATERIALS
IN INFORMATION ANALYSIS AND CLASSIFICATION

In a retrieval system, information is indexed in the categories outlined by a descriptor language such as the Intelligence Subject Code. This language, a list of words and phrases, is used to classify information in categories representing areas of interest to the user of the system. Usually the categories of the descriptor language are not determined by logically dividing the subject field, but by analyzing the type of questions asked of the system and the type of material indexed in it. For example, a logical analysis might suggest that the descriptor language should contain as many subtopics about one political party as another. However, if the Communist Party were of primary interest to the user of the system, information about its activities would be classified in more detail. Therefore, the descriptor language of a retrieval system often seems to be a highly arbitrary arrangement of topics; one which often conflicts with a common sense classification of the same subject matter. An early goal of training personnel to index information in a retrieval system is to have them abandon their own "common sense" concepts of classification and respond in terms of the categories defined by the descriptor language.

In addition to the difficulties that arise from the nature of the descriptor language, a second training problem is caused by the nature of the information to be indexed. The descriptor language of the system is a finite list of words and phrases descriptive of the system's categories, but the information to be indexed is expressed in an infinite number of ways. Although two documents might be phrased in completely different words, they could actually mean the same thing. Because dissimilar documents can fall in the same descriptor language category, classification of information in a retrieval system depends on the judgment and the accumulated knowledge of experienced indexers. In spite of the automation of many retrieval system operations, such as filing and storage, the input process has remained a human, and consequently, expensive, time-consuming, and often unreliable operation.

-2-

Although it is as yet impossible to automate initial classification of information in a retrieval system, it is possible to enhance the training of the individuals who make the input judgments. Already the procedural operations involved in indexing information for the Intellofax Retrieval System have been taught by auto-instructional techniques. Although this program dealt to some extent with both of the training problems mentioned here -- the categories of the descriptor language and the relationship between document wording and descriptor language terminology -- it emphasized the procedural operations of a specific retrieval system.

To extend the usefulness of these materials and further enhance training, it is proposed that auto-instructional materials be prepared to train judgment skills in using a descriptor language, in this case the Intelligence Subject Code, without reference to any particular retrieval system. This material would give new analysts, ones who had completed the Intellofax Program, the needed additional practice in making classification judgments. At present, this practice is provided by a lengthy combination of experience and training in analyzing actual documents under the close supervision of experienced indexers. Such supervision is costly and not necessarily an efficient means for rapidly producing satisfactory classification judgment because documents are coded as they are received, regardless of their relevancy to learning new skills. By providing guided and relevant classification experience, an auto-instructional program can substantially reduce the cost of on-the-job supervision of new personnel and decidedly accelerate the time when an analyst would be a productive member of his organization.

Since the Intelligence Subject Code is a widely used descriptor language, an instructional program which developed skill in classifying information according to its categories is applicable to the training of document analysts working with either automated or manual retrieval systems. In many of these systems, where there is a lack of trained supervisory personnel, the replacement of on-the-job experience by programmed instruction will mean a gain in the accuracy of indexing and consequently more reliable retrieval. The proposed auto-instructional program would provide training for both analysts who work with the Intellofax System and for those who will use the less complex systems which are also based on the Intelligence Subject Code.

-3-

The terminal behavior to be developed by the proposed program is the ability to classify a group of typical documents or document summaries into their correct Intelligence Subject Code categories. Thus, an adequate response in dealing with a to-be-coded document is the listing of appropriate Intelligence Subject Code categories in the form of subject codes, subject modifiers, area relationships, and area codes. This response is, however, only the final act in a long behavioral chain which begins when the analyst reads the document and ends only after many responses, which are usually described as "thinking" or "analysis," have occurred. In this situation, "thinking" implies that the words or context of the document serve as stimuli for covert responses on the part of the analyst. These responses are usually the names of some categories in the Intelligence Subject Code. For example, a document concerning "aluminum" immediately evokes a response on the part of the analyst of a possible classification category for the document. Other words in the document such as "research on aluminum," "foreign sales of aluminum," or "labor force in the aluminum industry" evoke other category responses. As the analyst scans or reads the document, it continues to evoke responses expressed in the categories of the Intelligence Subject Code. All of these responses, including the rejection of tentative categories and the substitution of others, occur subvocally and are not observable in the work of the experienced analyst. He may not even realize that they occur. He "thinks" about the document or "analyzes" it. However, these descriptions of the process used to classify documents are of little aid to the new analyst because of the nature of the Intelligence Subject Code and the abundance of synonyms in the English language.

The proposed auto-instructional program will not only give the student relevant practice in analyzing documents, but it will teach him the analysis process. It will be designed to teach him to "think" in terms of the Intelligence Subject Code. This will be done by shaping his responses to key words and phrases in documents. By breaking down the thought process that occurs in classifying a document, the program will develop his skill in each phase of the process. Then it will build his ability to perform the entire response chain beginning with the reading of the document and ending with its correct classification in the descriptor language.

It is proposed that the research project and the development of the programmed materials be carried out in the following six steps:

-4-

1. Behavioral analysis. Since the staff of the proposed project is already familiar with the contents of the Intelligence Subject Code, the initial development of objectives will emphasize an examination of the actual behavior involved in classifying documents in the descriptor language. During this period, the sponsor will collect fifty representative documents and fifty document summaries. Two skilled analysts will then classify these materials while being interviewed by two members of the project staff. The interview questions will be designed to make observable the thought processes involved in classification and to identify the stimulus dimensions affecting category choice. The interviews will then be recorded. When they are transcribed they will provide a list of behavioral patterns of skilled analysts expressed in terms of significant stimulus conditions and appropriate classification responses (suggested possible categories, indexing rules, rejected categories, etc.).

The project staff will use this list of behavioral patterns as a guide to classifying an additional 50 documents and 50 summaries provided by the sponsor. Any document which cannot be classified by reference to the list of behavior patterns will be referred to the expert analysts. From this extensive observation of actual classification behavior, a list of significant stimuli and appropriate responses will be derived which will form the basic outline of the auto-instructional program. It is estimated that an experienced analyst can probably classify the 50 documents and 50 summaries in about three days. An additional day should be allowed to classify documents which the project staff are unable to categorize.

The collection of the documents and document summaries, their classification by expert analysts, their transcription, the summary of the behaviors patterns, and the classification of the additional documents by the project staff will take approximately one month.

2. Development of the criterion test. The sponsor will select approximately fifteen documents and fifteen summaries from the two hundred that have been classified in Step 1. These will be chosen as representing typical and fairly difficult classification problems. These examples will not be used in the training material, but will form the criterion test to be given after the program to measure its effectiveness. The time required for developing this test is included in the time allowed for the first step above.

-5-

3. Development of the draft program. The programmed materials will be developed from the list of behavior patterns of skilled analysts described in Step 1 above. These patterns will be organized into procedural generalizations of the analytic process. These generalizations will form the outline of the program which will be designed to teach new trainees to use the same analytic procedures used by skilled analysts. The success of the program will be determined by the trainee's skill and accuracy in indexing the fifteen documents and fifteen summaries comprising the criterion test.

In the process of developing the program, draft versions will be tried out on appropriately representative subjects. Twelve or more subjects will be administered the program at this stage in its development, and revisions will be made in the program after tryout results are obtained from each individual subject or small group of individuals. The program also will be revised on the basis of student performance on the criterion test. The program will be designed so that no previous experience in indexing is required. This will be made possible by including an introductory unit. Students who have worked the Intellofax Program or have some on-the-job experience with Intelligence Subject Code will be able to skip this initial sequence and move ahead to advanced materials. By making the program self-contained, it will be applicable to other indexing systems and usable with tryout subjects who have not learned from the Intellofax Program.

Since there are few guidelines to use in preparing auto-instructional materials in "thinking" or "analysis," the present project will, to a large extent, develop new techniques to teach these skills. The final form of the program cannot be definitely described at the present time for it will be determined by analysis of the behavior patterns of the skilled analysts and by experience in student tryouts. It is estimated that the final program will consist of 500 to 600 frames and take from seven to ten hours of actual working time to complete. The final program will be prepared in a format which is suitable for inexpensive reproduction and which is suitable for general distribution to users of the Intelligence Subject Code.

It is estimated that program development will take approximately three months.

4. Twenty-five copies of the revised draft program will be submitted to the sponsor for use in training document analysts. As each student completes the program, he will be given the criterion test to measure the program's effectiveness.

-6-

The completed programs and the test results will be returned to the American Institute for Research for analysis.

5. Final revision. The final revision of the auto-instructional materials will be based on the data from the first field trial. Probable time for revision is one month, after which fifty copies of the final revision will be delivered to the sponsor.

6. Final field trial and report. The sponsor will use the final version of the program for training their own personnel and for training personnel from other departments. In addition, copies of the material will be distributed to departments which wish to train analysts at other locations. It is hoped that these departments will make available to the sponsor and the [redacted] both 25X1 response data and criterion test scores. The final report will summarize the student data collected in both the first and final field trials. In addition, the report will contain the list of behavior patterns developed in Step 1 and a description of techniques used to train classification skills.

The proposed time schedule provides that the initial version of the program will be delivered to the sponsor within four months of the initiation of the project. Although it is difficult to determine the time required for the field trials, it is estimated that the final report will be prepared within ten months of the contract date.

Personnel

The successful completion of the proposed project will depend, to a large extent, on the combination of the experience and competence of the project staff.

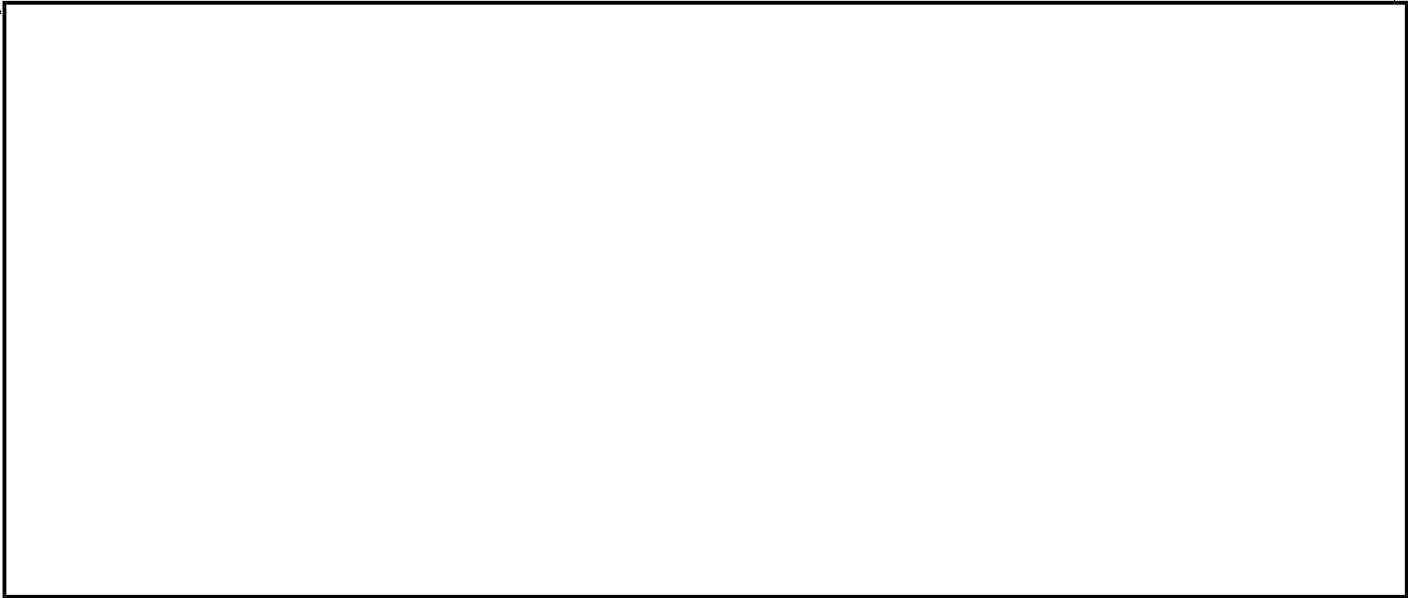
The project will be directed by [redacted] under the supervision of [redacted] 25X1 [redacted] 25X1 [redacted] 25X1 [redacted] 25X1

[redacted] has directed the current United States Government project concerned with the programming of indexing and document analysis skills for use with the Intellofax Retrieval System. [redacted] 25X1

[redacted] September 1962. He has had extensive work in directing the writing, editing, revision, and tryout of programmed materials in many subject areas for all education levels. He has trained Air Force personnel in programming techniques and is currently a consultant to Air Training Command on training problems. 25X1

25X1

- 7 -



Additional research assistance will be provided from among the professional staff of the [redacted] Individuals who participate on the project will be chosen on the basis of their relevant experience and training in psychological and educational research. The large staff of the [redacted] makes it possible to assign individuals to specific projects in such a way as to take maximum advantage of staff capabilities and to meet the needs of the project as exactly as possible.

25X1

25X1

Budget

This budget is for a period of ten months. The staff will be active on the project for six months; the other four months are allowed for tryout of materials. The budget is offered on a cost-reimbursable basis as an extension, or additive task order, to Contract XG-2527.

Salaries, Wages, and Leave

25X1

- Executive Research Scientist (part-time)
- Research Scientist (1/2 time, 6 months)
- Research Assistant (full-time, 6 months)
- Secretarial Assistance (1/2 time, 6 months)

Overhead (58.0% of Salaries and Wages)*

Miscellaneous Expenses

- Paid Subject 20 x 10 hours @ \$1.60
- Communications and Postage
- Reproduction of Reports and Materials

Travel 25X1

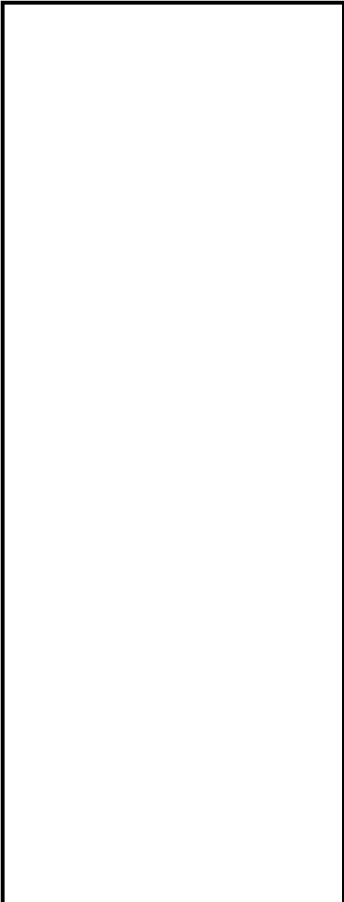
[Redacted]

8 days per diem [Redacted] 25X1

Subtotal

Fixed-Fee (6% of subtotal)**

TOTAL



* This is the suggested provisional rate for the period of the contract, and is subject to redetermination to the actual rate. The cognizant government auditing agency for the [Redacted]

25X1
25X1

** This fee is intended to cover such expenses as research staff conferences; providing a fund for fundamental research essential to our overall program, but which would not normally be included in specific contracts; providing funds for the procurement of new scientific equipment and the replacement of obsolete equipment; and establishing sufficient reserves to meet working capital requirements.

Administrative information

The person to be contacted for additional research details with respect to this proposal is: 25X1



The person to be contacted for additional administrative details with respect to this proposal is: 25X1



Submitted by,



25X1

12 April 1963