

ESD-TDR-64-177

AD-432 098

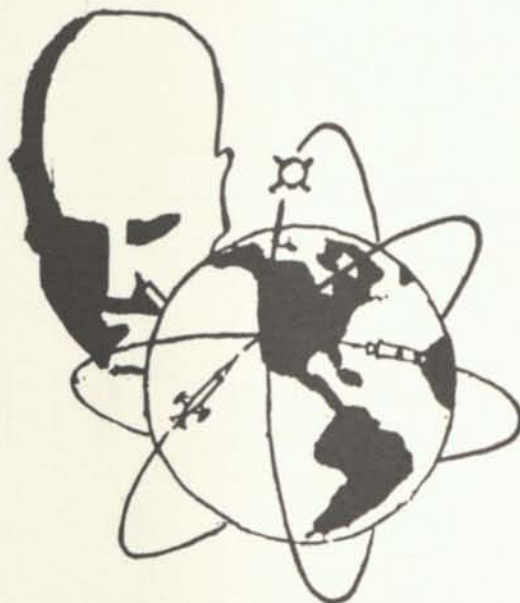
RESEARCH ON COMPUTER AUGMENTED INFORMATION MANAGEMENT

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-177

NOVEMBER 1963

Charles P. Bourne

DIRECTORATE OF COMPUTERS
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts



SRI Project 4506

(Prepared under Contract AF 19(628)2914 by the Stanford Research Institute,
Menlo Park, California)

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.

Qualified requesters may obtain copies from Defense Documentation Center (DDC). Orders will be expedited if placed through the librarian or other person designated to request documents from DDC.

Copies available at Office of Technical Services, Department of Commerce.

ESD-TDR-64-177

RESEARCH ON COMPUTER AUGMENTED INFORMATION MANAGEMENT

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-177

NOVEMBER 1963

Charles P. Bourne

ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L.G. Hanscom Field, Bedford, Mass.

SRI Project 4506

(Prepared under Contract AF 19(628)-2914
by the Stanford Research Institute, Menlo Park, California)

RESEARCH ON THE HISTORY OF THE UNITED STATES

THE HISTORY OF THE UNITED STATES

1877-1878

1879-1880

RESEARCH ON THE HISTORY OF THE UNITED STATES

THE HISTORY OF THE UNITED STATES

1881-1882

1883-1884

1885-1886

RESEARCH ON THE HISTORY OF THE UNITED STATES
THE HISTORY OF THE UNITED STATES
1887-1888

FOREWORD

This report was prepared by Stanford Research Institute, Menlo Park, California, on Air Force Contract AF 19(628)-2914 as SRI Project 4506, "Research on Computer-Augmented Information Management." This project is one of a series of projects in progress at SRI on the "Augmented Human Intellect" program. The principal investigators on this project were Mr. Charles P. Bourne, and Leonard J. Chaitin.

The author gratefully acknowledges the contributions to this project by Tom L. Humphrey, F. K. Tomlin, and Jack Machanik for their development of the basic symbol-manipulation processes used on this project. I also wish to thank Douglas Engelbart for his many useful suggestions and comments.

ABSTRACT

This report describes the results of one of a group of studies on the "Augmented Human Intellect" program at Stanford Research Institute sponsored by the Air Force Electronic Systems Division (ESD), Advanced Research Projects Agency (ARPA), Air Force Office of Scientific Research (AFOSR), and Stanford Research Institute (SRI). The work reported here was supported by ESD, and was concerned primarily with the design and development of an information-management subsystem that would operate in the environment of a cooperative man/computer system for the production of operating computer programs.

An experimental system was developed which utilized the AN/FSQ-32 computer in a time-sharing mode of operation, as well as a cathode ray tube display and associated input-output facilities for real-time machine use by a programmer. A number of symbol manipulation techniques (e.g., insert, delete) developed on the ARPA project were used for the editing and manipulation of text and programming material on the display. Methods of file organization were studied and established for the machine file. Procedures were established for obtaining and using machine-readable records from all of the project records (e.g., memos, correspondence, bibliographies, reports) generated by the user group, to serve as the basis for an experimental machine file. This machine file was established, along with appropriate file search routines, to permit the reading of the natural text of the file material. Studies were made of methods to automatically determine alternate search prescriptions, and display

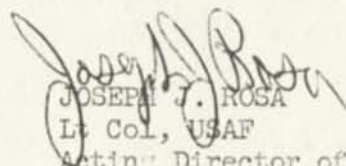
these to the inquiror. Procedures were established to perform some copy editing of the text material (e.g., check for correct spelling). Suggestions are given for additional work to be done in this area.

KEYWORD LIST

Computers
Programming
Information Retrieval
Documentation
Human Engineering
Human Factors
Display Systems

REVIEW AND APPROVAL

This technical documentary report has been reviewed and is approved.


JOSEPH M. ROSA
Lt Col, USAF
Acting Director of Computers
Deputy for Engineering & Technology

CONTENTS

FOREWORD	ii
ABSTRACT	iii
LIST OF ILLUSTRATIONS	vii
I INTRODUCTION	1
II ORGANIZATION OF THE MACHINE FILES (ESD-Supported).	7
A. Expected Use of the Machine Files.	7
B. Private Machine Files for Individual Users	9
C. The General Types of Magnetic Tape Files	11
D. Detailed Structure of the Magnetic Tape Files.	13
III BASIC PROCESS CAPABILITIES (Primarily ARPA-Supported).	17
A. Primary Commands	18
B. Secondary Commands	19
IV HIGHER PROCESSES (ESD-Supported)	22
A. Input Processing	22
B. File Searching	26
1. Discussion of Alternate Methods of File and Index Organization for File Searching.	26
2. The File-Search Command.	29
3. Discussion of Machine Determination, Display and Use of Alternate Search Terms.	31
C. Copy Preparation and Editing	34
D. Data Communication Facilities (ARPA-Supported)	36

CONTENTS (Continued)

V	PLANS FOR FUTURE WORK	37
	A. Requirements for Program Documentation.	37
	B. Current Awareness	37
	C. Retrieval Aids.	37
	D. Copy Editing.	38
	E. Additional Operating Experience	39
APPENDIX I	FACILITIES AND EQUIPMENT	40
	A. Computer Equipment (ARPA-Supported).	41
	1. The Q-32	41
	2. The CDC-160A	42
	B. Work Station (ARPA and SRI-Supported).	42
	1. CRT Display.	42
	2. Keyboard	44
	3. Light Pen.	44
	4. Five-Key Keysets	44
	5. Pushbuttons.	44
	6. Lights	45
	C. Data Preparation Facilities (SRI-Supported).	45
APPENDIX II	SYSTEM LANGUAGES	46
	A. Programming Languages.	47
	B. Media.	47
	C. File Material.	48

ILLUSTRATIONS

Fig. 1	Relationship Between the Users and Various Files.	10
Fig. 2	Representative File Item	15
Fig. I-1	Basic Work Station and Small CRT Display and Keyboard Unit.	43

I INTRODUCTION

For the past several years, Stanford Research Institute (SRI) has been engaged in a research study entitled the Augmented Human Intellect (AHI) Program. This has been a relatively large effort, with support for various aspects from Stanford Research Institute (SRI), Air Force Office of Scientific Research (AFOSR), Advanced Research Projects Agency (ARPA), and the Air Force Electronic Systems Division (ESD). SRI has provided in-house support since 1959, AFOSR has supported a one-man study since March 1961, ARPA has supported about a 5-man effort plus equipment since March 1963, and ESD has supported about a 2-1/2 man effort since May 1963. The intent of this report is to describe the work that has been done over a five-month period on the ESD portion of this joint program. This can best be done by providing the description of the ESD work in the context of the activities of the entire joint program. For that reason, an initial description of the entire program seems to be in order. The delineation of areas of support for various aspects of this total effort is shown by the labels on the major sections of this report.

The SRI role in this program has been to make available some of the necessary capital equipment and facilities (e.g., one of the large CRT displays) as well as to support some continuing study efforts. The AFOSR role has been to support a continuing study of the basic means for the augmentation of human intellect, as well as to develop the framework to show the relationship between the various major problem areas and indicate how a more integrated and coordinated research effort might be pursued.^{1,2} The ARPA role has been to support research aimed at computer

1. Engelbart, D.C., "Augmenting Human Intellect: A Conceptual Framework," Summary Report, Stanford Research Institute, Menlo Park, California SRI Project 3578, Contract AF 49(638)-1024 (October 1962), AD-289 565.
2. Engelbart, D.C., "A Conceptual Framework for the Augmentation of Man's Intellect," Vistas in Information Handling, pp. 1-29 (Spartan Books, Washington, D.C., 1963).

augmentation of a programmer (RECAP). In that regard, the ARPA project has as its objective the development of a cooperative man/computer system for the production of operating computer programs. In that project, real-time computer aids are being developed to improve the productivity and effectiveness of a computer programmer working over the gamut of programmer tasks--conceptual, design, writing, debugging, documentary, maintaining. This includes the development of some symbol-display and symbol-manipulation techniques that are also used on the ESD project. In addition, ARPA has made available the necessary computer facilities [the AN/FSQ-32 computer* at the System Development Corporation (SDC), Santa Monica, California] as well as the data transmission facilities and terminal equipment for the SRI-SDC data link. The general objectives of the ESD project were to develop the information-management sub-system for the ARPA project, using the computers, communication, and display facilities provided by ARPA and SRI, as well as to investigate more general problems of information management, such as the development of systems of hardware, software, concepts, and methods to permit computer-aided humans working in real time to analyze and structure information, to update the information structure, and to retrieve and compile significant information more quickly and comprehensively. The specific ESD objectives were:

- (1) To investigate, design, construct, and document information management subsystems that will operate in the environment of a cooperative, man-computer system for the production of operating computer programs. As used here, "information management" includes: file creation, storage, querying, retrieval, rearrangement, and alteration.
- (2) To investigate such information-management subsystems as (a) components of a system whose purpose is program production; and (b) experimental facilities for providing insight into more general information management processes. This dual point of view implies that an initial information-management subsystem will be made operational with minimum delay, but that significant insights and rejected alternatives will be considered and documented for possible future reconsideration.

* Referred to in this report as the Q-32.

For a continuing study of this sort, a serious effort should be made to consider the whole of the user's information needs and services when using an on-line system, rather than concentrating on any single aspect such as file searching or text editing. However, in order to establish an initial working system for the tests, much of the initial effort was spent on the implementation of a first working system in order to get some experience relevant to its on-line use. Efforts of the ESD project were directed to examining ways in which combinations of people, equipment, concepts, notation, and procedures could be assembled together to permit the human to more directly utilize the real-time computer capability in a way that would make a trained human a more effective manager of working information. Because this general approach was taken, the results of the research can perhaps be extended to other applications such as the development of more flexible information management systems for command and control systems.

The basic method of approach for the ESD project is experimental, with an effort made to obtain, wherever possible and as soon as possible, actual working experience on an operating system. In our method of approach, emphasis is placed on working with our own developments in our every-day tasks. It was felt that the development should be an evolutionary one, with design changes coming as a direct result of the practical experience obtained in the use of this system. It was expected that the programs would be written as quickly as possible, and considered in the same way as experimental laboratory equipment--something to be used to get experimental data, and not a published product. We assume that our programs would always be in a state of change as new features were incorporated in response to experience in the use of the programs.

As mentioned earlier, most of the effort on this initial five-month project was devoted to the design and implementation of an operating system to be used as the basis for some operational testing. Unfortunately, the complete operating system became available at a time that was too late to permit any extensive testing to be done and described in this report. However, much more system testing and experimentation is planned

for the next phase of this work. A more detailed description of plans for future work along these lines is given in Sec. V.

One of the major accomplishments of this project has been the initial planning for the total information subsystem. This has included the design of the computer files and a means for obtaining compatible machine-readable records of project records (memos, reports, bibliographies, etc.) that could be used concurrently with conventional manual procedures for engineering documentation. The machine file design was built around the concept of permitting a user to query and use a computer file in real time, using common files accessible to all users, as well as private machine files accessible only to the individuals that established them. More specific information regarding the organization of the machine files is given in Sec. II.

One other major accomplishment on this project was the design and development of a flexible system for file searching and display that would permit searches to be made on the natural English text of the file items as well as on a subject-heading basis with a more conventional indexing and classification approach. Procedures were outlined but not implemented to permit the computer to suggest alternate formulations of the search requests. This is described in more detail in Sec. IV.

One further accomplishment on this project was the planning and partial implementation of some comprehensive editing routines to be used for checking both input material for the file and material prepared by the user on-line at the work station. These editing routines include such features as spelling checks and text adjustment for line justification. This is described in more detail in Sec. IV.

As a guide to our planning, we found it useful to think of the information-management system as being divided into three parts: (1) information handling for the individual, (2) information handling between the persons within a group, and (3) information handling for the communication between the external environment and the group or the external environment and individuals. Preliminary work was done in all of these areas.

Another way to think of the information management system was in terms of the time constants of various types of information needs, uses, and services. For example, a system to organize, store, retrieve, and display the current and up-to-the-minute copies of an individual's working notes, as well as his team members' comments and reactions to them--would require a relatively fast response time. In contrast, a system for the retrieval of references on a desired subject might not need such a fast response time. It is in the area of needs for fast response times that the real-time systems have so much potential for aiding the human. The information systems with the fast response times become especially effective in those instances in which the human's query to the system must be re-phrased and given again in some modified form as a new query. In a real-time system this can be a relatively rapid man/machine dialogue.

In the present computer-augmented system, it was not envisioned that the user would always have a continual need for on-line computer facilities. In fact it was expected that there would be some instances in which the computer could be used in an off-line mode to work on particular problems such as file maintenance or input editing. Consequently, attention was given to alternate modes of operation for many of the procedures planned for implementation. For example, it would be possible to put a note or memo into the system file by typing it into the console in a real-time manner, or preparing it off-line on a Flexowriter for subsequent entry into the system. Plans were also made for allowing corrections to be made to the file items from either an off-line mode (Flexowriter tape with instructions) or on-line mode (operator making changes at the console).

At this point, there are indications that there may be two types of people associated with the system. There may be one or more working individuals (e.g., programmers, designers, controllers) who are using the system aids to perform useful work, and there may be the system operator (combination editor, file clerk, librarian, programmer) who maintains and operates the system. Not enough experience has been obtained yet to see how the working tasks and responsibilities might be

divided between these two classes of people. Help for both types of people, from the computer-augmented system techniques we are developing, would be of significant value to people in many fields.

With this description of the background and philosophy behind us now, we are in a much better position to describe the specific work performed on the ESD study. This will be done by describing first the elemental building blocks at our disposal (hardware, software, procedures, languages) and then describing how these were assembled into a system of useful operating processes for the user. A description of the facilities and equipment that were used are described in Appendix I. A description of the system languages is given in Appendix II.

II ORGANIZATION OF THE MACHINE FILES (ESD-Supported)

A. EXPECTED USE OF THE MACHINE FILES

From the beginning of the project it was felt that an important feature of this system would be the ability to store the users' notes and working records in the machine file in such a way that they would be easily and rapidly accessible. Consequently, a large portion of the effort on this study was focused on the development and implementation of a file system that would contain various types of information for the users and be organized in such a way that a means for the real-time retrieval and display of requested information would be provided as a natural working tool. It was hoped that this would provide a capability in several ways superior to the previous manual system. It was not expected that the user would rely entirely on the machine file and discard his present manual files, but it was expected that a combination of both approaches would be used.

The initial requirements for the file organization were that:

- (1) it would have to be relatively flexible in order to handle diverse types of material (e.g., text, programs, diagrams);
- (2) it would have to be compatible with the user's manual systems;
- (3) it would have to be implemented very rapidly;
- (4) it could accommodate later changes and modifications;
- (5) it would have to provide a relatively rapid and powerful search facility.

As a test of the utility of such a file, the members of the SRI staff working on this joint research program were to be used as test subjects, with all of their project notes and reports put into the machine file. In this way, the file would have been used as a definite part of our experimental system. Unfortunately, there was little opportunity to get enough operating experience with the files during the course of this brief project to determine what changes or modifications would be desirable.

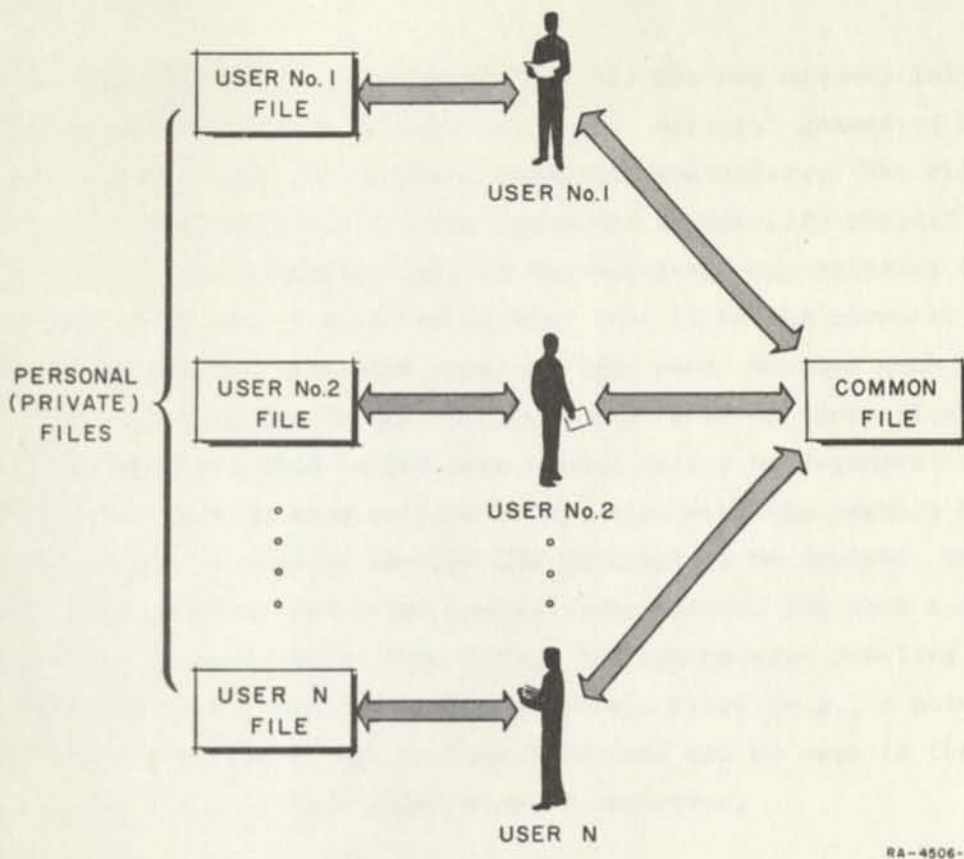
The file was established on magnetic tape for the Q-32. The initial test file consisted of the full text of 7 project memos and reports ("file items") and their associated indexing and descriptive information. This amounted to approximately 10,000 Q-32 computer words and was contained on a portion of a reel of magnetic tape for the IBM 729-IV tape transport with 556 characters/inch recording. At this rate, the single reel of tape could have held about 1400 different 3-page file items.

This test file of 7 items on magnetic tape consisted of continuous text material such as notes to the project file. Another 50 items were recorded on magnetic tape but failed to pass the input editing routine to get onto the test file. Approximately 115 more file items exist in punched paper tape form as a by-product of the regular typing work on this project, and will soon be converted to magnetic tape for the text file. The master bibliographic reference list of over 1200 references for this research program (compiled and transcribed on the AFOSR project) is also in punched paper tape form, waiting to be put in the test file. This means that a total of about 500 typewritten pages of project documentation are presently available in machine language to form the test corpus for subsequent studies. Some programs, with their associated text descriptions, are waiting to be entered into the system. Few diagrams were included since there are still a number of difficulties involved in handling graphic material and special symbols in the machine file and display system. Some diagrams were included by synthesizing them from the typed symbols for dashes, periods, and a few digits. This synthesis was generally a cumbersome operation and was only tried a few times. For most of the graphic material a reference in the display was made to a hard copy of the original illustration that will be kept in a reference collection at the work station. The RECAP project is planning to develop the means for allowing graphic displays (e.g., program flow charts) to be generated directly from the program code or from light pen operations by the user, and then stored in the machine file for subsequent processing or display.

B. PRIVATE MACHINE FILES FOR INDIVIDUAL USERS

The file system was designed to provide both a common file accessible to all potential users, as well as a group of personal (private) machine files that could be set up by the individual users in addition to the common file. There are at least three strong arguments in favor of allowing individuals, at their option, to develop and make use of personal and private machine files that are functionally separate from the common file. Assuming that an individual wants to have his material in the machine file, the first argument is that there may be some material (e.g., preliminary drafts or notes) that might be misinterpreted or misapplied if used in its present state by other people. The second argument is that perhaps the user simply doesn't like to have people looking over his shoulder or into his notebooks. The third argument is that the individual user may want to set up his own method of file organization and indexing (with his own special subject headings); this is awkward to implement with a common data pool because of the problems of separating and keeping track (for input, searching, etc.) of the different indexing terms applied to a particular item by several different people. At this point it seemed to be worthwhile to experiment with a facility in which the individual users could establish and maintain their own personal machine files in addition to their common data bank. Figure 1 shows the relationships between the users and the various files.

An entry in the header of the file item is currently used to determine who receives the initial distribution in printed form; it does not state restrictions on the subsequent access to that item. However, items that are stored in a personal file will normally be accessible only to the author or addressees. Unless otherwise noted, any memo or file item (regardless of who or how many people it was initially addressed to) is accessible to any user, and constitutes what is referred to in Fig. 1 as the common file. Thus a memo from USER #1 to all project members will be in the common file, and so will a memo from USER #2 addressed to USER #3 and USER #4.



RA-4506-1

FIG. 1 RELATIONSHIP BETWEEN THE USERS AND VARIOUS FILES

As an initial step to ensure that access to the private files is restricted to the authors or addressees and which can be made more complicated later if necessary, the search program will require the searcher to identify himself at the time of the search or at the time when he signs in on the machine for a period of continuous operation. Some means will also be provided to permit two or more users to search the private file jointly to recover material that might be in either of their private files. This would allow two people to work together at the console, for example, to assemble material for a joint report--without requiring separate searches to be made by each of them for the same types of material.

For each individual, a personal file for his own use may initially be built up from two types of material: (1) material generated by that user himself; and (2) material received from others. The file-searching programs will ask for the requestor to identify himself and will not display or otherwise call to his attention any material (except that generated by him or directed to him) that is in the personal files of other individuals. Material received from other sources such as general distribution can be put into personal files by addressing the copy to the personal file in the same manner as for self-generated material. This can be done on-line or off-line with the machine files so that there is no need to re-type the material to be copied. The special indexing terms and other special notes useful for that individual's file copy can be inserted at this time. The system also provides a means for sending material to other people's private files (e.g., a note from USER #1 to USER #2 ONLY), but no immediate need can be seen in this experimental system for this particular mode of operation.

C. THE GENERAL TYPES OF MAGNETIC TAPE FILES

Input to the Q-32 tape files comes primarily from paper tape. For future operation, these tapes will be prepared by the present 8-level Flexowriter, or the proposed 6-level CDC Standard tape preparation unit, or the punch on the CDC-160A. Two other sources of input to the tape files are the on-line entries from the work station, and the programs from the program tapes. Regardless of the source of the input, the items will eventually be written on magnetic tape.

Three basic principles have been established which influence our approach and procedures.

- (1) First, there is the idea that you can't (or shouldn't) re-write history. Consequently, the system will operate in such a way that a master record (called a history or archive record) will be maintained, and will include a copy of each file item (in correct form) as it appeared when it was first entered into the system. This history tape cannot be changed or modified. Copies can be made

of items on this tape, modified, and added to the system as an additional entry, but the original item cannot be changed on the history tape.

- (2) Secondly, as mentioned earlier, facilities have been provided that permit users to maintain their own personal machine files, which are accessible only to that individual. This requires the establishment of some special file controls.
- (3) Thirdly, there is the principle common to all magnetic tape files--that up-to-date backup or duplicate file copies must be maintained as insurance copies in the event that something happens to the main files.

One direct result of these principles has been the establishment of three separate tape files, each with different purposes and restrictions on use. The following files have been established. These need not be physically separate files, and may even be stored on the same magnetic tape.

- (1) History or archive file--As mentioned earlier, this file (currently consisting of 7 items) will include a copy of every item that was formally put in the system file either in the common or private files. Items on this file will be accessible to any user, with the constraints that have been established for private file items. This file is stored on a separate tape reel which is file-protected to prevent writing on that tape. It is expected that this file will be viewed rather infrequently, perhaps serving usefully for occasional reference, history, and security storage--but not meant to be an active working file. This is somewhat analogous to the use of microfilm copies of conventional office records within a business organization.
- (2) Working file--In order to permit some file purging, assembly, and modification to take place in order to make the file a more useful working tool, and in order to permit the users to update the first-, second-, and third-draft copies of notes as they work on them over a period of several days, it seems reasonable to provide a working file in which any file item could be changed or deleted or commented on by the users. Not enough working experience with the system has been obtained to indicate how the working file should be established or what controls should be placed on its use. One possible approach would be to permit each user to have his own separate tape, and assume the responsibility himself for maintaining his working file. Another possibility would be to

establish a common tape which would automatically receive a copy of every item that goes into the history file. Other approaches are possible, and more definite procedures will be established as more experience is obtained.

- (3) Private File--There are plans to establish a file for personal or private file items that are accessible only to the authors or the addressees. Copies of these items if formally entered into the system will also be maintained in the history file. The personal items, with appropriate distinguishing marks, may also be stored on the working tape. File maintenance routines have been prepared to periodically re-pack and clean up the working and private files. Under the new time-sharing system (TSS-1) SRI will have access to more than one tape on-line. Consequently, it will be possible to label a record as obsolete, and have it removed during the next clean-up run.

D. DETAILED STRUCTURE OF THE MAGNETIC TAPE FILES

Some of the specific features of the magnetic tape record are described here in order to aid the understanding of some of the system processes, described in a later section, that make use of the machine files. Only the main features of the record are described here.

The first basic thing to be noted about the tape files is that they are constructed on a unit-record basis. That is, each separate file item (e.g., a project report or a memo to the project file) is handled as a self-contained package of information independent of all the other items in the file. The second basic thing is the fact that each file item is accompanied by a "header" which furnishes descriptive and identifying information regarding that particular item. The information in the header is used in many of the file processing operations, as described in later sections. The average length of file item to date has been about two typewritten pages. At this rate, the high-density tapes (IBM 729-IV) on the Q-32 would permit approximately 1400 3-page items to be recorded on a single reel of tape, which is entirely adequate for the needs of this project for some time to come.

The header block consists of the following entries:

SOURCE
TITLE
COPIES TO
SEE ALSO ITEMS
FILE FOLDER
FORMAL DESCRIPTORS
ABSTRACT
TEXT

The header entries are prepared by the author of the file item, and also appear on the printed record that is distributed for the users' hard copy files. A representative example of the use of these header entries is shown in Fig. 2. The SOURCE entry consists of the joint author and date entries, and provides a unique, serialized identification code for each file item. Provision is made for handling multiple items by the same author on the same day, as well as items with joint authors. The TITLE entry is the title of this particular file item. The COPIES TO entry indicates who is to receive hard copies of the initial distribution. It is also used to identify items for the private machine files. The SEE ALSO ITEMS entry provides a reference or citation entry to previous or related file items on the same topic, and establishes one mechanism by which relevant file items on a particular topic can be brought together.* The FILE FOLDER entry is a place for each individual to record the subject headings or descriptive words to classify or index that file item into his private machine file. These headings may have meanings that are perhaps local and unique to his private file. The FORMAL DESCRIPTORS are the subjects headings or descriptors chosen from a master authority list to describe the contents of this file item. These FORMAL DESCRIPTORS will have a standard meaning for all of the users in the group, and will be used to index or classify file items for the common

* A citation tracing program is described later in Sec. V.

SOURCE: LJC 19 JUL 63

TITLE: REVISIONS AND ADDITIONS TO TAPE FILE FORMAT

COPIES TO: DCE, JHW, TLH, DCL, FKT, CPB, JWM, JJB

SEE ALSO ITEMS: LJC 1 JUL 63

FILE FOLDER: EOC WRITEUPS

FORMAL DESCRIPTORS:

ABSTRACT: Further modification to the magnetic tape file format.

TEXT: 1. The first word (WORD 1 in LJC 1 JUL 63) of the file item is to be used by the file programs and will consist of two items of information (n, and m times 1024). The letter n represents the number of computer words used to store the entire file item. At present, restrictions imposed by SDC limit us to the use of file items that are less than or equal to 8192 words. This means that longer file items will have to be broken into smaller items by the file programs. The letter m represents the number of physical records (i.e., blocks separated by inter-record gaps) that make up the file item. There cannot be more than 8 physical records per file item because of the restriction to 8192 words.

2. The first word of the abstract (WORD 85) takes the following form (in octal notation), where A (six octal digits) denotes the number of computer words used to store the entire abstract.

77 --- A --- 4 4 4 4

FIG. 2 REPRESENTATIVE FILE ITEM

machine file. These terms, which will represent a well controlled indexing vocabulary, have not been established yet. The ABSTRACT is a one- to three-sentence summary of that file item, and is useful for some of the later display and retrieval routines (e.g., as an initial response to a file search request) as well as for the initial distribution. The TEXT portion is the main body of the file item (text, programs, lists of references, etc.).

Another group of special markers is used on the magnetic tape in the TEXT portion to indicate the beginning and end of portions of program, or diagrams, or comments that were inserted after the item had been written on the working tape. These markers, for example, allow programs to be extracted from file items even when each of the program steps is followed by extensive text discussion. The markers also permit a file item to be displayed or printed with or without the associated comments, at the users' option. Special markers are also used to permit the user to assign a name (e.g., "rules") to a specified portion of a single file item, and permit that portion of the file item to be called up at some later time by that particular name. This feature (naming) is described in more detail in a later section.

III BASIC PROCESS CAPABILITIES (Primarily ARPA-Supported)

A number of basic processes have been designed and programmed to serve as the basic building blocks for the machine system. It was felt that these basic processes (e.g., elementary symbol manipulation instructions for the displays) would be immediately useful as experimental tools, and would also be used to help construct more complex processes. Much of the effort to develop these basic processes was supported by the ARPA project. These commands are to be initiated by the user at the work station. For the early experiments, the commands will be initiated primarily by typing into the keyboard in conjunction with appropriate light-pen and push-button operations.

Normal execution speeds of all of these commands are currently high enough that the user generally does not have to wait for the equipment to operate. The effect is similar to typing with a good electric typewriter. However, there have been instances in which delays of 5 to 10 seconds have been observed between the time that characters were entered into the keyboard and the time at which they were displayed on the display. This delay seems to be due to the particular time-sharing system now in use (TSS-0) and the other programs that are being run at that time.

During the real-time operation, it will be possible to insert data from the keyboard into any part of the displayed data. A special data input buffer is used and its contents are displayed at the bottom portion of the CRT display. Data in this register can be inserted into the place in the display noted by the light-pen action. The function of each of these basic processes is described below. Their detailed operation will be described later in a report of the ARPA project. More complex processes, utilizing these basic elements, are described in a later section.

The command set presently implemented is divided into two classes: Primary and Secondary commands. The primary section of the program services the display buffer (one of the Q-32 drums) and maintains a record

of the entered characters. None of the changes made to the displayed text or to the command being entered into the command buffer is made permanent until Secondary is called. Secondary is a much longer program which, when called, is read into the computer by the time-sharing executive program. This Secondary program fixes changes made in text by the primary program and executes any command that has been entered into the command buffer. In certain cases, Secondary may execute two commands. The available commands are described below, grouped into their respective classes. All of these commands have been put into operation except for the REPLACE and REMOVE commands which are almost finished.

A. PRIMARY COMMANDS

RESET PLACE--This command provides the capability of changing the position of character entry on the display to a new position designated by the light pen, and is activated by a console button and a light-pen action. After a RESET PLACE command characters entered from the keyboard will be displayed starting from the reset position, replacing any characters which previously were displayed in these positions. If this replacing action occurs in the text area, the changes will be made permanent upon the next call to Secondary. If the replacing action occurs in the command area, the structure of the entered command will be changed. Initiation of the RESET PLACE command will cause an underline to appear under the character position on the display to which the entry position has been reset. Successive RESET PLACE actions with no intervening character entry will have no effect upon the displayed or stored information.

DELETE CHARACTER--This command is used in conjunction with RESET PLACE to delete characters, principally from text. The DELETE CHARACTER command is actuated by a console button in the same manner as a character entry from typewriter or keyset. The character that is presently in the current entry position will be replaced on the display by a backwards slash and upon entry to Secondary will be deleted from the text. No closure of the display occurs until the action is fixed by Secondary.

BACKSPACE--This command is actuated by depression of the BACKSPACE button on the keyboard and is a special case of RESET PLACE wherein the current entry position is backed up by one position. If this command is given when the current position is at the top left of the display, no operation occurs.

CARRIAGE RETURN--Under normal conditions, when the end of a line of the display is reached while entering characters, a new line will automatically be started. If it is desired to start a new line regardless of the current position, the CARRIAGE RETURN key on the keyboard is depressed. As many lines will be skipped as carriage returns entered.

TABULATION--This command is the equivalent of a typewriter tab operation except that at present no possibility of modifying the tabsetting exists. Tabs are set every four character positions up to position 32, the end of the line. Further tab commands will have no action.

HANDSPEC--This command is associated with the five key keysets. It allows specification of whether the keysets are being operated in the left- or right-handed mode. It is actuated by entry of a symmetrical code on the keyset followed by a reset code that is non-symmetrical, thereby specifying which hand is being used.

B. SECONDARY COMMANDS

NAME--The NAME command allows a block of data within a single text item to be assigned a given name if no other block in that text item has previously been assigned that name. Names must be combinations of alphanumeric characters of arbitrary length that do not include spaces or punctuation. Names can be nested in an arbitrary manner within the text item. The same name can be assigned to several text items since names are local to an item. Hence names such as "Bibliography" and "Section 2" may be used to advantage. The command itself requires a pair of light-pen actions to designate the area to be named and a string of characters that constitute the name to be assigned.

DELETE--This command causes the text delimited either by a pair of light-pen actions or by a name to be deleted from the file item currently in the store. The deleted area is closed in and appropriate bookkeeping performed.

INSERT--This command causes insertion of a block of data delimited by either light pen or name, or a block of data entered from the keyboard before the position in text indicated by a light-pen action or at the end of a named block of text. The file item is pushed down to make room for the insertion, and appropriate bookkeeping is performed.

REPLACE--This command causes a block of data delimited by either light pen or name to be replaced by a block of data delimited by either light pen or name or by a block of data entered from keyboard. In effect, the text to be replaced is deleted and the replacing data inserted into the position previously occupied by the deleted text. The replacing text is left unchanged in its original position.

REMOVE--This command is similar to REPLACE except that the replacing data is deleted from its original position. In effect, REMOVE is an overwriting move data operation, whereas REPLACE is an overwriting copy data operation.

OUTPUT--This command allows outputting to the line printer the item that is currently in the core storage of the machine. Formatting of the item is rather crude. That is, tabs are treated in the same manner as on the display. Carriage returns start a new line, but if end of line is reached before a carriage return, a new line is started. Backspaces and underlines are ignored as are upper and lower cases.

SCAN--Two SCAN commands are available, SCANUP and SCANDOWN. As the name suggests, the command causes the displayed data to be scanned in such a manner that the display is like a window through which text is viewed, being moved either up or down by the command.

RESET--This command causes the entire display to be blanked and appropriate program parameters to be reset. Changes made in Primary by means of RESET PLACE and DELETE CHARACTER are ignored by this command.

RESET COMMAND AREA--This command clears the command buffer without affecting in any way the displayed text. No changes to the displayed text made by RESET PLACE and DELETE are fixed, but they are not reset, so that future calls to Secondary can fix these changes.

DO--This is normally used as a command activator at the completion of a command entry. However, if no command has been entered, it will merely fix the changes made in primary by RESET PLACE and DELETE, and return control to the operator with the changes displayed and deleted characters closed in rather than being displayed as backward slashes.

CONTINUE--The SHOW command (to be covered in detail later) and the SCAN commands can be performed in conjunction with other commands by use of the CONTINUE command. This command will perform the command in the command buffer and then perform whichever of the commands--SCANDOWN, SCANUP, or SHOW--has been last entered. In this way a user may scan an entire item, making changes as he goes, without having to alternately enter scan commands and data modification commands. Data may be displayed in any of several modes. That is, names of data blocks can be displayed or ignored, and/or comments (for which a provision has been made but for which no commands are presently available) can be displayed in full or merely be indicated as present. In the existing system, all names and comments are displayed in full.

IV HIGHER PROCESSES (ESD-Supported)

A number of more complex programs have been developed, or are in the process of being developed, that implement larger portions of the total operating system and utilize some of the more elementary building blocks described in the previous section. These programs are being developed on the ESD project to provide the information-management support for the programmer's work (conceptualizing, designing, writing, debugging, documenting, maintaining) within our Augmented Human Intellect Program. It is hoped that the experience gained with the use of these programs will help provide suggestions for more improved versions. As mentioned earlier, the SRI project staff will use these programs as regular working tools in order to obtain some realistic and practical evaluations. Several of these programs are described in more detail below.

A. INPUT PROCESSING

The machine file is a fundamental portion of the information-management system, and a considerable amount of effort has been spent in developing and improving the system and procedures by which information prepared by the users can easily and quickly be put into a machine file for subsequent processing and recall. The major portions of this file input system have been in operation for the SRI staff for several months. With the exception of some file items that will be prepared at the console at some time in the future, essentially all of the file material consists of items (e.g., project memos) that were initially prepared to be used manually. Figure 2 is a representative example of such a file item. One of the design problems on this project was the question of how to provide this material in machine-readable form for the machine files. The procedure put into effect for the present system was to have the items typed on a tape-punching typewriter so that a paper-tape record is obtained from the first keyboard operation. The typing was done on reproduction masters so that multiple copies could be made easily and distributed to

the staff members for their manual files. This paper tape was then used as the primary source of input for the machine files, and was subsequently put through a series of editing and processing operations to put it into correct form for the machine files.

The input processing starts with the transcription of information onto magnetic tape from the paper tape or the console. During this operation the data is transcribed character-for-character onto the magnetic tape without formatting it into the structure required for the file-searching programs. The only operation performed at this time is to convert the character codes from that of the Flexowriter into the code used by the Q-32 programming system, and to write those characters onto the magnetic tape. This step produces the input-editing tape. This process was done on a CDC-160A at the local CDC Service Bureau because the CDC-160A used at SRI on this project did not have a magnetic tape transport. However, another CDC-160A at SRI, with a magnetic tape transport, has recently been made available for this task. Eventually, the transcription will be done by reading the paper tapes on the CDC-160A at SRI and sending the data down to the Q-32 magnetic tapes in Santa Monica via the telephone data link.

The input editing routine uses this first input tape and performs three main tasks:

- (1) It structures the items into the format required for the tape file. This part of the program has been put into operation.
- (2) It performs some proofing and validity checking of the file item header. For example, a check will be made at this point to see if: (a) the correct abbreviations have been used for the months and people's names used in the header, (b) the header entry labels have been spelled correctly and have been prepared in the proper sequence, (c) the number of words in the TITLE or other entries have not exceeded their allowable number. Appropriate error messages are printed or displayed by this routine to indicate the source of any errors noted by the program. This part of the program has been put into operation.
- (3) It performs some copy editing of the file item. This part of the program has been planned, but has not been

implemented yet. Initially, the checking at this point will consist of looking for spelling errors in the header and text material by comparing each incoming word against a table of allowable words that have been accumulated to date, and noting the new entries.

Spelling errors are important in this system because they offer a loophole by which relevant but misspelled words could be missed by the text search program. Words encountered in the incoming file items that had not already been posted to the master word list (i.e., words that appear to have been misspelled) could be displayed to the input editor for on-line editing or printed out for off-line review. This "editor" may be an assigned individual serving as editor for the entire group, or it may be more desirable to have each author edit and clean up his own work--both methods will be tried. At the time of display, the editor will have the option and capability of: (1) adding a particular word to the master word list (for those displayed words which had not been previously noted, but which are spelled correctly and will probably be encountered in subsequent file items) and printing this word for later study and incorporation into the thesaurus of search terms; (2) asking for a display of the specific sentence or place in the text where the noted word was used (in the case of doubtful or obviously misspelled words) so that the correction can be made at that point if appropriate; (3) providing some positive action to indicate that in this instance the word is to remain in the text in its present form, with no action being taken to correct it or add it to the word list (for special cases such as misspelled words used as examples, special abbreviations, and other unusual symbol strings). This spelling check could probably catch many of the spelling errors, but would not be able to catch the words that were misspelled into other allowable forms (e.g., boat misspelled as goat). The allowable word table would be cumulative and made up empirically from the words that are actually used in the file, rather than all possibilities of words that might be used. A preliminary list of about 1200 acceptable words has already been assembled from the text of six file items to serve as the basis for this machine word list.

As mentioned earlier, any errors or discrepancies noted by the input editing routine will be called to the attention of the input editor. Several methods of display are available (e.g., the CRT displays, CDC-160A typewriter, off-line printer on the Q-32, and combinations thereof.) After these discrepancies have been settled, the entries on this input editing tape will then be posted to the history or archive file, as well as to the working or private files as appropriate. New words will have been collected and posted to the master word list. Any corrections or changes (e.g., corrections in spelling or format) must be made on the input-editing tape before it is posted to the history file. Some facility may also be provided for off-line corrections, but this would appear to be more difficult.

As part of the input-editing process, it will be relatively easy to make corrections or alterations to the incoming items in order to correct typing errors, etc. that were noted after the original hard-copy version was prepared and distributed. This would reduce the need to prepare extremely neat and correct Flexowriter tapes, and probably result in faster and easier preparation of input material.

It is expected that this spelling check routine might also be usefully employed as a separate editing routine to be employed when a user prepares a memo, report, or some other document at the console. There have been some thoughts of developing a special macro instruction, COPY EDIT, to call this routine into operation. Additional copy editing chores such as checking to see that quotes and parentheses are always used in pairs, could also be added to the program. This is described in more detail in a later section.

One other program being developed as part of the input spelling check program will accumulate tallies of the number of times that each different word has been used in the file. This tally will be relatively simple to obtain as an addition to the spelling check program, and will be helpful for three main tasks: (1) providing information to help choose which words would be included in the master word table in the event that some control must be exercised on the size of the table,

(2) providing information to decide what terms would be reasonable candidates for the development of a schedule of subject headings for the indexing of the common file, and (3) providing information that might be useful for the users in planning file-search strategy.

B. FILE SEARCHING

During the early planning stages of this study it was felt that the user would eventually need a capability for retrieving items from the machine files that satisfied any of a number of different search criteria such as: (1) specific denumerable terms in the headers (e.g., authors, dates, or report titles); (2) specific words or word groups used in the text (e.g., Q-32, data link, CRT); (3) general concepts (e.g., file preparation, man-machine communication). The development effort for the first programming system concentrated on providing a search system that would completely accommodate the first two types of search criteria and approximate the third. This section discusses file-search questions relevant to our activity and describes the programming system that has been developed to do the file searching.

1. Discussion of Alternate Methods of File and Index Organization for File Searching

As an aid and tool for the user, it seemed to be important to provide a capability to locate file items on the basis of specific words that were used in the text of the document or its header. The search criterion might be a single word, or even groups of words with specified logical and syntactical relationships. There are basically two different approaches to locating the file items that satisfy such a search criterion: (1) at the time of the search request, scan the entire text of all the file items to see which items use the specified word or words; (2) do some pre-processing in advance of the search requests, to prepare word indexes to the file items, and then search the indexes at the time of the search request. The second method results in faster searches, but requires more complicated preprocessing. There are several ways to implement these two different approaches, each with differences in

logical complexity, search response time, file maintenance problems, and other characteristics. It is of interest here to briefly describe a few of these approaches to show some of the choices that are available.

- (1) Scan the Complete Text in Serial Order--With this approach each incoming file item is added to the end of a master tape. No indexes are maintained, and no further input processing is required. A search is made by successively reading batches of items into core memory and doing a word-for-word search on the entire collection.
- (2) Scan an Abbreviated Text in Serial Order--With this approach, a search is made by reading the entire collection of abbreviated items into core memory and doing a word-for-word search using an appropriately abbreviated version of the text material, and each of these ways requires some pre-processing of the file item. One simple way is to represent the text by a list of the more significant words that occurred in the text (i.e., exclude common words such as a, the, is, at, be).

Another way would be to systematically abbreviate each word to a specified size (e.g., truncation to six characters). A further condensation could be obtained by using a coded representation of each word. A number of coding schemes are possible, for example: superimposed codes, products of prime numbers, and numeric word equivalents. However, the input processing for coded representations would be further complicated by the requirement for the development, maintenance, and use of coding tables or algorithms.

The actual word list may or may not be included with the coded equivalent and text material, depending upon the degree to which the coded expression can unambiguously describe the word list. That is, the coded expression may be used as a fast first screen to decide when it is necessary to scan the entire text for that particular item. The full word list, abbreviated word lists, or coded word lists can either be stored with the full text of the respective items, or kept separate in serial order to correspond with the serial order of the file items.

- (3) Examine a Composite Word Index Separate from the Text--This approach is the one taken by most reference retrieval programs. Each memo or file item is given a unique serial number and stored in serial order on a

text tape. Separate records are maintained on a drum or another tape to record the location of each significant word that has been encountered in the entire collection of text material. Each of these separate ("term") records includes the serial number of every memo or file item that used that term, and thus acts as an index to show the numbers of the documents in which each of these words was used. The input operation consists of adding the text of an incoming item to the end of the text tape, and posting the serial number of that item to the term records of every significant term that was used in that document. Searching is done by going to the term record for the word in question, and obtaining the serial numbers of the relevant documents. The serial numbers are then used to locate the actual text material on the text tape.

An examination of the search speeds of these methods and their variations showed that the index search (Item 3, above) provided the most rapid response, but that the times for all of the methods were fast enough, for the size of our potential collection, to be considered for our use. It was estimated that all of the methods, for example, could perform a complete search of our file in less than 30 seconds. Consequently, in keeping with our philosophy of trying to put a system into operation as soon as possible in order to obtain useful operating experience, the decision was made to implement the simplest (and slowest) approach of serial text scanning, while considering the use of one of the faster approaches at some later time. Regardless of whether or not a more effective index search program is developed, the user will still have need for a text-searching capability in any future system. It may turn out to be useful to have both the text-scanning and index search method in operation, with the index providing a first screening to answer some questions, and the text-scanning approach used as a backup to answer more complex questions not answered by the index. Since different methods may have different speeds for different search criteria, it might even be fruitful to have more than one different search technique operating at the same time (i.e., in direct competition) to answer the query in a time that is generally shorter than can be achieved with a single search technique.

2. The File-Search Command

A general search instruction has been developed that meets the initial requirements of: (1) being relatively easy for the user to generate, read, and interpret, (2) being compatible with the general RECAP programming system and conventions, (3) being relatively easy to implement in several stages, in a logical sequence, and (4) providing many of the initial features and capabilities that were planned for the information subsystem. The instruction is typed into the keyboard by the user and the program then searches the tape (e.g., history, working tapes) that is mounted on a designated transport. This command has been put into operation.

The instruction generally indicates what is to be displayed, describes the search field and search key, and notes any logical delimiters (e.g., AND, OR) that might be used as part of the search prescription. The instruction takes the following form:

```
SHOW ( A FROM B " C " D .)
```

The A (display) part of the instruction tells the program what portion of the selected file items is to be displayed. Allowable designations are SOURCE, TITLE, HEADER, ABSTRACT, TEXT, CONCORDANCE*. The SOURCE display will show only the SOURCE entry. The TITLE display will show the TITLE as well as the SOURCE. The HEADER display will include all elements of the file item except the ABSTRACT and TEXT. The ABSTRACT display will show the ABSTRACT as well as all the other parts of the header. The TEXT display will display all of the header and text material. The CONCORDANCE display shows only the first sentence in the text that satisfied the search criteria. This CONCORDANCE display provides a convenient means for the editor to use misspelled words as search keys (i.e., words noted during the input editing process) in order to locate the specific sentence in which they were used, so that corrections can be made at that point if appropriate.

* Refer again to Fig. 2 for an example of the order and relationship between these header entries.

The B part of the instruction tells the program what portions of each file item are to be examined during the search process. The search fields that may be designated are SOURCE, TITLE, FILE FOLDER, FORMAL DESCRIPTORS, ABSTRACT, and TEXT. The proper designations of the search field by the user can reduce the amount of machine processing required. For example, the search for the file items by a particular author needn't examine the text of the file item, since the SOURCE entry is sufficient. There are also plans to provide a means for designating a more inclusive field in which all of the search fields noted above would be examined during the same search.

The C part of the instruction provides the search key for the program. This search key can be a continuous symbol string of any length (e.g., one or two text words, or even a complete sentence; or perhaps a particular instruction in a program listing).

The D part of the instruction will tell the program what search logic is to be used, such as AND, OR, NOT, GREATER THAN, LESS THAN. This permits more complex forms of questions to be posed to the machine file (e.g., items with dates LESS THAN a certain number). The present search program has not implemented this D part yet, and is searching only on equality of symbol strings. Present plans call for this logic portion to be developed at an early date.

After the user initiates the search, the first item that satisfies the search criterion is displayed on the CRT in the prescribed manner. After viewing it, the user can press a "continue" button, which will cause the display of the next item that satisfies the search criteria. This stepping to sequential selected items is continued until the end of the file is reached, at which time an appropriate message is displayed.

A few examples should illustrate how this search instruction might be used.

<u>Instruction (as typed on the keyboard)</u>	<u>Interpretation</u>
SHOW (HEADER FROM SOURCE "TLH".)	Display the HEADER of all file items written by TLH.
SHOW (TITLE FROM SOURCE "JHW".)	Display the SOURCE and TITLES of all the file items written by JHW.
SHOW (TITLE FROM TEXT "RECAP".)	Display the SOURCES and TITLES of all the file items that use the word RECAP in the text of the file item.
SHOW (SENTENCE FROM TEXT "CAMPUTER".)	Display the one or more sentences in the file that use the word CAMPUTER.

The direct searching of natural text material has uncovered several problems, mostly in file searching, that had not been anticipated. These problems arise primarily because of the presence of special characters, (e.g., carriage return, tab, hyphen, upper and lower case shifts, backspace and underline) in the symbol string. These problems have been sidestepped by most of the other workers in this field by their use of special punched-card input procedures instead of taking the records that result from the natural typing of the text material. Procedures are being worked out to resolve or minimize some of these problems.

3. Discussion of Machine Determination, Display, and Use of Alternate Search Terms

The present file-search program has the capability to search the file for a specific word, or for any of several words if they are specified. This is expected to be useful in practice. It would perhaps be even more useful if some programming features were incorporated that would automatically generate alternate search instructions for the user after he had given an initial search key.

The need for this expanded capability was noted while performing a recent series of trials in the use of the Multiple Instantaneous Response File (MIRF) which has just been built by SRI and delivered to the Air

Force at the Rome Air Development Center.* The MIRF machine required an operator to type search terms (descriptors) at the console, and these would correspond to instructions for the search of the file (a large content-addressed memory). It was found that some descriptors gave responses for the singular form of the descriptor, but not for the plural form, and vice versa. There were also some problems with words that had common roots and meaning, but different word endings. This is of interest here because of the similarity of operation by the user in both the MIRF and the RECAP information management system. For both system it is unreasonable for the user to have to think of, and type in, the plural forms and other variants in spelling for every search key. This is perhaps less of a problem with the MIRF since it restricts itself to the use of a fairly well-defined and established descriptor vocabulary. This is not the case with the RECAP system in which the actual text material uses a great variety of words with no particular constraints. Present plans are to build some mechanism into the RECAP program to automatically, at the time of search, relieve the user of some of the burden of determining the variations in spelling and word form, and keying these forms in as alternative searches if desired.

* The final report to describe the delivered MIRF is currently being written and thus is unavailable. However, a description of some of the earlier work is given in Refs. 3,4, and 5, below.

3. Frei, E. H., and J. Goldberg, "A Method for Resolving Multiple Responses in a Parallel Search File," IRE Trans., PGEC-10, 4, pp. 718-721 (December 1961).
4. Goldberg, J., et al., "Multiple Instantaneous Response File," Final Report, SRI Project 3101, Contract AF 30(602)-2142. Stanford Research Institute, Menlo Park, California (August 1961), RADC-TR-61-233, AD-266 169.
5. Goldberg, J. and M. W. Green, "Large Files for Information Retrieval Based on the Simultaneous Interrogation of all Items," Large-Capacity Memory Techniques for Computing Systems, pp. 63-77 (The Macmillan Company, New York, 1962).

There are basically three types of tasks for such a program:

- (1) Store and maintain a file of related words
- (2) Display the related words as soon as a search word has been keyed into the system
- (3) Automatically construct search commands, and execute file searches at the user's option, using the alternate word forms.

These actions can be done in varying degrees of detail, with different amounts of user intervention. The lists and groupings of related words should be based on the words that are actually used in the machine file, and modified and updated on a continuing basis as a part of the input processing operations.

There are a number of ways in which the words could be associated together (some ways are more complex than others). Some of the kinds of groupings and capabilities that might be provided are suggested below. They are presented in the order of what is felt to be an increasing difficulty of implementation.

- (1) Noting word plurals (e.g., computer, computers; capacity capacities; goose, geese; thesaurus, thesauri). The rules for forming plurals appear to be complex, with many special cases. Hence it may be necessary to use a table look-up operation rather than an algorithm to determine the plurals.
- (2) Noting permissible variations in spelling the same word (e.g., Q-32, Q32; FEB, FEBRUARY; programming, programing; high speed, high-speed; mag tapes, magnetic tapes). Strict input editing could eliminate these variations by insisting on a uniform way of spelling, but it would seem to be easier and more flexible to let the program keep track of the variations. This will have to be a table look-up operation.
- (3) Noting variants in word endings (e.g., compute, computer, computers, computed, computation). The user may accomplish the same thing in many cases by using only the word root (e.g., comput) as the search key.
- (4) Noting synonyms or synonymous phrases (e.g., sun, solar; memory, storage; search, quiz, interrogation). Words that are used by the project staff in a synonymous manner can be noted and grouped together as alternate search terms. This will also require a table look-up operation.

- (5) Noting generic relationships. It is possible that some of the text words will be related in a generic way. That is, some will be a more general expression for other collections of words, or a less general expression and sub-set of another word. For example, the word "Q-32" might be noted as a term less general than the word "computer." Hence, the keying in of the word "Q-32" could cause display of the instruction, SEE ALSO COMPUTERS (i.e., a reference to a higher generic level). In the same way, the keying in of the word COMPUTER might also cause display of the instruction, SEE ALSO Q-32, CDC-160A, 7090, B-5000 (i.e., a reference to a lower generic level with more specific terms). We will not have a real feel for the importance of generic word associations until a word list is made up from the file material, and some experience is gained in the operational use of the regular file search facility.

The text material in our file is currently being studied to develop the word groupings of the first four types described above. Approximately 500 words have already been arranged in pairs, triplets, or larger groups of alternate forms. Programs are being written for such tables of words to be assembled and stored with the file-search program so that associated words can be automatically displayed to the user at the time that a search command is keyed into the system. At the time at which the display of alternate search terms is presented, the user will be able to specify which, if any, of the alternate terms are also to be taken by the program and assembled and executed as search terms.

C. COPY PREPARATION AND EDITING

As a part of their total operating system, the users need some aids to help them communicate their results and findings to other interested parties. This communication quite often takes place via formal written material such as project reports. Consequently, we feel it important for the information-management system to provide some assistance with this reporting function. Some aid in the assembly of source material for the reports is obtained by using the file search procedures. Some aid is obtained in the editing (e.g., rephrasing or modification of a sentence, or rearrangement of sentences within a section) by using the basic display

and symbol manipulation processes that are available. And as we mentioned earlier, it seems quite possible that the routine for checking the spelling now being developed for the input-editing task could also be used, on-demand, to edit report copy that is prepared on-line at the console. Several other useful copy-editing features could be added with a reasonable amount of effort. For example, it would be possible and relatively easy to include a test to see that all parentheses and quotation marks are used in pairs. It would not be difficult to include a line-justification routine to count the number of letters and spaces in a line, and adjust the spaces so that all the right-hand margins of the printed text are justified. That is, there is always a printed character at the extreme right-hand edge of every full line of text, in the same manner as the following paragraph.

This simple routine would generally result in improved composition of the printed report. As a further aid to the report composition, it would also be possible to incorporate a program to perform proper hyphenation and division of words between lines. However, the few programs that have been written and used for this purpose by commercial publishers have been rather complex. Consequently, this may be a feature that needn't be considered for the immediate system.

Finally, an effort is underway to provide a means by which the assembled material in the machine file can be printed out in a form that is of high enough quality to be distributed as an organization report. This particular report, for example, could have been typed out by the typewriter on the CDC-160A computer, with the manual entry of page numbers, the drawing of some figures, and the insertion of a few special characters and symbols. We need to have punched paper tape as a by-product of our regular typing operations, but have found that the quality of the Flexowriter printing is generally inferior to that of the IBM typewriter on the CDC-160A. Consequently, we developed the programs to list the Flexowriter tape on the IBM typewriter on the CDC-160A in order to get a higher quality printing. This would seem to be an around-about

way to get a printout from a paper tape, but it is being used as a prelude to more extensive machine operations such as line justification and hyphenation, and text editing on the basis of notations placed in the original text at the time of first typing.

D. DATA COMMUNICATION FACILITIES (ARPA-Supported)

In mid-October, a data-link facility was installed to connect the CDC-160A computer at SRI, Menlo Park to the Q-32 computer at SDC, Santa Monica, over a distance of about 400 miles. This data link facility included: (1) a CDC-8061 Digital Communications Terminal, Model 8523-B; (2) an AT&T type 201A Data set; and (3) an AT&T voice bandwidth communications circuit for both voice and data transmission. This facility permits transmission either way at the rate of 2000 bits per second between the CDC-160A and Q-32 computers. This will permit the work station to be operated at the SRI laboratories, but using the SDC computer facilities.

V PLANS FOR FUTURE WORK

A number of tasks have been identified for more serious attention during the next phase of study. The relative priorities for these tasks will be established in the near future as more experience is obtained with the tools that have been developed to date, and as we gain a better appreciation of the relative importance of these tasks to the total task of developing a total information system for the augmented programmer.

A. REQUIREMENTS FOR PROGRAM DOCUMENTATION

Some further investigation must be done to determine ways to improve the documentation of computer programs, both for the developer of a particular program, as well as any other person who may have a need to use or modify that program. Consideration must be given to such questions as: "What data is to be included in a program description, and in what detail?" "What tests can be made to determine what information is useful to include in the program write-up?"

B. CURRENT AWARENESS

Some attention must be given to the problem of keeping the workers better informed concerning both the work that their team members are performing, and the work being done outside the group. Some means must be provided (e.g., a selective dissemination of information system) to call relevant citations, abstracts, or articles on work done outside the group to the attention of the users within the group.

C. RETRIEVAL AIDS

A considerable amount of effort can be spent in this area. Some of the specific sub-tasks that can be suggested are:

- (1) Obtain a better indication of what the specific retrieval needs are for the working programmer.
- (2) Continue the present work on thesaurus development to include forms (e.g., word pairs or phrases) other than single words.
- (3) Establish a schedule of formal subject headings for the entire file collection (i.e., an authority list).
- (4) Expand the file-search program to include the features for specifying logical relationships between the search terms.
- (5) Expand the file-search program to allow the search criteria to be expressed in terms of word proximities (e.g., the two search terms have to occur together in the same sentence in order to satisfy the search).
- (6) Expand the file-search command to include grammatical or syntactical relationship of the text words.
- (7) Expand the file-search program to consider more of the word association possibilities and associative links that can be established.
- (8) Develop a program to perform a citation tracing and display of the relationships between and genealogy of the file items as indicated by their references to each other.

D. COPY EDITING

The procedures for copy editing can be expanded both for the checking of material that is being put into the file and for material that is being written on-line and intended to be printed out for delivery to interested parties outside of the system. Some of the specific tasks that can be suggested are:

- (1) Continue and improve the spelling check program.
- (2) Develop a program to check to see that quotation marks and parentheses are used in pairs.
- (3) Develop a program to perform justification and perhaps word hyphenation for material to be printed for external use.
- (4) Continue the development of the symbol-manipulation techniques so that words, sentences, and paragraphs can be easily modified or repositioned within a body of text.

- (5) Improve the means for handling special symbols and other graphic material in the machine language and display. This includes providing a means for determining what special symbols are recorded (e.g., carriage return, case shift) but not displayed.

E. ADDITIONAL OPERATING EXPERIENCE

In keeping with the basic method of approach established early in this project, we hope to obtain as much operating experience as possible during the next phase of study. This will provide us with the information necessary for the continued evolution and improvement of the total system. In addition to testing each of the specific programs and procedures that have been established, a number of exercises will be used to simultaneously test many of the features of the total system.

FACILITIES AND EQUIPMENT

The structure... facilities and equipment... (faint text)

Appendix I

FACILITIES AND EQUIPMENT

The structure... facilities and equipment... (faint text)

Footnote text at the bottom of the page.

Appendix I

FACILITIES AND EQUIPMENT

The equipment associated with this system is of four types: (1) computer equipment; (2) console or work station; (3) data preparation facilities; (4) data communication facilities. Each of these groups is described briefly in the following sections.

A. COMPUTER EQUIPMENT (ARPA-Supported)

1. The Q-32

The Q-32 computer at System Development Corporation in Santa Monica is the computer used for most of the studies in this experimental system. This is a large high speed computer with an extensive amount of core and auxiliary memory.* This system has been available to SRI at limited intervals on a time-sharing basis, with other SDC programs running at the same time. The particular SDC time-sharing system used for our testing (TSS-0) provided some restrictions to the SRI programs, such as: (1) limiting the amount of available core memory to 16,383 words out of the 65,000 words possible; (2) providing limited drum memory; (3) limiting the number of tape transports to one in an on-line mode, but any number in an off-line mode. The SRI console and display unit was installed and connected directly to the Q-32 at SDC for the major part of this project. The SDC schedules permitted the SRI unit to be used with the time sharing system for approximately 16 hours per week before September 30 and then 10 hours per week. At all other times it was disconnected from the Q-32. Future experimental work will take advantage of an improved time-sharing system now being put into operation at SDC.

* The Q-32 uses a 48-bit word size and has 65,000 words of core storage, 1/2 million words of drum storage, 18 magnetic tape transports, and a 2.75- μ sec execute time for most of the basic instructions.

2. The CDC-160A

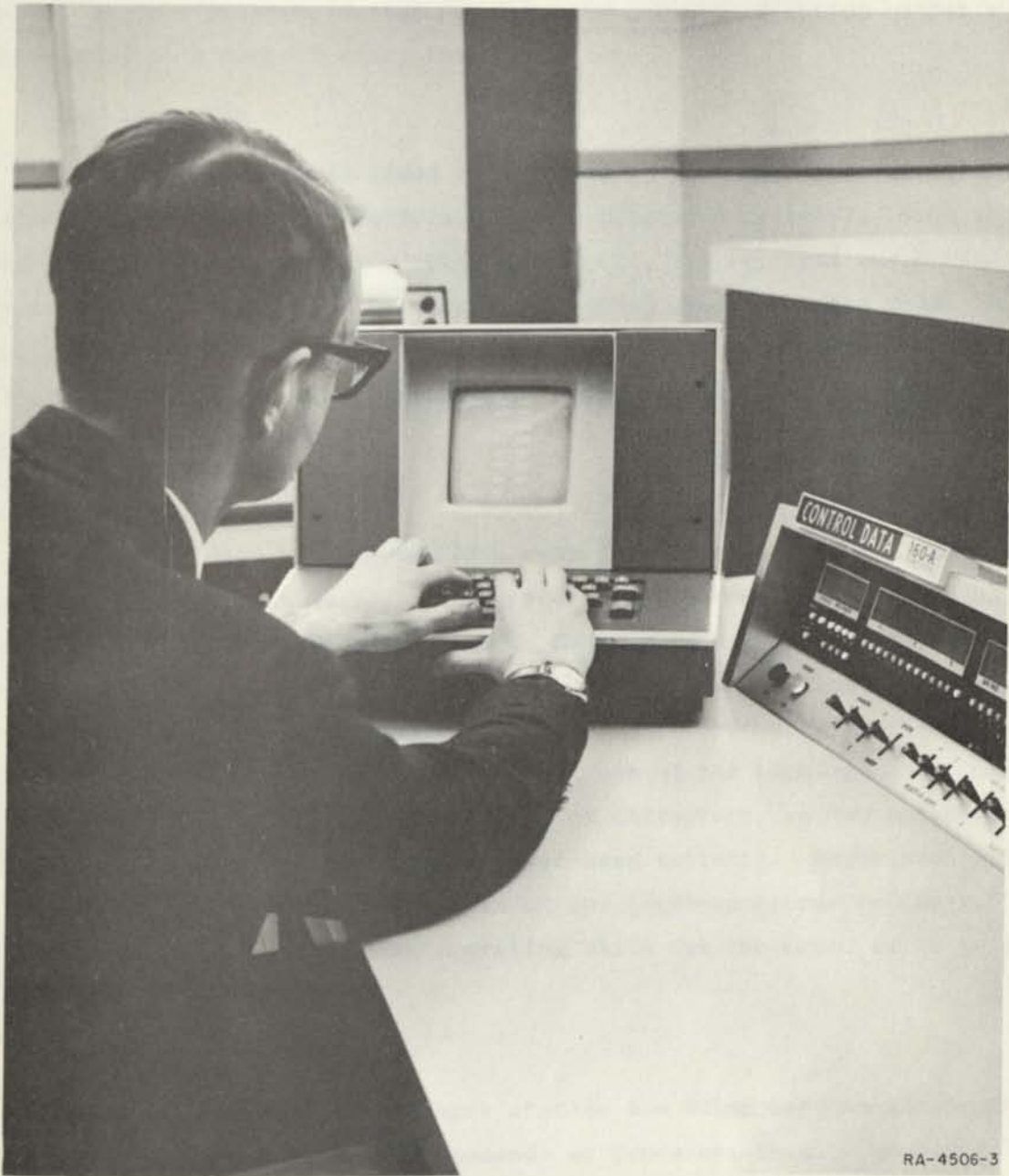
The CDC-160A computer was installed at SRI in Menlo Park, California, to serve as a minimum computer and display system. It also works as a satellite computer to the Q-32 to permit the work station to be operated at SRI via a telephone line data link to the Q-32. The CDC-160A will maintain the displays and handle other small local jobs, but will call on the Q-32 for some of the more complex tasks such as file searching. The CDC-160A is a relatively small system and includes a paper-tape reader, a paper-tape punch, and an electric typewriter.

B. WORK STATION (ARPA and SRI-Supported)

The work station (see Fig. I-1) is a console or desk type of station, with a cathode ray tube (CRT) display and a number of input-output facilities as described below.

1. CRT Display

The Control Data Corporation CDC-220 Computer Display Unit is used. It is a seven-inch-diameter CRT display, and presently displays 32 lines of alphanumeric characters with 32 characters per line. The present character set consists of upper case alphabetic characters, the digits 0-9, and a few special characters (. / \ * \$ + - =). The display is driven at a sufficient repetition rate to avoid any flicker. This system also includes a basic character generator that can be used to drive several displays simultaneously. A larger display tube (16-inch) is on order, and will be used at the work station in conjunction with the smaller display. This larger unit would permit a display of text material that approximates a full page of typewritten text, but would have the same character set since the same character generator would be used. Steps are being taken to increase the character set to include some extra characters (, ' ÷ () ?) so that the display includes the full set of characters for both the JOVIAL and SCAMP programming languages. Consideration is also being given for means to show a difference between upper and lower case characters on the display.



RA-4506-3

FIG. I-1 BASIC WORK STATION AND SMALL CRT DISPLAY AND KEYBOARD UNIT

2. Keyboard

A modified typewriter keyboard is present as a regular part of the CDC display unit. This keyboard unit enters data directly into the computer, where a programmed response can give a keyboard action either a data-entry or a control-entry interpretation.

3. Light Pen

A light pen has been added to the display unit to help specify and delimit operations to be performed on the displayed material. When the light pen detects light at a point on the CRT, the vertical and horizontal coordinates of that point (a total of 12 bits) are fed to the Q-32. A raster of position points is normally displayed on the CRT for use with the light pen. Preliminary experience has indicated that the light pen is somewhat awkward to use.

4. Five-Key Keysets

A pair of five-key keysets have been incorporated into the console as experimental input means. Binary coding of chorded hand strokes on a single keyset provide 31 input codes. Case shifting can provide as large a character repertoire as desired. The main reason for inclusion of the keyset at this time is to obtain one-handed keyboard-entry facility so the other hand can be free for concurrent use of the light pen. Either hand alone can input the complete range of characters, so two keysets gives considerable flexibility for other-hand activity. Experimentation with this keyset is going on as part of our AFOSR-sponsored projects. Two people currently have some operating skill (on the order of 20 to 30 words per minute).

5. Pushbuttons

Eleven pushbuttons on the work station are being used to interrupt the Q-32 and initiate specific commands or processes (e.g., "continue," "clear," "halt").

6. Lights

Four lights on the work station are used as indicators that can be switched on and off under program control by the Q-32.

C. DATA PREPARATION FACILITIES (SRI-Supported)

A Friden 8-level Flexowriter is used at SRI to prepare paper tapes for input to this system. It is capable of producing a tape from a key-stroking operation, of reproducing tape, and of providing a printed output from paper tape. Some consideration is being given to obtaining a 6-level tape preparation unit that is compatible with the code for the CDC-160A.

Appendix II
SYSTEM LANGUAGES

Appendix II

SYSTEM LANGUAGES

A. PROGRAMMING LANGUAGES

In order to effectively use the Q-32 in the present real-time time-sharing mode of operation the SCAMP programming language was chosen. Both JOVIAL and IPL-5 were considered but rejected in favor of the SCAMP language. This present SDC time-sharing system (TSS-0) was established just to give SRI initial on-line access to the Q-32, and no attempt was made to provide a generalized or universal system. SDC is currently developing a newer time-sharing system (TSS-2), available in November 1963, that will also permit JOVIAL, IPL-5, and other programming systems to be used.

The CDC-160A programs have been written in the OSAS Assembly System. Most of the programming to date has been done for the Q-32 system.

B. MEDIA

A number of different means are available for the storage and communication of operating programs and file material (paper tape from the Flexowriter or CDC-160A, magnetic tape, punched cards, internal memory of the Q-32 and CDC-160A, CDC keyboard and display unit; printed pages from CDC-160A typewriter, Flexowriter, and the Q-32 line printer). Each of these media generally has some differences from the others in character sets and codes. For example, the CRT display and the Q-32 line printer only provide a representation with upper case letters, whereas all of the other media can represent both upper and lower case characters.*

* Punched cards are a special media for intermediate storage that can carry the necessary shift codes to represent both upper and lower case characters.

The paper tape from the CDC-160A uses a different code (6-level) than the Flexowriter tape (8-level). Some symbols are available on the Flexowriter (e.g., underline, colon) that presently have no equivalent on the CRT display. For reasons such as these it was necessary to establish a "system code" as a common standard for the system, and translate all of the other codes into and out of this system code.

In addition to differences in notation and character sets, there are also some other differences, such as transfer speeds and print quality. For example, even though the Flexowriter and the IBM typewriter on the CDC-160A have essentially the same operating speeds and character sets, the IBM typewriter definitely provides a much better quality of printing--both in terms of type font design and in registration and impression of the printed characters. Consequently, machine-readable material that is to be prepared for distribution to parties outside the project team (e.g., project reports and correspondence), and which should be in a somewhat polished form, would be candidates for the CDC-160A typewriter, whereas internal project notes could be prepared on the Flexowriter. For extensive amounts of rough copy (e.g., report drafts), the Q-32 line printer might be perfectly adequate, even with its limited character set.

C. FILE MATERIAL

There are basically three types of data that were expected to be handled in this system: (1) text, (2) programs, (3) diagrams. There will, of course, be combinations of these three. Procedures have been developed to store, process, and display a considerable amount of text material in the present system (e.g., project memos, working notes, bibliographies, project reports, and some correspondence.) And since this project is aimed at helping the programmer, procedures have also been established to store, process, and display programs. There are provisions to permit extensive mixing of program and text material so that lengthy comments, examples, and other information can accompany and be inserted in lists of

program coding, but still be passed over in the event that the coding needs to be examined or executed without the accompanying text. Special markers are used for control purposes to delimit such things as the beginning and end of a piece of text, inserted comments, and programs, as well as to note special names assigned to the file item by the user.

INSTITUTE

Library of Congress

Department of Education

Office of Education

Division of Education

Section of Education

Unit of Education

Subunit of Education

Item of Education

File of Education

Volume of Education

Issue of Education

Page of Education

Section of Education

Unit of Education

Representative

Section of Education

Unit of Education

Item of Education

File of Education

Volume of Education

Issue of Education

STANFORD
RESEARCH
INSTITUTE

MENLO PARK
CALIFORNIA

Regional Offices and Laboratories

Southern California Laboratories

820 Mission Street
South Pasadena, California

Washington Office

808-17th Street, N.W.
Washington 6, D.C.

New York Office

270 Park Avenue, Room 1770
New York 17, New York

Detroit Office

1025 East Maple Road
Birmingham, Michigan

European Office

Pelikanstrasse 37
Zurich 1, Switzerland

Japan Office

c/o Nomura Securities Co., Ltd.
1-1 Nihonbashidori, Chuo-ku
Tokyo, Japan

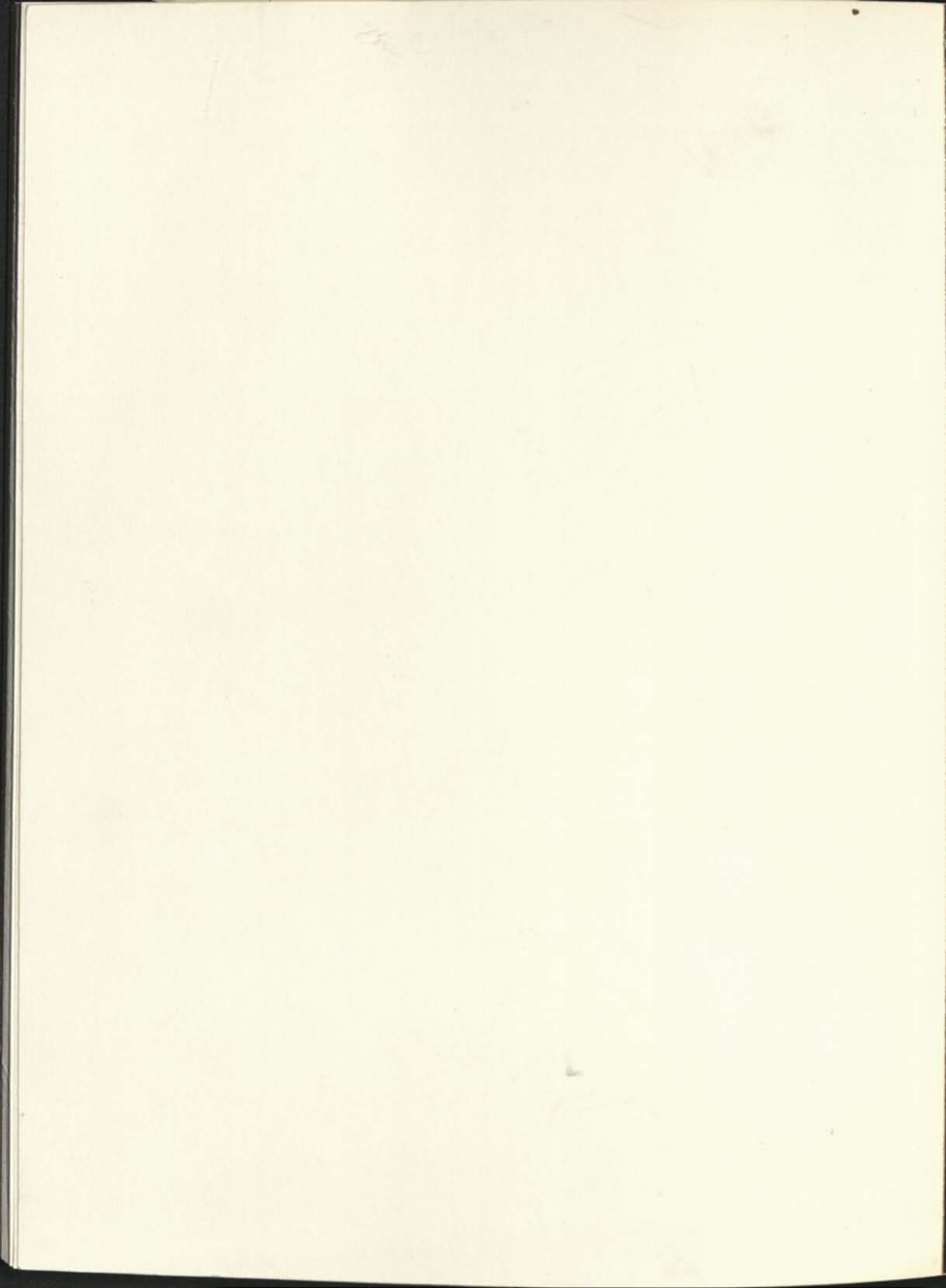
Representatives

Toronto, Ontario, Canada

Cyril A. Ing
Room 710, 67 Yonge St.
Toronto 1, Ontario, Canada

Milan, Italy

Lorenzo Franceschini
Via Macedonio Melloni, 49
Milano, Italy



To: Torben Meisling

Date: 7/21/66

From: Pat Conley

Location:

Subject: Projects Relating to Man-Computer Program

Answering:

cc (43 project (1))
MIT - Project MTC
ESD
RADC

Project 3578

Client: AFOSR Project began: 3/1/61 Project terminated: 6/30/64
Title: Research in Individual Information Handling
Proposals: EU 60-251 effective 3/1/61 \$26,924
EU 60-251 effective 2/20/62 26,924
ESU 62-85 effective 3/20/63 35,997
Total Project Funds: \$89,845
Contract No. AF 49(638)-1024
Project Leader: D. C. Engelbart

Online Project

Project 4385

Client: Office of Secretary of Defense
Project began: 2/4/63 Project terminated: 5/8/64
Title: Research Aimed at Computer Augmentation of a Programmer
Proposal: ESU 62-89 effective 2/4/63 \$195,000
ESU 62-89 Rev. A " 4/29/64 (?) 48,689
Total Project Funds: \$243,689
Contract No. SD-163
Project Leader: J. H. Wensley

Project 4506

Client: ESD Project began: 5/1/63 Project terminated: 12/31/63
Title: Research on Computer-Augmented Information Management
Proposal: ESU 63-7
Total Project Funds: \$39,716
Contract No. AF 19(628)2914
Project Leader: C. Bourne; Project Supervisor: D. C. Engelbart

Project 4590

Client: MIT Project began: 7/8/63 Project terminated: 8/31/63
Title: Services of D. C. Engelbart for technical assistance on Project MAC
Total Project Funds: \$6,550
Purchase Order No. SR 95209
Project Leader: D. C. Engelbart

STANFORD RESEARCH INSTITUTE

To: _____ Date: _____
From: _____ Location: _____
Subject: Projects Relating to Man-Computer Program-- Answering:
Page 2

Project 4784

Client: U. S. Naval Training Device Center
Project began: 12/31/63 Project terminated: 12/30/64
Title: Automatic Psychomotor Skill Training
Proposal: ESU 63-49
Project Funds: Sub 1 \$24,917 (D. C. Engelbart)
Sub 2 30,000 (Woodworth)
Total Project Funds: \$54,917
Contract No. N61339-1517
Project Leader: D. C. Engelbart

Project 4987

Client: ESD Project began: 4/1/64 Project terminated: 4/30/65
Title: Research on Computer Augmented Information Management
Proposal: ESU 64-11
Total Project Funds: \$59,638
Contract No. AF 19(628)-4088
Project Leader: D. Lincicome; later D. C. Engelbart

Project 5061

Client: NASA Project began: 6/11/64 Project terminated: 7/11/65
Title: A Research Study of Computer-Aided Human Control of Computer Display
Proposal: ESU 64-43
Total Project Funds: Sub 1 \$76,626 (W. K. English)
2 5,000 (J. Yarborough)
Total 999 \$85,626
Contract No. NAS1-3988
Project Leader: W. K. English

STANFORD RESEARCH INSTITUTE

To: _____ Date: _____
 From: _____ Location: _____
 Subject: Projects Relating to Man-Computer Program-- Answering:
 Page 3

Project 5150

Client: Office of Secretary of Defense
 Project began: 6/29/64 Project terminated: 3/31/66
 Title: Computer Facilitation of a Computer Programmer
 Proposals: ESU 64-74 \$58,643 effective 6/29/64
 ESU 65-7 39,900 " 4/26/65
 ESU 65-66 40,649 10/27/65
 Total Project Funds: \$139,192
 Contract No. SD-269
 Project Leader: D. C. Engelbart

Project 5919

Client: RADC Project began: 2/23/66 Project termination: 3/22/67
 Title: Computer Augmented Management System
 Proposal: ESU 65-106
 Project Funds: Initial to 9/30/66 \$60,000
 Rec'd 7/6/66 33,528
 Total Project Funds: \$93,528
 Contract No. AF 30(602)-4103
 Project Leader: W. K. English
 (See next page for additional information on this project)

Project 5890

Client: NASA-ARPA Project began: 2/8/66 Project termination: 4/7/68
 Title: Study for the Development of Human Intellect Augmentation Techniques
 Proposal: ESU 65-115
 Project Funds: Initial to 6/7/67 \$325,000
 Proposed addition 167,476
 Total Project Funds: \$492,476
 Contract No. NAS1-5904
 Project Leader: D. C. Engelbart

STANFORD RESEARCH INSTITUTE

To: _____ Date: _____
From: _____ Location: _____
Subject: Projects Relating to Man-Computer Program-- Answering:
Page 4

Project 5919 (continued)

Client: RADC

Project extended (in dollars (from ARPA) and scope to support the multiconsole time-sharing system)

Proposal: ESU 67-10

Additional Funds: \$565,676 (received May 1967)

Project 6631

Client: SOG

Project began: 5/16/67 Project terminated: 11/15/67
(extended 1/16/68)

Title: Feasibility Study of File Structuring and
Techniques for Data Storage and Retrieval

Proposal: ESU 67-31

Total Project Funds: \$23,997

Contract No. XG-2960(62-2269)67R

Project Leader: D. C. Engelbart

To: ESD File
From: D. C. Engelbart
Subject: Call from Lt. Herin

HC ESD
45 pages file
January 18, 1964

Herin reports that the reviewers of our first report feel that it needs "more substance," and that he wonders if we could beef it up. Mentioned Major Jeffreys said that a reader who wanted to make a system like ours wouldn't be able to from the description that we gave.

I expressed the feelings that

(a) much of the detail work of the system evolved out of the ARPA work, which would be reported in full detail at a later date and which wouldn't be appropriate in the ESD report (since ESD contributed only about one-fifth of what ARPA has), and

(b) we considered the report really as a progress report in the first leg of a continuing project--a leg of such short duration that one could hardly expect a comprehensive system development and description, and

(c) I feel that our emphasis is on exploring the utilization of the services we described--that there was no attempt at novelty in the implementation of these services, that we consider it as instrumentation for the research that we are trying to get at, etc.

I didn't exactly refuse to add more to the report, in fact I explicitly stated that I was only giving a point of view. Herin apparently feels my position was sound, and said he'd go back into the fray to champion our stand. I asked him to let me know if he was getting too bloody so that we could give a little to make it easier on him. But I did point out that we didn't have any more funds to put into this, from their support, and that I wasn't sure just when it could get done.

He mentioned another item (better news): Major Jeffreys mentioned something to him about getting another 8K to add to the 60K, for use on "communicating equipment" to the Q-32. In my last memo, on the call from Jeffreys I mention that they would interpret our proposed tapewriter equipment as being an ok expense here. Herin also said that it looked as if the paragraph I dictated over the phone would be accepted for the new work statement.

I mentioned that some of our people might be in the area, to visit them. He seemed to hope sincerely that they could. We agreed that face-to-face talking would be very helpful.

D. C. Engelbart

na

CC: J. D. Noe, R. C. Amara, T. L. Humphrey

He WS project file

January 13, 1964

Mr. Lee S. Christie
Principal Scientist
System Development Corporation
2500 Colorado Avenue
Santa Monica, California

Dear Mr. Christie:

I very much appreciated receiving your letter and draft Consultant's Report. It is extremely pleasing to find another kindred spirit, and I very much hope that we find an early opportunity for personal discussion. I have occasional business at SDC anyway, and may soon find the opportunity to combine such with a visit with you--but I think I wouldn't wait too long before I should try to make it there even without other business. You may know that SRI is part of the community that is tying in to the time-sharing system of the G-32 in the ARPA Command Laboratory there at SDC--it is actually our little local CSR that is getting service from them over a Data Link phone line.

We have started our experimental research program, based pretty much on the recommendations in my AFOSR report. You might be interested in visiting our laboratory to see what we are doing. I am sure that the rest of the staff here would welcome the opportunity to get acquainted with you. There are six full-time professional staff members, and quite a number of part-time participants. So if you have the chance to visit us, please be sure that you will be warmly welcomed.

Sincerely,

Douglas C. Engelbart
Sr. Research Engineer

DCE/na

to CBS project

SRI - SDC EXPERIENCE IN USE OF 2 KC DATA TRANSMISSION LINES
FROM

MENLO PARK TO SANTA MONICA, CALIFORNIA

A JOINT REPORT BY SRI AND SDC

I INTRODUCTION

This report contains conclusions which have been reached on the basis of SRI/SDC experience in setting up a data transmission facility between two computers. The various conclusions are separated in the main body of the text and examples and reasons are given for the particular conclusions. In all cases an attempt has been made to make the conclusions general so that they may be applicable to as wide a class of problems as possible. A section of this report details the specific experience which we gained from our attempts to set up such a facility. It lists the effort required in different phases of the activity, and the number of machine instructions which were required to implement the various functions. An appendix details the characteristics of the system which was implemented.

2. Conclusions

2.1 Specification

It was found essential in setting up a line (when the programming at the two different ends was carried out by two separate organizations) to specify accurately the system to be implemented. There should be no possibility of ambiguity in the specification. Any time used in removing ambiguity will be saved in the eventual check out of the program and the system. The system which we implemented contains certain ambiguities in the description.

For example, we used a particular Q-32 word as a longitudinal check sum which was defined in the original system specification as being, "the exclusive or of all words other than itself." At SRI this was interpreted in its literal fashion, namely, of all words, even including other control words in the specification (see Appendix). At SDC the "all words" was interpreted as being only applicable to the "data" words. The result was that the check sum was differently defined at the two ends, and when the system was checked out, the check sums were found to be wrong. Time and effort were spent in identifying this error, plus a modification of one program to correct the situation. This particular example shows that possible ambiguities in the specification could have been avoided by quoting examples of typical valid messages sent from one end to the other.

2.2 The most reliable communication should be used for control.

In the running of a system whereby data is exchanged between two computers, either in a real-time or non-real-time sense, (i.e., allied or not allied to human response time), it is imperative that the two computers should be able to understand what the state of the system is at the far end. A situation should not develop wherein both terminals are transmitting simultaneously, or both are simultaneously attempting to receive. Frequently these situations require human intervention to correct, and it is assumed that human intervention is to be reduced. Our conclusion is that the most reliable communication should be the one bit of information which says, "I have just transmitted to you" and which is implied at the distant end as being, "I have just received from you." Following this one bit of information, it should be perfectly clear to both ends as to which terminal is next to transmit, which

terminal is next to receive and what valid messages may be sent at this particular time. We have had difficulty in inferring correctly at one end that a transmission has occurred from the other end. Firstly, the sync pattern used is a short one, namely, "0000001." The probability that this particular set of bits would occur within a noise string of n bits is

$$\frac{n-6}{128.}$$

If a noise burst on the line exists of a fraction of a second, there is a high probability that somewhere within it sync pattern will be detected. The receiving computer will then infer incorrectly that a message has been sent. This particular trouble can be avoided by using a longer sync pattern. The general rule can be stated that if the sync pattern is a unique pattern of m bits long, and if a noise burst is of n bits long, then the probability that the noise will contain a valid sync pattern is $\frac{n-m+1}{2^m}$.

As an example of this, if the sync pattern were made 20 bits long, and noise bursts were an average of a thousand long, then the probability of a noise burst containing a sync pattern would be,

$$\frac{1000-20+1}{2^{20}} \doteq .001,$$

which is sufficiently small for the use to which we put the system.

Having increased the length of the sync pattern, there now exists a higher probability that the sync pattern contained in a valid message will be incorrectly received (and therefore not recognized) due to one or more bits of that sync pattern being changed due to errors on the line. This would mean that a valid message sent would not be received

resulting in a loss of knowledge concerning message exchange between the two terminals. This can be avoided by having several sync patterns in each message, let us say, for example, three sync patterns all of a sufficient length to avoid the troubles mentioned in the previous paragraph. The several sync patterns should not be identical as it will be necessary to distinguish whether a sync pattern which is recognized was the first, second or third. The use of several sync patterns would avoid the problem that a valid message sent which suffered error during transmission was not interpreted as a valid message.

Following the measures adopted or proposed in the two paragraphs above, it should be possible to place high reliability on the question as to whether a message has been received or not, even if the message contained errors. We should now adopt a third principle, which is that messages should follow a strict sequence. This could, for example, be a strict alternation of messages or could involve a number of messages from one end followed by one from the other end. Any sequence which is known in advance will suffice. Naturally, the simpler the sequence, the simpler the program. The system which we used had some ambiguity as to the sequence. For example, if an 05 message was sent from one end to another, (let us say from a to b), then it would be expected that b would send the next message to a. However, if an 03 message were sent, then the next message would be from a to b. As 05 or 03 messages are both possible following an 06 transmission from b to a, it becomes important that the computer distinguish between 03 or 05, or we would have a situation where both computers are transmitting or both are receiving.

As outlined in the paragraph above, the difference between an 03 or an 05 message in our system results in a loss of control of the system, i.e., the necessity for human intervention. This would be partially avoided by adhering to a strict sequence, however, it would also be necessary to have a system wherein the probability that one of these messages could be mistaken for the other was highly unlikely. It is therefore important that we state a further principle, namely that all control messages should be coded such that the codes have a large Hamming distance between each other and from any likely error messages, (i.e., cross talk, etc.). We might, therefore, for example, using our system as an example, code the 03 message as being 28 "0"s followed by 28 "1"s, and the 05 message as being alternating "0"s and "1"s. This would give a satisfactorily large Hamming distance between the two messages (and from error conditions of all "0"s or all "1"s).

2.3 Test Procedures

During tests of the programs at a particular end, it is essential that control of the tests rest at that end. No attempt should be made to carry out tests which are checking or trying to 'fault-find' equipment at both ends of the line. As far as possible a test of one end of the line should only involve the other end of the line in a function which had already been found to be reliable.

2.4 Start-Out Procedure

It was found necessary to have a rigorous start-up procedure. This involved programming the humans at time of cut-over from "voice" to "data" on the line. It is not necessary for this system to be complex, however, it must be adhered to at both ends of the line. A complete understanding should be present as to which end of the line should

transmit first and what type of message could be a valid opening message. In our case, this was achieved by using a special message (type 20) to indicate start of transmission. This is unnecessary and any specially pre-arranged sequence or single message could be used.

2.5 Program Construction

It is necessary in constructing programs for data transmission as with any other programs to achieve a ruggedness within the environment in which the program is to work. In the conventional computer program this is a question of the program not failing with different kinds of data. However, in the case of a data transmission program, it is necessary to insure that the programs will not be destroyed with different kinds of data, different human errors, different line errors and different computer errors. For example, a burst of noise on the line when it would normally be expected for Station A to transmit to Station B should not put the system into a state from which it cannot automatically pick up. This has been achieved in the system we used. For example, if when SRI is preparing to transmit and SDC is listening, a burst of noise occurs on the line, SDC will receive this and will immediately transmit an 05 message, signifying that the message received was in error and should be repeated. It is assumed that the burst of noise is unlikely to satisfy the various checks (horizontal and longitudinal parity). Following the 05 from SDC, it would again return to a condition of listening and waiting for a message from SRI. When eventually a message was sent from SRI, SDC would interpret it as being the second transmission of this message, however, this would produce no change in the eventual outcome of the operation. The system we have been using is rugged with respect to this particular error. This ruggedness is not achieved in the special case when SDC has just sent an 06 message to SRI and is

listening, waiting for a return of an 05 or 03 message. Any noise could be interpreted by SDC as either of these or if it detected that it was neither, it would not know whether to infer that it was noise in which case it should listen for an 03 or 05 again, whether it was an 05 in which case it should transmit an 06 message to SRI, or whether it was an 03 in which case it should listen and expect to receive an 06 message. This is a condition which has been discussed in Section 2.2, namely that there should be no doubt at any end as to the state of the system, and as to valid message which can next occur. The time span over which a computer is listening for an 03 or 05 is typically very short as acknowledgments are sent immediately upon receipt of last message so this particular occurrence has never been of the significant trouble to the SRI/SDC system.

A further condition which should be guarded against is that human interference should not put the programs into a state from which they cannot pick up. In the SRI/SDC system, this has been achieved at SDC but not wholly so at SRI. This is due to the nature of the equipment at both ends. In this particular case, it should be possible for a human to switch from "data" to "voice", to ring the other end of the line, to discuss some issue with them, to revert back to data, and to restart the computer at both ends, whereupon the programs will pickup and carry on from the last action. Broadly speaking, this has been achieved in the SRI/SDC system, however, it requires manual re-setting of the computers to a particular state depending upon the time at which the interference of a human occurred.

This section really is a reiteration of general programming philosophy in any system which is to have a good reliability, namely, the program should be designed to be rugged with respect to the environment in which they are working.

3. Experience with SRI/SDC System

The particular experience gained in the use of a data transmission system is difficult to extrapolate to other conditions (except in principles). For example, the noise and error characteristics of particular lines are very variable. For further example, we were operating using half-duplex channel and the way in which control would be exercised when a full duplex was used could quite probably be different. It is certainly true that no extrapolation could be made of any figures concerning reliability, error rate, etc. from the very small experience we have had in terms of the fact that we have used one and only one line between two fixed stations. This line has to our knowledge been of constant equipment configuration as far as the telephone network is concerned. It is unlikely that in general this can be achieved. We are fully aware that the telephone companies involved carried out extreme measures to assure a line of high quality. It is doubtful whether in all cases this can be achieved. We can say that the number of faults which occurred due to error conditions in the line was very, very small, in fact below the average level which was forecast by the telephone company for this facility. We were fortunate in this respect. It is therefore not our intention to try to quote figures concerning error rates, reliability, etc.

Our experience in setting up and using this system however, may be valid in terms of the information which we can state fairly clearly as to the amount of effort required to achieve the facility. At the 160A (SRI), the data transmission program requires about 1,000 instructions which took approximately three man weeks to write; the check out of the system took approximately three man months. At SDC the figures were (to be inserted by SDC) The check out mentioned in both cases is a check out to the point where the total system works. This includes all work which involves a programmer being available but which may nevertheless be fault finding in equipment, on the lines, or the interfaces between computers and line.

4. The SRI/SDC System

(It is my intention in this section to have SDC incorporate an up-to-date version of their system specification for data transmission system.)

VCC

LC LICKLIDER
US project

May 4, 1964

Dr. J. C. R. Licklider, Director
Behavioral Sciences
Command and Control Research
Advanced Research Projects Agency
Pentagon -- 3D-179
Washington 25, D. C.

Dear Dr. Licklider:

In compliance with your recent request, this letter proposes a modified work statement and budget for continuation of our ARPA-supported man-computer research. Our proposal ESU 64-1, Revised (as enclosed) is hereby modified. The new activity would be research on an exploratory on-line working system.

I propose using your support to evolve my own personal system of computer aids and to use them as much as possible to do my work. I need the full-time assistance of a programmer to develop, modify, and maintain the computer programs involved in this work. I wish to develop and use a ten-key "short-hand" keyset and code. I also want to experiment with voice-recording devices and associated clerical help in transcribing to computer-sensible form, plus transcription from printed data. These require technician and clerical help.

Attached is an itemized cost estimate, and below is a more detailed description of the work I plan.

THE COMPUTERS TO BE USED

I want to get some use from the facilities we have developed in the Q-32 and propose keeping linked to it until September 1, 1964. We are shifting the dataphone link from a leased line (at about \$1500/month) to a dial-up service costing about \$150/month to maintain and \$24/hr (\$0.40/min) to operate. I estimate about 50 hours usage over the four months.

I plan to shift to our 160A within the four months. We are planning to add auxiliary storage -- e.g., a tape and a drum can be added for essentially the same rental rate as our line to Santa Monica has been costing us. Charges for the 160A system will be on a pro-rated usage basis, and there will be five projects sharing its time and costs.

THE ON-LINE EXPERIMENTATION

For years, I have dreamed, talked, and written of increasing one's effectiveness by means of on-line computer aids. I want to use ARPA support to get on-line and begin bootstrapping myself into an ever more efficient system for doing my own work. Over the next nine months, approximately half my time will be involved in this work, divided roughly as follows: (a) two months' time available to work on the system -- to specify and supervise programming, to evaluate and respecify, etc. -- and at least (b) that much more time to exercise the system in doing real work with it. (See the next paragraph, though, for project-charging plans.) I can start with the on-line Z-Code system already developed for the 160A and soon get a quite usable small-scale text-manipulation facility. Larger-scale manipulations can be done on the Q-32.

To work with the system, I have available the equivalent of several months (see (b) in paragraph above) to devote to "computer-aided" work for experimental exercise of my system. Since other projects are expected to underwrite my time for doing this work on-line, the ARPA project will be charged only for the computer time involved. I consider this a good way to exercise and help evaluate the system. There is value to the objectives of ARPA in providing the experimental facilities for selected users who will, while working at other work, exercise the system in a manner profitable to ARPA's purpose. ARPA would receive full acknowledgment for the novel way in which the work was done, and for the lessons so learned, while the other sponsors would receive normal acknowledgment for the work itself. I hope, among other things, to write my papers and reports using our (referring to me and ARPA) computer-aided techniques.

Our NASA project is also slated to build up our facility by setting up basic input and editing functions (at the level of Delete, Insert, Move, etc.) on the 160A -- to keep improving them and experimenting with new devices and techniques, but always keeping serviceable the best generally usable configuration. These facilities are aimed to make it really fast and easy to compose and modify text (including program instructions) on a single-frame scale of activity. In my ARPA work, I shall harness the "best" (to my personal liking) of these facilities as they emerge, integrating them into my personal working system.

To describe what I plan to do for ARPA, it is best perhaps to say where I will start and then state the principles that will guide developments from there. I want first to make it easy for me to work on a body of text of perhaps ten pages (say 30,000 characters) in size. I will want to nail down techniques allowing me quick inspection, which will involve such as: (1) determining line length, character size, and formatting that seem best, (2) finding whether I like page-hopping, scroll-rolling or half-frame jumps for scanning, (3) developing a convenient "place-marker" system that makes it easy to set up or move markers, easy to keep track of the reasons for having left markers at given points, and easy to hop nimbly between markers.

Quick, detailed inspection will begin to provide for efficient study of my own or another's work. The next requirement is efficient extraction of relevant material in meaningful form. Then follows the requirement for efficient integration of added concepts -- compatibly inserting new text, modifying existing text, and modifying structure. The NASA project will supply good basic instruments, but there will still need to be significant exploration of procedures, conventions and working methods.

Beside and beyond the work directly implied by the above goals will be that associated with exploratory pursuit of both (a) crisper, more effective execution of the working processes that I find useful, and (b) procedures, conventions and methods newly available to me by virtue of the computer aids and of the way I will have learned to use them. The nature of the new procedures, conventions and methods that I have in mind to explore is described in some detail in my report, Augmenting Human Intellect: A Conceptual Framework, AFOSR-3223, October 1962, (Contract AF 49 (638)-1024). In particular, I am interested in the associative-linkage concepts in Section III A-3, and the argument-structuring concepts in Section III B-4 -- but there are a number of other notions in Section III in which I also am quite interested.

BINARY KEYSSETS

I have developed, learned, and experimented with five-key chording keysets, with which I can now transmit at around thirty words per minute. There are several directions I would like to pursue with this approach. One of them, using a one-handed keyset in conjunction with a light pen (or cursor) to give two-handed, on-line, selection and designation capability, will be explored under our NASA project.

Another direction, which I would like to pursue under ARPA sponsorship, is to explore the possibilities of higher-speed transmission by providing more bits per hand stroke.

I would like to build a ten-key one-handed binary keyset, with two keys per finger, 1023 unique hand-stroke codes. I want to use 255 of these codes (on eight keys), and to develop, learn, and use a simple shorthand code, in which single characters are in the ASCII code (the coding that I already use) as a compatible subset of the larger code, and in which common words, n-grams and phrases are included for single-stroke designation. There is much we can borrow from previous studies in designating a reasonable trial code. In programming for computer interpretation of this code (from off-line-produced tape record for on-line input), I would like to include the facility for pre-scribing and using abbreviations. (These two features -- the abbreviation and shorthand -- are described to some extent on pages 78 and 79 of my above mentioned AFOSR report.)

I plan to link these keyset features into my off-line work by making a coupling device which will punch eight-level code (with the paper-tape punch on my personal tapewriter - Model 33 ASR Teletype) when I operate this keyset off-line, so as to have a portable (long-wire) note-taking device. These experiments will require machinist and technician help.

May 4, 1964

VOICE RECORDING AND TRANSCRIPTION

I plan to experiment with both portable and fixed voice recorders, to capture certain types of data, ideas, etc., and to have a procedure for a secretary to transcribe these into computer-sensible form so that I can edit and integrate them during my on-line working sessions. I also expect to use a fair amount of secretarial help for transcribing printed references of special value so that I can work with the material on line. I expect to need technician help in modifying and wiring some of the voice-recording equipment.

OUTPUT FROM THE PROPOSED WORK

Nine months should produce a good start toward evolving the "way of life" that someday will let me harness computer aid in a way that significantly increases my working effectiveness. As that way of life evolves, there will be many products of value, but the primary value is expected from the increased knowledge of (1) what procedures, conventions, working methods, and skills can help a human harness real-time computer-aid to increase significantly his working effectiveness, and (2) what relative value to this user seem to be the various parameters of computer-system service. Another important though intangible product will be visible demonstration of a coordinated system of on-line aids actually being used to do real work with increased effectiveness.

After nine months elapsed time, the chief output will be a report to ARPA of the nature, "Here is a description of the system that I have developed and been using. Here is how I work now, this is what my experience has taught me, and here is what I will change or add." It will be one of my aims to be able to say, "This report was thought through, drafted and re-drafted, edited, and given final composition, entirely on line."

The functional characteristics of the terminal hardware and the computer programs will be assumed as being important to communicate. However, for some time I frankly expect that the programs themselves, serving an exploratory and evolving system, would be so patched-up and inefficient that description beyond the significant functional characteristics would be worthless. In a year or so, this exploratory user-system activity would likely settle upon fundamental configurations for data structuring and functional programming organization, at which time it would be valuable to publish a clean design for such a system.

A valuable product of the work is expected to be a growing perspective as to the value which the user can derive from different configurations of terminal, computational, and storage facility. This value is necessary to compare against cost when hardware-software system designers set specifications on future systems.

I hope this letter serves its purpose.

My best regards,

D. C. Engelbart
Program Head
Man/Machine Information Systems

DCE/na
Attachments -
Cost Estimate, Proposal

TO: ESD File

January 2, 1964

FROM: D. C. Engelbart

SUBJECT: Telephone discussion with Lt. Herin and Major Jeffries

Lt. Earle Herin called on Monday, December 30, to say that the work statement handed to their contracts people by Conway for the new contract with us had been rejected. It was the same statement used for our earlier contract and they wanted a different one ("we don't want to buy the same piece of goods all over again.") I told him I would prepare a new one and call him back with it. In working on the statement, the question arose as to the possibility of negotiating with ESD to raise the contract price to cover equipment costs that heretofore hadn't been charged them. After discussion with Roy and with Pete Browning, we decided to go ahead on that basis if Herin would think it reasonable.

Herin called me this morning before I had a chance to call him. I quoted the following paragraph to a stenographer, "Put into operation the information-management subsystem as designed and largely implemented during the first five-month contract period. Integrate its services into the research and programming system used by the project researchers, including both their off-line and on-line activities. Improve user conventions, methodology, and terminology as indicated by system analysis and evaluation, and similarly evolve changes in the information structuring and programs of the subsystem." I told Herin that it would seem all right to me either to put this in as Item 3 along with the Items 1 and 2 of the original work statement, or to put this paragraph in by itself as the whole work statement. He said that he would get together with his contracts man and work out a statement for the PR.

I also asked him to change the original work statement by deleting the following from the last sentence of Part B, "at no cost to the Electronic Systems Division." And explained how we felt we should start charging them for the direct equipment costs as incurred by the researchers on their project. Then I went on to ask his opinion of the likelihood

of our getting extra funds to pay for these expenses rather than to have to reduce the manpower. He said he didn't know, and would try to find out. I said we'd plan on writing our proposal as for the extra-fund possibility, at about 1K/month, unless we heard otherwise from them.

Then I got a call from his boss, Major Jeffries about two hours later. Told him the whole story about RECAP getting the O. K. to move back up here and establish the 160A and phone link between us and the Q-32, and how these direct charges to ARPA's project were in a different category from the free-use, no-charge deal with the Q-32, so that we felt it only fair for ESD to pay for the equipment time that its researchers used. Jeffries went along with this, but said that they were really up against it now for funds and he couldn't see at all their finding this extra money. Said he'd been looking for an extra 30K so that they could make this be a year's contract instead of an eight-month contract, but had given up on that, too.

Jeffries will change the cost-charge item in the work statement to allow expenses for "equipment used to communicate with the Q-32." I told him about our off-line tapewriter plans, and he said that an off-line piece of equipment like that would be considered as being used to communicate with the Q-32 and would therefore represent allowable charges. He apparently was worried about the reaction of his contracts types to this equipment-charge change.

Conclusion: No extra funds forthcoming, but probably can make the equipment charges to the contract at the expense of less manpower.

DCE/rt

cc: R. Amara
P. Browning
T. Humphrey
J. Noe

copy to CS program file

December 6, 1962

Mr. Ivan Sutherland
Electronic Systems Laboratory, Bldg. 32
Massachusetts Institute of Technology
Cambridge 39, Massachusetts

Dear Ivan:

I trust that you are getting your thesis finished on schedule. Can I be put on the distribution list, if there is to be such a thing?

And, another request. It would give my cause (of getting people around here stimulated in the possibilities for close man-computer cooperation) a great boost to be able to show them your film. You mentioned that there were going to be distribution copies made up. We'd be most happy to pay any handling and mailing expenses involved. Can you tell me how to go about getting a copy for a week or so, and how soon this might be arranged?

The possibilities are definitely looking bright for us to get going on an experimental program to pursue my peculiar approach to man-computer research. ARPA has a command-control research program within which our suggested approach for "augmenting" a computer programmer as a first experimental step seems to fit very nicely. And it looks as if we will have rather good availability of the transistorized super-AGE computer (Q32) that is in the Command Laboratory at SDC. That machine even tops the TX-2 in size. We have acquired a character-generator and display to tie to our CDC 160A for local experimentation, and 2.4 of us are hard at work in the initial stages of our research already (under temporary SRI sponsorship). There are lots of things we can be learning on the little machine. So life is getting quite exciting for me, after all the years of being able only to philosophize.

Hope you are making good progress. I will be interested in keeping in touch, so let me know what happens to you. And I really hope that you will be able and willing to keep working in this same field--you've made such a good start, and I think that it is an extremely important field in which to invest one's professional energy.

Best regards,

Douglas C. Engelbart
Senior Research Engineer

DCE:etm

copies A 40 pages BEL

SOURCE:

DCE 8 JUL 63-2

TITLE:

Decision Trees

COPIES TO:

CPB, JHW

SEE ALSO ITEMS:

FILE FOLDER:

FORMAL DESCRIPTORS:

ABSTRACT: Reminder that sequential selection techniques can be very powerful with on-line CRT and that we should be sure to implement representative examples of such in useful processes as soon as possible.

TEXT: We have all discussed this technique before. I note in the MAC discussions that it seems to be considered as a very promising basic technique. I speak of the sequential selection of alternate choices as presented in programmed sequence to the human as fast as he selects, he is presented with a set of succeeding choices whose relevance is based upon his prior selections. Sort of like the twenty-question approach to using a very small amount of information to designate a choice or arrive at a decision.

In the plans for the command structure of the generalized search command it has been considered that a one-level sort of selection scheme be used to help the human select his desired command structure. What I recommend is - that we isolate the apparent advantages of this sequential selection technique and seek some good applications for it within our systems.

I have previously mentioned to CPB its possible use in finding an index item in a hierarchical classification system of considerable depth and comprehensiveness. It might also be used, where cleverly programmed multiple-choice questions lead one to the appropriate descriptor.

copies A W. project files

SOURCE:

DCE 8 JUL 63-2

TITLE:

Decision Trees

COPIES TO:

CPB, JHW

SEE ALSO ITEMS:

FILE FOLDER:

FORMAL DESCRIPTORS:

ABSTRACT: Reminder that sequential selection techniques can be very powerful with on-line CRT and that we should be sure to implement representative examples of such in useful processes as soon as possible.

TEXT: We have all discussed this technique before. I note in the MAC discussions that it seems to be considered as a very promising basic technique. I speak of the sequential selection of alternate choices as presented in programmed sequence to the human as fast as he selects, he is presented with a set of succeeding choices whose relevance is based upon his prior selections. Sort of like the twenty-question approach to using a very small amount of information to designate a choice or arrive at a decision.

In the plans for the command structure of the generalized search command it has been considered that a one-level sort of selection scheme be used to help the human select his desired command structure. What I recommend is - that we isolate the apparent advantages of this sequential selection technique and seek some good applications for it within our systems.

I have previously mentioned to CPB its possible use in finding an index item in a hierarchical classification system of considerable depth and comprehensiveness. It might also be used, where cleverly programmed multiple-choice questions lead one to the appropriate descriptor.

copy to: WD project file (2)
DION (RAWID)
ESD

December 31, 1962

Lt. Melvin Conway
WSRC, Hanscomb Field
Bedford, Massachusetts

Dear Lt. Conway:

It was a real pleasure to receive your call this morning and to learn that ESD has an interest in our RECAP project (Research Aimed at Computer Augmentation of a Programmer). The fact that the estimated charges for the project, as listed in our proposal ESU 62-89 Revision No. II, represent a larger sum than you can invest should not be any great barrier.

As you may know, SRI has been investing its own basic research funds in this program for two years. But practically speaking, the Institute can't carry this burden indefinitely and can't afford the added burden represented by the contact, communication, and coordination work required for handling many small "subscribers." So I imagine that a lower-limit ante of something like 100K per year from the first few subscribers would represent our current attitude, at least until we had a total that was of the order of 300K. If you could fund something over 100K per year, we would be agreeable to working out an independent contract with you on a basis that would be compatible with a multi-sponsor program.

I am enclosing a list of the other agencies and contacts therein to whom we have submitted the RECAP proposal, and you are welcome to contact any of them if you would like to explore the possibility of going in jointly with one of them on a contract with us instead of letting a smaller independent contract as mentioned above. We have received negative letters from the AEC (9 and 10) and from NSA (7), relatively promising responses from ARPA (1) and NIH (2), and no response at all from the others. A site-inspection committee from NIH will visit us January 8, and their first committee meeting to act upon our proposal will meet the first of March.

With ARPA, we are in the rather definite stage of talking details, but on a basis that is more development oriented than we had proposed--and as I mentioned on the phone, to do all of our experimental work on a work station that we build and tie to the Q32 at SDC (400 miles from here). The way Licklider and his assistant, Dick Chapman, talked at our last meeting two weeks ago, we should reach a definite agreement (one way or another) with ARPA by the end of January. If the ARPA possibility actually matures, there would be several rather natural areas of complementary research still required to round out the RECAP project as we feel it should be covered. In such a case, an independent contract from you would work out very nicely. We can discuss such possibilities further with you if you are interested.

December 31, 1962

I should add also that Mr. Fred Dion (RAWID) in the Information Processing Laboratory at RADC has corresponded and visited with us over the past year and a half. He is supporting some work at Ramo-Wooldridge in mathematical manipulations on a real-time work station, and is associated with a real-time computer-display system that they are setting up in-house at Rome. I've never met Dr. Crocetti to whom somebody at RADC advised us to forward the proposal. I think, in fact, that he is in a different Directorate from Dion. I should think that Fred Dion would be your best initial contact there.

After attending the Command Control conference at Hot Springs, I gathered that there was a close connection between MITRE and ESD, although I couldn't quite understand the exact nature of it. However, I had several long talks with John Jacobs and Dave Brown of MITRE, who expressed quite a bit of interest in the program we are trying to launch, and to whom I subsequently sent copies of the AFOSR Summary Report (3223). Perhaps, if it affords any possibility of a source of more funds for you, it might do some good to contact them. Jacobs is one of the vice presidents, and Brown is in charge of a good share of the information-processing-system development.

I am really delighted that you are interested, and hope to hear from you again soon. We stand very willing to discuss any contract arrangement that is to your interest.

Sincerely

Douglas C. Engelbart
Senior Research Engineer

DCE:etm

Enclosure

1. Dr. J. C. R. Licklider
Advanced Research Products Agency
OSD Room 3E-163
The Pentagon
Washington 25, D. C.
2. Dr. Bruce Waxman, Executive Secretary
Advisory Committee on Computers in Research
Division of Research Grants
Department of Health, Education, and Welfare
Norfolk Building, Room 208
Bethesda 14, Maryland
3. Mr. J. R. Marvin
Director, Code 491
Department of the Navy
Office of Naval Research
Washington 25, D. C.
cc Mr. James R. King
Resident Representative
Office of Naval Research
Stanford University
Stanford, California
4. Dr. Richard H. Wilcox
Head, Information Services Branch
Department of the Navy
Office of Naval Research
Washington 25, D. C.
5. Rome Air Development Center
Griffiss Air Force Base
Rome, New York
Attn: Dr. C. P. Crocetti, RASH
6. Mr. Greg McClurg
Scientific Adviser to Colonel Hazeltine
Operations Research Division
Army Research Office
Arlington Hall, Virginia
7. Director of National Security Agency
Fort Meade, Maryland
Attn: R4 (Dr. H. H. Campaigne)
8. Mr. Herman Lowell
Office of Advanced Research and
Technology
National Aeronautics and Space
Administration
Washington 25, D. C.
cc Dr. T. L. E. Smull, Director
Office of Research Grants and
Contracts
NASA, Washington 25, D. C.
9. Dr. Sid Fernbach
Computation Department Head
Lawrence Radiation Laboratory
P. O. Box 808
Livermore, California
cc Mr. Russell Nanfelt, USAEC
San Francisco Operations Office
211 Bancroft Way
Berkeley, California
10. Dr. C. V. L. Smith
Atomic Energy Commission
Germantown, Maryland

copy in 60 project file
.ESD

7 March 1963

Roy Amara
Jerre Noe

D. C. Engelbart

Contact with ESD (located in Bldg. L at
MITRE) Feb. 28, 1963.

Personnel: Lt. Melvin Conway

Subject: Our proposal ESU 63-7

I talked with him for three and a half hours. Very informative interchange. Turns out that he has a Ph.D. from Case in mathematics, and has prime personal interest in programming systems. Has his own pet compiler design which he is independently (of the AF) getting several manufacturers interested in implementing. Also, at MITRE, there is a project called ADAM (Oliver Kinney, of MITRE, in charge) for which Conway is the AF project manager. It apparently is an on-line information storage and retrieval (not just retrieval) system--but Conway feels its characteristics wouldn't interest us because it is for use in a big operational system and would be quite "stiff" compared to our planned developments. He seemed to approve of what we planned to do and of our method of approach. He said the paperwork had gone to their purchasing department, but he wasn't too optimistic about its getting through by the first of April as he had hoped.

DCE/ms
cc: Cons

copy for OS project files (2)

AUTHOR: DCE
DATE: 27 MAY 63
SUBJECT: Auxiliary Storage for 160A

NORMAL DESCRIPTIONS:

SEE ALSO:

REFERRED FROM:

LINK NAME:

COPIES TO: JHW, DCL, TLH, RCA

ABSTRACT: Less-critical dependence on Q32 and phone link, plus possibility for time sharing 160A with other users, suggest adding drum and possibly cheap tapes.

TEXT: The 160A is to be set up as a Division facility, with AHI-program sharing its use with Jim Bliss and others. Not only must we share its use with other people, we stand to have our experimental time also dependent upon the proper working of the Q32, of the phone link and terminal equipment, and of the SDC time-sharing systems.

The chances of getting plenty of experimental on-line time would seem much greater if we added a drum (such as CDC's 32-thousand-word model) to the 160A: (1) we could have enough local store to do some interesting experiments and work here during those inevitable times when service from the Q32 isn't available to us for some reason, and (2) we could very likely put together a simple time-sharing system in which Jim Bliss (or other similar users) could have immediate-interrupt service during much of our experimental working time. It is unlikely that there would be room for the programming and data for both our use and Bliss's in the 8K of core memory alone to set up time sharing without the drum.

An added attraction for the drum is that it could enable us to take the burden of regenerating the displays off the 160A--freeing some of its core memory and some of its computing time. Also several AHI users could time share for some types of work, since there would be enough terminal equipment for two to work independently.

In view of the fact that the sharing of the 160A without real-time time sharing is predictably going to more than saturate the 160A by next Fall, just between Bliss and us, and since the addition of a drum and the time-sharing system would alleviate this (perhaps saving acquisition of another 160A) as well as provide considerable advantage to other ways, I feel that we should begin serious consideration now.

A somewhat independent possibility is to get some of the new DEC "low-cost" tape units. Small reels, with slow stop-start by the usual computer-tape standards, but very much cheaper--something like \$9K for the tape control unit and \$7.5K for each tape-transport unit containing two capstans (four reels) each. DEC can alter the control unit to adapt from any

computer, but they will probably go through the IBM-machine adaptations before they get to working on one for the 160A.

Consideration of this kind of storage is stimulated mostly by the problems as seen ahead for our having use of Q32 tape transports. For some of our processing, we will be wanting to modify the record on tape, and on the Q32 this can only be done by reading one tape, processing and writing on a second tape that becomes the modified record. If we had a couple of transports of our own, we could be sure of the ability to do this type of work. Also, our independence of SDC would be much improved.

(In line with our proposed use of our memo system, I invite others to toss comments, considerations, suggestions, etc., into the file, referenced to this memo.)

copy to CRS project files (2)

AUTHOR: DCE

DATE: 17 JUN 63

TITLE: File-entry Headers

COPIES TO: CPB

SEE ALSO ITEMS: CPB 6 JUN 63, CPB 1 MAY 63, CPB undated (about Feb 63, "Preliminary Suggestions for mdcc File Structure")

FILE FOLDER:

FORMAL DESCRIPTORS:

ABSTRACT: For memo entries, suggest numbering paragraphs, merging DATE and AUTHOR headings, giving general access to all inter-personal memos, keeping a chronological "journal" record of all entries, providing off-line FILE FOLDER assignments to prior entries, and considering independent personal indices to common-pool data entries.

TEXT: These are suggestions by me, not notice of changed procedure. CPB is the only one with authority to specify changes, and he has the concomitant responsibility for consistency, etc.

1. To make it easier to refer back to specific concept in a prior file entry, perhaps we could serially number every paragraph, independent of headings in the item. These numbers could be entered automatically by the transcription clerk, if a reasonably straightforward convention were established regarding what should be considered for our purposes as a paragraph. Probably wouldn't have to be applied too consistently to be very useful.

2. I also suggest that we combine present Author and Date headings, together with Day Serial Number, into what would become a single, unique, and explicit source code. For example, 2 DCE 17 JUN 63 would refer to the third entry I made today.

The following items are specifically relevant to CPB 6 JUN 63:

3. Regarding the COPIES TO header. What if I write a memo like this one, that principally only concerns two of us, and which might clutter up the notebooks of the rest of the program people--I would be tempted to do as I did, list only you under the COPIES TO heading. But when this gets into machine data base, I should like to have anybody who might have an interest have direct access to it. Shouldn't any memo other than a personal one be available to all who might be tracking down something at the on-line working station?

And, another point here. A person may have a personal-file item that at some time becomes relevant to an interchange--and as you suggested, he can easily transfer that item to the group-communication system. I suggest that in this case the DATE listed in the group file be the date at which the item was transferred from his personal file into the group file. I think that this sort of record control will be important to later untangling the history of ideas that represents the serious study of our conceptual development or of our practising methodology. Still retain as an item of data in that item what its date of entry in the personal system was.

4. The "journal" system is something I favor, where every entry made into either a personal or a group file is entered, in its chronological order relative to other entries, into a permanent "journal" file. I think that it is more important in the group file than in the personal file. A person should be able to copy out of the group journal any data he wants, manipulate it how he sees fit in his personal working domain, and then enter new items into the group file. I think we court disaster to allow changes in the existing prior-entry file, other than perhaps to list subsequent links to more-recent entries. The "safety" of this basic record is very vital.

5. About FILE FOLDER heading: Are you going to provide provision for people to make entries nowadays that establish FILE FOLDER assignments, for their personal Folder relationship, to prior entries? Or do we have to remember about our interest in doing so until we get on-line, where I assume there will be a procedure for establishing personal file folders. Perhaps no more than a personal entry to remind oneself to do this, so he can exercise the on-line procedures later--although if we planned these on-line procedures ahead a bit, we could probably establish a way to designate this from off-line entries. I am quite sure that this off-line designation capability would prove very handy to us later, when we might well want to do such things during our off-line meditation periods.

6. About personal-use descriptors for the FILE FOLDER: You mention that we would have false-drop difficulty if we add independent descriptors to the main-file FILE FOLDER heading of an entry, and posed this as an argument against having a common data pool. I assume that we could have a common entry file, and independent indexing files, to allow flexible sorts of personal file-folder procedures.

copy to 445 project files (2)

SOURCE:

DCE 8 JUL 63-3

TITLE:

CONTACT REPORT FROM M I T

COPIES TO:

JHW, TLH, DCL, FKT, CPB, LJC, JWM, JJB

SEE ALSO ITEMS:

FILE FOLDER:

FORMAL DESCRIPTORS:

ABSTRACT:

TEXT:

1. Regarding the enclosed literature (xerox copies). The schedule is for your interest. I would like generally to keep you all up to date on what is going on here in case you want some special information from me or have some special inputs you'd like to make.

Regarding Taylor's think piece on computer aided programming. I like to note that we have anticipated a good deal of his suggested aids in our planning, with the added stress on an integrated system in which the whole job could be compatibly done. You might note that I am going to be on the same session with him and some others next TUESDAY, and that if there is any chance of getting more definition from JHW AND CPB on this system of ours, it would help considerably to strengthen our position in this group talk; i.e., what could make a considerable difference would be a description of procedures that we actually are planning to implement for our first trial system. The strategy of establishing a launching point in the research effort by rapidly accomplishing a working system seems to be very compatible with the attitudes expressed by a number of people so far, and if I can pick that up and go on to describe a system, it would be most beneficial to our situation.

2. Regarding some feedback from Mel Conway and Lick in private conversation today:

a. Conway seemed generally to appreciate and approve the start we were making on the group-cooperation memo scheme, and the way we could apply information-management techniques in on-line updating of our structuring as our concepts and our system practices changed.

b. Lick was quite enthusiastic about said memo system as an example of unique computer aid to a research group. I brought it up in connection with some things I want to try here at MAC, and his interest was so immediate that I went on to describe more of what we were leading up to. My reaction to his reaction was to ask

if perhaps this was the kind of thing that he would consider a valuable output of our research (a working system that we used, with on-line scanning, extraction, manipulation, and, for compatible on-line off-line combination, outputting onto hard copy)--and he seemed quite positive and enthusiastic about saying yes. A good feeling, huh.

3. On the memos that I send, unless otherwise stated, I shall make a paper tape copy here, and you can xerox and distribute them at your end. If I goof on the format, as I will tonight because I didn't bring the standard header form with me--I plan to use either the Expensive Typewriter routing on the PDP-11, or the scope-edit system that I plan to implement here, to go through later and correct the tapes. (Incidentally, Lick says he wants to join me in getting some of these things going here on the PDPS).

copy to CRS project files (2)

AUTHOR: DCE
DATE: 17 JUN 63
TITLE: File-entry Headers
COPIES TO: CFB

SEE ALSO ITEMS: CFB 6 JUN 63, CFB 1 MAY 63, CFB undated (about Feb 63,
"Preliminary Suggestions for mdoc File Structure"

FILE FOLDER:
FORMAL DESCRIPTORS:

ABSTRACT: For memo entries, suggest numbering paragraphs, merging DATE and AUTHOR headings, giving general access to all inter-personal memos, keeping a chronological "journal" record of all entries, providing off-line FILE FOLDER assignments to prior entries, and considering independent personal indices to common-pool data entries.

TEXT: These are suggestions by me, not notice of changed procedure. CFB is the only one with authority to specify changes, and he has the concomitant responsibility for consistency, etc.

1. To make it easier to refer back to specific concept in a prior file entry, perhaps we could serially number every paragraph, independent of headings in the item. These numbers could be entered automatically by the transcription clerk, if a reasonably straightforward convention were established regarding what should be considered for our purposes as a paragraph. Probably wouldn't have to be applied too consistently to be very useful.
2. I also suggest that we combine present Author and Date headings, together with Day Serial Number, into what would become a single, unique, and explicit source code. For example, 2 DCE 17 JUN 63 would refer to the third entry I made today.

The following items are specifically relevant to CFB 6 JUN 63:

3. Regarding the COPIES TO header. What if I write a memo like this one, that principally only concerns two of us, and which might clutter up the notebooks of the rest of the program people--I would be tempted to do as I did, list only you under the COPIES TO heading. But when this gets into machine data base, I should like to have anybody who might have an interest have direct access to it. Shouldn't any memo other than a personal one be available to all who might be tracking down something at the on-line working station?

And, another point here. A person may have a personal-file item that at some time becomes relevant to an interchange--and as you suggested, he can easily transfer that item to the group-communication system. I suggest that in this case the DATE listed in the group file be the date at which the item was transferred from his personal file into the group file. I think that this sort of record control will be important to later untangling the history of ideas that represents the serious study of our conceptual development or of our practising methodology. Still retain as an item of data in that item what its date of entry in the personal system was.

4. The "journal" system is something I favor, where every entry made into either a personal or a group file is entered, in its chronological order relative to other entries, into a permanent "journal" file. I think that it is more important in the group file than in the personal file. A person should be able to copy out of the group journal any data he wants, manipulate it how he sees fit in his personal working domain, and then enter new items into the group file. I think we court disaster to allow changes in the existing prior-entry file, other than perhaps to list subsequent links to more-recent entries. The "safety" of this basic record is very vital.

5. About FILE FOLDER heading: Are you going to provide provision for people to make entries nowadays that establish FILE FOLDER assignments, for their personal Folder relationship, to prior entries? Or do we have to remember about our interest in doing so until we get on-line, where I assume there will be a procedure for establishing personal file folders. Perhaps no more than a personal entry to remind oneself to do this, so he can exercise the on-line procedures later--although if we planned these on-line procedures ahead a bit, we could probably establish a way to designate this from off-line entries. I am quite sure that this off-line designation capability would prove very handy to us later, when we might well want to do such things during our off-line meditation periods.

6. About personal-use descriptors for the FILE FOLDER: You mention that we would have false-drop difficulty if we add independent descriptors to the main-file FILE FOLDER heading of an entry, and posed this as an argument against having a common data pool. I assume that we could have a common entry file, and independent indexing files, to allow flexible sorts of personal file-folder procedures.

copy to IM project file (C)

SOURCE: DCE 8 JUL 63-1

TITLE: SUGGESTIONS TO IM PROJECT FOR SYSTEM CONSIDERATIONS

COPIES TO: CPB, JHW, RCA

SEE ALSO ITEMS:

FILE FOLDER:

FORMAL DESCRIPTORS:

ABSTRACT:

Suggest (a) hard-copy output procedures, (b) real-time data-base modification, and (c) documentation of planning and progress reporting.

TEXT:

Before I left, CPB and I discussed the need to review some of the possibilities for next-step implementation and testing. CPB was to list some that he had in mind. This memo lists a couple of items of this sort that have occurred to me.

a. For our group-cooperation system, based upon memo documentation, to be an effective process in our research activity, where unlike the PECAP system, we can't reasonably consider complete on-line system, it seems necessary to implement procedures for text that has been assembled by on-line manipulation to be printed out as hard copy for off-line use.

There seems to be two sorts of need here. Sometimes we would need to make corrections and modifications in rather large structures that we would like to have available for off-line perusal. Perhaps many pages. For this it would possibly be convenient to have line-printer output, so that we wouldn't worry too much if we needed to redo status statements frequently that were many pages long, or revise our system description writeups to keep them very current, etc.

Another sort of need wouldn't be fulfilled by a line printer, unless it were a very special high-quality machine with both upper and lower case etc. This is where we want rather decent text in more standard and acceptable format. Here our best current bet seems to be to output on the 160a on-line-typewriter.

In either case, there would have to be means for mapping certain CRT-display structures into some acceptable page formats, either in

standard manner or by special-designation processes. I should think that the typewriter output might be a choice for the first of these to implement (unless it proves easier on a line printer) because it seems more flexible, and because it may be a while before we have developed our methods of work to where we have much need for large-volume or fast printout.

b. Implementing the procedures for going back over an established data structure to update or modify its structure or search coding--to do this handily in on-line working mode - would seem to go along with the processes of Item (a) above and - with the generalized-search techniques being implemented now, to make a fairly complete system for the group-cooperation part of our IM project - at least for a first, demonstrative model.

c. But another type of IM system-concern, very important to the group-cooperation scheme, would be the methods by which the participants harnessed the documentation procedures. In particular, how to make more available to the other members of the group the useful information associated with your planning at different stages of your work, and with reports of the progress of your work. Other people have to make plans of their own which often can be affected by either (or both) of your current plans or your current status. They could be materially helped if it were convenient for them to zip into the record and uncover relevant information. Also, other people, from what they have in experience, perspective, skill, etc., that you don't relative to some aspect of your work, can often spot future-trouble spots, or possibilities for improvement in your work or in the ability of someone else to use your work if you change, etc., and there can be considerable value to you if other people can easily browse through certain aspects of your current record.

What I suggest here is that CPB begin to consider this problem and to try establishing pilot policies, methods, formats, etc. for people to use in their memo-file contributions. A kernel aspect of this type of-thing, within the basic pursuit of our program, would be the type of planning and status documentation for programmers to use as they proceed, within an interdependent system, to form an approach toward realizing the problem formulation, as they shape up their data-structuring and process-structuring plan, and as they move through designing the logic and - then the coding, through the debugging etc. One would think that once debugged, after working with good current records, there would be essentially no further work left for the documentation that enables later maintenance; pirating, expanding, modification, etc.

Categories of items, such as assumptions, objectives, considerations regarding pursuit of objectives, method of approach, pertinent facts, relevant reasoning, etc., can be fitted into logical relationships for reporting the planning and progress. To make first steps toward such methodology, consider what it is for instance that Tom and Len might like to know about the plans and progress of one another, things that could be ferreted out with about as much ease as they would experience in going back to check on one of their own earlier decisions. Or, consider what it is that Charlie or John would like to be able quickly and easily to find out about what Len or Tom has planned or done in order that they (CPB or JHW) might better juggle around with system-design planning, or with the planning of their project activities.

We know that records which would help other people find out these things would do a lot for the group progress. But the chances are that it might help the individuals themselves - the process of developing and updating these records might help greatly to make and to keep visible to oneself a clear picture of what it is he wants to do, how he is going to proceed, and just where he is now.

The chances are that it will be much easier to construct and update such records on line rather than off line - at least it will be much more feasible in a practical sense to implement a recording system of any significant sophistication for on-line work.

AP-432 098

RESEARCH ON COMPUTER AUGMENTED INFORMATION MANAGEMENT

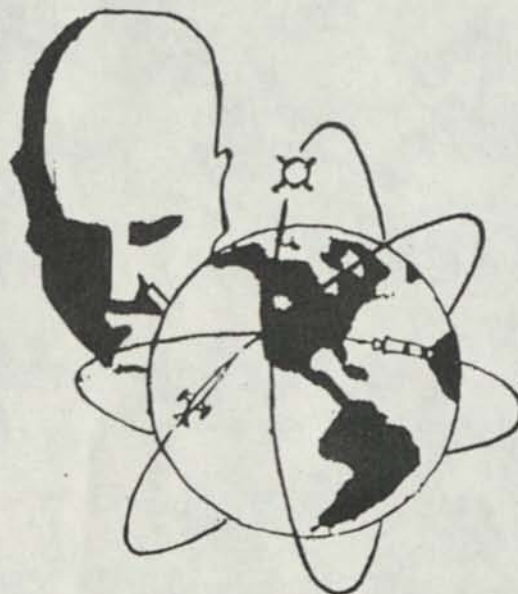
TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-177

NOVEMBER 1963

Charles P. Bourne

*1/4 bd. & cat.
for Library
School Libr.*

DIRECTORATE OF COMPUTERS
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts



SRI Project 4506

(Prepared under Contract AF 19(628)2914 by the Stanford Research Institute,
Menlo Park, California)

ESD-TDR-64-177

RESEARCH ON COMPUTER AUGMENTED INFORMATION MANAGEMENT

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-177

NOVEMBER 1963

Charles P. Bourne

"

ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L.G. Hanscom Field, Bedford, Mass.

SRI Project 4506

(Prepared under Contract AF 19(628)-2914
by the Stanford Research Institute, Menlo Park, California)

From: David Hemmendinger <hemmendd@union.edu>
To: bourmechasp <bourmechasp@aol.com>
Subject: Re: SRI online system
Date: Tue, Mar 19, 2013 2:21 pm

Charlie,

Len's email address is eljayinv@embarqmail.com . His address and phone (or his company's) are:
5660 Castle Drive
Tillamook, OR 97141-8371
503 842-7011

I've searched dtic.mil, but have found references only to your Nov. 1963 report, not the final project report. I expect that you have this information from your report: SRI Project #4506, AF contract 19(628)-2914. Searches for those numbers haven't yielded anything but your report. Given that your report was in late 1963, I'm surprised that the final project report would have been in 1963 rather than 1964, but if the latter, as you said, ntis.gov should list it.'

I'll send my paper draft, but that may not be for a while -- I'm still trying to untangle some bits of Arpanet prehistory and am coming up against conflicting memories and a lack of documentation.

David

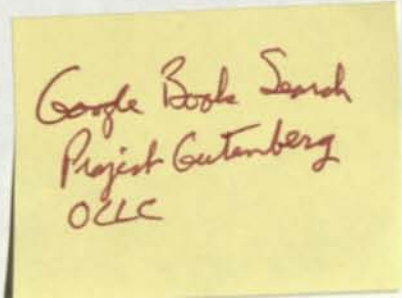
>This story is unfolding rapidly.

>

>Yes, I would appreciate anything further that you could send (citations, a copy of the final project report, Len's contact info, and a copy of your paper.).

>

>I'm in discussions now with the Computer History Museum out here to pass over some of my stuff, and I'd like to see it get into their collection.



Google Book Search
Project Gutenberg
OCLC

From: David Hemmendinger <hemmendd@union.edu>

To: bournechasp <bournechasp@aol.com>

Subject: Re: SRI online system

Date: Sun, Mar 17, 2013 5:05 pm

Charlie,

Thanks for writing. Since I wrote to you, I got a copy of your 1963 report, but it has only a brief paragraph on the network details and so doesn't add anything to what you have in your book, though the rest of the report was interesting to read.

I'll see if I can get the final project report from SRI.

I've also found information on line that might lead me to Len Chaitin -- it's surprising what one can turn up. A 2005 trip report by Bert Raphael, who may have overlapped you at SRI, mentions visiting Chaitin in Oregon, and another search turned up a Tillamook address and phone number that appears to be his.

I agree that it is somewhat surprising that you had good response time from your network connection to SDC. Another report by Larry Roberts, who experimented with a similar connection between Lincoln Labs and SDC in 1966, is that while it did work, it was slow and error-prone.

I'm still gathering information on some of the early networking details, but if you're interested, I can send you a copy of my paper when it's done.

Best wishes,

David Hemmendinger

>You mentioned your review of my book and the Conrad lecture, but you
>didn't mention any review of my original SRI report. That report has the
>complete story, and the details you seek may be in there. The citation
>in the book to a 1963 report of mine may, or may not, provide additional
>details. I've lost most of my early reports, so I can't help resolve
>that point. But I do recall that an SRI project report was written at
>the end of the project, probably co-authored by myself and the main
>programmer, Len Chaitin. That would be worth pursuing. I believe that
>it is available from the Dept. of Commerce's NTIS (National Technical
>Information Service). I tried a quick online search, but their public
>file only starts at 1964. You could also ask your library to contact
>the SRI Library in Menlo Park to see about getting a copy. I lost mine.
>

>Len Chaitin is the person who would have the answers to your questions. I
>lost track of him after I left SRI in 1966.

>

>We never got into anything like a regular use mode with this
>system, so there's no real experience re ease of use. It was really a
>proof-of-concept project. Considering that we'd never seen an online text
>search response before, it seemed pretty fast to us. And in retrospect,
>that was surprising, considering the several places where things could
>have slowed the resulting response quite considerably.

From: David Hemmendinger <hemmendd@union.edu>

To: bournechasp <bournechasp@aol.com>

Subject: Re: SRI online system

Date: Mon, Mar 18, 2013 4:18 pm

I wrote to Len Chaitin this afternoon and shortly afterwards, he called to talk about the 1963 work. It was helpful -- he was able to provide more details about the way that the SDC-SRI network was set up. He also said that he thought that it worked very well for the brief time that it was in operation.

When I remarked that it was probably the first long-distance computer network set up for interactive use, he said, "I hadn't thought of it that way, but you're right," and added that at the time, it just seemed like a useful thing to try to do, and to do as simply as possible. That seems to have been the attitude of SDC people too, which may help to account for how little was published about the networking.

Bert Raphael, who gave me Len's email address, asked me to send you greetings from his wife Ann and him -- he said that she worked with you in the 1960s.

Thanks again for your help.

David

Rates to soar. Residents in the new flood zone adopted by FEMA — which includes many Willows homeowners — should buy their insurance before the zone becomes effective April 21, said Ruben Nino, Menlo Park Director of Engineering.

Residents should make sure they buy the type of insurance which can be grandfathered in, he added. Buying now could mean a savings of \$800 a year. Once the new flood zones become effective, rates could reach up to \$1,200 a year.

Robert Redford. When Robert Redford recently spoke to the Peninsula Open Space Trust, (POST) he thrilled lots of women and some men with wise words about conservation and acting. But nothing topped his opening remark. First, POST President Audrey Rust introduced him with an admiring tribute. Followed by an enthusiastic introduction by Ambassador Bill Lane of Portola Valley. Followed by radio host Michael Krasny with a flattering profile.

When the spotlight finally focused on Mr. Redford, he said in his low-key, aw-shucks way, "For a while I thought I was going to get through this without having to talk."

Always SRI
Computer visionary. The Institooters, SRI International's own jazz band, played as Doug Engelbart, Atherton's computer visionary, was honored recently by the Menlo Park research and consulting firm.

Dr. Engelbart, who invented the computer mouse, received SRI's first Gibson Achievement Award, named after Dr. Wendell B. "Hoot" Gibson, SRI's third employee in 1946 and a guiding light ever since.

As director of SRI's center devoted to augmenting human intelligence, Dr. Engelbart and his team pioneered household computer tools, such as multiple windows, display editing, hypermedia, and on-screen video conferencing. His center was the second node of ARPAnet, the predecessor to the Internet.

This is not Dr. Englebart's first major award. He has received the Association for Computing Machinery's Turing Award, known as the Nobel Prize for computer scientists. He has been inducted into the National Inventor's Hall of Fame, and has received the Institute of Electrical and Electronics Engineers' John Von Neumann Award. He holds 20 patents.

Major change planned in school enrollment policy

■ Freshmen could pick high school of choice. School board criticized for not seeking input from community.

By MARJORIE MADER

Incoming freshmen will have a choice of which high school in the

Sequoia Union High School District they go to next fall if a major policy shift, recommended by the superintendent, is adopted by trustees on Wednesday, March 17.

Superintendent Jo Ann Smith's recommendation, crafted during a study session March 10 held to discuss open enrollment and ethnic balance, would change Sequoia's long-standing enrollment policy that supports having ethnic diver-

sity and balance at the four comprehensive high schools.

The proposal to open enrollment appears to have wide support among board members, but at least one trustee and the president of the teachers' association criticized the speed with which the district is moving. They urged the board to invite input from the district's teachers, parents and the community before making such a dra-

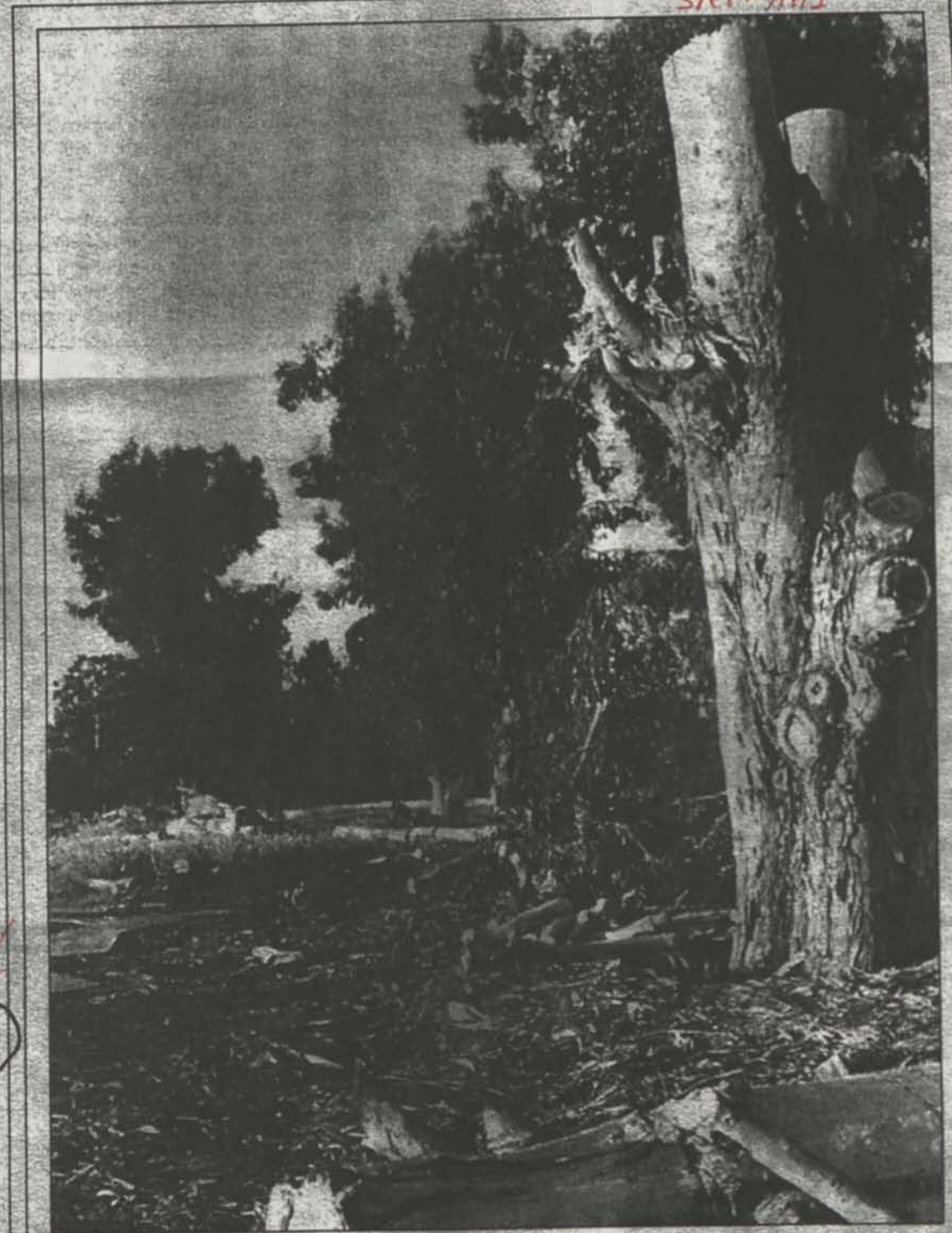


Photo by Carol Nie

Clearing the way

This was the scene last week at Ohlone Field off Sand Hill Road after workers felled dozens of trees — one of the first steps in Stanford's Sand Hill development project. The site will eventually house 628 apartments and 388 senior housing units.

3/17/99 The Almanac

From: Charlie Bourne <bournechasp@aol.com>
To: hemmendd <hemmendd@union.edu>
Subject: Re: SRI online system
Date: Tue, Mar 19, 2013 1:50 pm

David...

This story is unfolding rapidly.

Yes, I would appreciate anything further that you could send (citations, a copy of the final project report, Len's contact info, and a copy of your paper.).

I'm in discussions now with the Computer History Museum out here to pass over some of my stuff, and I'd like to see it get into their collection.

Thanks again.

Charlie Bourne
1619 Santa Cruz Ave.
Menlo Park, CA 94025
650/322-7101

-----Original Message-----

From: David Hemmendinger <hemmendd@union.edu>
To: bournechasp <bournechasp@aol.com>
Sent: Sun, Mar 17, 2013 5:05 pm
Subject: Re: SRI online system

Charlie,

Thanks for writing. Since I wrote to you, I got a copy of your 1963 report, but it has only a brief paragraph on the network details and so doesn't add anything to what you have in your book, though the rest of the report was interesting to read.

I'll see if I can get the final project report from SRI.

I've also found information on line that might lead me to Len Chaitin -- it's surprising what one can turn up. A 2005 trip report by Bert Raphael, who may have overlapped you at SRI, mentions visiting Chaitin in Oregon, and another search turned up a Tillamook address and phone number that appears to be his.

I agree that it is somewhat surprising that you had good response time from your network connection to SDC. Another report by Larry Roberts, who experimented with a similar connection between Lincoln Labs and SDC in 1966, is that while it did work, it was slow and error-prone.

I'm still gathering information on some of the early networking details, but if you're interested, I can send you a copy of my paper when it's done.

Best wishes,

David Hemmendinger

>You mentioned your review of my book and the Conrad lecture, but you
>didn't mention any review of my original SRI report. That report has the
>complete story, and the details you seek may be in there. The citation
>in the book to a 1963 report of mine may, or may not, provide additional
>details. I've lost most of my early reports, so I can't help resolve
>that point. But I do recall that an SRI project report was written at
>the end of the project, probably co-authored by myself and the main
>programmer, Len Chaitin. That would be worth pursuing. I believe that
>it is available from the Dept. of Commerce's NTIS (National Technical
>Information Service). I tried a quick online search, but their public

From: David Hemmendinger <hemmendd@union.edu>
To: bournechasp <bournechasp@aol.com>
Subject: Re: SRI online system
Date: Sun, Mar 17, 2013 5:05 pm

Charlie,

Thanks for writing. Since I wrote to you, I got a copy of your 1963 report, but it has only a brief paragraph on the network details and so doesn't add anything to what you have in your book, though the rest of the report was interesting to read.

I'll see if I can get the final project report from SRI.

I've also found information on line that might lead me to Len Chaitin -- it's surprising what one can turn up. A 2005 trip report by Bert Raphael, who may have overlapped you at SRI, mentions visiting Chaitin in Oregon, and another search turned up a Tillamook address and phone number that appears to be his.

I agree that it is somewhat surprising that you had good response time from your network connection to SDC. Another report by Larry Roberts, who experimented with a similar connection between Lincoln Labs and SDC in 1966, is that while it did work, it was slow and error-prone.

I'm still gathering information on some of the early networking details, but if you're interested, I can send you a copy of my paper when it's done.

Best wishes,

David Hemmendinger

>You mentioned your review of my book and the Conrad lecture, but you
>didn't mention any review of my original SRI report. That report has the
>complete story, and the details you seek may be in there. The citation
>in the book to a 1963 report of mine may, or may not, provide additional
>details. I've lost most of my early reports, so I can't help resolve
>that point. But I do recall that an SRI project report was written at
>the end of the project, probably co0-authored by myself and the main
>programmer, Len Chaitin. That would be worth pursuing. I believe that
>it is available from the Dept. of Commerce's NTIS (National Technical
>Information Service). I tried a quick online search, but their public
>file only starts at 1964. You could also ask your library to contact
>the SRI Library in Menlo Park to see about getting a copy. I lost mine.
>

>Len Chaitin is the person who would have the answers to your questions. I
>lost track of him after I left SRI in 1966.
>

>We never got into anything like a regular use mode with this
>system, so there's no real experience re ease of use. It was really a
>proof-of-concept project. Considering that we'd never seen an online text
>search response before, it seemed pretty fast to us. And in retrospect,
>that was surprising, considering the several places where things could
>have slowed the resulting response quite considerably.

From: David Hemmendinger <hemmendd@union.edu>
To: Bournechasp <Bournechasp@aol.com>
Subject: SRI online system
Date: Mon, Feb 11, 2013 6:45 pm

Mr. Bourne,

I'm doing some research on the early time-sharing system that ran on the Q-32 computer at System Development Corporation, and am writing a paper on it to send to the Annals of the History of Computing. I've read your accounts of having used the Q-32 for your online full-text search project in 1963, both in your book and in your 1999 Conrad Lecture. Your coauthor, Trudi Hahn, kindly gave me your address; I hope that it's all right to ask you about that project.

The network connection that you used between SRI and SDC was an early example of a computer network that was designed for interactive use (it may have been the earliest, though a 1965 one between SDC and MIT is commonly called the first). I'd be interested to know how convenient you found it to use. Other early networking experiments referred to the poor quality and low speed of telephone lines -- did you find that to be a problem?

I've just been in touch with Clark Weissman, one of the programmers of the SDC system. He thought that the network worked well on their side, though he said that had an awkward design, requiring two telephone lines, one for commands and one for data. Were you familiar with that feature, or was it hidden at a lower system level than the one at which you worked for your demonstration?

I realize that I'm asking about things that happened 50 years ago, but I'd be grateful for anything that you can recall.

Thanks,

David Hemmendinger	hemmendd@union.edu
Professor Emeritus	http://athena.union.edu/~hemmendd
Computer Science Dept.	+1 518 346 4489
Union College, Schenectady, NY 12308	FAX: +1 518 388 6789

From: Charlie Bourne <bournechasp@aol.com>
To: hemmendd <hemmendd@union.edu>
Subject: Re: SRI online system
Date: Sun, Mar 17, 2013 1:59 pm

David....

My apologies for the delayed response.

As you noted, 50 years is a long time to remember details.

You mentioned your review of my book and the Conrad lecture, but you didn't mention any review of my original SRI report. That report has the complete story, and the details you seek may be in there. The citation in the book to a 1963 report of mine may, or may not, provide additional details. I've lost most of my early reports, so I can't help resolve that point. But I do recall that an SRI project report was written at the end of the project, probably co-authored by myself and the main programmer, Len Chaitin. That would be worth pursuing. I believe that it is available from the Dept. of Commerce's NTIS (National Technical Information Service). I tried a quick online search, but their public file only starts at 1964. You could also ask your library to contact the SRI Library in Menlo Park to see about getting a copy. I lost mine.

Len Chaitin is the person who would have the answers to your questions. I lost track of him after I left SRI in 1966.

We never got into anything like a regular use mode with this system, so there's no real experience re ease of use. It was really a proof-of-concept project. Considering that we'd never seen an online text search response before, it seemed pretty fast to us. And in retrospect, that was surprising, considering the several places where things could have slowed the resulting response quite considerably.

Good luck with your exploration.

Charlie Bourne
1619 Santa Cruz Ave.
Menlo Park, CA 94025
650/322-7101

-----Original Message-----

From: David Hemmendinger <hemmendd@union.edu>
To: Bournechasp <Bournechasp@aol.com>
Sent: Mon, Feb 11, 2013 6:45 pm
Subject: SRI online system

Mr. Bourne,

I'm doing some research on the early time-sharing system that ran on the Q-32 computer at System Development Corporation, and am writing a paper on it to send to the Annals of the History of Computing. I've read your accounts of having used the Q-32 for your online full-text search project in 1963, both in your book and in your 1999 Conrad Lecture. Your coauthor, Trudi Hahn, kindly gave me your address; I hope that it's all right to ask you about that project.

The network connection that you used between SRI and SDC was an early example of a computer network that was designed for interactive use (it may have been the earliest, though a 1965 one between SDC and MIT is commonly called the first). I'd be interested to know how convenient you found it to use. Other early networking experiments referred to the poor quality and low speed of telephone lines -- did you find that to be a problem?

I've just been in touch with Clark Weissman, one of the

From: Charlie Bourne <bournechasp@aol.com>
To: hemmendd <hemmendd@union.edu>
Subject: Re: SRI online system
Date: Sun, Mar 17, 2013 1:59 pm

David....

My apologies for the delayed response.

As you noted, 50 years is a long time to remember details.

You mentioned your review of my book and the Conrad lecture, but you didn't mention any review of my original SRI report. That report has the complete story, and the details you seek may be in there. The citation in the book to a 1963 report of mine may, or may not, provide additional details. I've lost most of my early reports, so I can't help resolve that point. But I do recall that an SRI project report was written at the end of the project, probably co0-authored by myself and the main programmer, Len Chaitin. That would be worth pursuing. I believe that it is available from the Dept. of Commerce's NTIS (National Technical Information Service). I tried a quick online search, but their public file only starts at 1964. You could also ask your library to contact the SRI Library in Menlo Park to see about getting a copy. I lost mine.

Len Chaitin is the person who would have the answers to your questions. I lost track of him after I left SRI in 1966.

We never got into anything like a regular use mode with this system, so there's no real experience re ease of use. It was really a proof-of-concept project. Considering that we'd never seen an online text search response before, it seemed pretty fast to us. And in retrospect, that was surprising, considering the several places where things could have slowed the resulting response quite considerably.

Good luck with your exploration.

Charlie Bourne
1619 Santa Cruz Ave.
Menlo Park, CA 94025
650/322-7101

-----Original Message-----

From: David Hemmendinger <hemmendd@union.edu>
To: Bournechasp <Bournechasp@aol.com>
Sent: Mon, Feb 11, 2013 6:45 pm
Subject: SRI online system

Mr. Bourne,

I'm doing some research on the early time-sharing system that ran on the Q-32 computer at System Development Corporation, and am writing a paper on it to send to the Annals of the History of Computing. I've read your accounts of having used the Q-32 for your online full-text search project in 1963, both in your book and in your 1999 Conrad Lecture. Your coauthor, Trudi Hahn, kindly gave me your address; I hope that it's all right to ask you about that project.

The network connection that you used between SRI and SDC was an early example of a computer network that was designed for interactive use (it may have been the earliest, though a 1965 one between SDC and MIT is commonly called the first). I'd be interested to know how convenient you found it to use. Other early networking experiments referred to the poor quality and low speed of telephone lines -- did you find that to be a problem?

I've just been in touch with Clark Weissman, one of the

From: Charlie Bourne <bournechasp@aol.com>
To: hemmendd <hemmendd@union.edu>
Subject: Re: SRI online system
Date: Sun, Mar 17, 2013 1:59 pm

David....

My apologies for the delayed response.

As you noted, 50 years is a long time to remember details.

You mentioned your review of my book and the Conrad lecture, but you didn't mention any review of my original SRI report. That report has the complete story, and the details you seek may be in there. The citation in the book to a 1963 report of mine may, or may not, provide additional details. I've lost most of my early reports, so I can't help resolve that point. But I do recall that an SRI project report was written at the end of the project, probably co-authored by myself and the main programmer, Len Chaitin. That would be worth pursuing. I believe that it is available from the Dept. of Commerce's NTIS (National Technical Information Service). I tried a quick online search, but their public file only starts at 1964. You could also ask your library to contact the SRI Library in Menlo Park to see about getting a copy. I lost mine.

Len Chaitin is the person who would have the answers to your questions. I lost track of him after I left SRI in 1966.

We never got into anything like a regular use mode with this system, so there's no real experience re ease of use. It was really a proof-of-concept project. Considering that we'd never seen an online text search response before, it seemed pretty fast to us. And in retrospect, that was surprising, considering the several places where things could have slowed the resulting response quite considerably.

Good luck with your exploration.

Charlie Bourne
1619 Santa Cruz Ave.
Menlo Park, CA 94025
650/322-7101

-----Original Message-----

From: David Hemmendinger <hemmendd@union.edu>
To: Bournechasp <Bournechasp@aol.com>
Sent: Mon, Feb 11, 2013 6:45 pm
Subject: SRI online system

Mr. Bourne,

I'm doing some research on the early time-sharing system that ran on the Q-32 computer at System Development Corporation, and am writing a paper on it to send to the Annals of the History of Computing. I've read your accounts of having used the Q-32 for your online full-text search project in 1963, both in your book and in your 1999 Conrad Lecture. Your coauthor, Trudi Hahn, kindly gave me your address; I hope that it's all right to ask you about that project.

The network connection that you used between SRI and SDC was an early example of a computer network that was designed for interactive use (it may have been the earliest, though a 1965 one between SDC and MIT is commonly called the first). I'd be interested to know how convenient you found it to use. Other early networking experiments referred to the poor quality and low speed of telephone lines -- did you find that to be a problem?

I've just been in touch with Clark Weissman, one of the