

AD703444

AFOSR TR

FINAL REPORT

TO

INFORMATION SCIENCES PROGRAM

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

"USE OF A SOCIAL SCIENCE DATA BASE IN
INTERACTIVE INFORMATION RETRIEVAL"

(Grant AFOSR-69-1805)

D D C
RECEIVED
APR 9 1970
REGISTERED

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield, Va. 22151

Carl Beck
Department of Political Science
University of Pittsburgh
Principal Investigator

1. This document has been approved for
release and sale; its distribution is unlimited.

FINAL REPORT

In June, 1969, we began the development of an interactive information retrieval system that would allow researchers to query a complex, "ragged" dataset interactively and receive information in a form suitable to their research interests. Although it was expected that this would require one year to develop, the system has been developed and is now operating.

We have developed both a full file maintenance system and a full file search system. These two in combination constitute an information maintenance and retrieval system. Each part of the system--update and search--are described in this progress report. A full explication of the search segment of the system occurs in the attached manual.

A. Updating the File:

The update program is designed so that the length, width and content of the file can be manipulated interactively. We can add new sets of information, add specific information, change existing information, edit any card, or delete any card. During this process error messages are given in simple language. A copy program is part of this set of routines which enables us to create new files cheaply and efficiently at no risk to the existing file new information. As an example, in a recent run involving the insertion and editing of one thousand records the recopy of the entire tape and the printing of the entire tape cost \$15.00, a very low cost in view of the fact that the file consists of the equivalent of 35,000 punch cards.

B. Search Systems:

During the preceding six months we designed and implemented a search compiler. This system accepts as input the requests of a researcher who has direct communications with the search system via a remote computer terminal in an online computer system. The output from the search request is a

program in object code which is run directly on the machine. The researcher can enter his search request--which may consist of any number of statements --in a free form. He may also design his output format in any way that he wishes. Included in the system is the ability to create data items from information on the file as well as the ability to access any code category on the file with a unique keyword structure. The advantage of this research system is that it makes the data base fully accessible for any research purposes germane to the database. The development of the search system makes it unnecessary for the researcher to generate special purpose programs for each research request. In most cases the request for output can be satisfied within minutes after the research request is formulated.

SIGNIFICANT SCIENTIFIC AND TECHNICAL ACCOMPLISHMENTS

1. The construction of an update facility which assures the integrity of the original file. The technological developments that allowed this include the development of a structured control facility through the use of sophisticated programming techniques based on Assembler language.
2. The search segment of the CARESS compiler has done for information retrieval what the fortran compilers did for numerical calculations. The search logic implied by the users request is built directly into the search program rather than into an intermediate table. This technique has not been applied in information retrieval. It is amazingly powerful.
3. The creation of a practical, full-powered, transferable information retrieval system operating within the time sharing was done for a ridiculously low sum of money. This information retrieval system can be used on other data bases; it can be used on other machine configurations.

4. We have completed three months ahead of schedule the development of the interactive retrieval system with all of the capabilities outlined in the grant proposal. A manual is available for distribution.

TESTING OF THE SYSTEM

We have now tested this system under a number of conditions. The first test involved undergraduate students in a course on the Government and Politics of Eastern Europe. After a blackboard demonstration of the database and the retrieval program (fifteen minutes) an undergraduate student not acquainted with computers or computing accessed information on those persons purged in Czechoslovakia. The information that he retrieved went well beyond the literature. Subsequently a number of students have utilized the system by reference to the manual alone. Two other data bases are now being formatted in a similar manner. In the next six months CARESS will be demonstrated using portable terminals at meetings in Atlanta, Pittsburgh, Chicago, New York and San Francisco.

INFLUENCE OF SYSTEM ON INFORMATION RETRIEVAL

We have had the extensive inquires from persons interested in the data being accessed, and in the CARESS system. These inquiries come from students and faculty at other institutions as well as persons in the educational and health fields. We are able to demonstrate CARESS from any place which has an operative phone system, and we have been asked to do so at a wide variety of places.

List of Reports

Carl Beck, "Political Biography and the Analysis of Leadership in Communist Political System," Paper at the American Political Science Association Convention, September, 1969.

Gerard A. Johnson and Richard Z. Beizer, "Workshop in Information Retrieval," Workshop at the American Political Science Association Convention, September, 1969.

Edward H. Pechan, III, Career Retrieval System (CARESS): Search System, Published by the Archive on Political Elites in Eastern Europe, University of Pittsburgh, November, 1969.

Edward H. Pechan, III, "The Career Retrieval System: An Overview," submitted for publication

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security Classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

| | | | |
|---|---|---|--|
| 1. ORIGINATING ACTIVITY (Corporate author) University of Pittsburgh University Center for International Studies Pittsburgh, Pennsylvania 15213 | | 2a. REPORT SECURITY CLASSIFICATION UNCL | |
| | | 2b. GROUP | |
| 3. REPORT TITLE USE OF A SOCIAL SCIENCE DATA BASE IN INTERACTIVE INFORMATION RETRIEVAL | | | |
| 4. DESCRIPTIVE NOTES (Type of report and inclusive dates) scientific final | | | |
| 5. AUTHOR(S) (First name, middle initial, last name) Carl Beck | | | |
| 6. REPORT DATE 17 March 1970 | 7a. TOTAL NO. OF PAGES 4 | 7b. NO. OF REFS 0 | |
| 8a. CONTRACT OR GRANT NO. Grant AFOSR-69-1805 | 9a. ORIGINATOR'S REPORT NUMBER(S) | | |
| b. PROJECT NO. 9769-01 | 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR 70-0819TR | | |
| c. 61102F | | | |
| d. 681304 | | | |
| 10. DISTRIBUTION STATEMENT 1. This document has been approved for public release and sale; its distribution is unlimited. | | | |
| 11. SUPPLEMENTARY NOTES TECH OTHER | | 12. SPONSORING/MILITARY ACTIVITY Air Force Office of Scientific Research(SRI) Directorate of Information Sciences Arlington, Virginia 22209 | |
| 13. ABSTRACT CARESS is an interactive information retrieval system that allows any user to query a complex empirical ragged data set without expertise in computing. The system consists of a file maintenance component and a full file search system. The maintenance system assures the integrity of the files; the search logic implied by the users request is built directly into the search program. | | | |

DD FORM 1473

UNCLASSIFIED

Security Classification

14

KEY WORDS

LINK A

LINK B

LINK C

ROLE

WT

ROLE

WT

ROLE

WT

Data set, Ragged
Retrieval
Information
CARESS system
File maintenance
Update
Compilers