Reports and Technical Reports	All Technical Reports Collection

1981-10

TEAM evaluation

Barr, Donald Roy; Howard, Gilbert Thoreau

Monterey, California. Naval Postgraduate School

https://hdl.handle.net/10945/30243

NPS55-81-024PR

NAVAL POSTGRADUATE SCHOOL,
Monterey, California

TEAM EVALUATION

by

Donald R. Barr

Gilbert T. Howard

October 1981

Prepared for: Naval Electronic Systems Command
2511 Jefferson Davis Highway
Arlington, VA   20360

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral J. J. Ekelund   D. A. Schrady
Superintendent   Acting Provost

This report was prepared by:

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55-81-024PR	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) TEAM EVALUATION		5. TYPE OF REPORT & PERIOD COVERED Project Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Donald R. Barr Gilbert T. Howard		8. CONTRACT OR GRANT NUMBER(s) 1711804.6289
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT. PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Electronic Systems Command, Code 330 2511 Jefferson Davis Highway Arlington, VA 20360		12. REPORT DATE October 1981
		13. NUMBER OF PAGES 63
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office) Naval Ocean Systems Center, Code 122 San Diego, CA 92152		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Artificial Intelligence
TEAM
Database Retrieval
Natural Language Access

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

During the period 6 July to 25 September an experiment was devised to evaluate the current version of a program developed by SRI International called TEAM (Teachable English Access data Manager). The experiment involved use of the TEAM software in two modes, database administrator and database user, by 17 officer students in the C^3 curriculum at the Naval Postgraduate School. This report summarizes the experiences these students had in using TEAM, and discusses its strengths and weaknesses from the user's point of view.

TEAM EVALUATION

D. Barr
G. Howard

Department of Operations Research
Naval Postgraduate School
Monterey, CA 93940

ABSTRACT

During the period 6 July to 25 September an experiment was devised to evaluate the current version of a program developed by SRI International called TEAM (Teachable English Access data Manager). The experiment involved use of the TEAM software in two modes, database administrator and database user, by 17 officer students in the C^3 curriculum at the Naval Postgraduate School. This report summarizes the experiences these students had in using TEAM, and discusses its strengths and weaknesses from the user's point of view.

EXECUTIVE SUMMARY

The TEAM software is currently under development by
SRI International's Artificial Intelligence Group. The version
available for the experimentation trials conducted in early
September 1981 was not capable of accessing databases other
than those entered through the EDIT feature of TEAM itself.
The low success rate of answering queries experienced during
the experiment planning phase indicated a simple exercise for-
mat would be appropriate for the experiment. Seventeen officer
students in the Naval Postgraduate School C^3 curriculum par-
ticipated as subjects in the experiment. Each of these subjects
logged onto TEAM via the ARPANET to the TOPS 20 system at the
ACCAT laboratory, NOSC, San Diego. They successfully used the
natural language query system to access a simple data base,
previously prepared by NPS faculty. The subjects also attempted
to use the program ACQUIRE to set up the natural language sys-
tem for other users of this database. Most of the subjects
were not completely successful in this endeavor.

Overall, it appears that TEAM will be a good product,
having real potential for useful applications in information
retrieval from databases. However, the version used in the
experiment is incomplete; a number of improvements and exten-
sions are needed before TEAM can become a truly useful tool.

1. INTRODUCTION

The purpose of this report is to document the experiences of the experimenters and subjects during an evaluation experiment run during the summer of 1981, and to make suggestions and comments concerning apparent strengths and weaknesses of the software being evaluated, called TEAM.

TEAM (Teachable English Access data Manager) is a program developed by SRI International's Artificial Intelligence Group. The program uses artificial intelligence technology to provide a natural (English language) database access capability. It is written in LISP. From the user's point of view, the program consists of two parts. There is an acquisition part in which a "database administrator" teaches TEAM about the database to be accessed, including names and characteristics of fields in the database and English words associated with data in the fields. The second part of the program concerns natural language access, in which a "database user" (quite possibly different from the database administrator) retrieves data from the database.

For those familiar with LADDER, a major difference between that and TEAM is the way in which the grammar and lexicon constituting the natural language foundation is developed. In LADDER it is prepared for a specific database. In TEAM the software works interactively with the user to prepare the access mechanism for the database of interest. TEAM also has an edit and save capability for generating databases. It is planned that the natural language query system in TEAM will be

able to access either such "internal" databases or many "external" databases, developed without using TEAM. It is not possible, with the current version of TEAM, to access external databases, or to link different files in a common database.

In the spring of 1981, the investigators were asked by DARPA to conduct an evaluation of TEAM, with the cooperation of members of the AI unit at SRI. This was undertaken in the summer quarter at NPS, during the period 6 July to 25 September. Several meetings were held with the SRI personnel, including Daniel Sagalowicz, Barbara Grosz and Paul Martin. LCDR Ellen Roland at NPS assisted the authors and attended several of the meetings. The purpose of these meetings was to find out what the current version of TEAM should be capable of doing, and to discuss what kind of evaluation experiment might make sense for this version. The actual conduct and evaluation of the experiment was carried out independently by the authors at the C^3 laboratory at NPS.

2. THE EXPERIMENT

There are many questions about TEAM for which answers would be useful, including:

- Can TEAM accept truly natural language queries?
- To what extent must the user adapt his language to that understood by TEAM?
- To what extent is it necessary for the database administrator to know in advance what questions the user will ask?
- To what extent is it necessary for the user to know how the database administrator carried out the acquisition process?
- Can the database administrator create an interface that will handle a useful range of questions?
- How wide a range of questions can be handled?
- How wide a range of questions is it necessary that a natural language query system be able to handle?
- Is the prompting dialogue presented to the database administrator during the ACQUIRE process adequate? If not, how should it be modified?
- Can military officers who are not particularly knowledgeable in computer science and database theory properly prepare the acquisition phase in the roles of database administrators?
- How much training or experience might be necessary for users of TEAM?

- How does success with TEAM vary over databases of various types and sizes?

These questions range from specific issues involving the current implementation of the TEAM software, to questions about a TEAM-like concept. It was initially our goal to conduct a rather elaborate experiment using subjects with varying amounts of knowledge about a number of databases and the associated ACQUIRE sessions. It seemed possible to create "learning curves", success rates and times required which would represent how learning and initial skills and knowledge interrelated with success in using TEAM.

However, as we began to use TEAM ourselves early in our planning and experimental design phase, it became apparent that many of these goals were too ambitious. In addition, we had scheduling constraints which required the involvement of the NPS C^3 officer students as TEAM users to be completed before mid-September. Thus it appeared to us that the version of TEAM that would be available for the experiment with the C^3 student subjects would have deficiencies rendering it unable to support a sophisticated experiment. Among the "deficiencies" were:

- no access to exterior databases,
- no linking of database files (such a platform file and a weapons file in a ships database),
- various language deficiencies (no "and"; no "not", etc.),

- apparently unpredictable behavior, such as different results (in terms of query successes) for similar databases, and (on rare occasion) different success experience on a given query by different users,
- the documentation available was not user oriented.

Some of these were pointed out to us by the SRI researchers, others were discovered in the course of fairly extensive trial and error experience with TEAM. In addition, we found that our success rate (proportion of queries that seemed reasonable to us that were answered correctly) could not be made higher than about 50%, even with very careful ACQUIRE sessions for very simple databases.

During this period in which we were trying to learn how to use TEAM and to scope out an appropriate evaluation approach, a number of small databases were generated using the EDIT feature within TEAM. These very simple databases typically had four to six fields, with arithmetic, symbolic and feature fields represented. Among these databases were:

- car (several versions)
- plane
- student
- man
- auto
- officer
- oscope
- modl
- ABC.

As an exercise for the experiment subjects, and to generate databases for use in the experiment, an assignment was made in the C^3 laboratory class for each student to generate a small database and a number of queries which might be made of their databases. Just prior to this assignment, Barbara Grosz of SRI came to NPS and gave a seminar to these students about TEAM. A copy of the assignment sheet for this exercise is enclosed in the Appendix, together with two representative responses. The students submitted their work via the ARPANET, and a copy of these candidate databases was FTP'ed to SRI for review by members of the AI group. About a week later, we met with the SRI group for the purpose of selecting "appropriate" databases, from among the student's candidates, for use in the experiment. About half of the database candidat appeared to be useful, and four were selected for use.

The current state of development in the available versi of TEAM led us to conclude that we had to rescale the scope of the experiment. We decided it would be most useful to perform a simple exercise, in which the student-subjects would attempt to use TEAM first as database users, then as database adminis- trators. For the initial database user portion, the subjects would load an ACQUIRE session previously prepared and saved by us. In the second portion, the database administrator portion, the subjects would attempt to complete an acquire session for the same database, then submit queries through these acquisitio In both portions the subject would first ask "canned" queries,

which were known to be successful for the prerecorded ACQUIRE session. After the canned queries, the subjects would ask other queries of their own, in a "freeplay" manner. The successes and failures for all queries would be recorded by each subject, and a record of the session would be captured on the computer.

Using this evaluation approach, essentially an exercise with TEAM for each subject, does provide some useful information, including:

- feedback from officers concerning the strengths and weaknesses of the current version of TEAM,
- the amount of learning that takes place, and the amount of training that might be required,
- features of TEAM that should be modified or added,
- adequacy of the dialogue in the ACQUIRE portion of the TEAM software.

This approach does not provide much information about the general concept of using this approach to database retrieval, nor does it give much of an idea of the potential of later versions of TEAM to overcome some of the problems encountered with the present version.

The TEAM software was installed on the TOPS 20 at the ACCAT laboratory at NOSC, San Diego. It was accessed by the student subjects from the C^3 laboratory at NPS, via the ARPANET. Approximately 5 students were scheduled for each of four sessions, and these students individually logged onto TEAM and carried out the exercise as planned. A copy of the

9

instructions and data forms given each subject is shown in the Appendix. We were present during these sessions to assist the subjects in logging on and getting the TEAM software running, but the subjects were required to complete these database retrieval and acquire sessions by themselves, except in cases of serious problems requiring our intervention to get the exercise running properly again.

The subjects appeared to be keenly interested in the exercise and approached their tasks in a professional manner. We experienced some hardware problems, but in general, the exercise was completed by each subject without serious difficulty. The subjects had no problem understanding the general software structure, and they completed the "canned" queries with the previously prepared ACQUIRE without difficulty. However, in the freeplay portion, the success rate in getting answers to queries constructed by the subjects was low -- about 20%. In the second phase, where each subject prepared his own ACQUIRE, the success rate in getting correct answers to queries (even the canned queries) was quite low -- less than 10%. The subjects found some of the dialogue in the ACQUIRE session to be confusing; this is reported in more detail below.

3. <u>TEAM EVALUATION</u>

This section of the TEAM evaluation report is divided
into five parts. The first deals primarily with issues of
program flow and control and TEAM's interaction with the user.
Sections two through four are confined to comments specifically
related to the program dialog in ACQUIRE, VERBS, and EDIT. The
fifth part is a summary of questions that TEAM failed to answer.

i. Program Flow and Control

A. One of the students experienced a problem during an
ACQUIRE session in which he wanted to modify the current answers.
The problem was his own fault since he answered a question in-
correctly, but a solution to it could be useful for other rea-
sons. The problem arose when the student typed "!" and was asked
if he wanted to modify a previously constructed table. He erro-
neously answered "no" and was forced to go through the ACQUIRE
session for a new table having not yet saved the current session.
His fear was that he would lose the first ACQUIRE session. This
did not occur, but he would like to have been able to "bailout"
of the new ACQUIRE session. The ability to do so would also
allow the user to partially complete an ACQUIRE session and
return later to finish it.

B. One of the earlier versions of TEAM behaved non-
deterministically in that at the beginning of the session it
would answer a simple test question, but later it would not.
The prompt symbol was correct and there was no indication of
anything having gone wrong. This has apparently been fixed in
the latest version, but a similar problem sometimes occurs.
Sometimes after TEAM has failed to answer a question, the wrong
prompt is received (maybe 7: for example). Sometimes if the
user fails to notice this and asks another question, TEAM will
answer it. If the incorrect prompt is noticed and "bailout"
is typed you are returned to the executive level (outside of
TEAM).

Another puzzling problem occurred on September 14 just after the most recent version of TEAM was loaded at NOSC. Two terminals were logged in and using the same (new) ABC database. The question "who makes the ml234 scope" was asked at both terminals. On one terminal it was answered correctly. On the other the responses was that the system had detected a bug. The entire sequence of questions asked on each terminal is not available but the sequence immediately preceding this question is shown in Section 4.

C. The TEAM output should be modified to make the answer easier to find. Many users have no knowledge of LISP and do not like the current output format. The excess information also tends to separate the question from the answer and it is inconvenient to have to search for the question so you can remember what you asked.

D. The "bailout)" feature was a very helpful addition to TEAM. Before it was available, almost any error would force the user to spend a large amount of time in restarting TEAM, reloading the correct table and the user profile.

E. It is important that TEAM be modified so that it can link to external databases.

F. The ability to reenter and modify ACQUIRE was not available to us initially and the addition of that capability was a great help. That software seems to work well except for one puzzle. The ABC database contains a field "cost". Suppose the following questions and answers in ACQUIRE relate to that field:

13

Question	Answer
adjective	high
synonyms	large great
antonyms	small cheap

Now suppose the following questions are asked of TEAM:

	Question	Answer
a)	"What is the largest scope"	correct answer
	(meaning - largest cost scope)	
b)	"What is the smallest scope"	correct answer
c)	"What is the cheapest scope"	BUG

The interesting part about this example is that even if ACQUIRE is modified to make the antonyms "cheap small", TEAM will not answer question (c) above. It will answer (c) if ACQUIRE is originally constricted with the antonyms "cheap small".

    G.   Sometimes when TEAM fails to answer a question it asks "do you still want to go to the database?" This should be eliminated and the answer "no" assumed.

    H.   When TEAM fails to answer a question it gives the answer to the previous question. This should be suppressed.

ii.   ACQUIRE

Generally the students' ACQUIRE sessions seemed to go smoothly with only the usual kinds of questions and comments, some of which are noted below. There is one disturbing fact: even though everything appeared to go well, very few students had success in answering their questions using TEAM. There may be other reasons for this, but assuming that TEAM was working correctly and that the students are representative users, some of the cause of this failure must lie with the inability of ACQUIRE to elicit the proper responses. Out of the students who reported their results on Part II of the exercise, only three had any correct responses to the five sample questions. One of those got correct answers to 3 questions, another 4, and the third got 5 correct answers.

Several comments specifically related to the ACQUIRE program follow:

A.   The name ACQUIRE has no obvious meaning. Other possibilities include PREPARE, GRAMMAR, DIALOG, DEFINE, etc.

B.   Several students stumbled with the primary keys and convenient identifying fields. This is probably a minor point and may affect only first time users, but even more experienced users hesitate since they are not sure of the ramifications of their choices.

C.   The question "Name of file xxx's subject" might be more clear if stated as "what is the subject of this database--use singular."

D. The plural default is a good feature to save the user from too much typing.

E. Several students suggested that the "pronouns" question should be answered with one or more of the numbers rather than with the pronouns themselves.

F. The subject name in the pronoun question appear as plural. It should be singular.

G. The "human" question might be better as "does each entry in the database refer to a human?".

H. The obvious answer to the "name" question is the "name" field when the database contains such a field, but it may be incorrect. For example, in the oscilloscope database the name field referred to the manufacture's name not the scope's name.

I. The uninitiated user need not be bothered with the default question. The sophisticated user can be allowed access to the defaults some other way.

J. The question "?" feature of TEAM was very helpful and was used frequently by the students.

K. The amplified explanation on the "proper name" question is not helpful and should be clarified.

iii.   VERBS

The VERBS program is a good addition to the TEAM software. It resolves a number of problems faced with earlier versions of TEAM.

Some students were observed struggling with the VERBS questions to get the correct form of the verb, but the difficulty probably stems more from forgotten grammar than from deficiencies in TEAM. The explanations were generally satisfactory but required some study before answering.

One observation related to VERBS is that it is difficult to anticipate all the verbs that users might wish to use, and the database administrator is left with a feeling that he has not included all verbs that will be needed.

iv. EDIT

A. The name EDIT is not bad but the name DATA or DATABASE or even EDITDATA is slightly more descriptive.

B. For the terminals used at NPS, to escape from edit requires 2 control Q's. The first repeats the last entry, the second causes an exit from EDIT.

C. The first few students who used TEAM had a problem in trying to escape EDIT since their user profiles had reserved control Q for some other purpose. The students discovered this only after typing the data and then attempting to escape. In this case, it was not serious since the database was small, but it was lost when control C was used.

D. The directory of control characters displayed when EDIT is invoked is hard to read; and several students had trouble moving around in the file.

E. EDIT works very well for data entry and correction when the user is familiar with it. The spacing and prompting with the question mark makes data entry very convenient. It would be convenient to be able to return easily to the previous entry in order to make corrections.

v.   SUMMARY OF QUESTIONS THAT TEAM FAILED TO ANSWER

This section deals with specific databases and specific questions asked of TEAM. For each of the databases mentioned other questions were also asked, but the ones included here are selected to show the areas in which TEAM had difficulty. The databases are considered roughly in chronological order, but at the beginning of the evaluation we did not expect to be dealing with more than one version of the TEAM software, so we did not record the version that produced the responses shown below. This can be reconstructed from the dates associated with each file (table) in TEAM and by knowing the dates that new versions were loaded.

The following summary shows the database name in capital letters, a list of the fields and one sample entry followed by Q, R, and A which have the following meanings.

Q = The exact question asked.

R = The response by TEAM. In case where the exact response was not recorded the symbol x is used.

A = Our analysis of the reason for failure, or a comment.

Naturally, the questions by themselves are meaningless without the ACQUIRE session which supports the database. These sessions should still be available at ISIC and at NOSC.

<u>PLANE</u>

fields :	Model	Name	<u>ID</u>	Manufacturer	HP	Speed	Cost
example:	PA28-130	Cherokee	631	Piper	180	125	28

Q.   What is the fastest Cherokee?

R.   Bug-Soda.

A.   What is the fastest Cherokee plane.

Q.   What is the cost of the plane with ID 631?

R.   Wrong answer.

A.   TEAM ignored "with ID 631".

Q.   Which manufacture has the fastest plane?

R.   Soda.....

A.   That version of TEAM could not handle "has".

MAN

fields : Name   Height

example: Don   72

        Q. Who is not taller than Don?

        R. Wrong answer.

        A. TEAM ignored "not".

        Q. Who is 73 inches tall?

        R. ?

        A. TEAM can not handle this construction.

AUTO

fields : <u>Name</u>   <u>License</u>   <u>State</u>   <u>Transmission</u>

example: Ford   S123   Cal   A

Q.   How many Ford cars are there?

R.   It lists all cars.

A.   TEAM ignored the adjective FORD.

Q.   What is the state of the Datsun car.

R.   It lists all states.

A.   TEAM ignored the adjective Datsun. There are many
     examples of this. Others will not be mentioned.

Q.   What is the state of the Datsun name?

R.   ?

A.   ?

Q.   What is the transmission of S123?

R.   ?

A.   The license is the primary key and TEAM should be
     able to do this.

Q.   What is the name of the car with state Cal?

R.   TEAM lists all names.

A.   TEAM ignores terminal prepositional phrases beginning
     "with ...". There are many examples of this.

22

Q. Is S123 a smoothie?

R. ?

A. "Smoothie" was entered as a concrete noun associated
    with automatic transmission cars. Maybe TEAM needs
    "Is S123 a smoothie car".

Q. How many S123 cars are smoothies?

R. TEAM returned two interpretations - both failed.

A. ?

Q. How many cars have name Ford?

R. Failure.

A. TEAM will answer "how many cars have Ford name".

Q. How many cars with a name of Ford are there?

R. Failure.

A. ?

Q. How many S123 cars are there?

R. X

A. TEAM will answer "how many Cal cars are there".

Q. How many Cal cars have tag S123?

R. X

A. TEAM also failed on other similar questions. The
    problem may be with the verb has. This version of
    TEAM did not have VERBS.

Q. Is the ID of the Ford car S123?

R. Failure.

A. TEAM will answer "what is the name of the ID S123 car".

CAR

fields : Name   License   Weight   Transmission

example: Ford   UXL181   3100   A

    Q. What manual transmission car is the heaviest?

    R. Translating - integrating - succeeded - NIL

    A. ?

    Q. Is there a manual ford?

    R. Translating -- broken.....

    A. ?

    Q. Is there a ford with transmission of A?

    R. Can't be interpreted.

    A. ?

    An earlier version of the car database included
    a field called "doors" containing a number. TEAM
    could not answer the question "how many cars have
    four doors".

OFFICERS

fields : Name   Rank   YOS   Sex

example: Jones   CAPT   15   F

    Q. What is the rank of Jones?

    R. X

    A. X

    Q. How many officers are male?

    R. X

    A. "male" was entered as an abstract noun associated
        with field value M .

fields : <u>Student</u>   <u>Service</u>   <u>QPR</u>   <u>Code</u>

example: Smith   Navy   3.2   G

Q.   What is the branch of service of the student with the lowest QPR?

R.   TEAM gave all branches of service.

A.   This appears to be the same problem observed before, namely that the phrase "with...." is ignored.

Q.   Who has the highest score?

R.   Bug

A.   Score and QPR were synonyms. The problem may be with the verb "has". Several other questions containing "has" failed.

Q.   What is Smith's QPR?

R.   X

A.   We can not determine now if the possessive form was used or not.

Q.   Are there students with Code G?

R.   ...broken...

A.   X

Q.   What is the highest QPR?

R.   X

A.   The "adjective" good<u>est</u> also fails.

THE OSCILLOSCOPE DATABASES

An early part of our evaluation plan called for the use of four databases. In one part of the experiment each student was to interact with one of these databases both as a user and by doing a complete ACQUIRE session. For their "user session" the ACQUIRE was to have been "professionally" prepared by NPS faculty involved in the TEAM evaluation. One of the selected databases dealt with oscilloscopes and was used extensively in various forms for testing by the authors and finally as the subject database for the student exercise. A fairly extensive history is available for this database and it will be reported more extensively than the other databases already mentioned. One of the earliest attempts with an ACQUIRE session was called SCOPE. The success rate in answering questions using SCOPE was very low and the frustration was increased because at that time it was not possible to modify an existing acquire session. We could not tell if the lack of success was due to our own errors in ACQUIRE, or due to inadequacies in TEAM. A message was sent to SRI suggesting that they do ACQUIRE for the oscilloscope database. We do not know if this was done but shortly thereafter the program was changed to allow modification of ACQUIRE.

The next attempt with the oscilloscope data was named MOD1. With this table some degree of success was obtained, but an uncomfortably large fraction of questions were unanswered. At that time, when TEAM failed to answer, it often meant that

28

it was necessary for the user to enter "control C" and start again, although sometimes the command RETFROM (FEVAL) or RETFROM(TPLEX) was successful in returning the correct prompt.

Finally, the database ABC was constructed. The evaluation exercise planned for the students was designed to use the database ABC in two ways. First, they were to act as users asking ten "canned" questions as confidence builders and to familiarize them with TEAM. In the second phase the student were asked to do the ACQUIRE session for themselves and then ask the same ten questions. In both phases, freeplay questions of the students choice were also asked by each student.

When we prepared the ACQUIRE session for ABC, we recorded the answers given so that it could be repeated later if necessary. (This became necessary since ABC vanished from the list of files when the new software was installed at NOSC on September 13). The original ABC Table worked reasonably well but the ten canned questions were carefully selected because TEAM failed on many questions. This is documented in the remainder of this section.

When the ABC file disappeared on September 13, we simply repeated the ACQUIRE session from the notes made in our earlier session and we expected that the new ABC would be identical to the old. This was not so. In fact of the ten test questions previously prepared, the new ABC table would only answer four: (numbers 2, 4, 9, and 10 on the original student handout, part of which appears on page 36.)

29

One of the questions "list the hewlett scopes" was not answered with the new ABC in our first session but it was answered correctly in a later session.

A summary for SCOPE, MOD1, and ABC follows:

SCOPE

fields : <u>Name</u>   <u>Cost</u>   <u>Channel</u>   <u>Sensitivity</u>

example: Smith   555   1   10

    Q. How many channel 1 scopes are there?

    R. X

    A. TEAM also failed on

        "How many 1 scopes are there"

        "How many scopes have a channel of 1"

        "How many 1 channel scopes are there"

        "How many scopes are channel 1"

        "How many scopes are 1"

        "How many scopes have channel channel 1"

TEAM correctly answered the question "How many scopes are there"
indicating that it can interpret the construction "are there".

    Q. How many Smiths are there?

    R. Broken...

    A. TEAM probably needs "how many Smith scopes are there".

31

MOD1

fields : <u>Name</u>   <u>Model</u>   <u>Cost</u>   <u>Code</u>

example: Black   M1234   436   0 (for old)

      Q. Which scopes are new scopes.

      R. Wrong answer.

      A. TEAM counted the new scopes.

      Q. Give the name of all new scopes.

      R. X

      A. Maybe TEAM can not handle "give".

      Q. Find the model for any new Black scope.

      R. X

      A. TEAM may not be able to handle "find".

      Q. For the Black M123 what is the cost.

      R. X

      A. TEAM can not handle "for....."

      Q. Is the model M1234 a new scope.

      R. X

      A. TEAM also failed on "is M1234 new". The problem
          may be that TEAM requires "is the model M1234
          scope new".

Q. What Black scopes are new.

R. X

A. X

Q. What is the cost of M1234.

R. X

A. TEAM correctly answered this question in later
   sessions with the same MOD1 ACQUIRE.

The session from which all of the above questions were
taken also had other problems. For example, even though the
prompt was correct TEAM would not respond to the command "quit".

The next series of questions come from a separate
session with MOD1. Only some of the failures are listed.

Q. What scope has the highest cost.   BUG

Q. Is the HP2125 new.   BUG

Q. What is the model of the lowest price Black   BUG

Q. What Black scopes have code D.   BUG

Q. What is the cost of the highest price old
   scope.   BUG

Q. How many Black scopes are old.   BUG

Q. How many scopes with the name Black are old.   BUG

TEAM will answer "How many old scopes have the name
Hewlett." Notice the verb "have" is included.

Another session with the MOD1 database yielded the
following results.

33

```
Q.  What scope has a cost of 3500 dollars.        X

Q.  3500 dollars is the cost of what scope.       X

Q.  List the scopes with a cost of 3500.          X

Q.  Which scope costs 3500 dollars.               X

R.  X

A.  The verb cost was probably not known.
```

This last example raises a question that has probably been
addressed already in the design of TEAM. This database has
a field called "cost". The word "cost" is also a verb. Is
it necessary for the user in the ACQUIRE session and the VERB
session to avoid such conflict or does TEAM take care of this
internally?

The ABC database

The remainder of the section deals with the Database ABC and its successor "new ABC" which replaced it on September 14.

The ABC database is listed below on page 36.

The table ABC was used repeatedly in preparing the student's evaluation exercise. The ten questions finally selected for ABC and "new ABC" differ from the questions discussed with SRI as reasonable questions for this database. We were unable to successfully answer those questions and found it necessary to replace them with the ten questions shown on page 36. The new ABC database was finally used in the student exercise with yet another set of prepared question. These are shown in the appendix on page 53.

DATABASE ABC

The complete database is shown below.

NAME	MODEL	COST	CODE
BLACK	M1234	436	O
"	M1256	243	O
"	M2237	625	N
SIMPSON	SM113	556	N
"	SM1122	555	O
"	SS3363	999	N
HEWLETT	HP1020	3500	N
"	HP1021	3600	N
"	HP2125	4995	N
"	HP1025	2000	O

The database describes several oscilloscopes. The code field indicates if the scope is Old or New.

QUESTIONS

The following questions (and others) can be answered by TEAM for the oscilloscope database.

1. What is the price of M1234?*
2. What is the highest cost scope?
3. What is the cost of the lowest price scope?
4. Who is the manufacturer of the lowest price scope?
5. List the Hewlett scopes.
6. List the new Hewlett scopes.
7. Who is the manufacturer of M2237?
8. How many scopes have the name Hewlett?
9. What is the model of the lowest price scope?
10. Who makes the lowest price scope?

*The question mark is optional.

In many different terminal sessions the questions on page 36 and others were asked of TEAM, sometimes successfully, sometimes not. Not every session was documented but the following will indicate some of the difficulties we experienced.

August 25

Many questions led to the response BUG, or Translating, Integrating, Succeeded NIL, and on several occasions ATOM HASH TABLE FULL.

September 2

The table was modified to include "cheap" as the third antonym for the adjective "high" which modifies "cost", i.e. "small low cheap". The synonyms for high were "big great large".

    Q. What is the cheapest scope.

    R. BUG

    A. ?

    Q. What is the smallest scope.

    R. OK

    A. No obvious reason why the previous question failed.

    Q. What is the biggest scope.

    R. BUG

    A. ?

Q. What is the highest scope.

R. OK

A. Again, why isn't this like the preceding question?

These results suggest that if cheap were moved to become
the first antonym that TEAM might answer the question "what is
the cheapest scope". This change was made in the ACQUIRE pro-
gram by rejecting each of the present antonyms and replacing
them with "cheap small low" in that order. This change was
saved.

Q. What is the cheapest scope.

R. .....1 page left ---- stack overflow - makelexlist....

A. ?

After typing bailout, load, profile, etc. the following
questions were asked.

Q. What is the highest scope.

R. OK

A. This was accepted as confirmation that TEAM is still
    working.

Q. What is the smallest scope.

R. Stack overflow .....

A. X

Q.   List scope

R.   Unusual CDR ARG LIST

Q.   What are the scopes.

R.   OK

Q.   List the names

R.   OK

Q.   What is the smallest scope.

R.   Stack overflow .....

Q.   What is the highest scope.

R.   OK - the prompt returned was 7: .

Q.   How many Black scope are there.

R.   Not one of the attributes of this table.

A.   Sometimes TEAM answers correctly even when the

     prompt is #:

Q.   List the black scopes.

R.   Not one of the attributes (the correct prompt

     returned)

Q. List the cost of the scopes.

R. OK

Q. How many new scopes are there.

R. OK

Q. What is the cost of the old black scope.

R. Storage full ..... collecting lists .....

<u>new ABC</u> (September 14)

The new ABC database was created and tested on September 14 in preparation for the student's exercise. After it was created, it was tested by asking the ten previously pre-pared questions (page 36). The verbs program was also used when this database was created. Verbs such as "make" and "produce" were included.

Q. What is the price of M1234?

R. Not one of the attributes.

Q. What is the highest cost scope?

R. OK

Q. What is the cost of the lowest price scope?

R. Gives a list of black scopes.

Q. Who is the manufacturer of the lowest price scope?

R. OK

Q. List the Hewlett scopes.

R. ... DB.26 ...

A. This question was later answered correctly!

Q. List the new Hewlett scopes.

R. X

Q. Who is the manufacturer of M2237?

R. X

Q. How many scopes have the name Hewlett?

R. Wrong - it gives 16.

Q. What is the model of the lowest cost scope?

R. OK

Q. Who makes the lowest price scope?

R. OK

Other questions asked of the new ABC database included the
following:

Q. What is the biggest scope?   OK

Q. What is the cheapest scope?   OK

Q. What is the highest cost scope?   OK

Because of the unanticipated results with new ABC, two ter-
minals were logged on to ask questions. The details may be
available in the sessions automatically recorded at NOSC, but
the sequence of questions on each terminal was approximately
as shown below. The question "who makes M1234" was answered
correctly on one terminal but it failed on the other terminal
for some reason. Note that on the second terminal the question
was later answered correctly after "new ABC" was reloaded. In the

42

original session there was no obvious evidence that anything was wrong. Other questions were answered successfully as shown below.

Terminal 1

Q.	List the scopes.	OK
Q.	List the models - give models vs models.	OK
Q.	What is the price of M1234?	X
Q.	List the old scopes.	OK
Q.	Who makes M1234?	X
Q.	Who makes the M1234 scope?	OK
Q.	Who makes the cheapest scope.	OK
Q.	List the new Hewlett scopes.	X
Q.	How many scopes have the name Hewlett - wrong answer	X
Q.	How many old scopes are there?	OK
Q.	Is M1234 an old scope?	OK
Q.	What are the costs of Simpson scopes?	OK
Q.	Who is the manufacturer of SS3363?	X
Q.	Who is the manufacturer of the SS3363 scope?	OK
Q.	Who is the manufacturer of the Hewlett scopes - wrong answer	X
Q.	What is the cost of the lowest cost scope - wrong answer	X
Q.	Does the SM113 scope cost 556 dollars?	X

43

Q. Who makes the lowest price scope?   OK

Q. What is the price of M1234?   X

Q. What is the code of the lowest price scope?   OK

Q. What is the code of M1234?   Not attribute

Q. What is the cost of M1234?   Not attribute

Q. What is the cost of the M1234 scope?   OK

Q. Who makes M1234?   BUG

Q. Who makes the M1234 scope?   BUG

Q. What is the price of black?   BUG

Q. What is the cost of the M1234 scope?   OK

Q. List the Hewlett scopes.   OK

Q. List the new scopes made by Hewlett.   bailout

R. "made" is an unknown word.

A. The verbs session was checked and "made" should be OK.

Q. What new scopes has Hewlett made?

R. storage full, collecting lists.....

At this point the database was reloaded.

Q. Who makes the M1234 scope?

R. Correct answer.

A. Notice this is a question which failed above.

Q. What is the price of M1234?   bailout

Q. Who is the maker of M1234?   bailout

Q. Is M1234 used?   BUG

Q. Is the M1234 scope used?   BUG

5. SUMMARY AND CONCLUSION

• The ACQUIRE, VERBS and EDIT software is generally user-friendly. The ACQUIRE is not particularly easy for the beginner to use, for two reasons: lack of familiarity with the technical terminology used, and lack of understanding of the implications of some of the menu choices. The first difficulty is quickly overcome by experience with the system, and use of the "?" feature, which is generally quite well done. The second difficulty is not escaped as easily, because there is not immediate feedback to the user concerning his choices in the acquire session. There is thus no effective learning by experience with ACQUIRE, beyond learning the technical vocabulary. In our experience the second difficulty was amplified through periodic disruptions with new versions of software. This indicates a need for good, user oriented documentation for TEAM, and especially ACQUIRE.

• Success rates in answering queries are low. For very simple databases, and with ACQUIRE sessions modified, checked and improved over many sessions, we were not able to get more than about 50% success rates. The experimentation subjects achieved approximately a 10% success rate for their "free play" queries.*

• The experimentation subjects were somewhat skeptical about the utility of a natural language query system. They

---

*We do not know why the success rates are so low; it is probably a combination of factors including the database administrator's ACQUIRE choices, the database user's question format or content and deficiencies in TEAM.

felt some compromise between natural language and totally structured syntax might prove to be most "cost effective".

• A working knowledge of English grammar and of database structures is almost necessary for the TEAM user. This might require that some training or "refresher" materials on these subjects be made available to future TEAM users.

• The cooperation of the AI group at SRI was outstanding during the course of our evaluation activities. We are indebted to Daniel Sagalowicz, Barbara Grosz and Paul Martin for their help on this project. We also appreciate the cooperation of the ACCAT laboratory personnel during our evaluation trials on the TOPS 20 system.

APPENDIX 1: Students' Databases

Each student in the C^3 laboratory course was assigned the task of generating a database and a set of queries which could be made of the database. This Appendix contains the assignment instructions and a sample of two responses. The first candidate response, "Oscilloscopes", was used in the experiment (in somewhat modified form). The second candidate shown, "Geographic Database", could not be used because of its structure and the processing required in the proposed queries.

Homework Exercise - C^3 Lab Course
Due date - AUG 4

OVERVIEW:

This is the first step in a series of activities related to data-
base access systems. It is designed to help the student think
about database structure, content, and access.

ASSIGNMENT:

1. Define a database concerning any subject of interest to you.
Limit yourself to about five fields (columns). For each field
write a name for the field and provide a concise description of
its contents.

2. Write ten or more questions typical of those you think appro-
priate to ask of your database. Phrase them in exactly the gram-
matical form you think should be acceptable to a natural language
query system.

3. Show this database populated with three or more entries (rows).

FORMAT:

Please submit this by ARPANET message to R RICHARDS @ ISIE. Please
number the paragraph of your message to correspond to the assign-
ment numbers alone.

Student database #1:

OSCILLOSCOPES DATA BASE

MANUFACTURER NAME	MODEL NO.	COST ($)	CHANNELS #	BANDWIDTH (MHZ)	MAX INPUT SENSITIVITY (V/DIV)
B & K	1420	825	2	15	10
B & K	1432	855	2	15	2
B & K	1405	289	1	5	10
B & K	1466	560	1	10	10
B & K	1520	840	2	20	5
Tektronix	212	1350	2	0.5	10
Tektronix	213	1750	1	1	20
Tektronix	221	1325	1	5	5
Simpson	452	830	2	15	5
Simpson	454	675	2	15	5
Soltec	5101B	495	1	10	10
Soltec	5102B	640	2	10	10

QUESTIONS

1. List the name and model number of all available oscilloscopes which cost less than 1000 dollars.

2. Which one has the greatest bandwidth?

3. How many channels does the soltec model 5102B have?

4. How many models are in the tektronix line?

5. What is the bandwidth of the B & K 1520?

6. What is the maximum input sensitivity of the Simpson 452?

7. Which one has the largest maximum input sensitivity?

8. What is the cost of the tektronic model 213?

9. List all two channel models with a bandwidth greater than 10 MHZ?

10. What is the cost of the cheapest tektronix model?

GEOGRAPHIC DATA BASE

DATABASE DESCRIPTION

The purpose of the geography database is to allow users to ask questions about various international geographic features (such as cities, rivers, etc). Questions would sometimes only require the retrieval of several records and the printing of certain imbedded values. In other cases, a knowledge of the relationships between fields and records and the ability to compute simple mathematical relations would be necessary. The database only defines one record type (other organizations would be possible). Thus some fields would only apply to particular types of records. Climate factors, for example, would only be applicable to placename records. The fields in the database are:

FEATURE NAME - This field would contain the standard recognized name for the "feature". It might be the name of a city, a mountain, a lake or some other type of geographic entity.

TYPE - general type of entity: city, river, mountain etc.

LOCATION - The latitude and longitude coordinates of the entity. For geographically dispersed entities such as rivers we would select an arbitrary point such as the river source.

COUNTRY - National entity containing the specific feature.

50

REGION - This would differ from country to country. Thus a range of synonyms (e.g. state, province, SSR) would be necessary.

POPULATION - obviously only relevant to populated areas.

MEAN ANNUAL TEMPERATURE - Again not applicable to entities like rivers. Measured in degrees Celsius.

MEAN ANNUAL RAINFALL - see above. Measured in inches.

SIZE - A numeric field dependent on the TYPE field for interpretation. For a city it would mean square mile area; for a river, drainage area; for a mountain, height; etc.

TYPICAL QUERIES

The following are typical queries against a geography database. Note that some of them can be answered in a straight-forward manner directly from the database; others require a sophisticated "knowledge" system to produce the answer.

a. What is the average annual rainfall in California?

b. What is the highest mountain in Ethiopia?

c. Which country has the warmest average temperature?

d. Of all cities with population greater than one million, which has the greatest population density?

e. What lakes are located in the Uzbek SSR?

f. Which state in the U.S. is the dryest?

g. How far is Copenhagen from Moscow?

h. What Maryland cities have populations greater than 50,000?

i. What is the coldest country in the world?

j. How big is Mono Lake?

EXAMPLES OF DATABASE RECORDS

      For ease of entry, the rows of the database will be presented as columns:

FIELD	RECORD 1	RECORD 2	RECORD 3
Name	San Francisco	Mt. Whitney	Thames
Type	City	Mountain	River
Location	3745 N 12300W	3630NN 11800W	4250N0015W
Country	USA	USA	UK
Region	California	California	
Temperature	57	43	
Rainfall	20	180	

APPENDIX 2: EXERCISE INSTRUCTIONS, DATA FORMS

Experimentation subjects each logged onto TEAM and completed a database user phase and a database administrator phase. This Appendix contains a copy of the instructions and data forms issued to each subject prior to the exercise.

.

INTRODUCTION

      TEAM stands for Teachable English Access data Manager.
It is being developed by Stanford Research Institute for DARPA.

      In this exercise you will use TEAM in two ways:

1.   You will act as a user of the database system to retrieve
information.

2.   You will serve as the database administrator and will
interact with team to establish the vocabulary that TEAM
requires when it is trying to find data for a user.

For the first part of the exercise you will simply log on and ask
some prepared questions and some of your own if you wish (details
follow). The required vocabulary has already been prepared by
faculty members familiar with TEAM.

      In the second part you will deal with exactly the same
database but you will work with TEAM to provide the vocabulary.
This is done interactively by answering a series of questions
posed by TEAM. It is here that you must inform TEAM of the sub-
ject of the database, the fields (columns) it contains, the fact
that a "scope" and an "oscilloscope" are the same thing, and so
on. When the series of questions and answers is complete, you
should try accessing the database again using the vocabulary you
have created.

      There are several goals to this exercise:

1.   To acquaint you with TEAM.

2.   To give you an opportunity to work with a query system
from the database manager's point of view to help you
appreciate the complexity of natural language query
systems.

3.   To get your help in evaluating TEAM as it now stands.

4.   To get your suggestions for improvements in TEAM. That
is, what capabilities should be built in, and which are
most important.

The complete database for this exercise is shown below.

NAME	MODEL	COST	CODE
BLACK	M1234	436	O
"	M1256	243	O
"	M2237	625	N
SIMPSON	SM113	556	N
"	SM1122	555	O
"	SS3363	999	N
HEWLETT	HP1020	3500	N
"	HP1021	3600	N
"	HP2125	4995	N
"	HP1025	2000	O

The database describes several oscilloscopes for sale by a dealer. The code field indicates if the scope is Old or New.

QUESTIONS

The following questions (and others) can be answered by TEAM for the oscilloscope database.

1. What is the price of M1234?*
2. What is the highest cost scope?
3. What is the cost of the lowest price scope?
4. Who is the manufacturer of the lowest price scope?
5. List the Hewlett scopes.
6. List the new Hewlett scopes.
7. Who is the manufacturer of M2237?
8. How many scopes have the name Hewlett?
9. What is the model of the lowest price scope?
10. Who makes the lowest price scope?

*The question mark is optional.

1. Log on to the TOPS20 using TEAM as the directory and GROSZ as the password.

2. Type   TEAMTOP (CR).

3. Type   LOAD (CR),   answer   ABC (CR).   This is the oscillo-scope database.   Wait until you see   NIL   and the prompt 2_.

4. Type   PROFILE (CR),   answer   Y (CR),   GILFILE (CR).

At this point you should see a prompt sign consisting of a number followed by an underline, for example   3_.   The vocabulary and database were both loaded as   ABC,   and you should be able to ask questions about the database.

5. Type   WHAT IS THE COST OF M1234 (CR - wait for answer and prompt).

6. Continue to type any of the <u>attached</u> questions.*

When you tire of asking the canned questions you will want to experiment and ask other things. Feel free to do so but you must be aware of several things:

a)   The vocabulary previously established may be inadequate to deal with your question.

b)   The inner workings of TEAM may be unable to parse your sentence and formulate a query to the database.

c)   TEAM is a data retrieval system. It does not do computations on the data and it is not able to reason about the data. For example, you will not get an answer to questions like:
     "WHAT IS THE AVERAGE COST OF THE OLD SCOPES," or
     "ARE NEW SCOPES BETTER THAN OLD SCOPES?"

*The question mark is optional.

There are several things that can go wrong with the program in part 1 of the exercise. Sometimes when TEAM fails to answer a question, it will return with a different prompt consisting of a number followed by a semicolon. It may also return a message containing words like "broken" or "U.b.a." In these cases try either of the following commands to return to the lowest prompt:*

       RETFROM(FEVAL)   RETFROM(TPLEX)

If either of these is successful, you may continue asking questions. If not, it is best to CNTL C and return to step 2 above.

---

*A bailout function has now been added. When you see the prompt
 DBn or LANGn where n is a number between 1 and 100, you
 can also try BAILOUT) (CR) .

1. Whenever you have the correct prompt (a number followed by
   an underline), type ACQUIRE (CR) and answer the questions.
   In this case you will be constructing a new table not adding
   to an old one. If in doubt, this prompt can be obtained by
   typing CNTL C, then TEAMTOP (CR) .

The questions posed to you are supposed to be self explanatory,
but if you want more information type a question mark followed
by a carriage return. To modify previous answers, type an excla-
mation point followed by a carriage return. To see all previous
answers type two exclamation points and a carriage return.

2. When the dialog is complete TEAM will spend a short time
   "updating internal data structures." Issue the command
   SAVE (CR) then you will be asked to provide a name for the
   file you have created. Please name it with the first six
   characters of your last name.*

3. Type EDIT and answer the questions (default, then the type
   of terminal is 4). At this point the data base must be typed
   in. When finished, you must type CNTL Q two times. When
   the prompt sign returns, issue the command SAVE (CR).(Same name)

4. Type PROFILE (CR), answer Y (CR), GILFILE (CR).

5. Now you can ask questions of TEAM about the database you
   just entered.

 Please ask the first five of the canned questions in exactly
the form given. Record on the data sheet the response from TEAM.

You may not be satisfied with your previous answers in ACQUIRE.
These can be modified by typing ACQUIRE and answering the
questions asked. (Likewise for VERBS.)

*You can repeat steps 1 and 2 typing VERBs instead of ACQUIRE to
 teach TEAM about verbs that will be useful with your database
 (for example--make).

DATA FORMS

Part 1

Please record here all questions other than the "canned" questions
you asked TEAM to answer about the oscilloscope database. Record
also the answers received in a brief form.

Part 2

A. If you experienced any difficulty in the ACQUIRE session or
    if you did not understand the questions posed by TEAM or the
    explanations offered by TEAM, please explain the difficulty
    here.

After you completed the ACQUIRE session in part 2, what was TEAM's response to the first five canned questions?

Question	Correct Answer	Incorrect Answer	No answer from TEAM
1	☐	☐	☐
2	☐	☐	☐
3	☐	☐	☐
4	☐	☐	☐
5	☐	☐	☐

Comments:

SUMMARY

In addition to enriching the grammatical constructions that TEAM can handle, what enhancements do you feel would be most useful?

DISTRIBUTION LIST

NO. OF COPIES

Director, ACCAT Laboratory Naval Ocean Systems Center San Diego, CA 93152	1
Library, Code 0142 Naval Postgraduate School Monterey, CA 93940	2
Dean of Research Code 012 Naval Postgraduate School Monterey, CA 93940	1
Library, Code 55 Naval Postgraduate School Monterey, CA 93940	1
Professor J. M. Wozencraft Code 74 Naval Postgraduate School Monterey, CA 93940	1
Professor D. R. Barr Code 55Bn Naval Postgraduate School Monterey, CA 93940	15
Mr. William Dejka Naval Ocean Systems Center San Diego, CA 93152	1
CDR Ron Ohlander Information Processing Techniques Office Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209	1
Mr. Frank Deckleman Code 31011 Naval Electronics Systems Command Washington, DC 20360	1
SRI International Artificial Intelligence Center 333 Ravenswood Ave. Menlo Park, CA 94025 ATTN: Daniel Sagalowicz	1

NO. OF COPIES

SRI International Artificial Intelligence   1
  Center
333 Ravenswood Ave.
Menlo Park, CA 94025
ATTN: Barbara Grosz   1

SRI International Artificial Intelligence   1
  Center
333 Ravenswood Ave.
Menlo Park, CA 94025
ATTN: Paul Martin

U1997