

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) System Development Corporation Santa Monica, California		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP	
3. REPORT TITLE Data Base Operations Indicated by Universal Quantifiers			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical Report - May, 1972			
5. AUTHOR(S) (First name, middle initial, last name) Timothy C. Diller			
6. REPORT DATE 25 May 1972		7a. TOTAL NO. OF PAGES 10	7b. NO. OF REFS
8a. CONTRACT OR GRANT NO. DAHC15-67-C-0149 - ARPA		8b. ORIGINATOR'S REPORT NUMBER(S)	
b. PROJECT NO.			
c. ARPA Order #1327, Amendment No. 4		8b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d. Program Code No. 2D30 and 2P10.			
10. DISTRIBUTION STATEMENT Approved for public release; distribution unlimited.			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY	
13. ABSTRACT This paper shows that two different data management operations are required by the use of the English universal quantifiers "all", "each", and "every" in a question-answering system. Each operation is explained and illustrated with queries applicable to a data base of census information.			

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

AD743543

This document was produced by SDC in performance of contract DAHC15-67-C-0149

SERIES	BASE NO	VOL	REVISION
TM	4934	1000	00
AUTHOR <i>Tim Diller</i> Tim Diller			
TECHNICAL T.G. Williams			
RELEASE M.I. Bernstein			
for			
DATE		PAGE 1	
25 May 1972			



TECH MEMO

a working paper

System Development Corporation/2500 Colorado Ave./Santa Monica, California 90406

DATA BASE OPERATIONS INDICATED BY UNIVERSAL QUANTIFIERS

ABSTRACT

This paper shows that two different data management operations are required by the use of the English universal quantifiers "all", "each", and "every" in a question-answering system. Each operation is explained and illustrated with queries applicable to a data base of census information.

D D C
RECEIVED
MAY 19 1972
B

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution unlimited

25 May 1972

1

System Development Corporation
TM-4934/000/00

TABLE OF CONTENTS

<u>Section</u>		<u>Page</u>
1.	INTRODUCTION	2
2.	THE OPERATION "INCLUDED-IN"	2
3.	THE OPERATION "MATCH-RETURN"	3
4.	SAMPLE SENTENCES FOR MORE DETAILED INVESTIGATION	6
5.	CONCLUSION	8

1. INTRODUCTION

This paper shows that in an English Q-A system at least two data management operations are indicated by the universal quantifiers "ALL", "EVERY", "EACH", and the use of "THE" as a universal set indicator. It is also made obvious that an explicit representation of those universal quantifiers is essential in any data management input language. As the two types of data management operations are explained, a few illustrative sentences are given. A later section containing a fuller set of sentences is provided for perusal by anyone wishing to convert the abstract semantic representation into particular data management operators. The quantifier representation in the statements input to the data management system is assumed and not defended.

2. THE OPERATION "INCLUDED-IN"

The instances of the universal quantifier in this section map into a data management function that checks to see that the set universally quantified is a subset of some other set. The strings in the (B) parts of the following examples are a rather abstract level of semantic representation. The universal quantifier is indicated by 'UQ'. The sentence type is indicated by 'YNQ', 'WHQ', and 'IMP', which stand for yes-no question, WH question, and imperative respectively. Case roles of the arguments attached to the predicates are indicated by 'D' (DOMAIN), 'L' (LOCATIVE), 'N' (NEUTRAL), 'VAL' (VALUE) and 'S' (SOURCE). 'RREL' is used to indicate restrictive relative clause modification. And 'INT' stands for set intersection.

In the sentence in (1), the "CITY" set must be a subset of the "SMOGGY" property set (P-set) for a "YES" answer to be appropriate.

- (1) A. ARE ALL CITIES SMOGGY?
B. (YNQ (SMOGGY (D = (UQ (CITIES))))))

In the next sentence, the "CITY" set must be a subset of a derived subset of entries in the D column of the "LOCATED" relational set (R-set), that derived subset being those entries that are paired with an L equal to "CALIF".

- (2) A. ARE ALL OF THE CITIES LOCATED IN CALIF?
B. (YNQ (LOCATED (D = (UQ (CITIES)))(L = CALIF)))

Similarly, the next sentence requires the "CITY" set to be a subset of the subset of entries in the N column that are paired with D = John in the "VISIT" R-set.

- (3) A. DID JOHN VISIT EACH CITY?
B. (YNQ (VISIT (D = JOHN)(N = (UQ (CITY)))))

Finally, the next sentence requires the "CITY" set to be a subset of the subset of entries in the D column of the "POPULATION" function set (F-set).

- (4) A. DOES EVERY CITY HAVE A POPULATION GREATER THAN 50,000?
B. (YNQ (HAVE (D = (UQ (CITY))
(N = (POP (PREL = (GREATER (D = POP)
(VAL = 50,000))))))))

3. THE OPERATION "MATCH-RETURN"

The second type of data management operator called for by UQ's we shall term "MATCH-RETURN". When this operator is called, each entry in the set quantified is matched against another set one at a time while various operations are carried out on other columns in the array.

In the following sentence, each entry in the "CITY" set must be matched in the derived set of cities in the D column, and the value of L must be returned.

Viz.,

- (5) A. WHERE IS EACH CITY LOCATED?
B. (WHQ (LOCATED (D = (UQ (CITY)))(L = WH)))

Thus, for every city actually in the D column, a value for L will be returned. For those cities not in the D column, the response must be "NOT KNOWN".

Similarly, in the next sentence, the N column is matched against by each member of the "CITY" set, and the value of D (if it has an animate entry) is returned. If there is no match in the N column for a particular city, "NOT KNOWN" is returned.

- (6) A. WHO VISITED EACH CITY?
B. (WHQ (VISITED (D = WH, +ANIM)(L = (UQ (CITY)))))

As the following sentence shows, these questions are generally ambiguous. The "MATCH-RETURN" operator, however, is used to produce the answer in either reading. Consider the sentence in (7) while assuming a data structure such as that in (8).

- (7) A. HOW MANY ITEMS ARE FOUND IN EVERY BOX?
B. (WHQ (FOUND (D = WH-COUNT ITEMS)(L IN (UQ (BOX)))))

(8) FIND

E	D	L
1	CANDLE	RED-BOX
2	STRING	RED-BOX
3	WORM	RED-BOX
4	CANDLE	BLUE-BOX
5	BALL	BLUE-BOX

The answer to (7A) can be either "ONE", where the question is paraphrased as follows:

(9) HOW MANY ITEMS DO THE BOXES HAVE IN COMMON?

or "3 IN THE RED-BOX", and "2 IN THE BLUE-BOX", where the question is paraphrased as follows:

(10) HOW MANY ITEMS DOES EACH BOX HAVE?

The "MATCH-RETURN" operator would give the "ITEMS" in the D column for each of the "BOXES" in the L column. It would then be up to another function to either count the number of items connected to each box or count the number of items found in the intersection of the items connected to each box. It appears to be the case that the preferred reading with "EACH" is that in (10) while the preferred reading with "EVERY" is that in (9). I contend that both are ambiguous, however.

In YN questions the "MATCH-RETURN" operator (as well as the "INCLUDED-IN" operator) is needed when at least two noun phrases with UQ's are found. Viz.,

- (11) A. IS EVERY ITEM FOUND IN EVERY BOX?
B. (YNQ (FOUND (D = (UQ (ITEM)))(L IN (UQ (BOX))))))

- (12) A. ITEM B. BOX C. FIND

D
PENCIL
CANDLE
BALL

D
RED-BOX
BLUE-BOX

E	D	L
1	PENCIL	RED-BOX
2	BALL	RED-BOX
3	CANDLE	BLUE-BOX

The "MATCH-RETURN" operator must hand back all of the items for each box, and then the "INCLUDED-IN" operator must determine that the entries in the item set are in each of the sets handed back by the "MATCH-RETURN" operator.

The "MATCH-RETURN" operator must also be called upon in imperative sentences. Consider the significance of the 'UQ' in the following sentences.

- (13) A. PRINT EVERY CITY WITH A POP OVER 50,000!
 B. (IMP (PRINT (D = 2nd PERS. SING.)
 (N = (UQ (INT CITY (POP GR 50000))))))
 VS. C. PRINT A CITY WITH A POP OVER 50,000!

Further, F-sets that have a post-posed prepositional phrase containing a UQ cannot simply assume the UQ. This too must be represented explicitly. Viz.,

- (14) A. IS THE POP OF EVERY CITY GREATER THAN 50,000?
 B. (YNQ (GREATER (D = (INT POP (UQ (CITY))))
 (VAL = 50,000)))
 VS. C. IS THE POP OF ANY CITY GREATER THAN 50,000?
- (15) A. WHAT IS THE POP OF EVERY CITY?
 B. (WHQ (BE (D = WH) (N = (INT POP (UQ (CITY))))))
 VS. C. WHAT IS THE POP OF A CITY?

4. SAMPLE SENTENCES FOR MORE DETAILED INVESTIGATION

This section is intended as an aid to implementors of future data management systems. It will presumably provide a broad enough cross-section of the uses of UQ's to prohibit writing dead-ends into the system that will require major rewrite. The sentences that follow are grouped to deal with P-sets, T-sets, F-sets, and R-sets, in that order. It is also necessary to convert every proposition under scrutiny to its negated form and note the effect on the IL

representation and data management operation.

(16) YNQ: IS EVERY CITY SMOGGY?

WHQ: NOT POSSIBLE

DECL: EVERY CITY IS SMOGGY.

IMP:

- A. LIST EVERY CITY IN THE SMOGGY SET!
- B. STORE EVERY CITY IN THE SMOGGY SET!
- C. DELETE EVERY CITY FROM THE SMOGGY SET!

(17) YNQ: IS EVERY OBJECT A SQUARE?

WHQ: WHAT SHAPE IS EVERY OBJECT?

DECL: EVERY OBJECT IS A SQUARE.

IMP:

- A. LIST EVERY OBJECT
- B. STORE EVERY OBJECT AS A SQUARE!
- C. DELETE EVERY OBJECT FROM THE SQUARES!

(18) YNQ: IS THE POP OF EVERY CITY GREATER THAN 50,000?

WHQ: WHAT IS THE POP OF EVERY CITY?

DECL: THE POP OF EVERY CITY IS 50,000.

IMP:

- A. PRINT EVERY CITY WITH A POP OVER 50,000!
- B. STORE EVERY CITY WITH A POP OVER 50,000 IN THE CLEAR SET!
- C. DELETE EVERY CITY FROM THE POP SET!
- D. PRINT THE POP OF EVERY CITY!
- E. STORE THE POP OF EVERY CITY IN THE POP-VALUE SET!
- F. DELETE THE POP OF EVERY CITY!

25 May 1972

8 System Development Corporation
(last page) TM-4934/000/00

(19) YNQ:

- A. IS EVERY CITY LOCATED IN CAL?
- B. IS DOWNEY LOCATED IN EVERY STATE?
- C. DOES EVERY ITEM APPEAR IN EVERY BOX?

WHQ, TOP:

- A. WHO VISITED EACH CITY?
- B. WHERE IS EACH CITY LOCATED?
- C. WHEN DID EVERY ITEM APPEAR IN EVERY BOX?

WHQ, COMPL:

- A. WHO DID SAM SAY VISITED EVERY CITY?
- B. WHAT DID SAM SAY EVERY VISITOR COMMENTED ON?
- C. WHEN DID SAM SAY EVERY VISITOR INSPECTED EVERY SITE?

DECL, TOP:

- A. EVERY CITY IS LOCATED IN CALIF.
- B. DOWNEY IS LOCATED IN EVERY STATE.
- C. EVERY CITY IS LOCATED IN EVERY STATE.

DECL, EMB, RREL:

- A. WHAT ARE THE CITIES WHICH ARE LOCATED IN EVERY STATE?
- B. WHAT ARE THE CITIES WHICH EVERY VISITOR VISITED?
- C. WHAT WAS THE DAY ON WHICH EVERY VISITOR VISITED EVERY EXHIBIT?

DECL, EMB, COMPL:

- A. BILL EXPECTED EVERYONE TO BUY A SOUVENIER.
- B. BILL EXPECTED SOMEONE TO BUY EVERY SOUVENIER.
- C. BILL EXPECTED EVERYONE TO BUY EVERY SOUVENIER.

5. CONCLUSION

Several points have been brought out. First, more than one data management operation is needed for the English universal quantifier. Second, the semantic representation must invariably contain explicit indications of the universal quantifier. Third, any data management system must consider at least the sentences in section 4.0.