1. THE TROUBLE WITH TRIPLES

SETL NEWSLETTER NUMBER 168

ROBERT B. K. DEWAR 4/13/76

THIS NEWSLETTER DISCUSSES THE PROBLEMS OF REPRESENTATION OF MAPS IN THE SETL LANGUAGE. DISADVANTAGES OF THE CURRENT SCHEME ARE SHOWN AND SOME CHANGES PROPOSED. .

THESE CHANGES ARE TO BE INCORPORATED INTO THE NEW VERSION OF SETL.

1.1. DISCUSSION OF CURRENT SITUATION

CONSIDER THE FOLLOWING THREE SETS IN THE CURRENT FORMULATION OF SETL:

```
  F1 = (/ <1,2> , <2,3,4> /)
  F2 = (/ <1,2> , <2,<3,4>> /)
  F3 = F1 + F2
```

WHEN USED IN THE CONTEXT OF A MAP REFERENCE WITH TWO VARIABLES, THESE SETS ARE TREATED IDENTICALLY

```
  F1(2,3) = F2(2,3) = F3(2,3) = 4
```

HOWEVER, IN OTHER CONTEXTS, THESE SETS ARE DIFFERENT, EVEN THE NUMBER OF ELEMENTS IN F3 IS DIFFERENT FROM F1 AND F2. IN AN ITERATION THROUGH ANY OF THESE SETS, THE ORIGINAL ELEMENTS ARE RETRIEVED IN THE FORM IN WHICH THEY APPEAR IN THE ABOVE ASSIGNMENTS.

F3 IS A PARTICULARLY PECULIAR CASE IN THAT GIVEN THE ASSIGNMENTS:

```
  F4 = F3 LESS <2,3,4>
  F5 = F3 LESS <2,<3,4>>
```

F4 AND F5 STILL GIVE F4(2,3) = F5(2,3) = 4. NOTE THAT THE SITUATION ALSO OCCURS IF ONLY ONE ARGUMENT IS USED FOR THE MAP (F1(2) = F2(2) = <3,4>).

WE THUS HAVE A SITUATION IN WHICH AN ELEMENT OF A MAP MAY BE REPRESENTED IN TWO SEPARATE WAYS WHICH ARE CONSIDERED IDENTICAL IN A MAP CONTEXT, BUT NOT IN OTHER CONTEXTS. THE SITUATION BECOMES MORE COMPLEX WITH LONGER TUPLES. FOR EXAMPLE, THE FOLLOWING ELEMENTS ARE ALL IDENTICAL IN A 6-VARIABLE MAP REFERENCE:

<1,2,3,4,5,6,7>
<1,2,<3,4,5,6,7>>
<1,2,<3,<4,<5,<6,7>>>>>

NOTE HOWEVER, THAT THE FIRST TWO ARE IDENTICAL FOR A 3-VARIABLE MAP REFERENCE, BUT THE THIRD IS DIFFERENT.

THIS CONFUSION IS BOUND TO REFLECT ITSELF IN A SOMEWHAT INNEFICIENT REPRESENTATION FOR SETS WITH THE CAPABILITY OF SUPPORTING ALL OPERATIONS CORRECTLY. THIS IS BECAUSE AN EFFICIENT REPRESENTATION FOR MAP USE MUST TREAT THE VARIOUS POSSIBILITIES AS IDENTICAL, BUT WITHOUT LOSING THE DISTINCTIONS. THIS BECOMES EVEN HARDER IN THE CONTEXT OF THE ABOVE EXAMPLE WHERE ALL THREE ARE IDENTICAL IN ONE CONTEXT AND ONLY TWO ARE IN ANOTHER CONTEXT.

1.2. METHODS OF STORAGE OF MAPS

THERE SEEM TO BE THREE POSSIBILITIES FOR STORING SUCH COMPLEX STRUCTURES:

1) STORE THE SET IN TWO FORMS, ONE SUITABLE FOR MAP REFERENCES, ANOTHER SUITABLE FOR OTHER OPERATIONS. IN GENERAL, N REPRESENTATIONS MAY BE REQUIRED FOR MOST EFFICIENT REPRESENTATION WHERE N-TUPLES ARE INVOLVED.

   THIS SOLUTION GIVES EFFICIENT ACCESS IN ALL MODES, BUT IT IS FAIRLY CLEAR THAT IT IS NOT ACCEPTABLE BECAUSE THE EXPENSE OF MAIN- TAINING THE MULTIPLE REPRESENTATIONS IS TOO HIGH.

2) STORE THE SET WITH ALL 'SIMILAR' ELEMENTS COMBINED, EACH ELEMENT CONTAINING ENOUGH INFORMATION TO RECONSTRUCT THE ORIGINAL SET ELEMENTS.

   AS WE HAVE SEEN, N-TUPLES HAVE MANY POSSIBLE SIMILAR FORMS. GIVEN THE N TUPLE:

   <A1,A2,..........AN>

   THERE CAN BE ANOTHER < FOLLOWING ANY COMMA AND A SIMILAR FORM IS YIELDED, THUS THE NUMBER OF POSSIBILITIES IS 2**(N-1) WHICH CORRESPONDS TO (N-1) INDEPENDENT CHOICES OF PUTTING < AFTER THE (N-1) COMMAS.

   A GIVEN SET CAN CONTAIN OR NOT CONTAIN ANY OF THE 2**(N-1) FORMS, THUS EACH N-TUPLE MUST CONTAIN 2**(N-1) BITS TO INDICATE THE ELEMENTS PRESENT IN THE SET. THIS IS AN EXPLOSIVE NUMBER AND RULES THIS POSSIBILITY OUT WITHOUT NEEDING TO CONSIDER THE PROBLEMS INVOLVED IN ACCESSING AND MODIFYING SETS STORED IN THIS FASHION.

3) THE THIRD POSSIBILITY IS TO RECOGNIZE THAT THERE ARE TWO KINDS OF TUPLES IN THE SET:

   A) PAIRS (TUPLES OF 2 ELEMENTS)
   B) TUPLES OF MORE THAN 2 ELEMENTS

   SUPPOSE THAT FOR EACH HEAD VALUE X, TWO ENTRIES ARE MADE IN THE SET F:

   FXP = SET OF 2ND ELEMENTS OF ALL PAIRS WHOSE HEAD IS X

   FXT = SET OF TAILS OF ALL TUPLES WHOSE FIRST ELEMENT IS X

   THE VALUE OF F(/X/) IS FXP+FXT. THIS UNION MUST ACTUALLY BE COMPUTED IF BOTH SETS ARE NON-NULL, BUT ALMOST NO PROGRAMS USE SUCH MIXED FORMS SO A SIMPLE OPTIMIZATION PREVENTS THE COMPUTATION OF THIS UNION IN THE NORMAL CASE.

   THE SETS FXP AND FXT TREAT PAIRS AND TUPLES IN A CONSISTENT MANNER, SO THAT THE SPLITTING OCCURS AT EACH LEVEL. CONCEPTUALLY, THIS RESULTS IN DIVIDING THE SIMILAR ELEMENTS OF AN N TUPLE INTO 2**(N-1) SETS EACH OF WHICH IS EITHER NULL OR CONTAINS A SINGLE ELEMENT.

   THIS REPRESENTATION WILL BE QUITE EFFICIENT IN PRACTICE SINCE EITHER ONE OR THE OTHER, BUT NOT BOTH SETS WILL BE PRESENT. HOWEVER, DEALING WITH THE UNUSUAL CASES IS COMPLEX AND REQUIRES A LOT OF SPECIALIZED AND PROBABLY CLOSE TO USELESS CODE.

1.3. A SUGGESTED CHANGE

THE FORMULATION OF A SETL RUN TIME SYSTEM BECOMES MUCH SIMPLER IF THIS CONFUSION IS REMOVED. GOING BACK TO THE ORIGINAL EXAMPLE, WE MUST DECIDE WHETHER <2,3,4> OR <2,<3,4>> IS THE PREFERRED FORM FOR MAP ELEMENTS.

ONE OF THESE IS A TWO ELEMENT TUPLE, THE OTHER IS A THREE ELEMENT TUPLE. THE CURRENT SEMANTICS TREAT THESE CASES QUITE DIFFERENTLY, THE RESULT OF A MAPPING OPERATION IS THE TAIL OF THE TUPLE UNLESS IT CONTAINS EXACTLY TWO ELEMENTS, IN WHICH CASE IT IS THE SECOND ELEMENT.

THIS LEADS TO SUCH ANOMALIES AS:

```
F = (/ <A,B,C> /)
B = F(A) EQ <B,C>
```

B IS TRUE. UNLESS C IS UNDEFINED, IN WHICH CASE, B IS FALSE SINCE F(A) = B (NOT <B>).

WE HAVE THE RELATIONS:

```
F1(2,3) = F1(/2/)(3)
F2(2,3) = F2(/2/)(3)
```

THESE DERIVE FROM THE SET THEORETIC BASIS OF SETL AND WE DO NOT WISH TO INVALIDATE THEM. NOW IT IS CLEAR THAT:

```
F2(/2/) = (/ <3,4> /)
```

SINCE THE MAPPING OPERATION MUST WORK ON PAIRS. THIS MEANS THAT F2(2,3) MUST BE 4 SO THAT <2,<3,4>> MUST BE ALLOWED AS A MAP ELEMENT FORM.

WHAT ABOUT F1? CAN WE EXCLUDE THE TRIPLE <2,3,4> AS A MAP ELEMENT FORM? THE ANSWER TO THIS IS THAT WE CAN AND NO INCONSISTENCIES RESULT. IN THIS CASE,

```
F1(/2/) WILL BE THE NULL SET
F1(2,3) WILL BE UNDEFINED
```

SINCE ONE-VARIABLE MAPS ARE REPRESENTED AS SETS OF TWO-ELEMENT TUPLES, IT WAS NATURAL ENOUGH TO EXTEND THIS TO THE REPRESENTATION OF TWO-VARIABLE MAPS WITH THREE-ELEMENT TUPLES, BUT THIS EXTENSION IS NOT COMPATIBLE WITH OFHTER REQUIREMENTS ON MAP SEMANTICS.

IT IS THUS SUGGESTED THAT ONLY PAIRS BE RELEVANT TO THE MAP SEMANTICS. SETS MAY, AS USUAL, CONTAIN OTHER 'JUNK' ELEMENTS, BUT THESE DO NOT AFFECT MAP USE. TRIPLES AND LONGER TUPLES ARE SIMPLY SPECIAL CASES OF 'JUNK' ELEMENTS.

IF A 6-ELEMENT TUPLE IS TO BE PLACED IN A MAP IN SUCH A WAY AS TO ALLOW A 5-VARIABLE MAP REFERENCE, IT MUST APPEAR IN THE FORM:

```
<A1,<A2,<A3,<A4,<A5,A6>>>>>
```

1.4. IMPLEMENTATION ADVANTAGES

FIRST, THE STORAGE OF SETS IS GREATLY SIMPLIFIED, IT IS MERELY NECESSARY TO GATHER ALL PAIRS WITH A COMMON HEAD X AND STORE THE PAIR <X,F(/X/)>. THIS IS EFFICIENT BOTH FOR MAP REFERENCE AND MEMBERSHIP. IT IS A LITTLE INCONVENIENT FOR ITERATION, BUT IS MUCH BETTER THAN THE ALTERNATIVES WITHOUT THIS CHANGE.

ANOTHER ADVANTAGE IS THAT SETS OF LONG TUPLES, WHICH CAN OFTEN OCCUR IN NON-MAP CONTEXTS, ARE STORED IN AN EFFICIENT MANNER AND ARE NOT 'BROKEN UP' TO ALLOW FOR THE POSSIBILITY OF NON-EXISTANT MAP REFERENCES.

THE RUNTIME LIBRARY IS REDUCED IN SIZE SINCE THE OPERATIONS TO BE PERFORMED ARE SIMPLIFIED AS WELL AS MORE EFFICIENT.

1.5. SOME LANGUAGE EXTENSIONS

IF ELEMENTS ARE PLACED IN A MAP IN THE MODIFIED LANGUAGE, ONE OF THE FOLLOWING FORMS MUST BE USED (5-VARIABLE MAP EXAMPLE):

```
A WITH <A1,<A2,<A3,<A4,<A5,A6>>>>>
F(A1,A2,A3,A4,A5) = A6
```

NOTE: THE STATEMENT A WITH B MEANS A = A WITH B.

THE FIRST FORM IS A LITTLE AWKWARD, BUT IT CAN BE ARGUED THAT THE SECOND FORM IS MORE DESIRABLE IN ANY CASE. FORMING A MAP USING THE SECOND FORM INVOLVES AN ITERATION RATHER THAN THE USE OF A SET FORMER.

FOR MULTI-VALUED REFERENCES, THE TWO FORMS ARE:

```
A WITH <A1,<A2,<A3,<A4,<A5,A6>>>>>
F(/A1,A2,A3,A4,A5/) WITH A6
```

ITERATIONS POSE ANOTHER PROBLEM. IF X IS AN ELEMENT FROM A 5-VARIABLE MAP, THEN REFERENCES TO ITS COMPONENTS TAKE THE FORM:

```
X(1)
X(2)(1)
X(2)(2)(1)
X(2)(2)(2)(1)
X(2)(2)(2)(2)(1)
X(2)(2)(2)(2)(2)(1)
```

THIS ADMITTEDLY IS SOMEWHAT GRUESOME, ALTHOUGH MAPS OF THIS MANY VARIABLES ARE NOT COMMON. TWO NEW ITERATOR FORMATS ARE PROPOSED TO SOLVE THIS PROBLEM AS WELL AS MANY OTHER PROBLEMS WHERE THEY PROVIDE MORE ATTRACTIVE FORMS.

1) IN THE FORMS:

```
(/ P, Q->R ! C /)
?Q->R \ C
@Q->R \ C
```

Q MAY BE A GENERAL LEFT HAND SIDE. THE MEANING IS THAT ELEMENTS ARE DRAWN FROM THE SET AND ASSIGNED ONE AT A TIME TO Q USING THE NORMAL ASSIGNMENT SEMANTICS.

IN THE CASE OF ITERATING THROUGH A 2-VARIABLE MAP, WE COULD HAVE:

```
?<X,<Y,Z>>->R \ C(X,Y,Z)
```

THIS TESTS FOR AN ELEMENT IN R MEETING THE CONDITION C WITH ELEMENTS X,Y,Z WHICH ARE SET AND LEFT SET IF AN ELEMENT IS FOUND.

B) THE SECOND FORM IS SPECIALLY INTENDED FOR ITERATING THROUGH MAPS. IT IS SHOWN IN AN EXAMPLE IMITATING THE ONE GIVEN ABOVE:

```
?Z=F(X,Y) \ C(X,Y,Z)
```

ALL THE OBVIOUS EXTENSIONS TO OTHER CASES ARE ALLOWED. IN PARTICULAR, THIS FORM CAN BE USED TO ITERATE THROUGH TUPLES:

```
(@X=TUPLE(I)) BLOCK;
```

AND ALSO CAN BE USED WITH MULTI-VALUED MAPS TO OBTAIN IMAGE SETS INSTEAD OF INDIVIDUAL PAIRS:

```
(/ #B , B = F(/A/) \ A > 3 /)
```

TABLE OF CONTENTS