Bringing SQL to the Masses with Program Synthesis

Chenglong Wang, Alvin Cheung, Ras Bodik University of Washington



stackoverflow

SQL

vin each GROUP BY group?
gesta, lid like to select the first new of each set of news grouped with a "seture" are.
ve got a perchases table that ockalike this
H porchaves;
r 10081
5
2
for the list of the largest surchase (tistal.) made by each isustoner. Something
(id), purtomer, FBRSF(io/al)
tomer al DESC:
customer FIRST(total)
30e 5 511y 5
NAME SOOTA SAME - MARCH
ate values in a SQL table
a duplications with one field:
(DMM)(encl()
H1
Antigensi Joon Antigensi Joon
And good Loom Sobarni Loom
ol#Dasie.com
give us John, Sam, Tam, Tam because they all have the same (easy).
I want is to get dupicates with the same result and make.
to gat "Torn", "Torn",
eed this: I made a mistake, and allowed to insert duplicate mass and reacts read to sense attrange the duplicates, so I meet to drid them limit.
where they it is at 200 merce April 10 or 50 m
Anne Minist Ann. 20. Securit 2 - 20 120
Running Total in SQL Server
I
t, reliable MySOL hosting
following selfs (select TextTable)
relate somenalize
1387/06 - 2 1587/06 - 5
/Tite/WE 0
/15/35 34 /#ar/35 6
a query that returns a running tatal in class order, litta:
nedzte seneralne runnangzztak
(Jar/16) 3
23 BY 26 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
AF182-V86 7 15 JF197-V96 94 44
/flar/06 6 55
i ne vatour wrys o' doing this in SCI. Server 2003 / 2005 / 2005.
larly, intersected in this port of method that uses the aggregating set statement trict:











Observations

greatest-n-per-group

running-total

duplicates

A lot of common tasks require using complex SQL constructs.

Aggregation

Subquery

Exists/In-clauses

Many tasks can be concisely expressed with input-output exar

Idea: summarize our observation on StackOverflow

Transition: these problems can be concisely expressed with examples, can we build some system that allows users to ask question using examples only?



Programming by Example System



id	date	uid
1	12/25	1
2	11/21	3
4	12/24	2

oid	val
1	30
1	10
1	10
2	50
2	10

0ut	Out						
oid	date	uid	oid	MaxVal			
1	12/25	1	1	30			
4	12/24	2	2	10			





Idea: introduce what we want to do: build a PBE system.

Transition: let's first see what is the traditional algorithm to build such system.



Synthesis Algorithm: Value-directed Search

Input: 2, 2, Output: 6, Operators: add, mul



add(2, add(2, 2)) = 6add(2, mul(2, 2)) = 6

144	-one part indepart room	Kutilan AKL WAT HUAVE	er - bood Station		7	D0 = Jag Harry
9. 	+ 1 × × A keer	Б	c	D	E	
1	Full Name	First	Middle	Last		
2	Keith Austen Harper	Keith				
3	Jonas William Smith	Jonas				
4	Elizabeth Jane Arnold	Elizabeth				
5	Harry Robert Downing	Harry				
6	Mary Janet Wright	Mary				
7	Harold George Palmer	Harold				
8	Kathleen Gina Townsend	Kathleen				
9	George M. Miller	George				
10	Mary Elizabeth Patton	Mary				
11						
12						
13						
14						
15						
16						
17.						







Large tables



Insight: Decomposition



Insight: Decomposition



Pro: Smaller space of programs. Challenge: which ones to search for?

Search with Abstract Queries

Input: T1, T2, **Output:** Tout, **Operators:** abstract query operators



Goal: Select * From (Select * From T1 Where \Box) (Select id, Max(val) Join From T2 Where 🗆 Group By oid Having □) T3 **0**n

T6

id	date	uid	oid	MaxVal
1	12/25	1	1	30
2	11/21	3	1	30
4	12/24	2	1	30
1	12/25	1	1	10
2	11/21	3	1	10
4	12/24	2	1	10
1	12/25	1	2	50
2	11/21	3	2	50
4	12/24	2	2	50
1	12/25	1	2	10
2	11/21	3	2	10
4	12/24	2	2	10

U

oid	date	uid	oid	MaxVal
1	12/25	1	1	30
4	12/24	2	2	10



Instantiate Abstract Queries



4 12/24 2 2 10

Instantiate Abstract Queries



True + val < 50 + False + T1.uid = T3.oid False + val < 50 + False + T1.uid = T3.oidTrue + val == 50 + False + T1.uid = T3.oidTrue + val == 50 + MaxVal < 50 + T1.uid = T3.oid

> A intuitive solution that does no scale.

Transition: can we use properties of the abstract query to optimize this?







oid	date	uid	oid	MaxVal
1	12/25	1	1	30
4	12/24	2	2	10

Generating Solutions



. . .

```
Select *
From (Select * From T1 Where id <> uid)
Join (Select id, Max(val)
        From T2
        Where val <> 50
        Group By oid) T3
On T1.uid = T3.oid And T1.id <> T2.id
```

.

- Heuristically rank candidate queries.
 - Criteria: complexity, naturalness etc. •
- When the result is not desirable: ullet
 - Provide new input-output examples.

Ranking & Interaction



Evaluation



Conclusion

- Solution: An efficient two-phase synthesis algorithm.
- Evaluation: Able to solve 143/193 problems on StackOverflow.

Goal: Helping end users to program SQL with input-output examples.