Statistics for representative queries

NELL Dataset

Query	S1	S2	S3	S4	MS1	MS2	MS3	MS4
#Expressions	269	189	2750	901	469	93	3354	918
# Unique variables	1978	341	958	1024	1991	232	979	979

TPC-H Dataset

*Query	Q3	Q4	Q5	Q7	Q8	Q9	Q10
#Expressions	11895	41443	7850	5924	2602	319086	41253
# Unique variables	50782	197392	68797	16117	68797	698806	201725

* The only queries originally yielding read-once provenance were Q1 and Q6 over TPC-H. These queries were less interesting for algorithm comparison since they are join-free, hence their provenance is in the form of disjunctions and all algorithms needed very few probes to find/guess a variable evaluated to **True** per expression.