CONCORDIA UNIVERSITY
Dept. of Computer Science and Software Engineering

COMP6591 – Introduction to Knowledge-Base Systems

𝒫roject 𝒟escription
Submission through Moodle is due on December 2nd at 13:00 PM

PLEASE Note THE NEW "SECTION 4" BELOW with Title "IMPORTANT."

Maximum Points [45%]

In this project, you and your team are required to develop a running prototype for *Annotated Relational Algebra System* (or ARAS, for short), that implements the work of Green et al [PODS 2007] on provenance semirings, posted as paper P1 in the course web page. One input to ARAS will be an Annotated Relational Database D of some application, where the annotation "a" associated with each tuple "t" in each relation "R" in D could represent (1) the multiplicity of t (this means "R" is a bag), (2) the probability of t, (3) the certainty of t, relations for bag semantics), (4) provenance polynomial, or (5) 0 or 1 (for representing the standard case). The second input to the system will be any of the codes 1 to 5, where code i indicates the semantic of the annotations is as described in (i) above. The third and last input to the system is a positive relational algebra (RA) query over D in which the only operators allowed are projection π, select σ, union $\cup$, and natural join $\bowtie$. Note that these are the only operators allowed in Datalog. We use the terms project, select, union, and natjoin to express RA expressions. For example, the RA query expression $project < A, B > (q$ join r join s) means joing the relations q, r, and s and project on attributes A and B, and in so doing, consider the semantics of the annotations.

1 Performance Evaluation

Evaluate the performance of the system for correctness (that is correct set of annotated tuples) and speed, and scalability for large relations.

2 Presentations and Reports

In lectures 12 and 13 of the course, each team will be given a time slot to present their system, its performance, and the highlight of the working systems. Every one in the group should present part of the system. The final report is due on Dec. 10th at 23:55 PM. Devote a section in the report to report which team member contributed to which part(s) of the system.

The maximum points for a good, working system is 30, for report is 10, and for presentation is 5. We may consider additional 5 points for the best system developed.

3   What to Submit

You need to submit THREE files by the deadline mentioned above: (1) the project techincal report in PDF format (not more than 15 pages), (2) the source and supporting library and codes in a single zip file, and (3) the test dataset in a single zip file.

4   Important

Please note that the scope of the implementation is reduced to just the Green's paper on Provenance Semirings (liste as P1 in the course web page). However the scope of the technical paper and your in-class presentations includes the three papers: P1, P2 (on Annotated Reasoning), and P3 (on K-Relations). The PODs will examine the proper working of your ARAS system during the demos while your in-class presentations and technical paper will mainly be a survey of these papers together with a brief introduction of your implementation.