

CMSC 724, Spring 2015: Homework 6

Due Friday May 1, 2015, 11:59pm.

The homework is to be done by yourself. You are free to make reasonable assumptions about the problem if something is not fully specified – make sure to state any assumptions. Try to be concise in your answers. You can turn in a hard copy if you'd like before the deadline (slide under my office door if I am not there). Total weight = 5%.

1. R*-Trees are quite good at indexing 2-dimensional point data, and use rectangles as bounding predicates. Let's say we wanted to use "circles" as bounding predicates instead of rectangles. Explain why we may be able to fit more pointers in each interior node. Second, briefly sketch the "search" algorithm to find whether a given point exists in the database. What are the possible disadvantages of using this approach?
2. (Online Aggregation) Briefly explain the concept of "index striding" and what it is useful for.
3. What is a "stratified sample"? Explain why we must use stratified samples in an AQP system for answering queries with unpredictable query predicates (i.e., in scenarios where the predicate templates are known, but the exact values are not known).
4. (Data Cubes) Consider the example used in the paper, and let's say we have the view "(part, supplier, customer_city)" materialized (note that customer_city is a hierarchical dimension – see Figure 3). Can I use this view to answer a group by query on "(part_type, customer_nation)"? Briefly sketch how, or why not. What about a group by query on "(part, customer_city)" where I am only interested in sales for a particular supplier (say "suppl")? Here are the two queries in SQL (not valid queries – need some joins to make these work):

```
select part_type, customer_nation, sum(sales)
from ...
group by part_type, customer_nation;
```

```
select part, customer_city, sum(sales)
from ...
where supplier_id = suppl
group by part, customer_city;
```

5. (MADLib) Clearly explain the "Counted Iteration via Virtual Tables" approach, what problem it solves, and why it is not sufficient in general (to solve that problem).