Query Optimization – Homework 1

January, 2026

Submit your answer in a pdf file on Canvas.

- Write your name in the file.
- Use this template `https://www.overleaf.com/project/67f01a081d8c577a12f22353`

Grading is done using credit/partial-credit/no-credit; ignore the points below.
An asterix * indicates that the question may be more challenging.

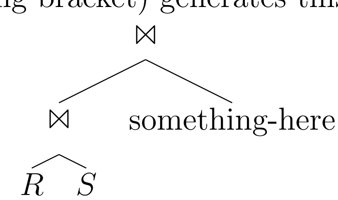
In this homework you will need to draw trees. Feel free to use any method you want to draw them. I am using the following package:

```
\usepackage{qtree}
```

from here `https://www.ling.upenn.edu/advice/latex/qtree/`. For example, the following Latex command:

```
\Tree[.$\Join$ [.$\Join$ $R$ $S$ ] something-here ]
```

(notice the space before each closing bracket) generates this tree:



1 Terms and Substitutions

1. (0 points)

   (a) Let $\Sigma = \{\oplus, \otimes, 0, 1\}$, where $\oplus, \otimes$ have arity 2, and 0, 1 have arity 0. Consider the following terms:

$t_3 = 1$

   i. Consider the Σ-algebra $(\mathbb{N}, +, *, 0, 1)$, where the functions $+, *, 0, 1$ have the standard interpretation. Let $\sigma : \{x, y, z\} \to \mathbb{N}$ be the substitution:

$$\sigma(x) = 5 \qquad \sigma(y) = 2 \qquad \sigma(z) = 4$$

   Compute $\sigma(t_1)$, $\sigma(t_2)$, $\sigma(t_3)$.

   ii. Consider the Σ-algebra $([0, \infty], \min, +, \infty, 0)$, where $\oplus$ is min, $\otimes$ is $+$, and $0, 1$ are $\infty, 0$ respectively. Let $\sigma : \{x, y, z\} \to \mathbb{N}$ be the substitution:

$$\sigma(x) = 5 \qquad \sigma(y) = 2 \qquad \sigma(z) = 4$$

   Compute $\sigma(t_1)$, $\sigma(t_2)$, $\sigma(t_3)$.

   iii. Consider the Σ-algebra $T(\Sigma, X)$, with the canonical interpretation of the operations $\oplus, \otimes, 0, 1$. Let $\sigma : X \to T(\Sigma, X)$ be the substitution:

$$\sigma(x) \stackrel{\text{def}}{=} t_2 \qquad \sigma(y) \stackrel{\text{def}}{=} 0 \qquad \sigma(z) \stackrel{\text{def}}{=} t_2$$

   Compute $\sigma(t_1)$, $\sigma(t_2)$, $\sigma(t_3)$.

   (b) For the same $\Sigma = \{\oplus, \otimes, 0, 1\}$, list (or describe) all identities satisfied by the Σ-algebra $T(\Sigma, X)$.

(c) Consider the algebra consisting of only the join operator: $\Sigma = \{\bowtie\}$. We assume that the join is the *natural join*, under set semantics. For example, $R(A, B) \bowtie S(B, C)$ is an eq-join on B, $R(A, B) \bowtie S(C, D)$ is the cross product, while $R(A, B) \bowtie T(A, B)$ is the set intersection. In this query language every query is a join expression, e.g. $Q = (R \bowtie S) \bowtie (R \bowtie T)$.

Recall that $Q_1 \equiv Q_2$ means that the queries Q_1, Q_2 return the same answer on any input database: $\forall D : Q_1(D) = Q_2(D)$.

Consider now the rule set RS-B0:

Commutativity:	$R \bowtie S \approx S \bowtie R$
Left Associativity:	$R \bowtie (S \bowtie T) \approx (R \bowtie S) \bowtie T$
Right Associativity:	$(R \bowtie S) \bowtie T \approx R \bowtie (S \bowtie T)$

We write, as usual, $Q_1 \approx Q_2$ to denote any consequence of the RS-B0 rules above, in other words $Q_1 \approx Q_2$ is the same as $Q_1 \overset{*}{\leftrightarrow} Q_2$.

Observe that RS-B0 is *sound*: $Q_1 \approx Q_2$ implies $Q_1 \equiv Q_2$. You don't need to prove soundness, but make sure you understand what it says.

i. Prove that the converse fails: $Q_1 \equiv Q_2$ does not imply $Q_1 \approx Q_2$. In other words, RS-B0 is not *complete*. To prove this, give an example of two queries Q_1, Q_2 such that $Q_1 \equiv Q_2$ but $Q_1 \not\approx Q_2$. In other words Q_1, Q_2 are equivalent in the sense that they return the same answer on any input database, yet we cannot prove their equivalence by using only the rule set RS-B0.

ii. Consider the rule set RS-T0 (the *trivial* rule set), with a single rule:

$$R \approx S$$

Indicate whether this rule set is sound, and whether it is complete.

iii. Extend RS-B0 with additional rule(s) so that it becomes a sound and complete set of rules. Your answer only needs to include the additional rule(s): in this question you do not need to provide the proof of soundness and completeness.

iv. * Now prove that the rule set from the previous question is complete.

2 Identities

2. (0 points)

(a) This is Exercise 4.11 from [1]. Let E be the set of the following identities:

$$f^5(a) \approx a \qquad\qquad f^3(a) \approx a$$

Prove that these identities imply: $f(a) \approx a$.
Here $f^5(a)$ means "apply f five times", i.e. $f(f(f(f(f(a)))))$.

(b) You landed on a planet in a galaxy far away, and their inhabitants hired use as a Database Administrator. Their Extraterrestrial Relational Algebra (ERA) works totally differently for what you are familiar with. They have three operations: a binary $\mathbin{⧖}$, and two unary operations σ_a, σ_b. You have no idea what they mean and what they do, but their databases run queries like $(R \mathbin{⧖} \sigma_b(S)) \mathbin{⧖} \sigma_a(S)$ just fine.

On the first day, your new employer asks you to check if these two queries are equivalent:

$$Q_1 = (R \mathbin{⧖} \sigma_a(S)) \mathbin{⧖} T \qquad\qquad Q_2 = (R \mathbin{⧖} \sigma_b(S)) \mathbin{⧖} T$$

Of course, you are clueless, but luckily you found an old user manual stating that the ERA satisfies the following identities:

$$(X \mathbin{⧖} Y) \mathbin{⧖} Z \approx X \mathbin{⧖} (Y \mathbin{⧖} Z)$$

$$\sigma_a(X \mathbin{⧖} Y) \approx \sigma_b(X) \mathbin{⧖} Y \qquad\qquad \sigma_a(X \mathbin{⧖} Y) \approx X \mathbin{⧖} \sigma_a(Y)$$

$$\sigma_b(X \mathbin{⧖} Y) \approx \sigma_a(X) \mathbin{⧖} Y \qquad\qquad \sigma_b(X \mathbin{⧖} Y) \approx X \mathbin{⧖} \sigma_a(Y)$$

You trust that these identities are sound. Complete your task of day one, by proving that $Q_1 \approx Q_2$.

3 Rule Sets

3. (0 points)

   (a) Use RS-B0 (commutativity, left-associativity, right-associativity) to rewrite the expression
   $$(R(A,B) \bowtie S(B,C)) \bowtie (T(C,D) \bowtie U(D,A))$$
   into
   $$(R(A,B) \bowtie U(D,A)) \bowtie (S(B,C) \bowtie T(C,D))$$
   without using any cross products in the intermediate steps.

References

[1] F. Baader and T. Nipkow. *Term Rewriting and All That.* Cambridge University Press, USA, 1999.