Chair of applied computer science III
Prof. Dr. Guido Moerkotte
Email: moerkotte@uni-mannheim.de
UNIVERSITY OF MANNHEIM

Query Optimization — Exercise sheet 3

## Exercise 1

Given the following schema:
$R_1(a, c, d, e)$
$R_2(a, b)$
$R_3(b, c)$
$R_4(d)$
$R_5(e)$

Determine and classify the query graph induced by each of the following queries. If we don't consider cross products, then how big is the search space for left-deep join trees and busy join trees? Are there deterministic algorithms that find the optimal join order in this query graph's search space in polynomial time?

### Exercise 1 a)

```
SELECT DISTINCT *
FROM R1, R2, R3, R4
WHERE R1.c = R3.c
  AND R2.b = R3.b
  AND R1.d = R4.d;
```

### Exercise 1 b)

```
SELECT DISTINCT *
FROM R1, R2, R3, R5
WHERE R1.a = R2.a
  AND R1.c = R3.c
  AND R1.e = R5.e
  AND R2.b = R3.b;
```

### Exercise 1 c)

```
SELECT DISTINCT *
FROM R1, R2, R3, R5
WHERE R1.a = R2.a
```

```
  AND R1.c = R3.c
  AND R1.d = R4.d
  AND R1.e = R5.e;
```

## Exercise 2

### Exercise 2 a)

Implement `GOO`. You may use the `QueryGraph` and `Tree` classes provided in the solution code.

### Exercise 2 b)

In *Introduction to the Design & Analysis of Algorithms* the author Anany Levitin desribes a greedy algorithm to compute Huffman encodings. Read chapter 9.4 Huffman Trees and Codes. Observe the similarity between the described algorithm and `GOO`. Adjust of your `GOO` implementation so that it can create Huffman trees for appropiate inputs. Note that it is easy to deduce the corresponding Huffman encoding from a Huffman tree.
Try to read it up in the book.
If you have no access to the book, you'll find a description at
`https://en.wikipedia.org/wiki/Huffman_coding#Compression`

## Exercise 3

For relation $R_1$ and $R_2$ you are given their cardinality $|R_i|$, the number of distinct values in an attribute $A$, denoted by $d(R_i.A)$, the mininimum and maximum value $min(R_i.A)$ and $max(R_i.A)$, and key/ foreign-key information.

Note: You will find the answers, as given in the solution, in the paper:
`https://www2.cs.duke.edu/courses/compsci516/cps216/spring03/papers/selinger-etal-1979.pdf`
This paper is the first published article on (cost-based) query optimization!

### Exercise 3 a)

How could you estimate the selectivity of the predicate $\sigma_{R_1.A=c}$, where $c$ is a constant value? Under what assumptions is your estimate accurate?

### Exercise 3 b)

How could you estimate the selectivity of the predicate $\sigma_{R_1.A>c}$, where $c$ is a constant value? Under what assumptions is your estimate accurate?

### Exercise 3 c)

How could you estimate the selectivity of the join predicate $R_1.A = R_2.B$? Under what assumptions is your estimate accurate?

### Exercise 3 d)

How could you estimate the selectivity of the conjunctive predicate $p_1 \wedge p_2$? Under what assumptions is your estimate accurate?