CSC3170 Tutorial 3

Haijin Wang

2023.2.14

1.

For some relation schema R with $K \subseteq R$, denote $p(K)$ as the selection predicate considering attributes K. If we have $\kappa \in K$, we may write $p(\{\kappa\})$ as $p(\kappa)$ for convenience. Some examples for $p(\kappa)$ are: $\kappa = v, \kappa < v, \kappa > v$. We shall add a supplementary definition for the selection operation, such that $\sigma_{p(\widetilde{K})}(r) = r$ if $\widetilde{K} \nsubseteq R$ with relation $r(R)$.

Now, suppose we get two relation schemas R and S. Relation schema R has (but may not only has) attributes α, β, i.e., $\{\alpha, \beta\} \subseteq R$, while Relation schema S has attribute β but doesn't have attribute α, i.e., $\alpha \notin S$ and $\beta \in S$.

Consider two non-empty relations $r(R)$ and $s(S)$ with the following two queries:

- Query 1: $\sigma_{p(\alpha)}(r \bowtie_{r.\beta=s.\beta} s)$
- Query 2: $\sigma_{p(\alpha)}(r) \bowtie_{r.\beta=s.\beta} s$

1.

(a) **Determine if these two non-identical queries are equivalent.** Please use relational algebra for rigorous proof (Hint: with the supplementary definition for operator select mentioned above, we suppose the select operator is of commutative property, that is $\sigma_{p_1}(\sigma_{p_2}) = \sigma_{p_2}(\sigma_{p_1})$; and we also suppose that the binary operator Cartesian product is of distributive property considering the unary operator select in this case, or say $\sigma_{p(\alpha)}(r \times s) = \sigma_{p(\alpha)}(r) \times \sigma_{p(\alpha)}(s)$ with $\alpha \in R$ or $\alpha \in S$ using formal expression).

1.

(b) **Tell which query is of higher efficiency.** Here we shall compare the number of computational operations of the two queries. We suppose the query execution engine will use sequential search when making loops or checking records satisfying some condition, and the operation of looping and checking will not be counted twice. Denote the number of tuples in r as m, i.e., $m = |r| \geq 1$, and the number of tuples in s as n, i.e., $n = |s| \geq 1$. Suppose in Query 1, the fraction of the number of returned tuples after selection $r.\beta = s.\beta$ over $r \times s$ to that of $r \times s$ is g, that is to say, we shall have the equation $g = |r \bowtie_{r.\beta = s.\beta} s| / |r \times s|$. To simplify the question, here we set $g = 1$. Suppose in Query 2, the fraction of the number of returned tuples after selection $p(\alpha)$ over r to that of the original r is f, that is, $f = |\sigma_{p(\alpha)}(r)| / |r|$. We shall have $0 \leq f \leq 1$ (Hint: calculate the difference of the number of operations and compare it with 0).

2.

Consider a university database, which stores the following information:

- students, including student-id, name, and program;
- instructors, including identification number, name, department, and title;
- courses, including number, title, credits, syllabus, and prerequisites;
- course offerings, including course number, year, semester, section number, instructor(s), timings, and classroom.

The enrollment of students in courses, and grades awarded to students in each course they are enrolled for should be represented. Supposing that a class meets only at one particular place and time, and that any given room allows multiple class meetings (through partitioning the room).

(a) Design an E-R diagrams for this situation, indicating the data attributes and stating clearly any assumptions that you make. You are suggested to use the notation (not the alternative ones) introduced in the slides.

(b) Construct relational schemas for the above E-R diagram.

3. Design a generalization–specialization hierarchy for a motor-vehicle sales company. The company sells motorcycles, passenger cars, vans, and buses. Assume that there are the two categories of vehicles: commercial and non-commercial. Note that each vehicle would attract a general sales tax, as well as an additional tax applicable to its category. You should determine the attributes of each entity type and indicate these attributes using appropriate notations. You should state any assumptions you make. Please at least draw the ER diagram, and it’s suggested to use the notation (not the alternative ones) introduced in the slides.

4-10. The following are related to Questions 4-10. Consider the following database schema, where the primary keys are underlined.

EMPLOYEE (Fname, Minit, Lname, Ssn, Bdate, Address, Sex, Salary, Super_ssn, Dno)
DEPARTMENT (Dname, Dnumber, Mgr_ssn, Mgr_start_date),
DEPT_LOCATIONS (Dnumber, Dlocation),
WORKS_ON (Essn, Pnumber, Hours),
PROJECT (Pname, Pnumber, Plocation, Dnum)
DEPENDENT (Essn, Dependent_name, Sex, Bdate, Relationship)

where Fname signifies first name; Minit, middle initial; Lname, last name; Ssn, Essn are the social security numbers; Super_ssn is the social security number of the supervisor; Dname, Dnum, Dno, Dlocation are department name, number and location (similarly for projects); while other attributes have an obvious interpretation.

4-10. 4. Retrieve the name and address of all employees who work for the “Research” department.

5. For every project located in ‘Stafford’, list the project number, the controlling department manager’s last name, address, and birth date.

6. Make a list of project numbers for projects that involve an employee whose last name is “Smith”, either as a worker or as a manager of the department that controls the project.

4-10. 7. Retrieve the names of employees who have no dependents.

8. List the names of managers who have at least one dependent.

9. Find all employees directly supervised by “James Borg”.

10. Find all employees directly supervised by those directly supervised by “James Borg”. Would it be possible to find all employees supervised by a given employee at all levels?