Lecture 9

Database Usage (and Construction)

Relational Algebra:
Summary

Relational Algebra I

- Select
  - $\sigma_{\text{<selection condition>}}(R)$
- Project
  - $\pi_{\text{<attribute list>}}(R)$
- Rename
  - $\rho_{\text{<new schema>}}(R)$
- Union
  - R U S

Relational Algebra II

- Intersection
  - R ∩ S
- Difference
  - R - S
- Cross Product
  - R x S
- Join
  - $R \bowtie_{\text{<join condition>}} S$

Relational Algebra III

- Natural Join
  - $R \bowtie S$
- Division
  - R ÷ S
- Grouping
  - $\gamma_{X,G}(R)$

Selection & Projection

name	branch	salary
Andersson	3	20000
Johnsson	3	25000
Larsson	3	32000
Persson	2	28000
Svensson	2	35000

Selection

$\sigma_{\text{name='Andersson'}}(R)$

name	branch	salary
Andersson	3	20000

"Horizontal" decompositions

name	branch	salary
Andersson	3	20000
Johnsson	3	25000
Larsson	3	32000
Persson	2	28000
Svensson	2	35000

Projection

$\pi_{\text{name, salary}}(R)$

name	salary
Andersson	20000
Johnsson	25000
Larsson	32000
Persson	28000
Svensson	35000

"Vertical" decompositions

Outer Join Operators

R1

name	branch
Andersson	3
Johnsson	3
Larsson	1

R2

branch	salary
3	20000
4	25000

$R1 \bowtie_{\text{LEFT}} R2$

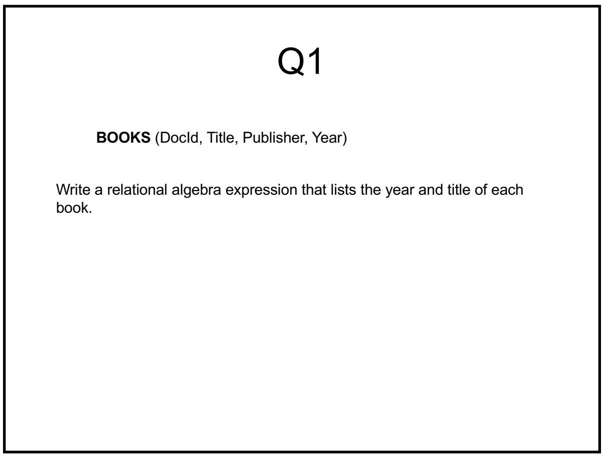
name	branch	salary
Andersson	3	20000
Johnsson	3	20000
Larsson	1	Null

$R1 \bowtie_{\text{RIGHT}} R2$

name	branch	salary
Andersson	3	20000
Johnsson	2	20000
Null	4	25000

$R1 \bowtie_{\text{FULL}} R2$

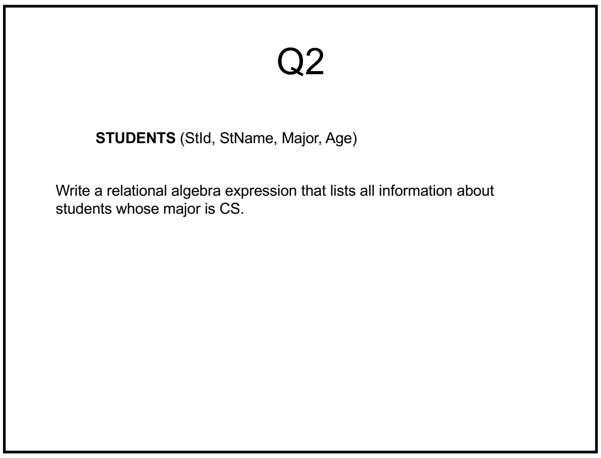
name	branch	salary
Andersson	3	20000
Johnsson	2	20000
Larsson	1	Null
Null	4	25000



Q1

BOOKS (DocId, Title, Publisher, Year)

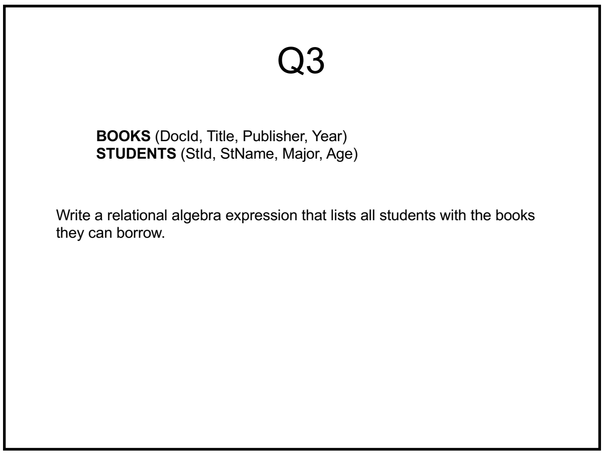
Write a relational algebra expression that lists the year and title of each book.



Q2

STUDENTS (StId, StName, Major, Age)

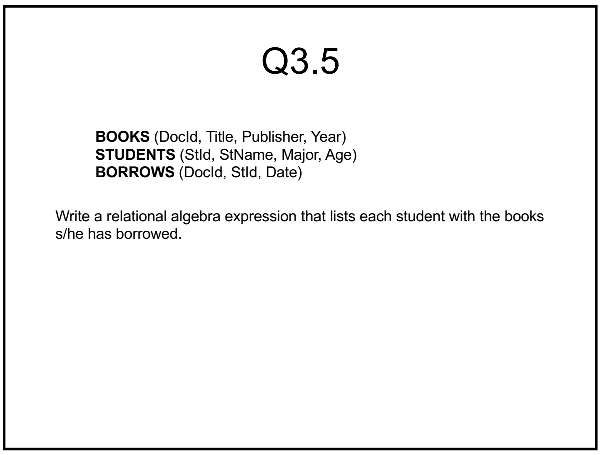
Write a relational algebra expression that lists all information about students whose major is CS.



Q3

BOOKS (DocId, Title, Publisher, Year)
STUDENTS (StId, StName, Major, Age)

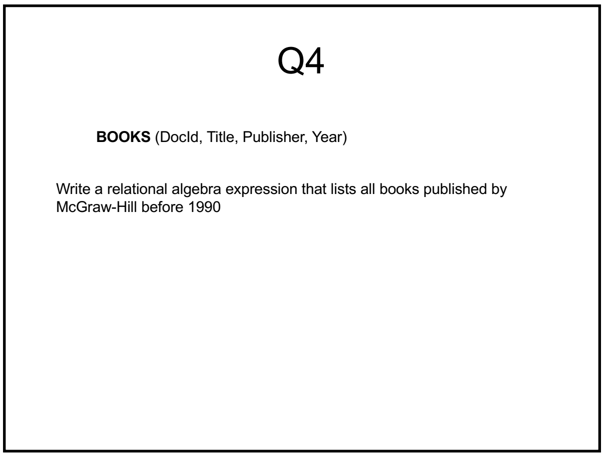
Write a relational algebra expression that lists all students with the books they can borrow.



Q3.5

BOOKS (DocId, Title, Publisher, Year)
STUDENTS (StId, StName, Major, Age)
BORROWS (DocId, StId, Date)

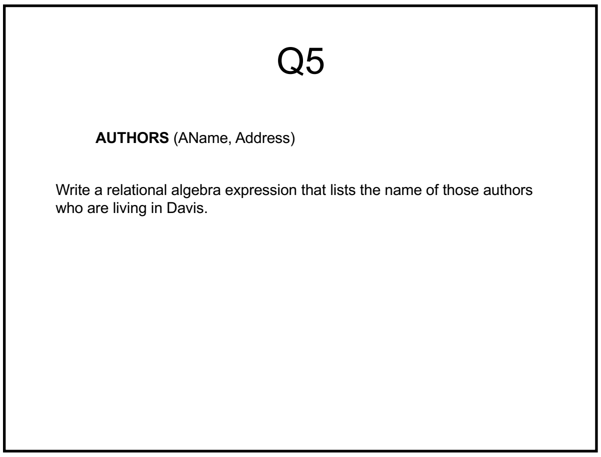
Write a relational algebra expression that lists each student with the books s/he has borrowed.



Q4

BOOKS (DocId, Title, Publisher, Year)

Write a relational algebra expression that lists all books published by McGraw-Hill before 1990



Q5

AUTHORS (AName, Address)

Write a relational algebra expression that lists the name of those authors who are living in Davis.

Q6

AUTHORS (AName, Address)

Write a relational algebra expression that renames “Aname” attribute in the relation AUTHORS to “Name”

Q7

STUDENTS (StId, StName, Major, Age)

Write a relational algebra expression that lists the name of students who are older than 30 and who are not studying CS

Q8

Wards (number, numBeds)
Patients (pid, name, year, gender)
PatientInWard (pid, ward)
Tests (pid, testTime, temperature, heartRate)

Write a relational algebra expression that finds the temperature and heart rate measured in each test carried out on patients born before 1950.

a) $\sigma_{\text{temperature,heartRate}}(\pi_{\text{year<=1950}}(\text{Patients}) \bowtie \text{Tests})$

b) $\pi_{\text{temperature,heartRate}}(\sigma_{\text{year<=1950}}(\text{Patients}) \bowtie \text{Tests})$

c) $\pi_{\text{temperature,heartRate}}(\sigma_{\text{year<1950}}(\text{Patients}) \bowtie \text{Tests})$

d) $\sigma_{\text{temperature,heartRate}}(\pi_{\text{year<1950}}(\text{Patients}) \bowtie \text{Tests})$

Q9

Departments (deptName, location)
Employees (empId, name)
WorksIn (employee, dept, location, percentage)

Write a relational algebra expression that finds the employee name and department name of employees who are assigned to work more than 50% in a department in Stockholm.

a) $\sigma_{\text{empId,deptName}}(\text{Employees} \bowtie_{\text{empId=employee}} (\sigma_{\text{percentage>50 AND location= Stockholm}}(\text{WorksIn})))$

b) $\pi_{\text{empId,deptName}}(\text{Employees} \bowtie (\pi_{\text{percentage>50 AND location= Stockholm}}(\text{WorksIn})))$

c) $\pi_{\text{empId,deptName}}(\text{Employees} \bowtie (\sigma_{\text{percentage>50 AND location= Stockholm}}(\text{WorksIn})))$

d) $\pi_{\text{empId,deptName}}(\text{Employees} \bowtie_{\text{empId=employee}} (\sigma_{\text{percentage>50 AND location= Stockholm}}(\text{WorksIn})))$

Q11

Exams (course, examDate, examTime)
Students (studentId, name)
registeredFor (student, course, examDate)

Write a relational algebra expression that finds the names of students who have registered for the exam in course ’TDA357’ on ’2010-12-18’.

a) $\pi_{\text{name}}(\text{Students} \bowtie (\sigma_{\text{course=‘TDA357’ AND examData=‘2010-12-18’}}(\text{registeredFor})))$

b) $\pi_{\text{name}}(\text{Students} \bowtie_{\text{studentId=student}} (\sigma_{\text{course=‘TDA357’ AND examData=‘2010-12-18’}}(\text{registeredFor})))$

c) $\pi_{\text{name}}(\text{registeredFor} \bowtie_{\text{studentId=student}} (\sigma_{\text{course=‘TDA357’ AND examData=‘2010-12-18’}}(\text{Students})))$

d) $\pi_{\text{name}}(\text{Students} \bowtie_{\text{studentId=student}} (\sigma_{\text{course=‘TDA357’ OR examData=‘2010-12-18’}}(\text{registeredFor})))$

Exams (course, examDate, examTime)
Students (studentId, name)
registeredFor (student, course, examDate)

Q12

Write a relational algebra expression that finds the average number of students who have registered for the exams in each course (for example, if there have been three exams in course ’TDA357’ and 100 students registered for the exam on the first occasion, 150 students registered for the second occasion and 80 students registered for the third occasion, then the average number of students registering for an exam in course ’TDA357’ would be 110).

The result should contain the course code and the average number of students registered for exams in that course, and the results should be sorted by course code.

a) $\tau_{\text{course}}(\gamma_{\text{AVG(nrSt)} \rightarrow \text{avgSt}}(\gamma_{\text{course, examDate, COUNT(student)} \rightarrow \text{nrSt}}(\text{registeredFor})))$

b) $\tau_{\text{course}}(\gamma_{\text{course, AVG(nrSt)} \rightarrow \text{avgSt}}(\gamma_{\text{course, COUNT(student)} \rightarrow \text{nrSt}}(\text{registeredFor})))$

c) $\tau_{\text{course}}(\gamma_{\text{course, AVG(nrSt)} \rightarrow \text{avgSt}}(\gamma_{\text{course, examDate, COUNT(student)} \rightarrow \text{Students}}(\text{registeredFor})))$

d) $\tau_{\text{course}}(\gamma_{\text{course, AVG(nrSt)} \rightarrow \text{avgSt}}(\gamma_{\text{course, examDate, COUNT(student)} \rightarrow \text{nrSt}}(\text{registeredFor})))$

Q13

Employees (empId, name, year, salary, entitlement, branch)
ParentalLeave (employee, startDay, startYear, endDay, endYear)

Write a relational algebra expression that finds the names of employees who had a period of parental leave that started in 2007.

a) $\pi_{\text{name}}\,(\text{Employees} \bowtie_{\text{empId=employee}} (\pi_{\text{startYear=2007}}\,(\text{ParentalLeave})))$

b) $\sigma_{\text{name}}\,(\text{Employees} \bowtie_{\text{empId=employee}} (\sigma_{\text{startYear>=2007}}\,(\text{ParentalLeave})))$

c) $\pi_{\text{name}}\,(\text{Employees} \bowtie_{\text{empId=employee}} (\sigma_{\text{startYear=2007}}\,(\text{ParentalLeave})))$

d) $\pi_{\text{name}}\,(\text{Employees} \bowtie_{\text{empId=employee}} (\sigma_{\text{startYear>=2007}}\,(\text{ParentalLeave})))$

Q14

Employees (empId, name, year, salary, entitlement, branch)
ParentalLeave (employee, startDay, startYear, endDay, endYear)

Write a relational algebra expression that finds the employee(s) in each branch who have the highest salary in their branch. The result should contain the employee’s name, the branch and the salary, and the result should be sorted by branch name.

a) $\tau_{\text{branch}}\,(\pi_{\text{name,branch,salary}}\,(\sigma_{\text{salary=maxSal}}\,(\text{Employees} \bowtie (\gamma_{\text{branch,MAX(salary)}\rightarrow\text{maxSal}}(\text{Employees})))))$

b) $\tau_{\text{branch}}\,(\pi_{\text{name,salary}}\,(\sigma_{\text{salary=maxSal}}\,(\text{Employees} \bowtie (\gamma_{\text{branch,MAX(salary)}\rightarrow\text{maxSal}}(\text{Employees})))))$

c) $\tau_{\text{branch}}\,(\pi_{\text{name,branch,salary}}\,(\sigma\,(\text{Employees} \bowtie (\gamma_{\text{branch,MAX(salary)}\rightarrow\text{maxSal}}(\text{Employees})))))$

d) $\tau_{\text{branch}}\,(\pi_{\text{name,branch,salary}}\,(\sigma_{\text{salary=maxSal}}\,(\gamma_{\text{branch,MAX(salary)}\rightarrow\text{maxSal}}(\text{Employees}))))$

Q15

Doctors (person_number, health_centre)
Appointments (appointment_id, patient, doctor, day, month, year)
doctor → Doctors.person number

Write a relational algebra expression that computes then names of the health centres, sorted in alphabetical order, where the patient with identification number ‘6006064444’ had appointments in year 2000 or more recently.

a) $\tau_{\text{health\_centre}}\,(\pi_{\text{health\_centre}}(\text{Doctors} \bowtie_{\text{person\_number=doctor}} (\sigma_{\text{patient="6006064444" AND year} \geq \text{2000}}(\text{Appointments}))))$

b) $\tau_{\text{health\_centre}}\,(\pi_{\text{health\_centre}}(\text{Appointments} \bowtie_{\text{person\_number=doctor}} (\sigma_{\text{patient="6006064444" AND year} \geq \text{2000}}(\text{Doctors}))))$

c) $\tau_{\text{health\_centre}}\,(\pi_{\text{health\_centre}}(\text{Doctors} \bowtie (\pi_{\text{patient="6006064444" AND year} \geq \text{2000}}(\text{Appointments}))))$

d) $\tau_{\text{health\_centre}}\,(\pi_{\text{health\_centre}}(\text{Doctors} \bowtie_{\text{person\_number=doctor}} (\sigma_{\text{patient="6006064444" AND year} \geq \text{2000}}(\text{Appointments}))))$

Q16

Doctors (person_number, health_centre)
Appointments (appointment_id, patient, doctor, day, month, year)
doctor → Doctors.person number

Write a relational algebra expression that computes, for each health centre, the total number of appointments at that health centre in each month of 2007.

a) $\pi_{\text{health\_centre, month, count(*)} \rightarrow \text{numApps}}\,(\text{Doctors} \bowtie_{\text{person\_number=doctor}} (\sigma_{\text{year=2007}}(\text{Appointments})))$

b) $\sigma_{\text{health\_centre, month, count(*)} \rightarrow \text{numApps}}\,(\text{Doctors} \bowtie_{\text{person\_number=doctor}} (\sigma_{\text{year=2007}}(\text{Appointments})))$

c) $\tau_{\text{health\_centre, month, count(*)} \rightarrow \text{numApps}}\,(\text{Doctors} \bowtie_{\text{person\_number=doctor}} (\sigma_{\text{year=2007}}(\text{Appointments})))$

d) $\gamma_{\text{health\_centre, month, count(*)} \rightarrow \text{numApps}}\,(\text{Doctors} \bowtie_{\text{person\_number=doctor}} (\sigma_{\text{year=2007}}(\text{Appointments})))$

Q17

Employees

name	branch	salary
Andersson	3	20000
Johnsson	3	25000
Larsson	3	32000
Persson	2	28000
Svensson	2	35000

Branches

number	city
1	Stockholm
2	Paris
3	London
4	Berlin
5	Rome

Write two different relational algebra expressions that evaluate to:

name	city
Larsson	London
Svensson	Paris

1) one of these relational algebra expressions **must use** the Cartesian product operator.

2) one of these relational algebra expressions **must not** use the Cartesian product operator.

Q18

Employees

name	branch	salary
Andersson	3	20000
Johnsson	3	25000
Larsson	3	32000
Persson	2	28000
Svensson	2	35000

Branches

number	city
1	Stockholm
2	Paris
3	London
4	Berlin
5	Rome

Write a relational algebra expression that computes the average salary at each branch, and sorts the results in increasing order of the average salary.

Q19

Offices (city, supplement)
Departments (city, dname, departmentHead)
Employees (empId, name, salary, dept, city)

Write a relational algebra expression that finds the employee identifier, name and total monthly salary of all employees (recall that the total monthly salary for an employee can be calculated by adding the city supplement to the employee's basic monthly salary). The results should be sorted by employee name.

a) $\tau_{\text{name}}\ (\pi_{\text{empId,name,supplement}}\ (\text{Employees} \bowtie \text{Offices}))$

b) $\tau_{\text{name}}\ (\pi_{\text{empId,name,salary}}\ (\text{Employees} \bowtie \text{Offices}))$

c) $\tau_{\text{name}}\ (\pi_{\text{empId,name,salary+supplement}}\ (\text{Employees} \bowtie \text{Offices}))$

d) $\tau_{\text{name}}\ (\pi_{\text{empId, salary+supplement}}\ (\text{Employees} \bowtie \text{Offices}))$

Q20

Offices (city, supplement)
Departments (city, dname, departmentHead)
Employees (empId, name, salary, dept, city)

Write a relational algebra expression that finds the names of cities where there is a sales department and, for each of these departments, the average basic salary of the employees in that department.

a) $\gamma_{\text{city,AVG(salary)}\rightarrow\text{avgSalary}}\ (\sigma_{\text{dept=''sales''}}\ (\text{Employees}))$

b) $\gamma_{\text{AVG(salary)}\rightarrow\text{avgSalary}}\ (\sigma_{\text{dept=''sales''}}\ (\text{Employees}))$

c) $\gamma_{\text{city}}\ (\sigma_{\text{dept=''sales''}}\ (\text{Employees}))$

d) $\gamma_{\text{city,AVG(salary)}\rightarrow\text{avgSalary}}\ (\sigma(\text{Employees}))$

Q21

Translate the following relational algebra expression to an SQL query:

$\pi_{\text{First.depatureTime,Second.arrivalTime}}\ ((\sigma_{\text{First}}\ (\text{Flights})) \bowtie_{\text{First.destinationAirport = Second.DepartureAirport}}\ (\sigma_{\text{Second}}(\text{Flights})))$

a)
```
SELECT first.departureTime, second.arrivalTime
FROM flights AS first JOIN flights AS second
ON first.destinationAirport = second.departureAirport;
```

b)
```
SELECT first.departureTime, second.arrivalTime
FROM flights AS first, flights AS second
WHERE first.destinationAirport = second.departureAirport;
```

c)
```
SELECT departureTime, arrivalTime
FROM flights JOIN flights
ON destinationAirport = departureAirport;
```

Q22

Words (string, lemma, class, description)

Write a relational algebra query that returns those strings whose class is ambiguous, i.e., can have two or more different values. An example is *läcker*, which is both the present tense of the verb *läcka* ("leak") and the singular real form of the adjective *läcker* ("delicious").

a) $\pi_{\text{A.string}}\ \sigma_{\text{A.string=B.string AND A.class<B.class}}\ (\sigma_{\text{A}}\text{Words} \times \sigma_{\text{B}}\text{Words})$

b) $\pi_{\text{A.string}}\ \sigma_{\text{A.string=B.string AND A.class<B.class}}\ (\rho_{\text{A}}\text{Words} \times \rho_{\text{B}}\text{Words})$

c) $\pi_{\text{string}}\ \sigma_{\text{string=string AND class<class}}\ (\text{Words} \times \text{Words})$

d) $\sigma_{\text{A.string}}\ \pi_{\text{A.string=B.string AND A.class<B.class}}\ (\rho_{\text{A}}\text{Words} \times \rho_{\text{B}}\text{Words})$

Q23

BOOKS (DocId, Title, Publisher, Year)

Write a relational algebra expression that finds the title of the oldest book.