CSE 311 Foundations of Computing I

Lecture 22

Relations, Database

Autumn 2012

Announcements

- Reading assignments
  - 7th Edition, Sections 9.3 and 13.3
  - 6th Edition, Section 8.3 and 12.3
  - 5th Edition, Section 7.3 and 11.3

Lecture highlights

Let A and B be sets,
A binary relation from A to B is a subset of $A \times B$

$S \circ R = \{(a, c) \mid \exists\ b$ such that $(a,b) \in R$ and $(b,c) \in S\}$

$R^1 = R$; $R^{n+1} = R^n \circ R$

Matrix and graph representation of relations

Paths in relations

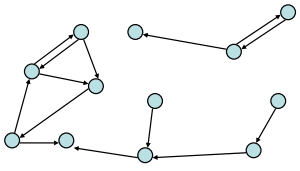
Let R be a relation on a set A. There is a path of length n from a to b if and only if $(a,b) \in R^n$

Connectivity relation

Let R be a relation on a set A. The connectivity relation R* consists of the pairs (a,b) such that there is a path from a to b in R.

$$R^* = \bigcup_{k=0}^{\infty} R^k$$

Transitive-Reflexive Closure



Add the minimum possible number of edges to make the relation transitive and reflexive

The transitive-reflexive closure of a relation R is the connectivity relation R*

How is  related to  ?

Anderson

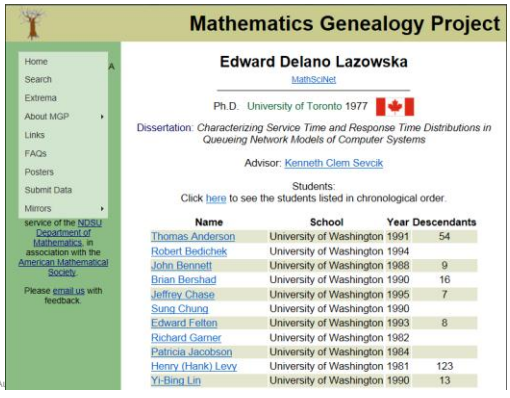
Copernicus

How is  related to  ?

Suciu

Copernicus

http://genealogy.math.ndsu.nodak.edu/




Anderson


Mayr


Bauer


Caratheodory


Minkowski


Klein


Lipschitz


Dirichlet


Fourier


Lagrange


Euler


Johann Bernoulli


Jacob Bernoulli


Leibniz


Weigel


Rheticus


Copernicus


Suciu


Tannen


Meyer


Fischer


Rogers


Church

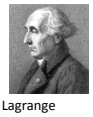

Veblen


Moore


H.A.Newton


Chasles


Poisson


Lagrange


Euler


Johann Bernoulli


Jacob Bernoulli

Nicolaus Copernicus
Georg Rheticus
Moritz Steinmetz
Christoph Meurer
Philipp Muller
Erhard Weigel
Gottfried Leibniz
Noclas Malebranache
Jacob Bernoulli
Johann Bernoulli
Leonhard Euler
Joseph Lagrange
Jean-Baptiste Fourier
Gustav Dirichlet
Rudolf Lipschitz
Felix Klein
C. L. Ferdinand Lindemann
Herman Minkowski
Constantin Caratheodory
Georg Aumann
Friedrich Bauer
Manfred Paul
Ernst Mayr
Richard Anderson

Nicolaus Copernicus
Georg Rheticus
Moritz Steinmetz
Christoph Meurer
Philipp Muller
Erhard Weigel
Gottfried Leibniz
Noclas Malebranache
Jacob Bernoulli
Johann Bernoulli
Leonhard Euler
Joseph Lagrange
Michel Chasles
H. A. (Hubert Anson) Newton
E. H. (Eliakim Hastings) Moore
Oswald Veblen
Alonzo Church
Hartley Rogers, Jr.
Patrick Carl Fischer
Albert Ronald da Silva Meyer
Val Tannen
Dan Suciu

n-ary relations

Let $A_1, A_2, \ldots, A_n$ be sets. An n-ary relation on these sets is a subset of $A_1 \times A_2 \times \ldots \times A_n$.

Relational databases

STUDENT

Student_Name	ID_Number	Office	GPA
Knuth	328012098	022	4.00
Von Neuman	481080220	555	3.78
Russell	238082388	022	3.85
Einstein	238001920	022	2.11
Newton	1727017	333	3.61
Karp	348882811	022	3.98
Bernoulli	2921938	022	3.21

Relational databases

STUDENT

Student_Name	ID_Number	Office	GPA	Course
Knuth	328012098	022	4.00	CSE311
Knuth	328012098	022	4.00	CSE351
Von Neuman	481080220	555	3.78	CSE311
Russell	238082388	022	3.85	CSE312
Russell	238082388	022	3.85	CSE344
Russell	238082388	022	3.85	CSE351
Newton	1727017	333	3.61	CSE312
Karp	348882811	022	3.98	CSE311
Karp	348882811	022	3.98	CSE312
Karp	348882811	022	3.98	CSE344
Karp	348882811	022	3.98	CSE351
Bernoulli	2921938	022	3.21	CSE351

What wrong?

Relational databases

STUDENT

Student_Name	ID_Number	Office	GPA
Knuth	328012098	022	4.00
Von Neuman	481080220	555	3.78
Russell	238082388	022	3.85
Einstein	238001920	022	2.11
Newton	1727017	333	3.61
Karp	348882811	022	3.98
Bernoulli	2921938	022	3.21

Better

TAKES

ID_Number	Course
328012098	CSE311
328012098	CSE351
481080220	CSE311
238082388	CSE312
238082388	CSE344
238082388	CSE351
1727017	CSE312
348882811	CSE311
348882811	CSE312
348882811	CSE344
348882811	CSE351
2921938	CSE351

Database Operations: Projection

Find all offices: $\Pi_{Office}(STUDENT)$

Office
022
555
333

Find offices and GPAs: $\Pi_{Office,GPA}(STUDENT)$

Office	GPA
022	4.00
555	3.78
022	3.85
022	2.11
333	3.61
022	3.98
022	3.21

Database Operations: Selection

Find students with GPA > 3.9 : $\sigma_{GPA>3.9}(STUDENT)$

Student_Name	ID_Number	Office	GPA
Knuth	328012098	022	4.00
Karp	348882811	022	3.98

Retrieve the name and GPA for students with GPA > 3.9 : $\Pi_{Student\_Name,GPA}(\sigma_{GPA>3.9}(STUDENT))$

Student_Name	GPA
Knuth	4.00
Karp	3.98

Database Operations: Natural Join

Student ⋈ Takes

Student_Name	ID_Number	Office	GPA	Course
Knuth	328012098	022	4.00	CSE311
Knuth	328012098	022	4.00	CSE351
Von Neuman	481080220	555	3.78	CSE311
Russell	238082388	022	3.85	CSE312
Russell	238082388	022	3.85	CSE344
Russell	238082388	022	3.85	CSE351
Newton	1727017	333	3.61	CSE312
Karp	348882811	022	3.98	CSE311
Karp	348882811	022	3.98	CSE312
Karp	348882811	022	3.98	CSE344
Karp	348882811	022	3.98	CSE351
Bernoulli	2921938	022	3.21	CSE351

Disambiguating Attribute Names

Names of students who share an office with Einstein:

$\Pi_{STUDENT1.Student\_Name}(STUDENT \bowtie_{STUDENT1.Office = STUDENT2.Office} (\sigma_{Student\_Name='Einstein'}(STUDENT)))$

STUDENT

Student_Name	ID_Number	Office	GPA
Knuth	328012098	022	4.00
Von Neuman	481080220	555	3.78
Russell	238082388	022	3.85
Einstein	238001920	022	2.11
Newton	1727017	333	3.61
Karp	348882811	022	3.98
Bernoulli	2921938	022	3.21

STUDENT

Student_Name	ID_Number	Office	GPA
Knuth	328012098	022	4.00
Von Neuman	481080220	555	3.78
Russell	238082388	022	3.85
Einstein	238001920	022	2.11
Newton	1727017	333	3.61
Karp	348882811	022	3.98
Bernoulli	2921938	022	3.21

Answer:

Student_Name
Knuth
Russell
Einstein
Karp
Bernoulli

Database Operations: Relational Algebra

Find the names and GPAs of all students who take 311

Find the names of all students who take at least two courses

Database Operations: SQL

Find the names and GPAs of all students who take 311

```
SELECT DISTINCT Student_name, GPA
FROM STUDENT Natural Join TAKES
WHERE Course = 'CSE311'
```

Find the names of all students who take at least two courses

```
SELECT DISTINCT Student_name, GPA
FROM STUDENT s, TAKES t1, TAKES t2
WHERE s.ID_number = t1.ID_number
    and s.ID_number = t2.ID_number
    and t1.course != t2.course
```

Relational Databases: Keys

- An attribute is a **key** if all its values in the database are always distinct

Student_Name	ID_Number	Office	GPA
Knuth	328012098	022	4.00
Von Neuman	481080220	555	3.78
Russell	238082388	022	3.85
Einstein	238001920	022	2.11
Newton	1727017	333	3.61
Karp	348882811	022	3.98
Bernoulli	2921938	022	3.21

Which attribute is the key?
Why is Student_Name not a key?

Relational Databases: Relationships

STUDENT

Student_Name	ID_Number	Office
Knuth	328012098	022
Von Neuman	481080220	555
Russell	238082388	022
Einstein	238001920	022
Newton	1727017	333
Karp	348882811	022
Bernoulli	2921938	022

PROJECT

PRJ_ID	Project_Name	Due_date
P331	"Flying cyphers"	11/2012
P004	"Virtual induction"	12/2012
P901	"Binary bots"	12/2012

WORKS_ON is a relationship between students and project

ID_Number	PRJ_ID
2921938	P004
2921938	P901
1727017	P901

Who works on what?

Types of Relationships in Relational Databases

- one-one:

- many-one

- many-many

What type is WORKS_ON?

ID_Number	PRJ_ID
2921938	P004
2921938	P901
1727017	P901