CSC343 Assignment1

Tianchang Mei 1004034109
Yize Zhao 1004343846

January 28, 2020

relations

- Collection(CID, date, SID)
  Tuples here represent entire collections from a field trip, where *CID* is the collection ID, *date* is the starting date of the field trip, and *SID* is the staff ID of the collector.
- Collected(CID, AN)
  A tuple here represents the fact that collection *CID* includes artifact number *AN*. A single collection usually contains multiple artifacts, and a single artifact may be aggregated from more than one collection.
- Artifact(AN, species, type, location, SID)
  Tuples here represent single artifact collected in the field. *AN* is the artifact number, *species* is the scientific species name, *type* is one of tissue, image, model, or live, *location* is where it was collected, and *SID* is the staff number of the technician who maintains this artifact.
- Published(AN, journal, date)
  A tuple here represents the fact that artifact *AN* was mentioned in scholarly publication *journal* with publication date *date*.
- Staff(SID, name, email, rank, date)
  These tuples represent a member of the institute's scientific staff. *SID* is the staff ID, *name* is their full name, *email* is their professional email, *rank* is one of: technician, student, pre-tenure, or tenured, and *date* is the date when they attained that rank.
- COL(family)
  A singleton tuple here means that *family* is a scientific zoological family name that appears in the Catalogue of Life.
- Genus(genus, family)
  A tuple here means that *genus* is in family *family*.
- Species(species, genus)
  A tuple here means that *species* is in genus *genus*.

our constraints

For each of the following constraints give a one sentence explanation of what the constraint implies, and why it is required.

- $\pi_{species}(Artifact) - \pi_{species}(Species) = \emptyset$.

  Explanation: There is at least one artifact for every species.

- $\pi_{rank}(Staff) \subseteq$ {'technician', 'student', 'pre-tenure', 'tenure'}.

  Explanation: The staff in this database are classified into four ranks: technician, student, pre-tenure and tenure.

- $\pi_{family}(Genus) - \pi_{family}(COL) = \emptyset$.

  Explanation: For each family in COL database, there is at least one genus in that family.

- $\pi_{genus}(Species) \subseteq \pi_{genus}(Genus)$.

  Explanation: Every species belongs to at least one genus.

- $\pi_{CID}(Collected) = \pi_{CID}(Collection)$.

  Explanation: All field trip collections must be documented as collected artifacts; and each collected artifact have information indicating in which field trip it was discovered.

- $\pi_{AN}(Artifact) = \pi_{AN}(Collected)$.

  Explanation: After arriving at the institute, every artifact collected from the field will join the institute's collection of all artifacts.

- $\pi_{SID}(Collection) \subseteq \pi_{SID}(Staff)$.

  Explanation: Every field trip collection is collected by a staff with staff number(SID).

- $\pi_{SID}(Artifact) \subseteq \pi_{SID}(Staff)$.

  Explanation: Every artifact is collected by a staff with staff number(SID).

- $\pi_{type}(Artifact) \subseteq \{'tissue', 'image', 'model', 'live'\}$

  Explanation: Every artifact is classified into one of four types: tissue, image, model or live.

- $\pi_{AN}(Published) \subseteq \pi_{AN}(Artifact)$

  Explanation: Only collected artifacts can be published.

queries

Write relational algebra expressions for each of the queries below. You must use notations from this course and operators:

$$\pi, \sigma, \rho, \bowtie, \bowtie_{condition}, \times, \cap, \cup, -, =$$

You may also use constants:

today (for current date) $\quad$ $\emptyset$ (for the empty set)

In your queries pay attention to the following:

- All relations are sets, and you may only use relational algebra operators covered in Chapter 2 of the course text.
- Do not make assumptions that are not enforced by our constraints above, so your queries should work correctly for any database that obeys our schema and constraints.

- Other than constants such as 23 or "lupus", a select operation only examines values contained in a tuple, not aggregated over an entire column.
- Your selection conditions can use arithmetic operators, such as $+, \leq, \neq, \geq, >, <$ and friends. You can use logical operators such as $\vee, \wedge$, and $\neg$, and treat dates and numeric attributes as numbers that you can perform arithmetic on.
- Use good variable names and provide lots of comments to explain your intentions.
- Return multiple tuples if that is appropriate for your query.

1. Rationale: Performance reviews include seeing how current the work is of staff who have held their current rank for a long time.

   Query: Find the most recent collection date of any artifact collected by a staff member who has held their current rank the longest. Keep ties.

   Answer:
   The staff that has certainly not held their current rank the longest:
   $NotLongest(SID) := \pi_{S1.SID}[\sigma_{S1.date>S2.date}[(\rho_{S1}Staff) \times (\rho_{S2}Staff)]]$
   The staff that have held the rank for the longest:
   $Longest(SID) := (\pi_{SID}Staff) - NotLongest$
   The collections that are not collected most recently by any staff:
   $NotRecent(CID) := \pi_{C1.CID}[\sigma_{C1.date<C2.date}[(\rho_{C1}Collection) \times (\rho_{C2}Collection)]]$
   The collections that are collected most recently by any staff:
   $MostRecent(CID) := (\pi_{CID}Collection) - NotRecent$
   desired tuples:
   $Answer(date) := \pi_{date}[Longest \bowtie (Collection \bowtie MostRecent)]$

2. Rationale: Staff who maintain every artifact in some collection should be considered favourably in performance reviews.

   Query: Find all staff who maintain all artifacts in at least one collection.

   Answer:
   The Collection number, Artifact number and staff number for all artifacts:
   $CidAnSid(CID, AN, SID) := \pi_{CID,AN,SID}(Collected \bowtie Artifact)$
   Collections maintained by multiple workers:
   $MultiStaffMaintain(CID) := \pi_{CID}[\sigma_{(t1.CID=t2.CID)\wedge(t1.SID\neq t1.SID)}(rho_{t1}CidAnSID \times \rho_{t2}CidAnSid)]$
   Collections maintained by only one worker:
   $SinglyHandledCollection(CID) := \pi_{CID}Collection - \pi_{CID}MultiStaffMaintain$
   Workers that maintain the above collections:
   $hiPerformStaff(SID) :=$
   $\pi_{SID}[\sigma_{CidAnSid.CID=SinglyHandledCollection.CID}(CidAnSID \times SinglyHandledCollection)]$

3. Rationale: An artifact collected and maintained by the same staff may have some special requirements that should be investigated.

   Query: Find all artifacts that were collected by the same staff who maintains them.

   Answer:
   A new table for every artifact with essential attributes (AN, SID) only and 1 added attribute CID
   $Artifact1(CID, AN, SID) :=$
   $\pi_{Collected.CID,Artifact.AN,Artifact.SID}(Collected \bowtie_{Collected.AN=Artifact.AN} Artifact)$
   Extended new table with 1 more added attribute: CSID (CollectionSID)
   $Artifact2(CID, CSID, AN, SID) :=$

$\pi_{Collection.CID,Collection.SID,Artifact1.AN,Artifact1.SID}(Collection \bowtie_{Collection.CID=Artifact1.CID} Artifact1)$
desired tuples:
$Answer(AN) := \pi_{AN}(\sigma_{CSID=SID}Artifact2)$

4. Rationale: Identify multi-talented field workers.

   Query: Find all staff who have collected at least 3 artifacts from every species in some family.

   Answer: cannot be expressed

5. Rationale: Which publications might have some specialized niche focus?

   Query: Find all publications that have used exactly 2 of our artifacts.

   Answer:
   The journal that has used at least 2 of our artifacts:
   $AtLeastTwice(journal) :=$
   $\pi_{P1.journal}[\sigma_{P1.AN \neq P2.AN \ \wedge \ P1.journal=P2.journal}(\rho_{P1}Published \times \rho_{P2}Published)]$
   The journal that has used at least 3 of our artifacts:
   $AtLeastThrice(journal) :=$
   $\pi_{P1.journal}[\sigma_{P1.AN \neq P2.AN \neq P3.AN \ \wedge \ P1.journal=P2.journal=P3.journal}(\rho_{P1}Published \times \rho_{P2}Published \times \rho_{P3}Published)]$
   The journal that have published exactly 2 of our artifacts:
   $ExactlyTwice(journal) := AtLeastTwice - AtLeastThrice$

6. Rationale: Identify motherlode locations.

   Query: Find all locations where at least one artifact from every family has been collected.

   Answer:
   All pairs of family and location
   $AllPairs(family, location) := COL \times \pi_{location}Artifact$
   Location and family of all found artifacts
   $LocFoundFamily(family, location) := [\pi_{family,location} \ (Genus \bowtie Species)] \bowtie Artifact$
   Location where entire families cannot be found
   $LocNotAllFamily(location) := \pi_{location}(AllPairs - LocFoundFamily)$
   Location where at least one artifact from every family have been collected
   $OneLocAllFamily(location) := \pi_{location}Artifact - LocNotAllFamily$

7. Rationale: Exclusively tissue sample collectors may need extra support for special reagents and shipping costs.

   Query: Find all staff who have collected only tissue samples.

   Answer:
   A table of all collectors and the types of artifacts they have collected:
   $AllCollector(CSID, type) := \pi_{Collection.SID,Artifact.type}[Collection \bowtie (Collected \bowtie Artifact)]$
   The staff who have collected more than 1 types of artifacts:
   $MultiCollector(CSID) := \pi_{A1.CSID}[\sigma_{A1.CSID=A2.CSID \ \wedge \ A1.type \neq A2.type}(\rho_{A1}AllCollector \times \rho_{A2}AllCollector)]$
   Collectors that have only collected one type of artifact:
   $SingleCollector(CSID) := \pi_{CSID}AllCollector - MultiCollector$
   Collectors who have only collected tissue samples:
   $Answer(CSID) := \pi_{CSID}[\sigma_{type='tissue'}(SingleCollector \bowtie AllCollector)]$

8. Rationale: Collection staff who should be encouraged to diversify their network.

   Query: Find all staff pairs who have worked only with each other on collections.
   Pairs of all Collection number and stuff number
   $CidSid(CID, SID) := \pi_{CID,SID}Collection$
   Pairs of workers who have worked together:

$HaveTogether(SID1, SID2) :=$
$\pi_{C1.SID,C2.SID}\ [\sigma_{C1.CID=C2.CID\ \wedge\ C1.SID\neq C2.SID}(\rho_{C1}CidSid \times \rho_{C2}CidSid)]$
Pairs of workers in which at lease one in the pair have worked with others
$NotAllTogether(SID1, SID2) :=$
$\pi_{C3.SID1,C3.SID2}(\sigma_{C3.SID1=C4.SID1\ \wedge\ C3.SID2\neq C4.SID2}(\rho_{C3}CrossTable \times \rho_{C4}CrossTable))$
Pairs of workers who have only worked with each other
$AlwaysTogether(SID1, SID2) := HaveTogether - NotAllTogether$

9. Rationale: Track the influence of a given staff member.

   Query: Staff member SID_1 is influenced by staff member SID_2 if (a) they have ever worked together on a collection or (b) if SID_1 has ever worked with a staff member who is influenced by SID_2. Find SIDs of staff members influenced by SID 42.

   Answer:
   Using the intermediate table $HaveTogether$ from Question 8 to obtain SID42's coworkers:
   $Coworkers(SID) := \pi_{SID2}(\sigma_{SID1=42}HaveTogether)$
   The people who have worked with SID42's coworkers:
   $CoCoworkers(SID) := \pi_{SID2}(HaveTogether \bowtie_{HaveTogether.SID1=Coworkers.SID} Coworkers)$
   The people influenced by SID42:
   $Answer(SID) := Coworkers \cup CoCoworkers$

your constraints

For each of these constraints you should derive a relational algebra expression of the form $R = \emptyset$, where R may be derived in several steps, by assigning intermediate results to a variable. If the constraint cannot be expressed in the relational algebra you have been taught, write "cannot be expressed."

1. No species is also a genus.
   Answer: $\pi_{genus}(Genus) \cap \pi_{species}(Species) = \emptyset$

2. No genus belongs to more than one family.
   Answer: $\sigma_{G1.genus==G2.genus\ and\ G1.family!=G2.family}(\rho_{G1}Genus \times \rho_{G2}Genus) = \emptyset$

3. All publications must be published after all artifacts they use have been collected.
   Answer:
   $DateAN(AN, date) := \pi_{AN,date}(Collection \bowtie Collected)$
   $\sigma_{DateAn.date>Published.date}(DateAn \bowtie Published) = \emptyset$

4. Students may not catalogue live artifacts.
   Answer: $(\pi_{SID}\sigma_{rank=='student'}Staff)\ \cap\ (\pi_{SID}\sigma_{type=='live'}(Artifact)) = \emptyset$