ZhenDi Pan 1003241823

Shuo Zhuang 1002264195

CSC343 Assignment 1
Relational Algebra

2020-02-07

1 Our constraints

1. $\pi_{species}(Artifact) - \pi_{species}(Species) = \emptyset$.
   All scientific species names of Artifact are contained in species names of table Species. It ensures all species names are included in Species.

2. $\pi_{rank}(Staff) \subseteq$ {'technician', 'student', 'pre-tenure', 'tenure'}.
   The rank of every staff must either be technician, student, pre-tenure or tenure. It declares the domain of the rank and ensures there is no other kind of staff.

3. $\pi_{family}(Genus) - \pi_{family}(COL) = \emptyset$.
   Every genus belongs to family that appears in the Catalogue of life. This ensures that every genus is a scientific zoological family that appears in the catalogue of life.

4. $\pi_{genus}(Species) \subseteq \pi_{genus}(Genus)$.
   Every genus in species are in genus Every species must belong to a family. This is required since we want every possible genus of species in Species are covered in Genus, which also implies it belongs to a family.

5. $\pi_{CID}(Collected) = \pi_{CID}(Collection)$.
   All collections contain artifacts. This ensures there is no meaningless collections, that is a collection with no artifacts at all.

6. $\pi_{AN}(Artifact) = \pi_{AN}(Collected)$.
   All Artifacts are collected. This ensures we don't have any artifact that isn't collected from a collection trip.

7. $\pi_{SID}(Collection) \subseteq \pi_{SID}(Staff)$.
   All collectors are a member of the institute's scientific staff. This is required since there shouldn't be any collector who is not a scientific staff, i.e a member of the institute.

8. $\pi_{SID}(Artifact) \subseteq \pi_{SID}(Staff)$.
   Every artifact must be maintained by a member of the institute's scientific staff. This ensures we don't have any artifact that is not maintained.

9. $\pi_{type}(Artifact) \subseteq \{'tissue', 'image', 'model', 'live'\}$
   The Artifact is one of tissue, image , model or live. This is required because we don't want a type of artifact that is different from these four types.

10. $\pi_{AN}(Published) \subseteq \pi_{AN}(Artifact)$
    Every artifact that was mentioned in scholarly publication must have already been collected from the field. This is required since a person cannot possibly publish something that hasn't been collected yet.

2 Queries

1. **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.

   allStaff is a table with only SID that contains all staffs' SID. YoungStaff contains all SIDs of staffs who have a staff elder than them

   $allStaff := \pi_{(SID)}(Staff)$
   $youngStaff := \pi_{(s1.SID)}[\sigma_{s1.date>s2.date}(\rho_{s1}(Staff) \times \rho_{s2}(\text{Staff}))]$
   $oldStaff := allStaff$- $youngStaff$

   We take a new table "newCollection"with two columns, the SID of the oldest staffs(if there is tie), and their dates of collection, there maybe multiple dates corresponding to one SID.

$newCollection := \pi_{(SID,date)}(oldStaff \bowtie Collection)$

Self join the newCollection to take all SID and their corresponding dates which at least one dates that more recent than that date. Lastly, we then subtract the daysEarly to get the most recent date with each SID. final is the desired query.

$daysEarly := \pi_{(n1.SID,n1.date)}[\sigma_{n1.SID=n2.SID \wedge n1.date<n2.date}(\rho_{n1}(newCollection) \times \rho_{n2}(newCollection))]$
$answer := newCollection - daysEarly$
$answer$

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

First we find out all CID and SID pairs to work with, then we do self join to take all CIDs with different SIDs

$staffCollect := \pi_{(CID,SID)}(Artfact \bowtie Collected)$
$Different := \pi_{(s1.SID,s1.CID)}[\sigma_{(s1.CID=S2.CID \wedge s1.SID \neq s2.SID)}(\rho_{s1}(staffCollect) \times \rho_{s2}(staffCollect))]$

Take Difference from staffCollect, that is the table which contains each collection maintained by the same staff, and take only SID column.

$answer := \pi_{(SID)}(staffCollect - Different)$
$answer$

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

First take only SID and AN colunms in Artifact, then natural join Collected with Art. We only want the part AN to be equal. Then we take only SID,AN,CID

$Art := \pi_{(AN,SID)}(Artifact)$
$artPart := \pi_{(SID,AN,CID)}(Art \bowtie Collected)$

Then we take the cartesian product artPart with Collection to get a new relation on condition that artPart.CID=Collection.CID and artPart.SID= Collection.SID, that will give us the final table we want.

$finalRel := \sigma_{((artPart.CID=Collection.CID) \wedge (artPart.SID=Collection.SID))}(artPart \times Collection)$
https://www.overleaf.com/project/5e3ce688017c950001949b01 Lastly we take the AN column from the final table, it gives all artifacts that were collected by the same staff.

$answer := \pi_{(AN)} finalRel$
$answer$

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

Self join the COllectedArtifact three times and pick out the rows with 3 different artifact from a speices.

$CollectedArtifact = \pi_{(CID,AN,SPECIES)} COLLECTED \bowtie ARTIFACT$

$ThreeCollectedArtifact = \sigma_{(C1.CID=C2.CID \wedge C2.CID=C3.CID \wedge C1.CID=C3.CID) \wedge (C1.species=C2.species \wedge C2.species=(C3.species) \wedge C1.AN \neq C2.AN \neq C3.AN)}(\rho_{C1} CollectedArtifact \times \rho_{C2} CollectedArtifact \times \rho_{C3} CollectedArtifact)$

Three natural joins to make the StaffSpeciesFamily relation:

$CIDwithSpecies = \pi_{(CID,species)} ThreeCollectedArtifact$
$StaffwithSpecies = \pi_{(SID,SPEICES)} Collection \bowtie CIDwithSpecies$
$SpecieswithFamily = \pi_{speices,family} Genus \bowtie Species$
$StaffSpeicesFamily = StaffwithSpecies \bowtie SpecieswithFamily$

Find all possible combinations to get the ones that we don't want, eliminate them from StaffSpeicesFamily, and we have our answer.

$staff = \pi_{(SID)StaffwithSpecies}$
$combinations = staff \times SpecieswithFamily$
$wrong = combinations - StaffSpeicesFamily$
$answer := \pi_{(SID)}(\pi_{(SID,family)} StaffSpeicesFamily - \pi_{(SID,family)} wrong)$
$answer$

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

   First we find all publications that collects at least 3 of our artifacts.

   $thrice := \pi_{(p1.journal)}[\sigma_{p1.journal=p2.journal \wedge p2.journal=p3.journal \wedge p1.AN<p2.AN \wedge p2.AN<p3.AN}(\rho_{p1}(Published) \times \rho_{p2}(Published) \times \rho_{p3}(Published))]$

   Find all publications that used exactly 2 of our artifacts.

   twice :=$\pi_{(p1.journal)}[\sigma_{p1.journal=p2.journal \wedge p1.AN \neq p2.AN}(\rho_{p1}(Published) \times \rho_{p2}(Published))]$

   Subtract thrice from twice to get all publications that have used exactly 2 of our artifacts.

   $answer := twice - thrice$
   $answer$

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

   First we take the speciesFamily by natural join Genus and Species

   $speciesFamily := \pi_{(species,family)}(Genus \bowtie Species)$

   We combine the Artifact table with speciesFamily. Only take columns (AN, location, family) and work on that relation.

   $workTable := \pi_{(location,family)}(\sigma_{Artifact.species=speciesFamily.species}(Artifact \times speciesFamily))$
   $allLocation := \pi_{(location)} workTable$
   $allFamily := \pi_{(family)} workTable$

   We combine allArt and allFamily to get every possible combinations, then subtract the workTable.

   $couldCollect := allLocation \times allFamily$
   $didnot := couldCollect - workTable$
   $answer := (\pi_{(location)} workTable) - (\pi_{location} didnot)$
   $answer$

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

   $answer := (\pi_{(SID)} Artifact) - (\pi_{SID}(\sigma_{type \neq 'tissue'} Artifact))$

8. **Query:** Find all staff pairs who have worked only with each other on collections.

   We first natural join Collected with Artifact to select all collections and their respective maintainer..
   $staffCollected := \pi_{(CID,SID)}(Collected \bowtie Artifact)$

   Self-join the table three times to work with.

   $WorkTable := (\rho_{s1}(staffCollected)) \times (\rho_{s2}(staffCollected)) \times (\rho_{s3}(staffCollected))$

   Find all collection that are maintained by 3 different staff, record the SIDs

   $ThreeCollectionTable := (\sigma_{s1.SID \neq s2.SID \wedge s2.SID \neq s3.SID \wedge s1.SID \neq s3.SID \wedge s1.CID=s2.CID \wedge s2.CID=s3.CID)} WorkTable)$
   $ThreeCollection(CID) := \pi_{(s1.CID)}(ThreeCollectionTable)$
   Find all collection that are maintained by 2 different staff, record the SIDs.

   $TwoCollection(CID) := \pi_{(t1.CID)}[\sigma_{(t1.SID \neq t2.SID \wedge t1.CID=t2.CID)}(\rho_{t1} staffCollected) \times (\rho_{t2} staffCollected)]$
   $TwoPerson(CID, PairOne, PairTwo) := \pi_{t1.CID,t1.SID,t2.SID}[\sigma_{(t1.SID \neq t2.SID \wedge t1.CID=t2.CID)}(\rho_{t1} staffCollected) \times (\rho_{t2} staffCollected)]$

   ExactlyTwo gives us all collections maintained by only two person, natural join it with TwoPerson, we get all the SID of the two maintainers

   $ExactlyTwo := TwoCollection - ThreeCollection$
   $Pair := ExactlyTwo \bowtie TwoPerson$

Now Pair is simply a relation which contains the collections that are maintained by two person and also the pairs of those collections: $Pair(\underline{CID}, PairOne, PairTwo)$

Lastly, we have to also eliminate the case where the collector is not one of those two person, so lastly we natural join it with collection.

$LastTable := \sigma_{(SID \neq PairOne \wedge SID \neq PairTwo)}(Pair \bowtie Collection)$
$answer := \pi_{(PairOne, PairTwo)}(LastTable)$

9. **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.

   This query cannot be expressed as it involves recursive steps, none of the operators we learned in class can do this.

3 Your Constraints

1. No species is also a genus.
   $(\pi_{species} Species) \cap (\pi_{genus} Genus) = \emptyset$
2. No genus belongs to more than one family.
   $[\sigma_{(g1.genus=g2.genus \wedge g1.family \neq g2.family)}(\rho_{g1} Genus \times \rho_{g2} Genus)] = \emptyset$
   Although this constraint is already implied by the relation. Since genus is a key, the case where a genus belongs to two family will never happen in our relations.
3. All publications must be published after all artifacts they use have been collected.
   $\pi_{AN}(Published) - \pi_{AN}(Collected) = \emptyset$
4. Students may not catalogue live artifacts.
   $\sigma_{(rank='student' \wedge type='live')}(Artifact \bowtie Staff) = \emptyset$