CSC343 - Assignment 1

alsaidia, chochan2

February 7th, 2020

Our Constraints - Description

- The species of an artifact has to be present in the Species relation, this is important since we cannot tell the genus and family of a species if it is not in Species.
- A staff member can only be a technician, student, pre-tenure, tenure, this is so staff can be easily categorized.
- The family of a genus has to be present in the COL relation, this is important to keep the database consistent.
- The genus of a species has to be present in the Genus relation, this is so any species can be associated with a family.
- All collections are in both the Collection and Collected relations, this is required so we can connect every artifact to a Collection tuple.
- All artifacts have been collected in a collection trip, this is important so we can connect every artifact to a collection.
- All collections are done by staff members, so we can access information about the staff member that has done the collection.
- Artifacts are maintained by staff members, so we can connect every artifact with the staff member maintaining it.
- An artifact can only be of type tissue, image, model, live, this is required because these are the types of artifacts defined.
- All publications are about artifacts in the Artifacts relation, this is required since only artifacts that exist can be published about.

Queries

For every query the answer is stored in relation **Out**

1. **Longest** $:= \pi_{SID}(Staff) - \rho_{R(SID)}(\pi_{s1.SID}((\sigma_{s1.date>s2.date}(\rho_{s1}(Staff)) \times (\rho_{s2}(Staff)))))$

   CollectLongest $:= Collection \bowtie$ **Longest**

   NotRecent $:= \pi_{c1.SID,c1.date}((\rho_{c1}$**CollectLongest**$) \bowtie_{c1.SID=c2.SID \wedge c1.date<c2.date} (\rho_{c2}$**CollectLongest**$))$

   Out $:= \pi_{SID,date}($**CollectLongest**$) - \rho_{R(SID,date)}($**NotRecent**$)$

Relation	Description
Longest(SID)	Contains the SID(s) of staff that held their rank the longest
CollectLongest$(CID, date, SID)$	Tuples here represents all collections of staff that held their rank the longest
NotRecent$(c1.SID, c1.date)$	Tuples here contain all dates of collections of staff that held their rank the longest except the date of their most recent collection
Out$(SID, date)$	Tuples here contain the SID of staff members who have held their rank the longest and the dates of their most recent collection

2. **AandC** $:= \pi_{AN,SID,CID}(Artifact \bowtie Collected)$

   Multi$(CID) := \pi_{c1.CID}((\rho_{c1}$**AandC**$) \bowtie_{c1.AN \neq c2.AN \text{ and } c1.CID=c2.CID \text{ and } c1.SID \neq c2.SID} (\rho_{c2}$**AandC**$))$

   Single $:= \pi_{CID}(Collection) -$ **Multi**

   Out $:= \pi_{SID}($**AandC** $\bowtie$ **Single**$)$

Relation	Description
AandC(AN, SID, CID)	Tuples here represent the artifacts with their CIDs and the staff member that is maintaining them. If an artifact is aggregated over multiple collections then there will be a tuple for every collection the artifact is in
Multi(CID)	Contains the CIDs of collections that are maintained by more than one staff member
Single(CID)	Contains the CIDs of collections that are maintained by a single staff member
Out(SID)	Contains the SID of staff who maintain every artifact in at least one collection

3. **Out** $:= \pi_{AN}(Artifact \bowtie Collected \bowtie Collection)$

Relation	Description
Out(AN)	Contains the AN of artifacts that are collected and maintained by the same staff member

4. **ArtiCollector**$(AN, CSID) := \pi_{AN,SID}(Collected \bowtie Collection)$

   ArtiSpecColl$(AN, species, SID) := \pi_{AN,species,CSID}(Artifact \bowtie$ **ArtiCollector**$)$

   Perm3 $:= \sigma_{a1.AN!=a2.AN \wedge a2.AN!=a3.AN \wedge a3.AN!=a1.AN}((\rho_{a1}$**ArtiSpecColl**$) \times (\rho_{a2}$**ArtiSpecColl**$) \times (\rho_{a3}$**ArtiSpecColl**$))$

   Atleast3 $:= \pi_{a1.SID,a1.species}(\sigma_{a1.species=a2.species \wedge a2.species=a3.species \wedge a1.SID=a2.SID \wedge a2.SID=a3.SID}($**Perm3**$))$

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

   AtLeast3Fam $:= (\rho_{R(SID,species)}$**Atleast3**$) \bowtie$ **SpecFam**

   StaffSpecFam $:= (\pi_{SID} Staff) \times$ **SpecFam**

   Out $:= \pi_{SID}((\pi_{SID} Staff) \times COL - \pi_{SID,family}($**StaffSpecFam** - **AtLeast3Fam**$))$

Relation	Description
ArtiCollector$(AN, CSID)$	Tuples here represent the artifacts a staff member have collected, where CSID is the SID of the staff member that collected a piece or all of an artifact AN
ArtiSpecColl$(AN, species, SID)$	Same as **ArtiCollector** but also has the species of the artifact
Perm3	Contains all possible permutations of 3 different artifacts tuples from **ArtiSpecColl**
AtLeast3$(a1.SID, a1.species)$	Tuples here represent staff members that have collected at least 3 different artifacts from some species
SpecFam$(species, family)$	Tuples here contain every species and the family it belongs to
AtLeast3Fam$(SID, species, family)$	Same as **AtLeast3** but also contains the family of the species
StaffSpecFam$(SID, species, family)$	All possible tuples of staff SIDs and (species,family) pairs
Out(SID)	Contains the SID of staff who maintain every artifact in at least one collection

5. **AtLeastTwo** $:= (\rho_{p1} Published) \bowtie_{p1.AN!=p2.AN \wedge p1.journal=p2.journal} (\rho_{p2} Published)$

   AtLeastThree $:= (\rho_a$**AtLeastTwo**$) \bowtie_{a.p1.AN!=p3.AN \wedge a.p2.AN!=p3.AN \wedge a.p1.journal=p3.journal} (\rho_{p3} Published)$

   Out $:= \rho_{R1(journal)}(\pi_{p1.journal}($**AtLeastTwo**$)) - \rho_{R2(journal)}(\pi_{a.p1.journal}($**AtLeastThree**$))$

Relation	Description
AtLeastTwice	Contains journals that used at least two artifacts
AtLeastThrice	Contains journals that used at least three artifacts
Out$(journal)$	Contains journals that used exactly two artifacts in their publications

6. **NotCollected** $:= (\pi_{location}(Artifact)) \times COL - \pi_{location,family}(Artifact \bowtie Species \bowtie Genus)$

   Out $:= \pi_{location}(Artifact) - \pi_{location}($**NotCollected**$)$

Relation	Description
NotCollected$(location, family)$	Tuples here represent locations family pairs where no artifact in *family* has been collected in *location*
Out$(location)$	Tuples here represent locations where artifacts from every family have been collected

7. **TissueCollector** $:= \pi_{SID}(\sigma_{type='tissue'}((\pi_{AN,type}Artifact) \bowtie Collected \bowtie Collection))$

   OtherCollector $:= \pi_{SID}(\sigma_{type!='tissue'}((\pi_{AN,type}Artifact) \bowtie Collected \bowtie Collection))$

   Out $:=$ **TissueCollector** $-$ **OtherCollector**

Relation	Description
TissueCollector(SID)	SIDs of staff that have collected tissue samples
OtherCollector(SID)	SIDs of staff that have collected non-tissue samples
Out(SID)	SIDs of staff that only collected tissue samples

Continues on next page

8. **CollectionMaintainer**$(CID, MSID) := \pi_{CID,SID}(Artifact \bowtie Collected)$

   AtLeastTwo $:= \sigma_{MSID1!=CSID}((\rho_{R1(CID,MSID1)}$**CollectionMaintainer**$) \bowtie (\rho_{R2(CID,date,CSID)}Collection))$

   AtLeastThree $:= \sigma_{MSID1!=MSID2 \wedge CSID!=MSID2}($**AtLeastTwo** $\bowtie (\rho_{A(CID,MSID2)}($**CollectionMaintainer**$))))$

   PairCollections $:= \pi_{CID}($**AtLeastTwo**$) - \pi_{CID}($**AtLeastThree**$)$

   PairWorkers $:= \pi_{CID,MSID1,CSID}($**PairCollections** $\bowtie$ **AtLeastTwo**$)$

   AloneWorkers$(SID) := \pi_{MSID}((\pi_{CID}($**CollectionMaintainer**$) - \pi_{CID}($**AtLeastTwo**$)) \bowtie$ **CollectionMaintainer**$)$

   NonPairWorkers $:=$ **AloneWorkers**$\cup (\rho_{R1(SID)}(\pi_{MSID1}($**AtLeastThree**$))) \cup (\rho_{R2(SID)}(\pi_{CSID}($**AtLeastThree**$)))$

   OnlyPairWorkers $:= \pi_{CID,MSID1,CSID}((\rho_{pw}$**PairWorkers**$) \bowtie_{pw.MSID1!=npw.SID \wedge pw.CSID!=npw.SID} (\rho_{npw}$**NonPairWorkers**$))$

   PairWorkersDup $:=$ **PairWorkers** $\cup (\rho_{pw(CID,MSID1,CSID)}(\pi_{CID,CSID,MSID1}($**PairWorkers**$)))$

   MultiPairBuf $:= (\rho_{pw1}$**PairWorkersDup** $\bowtie_{pw1.CID!=pw2.CID} \rho_{pw2}$**PairWorkersDup**$)$

   MultiPair $:= \sigma_{(pw1.MSID1=pw2.MSID1 \wedge pw1.CSID!=pw2.CSID) \vee (pw1.MSID1=pw2.CSID!\wedge pw1.CSID!=pw2.MSID1)}$**MultiPairBuf**

   MultiPairWorker$(SID) := \pi_{pw1.MSID1}$**MultiPair**

   Out $:= \pi_{pw.MSID1,pw.CSID}((\rho_{pw}$**OnlyPairWorkers**$) \bowtie_{pw.MSID1!=mpw.SID \wedge pw.CSID!=mpw.SID} (\rho_{mpw}$**MultiPairWorker**$)$

Relation	Description
CollectionMaintainer$(CID, MSID)$	Tuples of all maintainers of all collections; MSID maintains an artifact in CID
AtLeastTwo$(CID, MSID1, date, CSID)$	Collections that at least two staff members worked on, CSID is the SID of the staff member that collected the collection, MSID1 is the SID of the staff member that is maintaining the artifacts in the collection
AtLeastThree$(CID, MSID1, date, CSID, MSID2)$	Collections that at least three staff members worked on
PairCollections(CID)	Collections that exactly two staff members worked on
PairWorkers$(CID, MSID1, CSID)$	Collections that exactly two staff members worked on with the SIDs of the staff membes that worked on it
AloneWorkers(SID)	Workers that collected and maintain at least one collection
NonPairWorkers(SID)	Workers that worked on collections that were not worked on by pairs
OnlyPairWorkers$(CID, MSID1, CSID)$	Collections were staff members only work in pairs
PairWorkersDup	Contains the two possible permutations for the pairs of workers
MultiPairBuf	Dummy variable because the select(theta join) did not fit on one line
MultiPair	Contains collections that have staff that work on multiple pairs, and the pairs are different
MultiPairWorkers	Staff that work on multiple pairs of collections, and the pairs they work in are different
Out$(MSID, CSID)$	Pairs of staff that only worked with each other on collections

9. The word "influence" indicates that a mutual relationship in between the two workers must happen, for any worker that worker 42 worked, which is recursive, something that cannot be expressed in relational algebra. Therefore, it cannot be expressed.

your constraints

- $\pi_{species}(Species) \cap \pi_{genus}(Genus) = \emptyset$
- $\emptyset = \emptyset$ (Constraint already satisfied since Genus is a key table Genus where its family is defined)
- **CollectionDates** $:= (\pi_{AN,date}(Collection \bowtie Collected))$

  NotFirstCollection $:= \pi_{cd1.AN,cd1.date}((\rho_{cd1}\textbf{CollectionDates}) \bowtie_{cd1.AN=cd2.AN \wedge cd1.date>cd2.date} (\rho_{cd2}\textbf{CollectionDates}))$

  FirstCollection $:=$ **CollectionDates** $- (\rho_{R(AN,date)}\textbf{NotFirstCollection})$

  $Published \bowtie_{Published.AN=FirstCollection.AN \wedge Published.date<FirstCollection.date}$ **FirstCollection** $= \emptyset$

Relation	Description
CollectionDates$(AN, date)$	Tuples here represent an artifacts collection dates, an artifact can have multiple collection dates as it can be aggregated over multiple collections
NotFirstCollection$(cd1.AN, cd1.date)$	Tuples here represent an artifacts collection dates, the relation excludes the earliest collection date. If there is only one collection date for an artifact then it is not present in this relation
FirstCollection$(AN, date)$	Tuples here represent an artifacts earliest collection date

- $\sigma_{type='live' \wedge rank='student'}(Artifact \bowtie Staff) = \emptyset$