

# CSC343 Assignment 1

Dominic Casillano, Arun Jolly Marlier

February 2020

## 1 Our Constraints

1. The species name of each artifact must be present in the Species relation. This is required to ensure that the species name is genuine.
2. The rank of every staff member is either a technician, student, pre-tenure, or tenure. This constraint is required because these are the only possible ranks for the staff members at this specific institution.
3. Every family of a genus must appear in the Catalogue of Life. This is required to ensure that the family name of the genus exists.
4. The genus of a species must be contained in the Genus relation. This is to ensure that the genus of the species is real.
5. The collection IDs in the Collected and Collection relations must be equal. This must hold because if a collection exists, then there must be artifacts found in those collections, and therefore the CID for each collection must exist in both relations.
6. The artifact numbers in the Collected and Artifact relations must be equal. This is required because since Artifact tuples represent an artifact collected in the field, it must also show up in the Collected relation.
7. The SID of the collector of a collection must be part of the staff. This is required because collections in this relation should only be owned by members of the institution's scientific staff.
8. The SID of the staff member who maintains the artifact must be part of the Staff collection. This is required because you only want staff members handling the artifacts.
9. The type of an artifact must be either tissue, image, model, or live. This is required because these are the only possible types of artifacts in the collections.
10. The artifact number of artifacts in publications must be contained in the Artifact relation. This is to ensure that the artifact being mentioned within the article, was actually collected.

## 2 Queries

1.  $NotLongest := \pi_{S1.SID}(\sigma_{today-S1.date < today-S2.date}((\rho_{S1}Staff) \times (\rho_{S2}Staff)))$   
 $LongestMembers := (\pi_{SID}Staff) - NotLongest$   
 $NotRecentCollection := \pi_{C1.SID, C1.date}(\sigma_{C1.date < C2.date}((\rho_{C1}Collection) \bowtie_{C1.SID=C2.SID} (\rho_{C2}Collection)))$   
 $MostRecentCollection := (\pi_{SID, date}Collection) - NotRecentCollection$   
 $Answer := \pi_{date}(LongestMembers \bowtie MostRecentCollection)$

The first query holds all the SIDs of staff that has held their rank for a shorter period of time compared to another staff member. NotMostRecentCollection holds every collection date of each staff member that isn't their most recent artifact. Therefore, to get LongestMember and MostRecentCollection, we can use set subtraction. Then Answer contains the date of the most recent collection of the staff member(s) who have held their current rank the longest.

2.  $CollectionMaintainers := \pi_{Collected.CID, Artifact.SID, Artifact.AN} (Artifact \bowtie_{Artifact.AN=Collected.AN} Collected)$   
 $MoreThanOneMaintainer := \pi_{C1.CID, C1.SID, C1.AN} ($   
 $\rho_{C1}CollectionMaintainers \bowtie_{C1.CID=C2.CID \wedge C1.SID \neq C2.SID \wedge C1.AN \neq C2.AN}$   
 $\rho_{C2}CollectionMaintainers)$

$Answer := \pi_{SID}(CollectionMaintainers - MoreThanOneMaintainer)$

CollectionMaintainers holds the CID, artifact AN, and SID of the staff who maintains that artifact. MoreThanOneMaintainer holds the same information, but only for collections with more than one maintainer. Then, the SIDs CollectionMaintainers - MoreThanOneMaintainer are those who solely maintain every artifact in a collection.

3.  $\pi_{Collection.SID}(Collection \bowtie_{Collection.CID=Collected.CID \wedge Collection.SID=Artifact.SID} (Collected \bowtie_{Collected.AN=Artifact.AN} Artifact))$

The query contains all collections and artifacts of those collections where the collector of the collection and the maintainer of the artifacts is the same staff member. Then the SID of the staff member is projected, as required.

4.  $ArtifactToFamily := \pi_{Artifact.AN, Artifact.SID, Artifact.species, Genus.family} (Artifact \bowtie (Species \bowtie Genus))$   
 $ThreeOfSameSpecies := \pi_{A1.species, A1.family, A1.SID}$   
 $((\rho_{A1}ArtifactToFamily \bowtie_{A1.AN \neq A2.AN \wedge A1.species=A2.species \wedge A1.SID=A2.SID}$   
 $\rho_{A2}ArtifactToFamily) \bowtie_{A2.species=A3.species \wedge A1.AN \neq A3.AN \wedge A2.AN \neq A3.AN \wedge A2.SID=A3.SID}$   
 $\rho_{A3}ArtifactToFamily)$

5.  $AtLeastTwo := \pi_{P1.journal}(\rho_{P1}Published \bowtie_{P1.journal=P2.journal \wedge P1.AN \neq P2.AN} \rho_{P2}Published)$   
 $AtLeastThree := \pi_{P1.journal}((\rho_{P1}Published \bowtie_{P1.journal=P2.journal \wedge P1.AN \neq P2.AN} \rho_{P2}Published) \bowtie_{P2.journal=P3.journal \wedge P3.AN \neq P1.AN \wedge P3.AN \neq P2.AN} \rho_{P3}Published)$   
 $Answer := AtLeastTwo - AtLeastThree$

AtLeastTwo contains all the journals that use at least two artifacts. AtLeastThree contains all the journals that use at least three artifacts. Subtracting the first from the second results in all journals that use exactly two artifacts.

6.  $LocationAndFamily := \pi_{Artifact.location, Genus.family}Artifact \bowtie (Species \bowtie Genus)$   
 $AllLocationsAndFamilies := (\pi_{location}Artifact) \times (\pi_{family}Genus)$   
 $NotAllFamiliesInLocation := \pi_{AllLocationsAndFamilies.location}(AllLocationsAndFamilies - LocationAndFamily)$   
 $Answer := (\pi_{location}Artifact) - NotAllFamiliesInLocation$

LocationAndFamily stores the location and family of all artifacts. AllLocationsAndFamilies contains every combination of location and family. If a location contains all families, then subtracting the first query from the second query should result in the location being deleted from the resulting query. Therefore, the locations that remain after the subtraction are the ones that don't contain an artifact from every family. Subtracting this from every location of a found artifact results in all the locations where at least one artifact from every family was found.

7.  $NotTissueSamples := \pi_{Collection.CID, Collection.SID}(Collection \bowtie_{Collected \bowtie (\sigma_{type='image' \vee type='live' \vee type='model'}Artifact)})$   
 $CollectionsOnlyTissueSamples := (\pi_{CID, SID}Collection) - NotTissueSamples$   
 $Answer := \pi_{SID}CollectionOnlyTissueSamples$

NotTissueSamples contains the CID and collector of collections where at least one artifact from that collection is not a tissue sample. Then to find all the collections that consist only of tissue samples, we can subtract all collections from the collections that have at least one non-tissue sample. The Answer query contains the SIDs of the collectors of such collections, as required.

8.  $CollectorMaintainerPairs(SID1, SID2) := \pi_{Collection.SID, Artifact.SID}(Collection \bowtie_{Collection.CID=Collected.CID} (Collected \bowtie_{Collected.AN=Artifact.AN} Artifact))$   
 $MaintainerPairs(SID1, SID2) := \pi_{C1.SID, C2.SID}(\rho_{C1}(Collected \bowtie Artifact) \bowtie_{C1.CID=C2.CID \wedge C1.SID \neq C2.SID} \rho_{C2}(Collected \bowtie Artifact))$   
 $AllPairs := CollectorMaintainerPairs \cup MaintainerPairs$

$DiversifiedPairs := \pi_{A1.SID1, A1.SID2}(\rho_{A1}AllPairs \bowtie_{A1.SID1=A2.SID1 \wedge A1.SID2 \neq A2.SID2} \rho_{A2}AllPairs)$

$Answer = AllPairs - DiversifiedPairs$

CollectorMaintainerPairs contains all pairs of collectors and maintainers who have worked together on a collection, while MaintainerPairs contains all pairs of maintainers who maintain artifacts in the same collection. DiversifiedPairs contains all pairs of staff where at least one of them has worked with another person. Then the answer is found by subtracting the diversified pairs from all pairs since this removes all pairs of staff that have not worked only with each other and nobody else.

9.  $CollectorMaintainerPairs(SID1, SID2) := \pi_{Collection.SID, Artifact.SID}Collection \bowtie_{Collection.CID=Collection.CID} Collected \bowtie_{Collected.AN=Artifact.AN} Artifact$   
 $MaintainerPairs(SID1, SID2) := \pi_{C1.SID, C2.SID} \rho_{C1}(Collected \bowtie Artifact)$   
 $\bowtie_{C1.CID=C2.CID \wedge C1.SID \neq C2.SID} \rho_{C2}(Collected \bowtie Artifact)$   
 $PairsWith42 := \sigma_{SID1=42}(CollectorMaintainerPairs \cup MaintainerPairs)$   
 $AllPairs := CollectorMaintainerPairs \cup MaintainerPairs$   
 $SecondDegreeInfluenced := \pi_{AllPairs.SID2}AllPairs$   
 $\bowtie_{AllPairs.SID1=PairsWith42.SID2 \wedge AllPairs.SID2 \neq 42} PairsWith42$   
 $Answer := (\pi_{SID2}PairsWith42) \cup SecondDegreeInfluenced$

The first two queries are the same as in Q8. The union the two queries contains all pairs of staff who have worked together on a collection (shown in the AllPairs query). PairsWith42 contains all pairs where the SID of one of the staff is 42. SecondDegreeInfluenced takes all partners of the member with SID = 42 and returns all staff that these people have worked with. Then the union of PairsWith42 (taking the staff who worked with the SID = 42 member) and SecondDegreeInfluenced returns all of the staff who were influenced by SID 42, as required.

### 3 My Constraints

1.  $\pi_{species}Species \cap \pi_{genus}Genus = \emptyset$   
 If the intersection between the names of species and genus is empty, then we can conclude that no species is also a genus.
2.  $\rho_{G1}Genus \bowtie_{G1.genus=G2.genus \wedge G1.family \neq G2.family} \rho_{G2}Genus = \emptyset$   
 This expression checks that a genus does not have more than one family, as required.
3.  $Published \bowtie_{Published.AN=Artifact.AN \wedge Published.date < Collection.date} ((Collection \bowtie Collected) \bowtie Artifact) = \emptyset$   
 This expression checks that the published date of the publication is earlier than the collection date of the artifacts used in the publication, which should be equal to the empty set.

$$4. \sigma_{type='live'}Artifact \bowtie \sigma_{rank='student'}Staff = \emptyset$$

This expression means that joining all collected live artifacts with students who collected those artifacts should result in the empty set, as required.