CSC343: Assignment 1

Submitted by: Saminul Islam (1004511833), Sadman Hasan(1003918246)

Part 1: Our constraints

- $\Pi_{species}$ (Artifact) - $\Pi_{species}$ (Species) = Ø

  Explanation: The expression implies the 'species' column from the Artifacts table only contains values selected from the unique values from the 'species' column from the Species table and is required to ensure Artifacts of every type of species have been catalogued.

- Π_{rank} (Staff) ⊆ {'technician','student','pre-tenure','tenure'}

  Explanation: The rank column in the Staff table can only take one of the assigned values {'technician','student','pre-tenure','tenure'} and is required to maintain a hierarchy within the Staff.

- Π_{family} (Genus) - Π_{family} (COL) = Ø

  Explanation: The 'family' column from the Genus table only contains values selected from the unique values from the 'family' column from the COL table and is required to ensure every type of genus is associated with a family.

- Π_{genus} (Species) ⊆ Π_{genus} (Genus)

  Explanation: The 'genus' column from the Species table only contains values selected from the unique values from the 'genus' column from the Genus table and is required to ensure no unidentified species exist in the catalogue.

- Π_{CID} (Collected) = Π_{CID} (Collection)

  Explanation: This implies that each collection id in the relation 'Collected' must also be present in the relation 'Collection' and it is required to ensure every collection of an artifact is associated with a staff id.

- Π_{AN} (Artifact) = Π_{AN} (Collected)

  Explanation: This implies that each artifact in the relation 'Artifact' must also be present in the relation 'Collected' and it is required to ensure that all the collected artifacts are recorded.

- Π_{SID} (Collection) ⊆ Π_{SID} (Staff)

  Explanation: : This implies that every SID in relation collection must be present in the relation Staff and it means that the collector must be a staff but not all staff members have collected an artifact and it is required to track the collection to the staff who made it.

- Π_{SID} (Artifact) $\subseteq \Pi_{SID}$ (Staff)

  Explanation: This implies that every SID in relation artifact must be present in the relation Staff and that not all staff members necessarily collect/maintain artifacts and it is required to track the artifact to the staff related to it.

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

  Explanation: This implies that the attribute 'type' in the relation 'Artifact' can only contain the following values: {'tissue,'image','model','live'} and it is required to ensure proper categorization of the artifacts.

- Π_{AN} (Published) $\subseteq \Pi_{AN}$ (Artifact)

  Explanation: This implies that every artifact in the published relation must be present in the artifact relation and that not all artifacts are published in a journal and it is required to make sure an artifact is only published after it has been recorded in the system.

Part 2: Queries

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

Solution:

// we find the staff member id of the longest rank holder, the table minComplement below contains every member excluding the longest rankholder
minComplement := $\sigma_{staff\_A.date > staff\_B.date}$ (ρ_{staff_A} (Staff) x ρ_{staff_B} (Staff))

longestRankHolder := $\Pi_{SID, date}$ ((Staff x Staff) - minComplement)

// Now we need to join the table with Collection to find the most recent collection date
joinCollection := longestRankHolder $\bowtie_{longestRankHolder.SID=Collection.SID}$ Collection

Answer := Π_{date} ((joinCollection x joinCollection) - $\sigma_{join\_A.date < join\_B.date}$ (ρ_{join_A} (joinCollection) x ρ_{join_B} (joinCollection)));

2.) Find all staff who maintain all artifacts in at least one collection.

Solution:

//we retrieve the required staff data of staff who maintain artifacts
StaffData(CID,AN,SID):= $\Pi_{CID, AN, SID}$ (Collection $\bowtie$ Artifact)

//find the collections which are maintained by at least 2 staff members
AtleastTwo(CID):= $\Pi_{T1.CID}(\sigma_{T1.CID \neq T2.CID \wedge T1.SID \neq T2.SID}(\rho_{T1}\ StaffData \times \rho_{T2}\ StaffData))$

//we find the collections which are maintained by 1 staff by taking the difference between all the collections and collections which are maintained by two people
OneCollection(CID):= Π_{CID} (Collected)-AtleastTwo

//then we find the staff who maintains that collection by natural joining
Answer(SID):= Π_{SID} (StaffData ⋈ OneCollection);

3.) Find all artifacts that were collected by the same staff who maintains them.

Solution:

//retrieve the required information from collection and collection. Rename collector SID to differentiate from maintainer SID.
CollectData(CID,date,cSID):= $\Pi_{CID,\ date,\ SID}$ (Collection ⋈ Collected)

//get answer by selection tuples where collector SID = maintainer SID
Answer(AN):= Π_{AN} ($\sigma_{cSID=SID}$ (CollectData ⋈ Artifact));

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

Solution:

//We natural join the appropriate tables first
artifactData := Artifact ⋈ Species ⋈ Genus ⋈ COL

// Now we self-join the created table and select Staff IDs who have collected at least 2 artifacts from the same species, of every species from some family
atleastTwo := $\sigma_{artifacts\_A.AN \neq artifacts\_B.AN\ \wedge\ artifacts\_A.SID = artifacts\_B.SID\ \wedge\ artifacts\_A.family = artifacts\_B.family\ \wedge\ artifacts\_A.species = artifacts\_B.species}$ ($\rho_{artifacts\_A}$ (artifactData) x $\rho_{artifacts\_B}$ (artifactData))

// Now we can repeat the process to identify staff members who have collected at least three artifacts
atleastThree := $\sigma_{artifacts\_A.AN \neq artifacts\_B.AN\ \wedge\ artifacts\_A.SID = artifacts\_B.SID\ \wedge\ artifacts\_A.family = artifacts\_B.family\ \wedge\ artifacts\_A.species = artifacts\_B.species}$ ($\rho_{artifacts\_A}$ (atleastTwo) x $\rho_{artifacts\_B}$ (atleastTwo))

Answer := P_{SID} (atleastThree)

5.) Find all publications that have used exactly 2 of our artifacts.

Solution:

//get required data from published relation
Publication(AN,Journal):= $\Pi_{AN,Journal}$(Published)

//find publications that have used at least 2 artifacts

AtleastTwo(Journal):= $\Pi_{T1.Journal}(\sigma_{T1.Journal \neq T2.Journal \wedge T1.AN < T2.AN}(\rho_{T1}\ Publication \times \rho_{T2}\ Publication))$

//find publications that have used at least 3 artifacts

AtleastThree(Journal):= $\Pi_{T1.Journal}(\sigma_{T1.Journal \neq T2.Journal \neq T3.Journal \wedge T1.AN < T2.AN < T3.AN}(\rho_{T1}\ Publication \times \rho_{T2}\ Publication \times \rho_{T3}\ Publication))$

//difference between AtleastTwo – AtleastThree are the publications that have used exactly 2 artifacts

Answer(Journal):= AtleastTwo – AtleastThree;

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

Solution:

//Get cartesian product between families and locations

LocationsAndFamilies := COL x $\Pi_{location}$ (Artifact)

//Get all locations and families combination where artifacts where found

AllLocationsFamiliesCombo := $\Pi_{Artifact.location,\ Genus.family}(\sigma_{Artifact.species = Species.species \wedge Species.genus = Genus.genus}$ (Artifact x Species x Genus))

//Get all locations where artifacts from no family where not found

NoFamily := $\Pi_{location}$ (LocationsAndFamilies - AllLocationsFamiliesCombo)

//Get locations where at least one artifact from every family has been collected

Answer := $\Pi_{location}$ (Artifact) – NoFamily

7.) Find all staff who have collected only tissue samples.

Solution:

//retrieve the required information from collection and collection. Rename collector SID to differentiate from maintainer SID

CollectorData(cSID,AN):= $\Pi_{SID,\ AN,}$ (Collection ⋈ Collected);

//get data about what types of samples collectors have collected. Rename cSID to SID for convenience.

TypesData(SID,Type):= $\Pi_{cSID,type}$ (CollectorData ⋈ Artifact);

//find staff who have collected artifacts excluding tissue type.

NotTissues(SID):= $\Pi_{SID}(\sigma_{type \neq "tissue"}$ (TypesData));

//find staff who have collected tissue samples(including other sample types)

Tissues(SID):= $\Pi_{SID}(\sigma_{type = "tissue"}$ (TypesData));

//difference gives us staff who have collected ONLY tissue samples.
Answer(SID):= Tissues – NotTissues;

8.) Find all staff pairs who have worked only with each other on collections.

Solution:
//Assuming that a collector and maintainer of a collection do not qualify as working together and that only maintainers of a collection qualify as working together.
//Get staff IDs of collection maintainers
maintainers := $\Pi_{C1.SID}$ (Collected ⋈ Artifact)

//Get staff IDs of people who worked with at least two members
pairs := $\sigma_{C1.SID \neq C2.SID \wedge C1.CID = C2.CID}$ (ρ_{C1} (maintainers) x ρ_{C2} (maintainers))

// Get staff IDs of people who worked with at least three members
triplets := $\sigma_{C1.SID \neq C3.SID \wedge}$ $\sigma_{C2.SID \neq C3.SID \wedge C1.CID = C3.CID}$ (pairs x ρ_{C3} (maintainers))

// SID of staff who worked in groups of two or more
atleastTwo := $\Pi_{C1.SID}$ (pairs) ∪ $\Pi_{C2.SID}$ (pairs)

// SID of staff who worked in groups of three or more
atleastThree := $\Pi_{C1.SID}$ (triplets) ∪ $\Pi_{C2.SID}$ (triplets) ∪ $\Pi_{C3.SID}$ (triplets)

// Get staff who have exclusively worked with each other
Answer := atleastTwo - alteastThree

9.) Staff member SID1 is in influenced by staff member SID2 if (a) they have ever worked together on a collection or (b) if SID1 has ever worked with a staff member who is in influenced by SID2. Find SIDs of staff members in influenced by SID 42.

Solution: Cannot be expressed.

Part 3: Your Constraints

1. No species is also a genus.

Solution:

$\Pi_{species}$ (Species) ∩ Πgenus (Genus) = ∅

2. No genus belongs to more than one family.

Solution:

Πgenus ($\sigma_{gen\_a.genus = gen\_b.genus \wedge gen\_a.family \neq gen\_b.family}$ (ρ_{gen_a} (Genus) x ρ_{gen_b} (Genus))) = $\emptyset$

3. All publications must be published after all artifacts they use have been collected.

Solution:

ArtifactData(AN,CollectionDate):= $\Pi_{AN,date}$ (Collection $\bowtie$ Collected);
PublicationData(AN,PublishDate):= $\Pi_{AN,date}$(Published)
$\sigma_{CollectionDate>PublicationDate}$ (ArtifactData $\bowtie$ PublicationData) = $\emptyset$

4. Students may not catalogue live artifacts.

Solution:
$\sigma_{type="tissue" \wedge rank="student"}$ (Artifact $\bowtie$ Staff) = $\emptyset$