TEAM MEMBERS

1. **Rahul Banerjee (1004328382)**
2. **Pratyush Menon (1004282661)**

Our Constraints

a) For every tuple in Artifact, a tuple with the same species must be present in the relation Species; an artifact cannot have a species which is not recorded in the relation.

b) The attribute rank in relation Staff must contain a value belonging to the following set: {‘technician', 'student', 'pre-tenure', 'tenure’}, in order to ensure that every staff member has a rank and the rank can only be one of the abovementioned ranks since no other rank exists in the museum.

c) The value of attribute Family in relation Species must be such that there exists a tuple in COL with that specific value, as the relation COL contains all the families we know so it makes sense that the attribute genus in relation GENUS belongs to a known family.

d) This is a foreign key constraint which implies that all values for attribute genus in relation Species must be present in the relation Genus since the relation Genus records all the values of the attribute genus we know.

e) Each CID corresponds to one field trip by a staff member on a certain date, and must also have one or more artifacts collected on that date by that staff member - i.e. a field trip is only recorded by its CID if one or more artifacts were found.

f) Each artifact number in the Collected relation corresponds to an artifact number in the Artifact relation; this means that each artifact has a species, type, location, Staff ID, and Collection ID associated with it.

g) Every SID in the Collection relation is present in the Staff relation; this means that the only people who can go out on collection field trips and collect artifacts are staff of the zoological institute, but not all staff of the institute need to necessarily go on field trips.

h) Every SID in the Artifact relation is present in the Staff relation; this means that every artifact is maintained by a staff member of the zoological institute, but not all staff of the institute need to necessarily maintain artifacts.

i) The attribute type in relation Artifact must contain a value belonging to the following set: {‘tissue’, 'image', 'model', 'live’} - this means that there are only 4 types of artifacts that the institute collects and archives, so no other types are valid.

j) Every Artifact number in the Published relation is present in the Artifact relation; this means that every artifact which has appeared in a publication must have been found and categorized earlier, but not all artifacts which are found by the institute will necessarily appear in a publication.

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.**

   attainedDate := $\Pi_{SID,date}$ (Staff)

   /*

   attainedDate contains the SIDs of the staff and the dates they attained their rank

   */

   A1 := attainedDate

   A2 := attainedDate

   notLongestDuration := $\Pi_{A1.SID}$ ($\sigma_{A1.date > A2.date}$ (A1 × A2))

   /*

   notLongestDuration contains SIDs of the staff who attained their rank after some other staff

   */

   longestDuration = Π_{SID} (Staff) – notLongestDuration

   /*

   longestDuration contains the SID(s) of the staff that attained their rank the before any other staff, i.e the staff who have held their rank for the longest duration

   */

Assumption: The date attribute in relation Collection is the date an artifact with the CID was collected

T1 := Collection ⋈ longestDuration

/*

T1 contains the SID(s) of the staff who held their rank the longest and dates they collected an artifact and CID of artifact

*/

T2 := T1

T3 := T1

artifactNotRecent := $\Pi_{T2.SID,T2.date}$ ($\sigma_{T2.SID = T3.SID \wedge T2.date < T3.date}$ (T2 × T3))

/*

artifactNotRecent contains dates of all the artifacts which were discovered before some other artifact by a longest rank duration staff and staff's SID

*/

MostRecentArtifact := $\Pi_{SID,date}$ (T1) – artifactNotRecent

/*

MostRecentArtifact contains the SID of longest rank duration staff and date of most recent artifact collected by that staff

*/

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

T1 := $\Pi_{CID, SID}$ (Collected ⋈ Artifact)

T2 := $\Pi_{CID, SID}$ (Collected ⋈ Artifact)

/*

T1 and T2 contain CIDs and the SIDs of the staff who maintains each artifact in each collection

*/

GeneralCollections := Π_{CID} ($\sigma_{T1.CID = T2.CID \wedge T1.SID \neq T2.SID}$ (T1 × T2))

/*

GeneralCollections is the list of collections which have artifacts that are maintained by different staff members

*/

SpecializedCollections := Π_{CID} (T1) – GeneralCollections

/*

SpecializedCollections is the list of collections which have artifacts that are maintained by only one staff member

*/

SpecializedStaff := Π_{SID} (T1 ⋈ SpecializedCollections)

/*

SpecializedStaff is the list of staff which maintain those collections

*/

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

Maintainers := $\Pi_{CID, SID, AN}$ (Collected ⋈ Artifact)

/*

Maintainers is the list of staff which maintain each artifact with its corresponding CID

*/

ArtifactDetails := $\Pi_{Maintainers.AN, Maintainers.SID, Collectors.SID}$ (Maintainers $\bowtie_{Maintainers.CID = Collection.CID}$ Collection)

/*

ArtifactDetails is the list of artifacts and the staff that collected them and maintain them

*/

SpecialArtifacts := $\Pi_{Maintainers.AN}$ ($\sigma_{Maintainers.SID = Collection.SID}$ (T2))

/*

SpecialArtifacts is the list of artifacts that are maintained by their collector

*/

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

T1 := $\Pi_{family,species}$ (Genus ⋈ (Species ⋈ Artifact))

/*

T1 contains the species, family of the artifact and SID of staff who collected it

*/

SpeciesFamily := $\Pi_{family,species}$ (Genus ⋈ Species)

/*

SpeciesFamily contains all the known species and its families

*/

notAllSpecies := SpeciesFamily – T1

/*

notAllSpecies contains the species whose artifact have not been collected

*/

staffNotAllSpecies := Π_{SID} (notAllSpecies ⋈ Artifact)

/*

SID of staff who didn't collect artifact from every species in a family

*/

StaffAllSpecies := Π_{SID} Staff – staffNotAllSpecies

/*

StaffAllSpecies contains SIDs of staff who collected artifact from all species in some family

*/

T3 := $\Pi_{family,species,SID,AN}$ (Genus ⋈ (Species ⋈ (Artifact ⋈ StaffAllSpecies))

/*

T3 contains SID of staff who collected species from every family, the species of the artifact they collected and its family

*/

Answer := $\sigma_{S1.SID = S2.SID = S3.SID \wedge S1.AN \neq S2.AN \wedge S1.AN \neq S3.AN \wedge S2.AN \neq S3.AN}$ (ρ_{S1} (T3) X ρ_{S2} (T3) X ρ_{S3} (T3))

/*

Answer contains all SIDs who have collectes at least 3

artifacts from species in a family

*/

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

T1 := Published

T2 := Published

T3 := Published

atLeastTwice := $\Pi_{journal}$ ($\sigma_{T1.journal = T2.journal \wedge T1.AN \neq T2.AN}$ (T1 × T2))

/*

atLeastTwice contains all the journals which used at least two of our artifacts

*/

atLeastThrice := $\Pi_{journal}$ ($\sigma_{T1.journal = T2.journa = T3.journal}$

$\wedge$ T1.AN ≠ T2.AN $\wedge$ T1.AN ≠ T3.AN $\wedge$ T3.AN ≠ T2.AN

(T1 × T2 × T3))

/*

atLeastThrice contains all the journals which used at least three of our artifacts

*/

ExactlyTwo = atLeastTwice – atLeastThrice

/*

ExactlyTwo contains all the journals which were in the relation atLeastTwice but not in atLeastThrice, i.e journals which used only two of our artifacts

*/

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

T1 := $\Pi_{family,location}$ (Genus ⋈ (Species ⋈ Artifact))

/*

T1 contains families and the location where it was found

*/

T2 := (Π_{family} Col) × ($\Pi_{location}$ Artifact)

/*

T2 contains all the different pairs of pairs and locations

*/

NotAllFamilies = $\Pi_{locations}$ (T2-T1)

/*

NotAllFamilies contains all the locations from where at least one artifact from a family wasn't found

*/

AllFamilies := $\Pi_{location}$ Artifact – NotAllFamilies

/*

AllFamilies contains all the locations where at least one artifact from every family was discovered

*/

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

T1 := $\Pi_{CID, AN, type}$ (Collected ⋈ Artifact)

/*

T1 contains each artifact number and its corresponding collection ID and type

*/

$T2 := \Pi_{SID,\ type,\ AN}(T1 \bowtie Collection)$

/*

T2 contains each artifact number, its type, and the staff responsible for collecting it

*/

$TissueCollectors := \Pi_{SID}(\sigma_{type = \text{'tissue'}}(T2))$

$OtherCollectors := \Pi_{SID}(\sigma_{type \neq \text{'tissue'}}(T2))$

/*

T3 contains each SID who has ever collected a tissue sample, and T4 contains each SID who has ever collected a non tissue sample

*/

TissueSpecialists := TissueCollectors – OtherCollectors

/*

TissueSpecialists contains each SID which has collected a tissue sample but not any other type of sample

*/

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

ASSUMPTION : ONLY MAINTAINERS WORK TOGETHER

M1 := $\Pi_{SID, CID}$ (Collected ⋈ Artifact)

M2 := $\Pi_{SID, CID}$ (Collected ⋈ Artifact)

/*

M1 and M2 contain the SIDs of the staff who maintain artifacts in each collection

*/

T1(SID1, SID2) := $\Pi_{M1.SID, M2.SID}$ (M1 $\bowtie_{M1.CID = M2.CID, M1.SID \neq M2.SID}$ M2)

T2(SID1, SID2) := $\Pi_{M1.SID, M2.SID}$ (M1 $\bowtie_{M1.CID = M2.CID, M1.SID \neq M2.SID}$ M2)

/*

T1 and T2 contain staff who have maintained a collection together before

*/

T3 := $\sigma_{(T1.SID1 = T2.SID1 \wedge T1.SID2 \neq T2.SID2) \vee (T1.SID2 = T2.SID2 \wedge T1.SID1 \neq T2.SID1)}$ (T1 × T2)

DiversePeople(SID) := $\Pi_{T1.SID1}$ (T3)

/*

DiversePeople contains a list of all staff who have ever worked with more than one person on a collection

*/

AllStaff := Π_{SID} (Staff)

/*

AllStaff contains the SIDs of everyone who works at the institute

*/

OnlyPairs(SID) := AllStaff - DiversePeople

/*

OnlyPairs contains the SID of staff members who have only ever worked with one other person while maintaining collections

*/

9. **Staff member SID1 is 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 influenced by SID2. Find SIDs of staff members influenced by SID 42.**

   Not doable using Relation Algebra

Your Constraints

1) **No species is also a genus**

   nameSpecies (names) := $\Pi_{species}$ (Species)
   /*
   nameSpecies contains all the names of species under the attribute names
   */

nameGenus (names) := Π_{genus} (Genus)

/*

nameGenus contains all the names of Genus under the attribute names

*/

commonNames := nameSpecies ∩ nameGenus

/*

commonNames contains all the names common to both nameSpecies and name Genus

*/

commonNames = Ø

/*

the resulting intersection must be empty, i.e relation of names which is both a species and a genus must be empty

*/

2) No genus belongs to more than one family

T1 := Genus
T2 := Genus
T3 := T1 × T2

/*

T3 contains a self-join of relation Genus with itself

*/

atLeastTwo := $\Pi_{T1.genus}$ ($\sigma_{T1.genus=T2.genus \wedge T1.family \neq T2.family}$ (T3))

/*

atLeastTwo contains all the genus which belong to 2 or more families.

*/

AtLeastTwo = Ø

/*

atLeastTwo is set to null, i.e the relation containing all the genus with two or more families must be empty

*/

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

publishedDate(CID,cDate) = $\Pi_{CID,date}$ (Published ⋈ Collected)

/*

publishedDate contains the CID and the date it was published

*/

collectedDate(CID,pDate) = $\Pi_{CID,date}$ (Collection)

/*

collectedDate contains the CID and the date it was collected

*/

T1 := publishedDate ⋈ collectedDate

/*

T1 contains the CID of the collection, date it was collected and date it was published

*/

Answer := Π_{CID} ($\sigma_{pDate<cDtae}$ (T1))

/*

Answer contains all the CIDs such that it's publish date was before its collected Date.

*/

Answer := Ø

/*

The relation containing CIDs whose publish date is before the collected date must be Empty

*/

4) Students may not catalogue live artifacts.

artifactsStudents := $\Pi_{AN,type,rank}$ (Artifact ⋈ Staff)

/*

artifactStudents contains the attributes AN,type,rank after a natural join of Artifact and Staff

*/

Answer := Π_{AN} ($\sigma_{rank='student'^type='live'}$ (artifactStudents))

/*

Answer contains all the artifact IDs collected by a student whose # type is live

*/

Answer = Ø

/*

The relation containing Live Artifacts collected by students must be empty

*/

-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x **THE END** –x-x-x-x-x-x-x-x-x-x-x-x-x-x-