csc343 winter 2020
assignment #1: relational algebra

Xiu Yan Lu 1003878235
Yulin WANG 1003942326

Part1 our constraints

For each of the following constraints give a one sentence explanation of what the constraint implies, and why it is required.

- $\pi_{species}(Artifact) - \pi_{species}(Species) = \emptyset$.

  This implies that an artifact's species must appear in the species of Species, because the artifact's official species name must be recorded and appear in the COL(Catalogue of Life)

- $\pi_{rank}(Staff) \subseteq$ {'technician', 'student', 'pre-tenure', 'tenure'}

  Rank of a staff must be one of 'technician', 'student', 'pre-tenure' and 'tenure', these are the only obtainable ranks

- $\pi_{family}(Genus) - \pi_{family}(COL) = \emptyset$.

  Every family name that a genus belongs to is a scientific zoological family name that must appears in the Catalogue of Life

- $\pi_{genus}(Species) \subseteq \pi_{genus}(Genus)$.

  Each species belongs to and must belong to exactly one genus which is in Genus, but not necessarily all genus in Genus has a species in Species that belongs to that genus

- $\pi_{CID}(Collected) = \pi_{CID}(Collection)$.

  All of the CID from Collection belongs to Collected and all the CID from Collected belongs to Collection, every collection has at least one artifact and every artifact belongs to some collection

- $\pi_{AN}(Artifact) = \pi_{AN}(Collected)$.

  All AN in Artifact are in ANs of Collected, all AN in Collected are also in ANs of Artifact, every artifact must be aggregated from at least one collection

- $\pi_{SID}(Collection) \subseteq \pi_{SID}(Staff)$.

  Every SID in Collection is in SIDs of staff, which means all collectors in Collection must be a Staff in database

- $\pi_{SID}(Artifact) \subseteq \pi_{SID}(Staff)$.

  Every SID of maintainer in Artifact is one of the SIDs in Staff, which means all maintainers in Artifact must be a Staff in database

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

  The type of an artifact in Artifact must be one of 'tissue', 'image', 'model' or 'live'

- $\pi_{AN}(Published) \subseteq \pi_{AN}(Artifact)$

  The artifact that are mentioned in some publication in Published must also be an artifact collected in Artifact, but not all the artifacts have been published

Part2 queries

1. Rationale: Performance reviews include seeing how current the work is of staff who have held their current rank for a long time.

   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.

   -All the staff have ever been collectors

   $$Collectors(SID, date) := \Pi_{SID,date}((\Pi_{SID}(Collection)) \bowtie Staff)$$

   -collector who does not hold the longest current rank

   $$NotMinDate(SID) := \Pi_{C1.SID}\sigma_{C1.date>C2.date}[(\rho_{C1}Collectors) \times (\rho_{C2}Collectors)]$$

   -collector who holds the longest current rank

   $$MinDate(SID) := (\Pi_{SID}Collectors) - NotMinDate$$

   -all collections by this collector

   $$Collections(CID, date) := \Pi_{CID,date}(Collection \bowtie MinDate)$$

   -Not the most recent collection date

   $$NotMostRecent(date) := \Pi_{C1.date}\sigma_{C1.date<C2.date}[(\rho_{C1}Collections) \times (\rho_{C2}Collections)]$$

   -Answer:most recent collection date

   $$Answer(date) := (\Pi_{date}Collections) - (\Pi_{date}NotMostRecent)$$

2. Rationale: Staff who maintain every artifact in some collection should be considered favourably in performance reviews.

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

   -Collected Artifact with the maintainer information

   $$ArtifactMaintainer(CID, AN, SID) := \Pi_{CID,AN,SID}(Artifact \bowtie Collected)$$

-Collections that have different maintainers

$$DifferentMaintainer(CID, AN, SID) :=$$

$$\Pi_{A1.CID,A1.AN,A1.SID}\sigma_{(A1.CID<A2.CID)\wedge(A1.SID\neq A2.SID)}$$
$$(\rho_{A1}ArtifactMaintainer \times \rho_{A2}ArtifactMaintainer)$$

-Answer: SIDs of the staffs who maintain all artifacts in at least one collection

$$Answer(SID) := (\Pi_{SID}ArtifactMaintainer) - (\Pi_{SID}DifferentMaintainer)$$

3. Rationale: An artifact collected and maintained by the same staff may have some special requirements that should be investigated.

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

   -Artifact with collection CID

   $$ArtifactC(CID, AN, maintainer) := \Pi_{CID,AN,SID}(Artifact \bowtie Collected)$$

   -Artifact with collectors' staff ID

   $$ArtifactWorker(AN, maintainer, collector) := \Pi_{AN,maintainer,SID}(ArtifactC \bowtie Collection)$$

   -Answer:All artifacts have the same maintainer and collector

   $$Answer(AN) := \Pi_{AN}\sigma_{maintainer=collector}ArtifactWorker$$

4. Rationale: Identify multi-talented field workers.

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

   -Artifact with Collection information

   $$Artifacts(CID, AN, species) := \pi_{CID,AN,species}(Artifact \bowtie Collected)$$

   -Artifact with collector information

   $$NewArtifact(AN, collector, species) := \Pi_{AN,SID,species}(Artifacts \bowtie Collection)$$

   -Combine Species and Family information

   $$SF(species, family) := \Pi_{species,family}(Genus \bowtie Species)$$

   -Artifact with family information

   $$ArtifactSF(AN, collector, species, family) := \Pi_{AN,collector,species,family}(NewArtifact \bowtie SF)$$

   -All combos should have occurred

   $$ShouldHaveBeen(AN, collector, species, family) :=$$

   $$\Pi_{AN,collector,species,SF.family}\sigma_{ArtifactSF.family=SF.family}$$

$$[(\Pi_{AN,collector,family} ArtifactSF) \times SF]$$

-get the 'did not always' part

$$WereNotAlways(AN, collector, species, family) := ShouldHaveBeen - ArtifactSF$$

-Every species in some family

$$EverySpecies(collector) := (\Pi_{collector} ArtifactSF) - (\Pi_{collector} WereNotAlways)$$

-Answer: All staff who have collected at least 3 artifacts from every species in some family

$$Answer(collector) := \Pi_{E1.collector}$$

$$\sigma_{(E1.AN<E2.AN<E3.AN)\wedge(E1.collector=E2.collector=E3.collector)}$$
$$[(\rho_{E1} EverySpecies)] \times [(\rho_{E2} EverySpecies)] \times [(\rho_{E3} EverySpecies)]$$

5. Rationale: Which publications might have some specialized niche focus?
   Query: Find all publications that have used exactly 2 of our artifacts.

   -used at least 2 of our artifacts

   $$AtLeastTwo(journal) := \Pi_{P1.journal} \sigma_{(P1.AN \neq P2.AN)\wedge(P1.journal=P2.journal)} [(\rho_{P1} Published) \times (\rho_{P2} Published)]$$

   -used at least 3 of our artifacts

   $$AtLeastThree(journal) :=$$
   $$\Pi_{P1.journal} \sigma_{(P1.AN<P2.AN<P3.AN)\wedge(P1.journal=P2.journal=P3.journal)}$$
   $$[(\rho_{P1} Published) \times (\rho_{P2} Published) \times (\rho_{P3} Published)]$$

   -Answer: Exactly two artifacts

   $$Answer(journal) := AtLeastTwo - AtLeastThree$$

6. Rationale: Identify motherlode locations.
   Query: Find all locations where at least one artifact from every family has been collected.
   -Artifact with location and family information

   $$ArtifactInfo(AN, species, location, family) := \Pi_{AN,species,location,family} (Artifact \bowtie Genus \bowtie Family)$$

   -Checklist of locations and family

   $$Checklist(location, family) := \Pi_{location,family} (COL \times Artifact)$$

   -Families that does not appear at that location

   $$NotAppear(location, family) := Checklist - \Pi_{location,family} ArtifactInfo$$

   -Answer: Locations where at least one artfact from every family has been collected

   $$Answer(location) := (\Pi_{location} Artifact) - (\Pi_{location} NotAppear)$$

7. Rationale: Exclusively tissue sample collectors may need extra support for special reagents and shipping costs.

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

   -Collection with collector information

   $$Collector(CSID, AN, CID) := \Pi_{Collection.SID, Collected.AN, Collected.CID}(Collected \bowtie Collection)$$

   -Artifact with collector information and type

   $$ArtifactInfo(SID, AN, CID, type) := \Pi_{Collector.CSID, Collector.AN, Collector.CID, Artifact.type}(Collector \bowtie Artifact)$$

   -Staff who collected any other types

   $$NotTissue(SID) := \Pi_{SID}\sigma_{A.type \neq 'tissue'}(\rho_A ArtifactInfo)$$

   -Answer:Staff who collected only tissue

   $$Answer(SID) := (\Pi_{SID} ArtifactInfo) - NotTissue$$

8. Rationale: Collection staff who should be encouraged to diversify their network.

   Query: Find all staff pairs who have worked only with each other on collections. -Collection with collector information

   $$Collector(CSID, AN, CID) := \Pi_{Collection.SID, Collected.AN, Collected.CID}(Collected \bowtie Collection)$$

   -Artifact with collector information and maintainer information

   $$ArtifactInfo(CSID, CID, ASID, AN) :=$$

   $$\Pi_{Collector.CSID, Collector.CID, Artifact.SID, Artifact.AN}(Collector \bowtie Artifact)$$

   -Maintainer and Collector are the same person

   $$SamePerson(CSID, CID, ASID, AN) :=$$

   $$\Pi_{A1.CSID, A1.CID, A1.ASID, A1.AN}$$
   $$\sigma_{A1.CSID = A1.ASID}(\rho_{A1} ArtifactInfo)$$

   -Different maintainer and collector

   $$Distinct(CSID, ASID) := \Pi_{CSID, ASID}(ArtifactInfo - SamePerson)$$

   -A maintainer of an artifact that worked with multiple collector

   $$MultiCollector(CSID, ASID) := \Pi_{D1.CSID, D1.ASID}$$

   $$\sigma_{(D1.CSID \neq D2.CSID) \wedge (D1.ASID = D2.ASID)}(\rho_{D1} Distinct \bowtie \rho_{D2} Distinct)$$

   -A collector that worked with multiple maintainers

   $$MultiMaintaieor(CSID, ASID) := \Pi_{D1.CSID, D1.ASID}$$

   $$\sigma_{(D1.CSID = D2.CSID) \wedge (D1.ASID \neq D2.ASID)}(\rho_{D1} Distinct \bowtie \rho_{D2} Distinct)$$

   -Answer: Staff pairs who have worked only with each other on collectons

   $$Answer(CSID, ASID) := Distinct - MultiCollector - MultiMaintainer$$

9. Rationale: Track the influence of a given staff member.

   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.

   -Answer: cannot be expressed since the requirement(b) is recursively defined.

Part3 your constraints

For each of these constraints you should derive a relational algebra expression of the form $R = \emptyset$, where R may be derived in several steps, by assigning intermediate results to a variable. If the constraint cannot be expressed in the relational algebra you have been taught, write "cannot be expressed."

1. No species is also a genus.

$$\sigma_{S.species=G.genus}(\rho_S Species \times \rho_G Genus) = \emptyset$$

2. No genus belongs to more than one family.

$$\sigma_{(G1.genus=G2.genus)\wedge(G1.family\neq G2.family)}[(\rho_{G1} Genus) \times (\rho_{G1} Genus)] = \emptyset$$

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

   -Artifact with collection information

$$Artifacts(AN, CollectedDate) := \Pi_{AN,date}(Collected \bowtie Collection)$$

   -Add the publication information

$$PublishArtifact(AN, CollectedDate, PublishedDate) := \Pi_{AN,CollectedDate,date}(Artifacts \bowtie Published)$$

   -Answer: no publications are published before collected

$$\sigma_{PublishedDate<CollectedDate} PublishArtifact = \emptyset$$

4. Students may not catalogue live artifacts.

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