CSC 343 Assignment 1

Jerry He 1003979180

February 7, 2020

Our Constraints

- $\pi_{species}(Artifact) - \pi_{species}(Species) = \emptyset$.
  Every artifact belongs to a known species in the COL database; otherwise, artifacts could have unrecorded species names.

- $\pi_{rank}(Staff) \subseteq$ {'technician', 'student', 'pre-tenure', 'tenure'}.
  Every staff member must have a rank of technician, student, pre-tenure, or tenure; these are the only possible ranks at the various institutes.

- $\pi_{family}(Genus) - \pi_{family}(COL) = \emptyset$.
  Every genus belongs to a known family in the COL database; otherwise, a genus could belong to an unrecorded family.

- $\pi_{genus}(Species) \subseteq \pi_{genus}(Genus)$.
  Every species belongs to a known genus in the COL database; otherwise, a species could belong to an unrecorded genus.

- $\pi_{CID}(Collected) = \pi_{CID}(Collection)$.
  Every individual collected item belongs to an entire collection from a field trip; otherwise, information about where an artifact was collected could be unrecorded.

- $\pi_{AN}(Artifact) = \pi_{AN}(Collected)$.
  Every individual artifact/part of an artifact belongs to one or more collections from field trips; otherwise, information about where an artifact was collected could be unrecorded.

- $\pi_{SID}(Collection) \subseteq \pi_{SID}(Staff)$.
  Every collection was done by a registered staff member; otherwise, the one who gathered a collection could be unrecorded.

- $\pi_{SID}(Artifact) \subseteq \pi_{SID}(Staff)$.
  Every artifact was collected by a registered staff member; otherwise, the one who collected an artifact could be unrecorded.

- $\pi_{type}(Artifact) \subseteq \{'tissue', 'image', 'model', 'live'\}$
  Every artifact is either tissue, an image, a model, or a live colony; these are the only possible types of artifacts.

- $\pi_{AN}(Published) \subseteq \pi_{AN}(Artifact)$
  Every mention of an artifact in a journal must refer to an artifact in the COL database; otherwise, journals could refer to nonexistent artifacts.

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.

   - $NotLongest(SID) := \Pi_{S1.SID}[\sigma_{S1.date>S2.date}$
     $[(\rho_{S1}Staff) \times (\rho_{S2}Staff)]]$

     The staff members that have not held their current rank the longest.
   - $Longest(SID) := (\Pi_{SID}Staff) - NotLongest$

     The staff member(s) that have held their current rank the longest.
   - $NotMostRecent(date) := \Pi_{C1.date}[\sigma_{\substack{C1.date<C2.date\wedge \\ C1.SID=Longest\wedge \\ C2.SID=Longest}}$
     $[(\rho_{C1}Collection) \times (\rho_{C2}Collection)]]$

     The artifacts collected by the staff member(s) that are not the most recent.
   - $Answer(date) := (\Pi_{date}[\sigma_{SID=Longest}Collection)] - NotMostRecent$

     The artifacts collected by the staff member(s) that are the most recent.

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.

   - $CollectedAndArt(CID, SID) := \Pi_{CID,SID}[Collected \bowtie Artifact]$

     The CIDs of collections and the SIDs of the staff who maintain said collections.
   - $NotOneStaff(CID) := \Pi_{C1.CID}[\sigma_{\substack{C1.CID=C2.CID\wedge \\ C1.SID\neq C2.SID}}$
     $[(\rho_{C1}CollectedAndArt) \times (\rho_{C2}CollectedAndArt)]]$

     The collections that are not maintained by only one staff member.
   - $OneStaff(CID) := (\Pi_{CID}Collection) - NotOneStaff$

     The collections that are maintained by only one staff member.
   - $Answer(SID) = \Pi_{SID}[OneStaff \bowtie Collection]$

     The SIDs of the staff members that maintain all artifacts in at least one collection.

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.

   - $Answer := \Pi_{AN}[Artifacts \bowtie Collection \bowtie Collected]$

     The artifacts whose collection staff and maintaining staff are the same.

4. Rationale: Identify multi-talented field workers.

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

   - $CollAndArt(SID, AN, species) := \Pi_{SID,AN,species}$
     $[Collection \bowtie Collected \bowtie (\Pi_{AN,species} Artifact)]$

     The AN of an artifact, the SID of the collector, and which species the artifact belongs to.

   - $ThreeCollector := \Pi_{C1.SID,C1.species}[\sigma_{\substack{C1.AN \neq C2.AN \wedge \\ C2.AN \neq C3.AN \wedge \\ C1.AN \neq C3.AN \wedge \\ C1.SID = C2.SID \wedge \\ C2.SID = C3.SID \wedge \\ C1.species = C2.species \wedge \\ C2.species = C3.species}}$
     $[(\rho_{C1} CollAndArt) \times (\rho_{C2} CollAndArt) \times (\rho_{C3} CollAndArt)]]$

     The SIDs of the collectors who have collected at least three artifacts from a species (triple collectors) and the species they collected.

   - $FamAndSpecies := \Pi_{family,species}[Genus \bowtie Species]$

     All existing species and the families they come from.

   - $CollectorSpeciesAndFam := ThreeCollector \bowtie FamAndSpecies$

     The SIDs of the triple collectors, the species they collected, and the family those species belong to.

   - $CollectorAndFam := \Pi_{SID,family} FamAndSpecies$

     The SIDs of the triple collectors and the family of the species they collected.

   - $EveryCollectorSpeciesAndFam := CollectorAndFam \bowtie FamAndSpecies$

     The SIDs of the triple collectors, the family of the species they collected, and ALL the species in those families.

   - $NonOccurringCollectorAndFam := \Pi_{SID,family}$
     $[EveryCollectorSpeciesAndFam - CollectorSpeciesAndFam]$

The SIDs of the triple collectors, and the families where there was at least one species where the collector did not collect three artifacts from said species.

- $Answer := \Pi_{SID}[CollectorAndFam - NonOccuringCollectorAndFam]$

  The SIDs of the staff who have collected at least three artifacts from every species in some family.

5. Rationale: Which publications might have some specialized niche focus?

   Query: Find all publications that have used exactly 2 of our artifacts.

   - $AtLeastTwo(journal) := \Pi_{journal}[\sigma_{\substack{P1.journal=P2.journal \wedge \\ P1.date=P2.date \wedge \\ P1.AN \neq P2.AN}}$

     $[(\rho_{P1}Publications) \times \rho_{P2}Publications)]]$

     The journals that use two or more artifacts.

   - $AtLeastThree(journal) := \Pi_{journal}[\sigma_{\substack{P1.journal=P2.journal \wedge \\ P2.journal=P3.journal \wedge \\ P1.date=P2.date \wedge \\ P2.date=P3.date \wedge \\ P1.AN \neq P2.AN \wedge \\ P2.AN \neq P3.AN}}$

     $[(\rho_{P1}Publications) \times (\rho_{P2}Publications) \times (\rho_{P3}Publications)]]$

     The journals that use three or more artifacts.

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

     The journals that use exactly two artifacts.

6. Rationale: Identify motherlode locations.

   Query: Find all locations where at least one artifact from every family has been collected.

   - $FamiliesFound := \Pi_{family,location}[Artifact \bowtie Genus \bowtie Species]$

     The actual families found and the locations they were found at.

   - $AllFamsAndLocations := (\Pi_{location}FamiliesFound) \times COL$

     All combinations of families and locations given the real locations that families were found at.

   - $NonMotherlodeLocations := \Pi_{location}$
     $[AllFamsAndLocations - FamiliesFound]$

     The locations that are not motherlodes, i.e. the locations that did not have at least one artifact from every family collected.

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

  The motherlode locations, i.e. the locations where at least one artifact from every family was collected.

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.

   - $NotTissue(SID) := \Pi_{SID}[\sigma_{\substack{type='image' \vee \\ type='model' \vee \\ type='live'}} Artifact]$

     The SIDs of the staff who have collected non-tissue artifacts.

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

     The SIDs of the staff who have only collected tissue samples.

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.

   - $Collect(CID, date, SID, AN) := Collection \bowtie Collected$

     Combined Collection and Collected into Collect.

   - $CollectorsWorkedWithOthers(SID) := \Pi_{C.SID}$
     $[\sigma_{\substack{C.AN=Artifact.AN \wedge \\ C.SID \neq Artifact.SID}} [\rho_C[Collect] \times Artifact]]$

     The collectors who have worked with other maintainers.

   - $MaintainersWorkedWithOthers(SID) := \Pi_{Artifact.SID}$
     $[\sigma_{\substack{C.AN=Artifact.AN \wedge \\ C.SID \neq Artifact.SID}} [(\rho_C Collect) \times Artifact]]$

     The maintainers who have worked with other collectors. We do not need to check for maintainers who have worked with other maintainers since different maintainers implies that the collector is working with someone other than him/herself. Both the collector and the maintainers would be covered by the two cases above.

   - $LoneWolves(SID) := (\Pi_{SID} Staff) - CollectorsWorkedWithOthers - MaintainersWorkedWithOthers$

     The SIDs of staff who have only worked alone.

- $MoreThanOne1(SID) := \Pi_{C1.SID}[\sigma_{\substack{C1.SID=C2.SID \wedge \\ C1.AN=A1.AN \wedge \\ C1.SID \neq A1.SID \wedge \\ C2.AN=A2.AN \wedge \\ C2.SID \neq A2.SID \wedge \\ A1.SID \neq A2.SID}}[(\rho_{C1}Collect) \times (\rho_{A1}Artifact) \times (\rho_{C2}Collect) \times (\rho_{A2}Artifact)]]$

  The relation within the $\Pi_{C1.SID}$ consists of two copies of Collect, C1 and C2, and two copies of Artifact, A1 and A2. C1 and A1 have the same AN, and C2 and A2 have the same AN, i.e. C1 and A1 refer to the same artifact, and C2 and A2 refer to the same artifact. It is possible for C1, C2, A1, and A2 to refer to the same artifact. C1 and C2 have the same SID, i.e. the same staff member collected the one or two artifacts in a single tuple. A1 and A2 have SIDs different from each other and the collector. If a tuple exists, that means the staff from C1 has worked with staff from A1 and A2, which means that none of the three are from a pair that works exclusively with each other. The SIDs of A1 and A2 are found in the following two expressions.

- $MoreThanOne2(SID) := \Pi_{A1.SID}[\sigma_{\substack{C1.SID=C2.SID \wedge \\ C1.AN=A1.AN \wedge \\ C1.SID \neq A1.SID \wedge \\ C2.AN=A2.AN \wedge \\ C2.SID \neq A2.SID \wedge \\ A1.SID \neq A2.SID}}[(\rho_{C1}Collect) \times (\rho_{A1}Artifact) \times (\rho_{C2}Collect) \times (\rho_{A2}Artifact)]]$
- $MoreThanOne3(SID) := \Pi_{A2.SID}[\sigma_{\substack{C1.SID=C2.SID \wedge \\ C1.AN=A1.AN \wedge \\ C1.SID \neq A1.SID \wedge \\ C2.AN=A2.AN \wedge \\ C2.SID \neq A2.SID \wedge \\ A1.SID \neq A2.SID}}[(\rho_{C1}Collect) \times (\rho_{A1}Artifact) \times (\rho_{C2}Collect) \times (\rho_{A2}Artifact)]]$
- $ExclusiveStaff(SID) := (\Pi_{SID}Staff) - LoneWolves - MoreThanOne1 - MoreThanOne2 - MoreThanOne3$

  The SIDs of collection staff that have worked in an exclusive pair on collections.

- $Collectors(SID, AN) := \Pi_{SID,AN}[ExclusiveStaff \bowtie Collect]$

  The collectors in the exclusive pair (one of the members must be a collector).

- $Maintainers(SID, AN) := \Pi_{SID,AN}[ExclusiveStaff \bowtie Artifact]$

  The maintainers in the exclusive pair (at least one of the members must be a maintainer).

- $Answer(SID, SID) := \Pi_{Collectors.SID,Maintainers.SID}$
  $[\sigma_{\substack{Collectors.AN=Maintainers.AN \wedge \\ Collectors.SID \neq Maintainers.SID}}[Collectors \times Maintainers]]$

The SID pairs of all staff pairs who have worked only with each other on collections.

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.

   - Cannot be expressed. This is a recursive query.

Your Constraints

1. No species is also a genus.
   - $(\Pi_{Species} Species) \cap (\Pi_{Genus} Genus) = \emptyset$

2. No genus belongs to more than one family.
   - $(\rho_{G1} Genus) \bowtie_{\substack{G1.Genus=G2.Genus \wedge \\ G1.Family \neq G2.Family}} (\rho_{G2} Genus) = \emptyset$

3. All publications must be published after all artifacts they use have been collected.
   - $CollectedAndPub(CID, date) := \Pi_{CID,date}[Collected \bowtie Published]$
   - $Collection \bowtie_{\substack{Collection.CID=CollectedAndPub.CID \wedge \\ Collection.date>CollectedAndPub.date}} CollectedAndPub = \emptyset$

     There are no artifacts whose collection date is after the publication date of a journal referring to said artifacts.

4. Students may not maintain live artifacts.
   - $Students(SID) = \Pi_{SID}[\sigma_{rank='student'} Staff]$

     SIDs of students
   - $\sigma_{type='live'}[Artifact \bowtie Students] = \emptyset$