CSC343 Assignment 1: Relational Algebra

Ryan Chang, Le Rong Shi

Our Constraints

- $\pi_{species}(Artifact) - \pi_{species}(Species) = \emptyset$.
  All species of an artifact will be a species in Relation Species. This is because all artifacts should belong to a species in the Species relation.

- $\pi_{rank}(Staff) \subseteq$ {'technician', 'student', 'pre-tenure', 'tenure'}.
  The ranks of a staff can only be a subset of 'technician', 'student', 'pre-tenure', 'tenure'. This is because these are the defined ranks a staff can be.

- $\pi_{family}(Genus) - \pi_{family}(COL) = \emptyset$.
  The distinct family under relation Genus are the same ones under relation COL. This is because all Genuses should belong to a family in the COL relation.

- $\pi_{genus}(Species) \subseteq \pi_{genus}(Genus)$.
  For any species, its genus must be under the Genus relation. This is because any genus of a species should be defined under the Genus Relation.

- $\pi_{CID}(Collected) = \pi_{CID}(Collection)$.
  The distinct CIDs in relation Collected are equivalent to the distinct CIDs in relation Collection. This is because all CIDs in either Collected or Collection should match.

- $\pi_{AN}(Artifact) = \pi_{AN}(Collected)$.
  The distinct ANs in relation Artifact are equivalent to the distinct ANs in relation Collected. This is because all ANs in either Artifact or Collected should match.

- $\pi_{SID}(Collection) \subseteq \pi_{SID}(Staff)$.
  For any collector SID in relation Collection must be under the SID of relation Staff. This is because all collectors are staffs and therefore should be under relation Staff.

- $\pi_{SID}(Artifact) \subseteq \pi_{SID}(Staff)$.
  For any maintainer SID in relation Collection must be under the SID of relation Staff. This is because all maintainers are staffs and therefore should be under relation Staff.

- $\pi_{type}(Artifact) \subseteq \{'tissue','image','model','live'\}$
  The types of artifacts can only be 'tissue', 'image', 'model', or 'live'. This is because these are the pre-defined artifacts types.

- $\pi_{AN}(Published) \subseteq \pi_{AN}(Artifact)$
  For any published journal, the artifact it referenced to must be under relation Artifacts. This is because academic journal needs to refer to an artifact that is documented in the database.

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.

   - SID of staff who held current rank for longest.
     $LongestStaff := \pi_{SID}(Staff) - \pi_{s1.SID}\Big(\sigma_{s1.date>s2.date}(P_{s1}Staff \times P_{s2}Staff)\Big)$
   - CID of collection with most recent collection date.
     $recentCollect := \pi_{CID}(Collection) -$
     $\pi_{c1.CID}\Big(\sigma_{c1.date<c2.date}(P_{c1}Collection \times P_{c2}Collection)\Big)$
   - Final solution
     $\pi_{date}\Big((sIDLongestStaff \bowtie Collection) \bowtie recentCollect\Big)$

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.

   Cannot be expressed.

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.

   - All artifacts with its corresponding collection and its collector.
     $AllCollectionArtifacts := \Big(P_{C(CID,date,cSID)}Collection \bowtie Collected\Big)$
     $\bowtie P_{A(AN,species,type,location,mSID)}Artifacts$
   - Final Solution
     $\pi_{A.AN}\Big(\sigma_{cSID=mSID}(AllCollectionArtifacts)\Big)$

4. Rationale: Identify multi-talented field workers.

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

   Cannot be expressed.

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

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

   - All Journals of that have used at least two different artifacts.
     $AtLeastTwice(Journal) := \pi_{p1.Journal}\Big(\sigma_{p1.AN\neq p2.AN\wedge p1.Journal=p2.Journal}$
     $(P_{p1}Published \times P_{p2}Published)\Big)$
   - All Journals of that have used at least three different artifacts.
     $AtLeastThrice(Journal) := \pi_{p1.Journal}$
     $\sigma_{p1.AN\neq p2.AN\wedge p2.AN\neq p3.AN\wedge p1.AN\neq p3.AN\wedge p1.Journal=p2.Journal=p3.Journal}$
     $(P_{p1}Published \times P_{p2}Published \times P_{p3}Published)$
   - Final Answer
     $ExactlyTwice(Journal) := AtLeastTwice - AtLeastThrice$

6. Rationale: Identify motherlode locations.

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

   Cannot be expressed.

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.

   - Aggregate all information about a given artifact
     $AllItems := \Big(P_{A(AN,species,type,location,ASID)}Artifact \bowtie Collected\Big)$
     $\bowtie P_{C(CID,date,CSID)}Collection$
   - Get all the Collector's SID who have collected tissue samples.
     $TissueAndOther(SID) := \pi_{C.SID}\Big(\sigma_{type='Tissue'}(AllItems)\Big)$
   - Get all the Collector's SID who have collected NON-tissue samples.
     $Other(SID) := \pi_{C.SID}\Big(\sigma_{type\neq'Tissue'}(AllItems)\Big)$
   - Final Answer
     $OnlyTissueStaff(SID) := TissueAndOther - Other$

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.

   - All of collection with its corresponding artifacts
     $AllItems := \Big(\pi_{CID}(Collection) \bowtie Collected\Big) \bowtie Artifacts$
   - Pairs of staff members who have worked with on at least one collection together
     $PairsOfTwo := \pi_{a1.SID,a2.SID}\Big(\sigma_{a1.CID=a2.CID\wedge a1.SID<a2.SID}(P_{a1}AllItems\times P_{a2}AllItems)\Big)$
   - Trios of staff members who have worked with on at least one collection together
     $PairsOfThree := \sigma_{a1.CID=a2.CID\wedge a2.CID=a3.CID\wedge a1.SID<a2.SID\wedge a2.SID<a3.SID}$
     $(P_{a1}AllItems \times P_{a2}AllItems \times P_{a3}AllItems)$
   - Final Answer
     $PairsOfTwo - \pi_{a1.SID,a2.SID}(PairsOfThree)$
     $- \pi_{a1.SID,a3.SID}(PairsOfThree) - \pi_{a2.SID,a3.SID}(PairsOfThree)$

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.

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.

   - $\sigma_{g1.genus=g2.genus \wedge g1.family \neq g2.family}(P_{g1}Genus \times P_{g2}Genus)\big) = \emptyset$

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

   - $ArtifactDate(AN, date) := \pi_{AN,date}(Artifact \bowtie Collected \bowtie Collection)$
   - $JournalDate(AN, date) := \pi_{AN,date}(Published)$
   - $\sigma_{JD.AN=AD.AN \wedge JD.jdate > AD.adate}$
     $(P_{AD(AN,adate)}ArtifactDate \times P_{JD(AN,jdate)}JournalDate)\big) = \emptyset$

4. Students may not catalogue live artifacts.

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