

Assignment 1

Due: Thursday, 9 October, at 11:00 pm sharp!

About the domain

For this assignment, you will write queries and integrity constraints on a database for a museum.

The objects in a museum can be as small as an individual button or as large as a tank. Each one is identified by a unique “catalogue number”.

Donors make donations to the museum. A donation can consist of a single object or multiple objects. The individual objects are given unique “catalogue numbers” when they are entered into the database. This process is called “cataloguing”. Because a donation can have many items, the items are often catalogued on different days and by different people.

One step in cataloguing an object is to assign it to a category. In order for museums to share information about their collection, a standard classification system is used. In Canada and the United States, the standard is the “Chenhall” system, devised by Robert G. Chenhall. Like biological classification, the Chenhall system is based on a hierarchy. In this assignment, we will simplify the Chenhall system by limiting it to a three-level hierarchy: The most general level will be called “category”, the next will be “primary term”, and the most specific category will be called “secondary term”. In our database, when an item is catalogued, it is assigned a secondary term. For example, a man’s dress shoe would be categorized with secondary term “footware”, which falls under primary term “clothing”, which falls under category “personal artifacts”. To solve the problems in this assignment, you do not need to know any of the specific names for categories and terms in the Chenhall hierarchy. You just need to know that we will use the three levels called category, secondary term and primary term.

In addition to classifying objects, our database can record tags associated with an object. These are useful when searching in the museum’s database.

Schema

Relations

- Object(CN, date, name, description, type, length, width, height, who)

A tuple in this relation represents an object in the museum’s collection. *CN* is the object’s catalogue number, *date* is the date on which it was catalogued, *name* is a short name for the object, *description* is a paragraph-length description of the object, *type* is the Chenhall secondary term for the object, *length*, *width* and *height* give the object’s dimensions, and *who* is the ID of the staff member who catalogued the object.

- Tag(CN, phrase)

A tuple in this relation represents the fact that object *CN* has tag *phrase*.

- Donor(DID, surname, firstname, address, email)

A tuple in this relation represents a person who has donated to the museum’s collection. *DID* is the donor ID, *surname* and *firstname* are the donor’s names, and *address* and *email* are the donor’s surface mail and email addresses.

- Donation(NID, date, DID)
A tuple in this relation represents a donation of one or more objects to the museum's collection. *NID* is the donation ID, *date* is the date on which the donation was made, and *DID* identifies the donor who made the donation.
- Contains(NID, CN)
A tuple in this relation represents the fact that donation *NID* contains object *CN*.
- Staff(SID, surname, firstname, address, email, type, date)
A tuple in this relation represents a person on the museum's staff. *SID* is the staff member's ID, *surname* and *firstname* are the staff member's names, *address* and *email* are the staff member's surface mail and email addresses, *type* is the type of staff member he or she is, and date *date* is the start date of his or her position at the museum.
- Chenhall(category)
A tuple in this relation represents the fact that *category* is a category in the Chenhall classification system.
- PrimaryTerm(primary, category)
A tuple in this relation represents the fact that *primary* is a primary term within category *category* in the Chenhall classification system.
- SecondaryTerm(secondary, primary)
A tuple in this relation represents the fact that *secondary* is a secondary term under primary term *primary* in the Chenhall classification system.

Integrity constraints

- Object[type] \subseteq SecondaryTerm[secondary]
- Object[who] \subseteq Staff[SID]
- Tag[CN] \subseteq Object[CN]
- Donation[DID] \subseteq Donor[DID]
- Contains[NID] \subseteq Donation[NID]
- Contains[CN] \subseteq Object[CN]
- $\Pi_{\text{type}} \text{Staff} \subseteq \{\text{"permanent"}, \text{"temp"}, \text{"volunteer"}, \text{"intern"}\}$
- PrimaryTerm[category] \subseteq Chenhall[category]
- SecondaryTerm[primary] \subseteq PrimaryTerm[primary]

Part 1: Queries

Write the queries below in relational algebra. There are a number of variations on relational algebra, and different notations for the operations. You must use the same notation as we have used in class and on the slides. You may use assignment, and the operators we have used in class: $\Pi, \sigma, \bowtie, \bowtie_{\text{condition}}, \times, \cap, \cup, -, \rho$. Assume that all relations are sets (not bags), as we have done in class.

Additional points to keep in mind:

- Do not make any assumptions about the data that are not enforced by the original constraints given in the schema above. Your queries should work for any database that satisfies those constraints.
- Assume that every tuple has a value for every attribute. (For those of you who know some SQL, in other words, there are no null values.)
- Remember that the condition on a select operation may only examine the values of the attributes in one tuple (not whole columns), and that it can use comparison operators (such as \leq and \neq) and boolean operators (\vee , \wedge and \neg). Two relations in our schema have a date attribute. You may use comparison operators on such values. You may refer to the year component of a date d using the notation $d.\text{year}$.
- You are encouraged to use assignment to define intermediate results, and it's a good idea to add commentary explaining what you're doing. This way, even if your final answer is not completely correct, you may receive part marks.
- The order of the columns in the result doesn't matter.
- When asked for a maximum or minimum, if there are ties, report all of them.

Write queries for each of the following:

1. Find the tallest item(s) that is tagged with "Fenian Raids", "weapon", and "British". Report the catalogue number, donor's full name, and the donation date.
2. Let's say a small donation is one that included no more than two items. Find all donors who have made 3 or more small donations. (They may have made other donations that were larger.) Report the donor ID and the date of their first and last small donations.
3. Find all permanent staff who (a) have started on or after July 1st 2014, or (b) have catalogued items from at most 2 different donations. Report the staff ID and email address.
4. Let's say a staff member is "experienced" if they have catalogued something in every Chenhall category. Find all the experienced staff members. Produce a relation with attributes called "sID" and "primary". For each of the experienced staff members, include a row for every primary term under which they've catalogued an item.
5. Find all donations for which every item was catalogued by the same staff member and that staff member was of type "temp". Report the donation ID, as well as the staff ID, surname, and firstname of the staff person who catalogued it.
6. Find all objects that have no tags. Report the catalogue number, donor email, and the object name, as well as it's Chenhall secondary term, primary term and category.
7. Let's say that two people have catalogued together if they have each, on the same day, catalogued items that come from the same donation. Find all volunteers who have been on staff since 2000 or earlier, and have catalogued together with every permanent staff member. Report the volunteer name, address and start date.
8. Find all pairs of permanent staff members who have catalogued items from the same donation as each other, but never on the same day. Report their staff IDs, full names, and email addresses.
9. Find all donors who have, at least twice, made more than one donation in a year, but haven't donated anything since 2009. Report the donor's full name and address, and the date of their most recent donation.

10. Find all donors who have made only one donation, and it was before 2008, and their donation included some art: i.e., some item whose secondary term, primary term, or category included the substring “art”. Report their donor ID and every Chenhall category in which they’ve donated an item. Put the information into a relation with attributes called “DID” and “category”.

You may use the notation $\text{substring}(x, \text{“art”})$ within a selection condition to say that attribute x contains the substring “art”.

Part 2: Additional Integrity Constraints

Express the following integrity constraints with the notation $R = \emptyset$, where R is an expression of relational algebra.

You are welcome to define intermediate results with assignment and then use them in an integrity constraint.

1. Staff members of type “temp” cannot catalogue objects.
2. The first things volunteers are allowed to catalogue must be in Chenhall category “Personal Artifacts”. Once they have catalogued three such objects, they are allowed to also catalogue objects in Chenhall category “Furnishings”. They are never allowed to catalogue other sorts of items.
3. For donations made prior to July 1st, 2001, all items in the donation had to be catalogued by the same person.

Submission instructions

Your assignment must be typed; handwritten assignments will not be marked.

You may use any word-processing software you like. Many academics use LaTeX. It produces beautifully typeset text and handles mathematical notation well. Whatever you choose to use, you need to produce a final document in pdf format, and you must call it “a1.pdf” (with no capital letters).

You are allowed, and in fact encouraged, to work with a partner for this assignment. You must declare your team (whether it is a team of one or of two students) and hand in your work electronically using the MarkUs online system. Instructions for doing so are posted on the Assignments page of the course website. Well before the due date, you should declare your team and try submitting with MarkUs. You can submit an empty file as a placeholder, and then submit a new version of the file later (before the deadline, of course); look in the “Replace” column.

For this assignment, hand in just one file: a1.pdf. If you are working in a pair, only one of you should hand it in.

Once you have submitted, be sure to check that you have submitted the correct version; new or missing files will not be accepted after the due date.