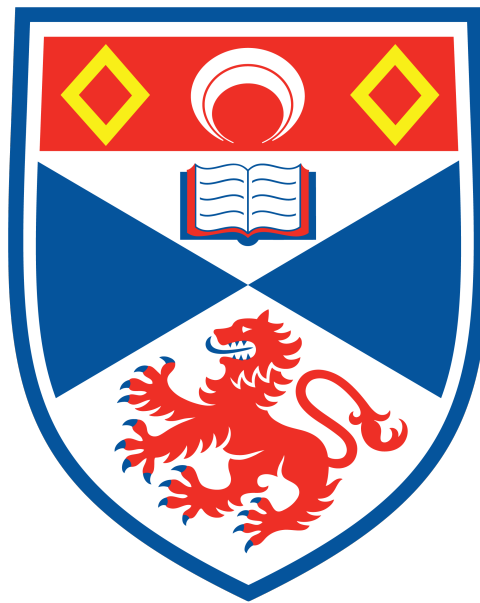


Report

Modelling Database Requirements with E-R Diagrams

210016568



UNIVERSITY OF ST ANDREWS

IS5102 - DATABASE MANAGEMENT SYSTEMS
ACADEMIC YEAR 2021/2, SEMESTER 1 - ASSESSMENT 1

Lecturer: Alexander Konovalov

Word Count: 1942 words

Due date: October 11, 2021

Abstract

This report documents the process in detail by which the problems given in IS5102 Database Management Systems assessment 1 are solved. It is divided into four parts, which are scenario analysis, specification of the data, entity-relationship diagram, and reflections.

I. SCENARIO ANALYSIS

This section will illustrate the existence of constraints and participation in the analysis process of the whole scenario. Also, the section will also explain in detail some attributes that are necessary for readers to fully understand the analysis process. After analysis, it will cover the potential questions from stakeholders.

i. entity set

i.1 organization entity set

Wildlife Protection Society can be regarded as an entity with a set of organization because it has several branches and a central office. It has two types of lower-level entity sets: branch and central office. For the higher entity set organization, an attribute with unique constraint named `organization_id` is the primary key. It is used to identify different branches or central office, whereby domain is integers starting from 0. To meet the needs for people to contact the different branches or central office of this organization, this entity set has an attribute called `contact_detail`, which includes `postal_address`, `phone_number` and `email_address`. Since different branches or central offices may have multiple phone numbers, `phone_number` is a multivalued composite attribute.

For the lower-level entity set branch, in addition to the attributes it inherited from higher-level entity set, it should have an attribute named `region` for assigning tasks within the organization according to the members' different addresses. The domain of `region` can be integer or abbreviation for avoiding spelling mistakes. Since central office and branches are two entity sets with different functions, disjoint constraint is applied between them.

i.2 member entity set

For the member of this society, it can be regarded as an entity set which can be divided into 2 lower-level entity sets: `registered_member` and `volunteer`. Some examples of the attributes in the member entity set include the following: `member_id` as primary key with unique constraint (domain is integers starting from 0), `name` (`first_name` and `last_name`), `gender`, contact details (`postal_address`, `phone_number` and `email_address`). Specifically, `phone_number` is a multivalued composite attribute, as discussed previously.

The entity set `registered_member` can be divided into two entity sets: `society_member` and `staff`. In addition to the attributes inherited from higher-level entity set, they also consist of attributes named `monthly_membership_fee` and `annual_salary`, respectively.

Between the `registered_member` and `volunteer`, maybe some society members prefer to do volunteer activities, so there should apply overlapping constraints.

For `society_member` entity set and `staff` entity set, a person cannot be an employee and a member at the same time, so disjoint constraint should be applied here.

i.3 observed_event entity set and observed_case weak entity set

In this section, the observation actions of members can be described as an entity set named `observed_event`. Here I use an attribute named `event_id` as the primary key and to apply

unique constraint. The domain of this attribute should be an integer starting from 0. The `interval_start_time` attribute is used to record the start time of the 15 minutes interval.

Because the observers may observe nothing or many species of butterfly, and the observation case has some attributes, I opted to use a weak entity set instead of multivalued composite attribute to illustrate the observation case of one butterfly species in a certain `observed_event` by a certain member. The special situation of this scenario needs to be illustrated. An observer finds the same butterfly species at different locations or multiple butterfly species at a single location, and this situation is treated as multiple cases. i.e., a case is an entity used to describe the number of a particular butterfly species at a fixed location. In this weak entity set named `observed_case`, the attribute `case_order` is the discriminator, it is imperative that this case happened in a certain observation event, i.e., it should be an integer starting from 0. The attributes `species`, `number`, `location` (including coordinates, postcode and county) record what kind of, how many, and where this butterfly species they found.

For this weak entity set, the strong entity set `observed_event` is its identifying set, the discriminator in weak entity set and the primary key in identifying entity set together form the primary key of this weak entity set.

i.4 butterfly entity set

This entity set is used to store the data from different kinds of butterfly. `Species` is the primary key applied unique constraint, the domain of it should be identified by a fixed and non-repeating number or abbreviation to prevent spelling errors. The attributes `english_name`, `latin_name`, `description`, and `photo_set` store the English name, Latin name, brief description, and photos. `Photo_set` should be a multivalued composite attribute because there may be more than one photo on the website.

ii. Relationship

Between entity set `organization` and `member`, the relationship is that an organization has some members. As one organization may have more than 1 member, but one member can only belong to 1 organization. Therefore, this relationship is one (organization) to many (member) in cardinality constraints. Every member must have an organization, and every organization must have at least one member to keeping branches and central office up and running, so both organization and member entity sets should be applied the total participation.

Between entity set `member` and `observed_event`, the relationship is that a member may do several `observed_events`. A member can do several `observed_events`, but an `observed_event` can only be done by one member. Thus, this relationship is one (member) to many (observed event) in cardinality constraints. Every `observed_event` must be done by a member, but not every member will join this activity. Thus, the member entity set should be applied partial participation, the `observed_event` should be applied total participation.

Between entity set `observed_event` and `observed_case`, `observed_event` is the identifying set of `observed_case`. The relationship is that in an `observed_event`, observer may find several cases. An `observed_event` can find several `observed_cases`, but an `observed_case` can only be found by one observed event. Thus, this relationship is one (observed event) to many

(observed_case) type in cardinality constraints. Every observed_case must be found by an observed_event, but not every observed_event will have a case of observation, maybe they cannot find any butterfly but just record a failure event. Thus, the observed_event entity set should be applied partial participation and the observed_case should be applied total participation.

Finally, all observed_cases are used to count the number of different species of butterfly. Species can be count by several cases, but a case can only contain the number of a certain species of butterfly observer counted. So, this relationship is one (butterfly) to many (observed_case) in cardinality constraints. Some already recorded butterflies may not be observed, and some observations may find some new species which is not in the list. So, two entity sets should apply partial participation.

iii. The potential questions from stakeholders

The first potential question may be that how to query the count of a certain butterfly species A. In this situation, user can select the table observed_case, show all the data of number when species is A, then add all numbers together.

How to deal with volunteers in two situation is also a possible question. The scenario requires that contact information be filled in only when volunteers submit their observations. Thus, the registration of volunteer does not need to collect contact details, until their first submission, they need to save their contact details into database.

Users may ask how to handle the discovery of new species. Since numbers are used to identify different species of butterflies in the database to avoid spelling errors, the option to submit new species should be added by asking the discoverer to submit a short description, photo, and other additional information, and then adding the new species to the butterfly table after review by the database administrator.

II. SPECIFICATION OF THE DATA

This section will illustrate the specification of data from the scenario.

The contact details of a certain organization or person will be stored in `contact_detail`, it contains `postal_address`, which is a string less than 80 lengths, `phone_number`, `phone_number` is a multivalued composite attribute, each value is a non-negative integer of 14 length (such as 0440XXYYYYZZZZ).

`Region` is used to help the system of regional branches dividing this society into different areas. It will be defined as a string no more than 5 lengths, as a kind of region or city abbreviation (e.g., London as LOND, St. Andrews as STA).

`Name` includes `first_name` and `last_name`, both two attribute are one string less than 10 length.

`Gender` should only have 4 types of value, male, female, unknown, and prefer not to say. It can be represented by integer from 0 to 3.

The `monthly_membership_fee` describes the fee membership paid every month, the currency unit is the pound sterling, domain should larger than 0, datatype is float.

The `annual_salary` is used to store data of staff's income every year. Its domain is same as `monthly_membership_fee`.

The `interval_start_time` records the start time of the 15 minutes interval, domain of it is a type of time format which can describe any time of the day, down to the seconds (HH:MM:SS).

The species of butterflies can be defined as a series of integer starting from 0, different number represent different species.

The coordinates are a pair of numbers including longitude and latitude, e.g. (X, Y).

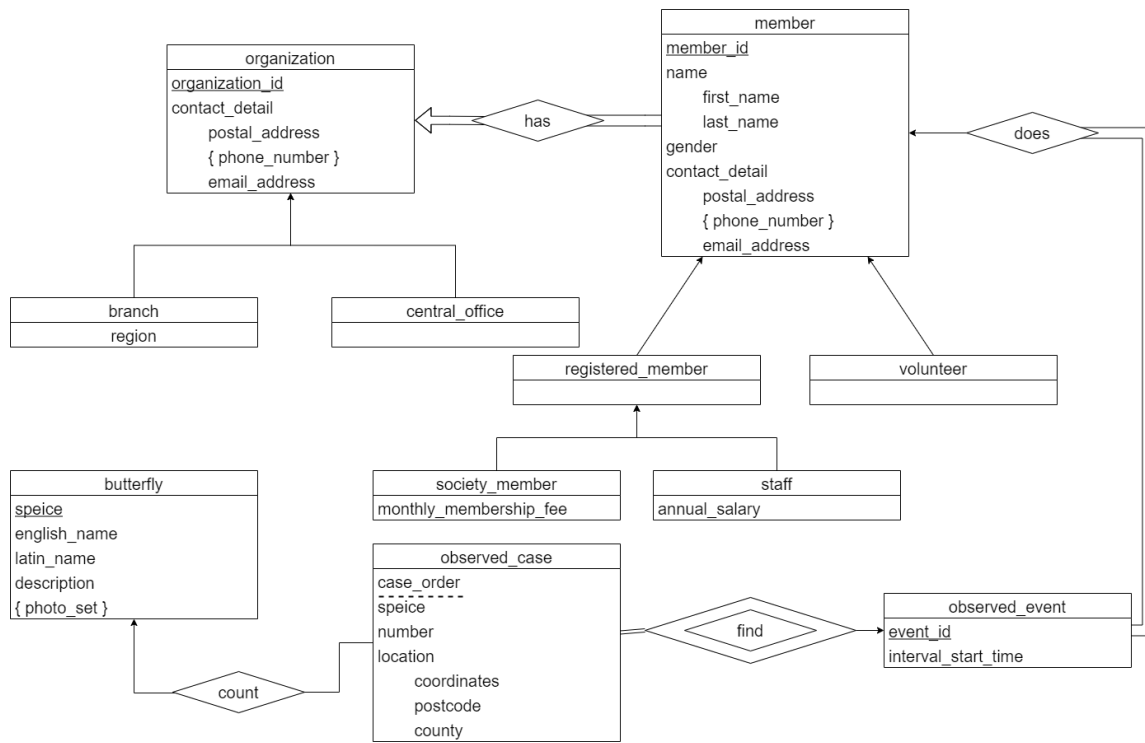
`County` should be defined as a string less than 10 length.

The English name and Latin name of butterfly are all strings less than 30 length.

The brief description of a certain butterfly species should be a string within 500 words.

The photo set of butterflies would be saved in a multivalued composite attribute, each value is a picture datatype.

III. ENTITY-RELATIONSHIP DIAGRAMS



IV. REFLECTION

In this assessment, I mainly encountered two main problems. The first is how to handle the observations of a single submission by a member. Since the same species may be found in multiple locations in a single observation, or multiple species may be found in a single location, the multivalued composite attribute does not satisfy all use cases. Initially, I did not consider splitting a single event into several cases and directly merged the events and cases in the current diagram together. However, then the species and location are still in the relationship of many-to-many, so it still did not fulfill the needs of the scenario. Finally, I repeatedly watched several recordings of the exercise class, replaced several entities in similar situations with entities in this assessment. Gradually I figured out that I should use the weak entity to describe the case of a single species being found in a certain location.

At the end of the drawing, I wasn't quite certain which entity set the butterfly entity set should have a relationship with. Initially, I figured the relationship should be "organization has butterfly list" and "butterfly list contains butterfly", but then there is no way for the number of butterflies to be recorded as an attribute to meet the need to display on the website. As a result, I tried to treat count as a verb rather than a noun, i.e., each case contains only one butterfly species that can be used as part of the total number of butterflies of that species. Thus, the relationship between the butterfly entity set and other entity sets should be "observed_case count butterfly".

However, in general, the whole process of finishing this assessment has been relatively smooth, except for the two problems mentioned above. The establishment of other entity sets and relationships are completed quite punctually, among which I think I applied overlapping constraint and disjoint constraint in dividing the member entity set into several lower entity sets, and completed the classification of the members of the animal protection association very accurately.