CS3101 - Databases
Assignment: P1 - DB Design

Matriculation Number: 160001362

29/10/2018

Contents

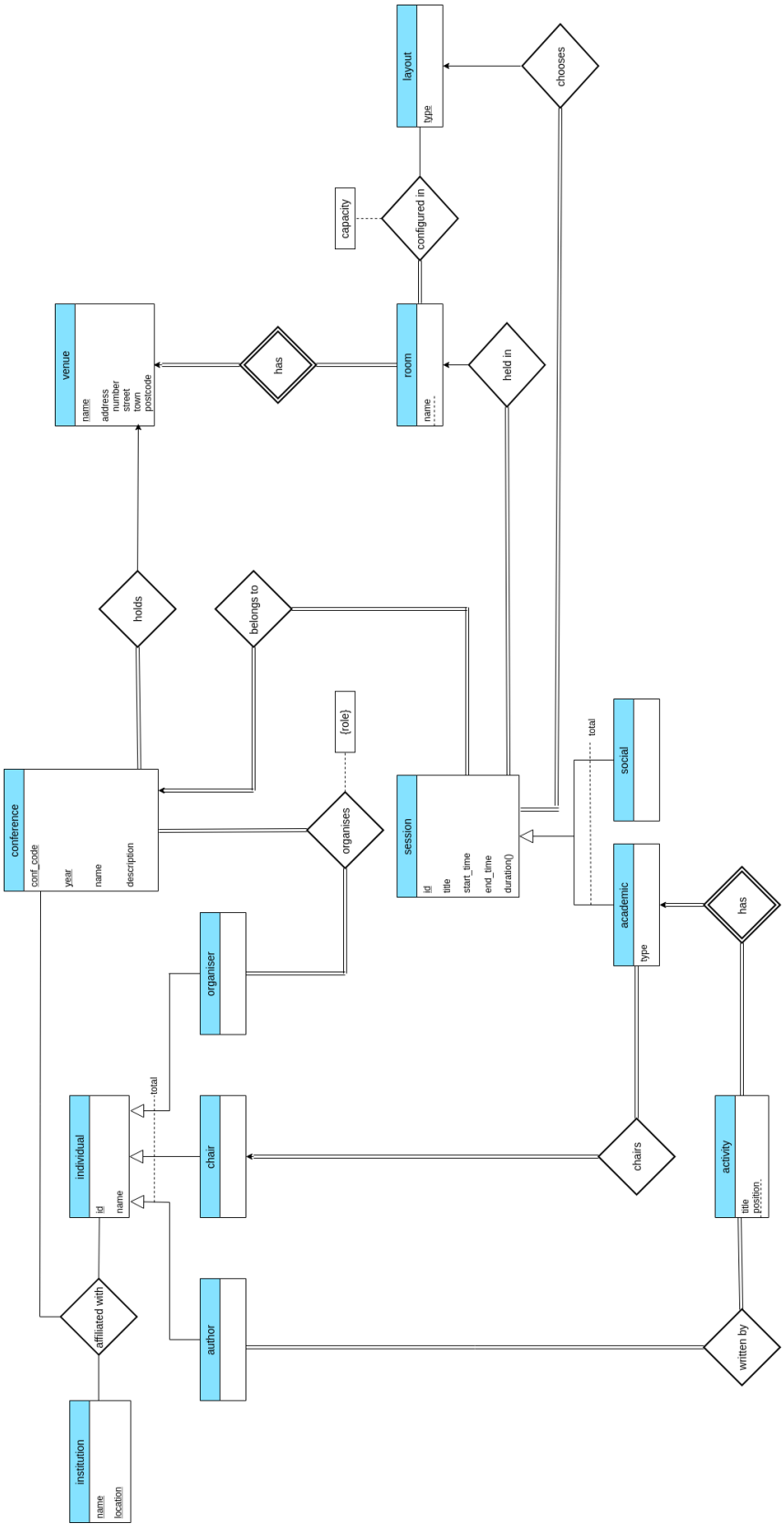

Figure 1: E-R diagram modelling ConMan conference management system.

1 E-R Model

1.1 E-R Diagram

See figure 1 for the entity-relationship diagram modelling the conference management scenario.

1.2 Underlying Assumptions

- The assumption has been made that all layouts can be utilised by all rooms (with varying capacities).
- The assumption has been made that an individual can have a multiple affiliations at a time, but that the institutions also share an affiliation with the conference the individual attends so that if the individual changes institutions, the old affiliation will still be reflected in the 'affiliated with' relationship.
- The assumption has been made that 'chair' and 'author' are distinct from the roles carried out by an organiser.
- The assumption has been made that a session is held in a single room and that a room can only hold one session at once.

1.3 Justification of Design Decisions

1.3.1 Primary Keys

- The candidate key 'conf_code' and 'year' was selected as the primary key for conference due to conference codes being unique year to year. The conf_code attribute allows any conference to be unique within a year, and the year allows it to be identifiable across any year.
- The candidate key 'name' was selected as the primary key for the venue entity as the name is unique according to the specification.
- The entity 'room' is a weak entity set. The attribute 'name' has been taken to be the discriminator (or partial key). The primary key of room is the discriminator 'name' combined with the name of the venue which is the primary key of the venue entity. This decision was made as the specification states that a room has a unique name in the context of its venue.
- The 'type' attribute of layout has been made the primary key of the entity, as it seems unrealistic to have two layouts with the same type.
- The candidate key of 'institution', 'name' and 'location', has been made the primary key of the entity as presumably there can be no institution with the same name in the same location. For example two University

of St Andrews in the United Kingdom, however there is a St Andrews University in North Carolina. Location rather than country allows for the specification of US states and other territories which people often use as descriptors.

- An id attribute has been added to the 'session' entity as the primary key to avoid having an additional weak entity set whose relations would have to reference many attributes.
- The entity 'activity' is a weak entity set. The discriminator 'position' has been selected as the partial key as position is unique in the context of an academic session.

1.3.2 Types of Attributes

- The discriminator 'name' has been chosen in the 'room' entity as each room has a unique name in the context of the venue.
- A composite attribute has been used for the address of the 'venue' entity, as address can compose values such as number, street, town, and postcode.
- The 'role' attribute of the relationship 'organises' has been made a multivalued attribute as an organiser can have multiple roles within a single conference.
- The 'duration' attribute of 'session' is an attribute derived from start_time and end_time.

1.3.3 Participation and Cardinality

- The relationship 'holds' between venue and conference is one-to-many from venue to conference. This is because a single venue may hold many conferences, but a conference must be held at a single venue.
- The relationship 'has' between venue and rooms is one-to-many from venue to room, as a venue must have one or more rooms but a room must belong to one venue.
- The relationship 'held in' between 'session' and 'room' is many-to-one from session to room as many sessions can use a single room throughout the day but a room does not have to host a session. A session must be held in a room.
- The relationship 'configured in' between 'room' and 'layout' is many-to-many as a room can be configured into many layouts and a layout can be used by many rooms. A room must be configured into a layout but a layout does not have to be used by a room.

- The relationship 'chooses' between 'session' and 'layout' is many-to-one from session to layout as many sessions must choose a single layout but a layout does not have to be chosen by a session.

- The relationship 'organises' between the 'organiser' and 'conference' entities is many-to-many as an organiser must organise one or more conferences (as the individual would otherwise not be an instance of organiser) and a conference must be organised by at least one organiser.

- The relationship 'affiliated with' is a ternary relationship between institution, individual and conference, and in all cases is many-to-many, as an individual can optionally have many affiliations with institutions and can attend many conferences, a conference can be affiliated with many institutions via the individuals that attend, and institutions can be affiliated with many individuals and many conferences via its attending staff.

- The relationship 'belongs to' between conference and session is one-to-many from conference to session. This is because a unique id is assigned to session so it must belong to a single conference, but a conference must be made up of a series of sessions.

- The 'chairs' relationship between 'chair' and 'academic' (session) is a one-to-many relationship. An academic session must have exactly one chair leading it, but a chair may lead many academic sessions (and must lead at least one, otherwise they would not be a chair).

- The relationship 'has' between 'academic' and 'activity' is one-to-many as each academic session must be associated with one or more activities, and an activity must be associated with one session.

- The relationship 'written by' between 'activity' and 'author' is many-to-many. An author must have written one or more activities (otherwise they wouldn't be made an author) and an activity must have one or more authors.

1.3.4 Modelling Relationships (e.g. entity set vs relationship set vs an attribute)

- The role of an organiser has been made an attribute of the relationship 'organises' between the organiser and the conference as it is related to both entities.

- The ternary relationship 'affiliated with' has been used to model the relationship between institutions, individuals, and conferences, in order to allow an individual to be affiliated with many institutions at once and across time while enabling the conference to still be associated with the individual and their old affiliation.

1.3.5 Placement of Attributes

- The attribute 'capacity' has been placed on the relationship 'configured in' as the capacity of a room relies on the layout that it utilises.
- The multi-valued attribute 'role' has been placed on the relationship 'organises' as roles relate both to the organiser who will undertake them but also to the conference which creates the need for the role and which the role also enables.
- The attribute 'capacity' has been placed on the relationship set 'configured in' between 'room' and 'layout' as it relies on both entity sets.

1.3.6 Specialisation/Generalisation

- The entities chair, organiser, and author are all specialisations of the generalised individual entity representing people involved in the conference. The completeness constraint for the inheritance is total, as all of the individuals modelled in the conference data in the specification are either chairs, authors, or organisers. Therefore there is no need to allow individuals to be instantiated. The inheritance is overlapping as entities can belong to more than one subclass entity set - this is modelled using multiple arrows.
- The entities 'academic' and 'social' are specialisations of the entity 'session' and have disjointed inheritance as a session can be either academic or social. The completeness constraint of the inheritance is total, as all sessions are academic or social sessions. This inheritance makes use of one arrow which splits off to academic and social as a result of the disjointed inheritance.

2 Relational Schema

2.1 Derived Relational Schema

The following relational schema differs slightly from the E-R diagram. This is because the E-R diagram was an attempt to clearly capture the scenario laid out in the specification. However the relational schema aims to more closely match the low level implementation of the database, so some minor changes have been made to reflect this.

Relational Schema		
Schema	Attributes	Types
conference	conf_code, year, name, description, venue_name*	integer, integer, string, string, string
venue	venue_name, building_number, street, town, postcode	string, integer, string, string, string
room	venue_name*, room_name	string, string
configured in	venue_name*, room_name*, layout_type*, capacity	string, string, string, integer
session	session_id, title, start_time, end_time, duration, venue_name*, room_name*, layout_type*, conf_code*, year*	integer, string, date/time, date/time, date/time, string, string, string, integer, integer
academic_session	session_id*, academic_session_type, individual_id*	integer, string, integer
layout	layout_type	string
individual	individual_id, name	integer, string
institution	inst_name, inst_location	string, string
affiliated_with	inst_name*,inst_location*, conf_code*, year*, individual_id*	string, string, integer, integer, integer
organises	individual_id*, conf_code*, year*, role	integer, integer, integer, string
written_by	position*, session_id*, individual_id*	integer, integer, integer
activity	session_id*, position, title	integer, integer, string

2.2 Justification of Design Decisions

2.2.1 Type of Attributes

- The composite attribute of 'address' in the 'venue' entity has been flattened down into the attributes 'building_number', 'street', 'town', and 'postcode'.
- The multivalued attribute 'role' has been made the discriminator (partial key) of the schema representing the 'organises' relationship set. The organises schema contains identifiers for individuals and conferences, and 'role' acts as an identifier for the specific role within a conference which the individual has taken on.
- The values in the relational schema have been assigned one of the following types - integer, string, or date/time:
  - Where the attributes are numeric such as 'size' or 'building_number' the type 'integer' has been assigned. The 'integer' type has also been assigned for codes such as 'conf_code' and 'session_id'. This assumes that the codes are numeric but equally these could be represented as strings if they can contain alphabetic characters.
  - Values that are text such as 'title', 'inst_name' or 'role' have been assigned the type 'string'.
  - Values that represent some form of date and/or time have been given the value date/time. These values include start_time, end_time and duration. The year is a single value so this has not been given the type date/time.

2.2.2 Placement of Foreign Keys

The following rules have been used to decide where foreign keys should be placed:

- Many-to-many relationship sets - Where entity sets share a many-to-many relationship, a schema is created for the relationship between the entities, containing foreign keys to each entity and any attributes that are associated solely with the relationship.
- Many-to-one/one-to-many relationship sets - Where entity sets share a many-to-one or one-to-many relationship, the foreign key of the 'one' schema is placed in the 'many' schema. For example 'individual_id' is placed into the 'academic_session' schema as a foreign key as a single individual chairs an academic session.
- One-to-one relationship sets - There are no one-to-one relationship sets in this representation of the scenario, but if there were, the foreign key could be placed in the schema representing either of the two entity sets.
- Weak Relationship Sets - For all relationship sets where one entity is a weak entity set, the foreign key of the identifying entity set is placed in the schema of the weak entity set. For example to identify a room in the context of a venue, the venue name must be in the schema of room as a foreign key.

2.2.3 Translation of Specialised/Generalised Entities

- Individual - In the E-R diagram, the specialisations of individual (author, chair, and organiser) were specified in order to more clearly model the scenario. However when modelling the schema, it is clear that the only necessary schema is that of individual, as the specialisations have no additional attributes and have overlapping inheritance. Not having the total inheritance constraint be enforced is not ideal, however it is a worthwhile trade off in this scenario to avoid having data duplication of individuals who fulfill more than one specialisation of the entity set 'individual'.
- Session - In the E-R diagram, the distinction between academic sessions and social sessions is explicit. This clearly models the scenario, but is not required for the relational schema or implementation. As social sessions do not have any additional attributes compared to the 'session' generalisation, it can simply be assumed that any session that does not have the 'type' attribute of an academic session is a social session, removing the need for disjointed inheritance. Therefore in the schema, the inheritance constraint of the session is actually partial instead of total, and the two required schema are session and academic_session.

2.2.4 Comparison With E-R Diagram

The E-R diagram for the most part has been directly translated into the relational schema. The only major departures, as discussed previously, are the changes made to the inheritance associated with the 'individual' and 'session' entity sets.

3 Normalisation

3.1 Design Methodology

REFLECT ON APPROACH DISCUSSED HERE: https://studres.cs.st-andrews.ac.uk/CS3101/Lectures/L4_E

3.2 Functional Dependencies

3.3 Discussion of Normal Form

4 Extensions

4.1 Analysis