Lecture Notes # 2

It is the aim to develop advanced skills in database design (conceptual view): analysis of requirements and draw of the ER diagram. [1].

Design Principles

- The design should be **faithful** to the specifications of the application, that is entity sets and their attributes should reflect reality, further whatever relationships are asserted should make sense given what we know about the part of the real world being modeled
- The design should **avoid redundancy**, that is specify everything only once
- The design should avoid introducing more elements than is **absolutely necessary**
- The design should choose only the **right relationships**, because adding every possibile relationship is not often a good idea
- The design should consider just the **right kind of elements** to represent a real-world concept

[1] The script is mainly based on "A First Course in Database Systems", J. Ullman, J. Widom. Mostly examples and exercises are created by the author of this script.

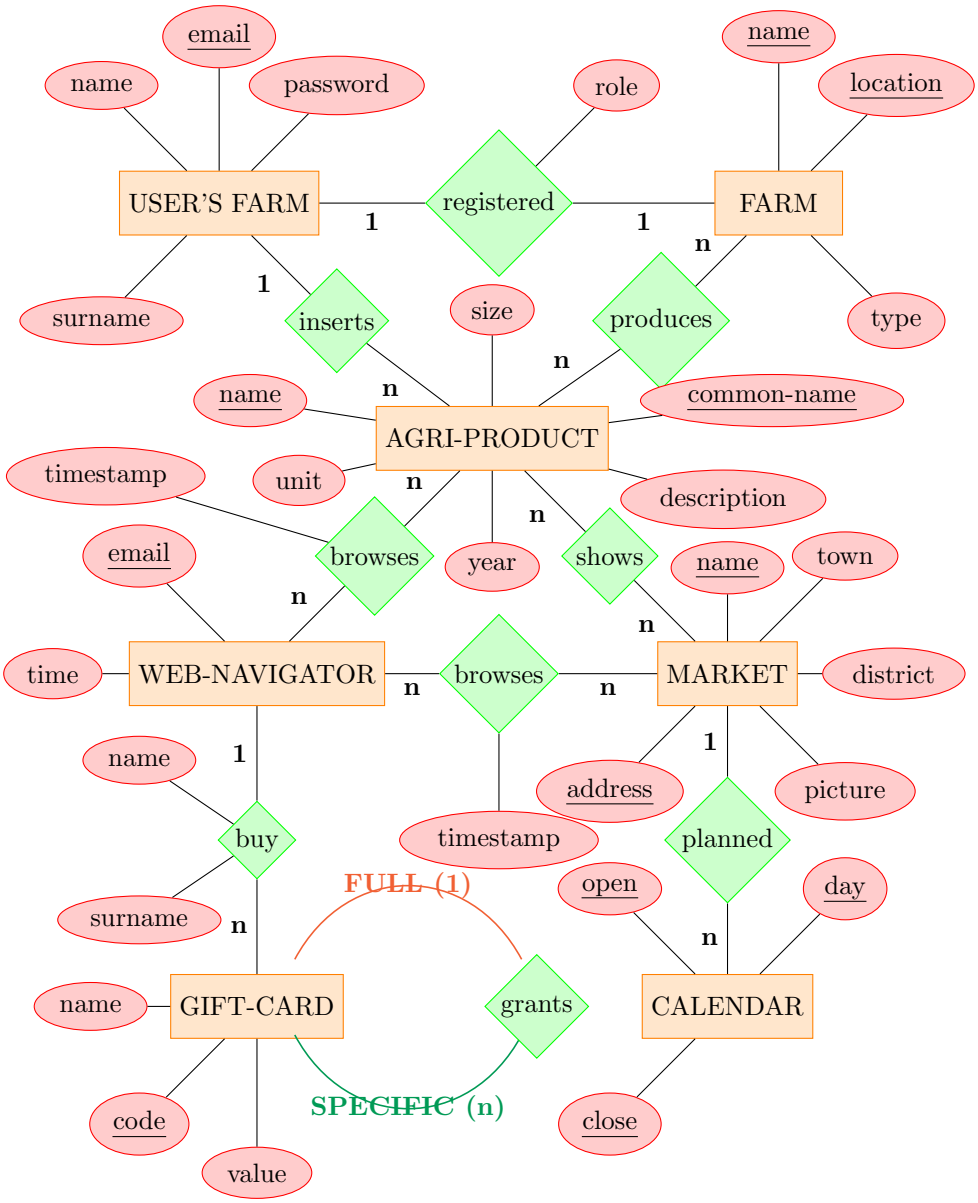
Database Design: From Requirements to E/R Diagram

Ex. # 1

"*Farmlands* is an initiative of associated farmers who periodically show their products in some markets at specified dates. Information about 'what, when and where' you can buy can be browsed by means of an 'app'. This year Web users can buy gift cards, accepted in all markets". The goal is to design a database to support the 'app' which promotes farmer products, keep tracks of user browsing, and sell gift cards.

Draw the E/R diagram that capture the requirements stated below. **Use "ID" as key only if strictly necessary**.

1. A **user's farm** has to register on the 'app', his name, surname, email and a password.
2. Only a user's farm can register the **farm**, entering the farm name, location, type (breeding, vineyard, fruit and vegetable, ...). User's farm has to specify his role in the farm (owner, administrative, ...).
3. The 'app' is the repository of different types of **agri-products**, for each one are stored, name, common name, size, unit, year of production, description of the production process, who has inserted the product. User's farm can add more agri-products for products of his farm not yet registered in the database.
4. The 'app' holds data about **markets**, a place where farmer can show their products for sale. Each market has name, town/village, district, address, picture.
5. Farmer can show food products at multiple markets, according to the **calendar** defined for each market.
6. People interested in agri-food markets information (what/where/when), is registered as **web navigator**. Since he/she can login through Facebook/Google, the email is registered, other than time (date + hour). After that he/she can browse the information. Navigator's browsing is tracked.
7. Currently **gift cards** are available. Gift cards are identified by fixed length code, and have a value (amount of money to spend). The name identifies the type of card, for example: junior, senior, Xmas card which grants special discounts, free entrance ..... **There exist cards which could have all grants**.
8. When a navigator selects a gift card and clicks on the button 'buy' he has to enter a person name and surname, who becomes the owner of the card. After the payment, by external bank services, the navigator receives a pdf file.


email
name
password
role
name
location
USER'S FARM
registered
FARM
1
1
n
1
surname
inserts
size
produces
type
n
n
name
common-name
AGRI-PRODUCT
timestamp
unit
n
description
n
email
browses
year
shows
name
town
n
n
time
WEB-NAVIGATOR
n
browses
n
MARKET
district
name
1
1
address
picture
buy
timestamp
planned
FULL (1)
open
day
surname
n
n
name
GIFT-CARD
grants
CALENDAR
SPECIFIC (n)
code
close
value

Ex. # 2

"Scientific association - society - (i.e. Società Italiana di Statistica) offers a variety of web tools to members aiming to promote (publish) *member scientific papers* and share these papers with fellow members of the society. Some limitations, depending on the membership subscription could be applied."

Draw the E/R diagram that capture the requirements stated below.

1. A member of the society is automatically a **user** of the web tools. He/she is identified by *name.surname*. Basically the starting and expiring membership dates are registered.

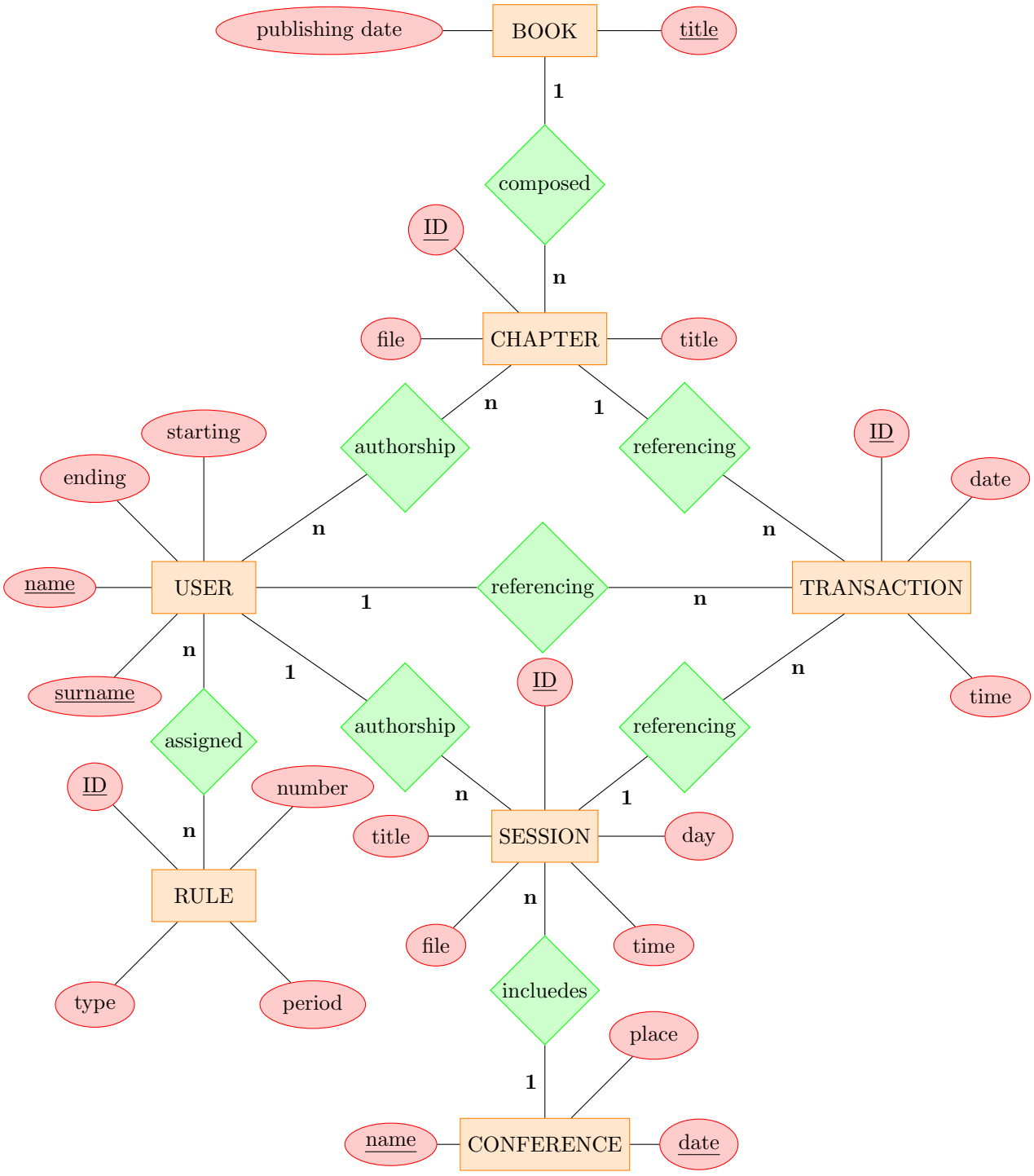
2. Scientific papers are classified in books and conference presentations:

   (a) A **book** has a title and a publishing date. A book is a collection of many papers (for instance grouped by theme) that becomes **chapters** of the book. A chapter has a title. Chapter authors are/were only members of the society, therefore *users could have the authorship relation*.

   (b) A **conference** has a name and a date/place where it was held. A conference has many **sessions** with presentations (typically slides). A session has a title, day and time. Session speakers are/were only members of the society, therefore *users could have the authorship relation*.

3. Users can access the society repository aiming to download scientific material: chapter (paper) or session presentation (slide). Some **rules** are assigned to each user that are the type, the number of items and period (for example, 3 papers daily) he/she is able to download.

4. Whenever a user download scientific material a **transaction** is registered. Each transaction is identified by date and time and further it is related to the user who downloaded and what he/she download, that is a chapter (paper) or a presentation session (slide).


publishing date
BOOK
title
1
composed
n
ID
file
CHAPTER
title
n
1
authorship
referencing
starting
ending
ID
date
n
n
name
USER
1
referencing
n
TRANSACTION
n
1
n
surname
assigned
authorship
ID
referencing
time
ID
number
n
1
title
SESSION
day
n
RULE
file
time
type
period
includes
place
1
name
CONFERENCE
date

Ex. # 3

"Just one click and you can locate and book the best cakes of the tradition, you identify your best favorite cake (panettone, torrone, pastiera, ...), locate one bakery close to you, check the availability, book the cake and finally get and taste it". The goal is to design a database to support the 'Web application' which allows user to "locate and book" the best favorite traditional cake.

Draw the E/R diagram that capture the requirements stated below.

1. **Traditional cakes** are stored with a common name, an additional name, the region of reference, year of the first preparation (when it is available), a picture.

2. Typically, many "variants" are available. Along time bakers have been created "modern" cakes, maybe just adding or replacing ingredients. Therefore these "variants" should be recorded using an additional name. In case of the traditional cake additional name takes "classic" as value.

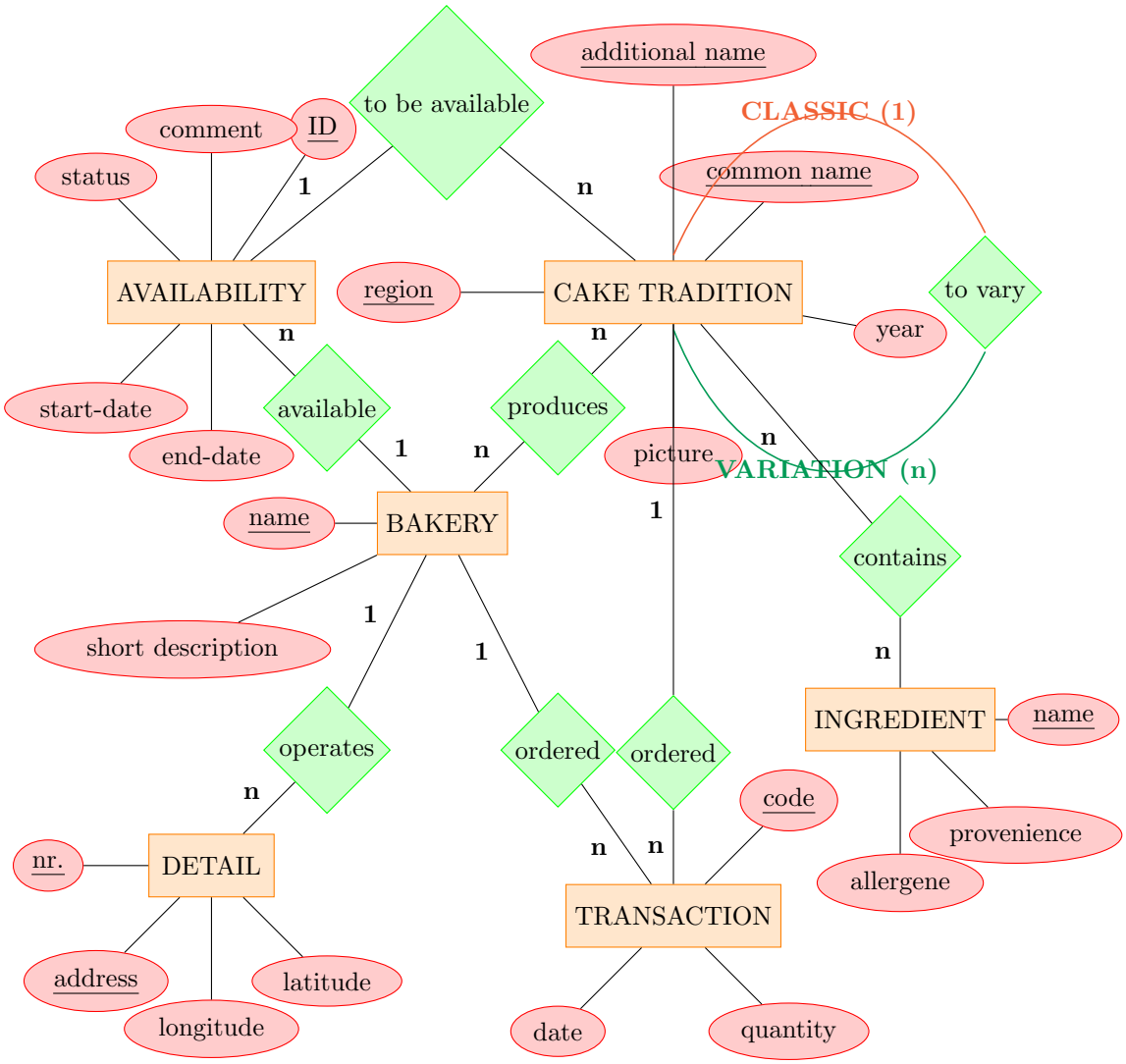
3. For each cake, and its variants, should be displayed the main **ingredients**, so the name, provenience, and if it could contain allergene (Yes or No).

4. **Bakeries** which offer cakes of the tradition are registered: name and short description.

5. A bakery could operate on the region with different shops, therefore the following **details** should be registered aiming to show to the user the best close shop, that is where to take the booked cake. Specifically details are address, geographical coordinates (longitude, latitude) in order to compute the distance from the user.

6. A bakery generally is not able to offer all cakes along the year, since for instance cakes are related to seasonality (panettone, torrone) or seasonal ingredients. Availability of cakes offered by each bakery must be registered in order to allow and accept bookings. **Availability** is described by a period: start date and end date, a comment, a status (True or False whenever this availability is confirmed or not).

7. Whenever a user locate a bakery, for its best favorite cakes and the user decides to order it a **transaction** is registered. The code of the transaction is then notified to the user.


additional name
to be available
comment
ID
CLASSIC (1)
status
common name
1
n
AVAILABILITY
region
CAKE TRADITION
to vary
year
n
n
start-date
available
produces
picture
n
end-date
1
n
VARIATION (n)
name
BAKERY
1
contains
1
short description
1
n
operates
ordered
ordered
INGREDIENT
name
n
code
provenience
nr.
DETAIL
n
n
allergene
TRANSACTION
address
latitude
longitude
date
quantity

The Modelling of Constraints

There are some other important aspects of the real world that can be captured by E/R notation throughout **constraints**. The example below shows few constrains.

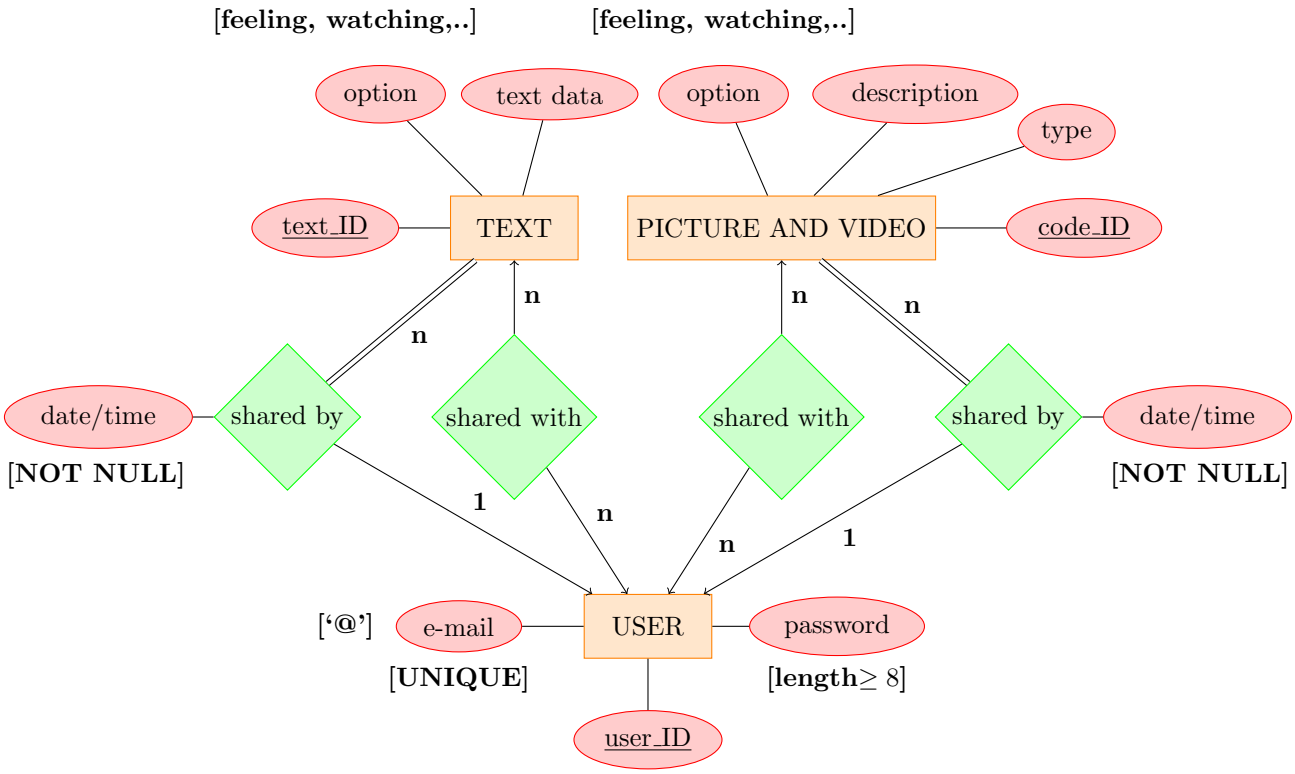


Figure 1: "FB, Post Text or Photo/Video" - E/R Diagram

- **Keys**, the attributes or sets of attributes that uniquely identify an entity within its entity set;
- **Single-value** requires that the values in certain context should be **unique** (city name in a nation should be unique, ...);
- **Participation** specify whether or not an entity must participate in a relationship **at least once**, that is for every entity in the entity set there exists a relationships in an instance relationship set;
- **Referential integrity** requires that a value referred (related) to, by means of a relationship, it exists (the capital of a nation should refer to an entity of the entity set city that exists);

- **Domains** requires that the value of an attribute must be in a set of values or a range of values (numbers, colors, categories, ...);
- General **assertions** are required to hold in the database (not null, is a number...);

In the E/R Diagram the attributes belonging to a **key** for an entity set is underlined. **Single-value** constraints could be represented placing a note beside the attribute (i.e. unique). Referential integrity constraints can be represented by a rounded arrowhead, pointing to the entity set is required the entity exits. [1] **Commonly, participation constraints are depicted using a double line from the entity set to the relationship set.**

Weak Entity Set

Whenever an entity set's key is composed of attributes some or all of which belong to another entity set, and it is said to be a "weak entity set". In case the key consists of zero or more of its own attributes, generally called "discriminator" (graphically the "discriminator" is dash-underlined) and key attributes from entity sets that are reached by a **many-one** *supporting* relationship (*many* examination records [*passed by*] *one* student, whenever examination record attributes are: date, course code, mark).

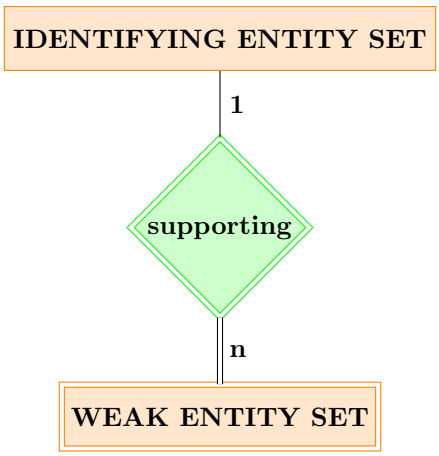


Figure 2: weak entity in relationship

Example:

A database needs to be developed that keeps track of compact disks, and songs recorded on them, together with authors and performers of the songs. Design the E/R diagram that captures the requirements stated below.

1. For each disk, we want to store the disk ID, the title, and the year of production. Disk ID's are unique. A song is recorded on some disk. It has a title and a number that indicates on which track it is recorded on the disk. For each disk, a song is uniquely identified by its track number.

2. For each person that may occur as a an author or a performer we want to store the person ID, the name, and the nationality. Person ID's are unique.

3. Each song has **at least** one author who is a person. There are different types of authorship, e.g., composer, text writer, or arranger. For each author of a song, we want to store the type of authorship.

4. Each song has **at least** one performer who is a person. There are different instruments for a performer, e.g., voice, piano, violin. We want to store for each performer the instrument they have played during the recording.

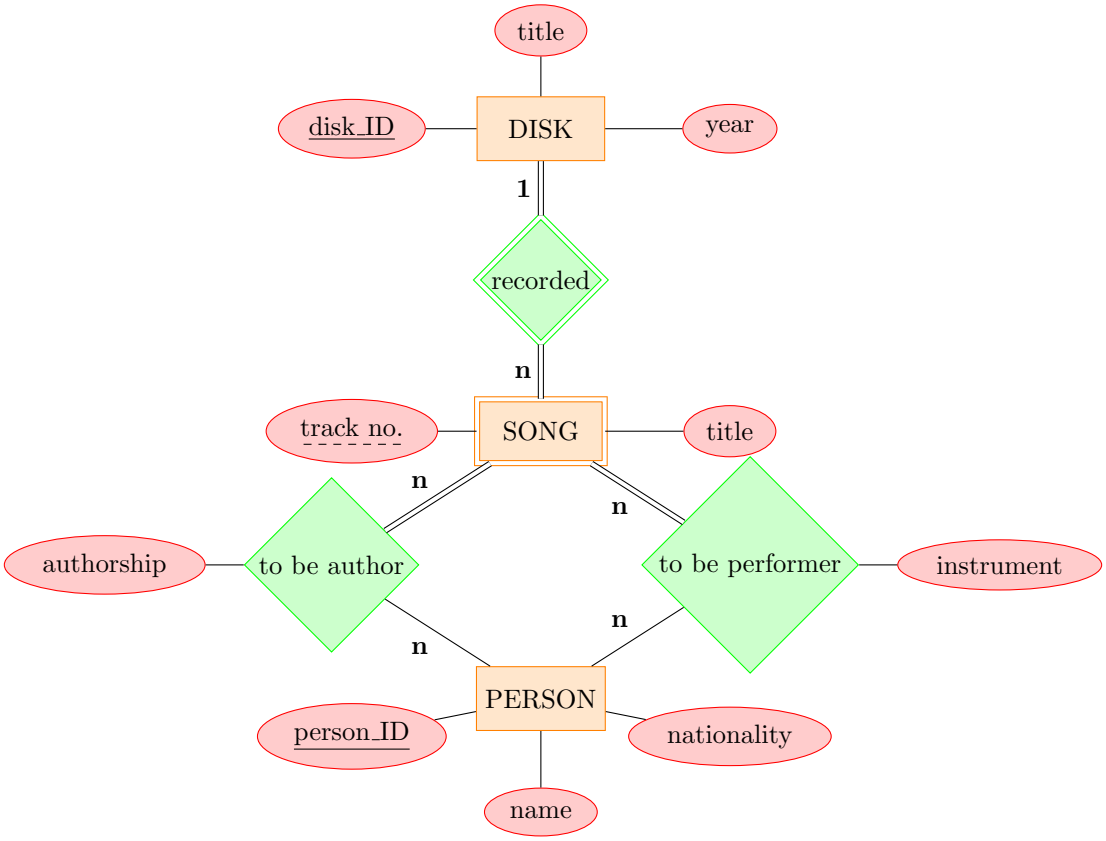


Figure 3: Disk - E/R Diagram

Ex. # 4

"It is very frequent that companies offer extra services to set a relationship with customers (or just interested people) through special cards. Cards are designed considering customer profiles. This type of offering provides data and information about customers, which are very helpful for marketing campaign."

Draw the E/R diagram that capture the requirements stated below.

1. Customer card requests implies to provide personal data: surname, name, date of birth, email and phone. Gathered data yields to the registration of a new **user** in the database.

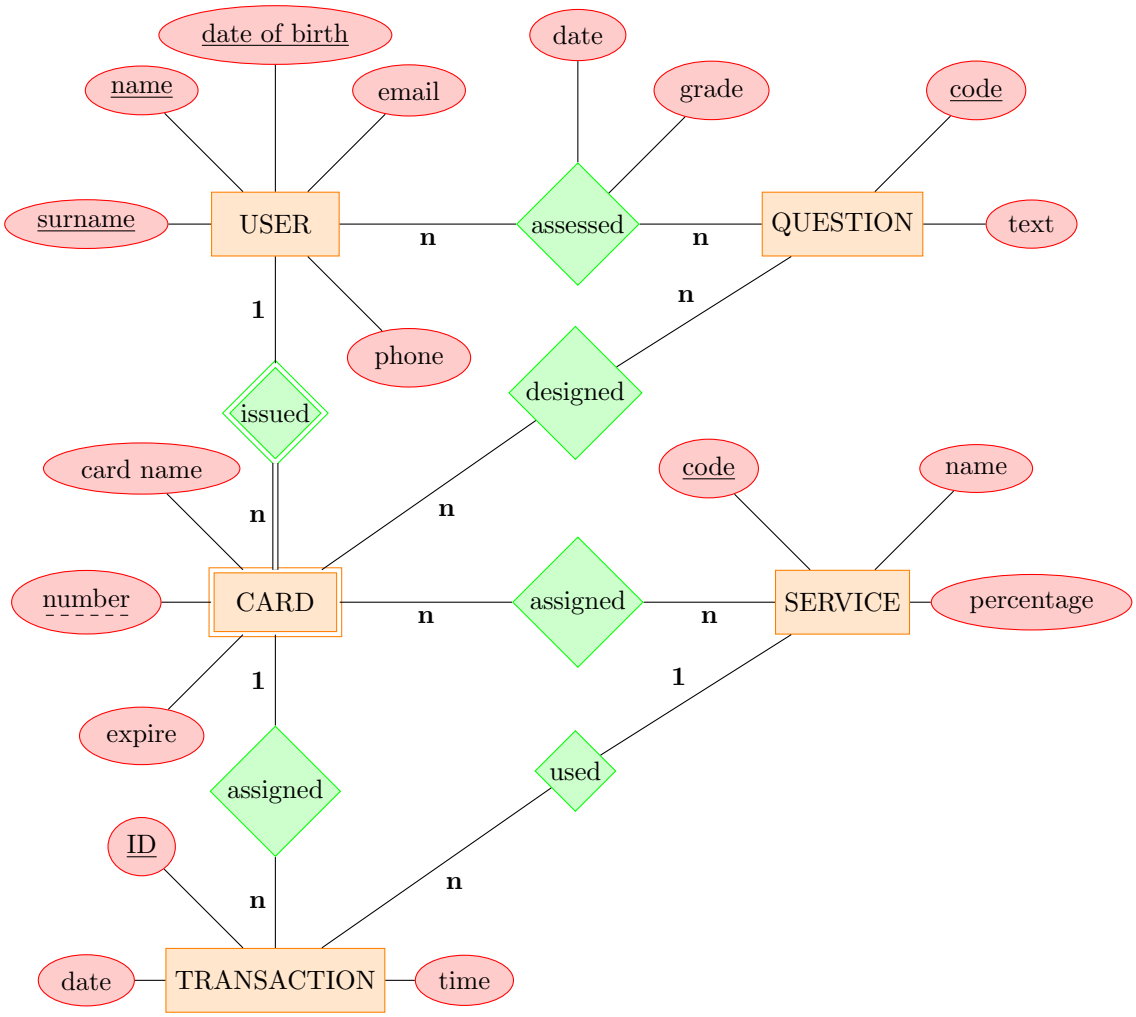
2. Card request implies to fill a questionnaire in order to have a profile of the user. All **questions** are coded. A user answers best fitting his profile based on his jobs, style of life, free time, and so on. All questions have a code, a name, and answers consist in a grade (how relevant it is in customer life 0 - 10).

3. A date must specify when this profile (answers) is set up considering that the profile can change in the future.

4. *Cards are specifically designed for specific profiles (for instance: top, senior, student, ...)*, but can be issued even if the user profile does not match. A **card** has a name, an expire date, and a number. A card is *uniquely* identified only when it is issued to the user, that is card number and surname, name, birth date, are combined. [*Tip*: it seems reasonable to design a weak entity set for card.]

5. A card grants access to extra **services** (free of charge), for example railway cards grant access to lounge rooms, obtain special or extra discounts, receive newsletters. All services has a code, a name, a percentage that could be a discount if applicable, null when not applicable. One or more services are typically associated to cards.

6. Each time the user uses a card a **transaction** is registered, that is date, time and used service.


date of birth
date
name
email
grade
code
surname
USER
n
assessed
n
QUESTION
text
1
n
phone
designed
issued
card name
code
name
n
n
number
CARD
n
assigned
n
SERVICE
percentage
1
1
expire
used
assigned
ID
n
n
date
TRANSACTION
time

Ex. # 5

"International business exhibitions, most frequently, propose events spread all over the city, but concentrated in some main areas. One most relevant example is the design week held in Milan yearly in spring. Each edition offers a huge variety of events, grouped by theme and located in different areas." The aim is to design a database to facilitate the planning of this type of exhibitions.

Draw the E/R diagram that capture the requirements stated below.

1. The most important entity is the **event**. The event is identified by a title, but it is essential to know where it is located, so we need the address of the location, the coordinates (in oder to visualize the event in a map), moreover when the event starts and ends.

2. Events are grouped by **districts**. A district identifies an urban area. Typically a district is identified by a name, but we need contacts and a web site address (if available).

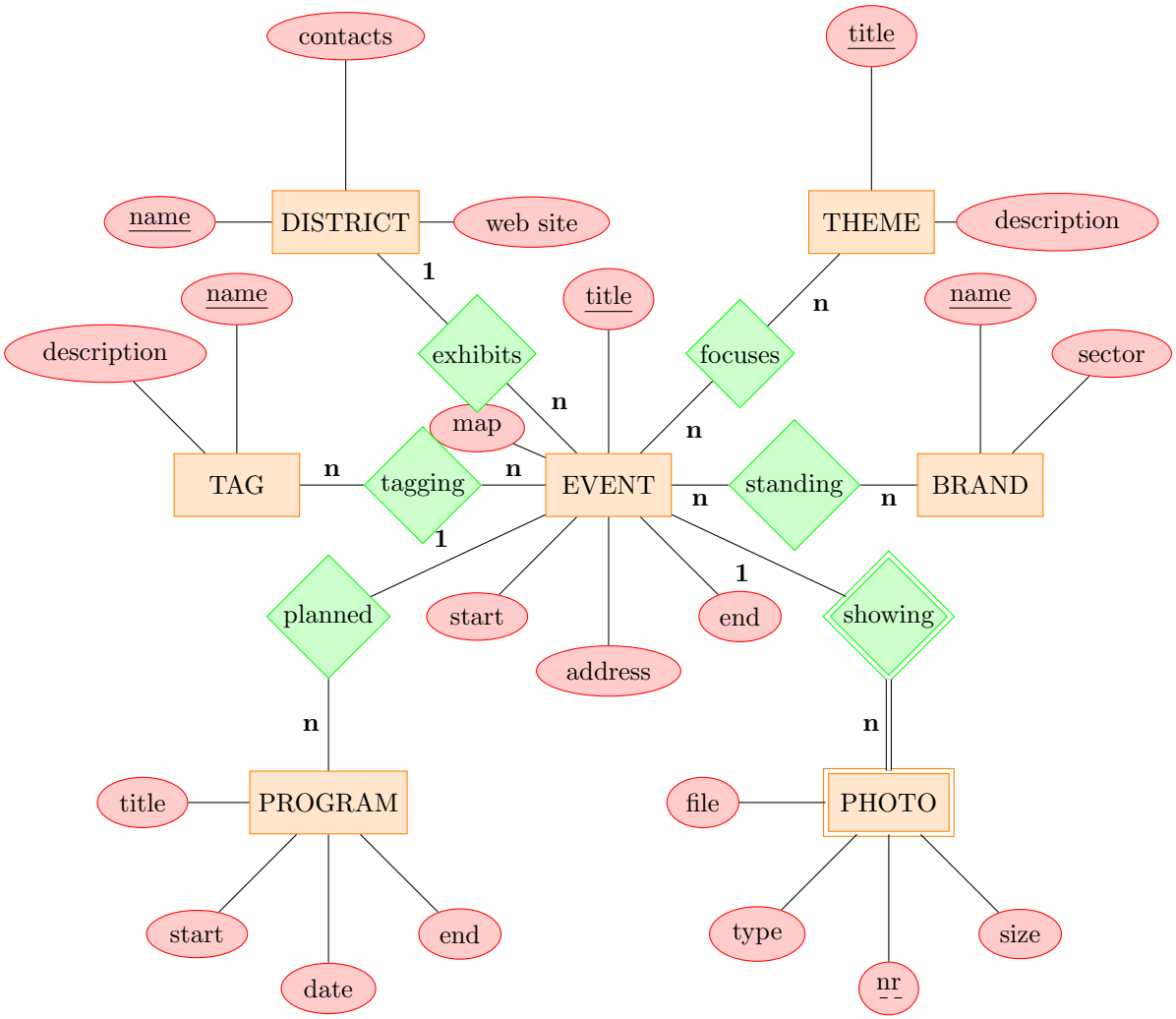
3. Moreover events are part of a **theme**. A theme is declared by a title and fully described.

4. One or more **brands** partecipate to one or more events. For each brand is needed the name, and the sector of activity.

5. One or more **tags** describe an event. Each *usable* tag has a name and a description.

6. Event presentation is well supported showing one or more **photos**. A picture (file) taken for an event has a sequential number, but it is important to register its size and photo type as well. Considering that the sequential number is not sufficient to identify a photo and moreover that a photo is strictly related to an event we assume to model this relationship by a week entity relationship.

7. Various activities are correlated to an event therefore the **program** of these activities are registered, in particular for each activity the date, start and end hours, and the title (for example *10 Wednesday, 19:30 - 22:30, reserved party*).


contacts
name
DISTRICT
web site
1
exhibits
name
description
TAG
n
tagging
n
map
title
EVENT
n
n
focuses
title
THEME
description
n
name
sector
BRAND
n
standing
n
1
planned
start
address
end
1
showing
n
n
title
PROGRAM
start
date
end
file
PHOTO
type
nr
size

Ex. # 6

"Recycling is a challenge to preserve earth and health". **"Recycl@ble"** is an app to help users to know if an object could be recycled, and in case of delivery at the municipality which benefits they have. The goal is to design a database supporting the 'app'.

Draw the E/R diagram that capture the requirements stated below. **Use "ID" as key only if strictly necessary**.

1. People interested in recycling something, is registered as **recycler**. Since he/she can login through Facebook/Google, the first time the email, password, name and surname are registered, other than the platform used to login. After that he/she can access to the functionalities offered by the 'app'. The user is always identified by this email.

2. The user logged can consult the list of all recyclable **objects**: name, description and if it is currently recyclable.

3. The 'app' is based on a sophisticated *Machine Learning* algorithm enabling detection among a set of images which is likely the object the user wish to recycle. When the user intends to recycle something selects from the list an object item and attaches a photo of the object, that is registered in the database.

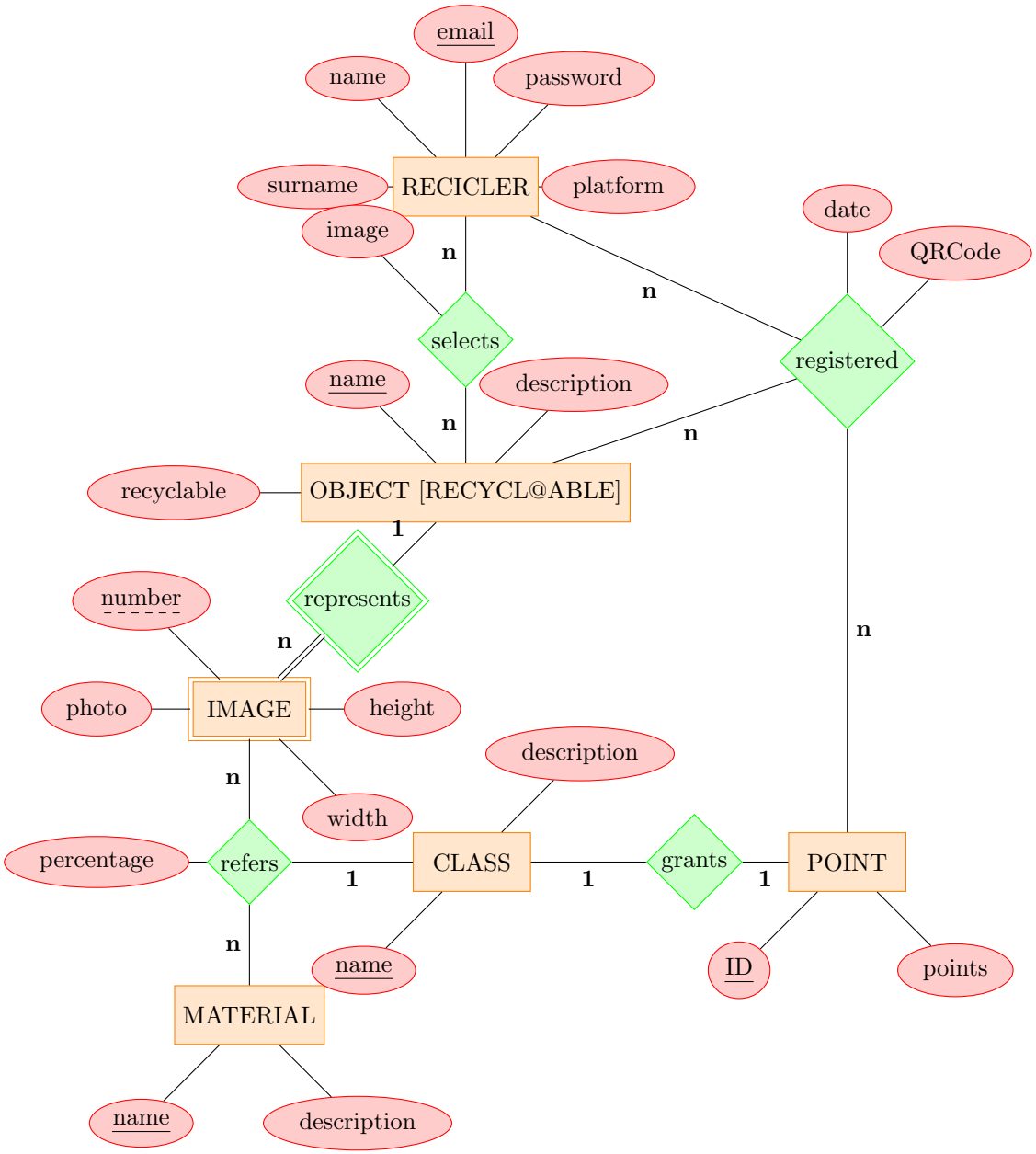
4. In order to identify a plausible recyclable object, the databases holds a high number of **images** for each object item in the list. An image has a number, specifying a sequence, size: width and height in pixel, the image itself.

5. Each image has associated one or more **materials** which should compose the object in the image. The association defines the percentage of material in the represented object.

6. An image with referred materials, depending on percentages, is classified. For example **full paper**, *paper and plastic*,... A **class** has name and description.

7. Depending on the class the user is entitled to know which benefit (transformed in **points**) receives if he/she delivers the object to municipalities for recycling.

8. Delivering the object the recycler receives a QRCode. Providing the QRCode to the 'app' the points are assigned and **registered** to the user - recycler -, that is object item, points, date, and QRcode.


email
name
password
surname
RECICLER
platform
image
date
QRCode
n
n
selects
registered
name
description
n
n
recyclable
OBJECT [RECYCL@ABLE]
1
number
represents
n
n
photo
IMAGE
height
n
description
width
percentage
refers
1
CLASS
1
grants
1
POINT
n
name
ID
points
MATERIAL
name
description

Ex. # 7

"Apps for smartphone are valid and friendly tools to offer cultural services". The municipality offers the great opportunity of free "cineforum", where citizen can choose movies they mostly like. An App that should accept movies preferences, member registration and participation is considered a valid solution. App should be supported by a database.

Draw the E/R diagram that capture the requirements stated below.

1. Everyone has the opportunity to vote one or more movies registered in the **catalogue** he/she likes the most. It represents a collection of movies, each one described by a title, year of realization, length, if it is available to show (yes/no), if it has already shown and when.

2. A movie can have one or more media visuals: poster, trailer, ... a **visual** has a type, a file, and could have many versions. A visual is strictly related to the movie, hence it should be designed has weak entity.

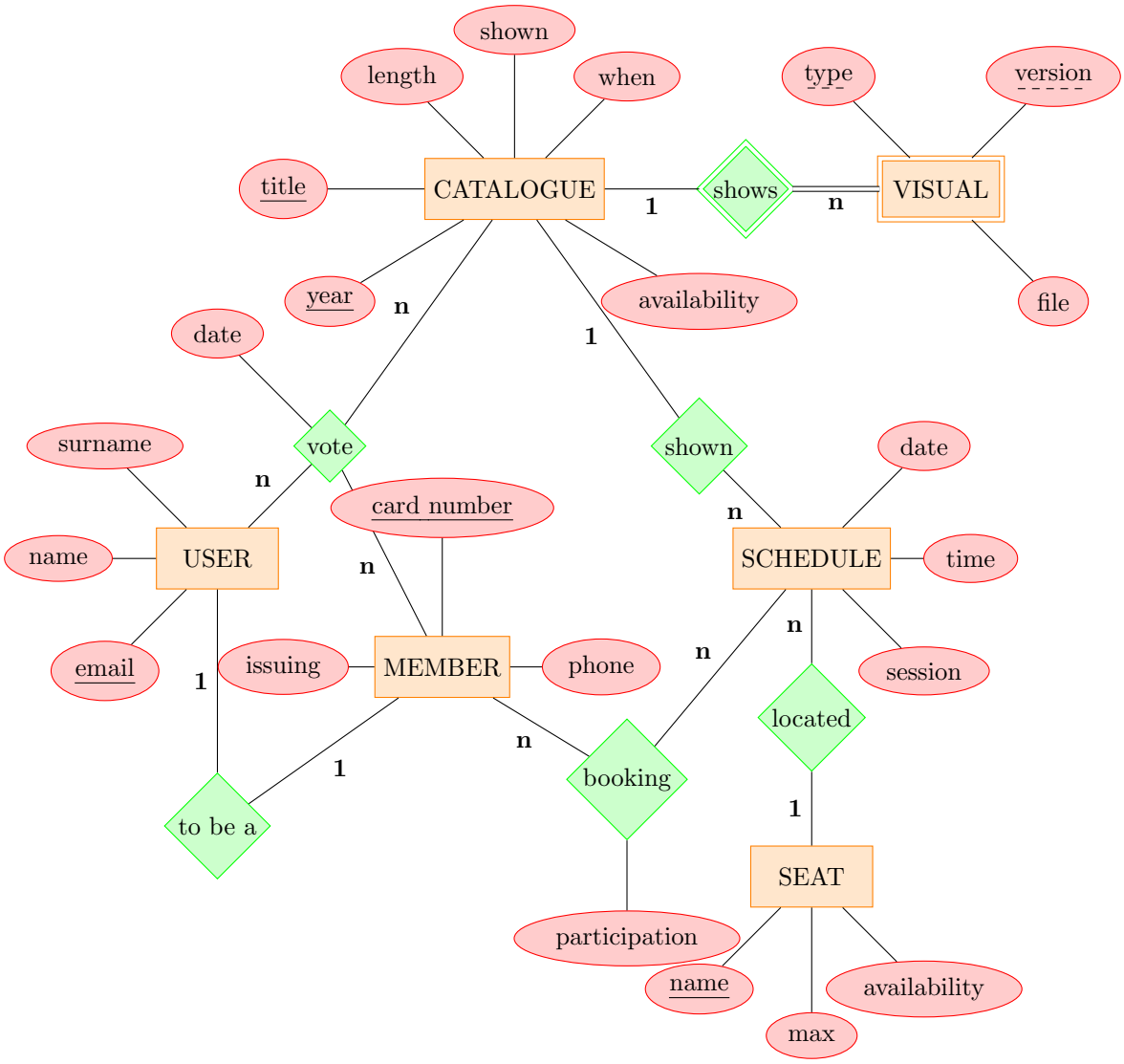
3. The app allows everyone, to vote his/her best movie in the catalogue - which can be voted just once - , it is sufficient to be a **user** of the App. To install the app you have to specify your name, surname and the email.

4. In order to book a seat at the cinema you must be **member**. The following data shall be registered: card number, name, surname, email and phone number, issuing date. **A member can continue to vote his/her preferite movies.**

5. Movies can be shown in various type of **seats**: park (open air movies), theatre, cinema, school, library, ... Each seat is identified by a name, and it is registered if it is currently available or temporarily not available, maximum number of seats.

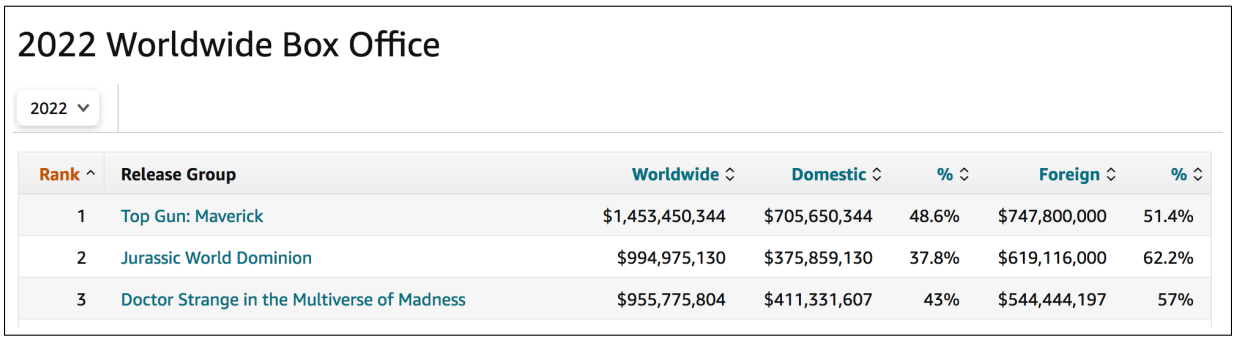
6. Every six months, the most voted movies are scheduled for the next weeks, considering if they are recently shown or not. **Schedule** identifies a movie to be shown in a seat at a specified date and time. The movie can be a part of a special session, in this case report the name of the session.

7. Members - not general users - can book a seat for one or more movie show (it is a very simple relationship). A member can go to the seat just showing a personal QRCode. The participation is registered (yes/no).


shown
length
when
type
version
title
CATALOGUE
1
shows
n
VISUAL
year
n
availability
file
date
1
surname
vote
shown
date
n
n
card number
name
USER
n
SCHEDULE
time
email
1
issuing
MEMBER
phone
n
n
session
located
n
1
booking
to be a
1
SEAT
participation
name
availability
max

XML vs. JSON

"Currently the Web shows a variety of data in effective way". In order to analyze the market of cinema consider the following worldwide box office data (year 2022), and describe it by means of the XML language and then by means of JSON notation.



2022 Worldwide Box Office

2022

Rank	Release Group	Worldwide	Domestic	%	Foreign	%
1	Top Gun: Maverick	$1,453,450,344	$705,650,344	48.6%	$747,800,000	51.4%
2	Jurassic World Dominion	$994,975,130	$375,859,130	37.8%	$619,116,000	62.2%
3	Doctor Strange in the Multiverse of Madness	$955,775,804	$411,331,607	43%	$544,444,197	57%

Figure 4: source: **https://www.boxofficemojo.com/year/world/**

```
<box-office year="2022">
 <ranking>
    <rank> 1 </rank>
    <movie> Top Gun: Maverick </movie>
    <amount currency = "$">
       <domestic> 705650344 </domestic>
       <foreign> 74780000 </foreign>
    </amount>
 </ranking>
 <ranking>
    <rank> 2 </rank>
    <movie> Jurassic World Dominion </movie>
    <amount currency = "$">
       <domestic> 375859130 </domestic>
       <foreign> 619116000 </foreign>
    </amount>
 </ranking>
</box-office>
```

Table 1: **XML file**

```
{
   "box-office": {
     "year": 2022,
     "ranking": [
        { "rank": 1,
           "movie": "Top Gun: Maverick",
           "amount": {
              "currency": "$",
              "origin": {
                 "domestic": 705650344,
                 "foreign": 74780000
              }
           }
        },
        { "rank": 2,
           "movie": "Jurassic World Dominion",
           "amount": {
              "currency": "$",
              "origin": {
                 "domestic": 375859130,
                 "foreign": 619116000
              }
           }
        }
      ]
   }
}
```

Table 2: **JSON file**

References

1. J. Ullman, J. Widom, A First Course in Database Systems - Third Edition, Pearson - Prentice Hall

2. P. Atezeni et Altri, Basi di dati: modelli e linguaggi di interrogazione, McGraw Hill