Statistical Methods for Database Integration
Examination
DATABASES

The exam consists in two parts:

1) Part A: **The exam is closed-book, closed-notes**;

2) Part B: **You are allowed to use lecture and labs notes**.

Each questions is assigned points expressed in cents.

PART A

Ex. 1

(a) **(10 points)** "The Web is widely used to promote artistic and natural heritage". This year "Fondo Ambiente Italiano" promotes and supports, with Web tools, insertion of information of Italian heritage, and gather appreciations, through votes. The Web page below displays the ranking of the first three places in Bologna area. Describe by means of the XML language the data: the ranking (rank and votes), location, type and title and other useful information. **Describe just one card at your choice!**

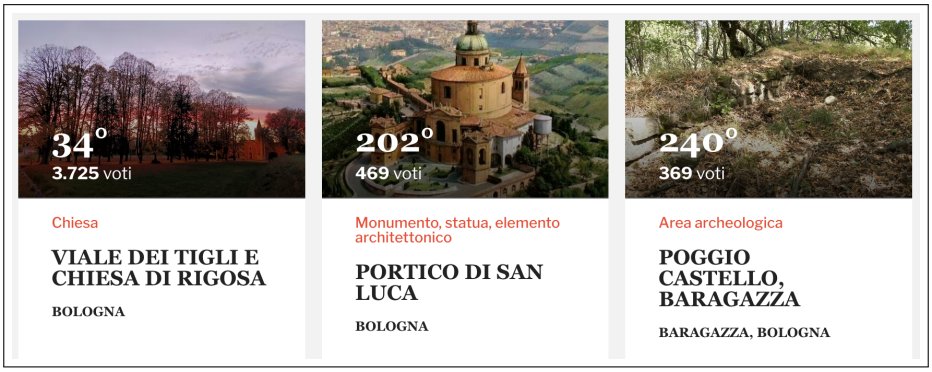


Figure 1: source: **https://www.fondoambiente.it**

[*Sol.*:]

```
<italian-heritage year="2022">
   <card>
      <ranking>
         <position unit="°"> 34 </position>
         <votes> 3725 </votes>
      </ranking>
      <category> Chiesa </category>
      <title> VIALE DEI TIGLI E CHIESA RIGOSA </title>
      <location> Bologna </location>
   </card>
</italian-heritage>
```

(b) **(10 points)** Now translate XML (semi-structured data model) into a relation (structured data model). **Specify carefully for each identified attribute the corresponding data domain (refer to the SQL data types or specific subsets of them)**.

[*Sol.*:]

```
italian-heritage(position: INTEGER,
                    votes: INTEGER,
                 category: VARCHAR(50), [Chiesa, Area archeologica, ...]
                    title: VARCHAR(100),
                 location: VARCHAR(30)
)
```

(c) **(Optional: 5 points) Scraping** is the process to gather data from Web pages. Describe shortly this process independently of the used programming language.

[*Sol.*: See teaching material]

Es. 2 - Data Modeling

(1) **(35 points)** "**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**.

(a) A **user's farm** has to register on the 'app', his name, surname, email and a password.

(b) 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, ...).

(c) 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, when the products of his farm are not yet registered in the database.

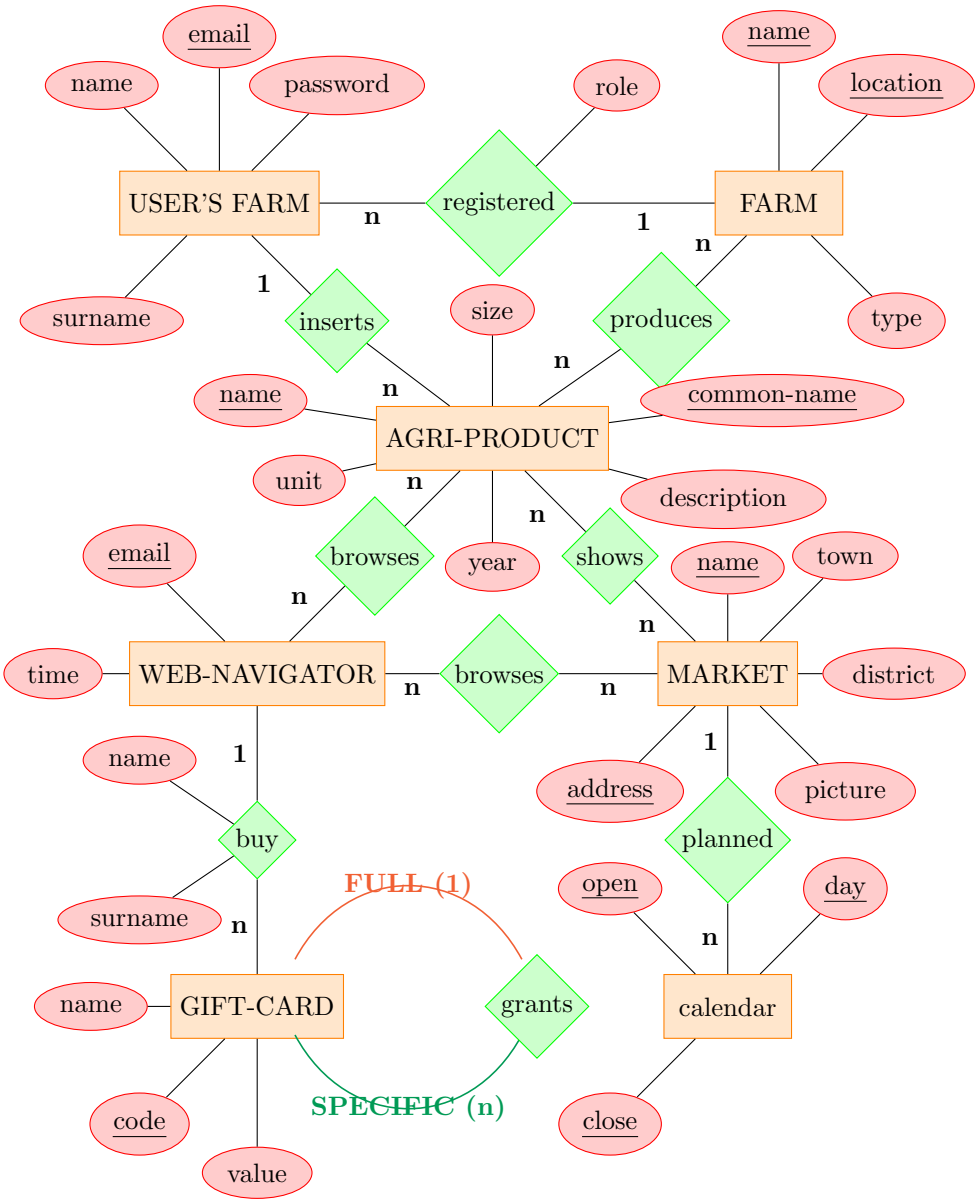
(d) The 'app' holds data about **markets**, places where farmer can show their products for sale. Each market has name, town/village, district, address, picture.

(e) Farmer can show food products at multiple markets, according to the **calendar** defined for each market.

(f) People interested in agri-foods/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.

(g) 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**.

(h) 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 the navigator receives an email with attached the card in a pdf file.


email
name
password
role
name
location
USER'S FARM
registered
FARM
n
1
1
n
surname
inserts
size
produces
type
n
n
name
common-name
AGRI-PRODUCT
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
planned
FULL (1)
open
day
surname
n
n
name
GIFT-CARD
grants
calendar
SPECIFIC (n)
code
close
value

(2) **(Optional: 5 points)**. Write the SQL statement to `CREATE` the "relations" that describe **Market** and **Calendar**.

```
CREATE TABLE Market(
   name VARCHAR(50),
   address VARCHAR(100),
   town VARCHAR(50),
   district VARCHAR(50),
   picture BLOB,
   PRIMARY KEY(name, address)
);

CREATE TABLE Calendar(
   day DATE,
   open TIME,
   close TIME,
   market_name VARCHAR(50),
   market_address VARCHAR(100),
   PRIMARY KEY (day, open, close),
   FOREIGN KEY (market_name, market_address) REFERENCES Market(name, address)
);
```

PARTE B

Es. 3 - SQL (45 points) Let assume the database "online-market".

(1) Region(name, description)

(2) Producer(name, description)

(3) Sheet(ID, description, Region.name, Producer.name)

(4) Ingredient(name, description)

(5) Made(Ingredient.name, Sheet.ID)

(6) Menu(name, description, main)

(7) Food(name, unit, weight, label, price, startDate, endDate, Menu.main_name, Sheet.ID)

(8) GiftBasket(name, description)

(9) BasketCombines(GiftBasket.name, Food.name, Food.unit, Food.weight)

(10) User(ID, date, time, network_info)

(11) Consulted(User.ID, Food.name, Food.unit, Food.weight, time)

(12) Selected(User.ID, Food.name, Food.unit, Food.weight, time, quantity)

Questions

1) Marketing has formulated the need of a report of users accesses long **summer time**, that is how many items (products) the user selected (or nothing), when browsing the shop.

   Report for each user, **identified by the network info**, the number of items selected (or nothing) in summer. Show number of selections from the largest to the smallest.

   [*Sol.*]

```
SELECT network_info, count(food_name) AS nr
  FROM user LEFT JOIN selected ON (user.ID = selected.ID)
 WHERE date >= '2022-06-22'
   AND date <='2022-09-21'
GROUP BY network_info
ORDER BY nr DESC;
```

2) We would like to create a new gift basket having in mind something of healthy and of tradition and for some of them to apply a special discount. List foods in the menu 'Olive Oil', but do not retrive those having the highest price. Considering the high quality of products we do not apply any discount. Then list foods that can be related to 'Bologna' tradition, and moreover all types of legumes. For those returns in a new column that you apply a 30% discount. [**Tip:** Write different selections and then combine returned tuples]

[*Sol.*]

```
(SELECT CONCAT(name, '-', weight, '-', unit) AS prod, price, NULL AS disc
   FROM food
  WHERE menu_name = 'Olive Oil'
    AND price NOT IN (SELECT MAX(price)
                        FROM food
                       WHERE menu_name = 'Olive Oil'))
UNION
(SELECT CONCAT(name, '-', weight, '-', unit) AS prod, price, '30%' AS disc
   FROM food
  WHERE name LIKE '%bologn%'
     OR menu_name = 'Legumes');
```

3) Aiming to update food sheets, marketing requests to list all sheets code, and producer name but only for producers who have **at least two** sheets (food products) published on the online-market.

[*Sol.*]

```
SELECT DISTINCT p1.producer_name AS producer, p1.ID AS code
  FROM sheet p1, sheet p2
 WHERE p1.producer_name = p2.producer_name
   AND p1.ID <> p2.ID
ORDER BY code;
```