



# 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)** “Waste reduction has great benefits for the environment”. According to OECD Environment at a Glance [2020] we compare the main indicators of treatment municipal waste: % **Landfilling**, % **Recycling**, **Total quantity of waste (Kg/hab)** in **each year**. Report data in a semi-structured data model (XML Language). **Describe only two years!**

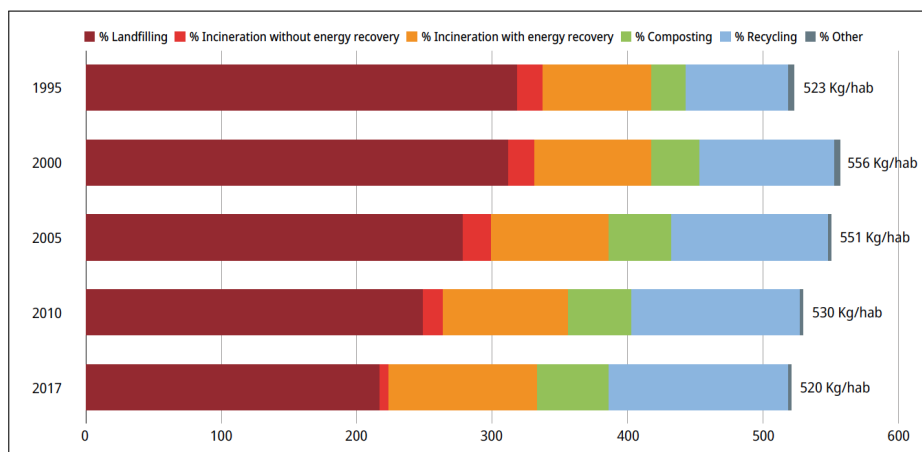


Figure 1: source: **Municipal Waste - OECD**



```
<municipal_waste>
  <time year="1995">
    <total unit="Kg/hab"> 523 </total>
    <percentages unit="%">
      <landfilling> 65 </landfilling>
      <recycling> 18 </recycling>
    </percentages>
  </time>
  <time year="2000">
    <total unit="Kg/hab"> 556 </total>
    <percentages unit="%">
      <landfilling> 62 </landfilling>
      <recycling> 23 </recycling>
    </percentages>
  </time>
</municipal_waste>
```

- (b) **(10 points)** The table tells the percentage of recyclable material we take out from electronic devices.

Device	Material	%
laptop	plastic	45%
	iron	25%
	silicon	30%
smartphone	plastic	65%
	silicon	35%

Create by means of a **SQL statement** a table holding the given data and **refer** to the table **Object** which holds some related data.

```
CREATE TABLE Object(
  name VARCHAR(20) PRIMARY KEY,
  description VARCHAR(100),
  recyclable BOOLEAN
);

CREATE TABLE Composition(
  object VARCHAR(20),
  material VARCHAR(30),
  percentage NUMERIC(5,2),
  FOREIGN KEY (object) REFERENCES Object(name)
);
```

- (c) **(Optional: 5 points)** NoSQL represents the next generation of Databases. What is the main difference with respect to **Relational Databases**?

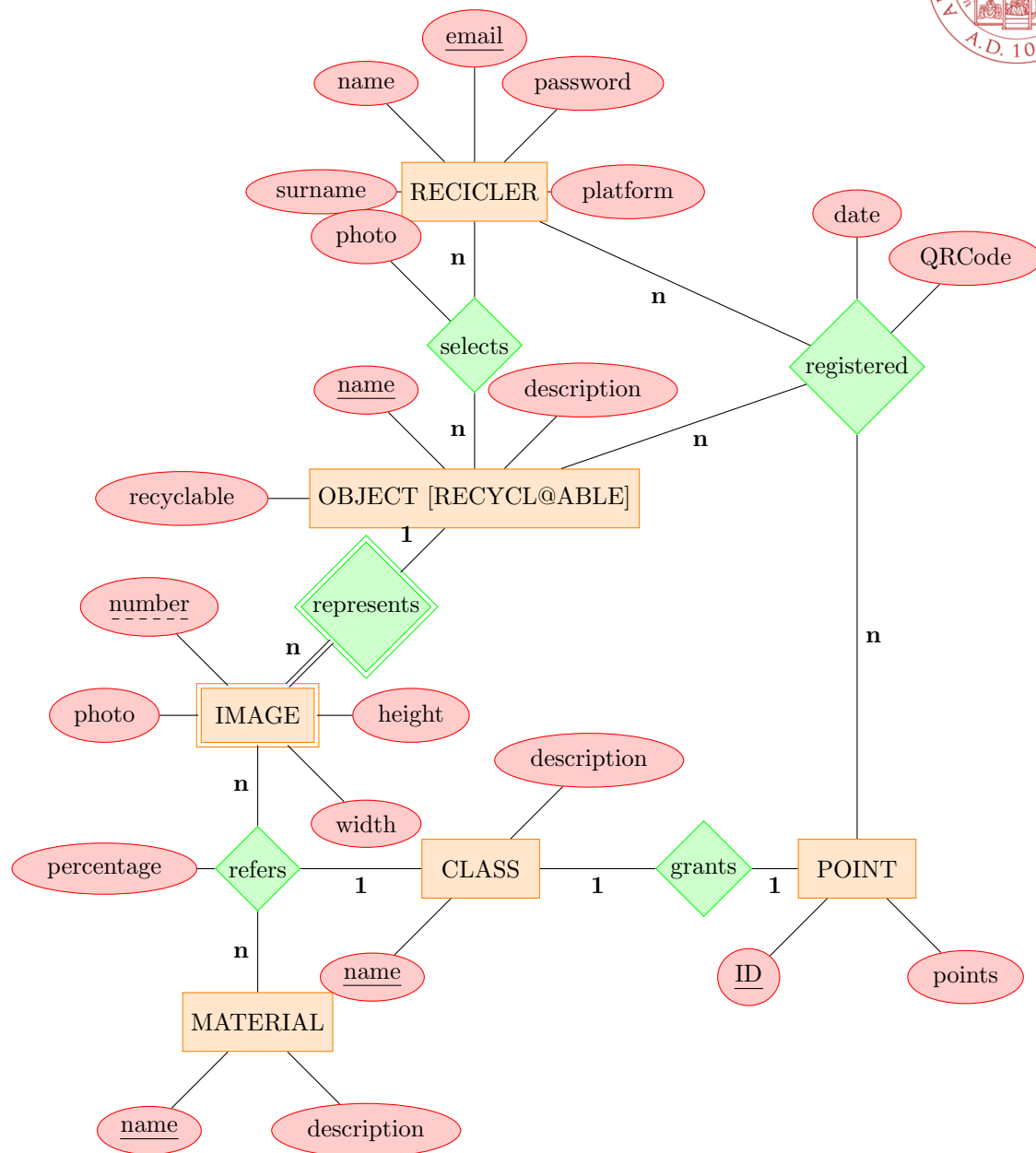
[Sol.: See teaching material]

**Es. 2 - Data Modeling**

- (1) **(35 points)** “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.

- (a) 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.
- (b) The user logged can consult the list of all recyclable **objects**: name, description and if it is currently recyclable.
- (c) 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.
- (d) 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.
- (e) 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.
- (f) An image with referred materials, depending on percentages, is classified. For example *full paper, paper and plastic,...* A **class** has name and description.
- (g) 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.
- (h) 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.



(2) (Optional: 5 points). Write “relations” of the **Relational Data Model** which describe **Class** and **Point**.

- Point(ID, points)
- Class(name, description, Point.ID)



## 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.menu\_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) We would like to create a particular gift basket having in mind something from Sicily and a selection of foods from each menu. List name, weight, unit of foods from 'Sicilia' then detect from each menu the product in glass jar (**unit in grams**) having the largest quantity. [Tip: The second selection should use in the condition the pair (menu\_name, weight)].

```
(SELECT name, weight, unit, 'Sicily'
  FROM food JOIN sheet ON sheet_ID = ID
 WHERE region_name = 'Sicilia')
UNION
(SELECT name, weight, unit, 'Big Glass'
  FROM food
 WHERE (menu_name, weight) IN (SELECT menu_name, MAX(weight)
                                FROM food
                                WHERE unit = 'g'
                                GROUP BY menu_name))
ORDER BY name, weight;
```



- 2) Marketing has formulated the request to determine in which day the online shop has been consulted the most (Clicks on some products). Report the date and the number of consultations (clicks). [Tip: Data to consider are available in **User** and **Consulted** tables].

```
SELECT date, count(*) AS nr, 'clicks'
FROM user JOIN consulted ON user.ID = consulted.ID
GROUP BY date
HAVING COUNT(*) >= ALL(SELECT count(*)
                        FROM user JOIN consulted ON user.ID = consulted.ID
                        GROUP BY date);
```

- 3) Marketing would like to know if - and in case the relevance - of organic foods.

A) Determine for each menu the number of organic products, and moreover for each menu the number of traditional products (non organic).

B) [extra points] Report data in table format:

Menu	# Organic	# Traditional
Olive Oil	2	1
Sauces and Pesto	1	4

```
SELECT O.menu_name, O.organic, T.traditional
FROM (SELECT menu_name, COUNT(*) AS organic
      FROM food
      WHERE name LIKE '%organic%'
      GROUP BY menu_name) O,
      (SELECT menu_name, COUNT(*) AS traditional
      FROM food
      WHERE name NOT LIKE '%organic%'
      GROUP BY menu_name) T
WHERE O.menu_name = T.menu_name;
```