Assignment 1

Submission instructions

The assignment should be submitted by pairs. You should submit one zip file. The archived file should contain 2 files:

- A Txt file named sql.txt that contains: the *AutoIncrement* trigger, *MaximalDistance* and the *SimCalculation* functions
- A Java class named *Assigment.java* that contains all the functions and comments

The archive file name should be of form id1_id2.zip .
Dead line: 30.11.2019

1. Development instructions

1.1. Oracle SQL Developer

1.1.1. Create a table MediaItems with the following columns:

- MID (NUMBER(9,0)) –primary key
- TITLE (VARCHAR2 (200))
- PROD_YEAR (NUMBER(4))
- TITLE_LENGTH (NUMBER(4))

1.1.2. Create a table Similarity with the following columns:

- MID1 (NUMBER(9,0)) – primary key, foreign key to MediaItems.MID
- MID2 (NUMBER(9,0)) – primary key, foreign key to MediaItems.MID
- SIMILARITY (FLOAT)

1.1.3. Create a trigger AutoIncrement

- On each insertion of the row into the table MediaItems, the trigger generate and update 2 values in the row: MID index and TITLE_LENGTH – length of the title.
- The first MID index should be 0.
- Do not use any sequence to generate MID

1.1.4. Create the Oracle function MaximalDistance

- The function does not receives any values
- The function return the maximal distance between all the items – a number
- Two items distance is the squared difference between the production years of the given items
- E.g:

$$d(a,b) = (\mathrm{PROD\_YEAR\_}a \;-\; \mathrm{PROD\_YEAR\_}b\,)^2$$

Maximal distance is the maximal result of the distance calculation between all the item pairs.

e.g. $F = \max(d(a,b)|\forall\, a,b\; \in MediaItems)$

1.1.5. Create an Oracle function SimCalculation which calculates the similarity between 2 media items.

- The function receives 2 MIDs and the number *maximal_distance*
- The function returns the similarity - a number in the range [0,1] (float)
- The similarity is defined as: $1 - \frac{two\_items\_distance}{maximal\_distance}$

1.2. Java

1.2.1. Write class constructor

- The constructor receives three values:
  1. Connection string
  2. DB username
  3. DB password
- The given parameters should be used for creation of the connection(s)

1.2.2. Write a Java function fileToDataBase

- The function receives the path of the file(String).
- The function reads the file content and inserts it to the *MediaItems* table
- The file is in CSV(comma separated values) format, the first value is the title of the item and the second value is the production year [see attached file: films.csv]

1.2.3. Write a Java function calculateSimilarity

- The function does not receive any values.
- The function calculates the similarity between every pair of items in the *MediaItems* table using the *SimCalculation* and *MaximalDistance* Oracle functions and inserts or updates the row in the *Similarity* table. The inserted row should contain MID1,MID2,SIMILARITY

1.2.4. Write a Java function printSimilarItems

- The function receives a long number mid.
- The function retrieves in ascending order from the database all the items that the similarity between them and the given item is at least 0.3.
- The function prints all the titles and the similarity value of the similar items (using System.out.println).

No need to submit the *main* function, but you can use it for the internal test, although JUnit is more convenient for that purpose.

Good Luck,
Noy Cohen-Shapira & Yoni Cohen.