Technologies for Information Systems

Part I (10 points)

prof. L. Tanca – June 22nd, 2018

Available time: 25 minutes

Last Name _____

First Name _____

Student ID _____ **Signature**_____

1. Define the concepts of valid time and transaction time in temporal databases also describing their advantages and disadvantages.
2. Describe what we mean by "lightweight approaches to data integration", using as an example the Mashup technique, and explain in which circumstances these approaches are appropriate, with advantages and disadvantages.

- **During this part of the exam, students are not allowed to use books or notes.**
- **Students should answer the theoretical questions using their own words, in order for the teachers to be able to assess their real level of understanding.**

Technologies for Information Systems

Part II (23 points)

prof. L. Tanca – June 22nd, 2018

Available time: 2h 00m

Last Name _____

First Name _____

Student ID _____ **Signature**_____

PoliScience is a platform offering the online purchase and download of scientific journal papers in electronic format. Registered users can access the platform and download papers by paying a fee. Note that the system charges every download, so if a user wishes to downolad again a paper he/she has already downloaded in the past, he/she is required to pay the fee again.

The management of PoliScience has now hired you to design a data warehouse to analyze the downloads performed by the users.

The following is the schema of the operational database used by PoliScience:

COUNTRY (CountryName, Continent)

USER (UserId, GivenName, Surname, Affiliation*, CountryName) *// Registered users may be affiliated to an institution (e.g., a university); Affiliation is null if this is not the case.*

PAPER (PaperId, Title, NumOfPages, PublicationDate, JournalName, Price)

DOWNLOAD (UserId, Date, Time, PaperId)

SUBTOPIC (SubtopicName, TopicName) *// Sample subtopic/topic pair: ("Databases", "Computer Science").*

AUTHOR (AuthorId, GivenName, Surname, BirthYear, CountryName)

PAPERSUBTOPIC (PaperId, SubtopicName, Relevance) *// A paper may have multiple subtopics, each with its relevance; the relevances of the subtopics to a paper sum to 1. For instance, a paper may be associated with the subtopic "Databases" with relevance 0.7 and with the subtopic "Software Engineering" with relevance 0.3.*

PAPERAUTHOR (PaperId, AuthorId) *// A paper may be authored by multiple co-authors.*

1. (3 points) Perform the reverse engineering of the given logical schema into a conceptual schema (Entity-Relationship model).

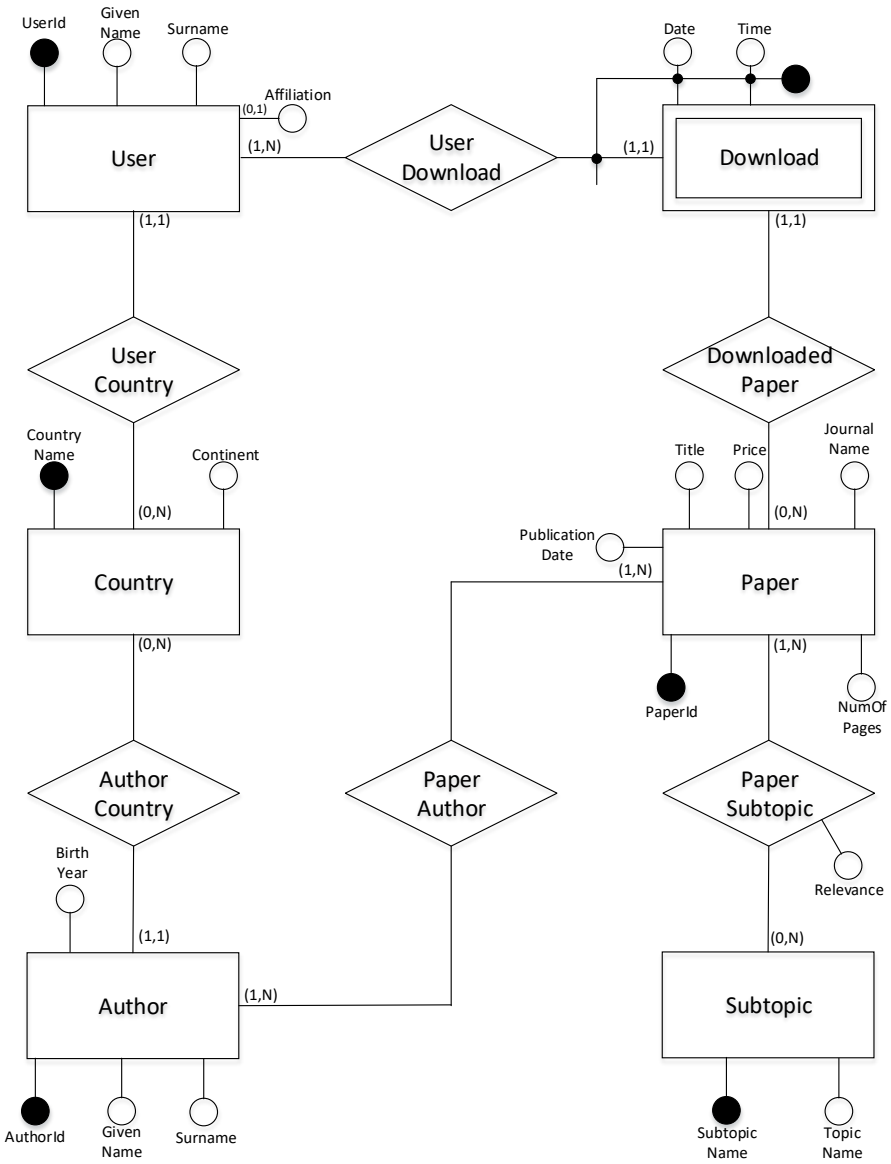
2. With respect to the produced ER diagram, discover the fact(s) that are useful specifically for answering the queries reported below. For each of these facts:
    a. (3 points) Produce the attribute tree (with pruning and grafting).
    b. (3 points) Produce the conceptual schema (fact schema).
    c. (2.5 points) Produce the glossary.

3. (3 points) Produce a logical schema consistent with the conceptual schema.

4. Write in SQL the following queries against the designed logical schema:
    a. (2 points) Aggregate the total income from European users by date, month and year (include in the answer the aggregations computed only by date, only by month and only by year).
    b. (2 points) Compute the number of downloads for each subtopic of “Computer Science”, weighted by the relevance of the subtopics to the papers.
    c. (2 points) Considering only Asian authors, compute the number of downloads for each author, user affiliation and journal name.
    d. (2.5 points) Compute the total income by paper (specify id and title) and user country, considering only the papers with at least one French author.

SOLUTION

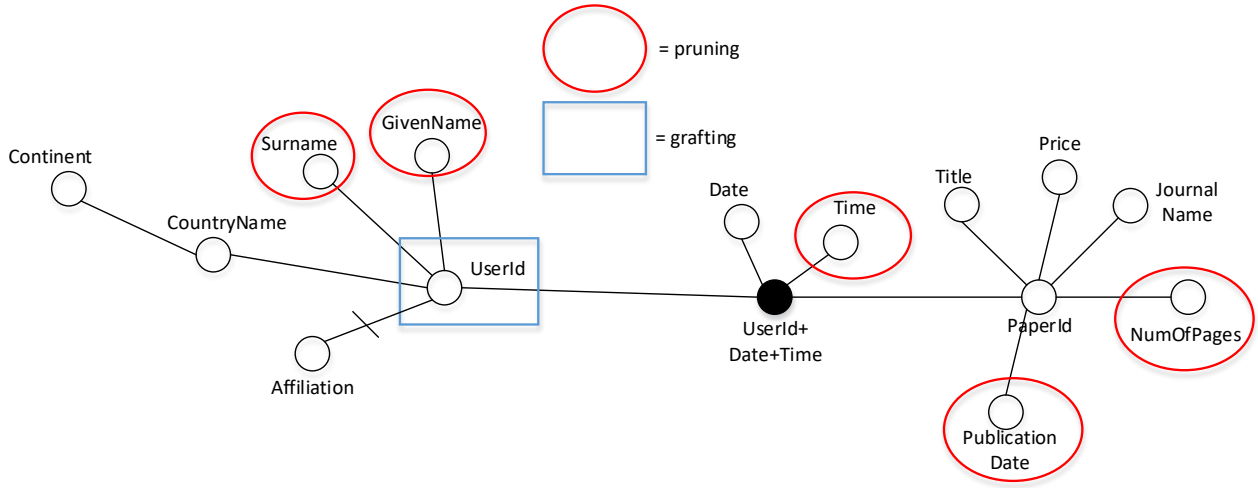
1. Reverse engineering



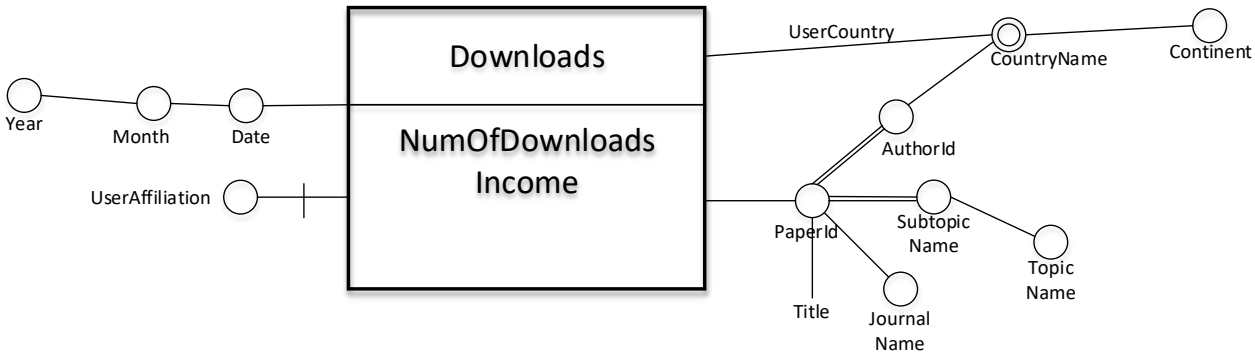
2. Conceptual design

Fact: Downloads (Download entity)

2a) Attribute tree



2b) Fact schema



2c) Glossary

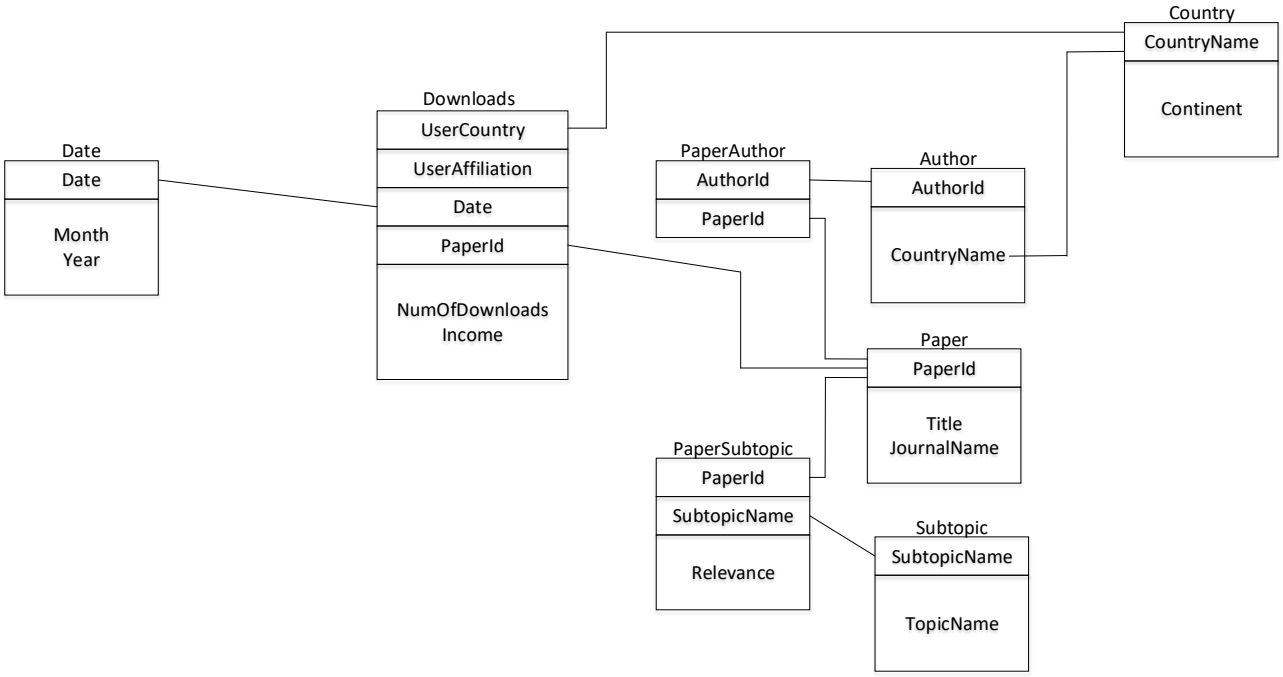
NumOfDownloads

SELECT D.PaperId, D.Date, U.Affiliation, U.CountryName, **COUNT**(*)
FROM Download **AS** D, User **AS** U
WHERE D.UserId=U.UserId
GROUP BY D.PaperId, D.Date, U.Affiliation, U.CountryName

Income

SELECT D.PaperId, D.Date, U.Affiliation, U.CountryName, **SUM**(P.Price)
FROM Download **AS** D, User **AS** U, Paper **AS** P
WHERE D.UserId=U.UserId **AND** D.PaperId=P.PaperId
GROUP BY D.PaperId, D.Date, U.Affiliation, U.CountryName

3. Logical design



4. Query answering

4a) Aggregate the total income from European users by date, month and year (include in the answer the aggregations computed only by date, only by month and only by year).

SELECT Da.Year, Da.Month, Da.Date, **SUM**(D.Income)
FROM Downloads **AS** D, Date **AS** Da, Country **AS** C
WHERE D.Date=Da.Date **AND** D.UserCountry=C.CountryName **AND** C.Continent='Europe'
GROUP BY Da.Year, Da.Month, Da.Date **WITH ROLLUP**

4b) Compute the number of downloads for each subtopic of "Computer Science", weighted by the relevance of the subtopics to the papers.

SELECT S.SubtopicName, **SUM**(D.NumOfDownloads*PS.Relevance)
FROM Downloads **AS** D, PaperSubtopic **AS** PS, Subtopic **AS** S
WHERE D.PaperId=PS.PaperId **AND** PS.SubtopicName=S.SubtopicName **AND** S.TopicName='Computer Science'
GROUP BY S.SubtopicName

4c) Considering only Asian authors, compute the number of downloads for each author, user affiliation and journal name.

```
SELECT A.AuthorId, D.UserAffiliation, P.JournalName, SUM(D.NumOfDownloads)
FROM Downloads AS D, Paper AS P, PaperAuthor AS PA, Author AS A, Country AS C
WHERE D.PaperId=P.PaperId AND P.PaperId=PA.PaperId AND PA.AuthorId=A.AuthorId AND
        A.CountryName=C.CountryName AND C.Continent='Asia'
GROUP BY A.AuthorId, D.UserAffiliation, P.JournalName
```

4d) Compute the total income by paper (specify id and title) and user country, considering only the papers with at least one French author.

```
SELECT P.PaperId, P.Title, D.UserCountry, SUM(D.Income)
FROM Downloads AS D, Paper AS P
WHERE D.PaperId=P.PaperId AND P.PaperId IN (
                                        SELECT PA.PaperId
                                        FROM PaperAuthor AS PA, Author AS A
                                        WHERE PA.AuthorId=A.AuthorId AND
                                                A.CountryName='France'
                                        )
GROUP BY P.PaperId, P.Title, D.UserCountry
```