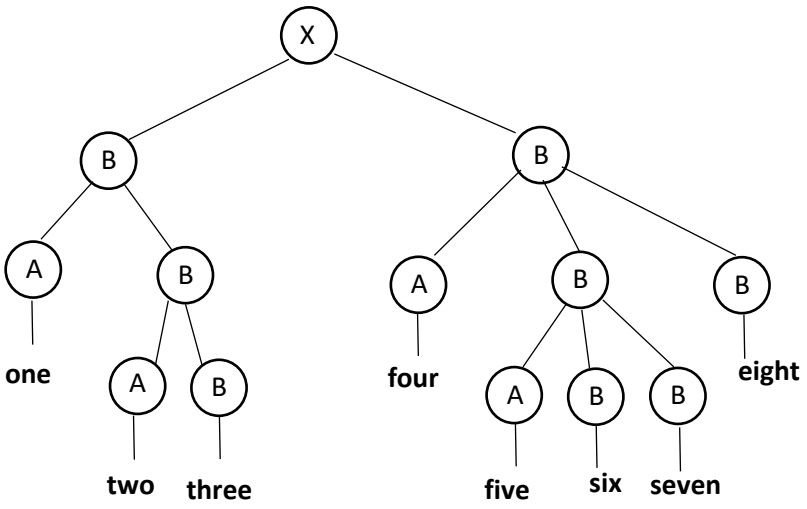
CENG 352 - Database Management Systems
Spring 2024
Written Assignment 1

I. XML and XPath (20 pts.)

Consider the following data in the form of a tree.



For each of the XPath expressions below indicate what they return when evaluated on the data represented above.

i. /X/B/*/text()
ii. //B[B/A]//text()
iii. //B[A]/B/*
iv. /X/B[B][A]/B/*/text()
v. //B[*/B[2]]/*/text()
vi. /B/B/preceding-sibling::text()
vii. //*[count(following-sibling::*) = count(preceding-sibling::*) + 1]/text()

II. Database Design (50 pts.)

2.1. Keys, Supekeys (6 pts.)

Consider the relation R(A,B, C, D, E) and the FD's {E->D, BD->A, AC->E}.

a) What are the non-trivial FD's that follows from the given FD's? List only the FD's with single attributes on the right side.
b) What are all the keys of R?
c) What are all the superkeys for R that are not keys?

2.2. BCNF Decomposition (14 pts.)

Given the schema R = {A, B, C, D, E, F, G, H, I, J} and the functional dependencies below:

{AE → BFGH, B → ADE, F → G, G → F}

a) Find all keys (Show your work).

b) Explain why R is not in BCNF.

c) Decompose R into a collection of BCNF relations.

d) Is your BCNF decomposition

i) dependency-preserving or not? Explain.

ii) lossless-join decomposition or not? Explain.

2.3. 3NF Decomposition (20 pts.)

Given the schema R = {A, B, C, D, E, F } and the functional dependencies below:

{ F → B, D → ABE, EF → D, C → E, B → AE, BF → A, AE → F, CF → ADE}

a) Find the minimal cover for the given set of functional dependencies.

b) Decompose relation R into 3NF using the minimal cover.

c) Is the decomposition also in BCNF? Explain.

2.4. 4NF Decomposition (10 pts.)

Consider the following relation that describes births at a hospital.

Birth(baby, mother, nurse, doctor, title)

At each birth, there is one baby (twins are represented by two births), one mother, any number of nurses, and any number of doctors. Each doctor has a title. For every baby there is a unique mother.

a) Indicate all functional and multi-valued dependencies that would be expected to hold in this relation. What is the key of the relation? (*6 pts.*)
b) Decompose relation Births into 4NF relations. (*4 pts.*)

III Finding Functional Dependencies (30 pts.)

You are given a ".csv" file which contains a table for a simple database. The table has some data anomalies due to redundancy. Your task is to load this table into PostgreSQL and identify the functional dependencies that cause anomalies by writing SQL statements. Once you identify the "bad" functional dependencies, your task is to normalize the table. Do the following:

1. Download and install PostgreSQL (if you haven't done so far) and download "sample.csv" file that is provided for you.

2. Create a table in the database and load the table with the data given in the given \.csv" file. You can use the \COPY" command of PostgreSQL. You can get detailed information from here.

3. Find all non-trivial functional dependencies in the table by writing appropriate SQL queries. Remember that a functional dependency is a constraint on a database instance. First, try to identify simple FDs like A → B, then try AB → C, etc. You should write an SQL query for each candidate FD. You can see if the FD holds or not by checking the answer of the query.

4. Decompose the table into BCNF tables using the FDs that you discovered. Create tables for normalized relations. Don't forget to create keys and foreign keys for the BCNF schema.

5. Load the new tables with the data from the original table. For this step, you should write SQL statements to load data into the new tables.

6. Dump created database into `eXXXXXXX.sql' file. You can dump the database using "pg dump". Detailed information can be found here.

What to turn in:

a) List of all FDs you identified and the corresponding SQL queries to discover them at the end of step 3 above.
b) List of all SQL statements to create normalized tables.
c) List of all SQL statements that load the contents of the tables.
d) Send "eXXXXXXX.sql" file along with the pdf. If there is no ".sql" file you will only get half the grade from this question. Also the dump version sometimes becomes corrupted, please check before sending.