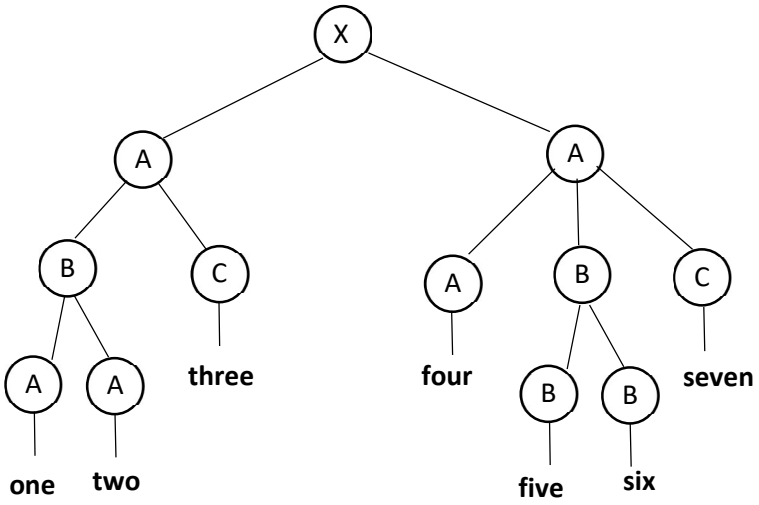
CENG 352
Database Management Systems
Spring 2023
Written Assignment 1

I. XML and JSON

1.1 XML

Consider the following data in the form of a tree.



a) Write down the XML document that describes this tree.

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

i. /X/A/*/text()
ii. //A[B/A]/C/text()
iii. /X/A[A]/B/*/text()
iv. //A/B/*/text()
v. /X/A[A][C]/B/*/text()
vi. //A[B]/B/*
vii. //A[A]//text()

1.2 JSON

Represent the three relations shown below as a single JSON document. (Try http://jsonlint.com/ for an easy way to check the validity of your JSON document.) There is more than one way to represent these relations in JSON. Begin your document with a collection of Customers, each containing a collection of products they ordered and the date at which those products are ordered. Then find a reasonable way to include products not ordered by any customer. In addition, explain whether this representation avoids redundancy.

Customers		
cid	**name**	**email**
1	John Smith	john@gmail.com
2	Jane Doe	jane@gmail.com
3	Bob Johnson	bob@yahoo.com

Orders			
id	**cid**	**pid**	**date**
1	1	100	2/15/23
2	1	101	2/18/23
3	2	101	2/20/23
4	2	102	2/22/23

Products		
pid	**pname**	**price**
100	iPhone14	999.00
101	Samsung S21	799.00
102	MacBookAir	1299.00
103	Bose Headphones	299.00
104	iPad Air	599.00

III. Database Design

3.1. BCNF Decomposition

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

{AB → CD, C → E, F → G, EF → A, AG → B}

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.

3.2 3NF Decomposition

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

{ ABC → CG, A → C, D → E, DE → G, DH → A, BH → G, CH → DE, DH → C}

(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.

3.3 Finding Functional Dependencies

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" property of PostgreSQL. You can get detailed information from here.

3. Find all 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 your will only get half the grade from this question. Also the dump version sometimes becomes corrupted, please check before sending.