Normalization: Normal Forms

Your company, Apasaja Private Limited, is commissioned by Toko Kopi Luwak to design relational schema for the management of their coffee beans, drinks, and cafes.

A coffee bean is fully identified by its unique brand name or a combination of its cultivar and region (since the same cultivar can be grown in different region). For instance, we may have a coffee bean named "The Waterfall" which comes from a Tabi cultivar grown in Colombia.

A drink can be made utilizing a particular coffee bean. The name of the drink is only unique for a particular coffee bean. This means, we can have an "Espresso" made with "The Waterfall" or "La Bella" (which is a Pacamara cultivar grown in Guatemala). The price of the drink is also recorded.

A branch identified by its branch name may then sell the drink. A drink may be sold by zero or more branches. A branch may sell zero or more drinks. Additionally, the address of the branch is also recorded. Lastly, for each drink sold by a branch, we record the quantity sold to see which branch is the most profitable.

We are only given an abstract schema for this application as follows.

$$R = \{A, B, C, D, E, F, G, H\}$$

$$\begin{aligned}\Sigma = \{\ & \{A\} \rightarrow \{C, E\},\ \ \{A, B\} \rightarrow \{D\},\ \ \{F\} \rightarrow \{H\},\ \ \{C, E\} \rightarrow \{A\},\ \ \{B, C, E\} \rightarrow \{D\},\\ & \{A, B, F\} \rightarrow \{D, G\},\ \ \{B, C, E, F\} \rightarrow \{G\}\ \}\end{aligned}$$

This tutorial continues from the computation of candidate keys and minimal cover in "Normalization: Functional Dependencies". You are advised to compute them before continuing.

Questions

Not all questions will be discussed during tutorial. You are expected to attempt them before coming to the tutorial. You may be randomly called to present your answer during tutorial. You are encouraged to discuss them on Canvas Discussion.

1. **Boyce-Codd Normal Form**.

   (a) Is R with Σ in 3NF?

   Comments:
   No.
   From "Normalization: Boyce-Codd Normal Form", we know that $\{A\} \rightarrow \{C\}$ is non-trivial and $\{A\}$ is not a superkey. Unfortunately, $\{C\}$ is a prime attribute. So, we have to look for other violations.
   Consider $\{A,B\} \rightarrow \{D\}$.

   - It is non-trivial (since $\{D\} \not\subseteq \{A,B\}$).
   - $\{A,B\}$ is not a superkey (since it is not a superset of either key $\{A,B,F\}$ or $\{B,C,E,F\}$).
   - D is not a prime attribute.

     Prime attributes are $\{A,B,C,E,F\}$.

   (b) Is R with Σ in BCNF?

   Comments:
   No.
   From Question 1a, we know that R with Σ is not in 3NF. Therefore, it cannot be in BCNF. However, let us verify this from the definition of BCNF. Obviously, we can consider $\{A,B\} \rightarrow \{D\}$, but let us consider a different functional dependency.
   Consider $\{A\} \rightarrow \{C\}$. This is a BCNF violation because:

   - It is non-trivial (since $\{C\} \not\subseteq \{A\}$), and
   - $\{A\}$ is not a superkey (since $\{A\}^+ = \{A,C,E\} \subset R$).

     $\{A\}$ is *also* not a superset of a key (keys are $\{A,B,F\}$ and $\{B,C,E,F\}$).

     $\Rightarrow$ This is a simpler way to check superkey if we have computed keys.

2. **Normalization**.

   (a) Decompose[1] R with Σ into a 3NF decomposition using the algorithm from the lecture.

   Comments:
   We can start from a canonical cover directly.

   $$\{\ \{A\} \rightarrow \{C,E\},\ \ \{F\} \rightarrow \{H\},\ \ \{C,E\} \rightarrow \{A\},\ \ \{B,C,E\} \rightarrow \{D\},\ \ \{B,C,E,F\} \rightarrow \{G\}\ \}$$

   For each functional dependency, we synthesize a fragment.

   $$\{\ \{A,C,E\},\ \{F,H\},\ \{A,C,E\},\ \{B,D,C,E\},\ \{B,C,E,F,G\}\ \}$$

   If there is any fragment that can be *subsumed*, we remove it from the result.

   $$\{\ \{A,C,E\},\ \{F,H\},\ \cancel{\{A,C,E\}},\ \{B,D,C,E\},\ \{B,C,E,F,G\}\ \}$$

   If none of the fragments contains a candidate key, we add one fragment for any key. Luckily, the key $\{B,C,E,F\}$ is a subset of $\{B,C,E,F,G\}$. So, we do not have to add another relation.

[1] Although it is a *synthesis* algorithm, the process is still decomposition.

$$\{\ \{A, C, E\},\ \{F, H\},\ \{B, D, C, E\},\ \{B, C, E, F, G\}\ \}$$

(b) Is the result dependency preserving?

Comments:
Yes. This is guaranteed by the algorithm.

(c) Is the result in BCNF?

Comments:
The 3NF synthesis only guarantees that the result is in 3NF, it may not be in BCNF. However, in this case, it is also in BCNF. We need to check each fragment. We leave the full check as an exercise given the projection below.

- $\{A, C, E\}$ with $\Sigma_1 = \{\ \{A\} \rightarrow \{C\}, \{A\} \rightarrow \{E\}, \{C, E\} \rightarrow \{A\}\ \}$
- $\{F, H\}$ with $\Sigma_2 = \{\ \{F\} \rightarrow \{H\}\ \}$
- $\{B, C, D, E\}$ with $\Sigma_3 = \{\ \{B, C, E\} \rightarrow \{D\}\ \}$
- $\{B, C, E, F, G\}$ with $\Sigma_4 = \{\ \{B, C, E, F\} \rightarrow \{G\}\ \}$

Note that it is not always the case that we are lucky to obtain a BCNF decomposition using 3NF synthesis, but it may happen.

Comments:
Note that **projection** is an error-prone process. We advise not to skip any steps when computing the projection. Otherwise, some functional dependencies that may be derived could be omitted by mistake. The 5 steps may be tedious, but they guarantee a correct result. Taking shortcuts does not guarantee correctness in all cases.

5 steps to compute projection of R with Σ onto X:

1. Find all subsets of attributes of X.
2. For each subset X', compute the closure (i.e., $\varphi_1 := \texttt{AttrClose}(X', \Sigma)$).
3. Keep only the relevant attributes (i.e., $\varphi_2 := \varphi_1 \cap X$).
4. Remove attributes that do not contribute new information (i.e., $\varphi_3 := \varphi_2 - X'$).
5. If φ_3 is not empty, form a functional dependency $X' \rightarrow \varphi_3$.

References

[1] S. Bressan and B. Catania. *Introduction to Database Systems*. McGraw-Hill Education, 2006. ISBN: 9780071246507.

[2] Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. *Database Systems: The Complete Book*. 2nd ed. Prentice Hall Press, 2008. ISBN: 9780131873254.

[3] Raghu Ramakrishnan and Johannes Gehrke. *Database Management Systems*. 2nd. USA: McGraw-Hill, Inc., 2000. ISBN: 0072440422.