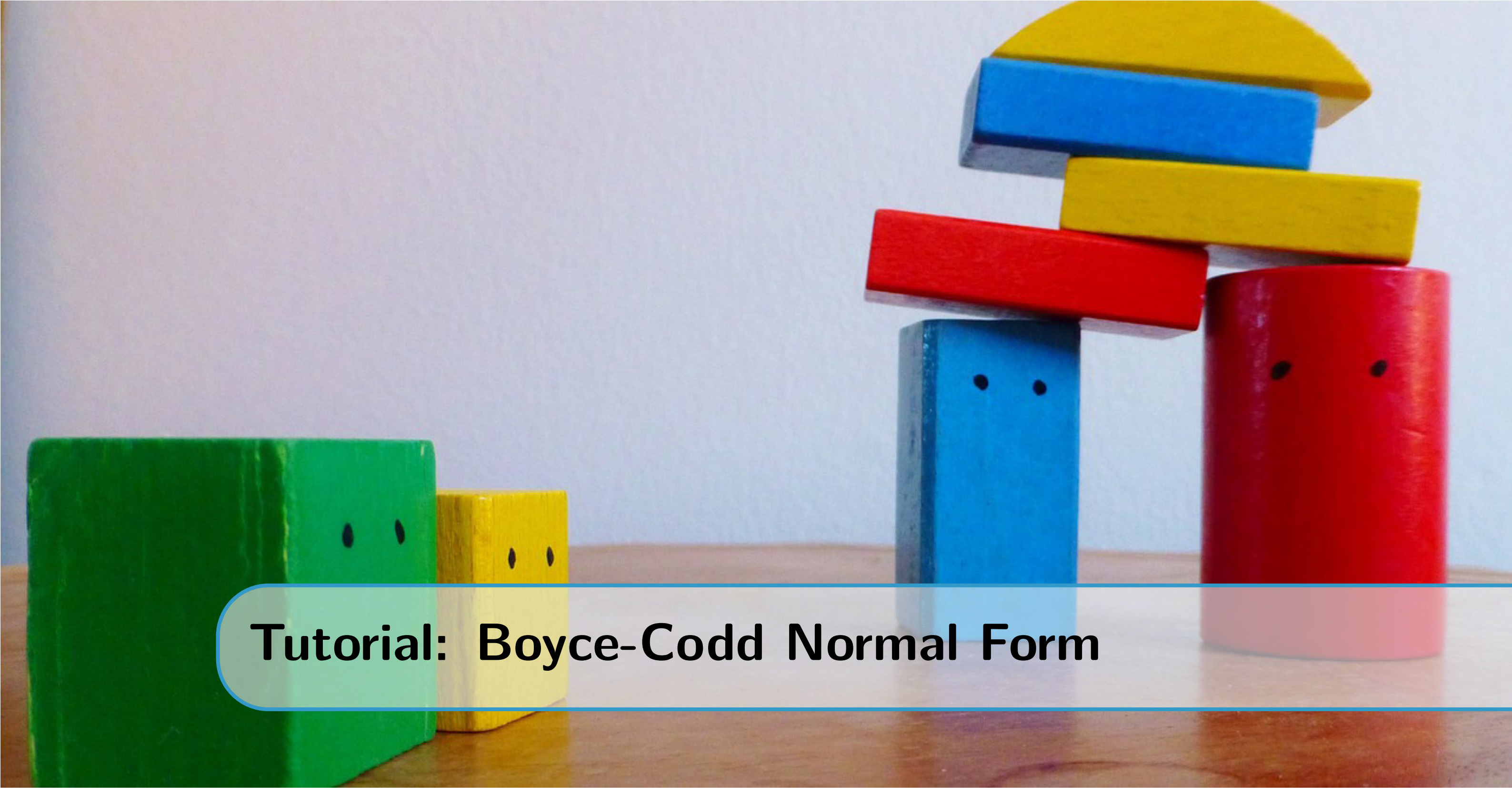

Tutorial: Boyce-Codd Normal Form

1. Consider the schema $R = \{A, B, C, D, E\}$ with the set of functional dependencies $\Sigma = \{\{A, B\} \rightarrow \{C\}, \{A, C\} \rightarrow \{D\}, \{E\} \rightarrow \{A, B, C, D\}\}$.

   Does the decomposition of R into $\delta = \{R_1(A, B, C), R_2(A, B, E), R_3(A, C, D)\}$ a lossless join decomposition? Justify your answer without using tableau method.

Solution: The decomposition of into more than two fragments is a lossless join decomposition if there is a sequence of lossless *binary* decompositions. Since there are 3 tables in the fragments, we need to perform binary decomposition twice where each decomposition from 1 table into 2 tables is a lossless join decomposition. As long as there is one possible way to do this, then the entire decomposition is lossless.

First, let us consider a failed attempt as you may potentially encounter some of these during your working.

1. Decompose $R(A, B, C, D, E)$ into $\{R_1(A, B, C), R_t(A, B, C, D, E)\}$.
   - Common attributes are $R_1 \cap R_t = \{A, B, C\} \cap \{A, B, C, D, E\} = \{A, B, C\}$.
   - $\{A, B, C\}^+ = \{A, B, C, D\} \supseteq R_1$, we have $\{A, B, C\} \rightarrow \{A, B, C\}$, therefore $\{A, B, C\}$ is the superkey of R_1 and the decomposition is lossless.
2. Decompose $R_t(A, B, C, D, E)$ into $\{R_2(A, B, E), R_3(A, C, D)\}$.
   - Common attributes are $R_2 \cap R_3 = \{A, B, E\} \cap \{A, C, D\} = \{A\}$.
   - $\{A\}^+ = \{A\}$ and neither $\{A\} \supseteq R_2$ nor $\{A\} \supseteq R_3$, then the decomposition is not a lossless join decomposition.

But this is just one possible decomposition, so we still need to consider decomposition. Now, let us consider a successful attempt. Note that this is not the only possible successful attempt.

1. Decompose $R(A, B, C, D, E)$ into $\{R_3(A, C, D), R_t(A, B, C, E)\}$.
   - Common attributes are $R_3 \cap R_t = \{A, C, D\} \cap \{A, B, C, E\} = \{A, C\}$.
   - $\{A, C\}^+ = \{A, C, D\} \supseteq R_3$, we have $\{A, C\} \rightarrow \{A, C, D\}$, therefore $\{A, C\}$ is the superkey of R_3 and the decomposition is lossless.
2. Decompose $R_t(A, B, C, E)$ into $\{R_1(A, B, C), R_2(A, B, E)\}$.
   - Common attributes are $R_1 \cap R_2 = \{A, B, C\} \cap \{A, B, E\} = \{A, B\}$.

- $\{A,B\}^+ = \{A,B,C,D\} \supseteq R_1$, we have $\{A,B\} \rightarrow \{A,B,C\}$, therefore $\{A,B\}$ is the superkey of R_1 and the decomposition is lossless.

3. Hence, the decomposition of R into $\delta = \{R_1, R_2, R_3\}$ is a lossless join decomposition.

2. Consider the schema $R = \{A,B,C,D,E\}$ with a set of functional dependencies $\Sigma = \{\{A\} \rightarrow \{E\}, \{A,B\} \rightarrow \{D\}, \{C,D\} \rightarrow \{A,E\}, \{E\} \rightarrow \{B\}, \{E\} \rightarrow \{D\}\}$.

(a) Is R in Boyce-Codd normal form with respect to Σ?

Solution: R is not in Boyce-Codd normal form with respect to Σ. Consider the following functional dependency $\{A\} \rightarrow \{E\}$. This violates the Boyce-Codd normal form property of R with respect to Σ because

- $\{E\} \not\subseteq \{A\}$ so $\{A\} \rightarrow \{E\}$ is non-trivial.
- $\{A\}^+ = \{A,B,D,E\}$ so $\{A\}$ is not the superkey of R.

Therefore, R is not in Boyce-Codd normal form with respect to Σ.

(b) Consider the following decomposition $\delta = \{R_1(B,D,E), R_2(A,C,E)\}$. Is δ in Boyce-Codd normal form with respect to Σ.

Solution: δ is not in Boyce-Codd normal form with respect to Σ. For δ to be in Boyce-Codd normal form with respect to Σ, all fragment must be in Boyce-Codd normal form with respect to Σ. Unfortunately, R_2 is not in Boyce-Codd normal form with respect to Σ.

- Consider $\{A\} \rightarrow \{E\}$ again. Since $\{E\} \not\subseteq \{A\}$ so $\{A\} \rightarrow \{E\}$ is non-trivial.
- $\{A\}^+ = \{A,B,D,E\}$ so $\{A\}$ is not the superkey of R_2.

Therefore R_2 is not in Boyce-Codd normal form with respect to Σ. So the decomposition is not a Boyce-Codd normal form decomposition with respect to Σ.

(c) If δ is not in Boyce-Codd normal form with respect to Σ, find a Boyce-Codd normal form decomposition of R with respect to Σ.

Solution: Since we know that δ is not in Boyce-Codd normal form with respect to Σ, we need to find a Boyce-Codd normal form decomposition. Using the algorithm from the lecture, we can arrive at the following decomposition.

- Given $\{A\}^+ = \{A,B,D,E\}$, we use $\{A\} \rightarrow \{A,B,D,E\}$ for decomposition of R into $R^1(A,B,D,E)$ and $R^2(A,C)$.
  - Given $\{E\}^+ = \{B,D,E\}$, we use $\{E\} \rightarrow \{B,D,E\}$ for decomposition R^1 into $R^3(B,D,E)$ and $R^4(A,E)$.
    - $R^3(B,D,E)$ is in Boyce-Codd normal form with respect to Σ.
    - $R^4(A,E)$ is in Boyce-Codd normal form with respect to Σ.
  - $R^2(A,C)$ is in Boyce-Codd normal form with respect to Σ.

Therefore, the answer involves R^2, R^3 and R^4. But it is good to rename them to use smallest number and subscript. So a possible Boyce-Codd normal form decomposition of R is

$$\delta_1 = \{R_1(A,C), R_2(B,D,E), R_3(A,E)\}$$

Note that Boyce-Codd normal form decomposition is not unique. There are choices that can be made on choosing the functional dependencies to be used for decomposition. Any functional dependencies that is a witness to a violation of the Boyce-Codd normal form property can be used. Generally, for the functional dependencies $X \rightarrow Y$ to be useful, we want X to be as small as possible and Y to be as large as possible. But this is merely a guideline.

Solution: The steps above are specific to the problem. In general, we want to compute the candidate keys and (*possibly*) the minimal cover. To compute the candidate keys, we use the two tips and tricks.

Since C is not on the right hand side of any functional dependencies, it must be in all keys.

Then we compute from the smallest cardinalities. The set of attributes must contain C.

- Singletons
  - $\{C\}^+ = \{C\}$ so $\{C\}$ is not a candidate key.
- Pairs
  - $\{A, C\}^+ = \{A, B, C, D, E\}$ so $\{A, C\}$ is a candidate key.
  - $\{B, C\}^+ = \{B, C\}$ so $\{B, C\}$ is not a candidate key.
  - $\{C, D\}^+ = \{A, B, C, D, E\}$ so $\{C, D\}$ is a candidate key.
  - $\{C, E\}^+ = \{A, B, C, D, E\}$ so $\{C, E\}$ is a candidate key.
- Triples. The set should not be a superset of $\{A, C\}$, $\{C, D\}$, or $\{C, E\}$ but should contain C. There are no such triples.

So the only candidate keys are $\{A, C\}$, $\{C, D\}$, and $\{C, E\}$. Using this, we can easily check if a functional dependency is a violation.

- Consider $\{C, D\} \rightarrow \{A, E\}$, we know $\{C, D\}$ is a superkey because it is a superset (*superset includes equality*) of a candidate key.
- Consider $\{A, B\} \rightarrow \{D\}$, we know $\{A, B\}$ is not a superkey because it is not a superset of any key.