Normalization

Mateusz Bulanda - Gorol

January 13, 2022

1 Identify all the functional relations in the database.

We have a table R with 13 columns (Country, LANGUAGE, area, national_day, country_code2, country_code3, Region, Continent, region_area, percent_of_region, YEAR, population, gdp) that we denote:

$$R = ABCDEFGHIJKLM$$

Now we can find the set of all functional dependencies:

$$F = \{A \to ACDEFGHIJ, C \to ACDEFGHIJ, E \to ACDEFGHIJ, F \to ACDEFGHIJ, G \to GHI, I \to GI, AB \to ABCDEFGHIJ, BC \to ABCDEFGHIJ, BE \to ABCDEFGHIJ, BF \to ABCDEFGHIJ, AK \to ACDEFGHIJKLM, CK \to ACDEFGHIJKLM, EK \to ACDEFGHIJKLM, FK \to ACDEFGHIJKLM, AL \to ACDEFGHIJKLM, CL \to ACDEFGHIJKLM, EL \to ACDEFGHIJKLM, FL \to ACDEFGHIJKLM, AM \to ACDEFGHIJKLM, CM \to ACDEFGHIJKLM, EM \to ACDEFGHIJKLM, FM \to ACDEFGHIJKLM, ABK \to ABCDEFGHIJKLM, CBK \to ABCDEFGHIJKLM, EBK \to ABCDEFGHIJKLM, FBK \to ABCDEFGHIJKLM, ABL \to ABCDEFGHIJKLM, CBL \to ABCDEFGHIJKLM, EBL \to ABCDEFGHIJKLM, FBL \to ABCDEFGHIJKLM, ABM \to ABCDEFGHIJKLM, CBM \to ABCDEFGHIJKLM, EBM \to ABCDEFGHIJKLM, FBM \to ABCDEFGHIJKLM\}$$

2 Find all the candidate keys of the database.

The set of key-candidate attribute can give us information about all attributes. The set of them is following:

$$\{ABK, CBK, EBK, FBK, ABL, CBL, EBL, FBL, ABM, CBM, EBM, FBM\}$$

Every element in above set is a belong to set of key attributes, so the set of key attributes is: $\{ABCEFKLM\}$, and the set of non-key attribute is: $\{DGHIJ\}$

3 Determine the highest normal form that is not violated. Normalize the database up to 3NF.

Let's simplify the set of functional dependencies:

$$F = \{A \to CDEFHGIJ, C \to ADEFGHIJ, E \to ACDFGHIJ, F \to ACDEGHIJ, G \to HI, I \to G, AK \to LM, CK \to LM, EK \to LM, FK \to LM, AL \to KM, CL \to KM, EL \to KM, FL \to KM, AM \to KL, CM \to KL, EM \to KL, FM \to KL, \}$$

As we can see, the relations: $\{A \to CDEFHGIJ, C \to ADEFGHIJ, E \to ACDFGHIJ, F \to ACDEGHIJ\}$ contain attributes D, G, I, J on the right hand side and they break $2NF$, so they are $1NF$. The relation $A \to D$ breaks the $2NF$ too. So if we split them, we get the result:

$$F_{1NF} = \{A \to CDEFHGIJ, C \to ADEFGHIJ, E \to ACDFGHIJ, F \to ACDEGHIJ\}$$

The $R_{1NF} = ABCDEFGHIJ$ an key candidate are: $\{A, C, E, F\}$. Moreover we can easly notice that they are $3NF$.

Now we need the second result of a split to the $2NF$ fuctional relations:

$$F_{2NF} = \{A \to CEF, C \to AEF, E \to ACF, F \to ACE, G \to HI, I \to G, AK \to LM, CK \to LM, EK \to LM, FK \to LM, AL \to KM, CL \to KM, EL \to KM, FL \to KM, AM \to KL, CM \to KL, EM \to KL, FM \to KL, \}$$

We can see that the candidate for the key are the same as in the previous exercise: $\{ABK, CBK, EBK, FBK, ABL, CB$ so they are the same as before splitting. In this case the $R_{2NF} = ABCEFKLM$ is the same as key attributes in previous exercise. The relations belong to F_{2NF} are $2NF$ because G and I don't belong to any key.

Now to find the $3NF$ we have to split the $G \to HI$ and $I \to G$ with the set F_{2NF}. In the result we have:

$$F_{3NF} = \{A \to CEF, C \to AEF, E \to ACF, F \to ACE, AK \to LM, CK \to LM, EK \to LM, FK \to LM, AL \to KM, CL \to KM, EL \to KM, FL \to KM, AM \to KL, CM \to KL, EM \to KL, FM \to KL, \}$$

The $R_{3NF} = ABCEFJKLM$ and the set of key-candidate is: $\{ABK, CBK, EBK, FBK, ABL, CBL, EBL, FBL, AB$ Here we haven't got any attribute which doesn't below to the key set, so these attributes are $3NF$.

Of course we have the second case of this split and in the result we get following set:

$$R^{'}_{3NF} = \{G ß HI, I ß G\}$$

This is of $3NF$ too.