# Normalization Theory sheet I

Gallo Tenis

February 6, 2025

*Exercises in normalization theory from J. Ullman's "Database Systems - The Complete Book". This week's topics are closures of functional dependencies sets, projections of FD's sets, minimal bases, BCNF and 3NF.*

## Section 3.2.

1. **3.2.5** Show that if a relation has no attribute that is functionally determined by all the other attributes, then the relation has no nontrivial FD's at all.
   **Solution.** Let $R$ be a relation schema. By hypothesis, for any set of attributes $X \in R$, $X^+ = X$. So if a nontrivial FD such as $X \longrightarrow Y$ existed, it would imply that $Y \nsubseteq X$, so $X^+ = XY$, a contradiction.

   □

2. **3.2.6** Let $X$ and $Y$ be sets of attributes. Show that if $X \subseteq Y$, then $X^+ \subseteq Y^+$, where the closures are taken with respect to the same set of FD's.
   **Solution.** Since $X \subseteq Y$ and $Y \subseteq Y^+$, then necessarily $X \subseteq Y^+$. Since $X \subseteq X^+$, we see that each element of $X^+ \cap X$ is in $Y^+$. With this in mind, we have shown the inclusion in the trivial DF's case.

   Now for the case of nontrivial dependencies $x \in X^+ \setminus X$ we have to show that $x \in Y^+$. Since $x$ is nontrivial then there exists some $z \in X$ such that $x \nsubseteq z$ and

   $$z \longrightarrow x.$$

   Since $z \in Y$ as well, we conclude that $x \in Y^+$.

   □

3. **3.2.7** Prove that $(X^+)^+ = X^+$.
   **Solution.** Since it is obvious that $X^+ \subseteq (X^+)^+$ (the closure of a FD set always contains itself), remains showing that $(X^+)^+ \subseteq X^+$.

   Suppose that there existed a FD $x \in (X^+)^+$ such that $x \notin X^+$. Then there exists an $y \in X^+$ such that

$$y \longrightarrow x.$$

So there must also exist a $z \in X$ such that

$$z \longrightarrow y,$$

hence, by transitivity $z \longrightarrow x$. We conclude that $x \in X^+$.

$\square$

4. **3.2.8** We say a set of attributes $X$ is *closed* (with respect to a given set of FD's) if $X^+ = X$. Consider a relation with schema $R(A, B, C, D)$ and an unknown set of FD's. If we are told which sets of attributes are closed, we can discover the FD's. What are the FD's if:

   (a) All sets of the four attributes are closed.

   (b) The only closed sets are $\varnothing$ and $\{A, B, C, D\}$.

   (c) The closed sets are $\varnothing$, $\{A, B\}$, and $\{A, B, C, D\}$.

   **Solution.**

(a) By exercise 3.2.5 $R$ has no non-trivial dependencies. So $\{A, B, C, D\}^+ = \{A, B, C, D\}$.

(b) We claim that there does exist at least one superkey of the relation. First we note that the unordered pairs of attributes is

$$\binom{4}{2} > 4,$$

so there must exist at least two pairs $X$, $Y$ such that $X \cap Y \neq \varnothing$ (by pigeonhole principle). In other words, since the closure of any attribute is at least two (they are not closed) we see that

$$X^+ = XZ \text{ and } Y^+ = YZ$$

or

$$X^+ = XZ \text{ and } Z^+ = ZY$$

So in the first case

$$X \longrightarrow Z \text{ and } Y \longrightarrow Z$$

then

$$XY \longrightarrow Z.$$

In fact, since neither $Z$ is closed, $Z \longrightarrow T$, hence by transitivity $XY \longrightarrow T$, we conclude that $(XY)^+ = XYZT = ABCD$.

In the second case, we get that $X$ is a superkey by transitivity.

So the FD's will look like $\{AB \longrightarrow C, C \longrightarrow D\}$ or $\{A \longrightarrow B, B \longrightarrow C, C \longrightarrow D, D \longrightarrow A\}$ with respect to both cases.

(c) Consider the sub relations $R_1 = \pi_{A,B}(R)$ and $R_2 = \pi_{C,D}(R)$. For $R_1$, since $A^+ \neq A$ and $B^+ \neq B$ but $(AB)^+ = AB$ it will suffice letting $A \longrightarrow B$ and $B \longrightarrow A$. In fact, this is the only way it can be done; as if we let $A \longrightarrow X$ with $X \subseteq R_2$ then

$$AB \longrightarrow A \longrightarrow X, \text{ so } (AB)^+ = ABX \neq AB.$$

For $R_2$ we have that $C^+ \neq C$ and $D^+ \neq D$ and that $(CD)^+ \neq CD$, since these are not closed. We then see that $CD$ is a superkey by seeing that since

$$CD \longrightarrow C,\ CD \longrightarrow D,\ C \longrightarrow X,\ D \longrightarrow Y \text{ (with } X \text{ and } Y \text{ not necessarily different),}$$

then

$$CD \longrightarrow XY.$$

And if $XY = X$, then since $(CD)^+ = CDX$, but since $CDX$ is not closed, by transitivity $(CD)^+ = CDXZ = ABCD$. Then, the FD's for $R_2$ will be of the form $\{C \longrightarrow D, D \longrightarrow A\}$, $\{C \longrightarrow A, D \longrightarrow B\}$ or $\{C \longrightarrow A, D \longrightarrow A, CD \longrightarrow B\}$.

□

5. **Exercise 3.2.11** Show that if an FD $F$ follows from some given FD's, then we can prove $F$ from the given FD's using Armstrong's axioms (defined in the box "A Complete Set of Inference Rules" in Section 3.2.7). *Hint:* Examine Algorithm 3.7 and show how each step of that algorithm can be mimicked by inferring some FD's by Armstrong's axioms.
**Solution.** Let $R = \{A_1, \ldots, A_n\}$ be a relation. Let $\{B_1, \ldots, B_m\}$ be a subset of $R$. Suppose $F$ had several attributes $\{C_1, \ldots, C_k\}$ on the right hand side, Algorithm 3.7 says $F$ holds even when splitting it into several one-attributed FD's, in other words

$$B_1, \ldots, B_m \longrightarrow F \iff$$

$$\begin{cases} B_1, \ldots, B_m \longrightarrow C_1 \\ \vdots \\ B_1, \ldots, B_m \longrightarrow C_k \end{cases}.$$

There are now some cases we are ought to study; first suppose that some $C_i$ were an element of $B_1, \dots, B_m$ for some $m \leq n$, then

$$\{C_i\} \subseteq \{B_1, \dots, B_m\} \implies B_1, \dots, B_m \longrightarrow C_i$$

More generally if we had a larger subset of $F$, say $\{C_1, \dots, C_q\}$ with $q \leq k$ and each element of it satisfying the above relation, then by union rule we can say that

$$\{C_1, \dots, C_q\} \subseteq \{B_1, \dots, B_m\} \implies B_1, \dots, B_m \longrightarrow C_1, \dots, C_q,$$

verifying reflexivity.

Let $\{B_1, \dots, B_m\} \longrightarrow C_i$ such that $C_i \longrightarrow C_j$ for some $i, j$. Then we see that

$$C_i \in \{B_1, \dots, B_m\}^+$$

so

$$B_1, \dots, B_m, C_i \longrightarrow C_j.$$

Meaning that $C_j \in \{B_1, \dots, B_m, C_i\}^+ = \{B_1, \dots, B_m\}^+$. In other words

$$B_1, \dots, B_m \longrightarrow C_i \text{ and } C_i \longrightarrow C_j \implies B_1, \dots, B_m \longrightarrow C_j,$$

verifying transitivity.

Lastly, since $B_1, \dots, B_m \longrightarrow F$, necessarily for any other subset $X$ of $R$, we have

$$B_1, \dots, B_m, X \longrightarrow B_1, \dots, B_m \longrightarrow F \text{ (by reflexivity)}$$

and

$$B_1, \dots, B_m, X \longrightarrow X. \text{ (by reflexivity)}$$

So by union rule

$$B_1, \dots, B_m X \longrightarrow FX,$$

verifying augmentation.

$\square$