

ON THE BOYCE - CODD NORMAL FORM FOR RELATION SCHEME

VU DUC THI, LUONG CAO SON

Abstract. The Boyce-Codd normal form (BCNF) is an essential normal form for relation schemes in the relational database. This normal form has been used in designing database systems. Keys and minimal keys are the important concepts of the relational datamodel. The set of minimal keys of relation scheme is Sperner system. In this paper we show a new necessary and sufficient conditions for an arbitrary relation scheme is in BCNF and its set of minimal keys is a given Sperner system.

1. INTRODUCTION

Now we start with some necessary definitions, and in the next sections we formulate our results.

Definition 1. Let $R = \{h_1, \dots, h_n\}$ be a relation over U , and $A, B \subseteq U$. Then we say that B functionally depends on A in R (denoted $A \xrightarrow[R]{f} B$) iff

$$(\forall h_i, h_j \in R)(\forall a \in A)(h_i(a) = h_j(a)) \Rightarrow (\forall b \in B)(h_i(b) = h_j(b)).$$

Let $F_R = \{(A, B) : A, B \subseteq U, A \xrightarrow[R]{f} B\}$. F_R is called the full family of functional dependencies of R . Where we write (A, B) or $A \rightarrow B$ for $A f/R B$ when R, f are clear from the context.

Definition 2. A functional dependency (FD) over U is a statement of the form $A \rightarrow B$, where $A, B \subseteq U$. The FD $A \rightarrow B$ holds in a relation R if $A \xrightarrow[R]{f} B$. We also say that R satisfies the FD $A \rightarrow B$.

Definition 3. Let U be a finite set, and denotes $P(U)$ its power set. Let $Y \subseteq P(U) \times P(U)$. We say that Y is an f -family over U iff for all $A, B, C, D \subseteq U$

- (1) $(A, A) \in Y$,
- (2) $(A, B) \in Y, (B, C) \in Y \Rightarrow (A, C) \in Y$,
- (3) $(A, B) \in Y, A \subseteq C, D \subseteq B \Rightarrow (C, D) \in Y$,
- (4) $(A, B) \in Y, (C, D) \in Y \Rightarrow (A \cup C, B \cup D) \in Y$.

Clearly, F_R is an f -family over U .

It is known [1] that if Y is an arbitrary f -family, then there is a relation R over U such that $F_R = Y$.

Definition 4. A relation scheme S is a pair $\langle U, F \rangle$, where U is a set of attributes, and F is a set of FDs over U . Let F^+ be a set of all FDs that can be derived from F by the rules in Definition 3.

Clearly, in [1] if $S = \langle U, F \rangle$ is a relation scheme, then there is a relation R over U such that $F_R = F^+$. Such a relation is called an Armstrong relation of S .

Definition 5. Let R be a relation over U , $S = \langle U, F \rangle$ be a relation scheme, Y be an f -family over U , and $A \subseteq U$. Then A is a key of R (a key of S , a key of Y) if $A \xrightarrow[R]{f} U$ ($A \rightarrow U \in F^+$, $(A, U) \in Y$). A is a minimal key of $R(S, Y)$ if A is a key of $R(S, Y)$ and any proper subset of A is not a key of $R(S, Y)$. Denote $K_R, (K_S, K_Y)$ the set of all minimal keys of $R(S, Y)$.

Clearly, K_R, K_S, K_Y are Sperner systems over U .

Definition 6. Let K be a Sperner system over U . We define the set of antikeys of K , denote by K^{-1} , as follows:

$$K^{-1} = \{A \subseteq U : (B \in K) \Rightarrow (B \not\subseteq A) \text{ and } (A \subseteq C) \Rightarrow (\exists B \in K)(B \subseteq C)\}.$$

It is easy to see that K^{-1} is also a Sperner system over U .

It is known [4] that if K is an arbitrary Sperner system plays the role of the set of minimal keys (antikeys), then this Sperner system is not empty (does't contain U). We also regard the comparison of two attributes to be the elementary step of algorithms. Thus, if we assume that subsets of U are represented as sorted list of attributes, then a Boolean operation on two subsets of requires at most $|U|$ elementary steps.

Definitions 7. Let $I \subseteq P(U)$, $U \in I$, and $A, B \in I \Rightarrow A \cap B \in I$. Let $M \subseteq P(U)$. Denote $M^+ = \{\cap M' : M' \subseteq M\}$. We say that M is a generator of I iff $M^+ = I$. Note that $U \in M^+$ but not in M , since it is the intersection of the empty collection of sets.

Denote $N = \{A \in I : A \neq \cap \{A' \in I : A \subseteq A'\}\}$.

In [6] it is proved that N is the unique minimal generator of I . Thus, for any generator N' of I we obtain $N \subseteq N'$.

Definition 8. Let R be a relation over U , and E_R the equality set of R , i.e. $E_R = \{E_{ij} : 1 \leq i < j \leq |R|\}$, where $E_{ij} = \{a \in U : h_i(a) = h_j(a)\}$. Let $T_R = \{A \in P(U) : \exists E_{ij} = A, \text{ no } \exists E_{pq} : A \subseteq E_{pq}\}$. Then T_R is called the maximal equality system of R .

Definition 9. Let R be a relation, and K a Sperner system over U . We say that R represents K if $K_R = K$.

The following theorem is known in [8].

Theorem 1. Let K be a relation, and K a Sperner system over U . We say that R presents K iff $K^{-1} = T_R$, where T_R is the maximal equality system of R .

Let $s = \langle U, F \rangle$ be a relation scheme over U . From s we construct $Z(s) = \{X^+ : X \subseteq U\}$, and compute the minimal generator N_s of $Z(s)$. We put $T_s = \{A : A \in N_s, \text{ no } \exists B \in N_s : A \subseteq B\}$. In [8] we presented the following result.

Proposition 1. Let $s = \langle U, F \rangle$ be a relation scheme over U . Then

$$K_s^{-1} = T_s.$$

Definition 10. Let $s = \langle R, F \rangle$ be a relation scheme over R . We say that an attribute a is prime if it belong to a minimal key of s , and nonprime otherwise. $s = \langle R, F \rangle$ is in the Boyce-Codd normal form if $A \rightarrow \{a\} \notin F^+$ for $A^+ \neq R, a \notin A$.

If a relation scheme is changed to a relation we have the definition of BCNF for relation.

2. RESULTS

In this section we show the following result. It is a new necessary and sufficient condition for an arbitrary relation scheme is in BCNF and its set of minimal keys is given Sperner system.

First we denote some following concepts.

Let K be a Sperner over U . Denote $T(K^{-1}) = \{A : \exists B \in K^{-1} : A \subseteq B\}$. $K_n = \{a \in U | \text{no } \exists A \in K, a \in A\}$. K_n is called the set of nonprime attributes of K . Then we have

Theorem 1. Let $s = \langle U, F \rangle$ be a relation scheme, K a Sperner system over U . Denote $K^{-1} = \{B - a : a \in B, B \in K^{-1}\}$ and $M^* = \{C : C \in K^{-1}, C \neq \cap \{B : B \in K^{-1}, C \subseteq B\}\}$. Then s is in BCNF and $K_s = K$ if and only if

$$\{U\} \cup K^{-1} \cup M^* \subseteq \{U\} \cup T(K^{-1}). \quad (*)$$

Proof. Assume that s is in BCNF and $K = K_s$. By Proposition 1 and from definitions of $Z(s)$, $T(K^{-1})$ we have the right-hand side of (*). Based on Proposition 1 we can see that $\{U\} \cup K^{-1} \subseteq$

$Z(s)$. According to definition of BCNF, if s is in BCNF then for every $B \in K^{-1}$, $B - a \in Z(s)$. Consequently, we obtain $M^* \subseteq Z(s)$.

Assume that we have (*). By Proposition 1 and according to definitions of $Z(s)$, $T(K^{-1})$, K^{-1} we obtain $K_s = K$. If there exists an $A \rightarrow \{a\} \in F^+$ and $A^+ \neq R$ and $a \notin A$. According to Proposition 1 there exists a $B \in K_n^{-1}$ such that $A^+ \subseteq B$. Clearly, $a \in B$ and $A \subseteq B - a$ hold. Consequently, we have $(B - a)^+ = B$. This is a contradiction. Thus, s is in BCNF. The proof is complete.

Clearly, the right-hand side and the left-hand side of (*) don't depend on s .

Note that if $s_1 = \langle R, F_1 \rangle$ and $s_2 = \langle R, F_2 \rangle$ are in BCNF then $K_{s_1} = K_{s_2}$ holds iff $F_1^+ = F_2^+$ holds.

From definition of $Z(s)$ and according to Proposition 1 and Theorem 1 the following corollary is obvious.

Corollary 1. *Let $s = \langle R, F \rangle$ be a relation scheme. Then s is in BCNF iff $\forall B \in K_s^{-1}$, $a \in B : (B - a)^+ = B - a$.*

REFERENCES

- [1] Armstrong W. W., Dependency Structures of Database Relationships, Information Processing 74, Holland Publ. Co., 580-583, 1974.
- [2] Beeri C., Bernstein P. A., Computational problems related to the design of normal form relational schemes, *ACM Trans on Database Syst.* **4** (1) (1979) 30-59.
- [3] Beeri C., Dowd M., Fagin R., Staman R., On the structure of Armstrong relations for functional dependencies, *J. ACM* **31** (1) (1984) 30-46.
- [4] Demetrovics J., Logical and structural investigation of relational datamodel, *MTA-SZTAKI Tanulmányok*, Budapest, **114** (1980) 1-97.
- [5] Demetrovics J., Thi V. D., Some results about functional dependencies, *Acta Cybernetica* **8** (3) (1988) 273-278.
- [6] Demetrovics J., Thi V. D., Relations and minimal keys, *Acta Cybernetica* **8** (3) (1998) 279-285.
- [7] Demetrovics J., Thi V. D., On keys in the relation datamodel, *Inform. Process Cybern. EIK* **24** (10) (1988) 515-519.
- [8] Demetrovics J., Thi V. D., Algorithm for generating Armstrong relations and inferring functional dependencies in the relational datamodel, *Computers and Mathematics with Applications*, Great Britain, **26** (4) (1993) 43-45.
- [9] Demetrovics J., Thi V. D., Some problems concerning Keys for relation schemes and relation in the relational datamodel, *Information Processing Letters*, North Holland, **46** (4) (1993) 179-183.
- [10] Demetrovics J., Thi V. D., Some computational problems related to the functional dependency in the relational datamodel, *Acta Scientiarum Mathematicarum* **57** (1-4) (1993) 627-628.
- [11] Demetrovics J., Thi V. D., Armstrong relation, functional dependencies and strong dependencies, *Comput. and AI* (submitted for publication).
- [12] Thi V. D., Investigation on Combinatorial Characterization Related to Functional Dependency in Relational Datamodel, Ph. D. dissertation, *MTA-SZTAKI Tanulmányok*, Budapest, **191** (1986) 1-157.
- [13] Thi V. D., Minimal keys and antikeys, *Acta Cybernetica* **7** (4) (1986) 361-371.
- [14] Demetrovics J., Thi V. D., Some results about normal forms for functional dependency in the relational datamodel, *Discrete Applied Mathematics* **69** (1996) 61-74.

Received June 1, 1998

Vu duc Thi - Institute of Information Technology
 Luong Cao Son - Informatic Centre of Office of Government