Serdica

Bulgariacae mathematicae publicationes

Сердика

Българско математическо списание

SOME ADDITIONAL PROPERTIES OF KEYS FOR RELATION SCHEMES

J. DEMETROVICS, HO THUAN

In this paper we prove some additional properties of keys and superkeys for relation schemes. Some of them and their variants have been proved, perhaps by different methods, and used to design an algorithm to find all keys for any relation scheme [3].

Introduction. In [1] some characteristic properties of keys for a given relation scheme $S=\langle\Omega, F\rangle$ have been investigated, in particular the necessary condition under which a subset X of Ω is a key.

In this paper we prove some additional properties of keys and superkeys for relation scheme. Some of them and their variants have been proved, perhaps by different methods, and used to design an algorithm to find all keys for any relation scheme [3].

The notation used here is the same as in [1] and [2]. The reader is required to know the basic notation of the relational model and functional dependency [4].

1. In this section we recall some notions and results which will be needed in sequel.

Let $S=\langle\Omega, F\rangle$ be a relation scheme, where $\Omega=\{A_1, A_2, \dots, A_n\}$ is the universe of attributes, and $F=\{L_i\to R_i \mid i=1, 2, \dots, k;\ L_i, R_i\subseteq\Omega\}$ is the set of functional dependencies (FDs). A relation R defined over the attributes $\Omega=\{A_1, A_2, \dots, A_n\}$ is said to be an instance of the relation scheme $S=\langle\Omega, F\rangle$ iff each functional dependency $f\in F$ holds in R.

Let us denote:

$$L=\bigcup_{i=1}^{k} L_i, \quad R=\bigcup_{i=1}^{k} R_i$$

$$C_i=\Omega\setminus L_i^+, \quad i=1, 2, \dots, k.$$

$\mathcal{J}=\{i \mid$ there is no j such that $L_i\supset L_j\}\subseteq\{1, 2, \dots, k\}$. Recall that for $X\subset\Omega$, $X^+=\{A \mid (X\to A)\in F^+\}$ is the closure of X w. r. t. F, where F^+ is the closure of F, i. e. the set of all FDs that can be inferred from the FDs in F by repeated application of Armstrong's axioms [5]. Without loss of generality, in this paper we assume that

$$L_i\cap R_i=\emptyset, \quad i=1, 2, \dots, k;\ L_i\neq L_j \text{ with } i\neq j;\ L\cup R=\Omega.$$

We have the following lemmas:

Lemma 1. [2]. *Let $S=\langle\Omega, F\rangle$ be a relation scheme, $X, Y\subseteq\Omega$, then $(X^+Y)^+ =(XY^+)^+=(XY)^+$.*

Lemma 2. [2]. *Let K be a key for $S=\langle\Omega, F\rangle$. Then $Z^+\cap(K\setminus Z)=\emptyset$ for all $Z\subseteq K$.*

Lemma 3. [2]. *For any $i\in\mathcal{J}$, L_i is a key for S iff $C_i=\emptyset$.*

2. We are now in a position to prove some properties of keys and superkeys for relation scheme which can be used for the design of algorithms to find the keys for a relation scheme.

Theorem (key representation). *Let $S=\langle\Omega, F\rangle$ be a relation scheme. Then any key K for S has the following form* $K=L_iX_i$ where $X_i\subseteq C_i$, $i\in\mathscr{I}$.

Proof. Let be given $K\in\mathscr{K}_S$ — the set of all keys for S. If $K=\Omega$, then obviously $K=L_iX_i$ $\forall i\in\mathscr{I}$. If $K\subset\Omega$, then by the algorithm to find the closure K^+ of K w. r. t. F, (see [4]), there exists L_j such that $L_j\subseteq K$.

Consequently, $\exists i\in\mathscr{I}$ such that $L_i\subseteq K$. Thus $K=L_iX_i$, $i\in\mathscr{I}$.

Now we have to prove that $X_i\subseteq C_i$. By Lemma 1, we have

(1) $$L_i^+X_i\subseteq(L_i^+X_i)^+=(L_iX_i)^+=K^+=\Omega=L_i^+C_i.$$

By Lemma 2:

$$L_i^+\cap(K\setminus L_i)=L_i^+\cap X_i=\emptyset.$$

On the other hand, it is clear that $L_i^+\cap C_i=\emptyset$. Hence, from (1) we have: $X_i\subseteq C_i$ The proof is complete.

Remark 1. This theorem can be considered as an immediate corollary of Theorem 2.2 [2]. But for our purposes, the representation given here is more appropriate.

Lemma 4. *Let $S=\langle\Omega, F\rangle$ be a relation scheme. If $C_i\neq\emptyset$ then with $j\neq i$ $C_i\cap L_jC_j\neq\emptyset$.*

Proof. Assume the contrary that $C_i\cap L_jC_j=\emptyset$. In that case, it is easy to see that: (see Fig. 1) $C_i\subseteq L_j^+\setminus L_j$ and $L_jC_j\subseteq L_i^+$. Thus we have $L_i^+\xrightarrow{*}L_jC_j$.*

This contradicts the hypothesis of the lemma stating that L_i is not a superkey (because $C_i\neq\emptyset$). The Lemma is proved.

Property 1. *Let $S=\langle\Omega, F\rangle$ be a relation scheme, and suppose that $|C_i|\leq 1$, $\forall i\in\mathscr{I}$. Then for any $i\in\mathscr{I}$, L_iC_i is a key for S iff there is no $j\in\mathscr{I}$ such that $L_iC_i\supset L_jC_j$.*

Proof. If L_iC_i is a key for S with $i\in\mathscr{I}$, then it is impossible to have a j such that $L_iC_i\supset L_jC_j$, because L_jC_j is itself a superkey for S.

We have only to consider the case $|C_i|=1$. (Otherwise by Lemma 3, L_i is a key). Suppose that L_iC_i contains strictly a key

$$K=L_i'C_i\subset L_iC_i,\quad L_i'\subset L_i.$$

By the theorem just proved above

$$L_i'C_i=L_jX_j,\quad j\in\mathscr{I},\ j\neq i.$$

The case $X_j=\emptyset$, by Lemma 3, it is clear that $C_j=\emptyset$ and $L_iC_i\supset L_i'C_i=L_jC_j$ (with $C_j=\emptyset$). The case $X_j\neq\emptyset$, obviously $X_j=C_j$. But this shows (in both cases) that $L_iC_i\supset L_jC_j$, a contradiction.

Remark 2. Property 1 is still true if the set $\mathscr{I}$ is replaced by the set $\{1, 2, \ldots, k\}$.

Corollary 1. *Let $S=\langle\Omega, F\rangle$ be a relation scheme with $|C_i|\leq 1$, $\forall i\in\mathscr{I}$. If for any $i\in\mathscr{I}$, $C_i\cap L_j=\emptyset$ $\forall j\neq i$, then L_iC_i is a key for S.*

Proof. From the conditions of corollary 1, it is evident that there is no $j\in\mathscr{I}$ such that $L_iC_i\supset L_jC_j$. Indeed, were this false and there exists a $j\in\mathscr{I}$ such that $L_iC_i\supset L_jC_j$. Since $C_i\notin L_j$ and $L_j\subset L_iC_i$, we have $L_j\subset L_i$, a contradiction. By property 1, L_iC_i is a key for S.

Property 2. *Let $S=\langle\Omega, F\rangle$ be a relation scheme. For any $i\in\{1, 2, \ldots, k\}$, if $|C_i|=1$ and $L_i\cap R_j=\emptyset$, $\forall j\neq i$ then L_iC_i is a key for S.*

Proof. From $L_i\cap R_j=\emptyset$, $\forall j\neq i$, it follows that $L_i\cap R=\emptyset$. Thus $L_i\subseteq L\setminus R$. Moreover [1]

$$L\setminus R\subseteq K,\quad\forall K\in\mathscr{K}_S.$$

* In the following instead of $(X\to Y)\in F^+$, $X\cup Y$ we write $X\xrightarrow{*}Y$ and XY, respectively.

On the other hand, L_iC_i is a superkey for S and $L_i^+ \subset \Omega$. This shows that L_iC_i is a key for S.

Remark 3. Let $S=\langle\Omega, F\rangle$ be a relation scheme. If $|C_i|=1$ then C_i is a subset of Ω which consists of a single prime attribute. Indeed, since L_iC_i is a superkey for S and $L_i^+ \subset \Omega$ (because $C_i \neq \emptyset$), it follows that L_iC_i must contain a key K which has the form $K=L_i'C_i$ where $L_i' \subseteq L_i$. This shows that C_i is a prime attribute.

Property 3. *Let $S=\langle\Omega, F\rangle$ be a relation scheme. Then $\forall i\neq j$, $i, j \in \{1, 2, \ldots, k\}$ $L_i(C_i \cap L_jC_j)$ is a superkey for S.*

Proof. The case $C_i=\emptyset$, we have $L_i(C_i \cap L_jC_j)=L_i$. But in that case, it is obvious that L_i is a superkey. We now consider the case $C_i \neq \emptyset$. By Lemma 4 we have $C_i \cap L_jC_j \neq \emptyset$.

It is clear that $L_i \xrightarrow{*} L_i^+$, $C_i \cap L_jC_j \xrightarrow{*} C_i \cap L_jC_j$. Consequently,

$$L_i(C_i \cap L_jC_j) \xrightarrow{*} L_i^+(C_i \cap L_j)(C_i \cap C_j).$$

On the other hand, we have: (see Fig. 2)

$$L_j=(L_j \setminus C_i)(C_i \cap L_j) \subseteq L_i^+(C_i \cap L_j),$$

$$C_j=(C_j \setminus C_i)(C_j \cap C_i) \subseteq L_i^+(C_i \cap C_j).$$

Hence $L_jC_j \subseteq L_i^+(C_i \cap L_j)(C_i \cap C_j)$.

Finally we have $L_i(C_i \cap L_jC_j) \xrightarrow{*} L_jC_j$ showing that $L_i\ (C_i \cap L_jC_j)$ is a superkey for S

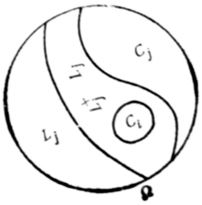

Fig. 1

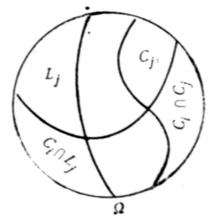

Fig. 2

Remark 4. The case $|C_i|=1$ we have: $L_i(C_i \cap L_jC_j)=L_iC_i$, $j\neq i$.

Remark 5. Let L_iX and L_jY be superkeys for S, $i\neq j$. In general $L_i(X \cap L_jY)$ is not a superkey. Let us consider the relation scheme:

$$\Omega=\{1\ 2\ 3\ 4\ 5\ 6\ 7\},$$

$$F=\{12 \to 36,\ 64 \to 52,\ 23 \to 17\}$$

It is easy to verify that 641 and 234 are superkeys.

$$(641)^+=6415237=\Omega,$$

$$(234)^+=2341765=\Omega.$$

On the other hand we have: $64(1 \cap 234)=64$, and $(64)^+=6452\neq\Omega$.

Property 4. *Let $S=\langle\Omega, F\rangle$ be a relation scheme. If $C_i\neq\emptyset$, $C_j\neq\emptyset$ then $T=L_i[(C_i \cap L_jC_j) \cap L_j(C_j \cap L_hC_h)]$ is a superkey for S, where $i\neq j$, $j\neq h$.*

Proof. By Property 3, it is clear that $L_j\ (C_j \cap L_hC_h)$ is a superkey for S.
Let us denote

$$Y=L_i^+[(C_i \cap L_jC_j) \cap L_j(C_j \cap L_hC_h)]=L_i^+(C_i \cap L_jC_j) \cap L_i^+L_j(C_j \cap L_hC_h).$$

Obviously $T \xrightarrow{*} Y$ and

$$L_i^+ L_j(C_j \cap L_h C_h) \supseteq L_j(C_j \cap L_h C_h).$$

On the other hand, $L_j C_j \subseteq L_i^+(C_i \cap L_j C_j)$, (From the proof of Property 3). It follows that $L_i^+(C_i \cap L_j C_j) \supseteq L_j(C_j \cap L_h C_h)$.
Hence $T \xrightarrow{*} Y \xrightarrow{*} L_j\ (C_j \cap L_h C_h)$ showing that T is a superkey for S.

Corollary 2. *With the same conditions as in Property 4, we have:*

$$(C_i \cap L_j C_j) \cap L_j(C_j \cap L_h C_h) \neq \emptyset.$$

Property 5. *Let K be any key for $S = \langle \Omega, F \rangle$ having the form $K = L_i X$, $X \subset C_i$. Then there exists $j_0 \neq i$ such that $K \subseteq L_i\ (C_i \cap L_{j_0} C_{j_0})$.*

Proof. Assume the contrary that $L_i X \nsubseteq L_i(C_i \cap L_j C_j)$, $\forall j \neq i$, or, equivalently $X \nsubseteq C_i \cap L_j C_j\ \forall j \neq i$. Then, for all $j \neq i$ there exists an attribute

$$A_{i_j} \in (L_j^+ \setminus L_j) \cap X \text{ (see Fig. 3)},$$

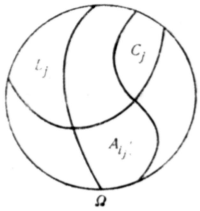

Fig. 3

Obviously we have: $L_i X \xrightarrow{*} L_i R_i X$. Then there must exist p such that $L_p \subseteq L_i R_i X$ (otherwise $L_i X \overset{*}{\leftrightarrow} \Omega$, a contradiction). Let $A_{i_p} \in (L_p^+ \setminus L_p) \cap X$ and let $X' = X \setminus \{A_{i_p}\}$. Since $A_{i_p} \notin L_p$, so $L_p \subseteq L_i R_i X'$. Therefore, it is easy to see that

$$L_i X' \xrightarrow{*} L_i R_i X' \xrightarrow{*} L_i R_i L_p R_p X' \xrightarrow{*} L_i R_i L_p^+ X'.$$

Moreover, $A_{i_p} \in L_p^+$. Consequently $L_i X' \xrightarrow{*} L_i X \xrightarrow{*} \Omega$, showing that $L_i X$ is not a key, a contradiction. The proof is complete.

Corollary 3. *The family $\{L_i(C_i \cap L_j C_j) \mid j \neq i,\ 1 \leq i,\ j \leq k\}$ can be used to find all keys for the relation scheme S.*

Property 6. *Let $S = \langle \Omega, F \rangle$ be a relation scheme. If $L_i(C_i \cap L_j C_j) = L_i C_i\ \forall j \neq i$, then either $L_i C_i$ is the unique key for S including L_i, or S has no key of the form $K = L_i X$ with $X \subseteq C_i$.*

This means that if $L_i C_i$ contains strictly any key of the form $L_i X$ then there exists $j_0 \neq i$ such that

$$L_i(C_i \cap L_{j_0} C_{j_0}) \subset L_i C_i.$$

In other words, $L_i C_i$ is a key for S iff

$$L_i(C_i \cap L_j C_j) = L_i C_i \quad \forall j \neq i$$

and there is no key of the form $K = L_i' X$ contained in $L_i C_i$ with $L_i' \subset L_i$, $X \subseteq C_i$.

Proof. Since $C_i \cap L_j C_j = C_i\ \forall j \neq i$, it follows that

$$C_i \cap (L_j^+ \setminus L_j) = \emptyset \quad \forall j.$$

Therefore if $A \in C_i$ then $\{A\} \cap (L_j^+ \setminus L_j) = \emptyset$ $\forall j$. Let A be any element of C_i, $A \in C_i$ and $X = C_i \setminus \{A\}$. It is easy to see that $L_i X \xrightarrow{*} L_i R_i X$. Since $L_i R_i \cap C_i = \emptyset$ (because $L_i R_i \subseteq L_i^+$), $A \in C_i$, $A \notin X$, it follows that $A \notin L_i R_i X$.

Now suppose that there exists $L_h \subseteq L_i R_i X$, $h \neq i$. Obviously $A \notin L_h$ and $L_i X$, $\xrightarrow{*} L_i R_i X \xrightarrow{*} L_i R_i L_h R_h X$. It is clear that $A \notin R_h$, otherwise $A \in (L_h^+ \setminus L_h)$, a contradiction. By repeating the same reasoning, we can prove that $L_i X \overset{*}{\leftrightarrow} \Omega$. This shows that either $L_i C_i$ is a key for S, or S has no key of the form $L_i X$, $X \subseteq C_i$. The proof is complete.

Corollary 4. *Let $S = \langle \Omega, F \rangle$ be a relation scheme. If $L_i(C_i \cap L_j C_j) = L_i C_i$ $\forall j \neq i$ and $L_i \cap R_j = \emptyset$, $\forall j \neq i$ then $L_i C_i$ is a key for S.*

Proof. The proof follows immediately by the application of Property 6 and by the fact that in this case L_i is contained in any key for S. [1] $(L_i \setminus R = L_i \subseteq L \setminus R \subseteq K$, $\forall K \in \mathcal{K}_S)$.

Remark 6. The case $L_i C_i$ is a key for S, we have:

$$T = L_i [(C_i \cap L_j C_j) \cap L_j (C_j \cap L_h C_h)] = L_i C_i.$$

Acknowledgement. The authors thank Vu Duc Thi for reading the paper and providing useful, comments.

REFERENCES

1. Ho Thuan, Le van Bao. Some results about keys of relational schemes. *Acta Cybernetica*, 7 1985, Fasc. 1, 99-113.
2. J. Demetrovics, Ho Thuan et al. Balanced relation scheme and the problem of key representation. MTA-SZTAKI Közlemenyek, 32, 1985, 51-80.
3. M. C. Fernandez. Determining the normalization level of a relation on the basis of Armstrong's axioms. *Computers and Artificial Intelligence*, 3, 1984, No. 6, 495-504.
4. J. D. Ullman. Principles of Database systems. 2 ed. Computer Science Press, 1982.
5. W. W. Armstrong. Dependency structures of database relationships. In: Information Processing'74. Amsterdam, 1974, 580-583.

Computer and Automation Institute
Hungarian Academy of Sciences
Hungary

Received 20. 3. 1986