DOI: 10.19884/j.1672-5220.202503008

Features of Prime Attributes in a Relation Scheme

XU Yulu, LIU Guohua*, YU Xiaoxue, LIU Changqi, ZHU Dongyan, ZHANG Limeng, HE Songda
School of Computer Science and Technology, Donghua University, Shanghai 201620, China

Abstract: Normal forms have a significant role in the theory of relational database normalization. The definitions of normal forms are established through the functional dependency (FD) relationship between a prime or nonprime attribute and a key. However, determining whether an attribute is a prime attribute is a nondeterministic polynomial-time complete (NP-complete) problem, making it intractable to determine if a relation scheme is in a specific normal form. While the prime attribute problem is generally NP-complete, there are cases where identifying prime attributes is not challenging. In a relation scheme $R(U, F)$, we partition U into four distinct subsets based on where attributes in U appear in F: U_1 (attributes only appearing on the left-hand side of FDs), U_2 (attributes only appearing on the right-hand side of FDs), U_3 (attributes appearing on both sides of FDs), and U_4 (attributes not present in F). Next, we demonstrate the necessary and sufficient conditions for a key to be the unique key of a relation scheme. Subsequently, we illustrate the features of prime attributes in U_3 and generalize the features of common prime attributes. The findings lay the groundwork for distinguishing between complex and simple cases in prime attribute identification, thereby deepening the understanding of this problem.

Keywords: normalization; key; prime attribute; nonprime attribute
CLC number: TP311.1 **Document code**: A
Article ID: 1672-5220(2025)06-0689-10

Open Science Identity (OSID)

0 Introduction

Codd[1-2] laid a foundation for relational database normalization theory by defining certain normal forms for a relation scheme. These normal forms served as clear objectives for database administrators to strive for. The first normal form, also defined by Codd, restricts attribute values but is unrelated to functional dependencies (FDs). The definitions of the second, third, and Boyce-Codd normal forms are based on the relationship of FD between a particular set of attributes and an attribute.

A relation scheme R is in second normal form (2NF) if every nonprime attribute is fully functionally dependent on each key. Similarly, it is in third normal form (3NF) if no nonprime attribute is transitively functionally dependent on any key. Lastly, R is in Boyce-Codd normal form (BCNF) if, for all disjoint nonempty sets of attributes X and Y in R, whenever X functionally determines Y, X must be a superkey of R.

To determine the normal form of a relation scheme (2NF, 3NF, or BCNF), one must first identify all keys and distinguish between prime and nonprime attributes. However, determining whether an attribute is a prime attribute is a nondeterministic polynomial-time complete (NP-complete) problem. Additionally, the number of keys can grow factorially or exponentially, depending on the number of attributes and FDs[3-5]. Therefore, the prime attribute problem poses a significant challenge in the determination of normal forms.

Although the problem of determining prime attributes is NP-complete, there are certain cases where recognizing prime and nonprime attributes is not difficult. Example 1 illustrates one such case.

Example 1 Consider a relation scheme $R(U, F)$, where U is the set of attributes, and $U = \{A, B, C, D, G, H\}$; F is the set of FDs, and $F = \{A \rightarrow BC, D \rightarrow H\}$.

In Example 1, the unique key for R is ADG. Therefore, the prime attributes are A, D, and G, while the nonprime attributes are B, C, and H. By examining the FDs in F, we observe that A and D only appear on the left-hand sides, whereas G does not appear at all. From this observation, we can deduce Proposition 1: if an attribute appears solely on the left-hand side or not at all in the FDs, it must be a prime attribute. Similarly, we note that B, C and H only appear on the right-hand sides of the FDs, leading us to conclude Proposition 2: if an attribute exclusively appears on the right-hand side of an FD, it must be a nonprime attribute. These propositions have been proven true[6].

To provide a more comprehensive analysis of the prime attribute problem, we present a proof using the partition of the attribute set. Assuming Proposition 1 and 2 hold, the problem of determining prime attributes becomes partially equivalent to testing an attribute's appearance in the set of FDs. The testing procedure primarily involves scanning F and checking whether the

Received date: 2025-03-07
* Correspondence should be addressed to LIU Guohua, email: ghliu@dhu.edu.cn
Citation: XU Y L, LIU G H, YU X X, et al. Features of prime attributes in a relation scheme[J]. *Journal of Donghua University (English Edition)*, 2025, 42(6): 689-698.

attribute is found solely on the left-hand or right-hand sides of the FDs. In the worst case, either side of an FD may contain $(m-1)$ attributes, where m is the number of attributes in set U. Thus, the comparison is performed at most $(m-1)$ times for one side of an FD. Consequently, the time complexity of the testing procedure is $O(m \times n)$, where n is the number of FDs in set F. This demonstrates that the prime attribute problem has polynomial-time solutions for certain special cases.

Like other NP-complete problems, the fact that some special cases can be solved in polynomial time does not show that the prime attribute problem is in the polynomial-time complexity class (P-class). The purpose of this paper is not to prove that the prime attribute problem is in the P-class, but rather to examine the difficult core of the prime attribute problem and generalize its features.

For a relation scheme $R(U,F)$, where U is the set of attributes and F is the set of FDs, the attributes in U can be divided into four distinct subsets: U_1, U_2, U_3, and U_4. These subsets are defined based on the appearance of attributes in F. Specifically, U_1 consists of attributes that solely appear on the left-hand sides of FDs, U_2 consists of attributes that solely appear on the right-hand sides of FDs, U_3 consists of attributes that appear on both sides of FDs, and U_4 consists of attributes that do not appear in F.

From the propositions mentioned above, we can infer that attributes belonging to U_1, U_2, or U_4 are the tractable aspects of the prime attribute problem. The remaining case is when an attribute belongs to U_3. Determining whether an attribute in U_3 is a prime attribute is actually more complex, making the case of U_3 the difficult core of the prime attribute problem.

In Example 1, if we include $BC \rightarrow DA$ in the set of FDs F, we observe that attributes A, B, and C appear on both sides of FD*s* in F. In this case, A, B, and C are prime attributes, while D is a nonprime attribute. Conversely, if we modify $BC \rightarrow DA$ to $BC \rightarrow D$, then even though B, C, and D appear on both sides of the FDs in F, they remain nonprime attributes. This example shows that an attribute in U_3 can be either a prime attribute or a nonprime attribute.

We aim to investigate the challenging core of the prime attribute problem, specifically addressing cases in which an attribute is in U_3. Additionally, we show the features of attributes in U_3 and make the following contributions:

1) we show the necessary and sufficient conditions for a key to be the unique key of a relation scheme;

2) we illustrate the features of prime attributes in U_3 and generalize the features of common prime attributes.

1 Related works

Several algorithms for finding keys of a relation scheme have been presented in previous studies. Fadous et al.[7] proposed a method for finding candidate keys using Boolean functions, improving efficiency in large databases by transforming FDs into Boolean functions. Hao et al.[8] proposed a method for finding all candidate keys using attribute relation tables, optimizing the traditional attribute closure approach with matrix operations. Wastl[9] introduced a linear derivation system using a Hilbert-style inference system to find all keys, providing an efficient way to derive minimal keys for relation schemes. Cordero et al.[10] presented a new approach for deriving all minimal keys using simplification logic, avoiding the need to atomize dependencies and improving computational efficiency. Demba[11] proposed KeyFinder, which introduces "core attributes" that must appear in keys if they occur only on the left-hand side of dependencies. In a recent extension of the key identification problem, Nakos et al.[12] introduced the Targeted Candidate Key model, which focuses on finding a minimum set of attributes to determine a specific target subset. This formulation highlights the complexity of attribute selection in reasoning and aligns with efforts to reduce the overhead of key enumeration.

Keys are the foundation for identifying prime attributes in a relation scheme. If all keys can be found, then the prime attribute problem is effectively solved. However, the number of keys may increase factorially or even exponentially, depending on the number of attributes and FDs[3-5]. Consequently, listing all keys to identify prime attributes becomes impractical. Lucchesi et al.[4] have proved the prime attribute problem to be NP-complete, confirming its computational difficulty. To address this issue, Kundu[13] presented a sufficient condition for determining whether an attribute is a prime attribute. This method avoids listing all keys, but instead involves constructing each subset W that does not contain attribute A. The time complexity of this approach is $O(2^{|U|-1} \times |F|)$. Building upon this work, Mannila et al.[14] refined the approach by restricting the subsets to max sets, thereby improving the time complexity. They showed a sufficient and necessary condition for an attribute to be a prime attribute. Specifically, consider a relation scheme $R(U,F)$, where U is the set of attributes, F is the set of FDs. For an attribute $A \in U$, let $\max(U,A)$ be the set of all maximal subsets $Y \subseteq U$ (with respect to $\subseteq$) such that $Y \nrightarrow A$. Then, A is a prime attribute if and only if for some $W \in \max(U,A)$, $(WA)_F^+ = U$. Testing an attribute based on this sufficient and necessary condition requires $O(m_A \times |F|)$ time, where $m_A = |\max(U,A)|$. Notably, neither Kundu's nor Mannila and Raiha's condition considers where attributes appear in FDs of F. However, the appearance of attributes plays a crucial role in solving the prime attribute problem. In certain cases, distinguishing between prime and non-prime attributes becomes straightforward. Feng et al.[6] discussed the relationship between a prime attribute and its appearance in F. They investigated three situations: an attribute A is a prime attribute when it only appears on the left-hand sides of FDs in F, or when it does not appear in F at all. Conversely, if it appears solely on the right-hand sides of FDs in F, then it is deemed a nonprime attribute. In these cases, the prime attribute problem is in P-class. Furthermore, Hao et al. [15] presented a sufficient and necessary condition for an attribute that appears on both sides of FDs in F to be a prime attribute. These previous works have laid the foundation for research on the issues we are addressing.

2 Preliminaries

2.1 Relation schemes and relations

In Codd's[16] relational database model, a database description is referred to as a schema, which includes descriptions for each relation in the database[17]. A relation scheme, which describes a single relation, consists of the relation name, its attributes, and a set of data dependencies. The notation for a relation scheme is $R(U, \Gamma)$, where R represents the relation name; U is the set of distinct attributes appearing in it, and $U = \{A_1, A_2, \cdots, A_n\}$; Γ denotes the set of data dependencies. Each attribute has a domain consisting of a set of possible values. It is important to note that different attributes may share the same domain. The main data dependency of concern in this paper is FD, which will be defined precisely in Section 2.2.

The contents of a relational database consist of a family of sets of tuples. Each set of tuples in the family is known as the state or extension of the corresponding relation scheme, and it can be considered as a relation in the mathematical sense[18]. Therefore, the state or extension of a relation scheme is also referred to as a relation. However, there is a slight difference between the relation as an extension and the relation in mathematics. The relation as an extension must be defined over the set of attributes of the relation scheme. This means that every element in a tuple corresponds to a distinct attribute in the attribute set of the scheme. In practical terms, a relation is often represented as a table, where each column corresponds to a distinct attribute and each row corresponds to a distinct entity. An <entity, attribute> entry in a relation represents a value associated with the entity, selected from the domain of the attribute. The set of tuples comprising a relation typically changes over time as entities are inserted, deleted, or modified, but its scheme remains relatively static throughout the lifecycle of the relation.

2.2 FDs and covers

Data dependencies are a specific type of semantic tool in databases. Some of these dependencies may result in update anomalies and internal inconsistencies in the data relations. Codd[1-2] observed that certain patterns of FDs among the attributes of a relation scheme can lead to undesirable properties in its extensions. This discovery led to the development of a series of normal forms for relations, namely 2NF, 3NF, and BCNF.

For a relation scheme R with U as the set of attributes, let A and B be attributes in U. We say that B is functionally dependent on A if, at any given time, there is at most one value of $b \in DOM(B)$ corresponding to a given value of $a \in DOM(A)$, where $DOM(A)$ and $DOM(B)$ represent the domains of attributes A and B respectively. If B is functionally dependent on A, we say that A functionally determines B. For convenience, we generally omit the "DOM" and use the notation $f: A \rightarrow B$ to denote the FD of B on A. f is called an FD.

The above definitions can be generalized for FDs over sets of attributes. If $X = \{A_1, A_2, \cdots, A_n\}$ and $Y = \{B_1, B_2, \cdots, B_p\}$ are sets of attributes in U, then $f: X \rightarrow Y$ means $f: DOM(A_1) \times DOM(A_2) \times \cdots \times DOM(A_n) \rightarrow DOM(B_1) \times DOM(B_2) \times \cdots \times DOM(B_p)$. We usually simplify the set notation and write $f: A_1, A_2, \cdots, A_n \rightarrow B_1, B_2, \cdots, B_p$, referring to X as the left-hand side of f and Y as the right-hand side of f.

Let $R(U, F)$ be a relation scheme, where U is the set of attributes and F is the set of FDs. Let X be a subset of U, A be any attribute in U, and $X \rightarrow A$ be an FD. We say that A is partially dependent on X if A is functionally dependent on a proper subset of X. In other words, there exists a set Y such that $Y \subset X$ and $Y \rightarrow A$. Attribute A is fully dependent on X if it is not dependent on any proper subset of X. A is transitively dependent on the set X if there exists a set Y of attributes of U such that $X \rightarrow Y$, $Y \nrightarrow X$, $Y \rightarrow A$, and A does not belong to Y. X and Y are sets of attributes in U. We say that Y is trivially dependent on X if Y is a subset of X.

Given a relation scheme $R(U, F)$, additional FDs can be deduced from F. Armstrong[19] presented a complete set of inference rules for FDs. Using these rules, we can derive additional FDs from F. The notation $F \vDash X \rightarrow Y$ indicates that $X \rightarrow Y$ is logically implied by F. The set of all FDs that are derivable from F by repeatedly applying the rules (including the FDs in F) is called the closure of F and is denoted by F^+. The sequence of FDs for deriving an FD is called the derivation of the FD, where each FD in the derivation is either in F, or follows from earlier ones by applying one of the inference Rule A1–A5 (i.e., reflexivity, augmentation, pseudotransitivity, union, and decomposition).

Given a set of attributes X, we define the closure of X (relative to F), denoted by X_F^+, to be the set of attributes that are functionally dependent on X. That is, X_F^+ contains the set of all attributes A such that $F \vDash X \rightarrow A$.

To model derivations, Beeri et al. [20] introduced a graph model, called F-based derivation trees (F-based DT).

Let F be a set of FDs, the set of F-based DTs is defined as follows:

Rule 1 If A is an attribute, then a node labeled with A is an F-based DT.

Rule 2 If T is an F-based DT with a leaf node labeled with A and the FD $B_1, B_2, \cdots, B_m \rightarrow A$ is in F, then the tree constructed from T by adding $B_1, B_2, \cdots, B_m$ as children of the leaf labeled with A is also an F-based DT.

Rule 3 A labeled tree is an F-based DT only if it so follows by a finite number of applications of Rules 1 and 2.

If the set of leaves' labels in a DT is contained in the set of attributes X, then we call it an X-DT. An X-DT with the root labeled by A is called a derivation tree for the FD $X \rightarrow A$.

Theorem 1[20] The FD $X \rightarrow Y$ is in F^+, where X and Y are sets of attributes, if and only if for each attribute A in Y, there exists an F-based X-DT with the root labeled by A.

Lemma 1[20] Let $X \rightarrow A$ be in F^+, where A is not in X. Then there exists a derivation of $X \rightarrow A$ from F in which reflexivity is not used and augmentation is either not used at all or used only in the last step.

Lemma 2[20] Let $X \rightarrow A$ be in F^+, where A is not in X. Then there exists a derivation of $X \rightarrow A$ from F in which only the rules pseudotransitivity, union, and restricted augmentation are used.

Lemma 3 $F \vDash X \rightarrow A$ if and only if there exists an FD $V \rightarrow W$ in F such that $A \in W$ and $V \subseteq X_F^+$.

Proof To prove the sufficiency of the asserted condition, we observe that since there exists an FD $V \rightarrow W$ in F such that $A \in W$ and $V \subseteq X_F^+$, we can deduce that $F \vDash V \rightarrow A$ by Rule A5, $F \vDash X \rightarrow V$ by the definition of closure of X, and $F \vDash X \rightarrow A$ by Rule A3.

To prove the necessity of the asserted condition, we assume, by way of contradiction, that $F \vDash X \rightarrow A$ but there does not exist any FD $V \rightarrow W$ in F such that $A \in W$. According to the definition of a derivation of an FD, every FD in the derivation is either a member of F or follows from previous FDs in the derivation by applying one of the inference rules A1–A5. Since $F \vDash X \rightarrow A$, $X \rightarrow A$ must be in F^+. According to Lemma 1 and Lemma 2, in the derivation of $X \rightarrow A$ from F, the FD is either a member of F or follows from the previous FDs in the derivation by an application of pseudotransitivity, union, and a restricted augmentation. As per our assumption, there does not exist any FD $V \rightarrow W$ in F such that $A \in W$. Therefore, in the derivation of $X \rightarrow A$, the right-hand side of every member of F does not contain A. Since the other FDs in the derivation of $X \rightarrow A$ are derived from previous FDs in the derivation by an application of pseudotransitivity, union, and a restricted augmentation, they also do not contain A when the right-hand sides of the previous FDs in the derivation do not contain A. The procedure of the derivation begins with the FDs in F, where the right-hand sides of these FDs do not contain A. The right-hand sides of the other FDs based on them also do not contain A. Thus, in the derivation, there does not exist an FD whose right-hand side contains A. This shows that there does not exist a derivation of $X \rightarrow A$, namely, $X \rightarrow A$ cannot be derived from F, which contradicts the assumption that $F \vDash X \rightarrow A$. Therefore, if $F \vDash X \rightarrow A$, then there exists an FD $V \rightarrow W$ in F such that $A \in W$. If V is not contained in X_F^+, then $F \vDash X \rightarrow V$ exists, showing that $X \rightarrow A$ cannot be derived using X. This contradicts the assumption of $F \vDash X \rightarrow A$.

Armstrong's completeness result can be stated as follows: For any set F of FDs, there exists a relation in which all the FDs in F^+ are valid and no other FD is valid. However, it is advisable to avoid directly dealing with the closure of a given set of FDs for several reasons. Firstly, even relatively small sets may have a closure that contains an overwhelming number of FDs. Additionally, the closure often includes redundant information since all the FDs in the closure can be derived from subsets of the closure, namely the original set.

Formally, if F is a set of FDs, a cover of F is any set of FDs that has the same closure as F. It is important to note that a cover does not have to be a subset of F. A cover is nonredundant if it does not contain any proper subset that is also a cover. An FD f in a set of FDs F is redundant if $F - \{f\} \vDash f$. Thus, a cover is redundant if and only if it contains a redundant FD.

Let $|F|$ represent the number of FDs in F, specifically the cardinality of F. F is the minimum if there is no set G where $|G| < |F|$ and $F^+ = G^+$. F is L-minimum if it is minimum and for every FD $X \rightarrow Y$ in F, there is no proper subset X' of X such that $X' \rightarrow Y$ is in F^+. F is LR-minimum if it is both L-minimum and replacing FD $X \rightarrow Y$ in F with $X \rightarrow Y'$, where Y' is a proper subset of Y, alters the closure of F. $|F|_A$ denotes the number of attributes appearing in a set of FDs F, counting repetitions. F is optimal if there is no set G where $|G|_A < |F|_A$ such that $F^+ = G^+$.

Given a set of FDs F, if $F \vDash X \rightarrow Y$, and $Y \rightarrow X$, then the sets of attributes X and Y are equivalent, denoted by $X \leftrightarrow Y$. The set of all FDs in F with left-hand sides equivalent to X is denoted by $E_F(X)$, and $|E_F(X)|$ represents the cardinality of $E_F(X)$. The set of left-hand sides of FDs in $E_F(X)$ is denoted by $e_F(X)$. The family consisting of all $E_F(X)$ is denoted by E_F, and $|E_F|$ is the cardinality of E_F. For the E_F, $E_F(X_i) \cap E_F(X_j) = \emptyset$ and $E_F(X_1) \cup E_F(X_2) \cup \ldots \cup E_F(X_{|F|}) = F$, where $1 \leqslant i, j \leqslant |F|$. Therefore, the E_F forms a partition of F.

If X in an $E_F(X)$ is a key, it is denoted as $E_F(X_K)$. If the attributes in an $E_F(X)$ are all prime attributes, it is denoted as $E_F(X_P)$. If the attributes in an $E_F(X)$ are all nonprime attributes, it is denoted as $E_F(X_N)$. If both nonprime attributes and prime attributes are present together in an $E_F(X)$, it is denoted as $E_F(X_M)$.

Example 2 Let $R(U, F)$ be a relation scheme, where $U = \{A, B, C, D, G, H\}$ is the set of attributes and $F = \{A \rightarrow BC, BC \rightarrow A, BCD \rightarrow H, H \rightarrow G, G \rightarrow H\}$ is the set of FDs.

The keys in Example 2 are BCD and AD. The prime attributes are A, B, C, and D, while the nonprime attributes are G and H. Therefore, $E_F(A_P) = \{A \rightarrow BC,$

$BC \to A\}$, $E_F(BCD_K) = \{BCD \to H\}$, and $E_F(H_N) = \{H \to G, G \to H\}$. The $e_F(X)s$ are $e_F(A) = \{A, BC\}$, $e_F(BCD) = \{BCD\}$, and $e_F(H) = \{H, G\}$, and $E_F = \{E_F(A), E_F(BCD), E_F(H)\}$. The set F is nonredundant, minimum, L-minimum, and LR-minimum. However, it is not optimal because there is another set of FDs $G = \{A \to BC, BC \to A, AD \to H, H \to G, G \to H\}$ such that $|G|_A < |F|_A$ and $F^+ = G^+$.

For the cover of a set of FDs, Maier[21] has shown the following results:

1) given a set of FDs G, finding a minimum cover F for G can be done in $O(np^2)$ time, where n is the length of G (in attribute symbols) and p is the number of FDs in G;

2) given a set of FDs G, L-minimum and LR-minimum covers for G can be found in $O(np^2 + n^2)$ time, where n is the length of G (in attribute symbols) and p is the number of FDs in G;

3) the optimal cover problem is NP-complete.

Since LR-minimum cover can be found in $O(np^2 + n^2)$ time, in this paper, we will use LR-minimum cover as the set of FDs of a relation scheme directly.

2.3 Keys and prime attributes

Consider a relation scheme $R(U, F)$, where U is the set of attributes, F is the set of FDs, and K is a subset of U. A subset K of U is considered a superkey of R if every attribute in U functionally depends on K. If a subset K of U is both a superkey and does not properly contain any other superkey of R, then it is considered a key of R. An attribute A is considered a prime attribute if it belongs to any key of R; otherwise, it is classified as a nonprime attribute.

Example 3 Consider the relation scheme $R(U, F)$, where $U = \{A, B, C, D, E, G\}$ is the set of attributes and $F = \{A \to BC, BC \to A, BCD \to E, E \to G, G \to BCD\}$ is the set of FDs.

In Example 3, prime attributes are A, B, C, D, E, and G, and the keys are AD, BCD, E, and G.

Lucchesi et al.[4] presented an algorithm that lists all keys. The time complexity of this algorithm is polynomial in the number of attributes and FDs, and linear in the number of keys. However, the number of keys can exponentially increase based on the number of attributes and FDs. In such cases, the algorithm's time complexity will be exponential in the size of the input. Therefore, the task of key finding is intrinsically challenging.

Lucchesi and Osborn also proved that the following two problems are NP-complete:

1) the prime attribute problem: Given an attribute A, determine whether it belongs to any key;

2) the key of cardinality m problem: Given an integer $m > 1$, decide if there is a key with a cardinality less than m.

These findings are relevant to the task of finding normal forms for relation schemes. For $R(U, F)$, which is in one of 2NF, 3NF, or BCNF, the prime attribute and nonprime attribute must obey specific properties. The NP-completeness of the prime attribute problem implies that this task may also be computationally challenging[4,17,22-23].

3 Features of Prime Attributes

Although the prime attribute problem is NP-complete, there are cases where it is easy to distinguish between prime and nonprime attributes. The appearance of attributes in the set F is a crucial factor in solving the prime attribute problem. Earlier in the paper, we show that when an attribute A solely appears on the left-hand side of FDs in F, or does not appear at all, or only appears on the right-hand side of FDs in F, the time required to test is $O(m \times n)$. Here, m represents the number of attributes in the set U, and n denotes the number of FDs in F. Thus, in these specific scenarios, the prime attribute problem falls into the P-class. We will thoroughly analyze all possible situations of attribute appearance in F and present the features of a prime attribute and nonprime attribute in each situation[6].

We are aware that for a relation scheme $R(U, F)$, the possibility of an attribute becoming a prime attribute is related to the sets U_1, U_2, U_3, and U_4, based on the appearance of attributes in F. The following theorems will demonstrate the relationship between the possibility of an attribute becoming a prime attribute and the set $U_i (1 \leqslant i \leqslant 4)$.

Theorem 2 Let $R(U, F)$ be a relation scheme, U is the set of attributes, F is the set of FDs.

1) Attributes in U_1 or U_4 are prime attributes, and $U_1 \cup U_4$ is included in any key of R.

2) Attributes in U_2 are nonprime attributes.

3) K is the unique key of R, if and only if $K \cap U_3 = \emptyset$.

Proof For 1), let's assume, by contradiction, that K is any key of R, $A \in U_1$ or $A \in U_4$, and A is a nonprime attribute. Clearly, A cannot be in K. Since $A \in U_1$ or $A \in U_4$, there are no nontrivial FDs $X \to A$, as per Lemma 3. Consequently, A does not appear on the right-hand sides of any FDs in F. This implies that apart from A itself, no other attribute or attribute set can functionally determine A. Since assuming A is not in K, K does not functionally determine A. Thus, K is not a key, contradicting the assumption that K is a key of R. Therefore, A must be in K and is thus a prime attribute. Since $A \in U_1$ or $A \in U_4$, all attributes in U_1 or U_4 are prime attributes, and $U_1 \cup U_4$ is included in any key of R.

For 2), assume, by way of contradiction, that $A \in U_2$, A is a prime attribute, K is any key of R, and $A \in K$. Since $A \in U_2$, there is a nontrivial FD $X \to A$, but there is no nontrivial FD $Y \to X$ and $A \in Y$. As K is a key, $K \to X$. Since there is no nontrivial FD $Y \to X$ and $A \in Y$, it follows that $(K - A) \to X$ and $(K - A) \to A$. This indicates that K is not a key, contradicting the assumption that K is a key of R. Therefore, A is a nonprime attribute. Furthermore, since A is any attribute in U_2, all attributes in U_2 are nonprime attributes.

For 3), to prove that the asserted condition is sufficient, assume, by way of contradiction, that K is a key of R and $K \cap U_3 = \emptyset$, while K' is another key of R where $K \neq K'$. According to 1), we know that $U_1 \cup U_4 \subseteq K$ and $U_1 \cup U_4 \subseteq K'$. From 1) and $K \cap U_3 = \emptyset$, we can deduce that $K = U_1 \cup U_4$. Therefore, $K \subseteq K'$. However, since $K \neq K'$, this implies that K' is not a key, contradicting the assumption that K' is a key of R. Hence, K is the unique key of R if $K \cap U_3 = \emptyset$.

For 3), to prove that the asserted condition is necessary, assume, by way of contradiction, that K is the unique key of R, $K \cap U_3 = \emptyset$, and $A \in K \cap U_3$. Since $A \in K \cap U_3$, it follows that $A \in U_3$. According to the definition of U_3, there exists a nontrivial FD $X \rightarrow A$, where $A \notin X$. Thus, $((K - \{A\}) \cup X)_F^+ = K_F^+$, which shows that $(K - \{A\}) \cup X$ is a superkey. Therefore, we can obtain another key K' from $(K - \{A\}) \cup X$. If $K' \cap X = \emptyset$, then X is redundant in $(K - \{A\}) \cup X$, meaning that $F \vDash (K - \{A\}) \rightarrow X$. From $(K - \{A\}) \rightarrow X$ and $X \rightarrow A$ by Armstrong's pseudotransitivity rule, we can deduce that $F \vDash (K - \{A\}) \rightarrow A$. This demonstrates that A is redundant in K, thereby proving that K is not minimal. Contradicting the assumption, this means that K is not a key. Hence, $K' \cap X \neq \emptyset$. From this, we can conclude that K' is different from K, indicating that K is not the unique key of R and contradicting our initial assumption. Therefore, $K \cap U_3 = \emptyset$.

Example 4 Consider a relation scheme denoted as $R(U, F)$, where $U = \{A, B, C, D, E, G\}$ is the set of attributes and $F = \{A \rightarrow BC, BC \rightarrow A, BCD \rightarrow E\}$ is the set of FDs.

In Example 4, at first, we can distinguish between prime and nonprime attributes by determining the keys of the scheme R, as per the definitions. The keys of R are ADG and $BCDG$, thereby identifying E as the nonprime attribute and A, B, C, D, and G as the prime attributes, where D and G are in all keys of R.

On the other hand, we can differentiate prime and nonprime attributes using Theorem 2. Based on the appearance of attributes in F, we have $U_1 = \{D\}$, $U_2 = \{E\}$, $U_3 = \{A, B, C, D\}$, and $U_4 = \{G\}$. By applying the decision condition of Theorem 2, we conclude that E is the nonprime attribute, while D and G are the prime attributes in all keys of R. This conclusion aligns with the distinction made using the definitions.

Theorem 3 Let $R(U, F)$ be a relation scheme, where U is the set of attributes; F is the set of FDs; $A \in U$. If A is a nonprime attribute, it must appear on the right-hand side of at least one FD in F.

Proof We assume, by way of contradiction, that A is a nonprime attribute but does not appear on the right-hand side of any FDs in F. According to the definition of the partition (U_1, U_2, U_3, U_4), $A \in U_1$ or $A \in U_4$. From Theorem 2, we can infer that A is a prime attribute, which contradicts our assumption that A is a nonprime attribute. Therefore, A must appear on the right-hand side of at least one FD in F.

Example 5 Consider a relation scheme $R(U, F)$, where $U = \{A, B, C, D, E, G\}$ is the set of attributes, and $F = \{A \rightarrow BC, BC \rightarrow A, BCD \rightarrow E, E \rightarrow G, G \rightarrow E\}$ is the set of FDs.

In Example 5, the keys for R are AD and BCD. Therefore, the nonprime attributes of R are E and G. Among the FDs in F, we have $BCD \rightarrow E$, $E \rightarrow G$, and $G \rightarrow E$, indicating that both E and G appear on the right-hand sides of FDs in F. This observation aligns with Theorem 3.

Theorem 4 Let $R(U, F)$ be a relation scheme, where U is the set of attributes and F is the set of FDs. Let $A \in U$. If A is a prime attribute and appears in F, it must appear on the left-hand side of at least one FD in F.

Proof Assume, by way of contradiction, that A is a prime attribute, but it does not appear on the left-hand side of any FD in F. Therefore, $A \in U_2$. By Theorem 2, it is known that A is a nonprime attribute, which contradicts the assumption that A is prime. Thus, A must appear on the left-hand sides of some FDs in F.

In Example 5, all keys are AD and BCD, so the prime attributes of R are A, B, C, and D. In F, there are FDs $A \rightarrow BC$, $BC \rightarrow A$, and $BCD \rightarrow E$, indicating that A, B, C, and D appear on the left-hand sides of FDs in F. This conclusion aligns with Theorem 4.

Theorem 5 Let $R(U, F)$ be a relation scheme, where U is the set of attributes, F is the set of FDs. $X \rightarrow Y \in F$, and K is a key of R. If $X_F^+ \cap K \neq \emptyset$, X must contain prime attributes.

Proof Let $V = X_F^+ \cap K$, implying that $V \subseteq X_F^+$. Based on the definition of X_F^+, it can be inferred that $F \vDash X \rightarrow V$. By replacing V with X in K, it can be concluded that $((K - V) \cup X)_F^+ = K_F^+$, which indicates that $(K - V) \cup X$ is a superkey. Therefore, another key K' can be obtained from $(K - V) \cup X$. If $K' \cap X = \emptyset$, then X is redundant in $(K - V) \cup X$, namely, $F \vDash (K - V) \rightarrow X$. Consequently, $(K - V) \rightarrow V$ is derived using Armstrong's pseudotransitivity rule from $(K - V) \rightarrow X$ and $X \rightarrow V$. This reveals that V is redundant in K, and as per the definition of the key, K is not minimal. Thus, K is not a key, contradicting the fact that K is a key. Therefore, $K' \cap X \neq \emptyset$. Since K' is a key, $K' \cap X$ must contain prime attributes. As $K' \cap X \subseteq X$, it can be inferred that X must contain prime attributes.

In Example 5, all keys are AD and BCD, therefore the prime attributes of R are A, B, C, and D. In F, we have $A \rightarrow BC$, $BC \rightarrow A$, $BCD \rightarrow E$, $E \rightarrow G$, and $G \rightarrow E$. The closures of the left-hand sides of these FDs are as follows: $A_F^+ = ABC$, $BC_F^+ = ABC$, $BCD_F^+ = ABCDEG$, $E_F^+ = EG$, and $G_F^+ = EG$. From this, we can see that $A_F^+ \cap BCD = BC \neq \emptyset$, $BC_F^+ \cap AD = A \neq \emptyset$, $BCD_F^+ \cap AD = AD \neq \emptyset$, and the left-hand sides A, BC, and BCD contain the prime attributes A, BC, and BCD respectively. This conclusion is consistent with Theorem 5.

Theorem 6 Let $R(U, F)$ be a relation scheme where U is the set of attributes and F is the set of FDs. If $X \to Y \in F$ and $A \in Y$, where A is a prime attribute, then X must contain prime attributes.

Proof Let K be a key of R, $A \in K$. Since $X \to Y \in F$, and $A \in Y$, we can deduce that $A \in X_F^+$, so, $X_F^+ \cap K \neq \emptyset$. According to Theorem 5, X must contain prime attributes.

In Example 5, the prime attributes of R are A, B, C, and D. In F, the right-hand sides of $A \to BC$ and $BC \to A$ contain the prime attributes A, B, and C. The left-hand sides A and BC are also prime attributes. This conclusion is consistent with Theorem 6.

Lemma 4[13] Let $R(U, F)$ be a relation scheme where U is the set of attributes and F is the set of FDs. Let $W \subseteq U, A \in U$. If $(WA)_F^+ = U$ and $W_F^+ \neq U$, then A is a prime attribute.

Theorem 7 Let $R(U,F)$ be a relation scheme where U is the set of attributes and F is the set of FDs. If attribute A belongs to U_3, and A is a prime attribute if and only if for any $X \to A \in F^+$, there must exist an FD $V \to B \in F^+$ where $A \in V$ and $B \in X$.

Proof To prove the sufficiency of the asserted condition, we assume, by way of contradiction, that attribute A belongs to U_3 and for any $X \to A \in F^+$, there must exist an FD $V \to B \in F^+$, where $A \in V$, $B \in X$, but A is not a prime attribute. Let K be a key of R, so $K \to A \in F^+$. Based on this assumption, there must be an FD $V \to B \in F^+$, where $A \in V, B \in K$. Therefore, $(K - B) \cup V$ is a superkey. From $(K - B) \cup V$, we can obtain a set of attributes K', which is also a superkey, by keeping A and removing other redundant attributes. If $(K' - A)$ is a superkey, there must be an FD $W \to A \in F^+$, where $W \subseteq (K' - A)$. By our assumption, there must be an FD $V' \to B' \in F^+$, where $A \in V', B' \in W$. Thus, $((K' - A) - B') \cup V'$ is a superkey. By keeping A and removing other redundant attributes from $((K' - A) - B') \cup V'$, we can obtain a set of attributes K'', which is also a superkey. If $(K'' - A)$ is a superkey, there must be an FD $W' \to A \in F^+$, where $W \subseteq (K'' - A)$. According to our assumption, there must be an FD $V'' \to B'' \in F^+$, where $A \in V'', B'' \in W'$. Therefore, $((K'' - A) - B'') \cup V''$ is a superkey. Since the number of sets of attributes that functionally determine A is at most $2^{|U|}$, the above process is finite. Let K_l be the last key obtained by removing A, we can obtain a superkey K_l' that includes A using the aforementioned process. Since K_l is the last key obtained by removing A, there is no set of attributes in $(K_l' - A)$ that can functionally determine A. Thus, $(K_l' - A)_F^+ \neq U$. According to Lemma 4, A is a prime attribute, which contradicts the assumption that A is a nonprime attribute.

To prove that the asserted condition is necessary, we note that since attribute A belongs to U_3, A must appear on the right-hand side of some FD in F. According to Lemma 3, there exists an FD $X \to A \in F^+$. As A is a prime attribute, it must be part of a key K. Hence, $K \to X \in F^+$, so there must be an FD $K \to B \in F^+$, where $A \in K$, and $B \in X$.

In Example 5, the attributes in U_3 are A, B, C, E, and G, whereas the prime attributes of R are A, B, and C. In F, attributes A, B, and C appear on both the left and right-hand sides of the FDs $A \to BC$ and $BC \to A$. According to Lemma 3, for any $X \to A$, $X' \to B$, $X'' \to C \in F^+$, there must exist FDs $V \to P$, $V' \to P'$, $V'' \to P'' \in F^+$, where $A \in V$, $P \in X$, $B \in V'$, $P' \in X'$, $C \in V''$, and $P'' \in X''$. This conclusion is consistent with Theorem 7.

Corollary Let $R(U, F)$ be a relation scheme, where U is the set of attributes and F is the set of FDs. Attribute A is a prime attribute if and only if for any $X \to A \in F^+$, there must be an FD $V \to B \in F^+$, where $A \in V$ and $B \in X$.

Proof To prove the sufficiency of the asserted condition, if for any $X \to A \in F^+$, there must be an FD $V \to B \in F^+$, where $A \in V$ and $B \in X$, then A must be in U_1, U_3, or U_4. The reason for this is that attributes in U_2 only appear on the right-hand sides of FDs in F, and there are no FDs whose left-hand side contains attributes in U_2 in F^+. According to Theorem 2, attributes in U_1 and U_4 are prime attributes. If attribute A is in U_3, since for any $X \to A \in F^+$, there must be an FD $V \to B \in F^+$, where $A \in V$ and $B \in X$, according to Theorem 7, attribute A in U_3 is a prime attribute.

To prove the necessity of the asserted condition, we consider attribute A as a prime attribute. It can be in U_1, U_3, or U_4. If A is in U_1 or U_4, Lemma 3 states that there are no FDs of the form $X \to A \in F^+$, satisfying the condition naturally. If A is in U_3, Theorem 7 states that for any $X \to A \in F^+$, there must be an FD $V \to B \in F^+$ where $A \in V$ and $B \in X$.

In addition to FDs, prime attributes and nonprime attributes also have the following features in $E_F(X)$, namely, subsets of the set of FDs F.

Theorem 8 Let $R(U, F)$ be a relation scheme, where U is the set of attributes and F is the set of FDs. For an $E_F(X)$, if X contains prime attributes, all other elements in $e_F(X)$ also contain prime attributes. On the other hand, if X does not contain prime attributes, $E_F(X)$ must be an $E_F(X_N)$.

Proof In the case of $E_F(X)$ containing prime attributes, each element in the $e_F(X)$ is equivalent to X. Therefore, their closures, being equivalent to X_F^+, also contain prime attributes. Theorem 5 supports this conclusion by stating that other elements in $e_F(X)$ contain prime attributes. Conversely, when X does not contain prime attributes, other elements in $e_F(X)$ also do not contain prime attributes. Otherwise, according to Theorem 5, X would need to contain prime attributes. Since elements in $e_F(X)$ do not contain prime attributes, Theorem 6 confirms that each right-hand side of FDs in $E_F(X)$ does not contain any prime attributes. As both the left and right-hand sides of FDs in $E_F(X)$ do not contain any prime attributes, $E_F(X)$ must be an $E_F(X_N)$.

In Example 5, all keys are AD and BCD, thus the prime attributes of R are A, B, C, and D, while the nonprime attributes are E and G. $E_F = \{E_F(A), E_F(BCD), E_F(G)\}$, where $E_F(A) = \{A \rightarrow BC, BC \rightarrow A\}$, $E_F(BCD) = \{BCD \rightarrow E\}$, $E_F(G) = \{E \rightarrow G, G \rightarrow E\}$, $e_F(A) = \{A, BC\}$, $e_F(BCD) = \{BCD\}$, $e_F(G) = \{E, G\}$. In $e_F(A)$, A contains a prime attribute, and the other element, BC, has prime attributes B and C, confirming Theorem 8. In $e_F(G)$, G does not contain a prime attribute, and the attributes in $E_F(G)$ are E and G, both of which are nonprime attributes, aligning with Theorem 8.

Theorem 9 Let $R(U, F)$ be a relation scheme, U is the set of attributes, F is the set of FDs. If there is an $E_F(X_K)$ in the E_F, it must be unique.

Proof Let us assume, by way of contradiction, that there exists an $E_F(X'_K)$ in the E_F, where X' is a key of R, but $E_F(X_K) \neq E_F(X'_K)$. Since X' and X are both keys of R, it implies that $(X')^+_F = (X)^+_F$, which indicates that X' is equivalent to X. Since X' and X are both left-hand sides of FDs, they must appear in the same $E_F(X)$, specifically $E_F(X') = E_F(X)$. This contradicts the assumption that $E_F(X_K) \neq E_F(X'_K)$. Therefore, $E_F(X_K)$ must be unique.

In Example 5, all keys are AD, BCD, $E_F = \{E_F(A), E_F(BCD), E_F(G)\}$, where $E_F(A) = \{A \rightarrow BC, BC \rightarrow A\}$, $E_F(BCD) = \{BCD \rightarrow E\}$, $E_F(G) = \{E \rightarrow G, G \rightarrow E\}$. According to the definition of $E_F(X_K)$, we can determine that $E_F(BCD)$ is an $E_F(X_K)$, and there are no other $E_F(X_K)$ in the E_F. Thus, $E_F(BCD)$ is unique, which aligns with Theorem 9.

Theorem 10 Let $R(U, F)$ be a relation scheme, U is the set of attributes, F is the set of FDs. If any attribute in U_2 appears in an $E_F(X)$, the $E_F(X)$ is not an $E_F(X_P)$.

Proof According to Theorem 2, all attributes in U_2 are nonprime attributes. Hence, if there are nonprime attributes appear in $E_F(X)$, it implies that $E_F(X)$ cannot be an $E_F(X_P)$.

4 Time Efficiency Analysis

In the problem of determining prime attributes in a relation scheme $R(U, F)$, traditional methods, such as those based on key enumeration algorithms[4,24], require first solving for all keys before determining the prime attributes. The worst-case time complexity of these methods is $O(2^{|U|} \times |F|)$. Kundu[13] proposed an improved approach that avoids full key enumeration, optimizing the time complexity to $O(2^{|U|-1} \times |F|)$. Our study, based on the theory of attributes partitioning and prime attribute's feature analysis, can further optimize the time complexity.

Assuming that the probability of an attribute $A \in U$ belonging to U_1, U_2, U_3, or U_4 is $\frac{1}{4}$. When $A \in U_1 \cup U_2 \cup U_4$, the time complexity for determining the prime attributes is reduced to $O(|F|)$. when $A \in U_3$, the time complexity remains at $O(2^{|U|-1} \times |F|)$. Based on this, the average time complexity is derived as

$$\frac{1}{4}(3 \times |F| + 2^{|U|-1} \times |F|) = O(2^{|U|-3} \times |F|).$$

The optimal time complexity of the algorithm occurs when $A \in U_1 \cup U_2 \cup U_4$, reaching $O(|F|)$, while the worst-case scenario (when $A \in U_3$) maintains a complexity of $O(2^{|U|-1} \times |F|)$, similar to the traditional methods. The results demonstrate that the attribute partitioning method proposed in this study effectively reduces the search space, improving the efficiency of prime attribute determination.

5 Conclusions

For a relation scheme $R(U, F)$, we partition U into four disjoint subsets: U_1, U_2, U_3, and U_4, based on the roles of attributes in F. Our study identifies U_3 as the critical subset for the prime attribute problem, where the problem is NP-complete. We further analyze and generalize the features of prime attributes in U_3, as well as the broader characteristics of common prime attributes. Our findings clarify the computational complexity of the prime attribute problem across these subsets and highlight the unique challenges posed by U_3 in comparison with U_1, U_2 and U_4. In addition, our results on the necessary and sufficient conditions for the uniqueness of a key provide valuable insights into optimizing relation schemes for both theoretical and practical purposes. The generalization of prime attribute features offers a foundational framework that could inform the analysis of other computational problems in database theory.

These contributions advance the theoretical understanding of the prime attribute problem while also supporting practical applications such as database normalization, scheme design, and optimization in relational database systems.

References

[1] CODD E F. Further normalization of the data base relational model[J]. *Data base systems*, 1972, 6(1972): 33-64.

[2] CODD E F. Recent investigations in relational data base systems[M]. New York: IBM Thomas J. Watson Research Division, 1974.

[3] DEMETROVICS J. On the number of candidate keys[J]. *Information Processing Letters*, 1978, 7(6): 266-269.

[4] LUCCHESI C L, OSBORN S L. Candidate keys

for relations [J]. *Journal of Computer and System Sciences*, 1978, 17(2): 270-279.

[5] YU C T, JOHNSON D T. On the complexity of finding the set of candidate keys for a given set of functional dependencies [J]. *Information Processing Letters*, 1976, 5(4): 100-101.

[6] FENG Y C, PANG C Y. Algorithm for finding candidate key word [J]. *Journal of Huazhong University of Science and Technology* (*Natural Science Edition*), 1986, 14 (5): 655-658. (in Chinese)

[7] FADOUS R, FORSYTH J. Finding candidate keys for relational data bases [C]//Proceedings of the 1975 ACM SIGMOD International Conference on Management of Data, San Jose, California. New York: ACM, 1975: 203-210.

[8] HAO Z X, LIU G H, LIU C L. A method for finding all candidate keys of relational schema based on attributes relative table [J]. *Journal of Computer Research and Development*, 1994, 31 (6): 6-13. (in Chinese)

[9] WASTL R. Linear derivations for keys of a database relation scheme [J]. *Journal of Universal Computer Science*, 1998, 4(11): 883-897.

[10] CORDERO P, ENCISO M, MORA A. Automated reasoning to infer all minimal keys [C]//IJCAI, Beijing, China. Palo Alto: IJCAI, 2013: 817-823.

[11] DEMBA M. KeyFinder: an efficient minimal keys finding algorithm for relational databases [J]. *Inteligencia Artificial*, 2021, 24 (68): 37-52.

[12] NAKOS V, NGO H Q, TSOURAKAKIS C E. Targeted least cardinality candidate key for relational databases [EB/OL]. (2024-08-24) [2025-04-05]. https://arxiv. org/abs/2408. 13540vl.

[13] KUNDU S. An improved algorithm for finding a key of a relation [C]//Proceedings of the Fourth ACM SIGACT-SIGMOD Symposium on Principles of Database Systems, Portland, Oregon, USA. New York: ACM, 1985: 189-192.

[14] MANNILA H, RAIHA K J. Practical algorithms for finding prime attributes and testing normal forms [C]//Proceedings of the Eighth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Philadelphia, Pennsylvania, USA. New York: ACM, 1989: 128-133.

[15] HAO Z X, LIU G H. An algorithm to find out all candidate keys of relation scheme [J]. *Chinese Journal of Computers*, 1991 (4): 300-307. (in Chinese)

[16] CODD E F. A relational model of data for large shared data banks [J]. *Communications of the ACM*, 1970, 13(6): 377-387.

[17] BEERI C, BERNSTEIN P A, GOODMAN N. A sophisticate's introduction to database normalization theory [M]//Readings in artificial intelligence and databases. San Francisco: Morgan Kaufmann, 1989: 468-479.

[18] HODEL R E. An introduction to mathematical logic [M]. New York: Dover Publication, 2013.

[19] ARMSTRONG W W. Dependency structures of data base relationships [C]//Information Processing 74. Amsterdam: North-Holland, 1974: 580-583.

[20] BEERI C, BERNSTEIN P A. Computational problems related to the design of normal form relational schemas [J]. *ACM Transactions on Database Systems*, 1979, 4(1): 30-59.

[21] MAIER D. Minimum covers in relational database model [J]. *Journal of the ACM*, 1980, 27(4): 664-674.

[22] JOU J H, FISCHER P C. The complexity of recognizing 3NF relation schemes [J]. *Information Processing Letters*, 1982, 14 (4): 187-190.

[23] OSBORN S L. Testing for existence of a covering Boyce-Codd normal form [J]. *Information Processing Letters*, 1979, 8(1): 11-14.

[24] LIU G H, HAO Z X, CHEN Z J. A quick replacing algorithm for finding all candidate keys of a relation scheme [J]. *Chinese Journal of Computers*, 1998, 21 (10): 890-895. (in Chinese)

关系模式的主属性特征研究

徐雨露，刘国华*，于晓雪，刘长奇，朱东燕，张丽萌，何松达
东华大学 计算机科学与技术学院，上海 201620

摘 要：范式在关系数据库规范化理论中具有重要作用，其定义通过主属性或非主属性与键之间的函数依赖（functional dependency，FD）关系来确立。然而，判定属性是否为主属性是一个非确定性多项式时间完全问题（nondeterministic polynomial-time complete，NP-complete），这使得确定关系模式所属范式的复杂度较高。尽管如此，在特定情况下，识别主属性并非复杂任务。为此，该文将关系模式 $R(U, F)$ 中的属性集 U 根据其在函数依赖集 F 中的出现情况，划分为四个子集：U_1（仅出现在 FDs 左部的属性）、U_2（仅出现在 FDs 右部的属性）、U_3（同时出现在 FDs 左右部的属性）和 U_4（未出现在 F 中的属性）。通过深入分析，阐明了候选键作为关系模式唯一键的充要条件，探讨了 U_3 中主属性的特征，并归纳了一般主属性的特征。该研究有助于有效区分主属性识别中的复杂与简单情形，进一步加深对这一问题的理解。

关键词：范式；键；主属性；非主属性