Computer Science Journal of Moldova, vol.**17**, no.2(50), 2009

# Determination of the normalization level of database schemas through equivalence classes of attributes

Cotelea Vitalie

**Abstract**

In this paper, based on equivalence classes of attributes there are formulated necessary and sufficient conditions that constraint a database schema to be in the second, third or Boyce-Codd normal forms. These conditions offer a polynomial complexity for the testing algorithms of the normalizations level.

**Keywords:** Relational database schema, functional dependencies, equivalence classes of attributes, normal forms, polynomial algorithms.

## 1 Introduction

The anomalies that appear during database maintaining are known as insertion, update and deletion anomalies. These are directly related to the dependencies between attributes. A rigorous characterization of the quality grade of a database schema can be made through the exclusion of mentioned anomalies, with consideration of attributes dependencies, which offers the possibility to define some formal techniques for design of desirable relation schemes.

The process of design of some relation scheme structure with intend to eliminate the anomalies, is called *normalization*. Normalization consists in following a set of defined rules on data arrangement with the scope to reduce the complexity of scheme structures and its transformation into smaller and stable structures which will facilitate data maintenance and manipulation. There exist several normalization levels that are called *normal forms*.

The normal forms based on functional dependencies are *first normal form* (1NF), *second normal form* (2NF), *third normal form* (3NF) and *Boyce-Codd normal form* (BCNF). These forms have increasingly restrictive requirements: every relation in BCNF is also in 3NF, every relation in 3NF is also in 2NF and every relation in 2NF is in 1NF. A relation is in 1NF if every attribute contains only atomic values. 2NF is mainly of historical interest. 3NF and BCNF are important from a database design standpoint [1].

For example, the design of a 3NF database schema, through the synthesizing method, can be performed in a polynomial time [2]. Unfortunately, the problem of determination of the normalization level is known to be NP-complete [3, 4], because normalization testing requires finding the candidate keys and nonprime attributes. Firstly, the definitions of normal schemes (second, third or BCNF) contain the notion of key. But it is known that a relation can have an exponential number of keys under the number of all attributes of its scheme [5]. Secondly, the definitions of normal forms use the notions of prime and nonprime attributes, which are also related to key.

The problem of prime and nonprime attributes finding has been solved in a polynomial time [6]. In this paper necessary and sufficient conditions for a scheme to be in 2NF, 3NF or BCNF are defined. These conditions are described in terms of redundant and nonredundant equivalence classes of attributes and the computation of these classes can be performed in polynomial time [6]. Therefore, the determination of normalization level of a scheme is also polynomial. Thus a database designer may work in terms of attributes sets and data dependencies, and not in terms of keys. This approach can be a part of the database analysis and design toolset, i.e. for the automation of database design and testing.

In Section 2, most of the definitions needed in this paper are presented. In Section 3, several properties for equivalence classes of attributes, proved in [6], are given.

Besides this, the correlation is proven between nonredundant classes of attributes and the right and left sides of functional dependency that is inferred from a given set of functional dependencies (Theorem 3). In Sections 4, 5 and 6 there are presented necessary and sufficient conditions (Theorems 4-6), in terms of equivalence classes of attributes, for a relation scheme to be in 2NF, 3NF or BCNF, respectively. The final section is about algorithmic aspects, where it is shown that the determination of the normalization level of database schemas can be performed in polynomial time.

## 2 Preliminary notions

In this and in the next section, that will be as concise as possible, some definitions and statements used in this paper are presented.

Let $Sch(R, F)$ be a relation scheme, where $F$ is a set of functional dependencies defined on a set $R$ of attributes. The set of all functional dependencies implied by a given set $F$ of functional dependencies is called the *closure* of $F$ and is denoted as $F^+$, that is $F^+ = \{V \rightarrow W | F| = V \rightarrow W\}$ [7].

If $F$ is a set of functional dependencies over $R$ and $X$ is a subset of $R$, then the closure of the set $X$ with respect to $F$, written as $X^+$, is the set of attributes $A$ such that $X \rightarrow A$ can be inferred using the Armstrong Axioms, that is $X^+ = \{A | X \rightarrow A \in F^+\}$ [7].

Armstrong's Axioms are *sound* in that they generate only functional dependencies in $F^+$ when applied to a set $F$. They are *complete* in that repeated application of these rules will generate all functional dependencies in the closure $F^+$ [1].

Let $X$ and $Y$ be two nonempty finite subsets of $R$. The set $X$ is a determinant for $Y$ with respect to $F$ if $X' \rightarrow Y$ is not in $F^+$ for every proper subset $X'$ of $X$.

If $X$ is a determinant for $R$ with respect to $F$, then $X$ is a key for relation scheme $Sch(R, F)$. Note that some relation scheme may have more than one key.

An attribute $A$ is *prime* in $Sch(R, F)$ if $A$ is contained in some key of $Sch(R, F)$. Otherwise $A$ is nonprime in $Sch(R, F)$.

In what follows, it will be assumed that the set $F$ of functional dependencies is reduced [7].

Given a relation scheme $Sch(R, F)$, the set $F$ can be represented by a graph, called contribution graph [6] for $F$ and denoted by $G = (S, E)$, where:

- for every attribute $A$ in $R$, there is a vertex labeled by $A$ in $S$;
- for every functional dependence $X \to Y$ in $F$ and for every attribute $A$ in $X$ and every $B$ in $Y$ there is an edge $a = (A, B)$ in $E$ that is directed from vertex $A$ to vertex $B$.

Let $G = (S, E)$ be divided into strongly connected components. The relation of strong connectivity is an equivalence relation over the set $S$. So, there is a partition of set of vertices $S$ into pairwise disjoint subsets, that is, $S = \bigcup_{i=1}^{n} S_i$.

Let $S_1, ..., S_n$ be the strongly connected components of a graph $G = (S, E)$. Then the condensed graph [8] of $G$, $G^* = (S^*, E^*)$ is defined as follows:

$S^* = \{S_1, ..., S_n\}$ and

$E^* = \{(S_i, S_j) | i \neq j,\ (A, B)\ in\ E,\ A \in S_i\ and\ B \in S_j\}$.

Evidently the condensed graph $G^*$ is free of directed circuits. Over the set $S^*$ of vertices of graph $G^*$ a strict partial order is defined. Strict partial orders are useful because they correspond more directly to directed acyclic graphs. Vertex $S_i$ precedes vertex $S_j$, if $S_j$ is accessible from $S_i$.

From the ordered sequence of sets $S_1, ..., S_n$ a sequence of ordered nonredundant sets can be built $T_1, ..., T_n$, where $T_1 = S_1$ and $T_j = S_j - (\bigcup_{i=1}^{j-1} T_i)_F^+$ for $j = \overline{2, n}$. All empty sets are excluded from the sequence and a sequence of nonempty sets $T_1, ..., T_m$ is obtained, keeping the precedence of prior sets.

**Lemma 1. [6].** If $X \to Y \in F^+$ and $X$ is a determinant of $Y$ under $F$, then for every attribute $A \in (X - Y)$ there is an attribute $B \in Y$ so that in the contribution graph $G$ there exists a path from vertex $A$ to vertex $B$ and for every attribute $B \in (Y - X)$ there exists in $X$ an attribute $A$, from which the vertex $B$ can be reached.

## 3 Some properties of equivalence classes of attributes

In this section a brief overview of several properties of equivalence classes of attributes is given. And their proofs are presented in [6].

**Theorem 1.** ([**6**]**,** Theorem 2). Set $X$ is a determinant of set $S_1 \bigcup ... \bigcup S_n$ under $F$, if and only if $X$ is determinant of set $T_1 \bigcup ... \bigcup T_m$ under $F$.

**Lemma 2.** ([**6**], Lemma 3). If $X$ is a determinant under $F$ of set $T_1 \bigcup ... \bigcup T_m$, then $Z$, where $Z = X \bigcap (T_1 \bigcup ... \bigcup T_j)$ and $j = \overline{1, m}$, is a determinant for $T_1 \bigcup ... \bigcup T_j$ under $F$.

**Theorem 2.** ([**6**], Theorem 4). If set of attributes $X$ is a determinant of set $T_1 \bigcup ... \bigcup T_m$, then $X \bigcap T_i \neq \emptyset$, where $i = \overline{1, m}$.

**Corollary 1.** ([**6**], Corollary 3). If an attribute $A$ in $S_1 \bigcup ... \bigcup S_n$ is prime in scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$, then $A \in \bigcup_{i=1}^{m} T_i$.

**Corollary 2.** ([**6**], Corollary 4). If an attribute $A$ in $S_1 \bigcup ... \bigcup S_n$ is nonprime in scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$, then $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$.

**Theorem 3.** Let $X \to Y \in F^+$, where $X$ is a determinant for $Y$ under $F$ and $X, Y \subseteq T_1 \bigcup ... \bigcup T_m$. For a $T_j$, where $j = \overline{1, m}$, the following takes place: if $Y \bigcap T_j \neq \emptyset$, then $X \bigcap T_j \neq \emptyset$.

***Proof.*** The soundness of this statement is proven by contradiction: let $Y \bigcap T_j \neq \emptyset$, but $X \bigcap T_j = \emptyset$. Evidently that $X \subseteq T_1 \bigcup ... \bigcup T_{j-1} \bigcup T_{j+1} \bigcup ... \bigcup T_m$ and $X \to (Y \bigcap T_j) \in F^+$. Let $X'$, where $X' \subseteq X$, is a determinant for $Y \bigcap T_j$ under $F$. According to Lemma 1, on the contribution graph of set $F$ of dependencies, from every vertex labeled with an attribute in $X'$ there exists a path to a vertex labeled with an attribute in $Y \bigcap T_j$. Thereby, $X' \subseteq T_1 \bigcup ... \bigcup T_{j-1}$. But in this case, $T_j$ is redundant. A contradiction has been reached.

Using above structures and statements it will be shown that the problem of determination of the normalization level has polynomial complexity. In the following sections, in terms of equivalence classes of attributes, sufficient and necessary conditions for a relation scheme to be in a normal form are presented.

## 4 Second normal form

Thus, the relation scheme in the 2NF can be defined:

**Definition 1. [9].** Scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ is in the 2NF under a set of functional dependencies $F$, if it is in 1NF and each nonprime attribute in $\bigcup_{i=1}^{n} S_i$ doesn't partially depend on every key for $Sch$. Database schema is in the 2NF, if each constituent relation scheme is in the 2NF.

**Definition 2. [9].** Let $X \to A \in F$ be a nontrivial functional dependency (namely $A \notin X$). An attribute $A$ is called *partially dependent* on $X$, if there exists a proper subset $X'$ of set $X$, such that $X' \to A \in F^+$. If such a proper subset doesn't exist, then $A$ is called that *completely depends* on $X$.

**Proposition 1.** If set of attributes $X$ is a determinant for attribute $A$ under set of attributes $F$, then $A$ completely depends on $X$.

The next theorem gives a characterization of the 2NF in terms of equivalence classes of attributes.

**Theorem 4.** Relation scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ is in the 2NF, if and only if it is in the 1NF and for every $T_j$, $j = \overline{1,m}$, $(\bigcup_{i=1}^{m} T_i - T_j)^+ = \bigcup_{i=1}^{m} T_i - T_j$ takes place.

***Proof. Necessity.*** Let scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ be in the 2NF. Then every nonprime attribute $A$, that is a member of set $\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i$ completely depends on every determinant $X$ of set $S_1 \bigcup ... \bigcup S_n$. According to Theorem 1, $X$ is a determinant of set $T_1 \bigcup ... \bigcup T_m$. In addition, $X \subseteq T_1 \bigcup ... \bigcup T_m$. Assuming to the contrary, that $Sch$ is in the 2NF, but there is an attribute $A \in (\bigcup_{i=1}^{m} T_i - T_j)^+$ such that $A \notin (\bigcup_{i=1}^{m} T_i - T_j)$. There are two cases: either $A \in T_j$, or $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$.

Let $A \in T_j$. From the construction of contribution graph, follows that $(T_1 \bigcup ... \bigcup T_j) \to A \in F^+$. Because $A \in (\bigcup_{i=1}^{m} T_i - T_j)^+$, namely $A \in (\bigcup_{i=1}^{j-1} T_i - T_j)^+$, then $(T_1 \bigcup ... \bigcup T_{j-1}) \to A \in F^+$. But this contradicts the fact that set $T_j$ is nonredundant.

Let $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$. If $X$ is a determinant for $T_1 \bigcup ... \bigcup T_m$ under $F$, taking into account Lemma 2, $X \bigcap (T_1 \bigcup ... \bigcup T_{j-1})$ is a determinant for $T_1 \bigcup ... \bigcup T_{j-1}$. So that $(T_1 \bigcup ... \bigcup T_{j-1}) \to A \in F^+$, then $(X \bigcap (T_1 \bigcup ... \bigcup T_{j-1})) \to A \in F^+$. In other words, the nonprime attribute $A$ partially depends on determinant $X$. That is $A$ partially depends on key $X$, fact that contradicts the assumption that scheme $Sch$ is in the 2NF.

***Sufficiency.*** Let scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ be in the 1NF and for every $T_j$, $j = \overline{1,m}$, the following equality takes place: $(\bigcup_{i=1}^{m} T_i - T_j)^+ = \bigcup_{i=1}^{m} T_i - T_j$. It will be proven that scheme $Sch$ is in the 2NF. Two cases are possible: either $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i) = \emptyset$, or $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i) \neq \emptyset$.

If $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i) = \emptyset$, then scheme doesn't contain nonprime attributes and, therefore, scheme is in the 2NF and it is even in the third.

If $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i) \neq \emptyset$, that is in the case when set of nonprime attributes is not empty, results that every nonprime attribute $A$ completely depends on $T_1 \bigcup ... \bigcup T_m$, furthermore it completely depends on determinant $X$ under $F$ of set $T_1 \bigcup ... \bigcup T_m$. So, the scheme is in the 2NF.

**Corollary 3.** Scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ is in the 2NF, if and only if for every $i = \overline{1,m}$ $T_i = S_i$ holds.

***Proof.*** The soundness of this statement follows from the fact that $(\bigcup_{i=1}^{m} T_i - T_j)^+ = \bigcup_{i=1}^{m} T_i - T_j$ takes place when $T_i = S_i$ holds for every $i = \overline{1,m}$ and vice versa.

# 5 Third normal form

In this section a characterization of the 3NF is given through the equivalence classes.

**Definition 3. [9].** Scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ is in 3NF under a set of functional dependencies $F$, if it is in the 1NF and every nonprime attribute doesn't transitively depend on a key of scheme $Sch$. Database schema is in the 3NF, if every constituent relation scheme is in the 3NF.

**Definition 4. [10].** Let scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$, $V, W \subseteq \bigcup_{i=1}^{n} S_i$ and $A \in \bigcup_{i=1}^{n} S_i$. It is considered that the attribute $A$ *transitively depends* on $V$ through $W$, if the following conditions are all satisfied:

1. $V \to W \in F^+$;
2. $W \to V \notin F^+$ (namely $V$ doesn't functionally depend on $W$);
3. $W \to A \in F^+$;
4. $A \notin VW$.

**Theorem 5.** Relation scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ is in the 3NF, if and only if $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ is a determinant for $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ under $F$.

***Proof. Necessity.*** Let scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ be in the 3NF. Then scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ is also in the 2NF and each attribute $A$, where $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$, fully functionally depends on key $X$ of scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$, namely it is fully functionally dependent on determinant $X$ of set $T_1 \bigcup ... \bigcup T_m$ under $F$. In addition, no attribute $A$, where $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$, transitively depends on $X$. That is, there doesn't exist any dependency $W \to A \in F^+$, such that $W \subseteq (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ and $A \notin XW$. Therefore, dependency $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i) \to (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ is reduced on the left side, fact that confirms that $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ is a determinant for $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ under $F$.

***Sufficiency.*** Assume $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ is a determinant for $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ under $F$. Let $X$ be a determinant of set $T_1 \bigcup ... \bigcup T_m$ under $F$. According to Theorem 1, $X$ is a determinant of set $S_1 \bigcup ... \bigcup S_n$. Hence, $X \to (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i) \in F^+$ holds. Because $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ is a determinant for $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ under $F$, then there doesn't exist any dependency $W \to A \in F^+$, such that $W \subseteq (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$, $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ and $A \notin XW$. That is, all nonprime attributes $A$ don't depend transitively on determinant $X$.

## 6 Boyce-Codd normal form

The concept of BCNF is refined from the notion of 3NF. In the determination of a database schema being in BCNF, a given set $F$ of functional dependencies is used.

**Definition 5. [11].** Relation scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ is in BCNF under set $F$ of functional dependencies, if it is in the 1NF and for every nontrivial dependency $V \to A \in F^+$ $V \to \bigcup_{i=1}^{n} S_i \in F^+$ takes place, that is, the left side of each functional dependency functionally determines all attributes of scheme.

**Theorem 6.** Scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ is in the normal form Boyce-Codd, if and only if it is in the 3NF and for every $T_j$, $j = \overline{1,m}$, the set of attributes $(\bigcup_{i=1}^{m} T_i - T_j)$ is a determinant for $(\bigcup_{i=1}^{m} T_i - T_j)$ under $F$.

***Proof. Necessity.*** Let scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ be in BCNF. Then for every nontrivial functional dependency $V \to A \in F^+$, that is, the case when $A \notin V$, $V \to \bigcup_{i=1}^{n} S_i \in F^+$ holds. Based on the reflexivity rule, $(\bigcup_{i=1}^{m} T_i - T_j) \to (\bigcup_{i=1}^{m} T_i - T_j) \in F^+$. If it's supposed that $(\bigcup_{i=1}^{m} T_i - T_j)$ is not a determinant for $(\bigcup_{i=1}^{m} T_i - T_j)$ under $F$, that is, if there exists a set of attributes $V \subset (\bigcup_{i=1}^{m} T_i - T_j)$, so that $V \to (\bigcup_{i=1}^{m} T_i - T_j) \in F^+$, then the last dependency is not trivial. By the definition of BCNF $V \to \bigcup_{i=1}^{n} S_i \in F^+$ holds. But this functional dependency contradicts the fact that every determinant $X$of set $\bigcup_{i=1}^{n} S_i$ and consequently a set $\bigcup_{i=1}^{m} T_i$ contains, according to Theorem 2, attributes in $T_j$ $j = \overline{1,m}$ too, namely $X \bigcap T_j \neq \emptyset$ for $j = \overline{1,m}$.

***Sufficiency.*** Let scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ be in the 3NF and for every $T_j$, $j = \overline{1,m}$, the set of attributes $(\bigcup_{i=1}^{m} T_i - T_j)$ is a determinant for $(\bigcup_{i=1}^{m} T_i - T_j)$ under $F$. It will be proven that scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ is in BCNF.

Let scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ not be in BCNF. In this case, there is a nontrivial functional dependency $V \to A \in F^+$, so that $V \to \bigcup_{i=1}^{n} S_i \notin F^+$ holds. Then it can be stated that $V \to \bigcup_{i=1}^{m} T_i \notin F^+$. Without constraining the generality, let $V$ be a determinant for $A$ under $F$. From the construction of the contribution graph and from the fact that dependency $V \to A$ is reduced, three cases can be examined (other cases don't exist):

1. $V \subseteq (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ and $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$, that is, left and right sides are formed just from nonprime attributes,

2. $V \subseteq \bigcup_{i=1}^{m} T_i$ and $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ - the left side is formed from prime attributes, and the right side consists of a nonprime attribute.

3. $V \subseteq \bigcup_{i=1}^{m} T_i$ and $A \in \bigcup_{i=1}^{m} T_i$, that is, left and right sides are formed just from prime attributes,

Suppose that $V \subseteq (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ and $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$, then nonprime attribute $A$ would transitively depend through $V$ on every determinant of set $\bigcup_{i=1}^{n} S_i$ and then scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ will not be in the 3NF, which contradicts the hypothesis.

If it is considered that $V \subseteq \bigcup_{i=1}^{m} T_i$ and $A \in (\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$, then nonprime attribute $A$ would partially depend on a determinant of set $\bigcup_{i=1}^{n} S_i$ therefore scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ will not be in the 2NF, that is, neither in the third, fact that contradicts the hypothesis.

If it is considered that $V \subseteq \bigcup_{i=1}^{m} T_i$ and $A \in \bigcup_{i=1}^{m} T_i$, then the set $(\bigcup_{i=1}^{m} T_i - T_j)$ of attributes is not a determinant for $(\bigcup_{i=1}^{m} T_i - T_j)$ under $F$. Indeed, let $m > 1$ and $A \in T_k$. Then by Theorem 3 and the construction way of drawing the contribution graph, it can exist two cases either $V \subseteq T_k$, or $V \not\subset T_k$, but $V \subseteq (T_l \bigcup T_{l+1} \bigcup ... \bigcup T_k)$, where $V \bigcap T_i \neq \emptyset$, $i = \overline{l,k}$. Evidently $m > k - l + 1$, but in this case there exists a $T_j$, where $V \bigcap T_j = \emptyset$, so that $(\bigcup_{i=1}^{m} T_i - T_j)$ is not a determinant for $(\bigcup_{i=1}^{m} T_i - T_j)$ under $F$, because $((\bigcup_{i=1}^{m} T_i - \{A\}) - T_j) \to (\bigcup_{i=1}^{m} T_i - T_j) \in F^+$.

## 7 Algorithms' complexities

Based on the above characterization, the polynomiality of the normal form testing problem can be proved. A few comments about the complexity of the algorithms for finding the normal form of scheme are made below.

Both construction of equivalence classes of scheme's attributes and redundancy elimination from these classes have a complexity $O(|R| \cdot ||F||)$ [6].

It is not hard to calculate the complexity of algorithms that determine whether a scheme is in the second, third or Boyce-Codd normal form. That is, $|R| \cdot ||F||$ for each of these algorithms. This is explained through the fact that the complexity of calculation of the classes of nonredundant attributes exceeds the complexity of calculation of the verification conditions that determine if a scheme is in one of the enumerated forms.

Thus, if nonredundant classes $\bigcup_{i=1}^{m} T_i$ are built, then calculation of the condition for the scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ to be in the 2NF (that is, if for every $T_j$, $j = \overline{1,m}$, $(\bigcup_{i=1}^{m} T_i - T_j)^+ = \bigcup_{i=1}^{m} T_i - T_j)$ requires a time $O(|NonRedEquivClasses| \cdot ||F||)$. Therefore the time is $O(|R| \cdot ||F||)$.

Computation of the condition for the scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ to be in the 3NF (that is, if $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i)$ is a determinant for $(\bigcup_{i=1}^{n} S_i - \bigcup_{i=1}^{m} T_i))$ requires a time $O(||F||)$.

Similarly, verification of the condition for the scheme $Sch = (\bigcup_{i=1}^{n} S_i, F)$ to be in BCNF (that is, if for every $T_j$, $j = \overline{1,m}$, set of attributes $(\bigcup_{i=1}^{m} T_i - T_j)$ is a determinant for $(\bigcup_{i=1}^{m} T_i - T_j)$ under $F$) requires a time $O(|R| \cdot ||F||)$.

## References

[1] Ramakrishnan, Raghu and Gehrke, Johannes. *Database Management Systems.* Second Edition, McGraw-Hill Higher Education, 2000, 900 pp.

[2] Bernstein, Philip A. *Synthesizing Third placeNormal Form Relations from Funcional Dependencies.* ACM Trans. Database Syst., V.1, N 4, 1976, p.277–298.

[3] Beeri, C.; Bernstein, P.A. *Computational Problems Related to the design of placeNormal Form Relations Schemes.* ACM Trans. Database Syst., V.4, N 1, March 1979, p.30–59.

[4] Jou, J.H., Fisher, P.C. *The complexity of recognizing 3FN relation schemes.* Inform. Process. Letters 14, 1982, p.187–190.

[5] Yu C.T., Johnson D.T. *On the complexity of finding the set of candidate keys for a given set of functional dependencies.* Information Processing Letters, V.5, N.4, 1978, p.100–101.

[6] Cotelea, Vitalie. *An approach for testing the primeness of attributes in relational schemas.* Computer Science Journal of Moldova, Chisinau, Vol.17, Nr.1(49), 2009, p. 89–99.

[7] Maier, D. *The theory of relational database.* Computer Science Press, 1983, 637 p.

[8] Even, Shimon. *Graph Algorithms.* Computer Science press, 1979, 250 p.

[9] Codd, E.F. *Further Normalization of the data base relational model.* Data Base Systems, R. Rustin (ed), Prentice Hall, 1972, p. 33–64,.

[10] Chao-Chih Yang. *Relational Databases.* Englewood Cliffs. NJ, Prentice-Hall, 1986, 260 p.

[11] Codd, E.F. *Recent Investigation in Relation Data Base Systems.* IFIP Congress, 1974, p. 1017–1021.

Vitalie Cotelea Received May 27, 2009

Vitalie Cotelea
Academy of Economic Studies of Moldova
Phone: (+373 22) 40 28 87
E-mail: *vitalie.cotelea@gmail.com*