COMPLETENESS OF ARMSTRONG'S AXIOMATIC

Dmitriy Buy, Anna Puzikova

Abstract: This paper presents a rigorous and convincing proof that Armstrong's axiomatic system (as the foundation of relational database normalization theory) is complete and sound within the paradigm of mathematical logic: the relations of syntactic and semantic entailment are introduced and it is shown that they coincide. The properties of set-theoretic data structure functional constraints have been used as mathematical framework.

Keywords: scheme, functional dependence, semantically deduced, normalization

Introduction

The dependability of hardware-software systems is largely defined by the reliability of relational databases in use. In its turn, the reliability of the above-mentioned databases is dependant on their logical design. To eliminate the known anomalies (update, insertion, deletion) the database normalization is required. The process of normalization is based upon functional dependencies theory the foundation of which is made by Armstrong's axioms and their completeness. The overview of research sources has shown that Armstrong's axiomatic system lacks the proof of completeness that would comply with mathematical rigor; in fact, this vitally important result has a seemingly announcement status. The situation is aggravated by the fact many widely spread CASE-means support normalization. Furthermore, the assertion about the hardware-software system dependability without utilizing formal methods, in the first place mathematical ones, is out of the question.

The purpose of this paper is the construction of a comprehensive and exhaustive proving of a well-known result in the theory of relational databases – Armstrong's axiomatic system completeness result for functional dependencies.

Let A – a set of attributes (names), t – a chart (table), R – a scheme of the chart (table) t (an arbitrary finite set of attributes), X, Y, W, Z– subsets of scheme R, s, s_1, s_2 – the rows of chart t. Let functional dependence[1] $X \to Y$, is taking place at chart t, if for two arbitrary rows s_1, s_2 of table t which coincide on the set of attributes X, there is their equality on the set of attributes Y, that is:

$$(X \to Y)(t) = True \overset{def}{\Leftrightarrow} \forall s_1, s_2 \in t\left(s_1|X = s_2|X \Rightarrow s_1|Y = s_2|Y\right). \qquad (1)$$

Let table t of the scheme R be the model of a set of functional dependences F, if each functional dependence $X \to Y \in F$ is carried out at table t:

$$t\ model\ F \overset{def}{\Leftrightarrow} \forall (X \to Y)(X \to Y \in F \Rightarrow (X \to Y)(t) = true). \qquad (2)$$

Semantic Succession

Functional dependence $X \to Y$ is semantically deduced ($|=$) from the set of functional dependences F, if at each table $t(R)$ which is the model of a set FD F, functional dependence $X \to Y$ is carried out as well:

$$F \models X \to Y \overset{def}{\Leftrightarrow} \forall t(R)\ (t\ model\ F) \Rightarrow (X \to Y)(t) = true). \qquad (3)$$

Lemma 1. $\forall t\ (X \to Y)(t) = true$, $Y \subseteq X$ (*axiom of reflexive property* according to William Armstrong).

Proof. Let us consider the rows of table t for which $s_1|X = s_2|X$ is carried out. We restrict both parts of this equality according to the set Y: $(s_1|X)|Y = (s_2|X)|Y$. According to the property of restriction operator $(U|Y)|Z = U|(Y \cap Z)$[1]

it follows $s_1|(X\cap Y)=s_2|(X\cap Y)$. Consequently, and from the condition $Y\subseteq X$ we get $s_1|Y=s_2|Y$, so $(X\to Y)(t)=true$ □

Corollary fact. $\emptyset\models X\to Y$, $\forall Y\subseteq X$ (such functional dependence is called *trivial*).

Lemma 2. $(X\to Y)(t)=true\Rightarrow(X\cup Z\to Y\cup Z)(t)=true$, $\forall Z\subseteq R$ (the rule of *completion).*

Proof. Let $s_1\mid(X\cup Z)=s_2\mid(X\cup Z)$ is carried out for some rows of table *t*, hence, according to the axiom of reflexive property, we get $s_1\mid X=s_2\mid X$ and $s_1\mid Z=s_2\mid Z$. Using the property of distributive restriction as for the uniting[1], previous equalities and condition $s_1\mid X=s_2\mid X\Rightarrow s_1\mid Y=s_2\mid Y$, it follows: $s_1\mid(Y\cup Z)=s_1\mid Y\cup s_1\mid Z=s_2\mid Y\cup s_2\mid Z=s_2\mid(Y\cup Z)$. □

Corollary fact. $F\models X\to Y\Rightarrow F\models X\cup Z\to Y\cup Z$.

Proof. $F\models X\to Y\overset{def}{\Leftrightarrow}\forall t(R)$ $(t\ model\ F\Rightarrow\ (X\to Y)(t)=true)$. According to the rule of completion, it follows $(X\cup Z\to Y\cup Z)(t)=true$, from this follows $F\models X\cup Z\to Y\cup Z$ □

Lemma 3. $(X\to Y)(t)=true$ & $(Y\to Z)(t)=true\Rightarrow(X\to Z)(t)=true$ (the rule of *transitivity*).

Proof. From the conditions $(X\to Y)(t)=true$ and $(Y\to Z)(t)=true$, we get $\forall s_1,s_2\in t$ $(s_1|X=s_2|X\Rightarrow s_1|Y=s_2\mid Y$ & $s_1\mid Y=s_2\mid Y\Rightarrow s_1\mid Z=s_2\mid Z$). From this follows $\forall s_1,s_2\in t$ $(s_1|X=s_2|X\ \Rightarrow\ s_1|Z=s_2|Z)$ and $(X\to Z)(t)=true$ □

Corollary fact. $F\models X\to Y$ & $F\models Y\to Z\Rightarrow F\models X\to Z$.

Proof. $F\models X\to Y$ & $F\models Y\to Z$ $\overset{def}{\Leftrightarrow}$ $\forall t(R)$ $(t\ model\ F)$ $\Rightarrow$ $(X\to Y)(t)=true$ & $(Y\to Z)(t)=true$. According to the rule of transitivity, it follows $(X\to Z)(t)=true$. From this follows $F\models X\to Z$ □

Syntactic Succession ($\mid-$)

Let us assume that functional dependence $X\to Y$ is syntactically derived from F (indentation: $F\mid -X\to Y$), if there is a finite succession $\varphi_1,\varphi_2,...,\varphi_{m-1},\varphi_m$, and it is such that $\varphi_m=X\to Y$ and $\forall i=\overline{1,m-1}$ each φ_i is either the axiom of reflexive property or refers to *F* or is derived with the help of some rule of deducing from the latter in this succession φ_j,φ_k, $j,k<i$[3]. Let defined certain set of functional dependences *F*. *Closure* $[F]$ is a set of all FD that are syntactically deduced from *F*:

$$[F]\overset{def}{=}\{X\to Y\mid F\mid -X\to Y\}. \tag{4}$$

Lemma 4. Correlations are carried out (properties of closure of functional dependences set):

1. $F\subseteq[F]$.

Proof. Let FD $X\to Y\in F$, then $F\mid -X\to Y$ with a number of steps of proving $k=1$, hence, $X\to Y\in[F]$ □

2. $[[F]]=[F]$.[2]

Proof. According to property 1, we have $[F]\subseteq[[F]]$. Let us prove the reverse inclusion $[[F]]\subseteq[F]$. Let $X\to Y$ – arbitrary FD, such that $X\to Y\in[[F]]$. Then there is a finite succession FD $\varphi_1,\varphi_2,...,\varphi_{m-1},\varphi_m$, such that $\varphi_m=X\to Y$ and $\forall i=\overline{1,m-1}$ each φ_i is either the axiom of reflexive property or refers to $[F]$ or is derived with the help of some rule of deducing from the previous in this succession FD φ_j, φ_k, $j,k < i$. Let us make a new succession according to such rules:

- if $\varphi_i \in F$ then φ_i is the axiom of reflexive property, then let us write down this FD without any changes;
- if $\varphi_i \notin F$ and $\varphi_i \in [F]$, then according to the definition of closure this FD has a finite succession of deducing $\mu_1, \mu_2, ..., \mu_p$ from F. Instead of FD φ_i let us insert the succession of its deducing;
- if φ_i is derived according to any rule of deducing from the previous in this succession FD φ_j, φ_k, $j,k < i$, at that according to the first two points φ_j, φ_k either refer to F or have the sequence of deducing from F then also we write down FD φ_i without any changes. Created in such a way the succession is a succession of deducing FD $X \to Y$ from F, that is $F \mid -X \to Y$, hence, $X \to Y \in [F]$ □

Conclusion. Closure $[F]$ is the minimal set which has F such that with application of Armstrong's axioms to it is impossible to obtain any functional dependence which wouldn't refer to $[F]$.

From axioms and rules of deducing indicated above in order of simplification of practical calculation of closure [F] set FD F is possible to get other rules of deducing.

Lemma 5. $\{X_1 \to Y_1,\ X_2 \to Y_2\} \mid -\ X_1 \cup X_2 \to Y_1 \cup Y_2$ (the rule of *composition*)[4].

Proof. Let us make a succession of deducing for FD $X_1 \cup X_2 \to Y_1 \cup Y_2$:

1. $X_1 \to Y_1$;
2. $X_1 \cup X_2 \to Y_1 \cup X_2$ (with 1 according to the rule of completion);
3. $X_2 \to Y_2$;
4. $Y_1 \cup X_2 \to Y_1 \cup Y_2$ (with 3 according to the rule of completion);
5. $X_1 \cup X_2 \to Y_1 \cup Y_2$ (with 2 and 4 according to the rile of transitivity). □

Lemma 6. $X \to Y_1 \cup Y_2 \mid -\ X \to Y_1\ \&\ X \to Y_2$ (the rule of *decomposition*)[4].

Proof. Taking into consideration inclusions $Y_1 \subseteq Y_1 \cup Y_2$ and $Y_2 \subseteq Y_1 \cup Y_2$ having applied to FD $X \to Y_1 \cup Y_2$ the axiom of reflexivity we get: $X \to Y_1\ \&\ X \to Y_2$. □

Closure [X] of a set *X* is the union of right parts of all FD collating of the form $X \to Y$, which are syntactically derived from the *F* set:

$$[X]_F \stackrel{def}{=} \bigcup_{X \to Y \in \{F\}} Y \ . \qquad (5)$$

Lemma 7. Correlations are carried out (properties of closure of X set):

1. $X \subseteq [X]$.

Proof. Let us make a succession of deducing which has one FD $X \to X$ (according to the axiom of reflexivity). Hence, $F \mid -X \to X$ and according to the definition of closure of X set follows $X \subseteq [X]$ □

2. $F \mid -X \to [X]$.

Proof. Let us make a succession which includes all FD $X \to Y$, such as that $F \mid -X \to Y$. According to the result of lemma 5, FD $X \to \bigcup_{X \to Y \in [F]} Y$ is carried out, which is equal to FD $X \to [X]$. From the created succession of deducing follows that $F \mid -X \to [X]$.▪

3. $X \to Z \notin [F] \Rightarrow Z \square [X] \subset R$.

Proof by contradiction. Let us assume that $X \to Z \notin [F] \Rightarrow Z \subseteq [X] \subset R$, then according to the definition of closure of X set, it follows that: $\exists n: Y_1, Y_2, ..., Y_n$, for which $Z \subseteq Y_1 \cup Y_2 \cup ... \cup Y_n$, at that $X \to Y_1, X \to Y_2, ..., X \to Y_n \in [F]$. From this follows that $\exists i$, $i = \overline{1,n}$ such as that $Z = Y_i$, hence $X \to Z \in [F]$, which contradicts the assumption □

Statement. If FD $X \to Y$ is syntactically deduced from the set FD *F*, then FD $X \to Y$ is deduced from *F* semantically:

$$F \mid -X \to Y \Rightarrow F \models X \to Y. \qquad (6)$$

The proof is carried out by the method of induction at the length of proving which is the quantity of elements in this succession:

1. Basis: $i = 1$. $\varphi_1 = X \to Y$, it means that FD $X \to Y$ is either an axiom or $X \to Y \in F$.
2. Let us assume that for every i, $i = \overline{1, k-1}$ implication is carried out, that is for every FD from the succession of deducing $\varphi_1, \varphi_2, ..., \varphi_{k-1}$, semantic succession $F \models \varphi_i$ takes place.
3. Inductive step: we check the implementation of statement for $i = k$ of sequence of deducing $\varphi_1, \varphi_2, ..., \varphi_k$, such that accomplishes $F \models \varphi_i$ $\forall i = \overline{1, k-1}$ (according to assumption) and $\varphi_k = X \to Y$. Let us consider the cases:

3.1. If $X \to Y$ is an axiom hence, according to the corollary to lemma 1 it follows $\varnothing \models X \to Y$.

3.2. If $X \to Y \in F$, then $F \models X \to Y$ is carried out trivially.

3.3. $X \to Y$ is deduced from certain φ_j, $j = \overline{1, k-1}$ according to the rule of completion. According to the corollary to lemma 2, it follows: $F \models \varphi_j \Rightarrow F \models X \to Y$.

3.4. $X \to Y$ is deduced from certain φ_j, φ_g, $j, g < k$ according to the rule of transitivity. According to the corollary to lemma 3, it follows: $F \models \varphi_j, F \models \varphi_g \Rightarrow F \models X \to Y$ □

Statement. If FD $X \to Y$ is syntactically deduced from the set FD *F*, then FD $X \to Y$ is deduced semantically from *F*: $F \models X \to Y \Rightarrow F \mid -X \to Y$;

or the same: $F \models X \to Y \Rightarrow X \to Y \in [F]$.

Proof by contradiction. Let the given set of functional dependences is *F*. Let us show that if FD $X \to Y \notin [F]$, then FD is not deduced semantically from *F*. We build a model where:

1. Every FD which belongs to $[F]$ is carried out;
2. If FD does not belong to closure it is not carried out.

Let model t of the scheme $R = \{A_1, A_2, ..., A_n\}$ comprise two rows $t = \{s_1, s_2\}^2$, where $s_1(A_i) = a_i, i \in \overline{1,n}$,

$$s_2(A_i) = \begin{cases} a_i, & if \quad A_i \in [X]; \\ b_i, & b_i \neq a_i, \; if \quad A_i \notin [X]. \end{cases} \qquad (7)$$

A_1	A_2	...	...	A_n
a_1	a_2	...	...	a_n

Let $X \to Y$ – is an arbitrary functional dependence which does not belong to $[F]$. Let us show that this FD is not carried out at *t*. Inclusion of $X \subseteq [X]$ (property 1 of lemma 7) means that an *X* set is made of all attributes which refer to closure $[X]$, hence, the meaning of the rows on this attributes coincide. From this follows that $s_1 | X = s_2 | X$. At that $X \to Y \notin F$, then $Y \square [X]$ (property 3 of lemma 7), that is there is an attribute $A \in Y$ & $A \notin [X]$, on which the rows s_1 i s_2 acquire different meanings, hence it follows $s_1 | Y \neq s_2 | Y$, consequently $(X \to Y)(t) = false$.

Let us show that arbitrary FD $W \to Z \in [F]$ is carried out at t.

If $W \square [X]$ then $s_1 | W \neq s_2 | W$. $W \subseteq [X]$. Let's make a succession:

1. $F | -X \to [X]$ (property 2 of lemma 7);
2. $[X] \to W$, $\forall W \subseteq [X]$ (axiom of reflexive property);
3. $X \to W$ (with 1 and 2 according to the rile of transitivity);
4. $F | -W \to Z$;
5. $X \to Z$ (with 3 and 4 according to the rile of transitivity).

From the meaning of $[X]$ set X follows $X \to Z \Rightarrow Z \subseteq [X]$, hence, for the created model t : $s_1 | Z = s_2 | Z$. By analogy, taking into consideration inclusions $W \subseteq [X]$, it follows $s_1 | W = s_2 | W$, hence: $(W \to Z)(t) = true$ □

Theorem. The relations of semantic and syntactic succession coincide.

$F \models X \to Y \Leftrightarrow F \,|\!- X \to Y$, or the same: $F \models X \to Y \Leftrightarrow X \to Y \in [F]$.

Conclusion

The paper presents a rigorous proving that Armstrong's axiomatic system is complete and sound by way of set-theoretic data structure application. The methods of completeness establishment consists in introducing the two relations based on functional dependencies – the relations of syntactic and semantic entailment – and the establishment of these relations coincidence. The inclusion of syntactic entailment relation into semantic entailment relation ensures correctness of Armstrong's axiomatic system; reverse inclusion secures completeness of the axiomatic system under study. Further research consists in constructing a comprehensive and rigorous theory of relational databases normalization: the definitions of normal forms are to be specified and mathematical results of the correctness of normalization algorithms are to be obtained. This theory will make a reliable foundation of normalization theory, will substantiate the existing normalization algorithms and will show ways of obtaining all more efficient possible normal forms. The results obtained can find their application in the dependability analysis and complex system construction the foundation of which is made by the relational databases.

Notes:

1. Buy Dmitro B., Kahuta Nadiya D., *"Properties of set-theoretic constructions of complete appearance, limitation, cofinality, to compatibility"* (paper presented at the all Ukrainian conference of scientific "Modern problems of the applied mathematics and informatics", 3-5 October 2006, Lvov), 25.
2. Maier David, *The theory of relational Databases* (Moscow: World, 1987), 608.
3. Lyndon Roger, *Notes on logic* (Moscow: World, 1968), – 128.
4. Chris Date, *An introduction to Database Systems* (Moscow: Williams Publishing House, 2005), 1328.
5. Pasichnik V.V., Reznichenko V.A., *Organization of Database and knowledge* (K.: BHV, 2006), – 384.

Authors:

DMITRIY BUY, doctor of Science (DSc), senior Researcher, professor of Theory and Technology of Programming Department. Head of the Research Society "Programming Problems" of the School of Cybernetics at the Taras Shevchenko Kyiv National University. Author of 135 research papers and a monograph. A senior doctoral dissertation on 'Theory of Programming Algebras of Composite Type and Its Application "(2002).

Scientific interests: relational database, theory of fixed points, compositional semantics of SQL language, modern CASE-tools, formal methods of software systems designing.

ANNA PUZIKOVA, the assistant of computer sciences department. Scientific interests: mathematical models of the theory of databases.