

Обобщение типизированных зависимостей включения с неопределенными значениями

ЗЫКИН С.В.

ИМ СО РАН (ОмФ)

2023

Ссылочная целостность данных

№ студента	№ группы	ФИО студента
1	1	Иванов И.И.
1	2	Сидоров С.С.
2	1	Петров П.П.
2	2	Ковалев К.К.
3	2	Стоянов В.С.
4	2	Бессараб О.П.
5	2	Нефедов К.А.
6	2	Вяткин М.С.
7	2	Драница А.А.
8	2	Алексенко С.В.
9	2	Белоусов П.О.
10	2	Ефимов Е.С.

№ студента	№ группы	№ предмет	Оценка
1	NULL	1	5
1	2	1	4
1	2	2	3
1	2	3	5
1	2	5	4
NULL	1	2	4
2	2	2	3
2	2	3	4
NULL	2	5	4
4	2	3	4
4	2	4	4
4	NULL	5	3
5	2	1	4
5	2	2	5
5	2	3	4
7	2	3	3
7	2	4	3
7	2	5	3
8	2	6	6
9	2	1	3



Обзор существующих работ

- Casanova M., Fagin R., Papadimitriou C. Inclusion Dependencies and Their Interaction with Functional Dependencies Journal of Computer and System Sciences, 28:1 (1984), 29–59.

Обзор существующих работ

- Casanova M., Fagin R., Papadimitriou C. Inclusion Dependencies and Their Interaction with Functional Dependencies Journal of Computer and System Sciences, 28:1 (1984), 29–59.
- Пусть $U = \{A_1, A_2, \dots, A_n\}$ – множество атрибутов, определенных в БД, $[R_i]$ – множество атрибутов, на которых определено отношение R_i , $[R_i] \subseteq U$, $1 \leq i \leq k$,
 $\mathfrak{R} = (R_1, R_2, \dots, R_k)$ – БД, $S = \{[R_1], [R_2], \dots, [R_k]\}$ – схема БД.

Определение 1

Пусть $[R_i]$ и $[R_j]$ – схемы отношений (не обязательно различные), $V \subseteq [R_i]$ и $W \subseteq [R_j]$, $|V| = |W|$, тогда соотношение $R_i[V] \subseteq R_j[W]$ называется зависимостью включения.

Обзор существующих работ

- Casanova M., Fagin R., Papadimitriou C. Inclusion Dependencies and Their Interaction with Functional Dependencies Journal of Computer and System Sciences, 28:1 (1984), 29–59.
- Пусть $U = \{A_1, A_2, \dots, A_n\}$ – множество атрибутов, определенных в БД, $[R_i]$ – множество атрибутов, на которых определено отношение R_i , $[R_i] \subseteq U$, $1 \leq i \leq k$,
 $\mathfrak{R} = (R_1, R_2, \dots, R_k)$ – БД, $S = \{[R_1], [R_2], \dots, [R_k]\}$ – схема БД.

Определение 1

Пусть $[R_i]$ и $[R_j]$ – схемы отношений (не обязательно различные), $V \subseteq [R_i]$ и $W \subseteq [R_j]$, $|V| = |W|$, тогда соотношение $R_i[V] \subseteq R_j[W]$ называется зависимостью включения.

- В определении $|V|$ – мощность множества V , $R_i[V] = \pi_V(R_i)$ – проекция отношения R_i по атрибутам V . Зависимость включения считается типизированной, если $V = W$, в противном случае – нетипизированной.

Обзор существующих работ

Система аксиом зависимостей включения:

- **IND1)** (рефлексивность): $R_i[X] \subseteq R_i[X]$, если X последовательность отдельных атрибутов R_i .
- **IND2)** (проецирование и перестановка): если $R_i[A_1, \dots, A_m] \subseteq R_j[B_1, \dots, B_m]$, тогда $R_i[A_{i_1}, \dots, A_{i_q}] \subseteq R_j[B_{i_1}, \dots, B_{i_q}]$ для каждой последовательности i_1, \dots, i_q различных целочисленных значений из множества $\{1, \dots, m\}$.
- **IND3)** (транзитивность): если $R_i[X] \subseteq R_j[Y]$ и $R_j[Y] \subseteq R_l[Z]$, тогда выполнено $R_i[X] \subseteq R_l[Z]$.
Надежность: $\Sigma \vdash \sigma \Rightarrow \Sigma \models \sigma$
Полнота: $\Sigma \models \sigma \Rightarrow \Sigma \vdash \sigma$

Обзор существующих работ

- Chandra A. K., Vardi M. Y., The Implication Problem for Functional and Inclusion Dependencies is Undecidable, SIAM Journal on Computing, 14:3 (1985), 671–677.

Обзор существующих работ

- Chandra A. K., Vardi M. Y., The Implication Problem for Functional and Inclusion Dependencies is Undecidable, SIAM Journal on Computing, 14:3 (1985), 671–677.
- Cosmadakis S.S., Kanellakis P.C., Vardi M.Y. Polynomial-time implication problems for unary inclusion dependencies, J. ACM, 37:1 (1990), 15–46.

Обзор существующих работ

- Chandra A. K., Vardi M. Y., The Implication Problem for Functional and Inclusion Dependencies is Undecidable, SIAM Journal on Computing, 14:3 (1985), 671–677.
- Cosmadakis S.S., Kanellakis P.C., Vardi M.Y. Polynomial-time implication problems for unary inclusion dependencies, J. ACM, 37:1 (1990), 15–46.
- Levene M., Vincent M.W. Justification for Inclusion Dependency Normal Form, IEEE Transactions on Knowledge and Data Engineering, 12:2 (2000), 281–291.

Обзор существующих работ

- Chandra A. K., Vardi M. Y., The Implication Problem for Functional and Inclusion Dependencies is Undecidable, SIAM Journal on Computing, 14:3 (1985), 671–677.
- Cosmadakis S.S., Kanellakis P.C., Vardi M.Y. Polynomial-time implication problems for unary inclusion dependencies, J. ACM, 37:1 (1990), 15–46.
- Levene M., Vincent M.W. Justification for Inclusion Dependency Normal Form, IEEE Transactions on Knowledge and Data Engineering, 12:2 (2000), 281–291.
- Gómez-López F. T., Gasca R. M., Pérez-Álvarez J. M. Compliance validation and diagnosis of business data constraints in business processes, Information Systems, 48 (2015), 26–43.

Основы формальной теории

Определение 2

Кортеж $t_i[X]$ **соответствует** кортежу $t_j[X]$ по атрибутам X ($t_j[X] \preceq t_i[X]$), если $t_i[A_l] \neq \text{Null}$, тогда $t_j[A_l] = t_i[A_l]$ или $t_j[A_l] = \text{Null}$; если $t_i[A_l] = \text{Null}$, тогда $t_j[A_l] = \text{Null}$ для любого атрибута $A_l \in X$.

Основы формальной теории

Определение 2

Кортеж $t_i[X]$ соответствует кортежу $t_j[X]$ по атрибутам X ($t_j[X] \preceq t_i[X]$), если $t_i[A_l] \neq \text{Null}$, тогда $t_j[A_l] = t_i[A_l]$ или $t_j[A_l] = \text{Null}$; если $t_i[A_l] = \text{Null}$, тогда $t_j[A_l] = \text{Null}$ для любого атрибута $A_l \in X$.

Определение 3

Зависимость включения $\sigma = R_j[X] \subsetneq R_i[X]$ от главной таблицы $R_i[X]$ к подчиненной таблице $R_j[X]$ по атрибутам X существует, если для любого кортежа $t_j[X] \in R_j[X]$ имеется соответствующий кортеж $t_i[X]$ в отношении $R_i[X]$. Такую зависимость будем называть типизированной с допущением неопределенных значений.

Аксиомы зависимостей включения

- **INN1)** (рефлексивность): если $X \subseteq [R_i]$, тогда $R_i[X] \subsetneq R_i[X]$;
- **INN2)** (проекция): если $R_j[Y] \subsetneq R_i[Y]$ и $X \subseteq Y$, тогда $R_j[X] \subsetneq R_i[X]$;
- **INN3)** (транзитивность): если $R_j[X] \subsetneq R_i[X]$ и $R_i[X] \subsetneq R_l[X]$, тогда выполнено $R_j[X] \subseteq R_l[X]$.

Теорема 1

*Система аксиом **INN1–INN3** непротиворечива.*

Теорема 2

*Система аксиом **INN1–INN3** полна.*

Построение замыканий

Определение 4

Замыканием отношения R_i на множестве зависимостей Σ относительно атрибутов X будем называть множество отношений $R_i^+[X]$, где $R_j \in R_i^+[X]$, если зависимость $\sigma = R_j[X] \subsetneq R_i[X]$ выводима из Σ за счет аксиом **INN1–INN3**, то есть $\Sigma \vdash \sigma$.

Разработан алгоритм **CLOSURE** построения замыкания R^*

Теорема 3

Алгоритм **CLOSURE** корректно формирует множество $R_i^+[X]$.

В теореме показано, что $R_j \in R_i^+[X]$ тогда и только тогда, когда $R_j \in R_i^*[X]$, или $R_i^+[X] = R_i^*[X]$.

Пример схемы БД

- $R_1 =$ Внутренние клиенты (**№ клиента**, № сотрудника),
- $R_2 =$ Внешние клиенты (**№ клиента**, ФИО клиента, Адрес клиента),
- $R_3 =$ Счета клиентов (**№ счета**, № клиента, Остаток средств),
- $R_4 =$ Заказы клиентов (**№ заказа**, № клиента, Содержание заказа).

Атрибут “№ клиента” является первичным ключом в отношениях R_1 и R_2 , в отношениях R_3 и R_4 этот атрибут является внешним ключом.

Все четыре отношения должны быть объединены одним ограничением целостности, в котором R_1 и R_2 являются главными отношениями, а отношения R_3 и R_4 – внешними (ребром ультра графа).

В отношении R_3 атрибут “№ клиента” может принять *Null* значение (остаток не востребоваанных средств). В отношении R_4 атрибут “№ клиента” может принять *Null* значение (предзаказ).

Предварительные сведения

Схему БД в примере невозможно получить, используя классическую теорию проектирования. Решение указанной проблемы предложено в работе:

- Зыкин С.В. Области определения функциональных зависимостей в базах данных // Труды Института математики и механики УрО РАН, Том 23, № 3, 2016. - Стр. 117-129.

Определение 5

Пусть R произвольная реализация отношения на множестве атрибутов U и $X \subseteq U$. Множество кортежей $t \in R$ будем называть областью определения $dom(X)$ множества атрибутов X , если $t[A_j] \neq \text{нов}$ для всех атрибутов $A_j \in X$.

Утверждение 1

1. Если $X = \cup_{i=1}^m X_i$, тогда $dom(X) = \cap_{i=1}^m dom(X_i)$.
2. Пусть $Y \subseteq X$, тогда $dom(X) \subseteq dom(Y)$.

Предварительные сведения

Определение 6

Областью определения $\text{dom}(X \rightarrow Y)$ функциональной зависимости $X \rightarrow Y$ называется множество кортежей $t \in R$, для которых $t[A_j] \neq \text{нов}$ если $A_j \in X \cup Y$, и для любой пары кортежей $t_1, t_2 \in \text{dom}(X \rightarrow Y)$ если $t_1[X] = t_2[X]$, то $t_1[Y] = t_2[Y]$.

Утверждение 2

Если $F = \{F_1, F_2, \dots, F_k\}$ множество функциональных зависимостей, то $\text{dom}(F) = \bigcap_{i=1}^k \text{dom}(F_i)$.

Утверждение 3

Для любой зависимости $X \rightarrow Y$ выполнено $\text{dom}(X \rightarrow Y) = \text{dom}(X) \cap \text{dom}(Y)$.

Предварительные сведения

Обобщение системы аксиом функциональных зависимостей:

- **A1***. (Рефлексивность) Если $Y \subseteq X \subseteq U$, то $X \rightarrow Y$ и $dom(X \rightarrow Y) = dom(X)$.
- **A2***. (Пополнение) Если $X \rightarrow Y$ и $Z \subseteq U$, то $XZ \rightarrow YZ$ и $dom(XZ \rightarrow YZ) = dom(X) \cap dom(Y) \cap dom(Z)$.
- **A3***. (Транзитивность) Если $X \rightarrow Y$ и $Y \rightarrow Z$, то $X \rightarrow Z$ и $dom(X \rightarrow Z) = dom(X) \cap dom(Z)$.

Замечание

Для представленной системы аксиом доказана полнота и непротиворечивость.

Разработаны алгоритмы построения замыканий и построения минимального покрытия

Разработан алгоритм формирования избыточной и непротиворечивой схемы БД,

Обобщенные зависимости включения

Определение 7

Обобщенная типизированная зависимость включения $\sigma = W[X] \subseteq V$ между главными отношениями V и внешними отношениями W по атрибутам X выполнена, если для каждого отношения $R_i \in W$ и каждого кортежа $t_i \in R_i$ существует отношение $R_j \in V$ и существует соответствующий кортеж $t_j \in R_j$, то есть выполнено условие: $t_i[X] \preceq t_j[X]$.

Обобщенные зависимости включения

Определение 7

Обобщенная типизированная зависимость включения $\sigma = W[X] \subsetneq V$ между главными отношениями V и внешними отношениями W по атрибутам X выполнена, если для каждого отношения $R_i \in W$ и каждого кортежа $t_i \in R_i$ существует отношение $R_j \in V$ и существует соответствующий кортеж $t_j \in R_j$, то есть выполнено условие: $t_i[X] \preceq t_j[X]$.

Замечание

Зависимость в примере будет иметь вид: $\{R_3, R_4\}[X] \subsetneq \{R_1, R_2\}$, где X атрибут “№ клиента”.

Аксиомы обобщенных зависимостей включения

- **A1)** (рефлексия) если $W \subseteq V$, тогда $W[X] \subseteq V$;
- **A2)** (проекция) если $W[X] \subseteq V$ и $Y \subseteq X$, тогда $W[Y] \subseteq V$;
- **A3)** (объединение) если $W[X] \subseteq V$ и $S[X] \subseteq V$, тогда $\{W \cup S\}[X] \subseteq V$;
- **A4)** (транзитивность) если $W[X] \subseteq S$ и $S[X] \subseteq V$, тогда $W[X] \subseteq V$.

Здесь V, W, S – множества отношений (таблиц) БД; X, Y – множества атрибутов.

Аксиомы обобщенных зависимостей включения

- **A1)** (рефлексия) если $W \subseteq V$, тогда $W[X] \subseteq V$;
- **A2)** (проекция) если $W[X] \subseteq V$ и $Y \subseteq X$, тогда $W[Y] \subseteq V$;
- **A3)** (объединение) если $W[X] \subseteq V$ и $S[X] \subseteq V$, тогда $\{W \cup S\}[X] \subseteq V$;
- **A4)** (транзитивность) если $W[X] \subseteq S$ и $S[X] \subseteq V$, тогда $W[X] \subseteq V$.

Здесь V, W, S – множества отношений (таблиц) БД; X, Y – множества атрибутов.

- Надежность

Теорема 4

*Система аксиом **A1–A4** непротиворечива.*

Полнота системы аксиом

- С использованием аксиом **A1–A4** выводимо правило декомпозиции:

Утверждение 4

если $W[X] \subsetneq V$ и $S \subseteq W$, тогда $S[X] \subsetneq V$.

Полнота системы аксиом

- С использованием аксиом **A1–A4** выводимо правило декомпозиции:

Утверждение 4

если $W[X] \subsetneq V$ и $S \subseteq W$, тогда $S[X] \subsetneq V$.

- С использованием аксиом **A1–A4** выводимо правило пополнения:

Утверждение 5

если $W[X] \subsetneq V$ и $V \subseteq S$, тогда $W[X] \subsetneq S$.

Полнота системы аксиом

- С использованием аксиом **A1–A4** выводимо правило декомпозиции:

Утверждение 4

если $W[X] \subsetneq V$ и $S \subseteq W$, тогда $S[X] \subsetneq V$.

- С использованием аксиом **A1–A4** выводимо правило пополнения:

Утверждение 5

если $W[X] \subsetneq V$ и $V \subseteq S$, тогда $W[X] \subsetneq S$.

- Полнота

Теорема 5

*Система аксиом **A1–A4** полна.*

Дальнейшие исследования

- Доказать отсутствие взаимодействия обобщенных зависимостей включения и функциональных зависимостей, ограниченных областями определения.

Дальнейшие исследования

- Доказать отсутствие взаимодействия обобщенных зависимостей включения и функциональных зависимостей, ограниченных областями определения.
- Разработать алгоритм построения замыкания обобщенных зависимостей включения и алгоритм построения минимального покрытия зависимостей.

Дальнейшие исследования

- Доказать отсутствие взаимодействия обобщенных зависимостей включения и функциональных зависимостей, ограниченных областями определения.
- Разработать алгоритм построения замыкания обобщенных зависимостей включения и алгоритм построения минимального покрытия зависимостей.
- Разработать единый подход к формированию схемы базы данных на основе функциональных зависимостей, ограниченных областями определения, зависимостей соединения и обобщенных зависимостей включения.

Дальнейшие исследования

- Доказать отсутствие взаимодействия обобщенных зависимостей включения и функциональных зависимостей, ограниченных областями определения.
- Разработать алгоритм построения замыкания обобщенных зависимостей включения и алгоритм построения минимального покрытия зависимостей.
- Разработать единый подход к формированию схемы базы данных на основе функциональных зависимостей, ограниченных областями определения, зависимостей соединения и обобщенных зависимостей включения.
- Разработать теорию и алгоритмы коммутативных межмодельных преобразований для мультимодельных баз данных.

Спасибо за внимание