

Александр Александрович КРОПОТИН¹
Александр Григорьевич ИВАШКО²

УДК 510.5; 510.22; 519-6

**РЕАЛИЗАЦИЯ МЕТОДА ИДЕНТИФИКАЦИИ
СЕМАНТИЧЕСКИХ КОНФЛИКТОВ МЕТАДАНЫХ И
НЕСООТВЕТСТВИЯ ИНТЕГРИРУЕМЫХ ДАННЫХ
ИСХОДЯ ИЗ ИХ СЕМАНТИЧЕСКОГО ОПИСАНИЯ**

¹ старший преподаватель
кафедры программной и системной инженерии,
Институт математики и компьютерных наук,
Тюменский государственный университет
a.a.kropotin@utmn.ru

² доктор технических наук, профессор,
директор Института математики и компьютерных наук,
Тюменский государственный университет
a.g.ivashko@utmn.ru

Аннотация

В работе предлагается комплекс программ, который позволяет строить концептуальные схемы «сущность — связь» реляционных баз данных и проверять их согласованность, выявлять семантические конфликты. Данный комплекс программ может быть применен на этапе проверки согласованности схем при интеграции реляционных баз данных. Применение комплекса программ может существенно повысить эффективность процесса интеграции реляционных баз данных. Комплекс программ основан на применении технологий баз знаний для представления знаний о сущностях и связях предметной области, которые описаны в реляционных базах данных. Научная значимость работы заключается в реализации оригинальных математических моделей и алгоритма онтологического представления концептуальных схем «сущность — связь» в виде комплекса

Цитирование: Кропотин А. А. Реализация метода идентификации семантических конфликтов метаданных и несоответствия интегрируемых данных исходя из их семантического описания / А. А. Кропотин, А. Г. Ивашко // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2017. Том 3. № 2. С. 115-127.

DOI: 10.21684/2411-7978-2017-3-2-115-127

программ, который позволяет идентифицировать семантические конфликты метаданных и несоответствия интегрируемых реляционных баз данных, предлагает функции преобразования концептуальных схем реляционных баз данных, метод выявления их семантических конфликтов и несоответствия данных.

Во введении представляется актуальность проблемы и постановка задач, которые решались в ходе выполнения работы. В основной части работы кратко приводится описание онтологической модели концептуальных объектов, включая ее расширение, в виде модели объединения онтологий концептуальных объектов, которые разработаны для решения поставленной во введении задачи. Также в основной части работы приведен алгоритм онтологического представления концептуальных схем «сущность — связь» и описание архитектуры комплекса программ, который реализует представленные модели и алгоритм. С целью апробации комплекса программ проверки согласованности концептуальных схем «сущность — связь» в работе приведены результаты вычислительных экспериментов на опытных данных, которые представляют различные случаи возникновения семантических конфликтов метаданных, а также доказано, что данный комплекс программ адекватно определяет семантические конфликты концептуальных схем «сущность — связь» реляционных баз данных. В заключении приводятся выводы по изложенному в работе материалу.

Ключевые слова

Базы данных, математическое моделирование, имитационное моделирование, технологии искусственного интеллекта, технологии семантической паутины, выявление семантических конфликтов, выявление несоответствия данных, процесс интеграции данных, онтологическое моделирование данных, методы интеграции реляционных баз данных.

DOI: 10.21684/2411-7978-2017-3-2-115-127

Введение

Развитие информационных технологий привело к появлению огромного количества разнообразных информационных систем. Такие системы в основном предназначены для автоматизации бизнес-процессов, хранения персональных и справочных данных, а также формирования отчетов на основании хранимых данных [1]. Часто оказывается, что эти системы дублируют друг друга. Обеспечение интероперабельности и интеграция информационных систем являются наиболее актуальными задачами [10].

В связи с тем, что объем обрабатываемых данных постоянно растет, специалистам в области информационных технологий необходимо обеспечить эффективный обмен информацией между информационными системами, но при этом избежать дублирования и потерь [1, 5, 10]. Все чаще и чаще разработчики информационных систем сталкиваются с множеством слабо связанных и разрозненных источников информации, и им приходится постоянно производить рутинные действия по выполнению низкоуровневых операций с данными в гетерогенной среде. Особое место в этом списке занимают вопросы интеграции данных, которые могут являться несогласованными и приводить к возникнове-

нию структурных и семантических конфликтов. Традиционные методы решения этой задачи в основном работают на основе связей между элементами систем, построенных вручную, и направлены на решение задач для каждого конкретного случая [1, 10-11]. При этом изменения структуры одной из систем могут привести к нарушению работы программного комплекса и требуют повторной интеграции [1, 5]. Таким образом, традиционные методы не предоставляют достаточно гибкого решения для создания системы интеграции данных [10-11].

На данный момент существует множество различных методов для решения поставленной задачи, но наиболее эффективным из них является применение технологий семантической паутины — технологии связанных данных [1, 5, 7, 10-12, 15-16].

Такие технологии обладают специальными стандартами машинного представления баз знаний и программными продуктами, реализующими алгоритмы логического вывода [5]. Однако комплексного решения в виде программного продукта, который смог бы обеспечить процесс выявления несоответствия интегрируемых данных с применением технологий семантической паутины и разработанных для этого методов, на сегодняшний день не существует.

При всем этом многие из данных методов решения проблемы интеграции данных направлены на то, чтобы игнорировать возможные конфликты этих данных и обеспечить совместный доступ к ним. Хотя в процессе интеграции могут возникать и некоторые конфликты, которые вызваны неоднородностью данных источников [1, 5, 15-16].

Таким образом, целью данной работы является разработка и апробация комплекса программ, который бы реализовывал математические методы и алгоритмы онтологического представления реляционных баз данных и выявления семантических конфликтов метаданных и несоответствия данных.

Основная часть

В качестве решения поставленной задачи предлагается программное обеспечение под полным наименованием The software of checking consistency of conceptual schemes и обозначением C_3S . Программное обеспечение C_3S реализует онтологическую модель концептуальных объектов [8, 15-16] и алгоритм [15] построения терминологических аксиом и утверждений дискреционной логики исходя из семантического представления схем реляционных баз данных.

Онтологическая модель концептуальных объектов предоставляет правила описания концептуальных схем «сущность — связь» реляционных баз данных [6, 13-14] в виде аксиом и утверждений дескрипционной логики $\mathcal{SHOIQ}(\mathcal{D})$ [3-4].

Каждый регулярный тип сущности представляется через набор терминологических аксиом, который определяет атомарный концепт A_n (1):

$$\{A_n \sqsubseteq T \sqcap \neg(A_{n+1} \sqcup \dots \sqcup A_m) \sqcap \forall E. \perp\}. \quad (1)$$

Слабый тип сущности и тип значения представляются подобным образом (1), с отличием в определении локального ограничения концепта A_n (1) на диапазон роли E (1), которое представляет зависимость существования. Тогда

концепты будут принадлежать одной области интерпретации и их семантика будет сопоставима. Каждая сущность или значение атрибута представляется через утверждение, которое определяет индивид атомарного концепта (1), причем первые дополнительно объявляются как взаимно отличные.

Каждый тип связи с отображением 1:n представляется через набор терминологических аксиом, который определяет абстрактные роли R_i, R_j^i (2):

$$\begin{aligned} & \{R_i \sqsubseteq U, R_i \sqsubseteq \neg R_i^-, \\ & R_j^i \sqsubseteq U, R_j^i \sqsubseteq \neg R_j^{i-}, R_i \equiv R_j^{i-} \\ & \top \sqsubseteq \leq 1R_i \sqcap \forall R_i. A_m \sqcap \forall R_j^i. A_n, \\ & \exists R_i. \top \sqsubseteq A_n, \exists R_j^i. \top \sqsubseteq A_m, \\ & A_n \sqsubseteq \exists R_i. \top, A_m \sqsubseteq \exists R_j^i. \top \}. \end{aligned} \quad (2)$$

Тип связи с отображением 1:1 представляется подобным образом (2), с отличием в определении одной симметричной абстрактной роли $R_i \sqsubseteq R_i^-$ (1), которая в совокупности с остальными аксиомами представляет отображение 1:1.

Каждый не уникальный атрибут представляется через набор терминологических аксиом, который определяет абстрактные роли F_i, F_j^i и конкретную роль T_h^i (3):

$$\begin{aligned} & \{F_i \sqsubseteq U, F_i \sqsubseteq \neg F_i^-, T_h^i \sqsubseteq U, \\ & F_j^i \sqsubseteq U, F_j^i \sqsubseteq \neg F_j^{i-}, F_i \equiv F_j^{i-}, \\ & \top \sqsubseteq \leq 1F_i \sqcap \leq 1T_h^i \sqcap \forall F_i. A_m \sqcap \forall F_j^i. A_n \sqcap \forall T_h^i. D^h, \\ & \exists F_i. \top \sqsubseteq A_n, \exists F_j^i. \top \sqsubseteq A_m, \exists T_h^i. \top \sqsubseteq A_m \\ & A_n \sqsubseteq \exists F_i. \top, A_m \sqsubseteq \exists F_j^i. \top, A_m \sqsubseteq \exists T_h^i. \top \}. \end{aligned} \quad (3)$$

Уникальный атрибут представляется подобным образом (3), с отличием в определении одной симметричной абстрактной роли $F_i \sqsubseteq F_i^-$ (1), которая в совокупности с остальными аксиомами представляет ограничение на уникальное отображение атрибута.

Ключ сущности представляет собой набор терминологических аксиом, который определяет локальное ограничение на ключевые роли (2)-(3) атомарного концепта (1): $HasKey(A_n F_i \sqcap, \dots \sqcap F_h^i)$. Транзитивные связи представляются через набор терминологических аксиом, который определяет транзитивную абстрактную роль S как суперроль для всех остальных абстрактных ролей (2)-(3): $S \sqsubseteq U, S \circ S \sqsubseteq S, R_i \sqsubseteq S, \dots F_j \sqsubseteq S$.

Алгоритм построения терминологических аксиом и утверждений дискреционной логики [2-4] исходя из семантического представления схем реляционных баз данных CS_p, \dots, CS_h

Такой алгоритм в общем случае представлен в листинге 1.

Алгоритм (листинг 1) позволяет преобразовывать схемы «сущность — связь» реляционных баз данных в терминах предложенной онтологической модели представления концептуальных объектов. Наихудшее время работы такого алгоритма, если не учитывать реализацию переборных и присвоить каждому его

шагу стоимость и количество добавляемых для выполнения одной итерации терминологических аксиом или утверждений, составляет: $T_1 = (a + b \times j) \times n + c \times i + (d + (e \times 2 \times j) + f \times i) \times k + 1$, где a, b, c, d, e, f — константы, которые определяются выбранными значениями стоимости шагов алгоритма; n, i, j, k — количество типов сущности, атрибутов, типов связи и сущностей соответственно.

Метод проверки согласованности концептуальных схем позволяет свести процесс выявления семантических конфликтов метаданных и несоответствия данных, возникающих при их интеграции, к решению алгоритмических проблем согласованности и классификации терминологии и согласованности онтологии.

Программное обеспечение C_3S может использоваться как программный комплекс, развернутый на одной рабочей станции пользователя, или как комплекс программ, развернутых на различных рабочих станциях, поддерживает различные архитектуры и системы, ограничения которых зависят от ограничений входящих в состав программ и их размещения. Программное обеспечение C_3S реализовано на языке Java, основным преимуществом которого можно назвать независимость функционирования разрабатываемого комплекса от различных платформ. В качестве инструмента разработки комплекса программ и его реализации использо-

Листинг 1

Псевдокод алгоритма построения терминологических аксиом и утверждений дискреционной логики исходя из семантического представления схемы реляционной базы данных

Listing 1

Pseudocode algorithm for constructing terminological axioms and discretionary logic statements based on the semantic representation of the relational database schema

ДЛЯ КАЖДОГО типа сущности $\forall B_n^l \in CS_l \triangleright$
 ○○○**ВЫПОЛНИТЬ ОПРЕДЕЛИТЬ** атомарный концепт A_n , через набор терминологических аксиом вида (1), соответствующий *регулярному* или *слабому типу сущности*
 ○○○**ДЛЯ КАЖДОГО** атрибута типа сущности $\forall H_i^l: B_n^l \rightarrow V_m^l \in CS_l \triangleright$
 ○○○○○○**ВЫПОЛНИТЬ ОПРЕДЕЛИТЬ** атомарный концепт A_m , через набор терминологических аксиом вида (1), соответствующий *типу значения*
 ○○○○○○**ВЫПОЛНИТЬ ОПРЕДЕЛИТЬ** атомарные роли F_i, \dots, T_h^i , через набор терминологических аксиом вида (3), соответствующий *уникальному* или *не уникальному атрибуту* \triangleleft
 ○○○**ВЫПОЛНИТЬ ОПРЕДЕЛИТЬ** ключевые роли F_i, \dots, T_h^i концептов A_n, A_m , через терминологические аксиомы вида $HasKey(A_n F_i \sqcap, \dots \sqcap F_j^i), HasKey(A_m T_h^i)$ соответствующие ключу сущности $H_k^l = \{H_i^l, \dots, H_j^l\} \triangleleft$
ДЛЯ КАЖДОГО типа связи $\forall G_i^l: B_n^l \times B_m^l \in CS_l \triangleright$
 ○○○**ВЫПОЛНИТЬ ОПРЕДЕЛИТЬ** атомарные роли R_i, \dots, R_j^i , через набор терминологических аксиом вида (2), соответствующий *отображению 1:1* или *1:n* \triangleleft

вались среды IntelliJ IDEA и Eclipse Neon [9, 17]. Программное обеспечение C_3S представляет собой комплекс из семи программ:

1. Интегрированная среда разработки Eclipse — применяется в качестве среды исполнения всех остальных программ комплекса.

2. Среда моделирования Rarugus — реализует функцию отображения информации о сущностях и связях реляционной базы данных в виде диаграммы классов по стандарту UML.

3. Интерфейс прикладного программирования OWL API — применяется для взаимодействия с онтологиями на языке OWL.

4. Программа онтологического представления концептуальных схем «сущность — связь» C_2O — написана специально для реализации функции конвертации диаграммы классов из среды моделирования Rarugus в онтологию на языке OWL. Представляет из себя консольное приложение и может исполняться отдельно от комплекса.

5. Плагин Eclipse RCP C_2OP — написан специально для интеграции программы C_2O в программный комплекс. Предоставляет графический интерфейс для взаимодействия с программой C_2O из среды Eclipse, который позволяет вызывать функции программы C_2O .

6. Редактор онтологий Protégé — применяется как средство для открытия и объединения онтологий на языке OWL.

7. Программа рассуждений над онтологиями Hermit — реализует алгоритмы и структуры данных, необходимые для интегрирования баз знаний, выполняет операции по проверке их согласованности, выполнимости концептов и классификации понятий.

Архитектура C_3S отображена на рис. 1.

Проверка согласованности концептуальных схем комплексом программ C_3S может быть представлена четырьмя ключевыми этапами (рис. 1):

1. Получение концептуальных схем «сущность — связь» интегрируемых реляционных баз данных или, в качестве альтернативы, описание таких схем при помощи среды моделирования Rarugus.

2. Конвертация концептуальных схем «сущность — связь» в онтологию посредством выбора таких схем и выполнения программного компонента C_2O в исполняемой среде C_2OP .

3. Интеграция нескольких онтологий в одну по правилам расширенной онтологической модели представления концептуальных объектов при помощи редактора онтологий Protégé.

4. Анализ онтологии в соответствии с предложенным методом применения решения алгоритмических проблем согласованности и классификации терминологии и согласованности онтологии для выявления семантических конфликтов метаданных и несоответствия данных, посредством вызова соответствующих функций Hermit в Protégé и анализа результатов.

Эффективность программного комплекса была проверена специально организованными и проведенными вычислительными экспериментами, которые

проводились на основе 204 концептуальных схем баз данных, представляющих 102 возможных случая возникновения конфликтов атрибут — атрибут, атрибут — сущность и сущность — связь. Такая оценка включала три этапа:

1. Подготовка опытных данных. Такие данные представлены наборами концептуальных схем «сущность — связь» реляционных баз данных, каждый из которых представляет различные случаи возникновения соответствующего конфликта двух таких схем. На основании предложенной онтологической модели представления концептуальных объектов и алгоритма построения терминологических аксиом и утверждений исходя из семантического представления реляционных баз данных каждого из такого набора построены онтологии по стандарту OWL 2.0.

2. Проведение вычислительных экспериментов. Такие эксперименты направлены на выполнение задач табличного алгоритма на опытных данных и выявление моделируемых ими конфликтов, в соответствии с предложенным методом применения дескрипционного исчисления для выявления семантических конфликтов и несоответствия данных при их интеграции: выполнение задач проверки согласованности и классификации терминологии для выявления терминологических аксиом конфликтов атрибут — атрибут, атрибут — сущность, сущность — связь, поведения, наследования и ключа; выполнение задачи со-

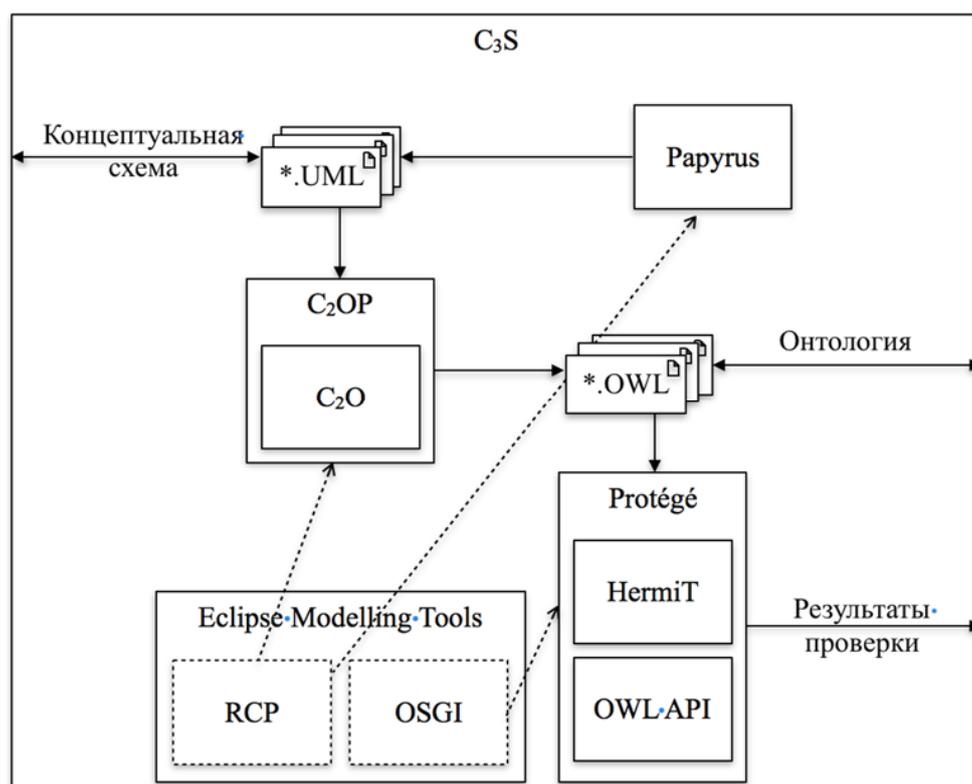


Рис. 1. Архитектура программного обеспечения C₃S

Fig. 1. The C₃S Software Architecture

гласованности онтологии целиком для выявления утверждений конфликтов формата значений, представления значений и актуальности значений; выполнение задач согласованности терминологии и согласованности онтологии для выявления терминологических аксиом и утверждений конфликтов зависимости, типов данных, допустимых значений и разрешенных значений.

3. Оценка результатов вычислительных экспериментов. Такая оценка основана на сопоставлении количества реально существующих в наборах опытных данных, семантических конфликтов концептуальных схем «сущность — связь» и выявленных в ходе вычислительных экспериментов соответствующих несоответствий терминологических аксиом и/или утверждений.

Результаты вычислительного эксперимента показали, что в 120 возможных случаях возникновения конфликтов верно выявлена семантическая несогласованность в 81 из 99 случаев. Результаты такого вычислительного эксперимента представлены на рис. 2.

Такая оценка (рис. 2) показала, что:

1. Разработанные модели онтологического представления концептуальных объектов и алгоритм построения терминологических аксиом и утверждений исходя из семантического представления реляционных баз данных ориентированы на онтологическое моделирование только семантической составляющей таких баз, исключая структурное их представление.

2. Предложенный метод применения дескрипционного исчисления для выявления семантических конфликтов схем баз данных и несоответствия данных позволяет выявлять только семантическую несогласованность таких схем, иг-

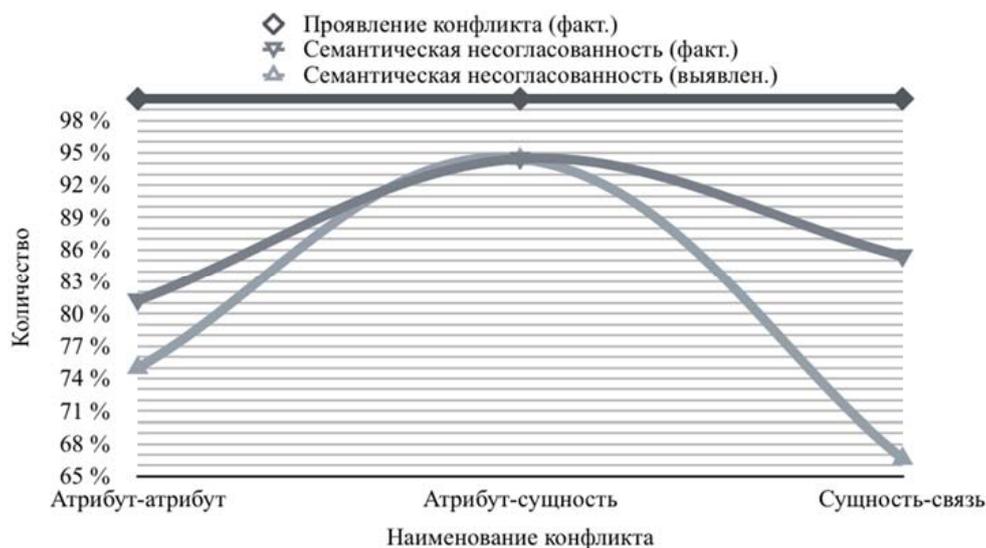


Рис. 2. Результаты вычислительного эксперимента на моделях конфликтов атрибут — атрибут, атрибут — сущность и сущность — связь

Fig. 2. The results of the computational experiment on conflict models: attribute — attribute, attribute — entity and entity — relationship

норируя их структурные различия. Такая применимость метода обусловлена поставленными задачами и результатами исследования проблем интеграции, которые показали, что ввиду семантической согласованности структурная гетерогенность разрешима стандартными инструментами интеграции данных.

3. Предложенный метод не может быть применен при жесткой типизации данных и не распространяется на конфликты, образованные различиями только ограничений кардинальности эквивалентных множеств связей. Такое ограничение метода обусловлено различным предположением об открытости и замкнутости мира, которое принято в онтологиях и базах данных соответственно.

Заключение

В работе представлен специализированный программный комплекс, который позволяет строить концептуальные схемы «сущность — связь» реляционных баз данных и проверять их согласованность, выявлять семантические конфликты. Была проведена апробация программного комплекса на вычислительном эксперименте и доказано, что он адекватно определяет семантические конфликты концептуальных схем «сущность — связь» реляционных баз данных, хотя и имеет некоторые ограничения.

Такие ограничения разработанного программного комплекса во многом обуславливаются поставленными задачами и результатами исследования проблем интеграции.

СПИСОК ЛИТЕРАТУРЫ

1. Бубарева О. А. Модель, алгоритмы и программное обеспечение интеграции данных информационных систем на основе онтологий (на примере вуза): дис. канд. тех. наук / О. А. Бубарева. Бийск, 2014. 137 с.
2. Григорьев А. В. Математические методы и алгоритмы определения согласованности баз знаний: дис. канд. тех. наук / А. В. Григорьев. Тюмень, 2013. 115 с.
3. Золин Е. Е. Глава 3. Факты и ABox / Е. Е. Золин // *Дескрипционная логика* (лекции). URL: http://pcs.math.msu.su/~zolin/dl/pdf/DL_03_ABox.pdf (дата обращения: 11.05.2016).
4. Золин Е. Е. Глава 6. Расширения логики / Е. Е. Золин // *Дескрипционная логика* (лекции). URL: http://pcs.math.msu.su/~zolin/dl/pdf/DL_06_ALCOIQ.pdf (дата обращения: 11.05.2016).
5. Когаловский М. Р. Методы интеграции данных в информационных системах / М. Р. Когаловский // *Сборник трудов Третьей Всероссийской конференции «Стандарты в проектах современных информационных систем»*. 2003. Москва, Россия. 23-24 апреля. С. 1-8.
6. Коннолли Т. Базы данных: Проектирование, реализация и сопровождение. Теория и практика / Т. Коннолли, К. Бегг; пер. с англ. и ред. Р. Имамутдиновой, К. Птицына. 3-е изд. М.: Вильямс, 2003. 1436 с.
7. Кропотин А. А. Применение табличного алгоритма для верификации бизнес-процессов / А. А. Кропотин, А. Г. Ивашко, А. В. Григорьев, Е. О. Овсянникова //

- Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2013. № 7. С. 202-213.
8. Кропотин А. А. Применение формализма дескрипционной логики для выявления семантических конфликтов концептуальных схем сущность-связь / А. А. Кропотин // Интеллект. Инновации. Инвестиции. Оренбургский государственный университет. 2016. Вып. 7. С. 93-98.
 9. Продукты и разработки: IntelliJ IDEA // JetBrains.
URL: <https://jetbrains.ru/products/idea> (дата обращения: 14.07.2016).
 10. Семерханов И. А. Методы и алгоритмы автоматизированной интеграции информационных ресурсов на основе онтологического подхода: дис. канд. тех. наук / И. А. Семерханов. СПб., 2014. 140 с.
 11. Семерханов И. А. Интеграция информационных систем при помощи связанных данных / И. А. Семерханов, Д. И. Муромцев // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 5 (87). С. 123-128.
 12. Семерханов И. А. Интеграция реляционных баз данных с использованием rdf/owl / И. А. Семерханов, Г. В. Варгин // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 1. СПб.: НИУ ИТМО. С. 117-118.
 13. Chen P. P-S. The Entity Relationship Model — Toward a Unified View of Data / P. P-S. Chen // Software Pioneers. Contributions to Software Engineering. Springer Berlin Heidelberg, 2002. Pp. 311-339. DOI 10.1007/978-3-642-59412-0_18
 14. Codd E. F. Extending the Database Relational Model to Capture More Meaning / E. F. Codd // ACM Transactions on Database Systems. 1970. Vol. 4. No 4. Pp. 397-434. DOI 10.1145/320107.320109
 15. Kropotin A. A. Database Schema Method for Automatic Semantic Errors Resolving During Information Systems / A. A. Kropotin, A. V. Grigoryev, A. G. Ivashko // Integration, Informal Proceedings of the 2016 10th IEEE International Conference on Application of Information and Communication Technologies, AICT2016, 2016. Azerbaijan, Baku. 12-14 October.
 16. Kropotin A. A. Realization of the Ontologically Based Method for Checking Structural Inconsistencies of Relational Databases / A. A. Kropotin, A. V. Grigoryev, Y. V. Bidulya, A. G. Ivashko, N. S. Durynin // Proceedings of the 27th International DAAAM Symposium on Intelligent Manufacturing and Automation 2016. 2016. Vol. 27. Mostar, Bosnia and Herzegovina. 26-29 October. Pp. 762-767.
DOI 10.2507/27th.daaam.proceedings.110
 17. Neon // Eclipse. URL: <http://www.eclipse.org/neon> (дата обращения: 14.07.2016).

Alexander A. KROPOTIN¹

Alexander G. IVASHKO²

**IMPLEMENTATION OF A METHOD
FOR IDENTIFYING SEMANTIC CONFLICTS
OF METADATA AND INCONSISTENCY OF MERGING DATA
BASED ON THEIR SEMANTIC DESCRIPTION**

¹ Senior Lecturer,
Department of Software and Systems Engineering,
Institute of Mathematics and Computer Science,
Tyumen State University
a.a.kropotin@utmn.ru

² Dr. Sci. (Tech.), Professor, Director,
Institute of Mathematics and Computer Science,
Tyumen State University
a.g.ivashko@utmn.ru

Abstract

This paper proposes a set of programs that allows to construct conceptual schemes of the “entity — relationship” of relational databases and verify their consistency, to identify semantic conflicts. This set of programs can be applied at the stage of checking the consistency of schemes when integrating relational databases. Application of the program complex can significantly improve the efficiency of the process of integration of relational databases. The complex of programs is based on the application of knowledge base technologies to represent knowledge about the entities and domain relationships that are described in relational databases. The scientific significance of the work lies in the implementation of the original mathematical models and the ontological representation of conceptual schemes of “entity — relationship” in the form of a complex of programs that allows identifying semantic conflicts of metadata and inconsistencies of integrable relational databases and offers functions for transforming conceptual schemes of relational databases and a method for identifying their semantic conflicts and data inconsistencies.

Citation: Kropotin A. A., Ivashko A. G. 2017. “Implementation of a Method for Identifying Semantic Conflicts of Metadata and Inconsistency of Merging Data Based on Their Semantic Description”. Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy, vol. 3, no 2, pp. 115-127.

DOI: 10.21684/2411-7978-2017-3-2-115-127

In the introduction, the urgency of the problem and the formulation of tasks that were decided in the course of the work are presented. The main part of the work briefly describes the ontological model of conceptual objects, including its extension, in the form of a combining of ontologies of conceptual objects that are designed to solve the problem posed in the introduction. Also, in the main part of the work the algorithm of ontological representation of conceptual schemes is presented the essence of the connection and description of the architecture of the program complex, which is represented by models and algorithm. With the purpose of approbation of the complex of programs for checking the consistency of conceptual schemas of “entity — relationship”, the results of computational experiments on experimental data that represent various cases of the emergence of semantic conflicts of metadata are presented and it is also proved that it adequately determines the semantic conflicts of conceptual schemes of the “entity — relationship” of relational databases.

Keywords

Databases, mathematical modeling, simulation modeling, artificial intelligence technologies, semantic web technologies, identifying semantic conflicts, identifying data inconsistencies, data integration process, ontological data modeling, methods of integration of relational databases.

DOI: 10.21684/2411-7978-2017-3-2-115-127

REFERENCES

1. Bubareva O. A. 2014. “Model', algoritmy i programmnoe obespechenie integratsii dannykh informatsionnykh sistem na osnove ontologii (na primere vuza)” [Model Algorithms and Software for Basis Integration of Information Systems on the Basis of Ontologies (on the Example of the University)]. Cand. Sci. (Tech.) diss. Biysk.
2. Grigoriev A. V. 2013. “Matematicheskie metody i algoritmy opredeleniya soglasovannosti baz znaniy” [Mathematical Methods and Algorithms for Determining the Consistency of Knowledge Bases]. Cand. Sci. (Tech.) diss. Tyumen.
3. Zolin E. E. 2016. “Fakty i ABox” [Facts and ABox]. Ch. 3 of Deskriptcionnaya logika (lektsii) [Descriptive Logic (Lectures)]. http://pcs.math.msu.su/~zolin/dl/pdf/DL_03_ABox.pdf (accessed on 11 May 2016).
4. Zolin E. E. 2016. “Rasshireniya logiki” [Extensions of Logic]. Ch. 6 of Deskriptcionnaya logika (lektsii) [Descriptive Logic (Lectures)]. http://pcs.math.msu.su/~zolin/dl/pdf/DL_06_ALCOIQ.pdf (accessed on 11 May 2016).
5. Kogalovsky M. R. 2003. “Metody integratsii dannykh v informatsionnykh sistemakh” [Methods of Data Integration in Information Systems]. Proceedings of the Third All-Russian Conference “Standarty v proektakh sovremennykh informatsionnykh sistem” [Standards in Modern Information Systems Projects] (Moscow, Russia. 23-24 April), pp. 1-8.
6. Connolly T., Begg C. 2003. Bazy dannykh: Proektirovanie, realizatsiya i soprovozhdenie. Teoriya i praktika [Databases: Design, Implementation and Maintenance. Theory and

- Practice]. Translated from English and edited by R. Imamutdinova, K. Ptitsyn. 3rd edition. Moscow: Vil'yams.
7. Kropotin A. A., Ivashko A. G., Grigoriev A. V., Ovsyannikova E. O. 2013. "Primenenie tablitsnogo algoritma dlya verifikatsii biznes-protsessov" [Application of the Table Algorithm for Verification of Business Processes]. Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy, no 7, pp. 202-213.
 8. Kropotin A. A. 2016. "Primenenie formalizma deskriptivnoy logiki dlya vyyavleniya semanticheskikh konfliktov kontseptual'nykh skhem sushchnost'-svyaz'" [Use of the Formalism of Descriptive Logic to Identify Semantic Conflicts of Conceptual Schemes of Entity-Relationship]. Intellekt. Innovatsii. Investitsii. Orenburgskiy gosudarstvennyy universitet [Intelligence. Innovation. Investments. Orenburg State University], no 7, pp. 93-98.
 9. JetBrains. 2016. "Produkty i razrabotki: IntelliJ IDEA" [Products and Development]. <https://jetbrains.ru/products/idea> (accessed on 14 July 2016).
 10. Semarkhanov I. A. 2014. "Metody i algoritmy avtomatizirovannoï integratsii informatsionnykh resursov na osnove ontologicheskogo podkhoda" [Methods and Algorithms of Automated Integration of Information Resources on the Basis of the Ontological Approach]. Cand. Sci. (Tech.) diss. St. Petersburg.
 11. Semerkhanov I. A., Muromtsev D. I. 2013. "Integratsiya informatsionnykh sistem pri pomoshchi svyazannykh dannykh" [Integration of Information Systems with Related Data]. Scientific and Technical Herald of Information Technologies, Mechanics and Optics, no 5 (87), pp. 123-128.
 12. Semerkhanov I. A., Vargin G. V. 2013. "Integratsiya relyatsionnykh baz dannykh s ispol'zovaniem rdf\owl" [Integration of Relational Databases Using rdf\owl]. Scientific and Technical Herald of Information Technologies, Mechanics and Optics, no 1, pp. 117-118. St. Petersburg: NIUITMO.
 13. Chen P. P-S. 2002. "The Entity Relationship Model — Toward a Unified View of Data". Software Pioneers. Contributions to Software Engineering. Springer Berlin Heidelberg, 2002. Pp. 311-339. DOI 10.1007/978-3-642-59412-0_18
 14. Codd E. F. 1970. "Extending the Database Relational Model to Capture More Meaning". ACM Transactions on Database Systems, vol. 4, no 4, pp. 397-434. DOI 10.1145/320107.320109
 15. Kropotin A. A., Ivashko A. G., Grigoryev A. V. 2016. "Database Schema Method for Automatic Semantic Errors Resolving During Information Systems". Integration, Informal Proceedings of the 2016 10th IEEE International Conference on Application of Information and Communication Technologies, AICT2016. Azerbaijan, Baku. 12-14 October.
 16. Kropotin A. A., Grigoryev A. V., Bidulya Y. V., Ivashko A. G., Durynin N. S. 2016. "Realization of the Ontologically Based Method for Checking Structural Inconsistencies of Relational Databases". Proceedings of the 27th International DAAAM Symposium on Intelligent Manufacturing and Automation 2016, vol. 27 (Mostar, Bosnia and Herzegovina. 26-29 October), pp. 762-767. DOI: 10.2507/27th.daaam.proceedings.110
 17. Eclipse. 2016. Neon. <http://www.eclipse.org/neon> (accessed on 14 July 2016).