

# МЕТОДЫ МОДЕЛИРОВАНИЯ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ ПРЕДМЕТНОЙ ОБЛАСТИ ПРИ ПОСТРОЕНИИ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В КОРПОРАТИВНЫХ СИСТЕМАХ

УДК 519.863.5

---

## КОЧУЕВА Зоя Анатольевна

к.т.н., доцент кафедры интеллектуальных компьютерных систем Национального технического университета «Харьковский политехнический институт».

**Научные интересы:** автоматизированная обработка естественного языка, информационно-поисковые системы, автоматизированные библиотечные системы.

## БОРИСОВА Наталья Владимировна

к.т.н., доцент кафедры интеллектуальных компьютерных систем Национального технического университета «Харьковский политехнический институт».

**Научные интересы:** автоматизированная обработка естественного языка, извлечение предметных знаний.

### ПОСТАНОВКА ПРОБЛЕМЫ

Ставится задача разработки распределенной системы поддержки принятия решений (СППР), обеспечивающей интеграцию данных и механизмы устранения семантических аномалий [1], возникающих при обработке данных, хранимых в базе данных (БД), согласно нормализованной логической схеме [3-5] в больших предметных областях (ПО), свойственных СППР для корпорации. Подобные аномалии зачастую обусловлены различием лексического и семантического значений имен атрибутов БД у пользователей различных специальностей (например, массивы данных, характеризующие качество воды, экологи называют загрязнением по химическим элементам, а химики концентрацией соответствующих вредных веществ. При этом пользователям БД необходимо отличать данные о концентрации вредных веществ на входе и выходе различных очистительных сооружений). Использование первичных ключей и функциональных зависимостей не достаточно для поддержания механизма определения семантики множеств значений атрибутов вне зависимости от их лексических значений (написания). Решение задачи предполагает разработку методов моделирования

информационных объектов предметной области [1, 2, 4], методов и алгоритмов определения сравнимости и эквивалентности множеств значений атрибутов на основе модели ПО, что позволит не только осуществлять интеграцию данных в БД, но и избежать семантических аномалий в операциях реляционной алгебры [3-5].

### АНАЛИЗ ПОСЛЕДНИХ ИССЛЕДОВАНИЙ И ПУБЛИКАЦИЙ

Ввиду отсутствия в БД достаточных средств (кроме ключей и функциональных зависимостей) интерпретации смысла атрибутов схемы базы данных, попытки создания СППР в больших ПО привели к необходимости разработки методов интеграции данных на основе анализа ПО и способов устранения семантических аномалий, возникающих при работе с базой данных посредством реляционной алгебры. К первым работам в этой области относят расширенную реляционную модель Кодда [1] и ER-модель Чена [6], которые были признаны несовершенными авторами моделей, а также работы Цаленко М.Ш. [1] и Трахтенгерца Э.А. [7].

### ФОРМУЛІРОВАКА ЦЕЛЕЙ СТАТТІ

Стаття присвячена вопросам моделювання ПО засадами теорії типів з метою розробки на цій основі механізмів інтерпретації семантичного значення даних, зберіганих в БД при умові багатозначності лексических і семантиеских значення імен атрибутів. Пути рішення цієї проблеми состолять в необхідності моделювання ПО з використанням діаграми Чена [6] на першом етапі, оптимізації глобальної схеми бази даних на другому етапі [2-5], з послующою розробкою механізмів інтерпретації семантиеского значення даних, зберіганих в БД при умові багатозначності лексических і семантиеских значення імен атрибутів на третьому етапі, а також в побудові ефективних алгоритмів визначення порівняності даних на четвертому етапі.

### ІЗЛОЖЕННЯ ОСНОВНОГО МАТЕРІАЛУ ІСЛЕДОВАНИЯ

Целю створення систем підтримки прийняття стратегіеских рішень управління являється автоматизація процесів підготовки прийняття наукообоснованих стратегіеских рішень з використанням сучасних інформаційних технологій для ефективного використання інформаційного ресурса в великих ПО.

Розробка приложень БД основана на результатах аналізу ПО і концептуальних моделях ПО і БД. Процес проектування БД включає наступні етапи:

1) змістательний аналіз ПО і вимог польователів. Перший етап являється неформальним і завершається виявленням інформації, котрою повинна містити БД і визначенням відношеня між компонентами цієї інформації;

2) концептуальне проектування. Другим і наступними етапами являються формальними. Етап завершається побудовою формалізованої концептуальної моделі ПО, визначеної в одній з нотаций. Традиційно для цього використовується модель «сутність – зв'язь» (ER-модель), в котрій визначаються формальні об'єкти і формальні правила підтримання цілостності даних;

3) логіеское проектування. На цьому етапі концептуальна модель в термінах ER-моделі відображається в реляційну модель;

4) проектування реалізації. На цьому етапі здійснюється реалізація отриманої на третьому етапі моделі в вигляді схеми БД з використанням мови визначення даних. В результаті, БД починає своє фізическое існування;

5) побудова формальної семантиескої моделі БД для аналізу семантики атрибутів.

Для побудови формалізованої концептуальної моделі ПО на другому етапі використовуються ER-модель [6] і відповідующа їй структурна компонента ER-моделі. В моделі використовується принцип семантиеского моделювання, передполагаючий, що світ состоить з об'єктів двох видів – сутностей і зв'язей. Сутність (entity) – це деякий неделимый (атомарний) об'єкт ПО, котрий може бути чітко ідентифікований. Зв'язь (relationship) – це агрегат сутностей. Типи сутності асоціюються засадою деяких іменованих типів зв'язи (агрегація на рівні типів). Сутності і зв'язи мають деякими властивостями. Властивість (property) – це елемент інформації, описуючий сутність або зв'язь. Сутності або зв'язи одного типу мають общими властивостями. Значення властивостей кожного типу належить деякому множетву значення (домену). Таким образом, сутність і зв'язь можна розглядати як агрегат значення властивостей, а тип сутності або зв'язи – як агрегат властивостей. Ограничення цілостності в ER-моделі задаються ключевими властивостями, типами відображеня, залежностями по існуванню, а також визначеннями допустимих множетв значення засадою формул логіески першого порядку. Одним з методів представлення семантиескої моделі БД в графіеском вигляді являється ER-діаграма (ERD). Вона наглядно відображає основні компоненти проекту БД. Процес відображеня семантиескої моделі в логіеску схему на третьому етапі багатоваріантний, в вигляді чого його не можна повністю автоматизувати. В простейшем випадку сильний тип сутності відображається в базове відношеня; кожне з них буде мати потенціальний ключ, відповідуючий ключу сутності. Властивості сутності відображаються в атрибути відношеня. Використання реляційної моделі на четвертому етапі являється перспективним. Етап побудови формальної семантиескої моделі БД для аналізу семантики атрибутів на п'ятому етапі зв'язан з роз-

работкой моделей семантической эквивалентности данных в реляционных базах данных, критериев и методов семантической сравнимости атрибутов БД. Решение этой проблемы позволяет не только эффективно осуществлять интеграцию данных и поддерживать ограничения целостности, но и служить инструментом нормализации лексики описания ПО.

В статье предлагается формализовать объекты предметной области ER-диаграммы. Будем рассматривать объекты ПО как структурированные типы данных, которые могут взаимодействовать с другими структурами фрагмента посредством своих элементов. Описать такой объект можно при помощи признаков или смыслов  $P = \langle RTO_1, RTO_2, \dots, RTO_n \rangle$ . Признак можно рассматривать с точки зрения его роли в классификации типов объектов таким образом, что признаком  $P$  можно назвать кортеж предикатов  $P = \langle RTO_1, RTO_2, \dots, RTO_n \rangle$ , а значением признака, задающим тип объекта  $X$ , определенную последовательность из нулей и единиц вида  $\langle |RTO_1(X), \dots, RTO_n(X)| \rangle$ , где  $|RTO(X)| = 1$ , если  $RTO(X)$  истина и нуль, если  $RTO(X)$  – ложь. Таким образом, каждому значению кортежа соответствует тип объекта с именем на множестве типов объектов  $O = \{O_i\}$ , соответствующего множеству значений кортежа  $P$ . Итак, для объектов данного типа применим интенциональный способ задания множества, когда само множество задается некоторым свойством, выделяющим часть элементов универсума. При таком подходе необходимо убедиться, что всякий элемент универсума, обладающий первым свойством, обладает вторым и наоборот. Все свойства  $RTO_1(X), \dots, RTO_n(X)$  являются интенционально равными, то есть каждое из них влечет выполнение другого. Интенциональные связи между свойствами можно рассматривать как связи между понятиями в некоторой системе знаний.

При задании множества типов используется экстенциональный способ задания множества, состоящий в том, что некоторым образом указываются все элементы универсума, которые принадлежат данному множеству. Экстенциональные связи – это связи между объектами универсума. Тот факт, что интенциональные

связи между понятиями согласуются с экстенциональными связями между обозначенными ими объектами является важным свойством мира, в котором мы живем. Сталкиваясь с объектами различной природы, мы классифицируем их, применяем к ним одни и те же способы обработки. Моделирование некоторой ПО будет осуществляться проще, если инструмент моделирования будет позволять разбивать моделируемые объекты на типы и манипулировать с объектами каждой группы наиболее адекватными средствами.

Пусть  $O = \{O_i\}$  – множество типов объектов ПО, соответствующее, например, ER-диаграмме согласно инфологической модели ПО.

Обозначим через  $O_i$  имя структурированного типа данных, отвечающего некоторому значению признака  $P$ , на подмножестве объектов ПО и свяжем его с понятием абстрактного типа так, что имени  $O_i$  будет отвечать абстрактный тип данных  $O_i$ . Каждый элемент  $O_i \in O$ , где  $O$  – множество типов объектов рассматриваемой ПО, определим в виде  $O_i(\{RTO_1, \dots, RTO_n\}; \Omega_i)$ , понимая при этом тип объекта как множества данных со смыслами элементов структурированного типа  $RTO_1, RTO_2, \dots, RTO_n$  соответственно определенных посредством операции  $\Omega_i$  интерпретации типа.

Подтип данного типа представляет собой подмножество множества значений образующего типа и функция соответственно определена на множестве значений родового типа. Таким образом, при образовании подтипа накладываются ограничения на множество значений данных образующего типа. Анализ ПО заключается в организации данных в структуры. Результатом является модель среды. Ее графическая интерпретация традиционно задается ER-моделью, а формализованное описание для последующего анализа семантики данных БД предлагается авторами статьи. Под средой понимается тройка  $\langle O, RTO, A \rangle$ , где  $O$  – множество объектов,  $RTO$  – множество смыслов (свойств) на этих объектах,  $A$  – множество операций, которые можно выполнить с элементами множеств  $O$  и  $RTO$  при решении задач.

Итак ПО состоит из множества  $O$  информационных объектов  $O_i$ :  $O = \{O_1, O_2, \dots, O_i, \dots, O_r\}$ ,  $i = 1, 2, \dots, r$ ,

каждый из которых является параметризованным (родовым) типом данных вида:  $O_r(\{RTO_1^r, RTO_2^r, \dots, RTO_s^r\}; \Omega_r)$ , где  $\Omega_r$  – множество операций на  $RTO_i$ . В свою очередь каждый из параметров  $RTO_k^r$  на объекте  $r$  может быть представлен множеством своих значений  $\{A_1, A_2, \dots, A_k, \dots, A_q\}$ , в частности множествами значений имен атрибутов БД. Последнее позволяет задавать критерии сравнимости (в частном случае эквивалентности атрибутов РБД).

В этом случае родовой тип данных будет иметь вид:

$$O_r(\{ \langle RTO_1^r, \{A_1, A_2, \dots, A_m\} \rangle, \dots, \langle RTO_s^r, \{A_1, \dots, A_{ns}\} \rangle \}; \Omega_r).$$

## ВЫВОДЫ

Развитие систем управления в больших предметных областях с неизбежностью приведет к необходимости хранения смысла данных и построения баз метаданных. Предлагаемая концепция построения СППР

позволяет принимать эффективные стратегические решения в больших предметных областях корпоративных СППР, где ставятся задачи оценки качества земель, рационального использования природных ресурсов, управления развитием территорий и обеспечивается высокий уровень организации обмена данными между пользователями различных специальностей и уровней иерархии, использующих профессионально отличающую лексику языка описания ПО.

Предлагаемый подход предполагает поэтапное решение проблемы анализа семантики данных БД. В статье рассмотрен первый этап решения этой задачи – моделирование ПО средствами теории типов [7]. Этот метод дает возможность разрабатывать способы семантического сравнения множества значений атрибутов реляционной БД и эффективные алгоритмы сравнимости методом поиска в таблице на языке СУБД.

## ЛИТЕРАТУРА

1. Calenko M.Sh. Modelirovanie semantiki v bazah dannyh /M.Sh. Calenko. – М.: Nauka, 1989. – 285 s.
2. Zamulin A.V. Tipy dannyh v jazykah programmirovaniya i bazah dannyh /A.V. Zamulin. – Novosibirsk: Nauka, 1987. – 608 s.
3. Buslik M.M. Global'nye shemy reljacionnyh baz dannyh /M.M. Buslik. – Н.: HNURJe, 2002. – 67 s.
4. Dejt K. Vedenie v sistemy baz dannyh: per. s angl. /K. Dejt. – М.: Nauka, 2001. – 1071 s.
5. Mejer D. Teorija reljacionnyh baz dannyh: per. s angl. /D. Mejer. – М.: Mir, 1987. – 608 s.
6. Chen P.P.-S., The entity Relationship Model: Towards a Unified View of Data. – ACM Trans. Database Syst., 1976, #1, P.9-36.
7. Trahtengerc Je.A. Sub'ektivnost' v komp'juternoj podderzhke upravlencheskih reshenij /Je.A. Trahtengerc. – М.: SINTEG, 2001.

**Рецензент:** д.т.н., проф. Шаронова Н.В.,  
НТУ «Харьковский политехнический институт».