

МЕТОД ПРЕОБРАЗОВАНИЯ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ В ОНТОЛОГИЮ ИНФОРМАЦИОННОГО ПОЛЯ ОРГАНИЗАЦИОННОЙ СИСТЕМЫ

Горковский М.В.

Научный руководитель канд. техн. наук Якунин Ю.Ю.

Сибирский федеральный университет

Современные проблемы управления в социальных и экономических системах (активные системы) поставили новые задачи исследования, которые в свою очередь породили различные теории и подходы моделирования активных систем. К одной из таких проблем можно отнести сложность структуры и количества информации, идентификации важных критериев и получения зависимостей с управляющими воздействиями. Один из подходов моделирования активных систем [1] выделяет субъект и объект управления, взаимодействующие через информационное поле – носитель информации объекта и системы в целом. Информационное поле содержит в себе как структурированную информацию, которая может храниться в базах данных, так и неструктурированную (знания, представления об информации каждым отдельным индивидуумом).

Для описания структуры информационного поля выделяются 3 категории информации [2]: показатели – количественные характеристики объекта, классификаторы – качественные характеристики и фактические данные, выражающие значения показателей в отношении одного или нескольких классификаторов. Такая структура представляется в виде трехдольного графа (рис. 1), если отбросить левые и правые части, представляющие собой деревья.

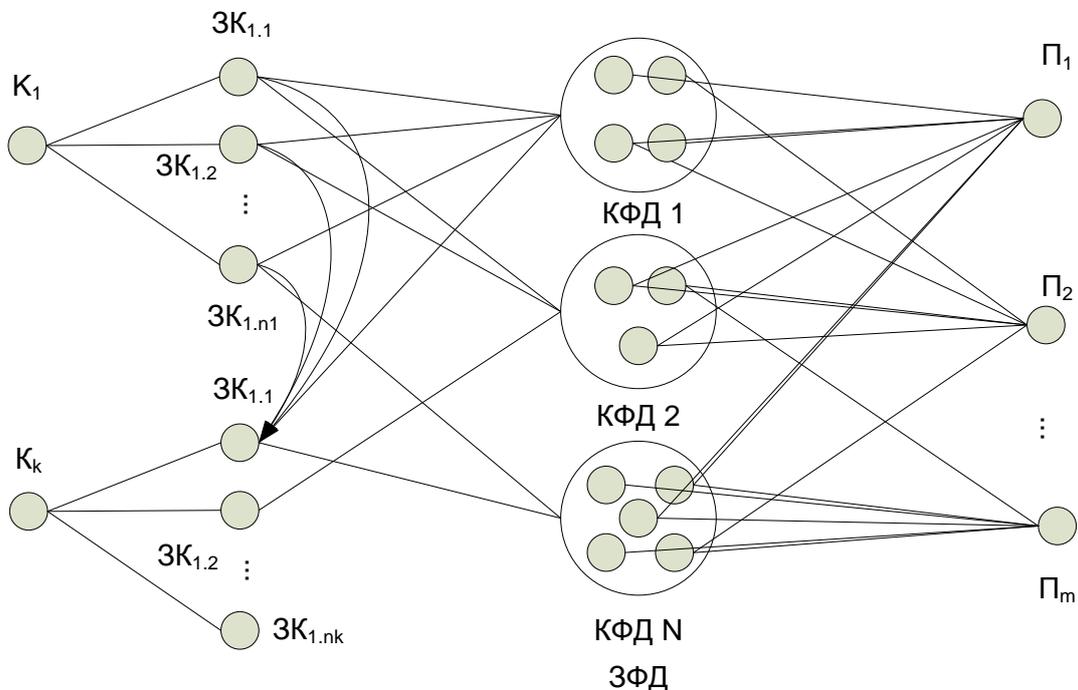


Рисунок 1 – трехдольный граф фактических данных

Левая доля представляет собой множество вершин последнего уровня графа классификаторов, состоящего из справочных данных предметной области. Правая доля представляет собой нижний уровень графа показателей, т.е. множество показателей имеющих связь со значениями фактических данных. Кластер фактических данных

(средняя доля) состоит из одного или нескольких значений фактических данных, которые используются для описания событий организационной системы имеющих несколько числовых характеристик. Например, при продаже товара необходимо учитывать количество проданного товара, цену продажи, цену по прайсу и др. Каждое значение такого фактического данного в кластере может быть связано с одним или несколькими показателями для определения сложных числовых характеристик события.

Данные в организационных системах чаще всего хранятся в базах данных, имеющих реляционную структуру. Таким образом, основным источником для построения информационного поля организационной системы будут являться реляционные БД, что порождает задачу получения методов преобразования информации из реляционных БД в онтологию информационного поля.

Автором предложены правила представления данных в реляционной модели и их сопоставления с правилами построения онтологии информационного поля организационной системы, а также разработан алгоритм построения онтологии на основе реляционной базы данных.

Чтобы выявить соответствия между реляционной таблицей и онтологией, необходимо найти в ней качественные и количественные показатели. Под количественными показателями будем понимать такие поля, которые содержат в себе числа и выражают количество некоторых условных единиц. Под качественными показателями понимаются сущности, описанные в таблице. Значения полей-классификаторов, в рамках одной таблицы, будут формировать значения классификаторов, а значения полей-показателей – фактические данные, которые будут объединяться в отдельный кластер для каждой строки таблицы.

Первичные и внешние ключи требуют особого подхода. Чтобы не создавать множество лишних связей, за состоянием каждой из которых придется следить, целесообразно первичные ключи считать классификаторами и привязывать к ним соответствующие значения классификаторов. Внешние ключи должны представлять собой связи и не будут иметь в онтологии явного отражения в виде отдельной вершины графа. Такой подход позволяет обрабатывать, в том числе таблицы, хранящие отношения типа «многие-ко-многим», представляющие собой таблицу без первичного ключа, состоящую из двух полей, содержащих внешние ключи.

Процесс классификации можно автоматизировать. Для этого предлагается использовать домены, созданные архитектором базы данных. Домены служат для группирования сущностей, значения которых лежат в рамках одного множества (например, множества действительных чисел), по определенному признаку. Их использование в базе данных говорит о том, что при ее создании уже был проведен анализ и выявление необходимых групп данных, имеющих общие свойства, иными словами, была решена схожая задача. Пользователю системы необходимо определить соответствия между доменами и категориями информации. При отсутствии доменов классификация проводится на основе типа данных поля, как единственной однозначной характеристики. Показателями следует считать поля, имеющие численные типы данных (INT, FLOAT и т.д.), классификаторами – все остальные.

Таким образом, можно сформулировать следующий алгоритм преобразования данных.

Для каждой таблицы, если она не обрабатывалась, выполнить:

1. Если таблица имеет внешние ключи, выполнить шаги 1-5 для каждой связанной таблицы.
2. Создать классификатор и категорию показателей по имени таблицы.

3. Определить категории информации, содержащиеся в полях таблицы, используя их домены (типы данных).
4. Для каждой строки в таблице:
 - 4.1 Создать значение классификатора, объединив все значения полей-классификаторов.
 - 4.2 Создать кластер, поместить в него фактические данные, связать с показателями и значениями классификаторов (в том числе по внешним ключам)
5. Отметить таблицу как обработанную

Для этого алгоритма не имеют значения порядок, в котором обрабатывать таблицы, и сложность связей. Единственное условие его работы – база данных должна быть приведена к третьей нормальной форме (3НФ). Это связано с тем, что в 3НФ информация об объектах не дублируется в рамках одной таблицы.

В качестве примера, рассмотрим реляционную базу данных, состоящую из трех таблиц.

Табл. 1

Студент

| Идентификатор | Имя | Дата рождения | Идентификатор группы |
|---------------|------|---------------|----------------------|
| 1 | Иван | 24.12.1981 | 1 |
| 2 | Петр | 23.02.1990 | 1 |

Табл. 2

Предмет

| Идентификатор | Предмет |
|---------------|----------|
| 1 | Биология |

Табл. 3

Группа

| Идентификатор | Группа |
|---------------|---------|
| 1 | BT15-01 |

Табл. 4

Успеваемость студента

| Идентификатор студента | Идентификатор предмета | Средний балл | Прослушано часов |
|------------------------|------------------------|--------------|------------------|
| 1 | 1 | 4,8 | 100 |
| 2 | 1 | 5 | 75 |

Таблица 1 содержит первичный ключ, два поля-классификатора и внешний ключ. Таблицы 2 и 3 имеют одинаковую структуру – первичный ключ и поле-классификатор. Таблица 4 имеет два поля, содержащие внешние ключи, и два поле-показателя.

Как уже упоминалось выше, первичные и внешние ключи требуют специального подхода. Являясь, по смыслу, классификаторами, они, тем не менее, редко несут в себе какое-то семантическое значение. В общем случае, целесообразно представлять их в графе не в виде отдельных узлов, а в виде связей между вершинами.

Поскольку первая таблица имеет внешний ключ, то сначала должна быть обработана таблица 3, на которую он указывает. Таким образом, сначала будет создан классификатор «группа», содержащий значение «BT15-01», затем классификатор «студент», содержащий два значения: «Иван,24.12.1981» и «Петр,23.02.1990», затем

значения классификатора «студент» будут связаны со значением классификатора «группа».

При обработке таблицы 2 будет построен классификатор «предмет», содержащий значение «биология».

Четвертая таблица содержит в себе информацию о связи между таблицами 1 и 2 и два поля-показателя. Таким образом, после обработки алгоритмом таблицы 4, построится два кластера, каждый из которых будет содержать два значения, связанных с показателями «средний балл» и «прослушано часов» (рис. 2).

Приведенный пример показывает, что данные из реляционных таблиц полностью можно представить с помощью онтологии информационного поля.

Исследование структуры информационного поля – важный шаг в развитии предложенного в [1] подхода к управлению активными системами. Описанный алгоритм, как один из этапов исследования, предназначен для обработки формализованных данных, имеющих реляционную структуру. Способы преобразования других формализованных данных, имеющих другую природу, например, объектную или документно-ориентированную, требуют отдельных исследований.

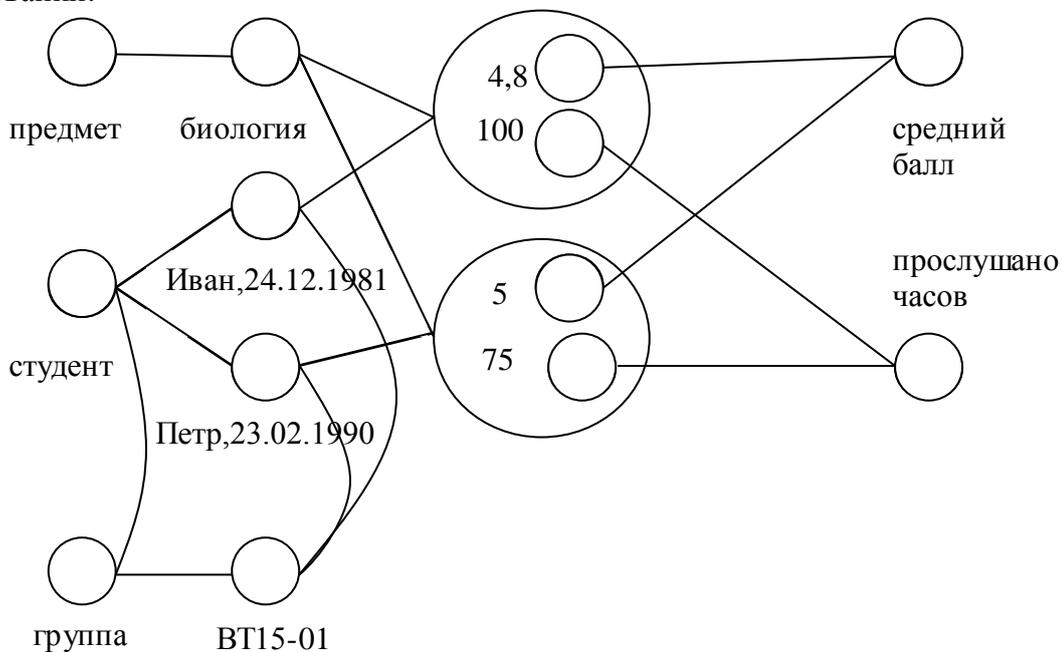


Рисунок 2 – трёхдольный граф фактических данных

Библиографический список

1. Якунин, Ю.Ю. Системный подход к моделированию организаций / В.М. Журавлёв, Ю.Ю. Якунин // Вестник сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнёва. – Красноярск: СибГАУ, 2010. – Вып. 6 (32). – С. 46-52.
2. Якунин, Ю.Ю. Алгоритмы расчёта комплексных показателей в динамических структурах представления данных / Ю.Ю. Якунин, А.А. Городилов // Вестник сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнёва. – Красноярск: СибГАУ, 2010. – Вып. 1 (27). – С. 59-63.