

УДК 004.62

**ИНТЕГРАЦИЯ ИС НА ОСНОВЕ ОБЪЕДИНЕНИЯ БАЗ ДАННЫХ
НА КОНЦЕПТУАЛЬНОМ УРОВНЕ АРХИТЕКТУРЫ**

Глава М. Г.

Ст. преп. каф. ЭКИТ

Одесский национальный политехнический университет, УКРАИНА

АННОТАЦИЯ. Рассмотрены проблемы и существующие методы интеграции ИС, проанализированы их недостатки. Предлагается технология объединения реляционных БД, как моделей ПрО, которая позволит автоматизировать создание концептуальной схемы ПрО (j+1)-го уровня, что даст возможность формировать интегрированные базы и хранилища данных ИС с экономией трудовых и временных затрат

Введение. Существенным конструктивным фактором успеха бизнеса являются современные информационные технологии. В современном мире уровень информатизации достиг такого уровня, что представить предприятие либо организацию любой сферы деятельности и размера без использования информационной системы (ИС), практически невозможно. Вопрос «Повлияет ли изменение бизнес-процессов предприятия на работу ИС» становится риторическим. Становится актуальным изучение вопросов реинжиниринга не только бизнес-процессов, но и самой информационной системы предприятия. Ведь любая информационная система оперирует большим количеством важной информации, которую необходимо сохранить в процессе реинжиниринга.

Цель работы. При реинжиниринге ИС специалисты сталкиваются с вопросами и проблемами интеграции данных [1], следовательно, существующие подходы к интеграции ИС требуют развития/доработки. Возникает необходимость в создании технологии, обеспечивающей интеграцию ИС на основе объединения баз данных (БД) на концептуальном уровне архитектуры.

Основная часть работы. Для создания общей концептуальной модели, объединяющей в себе представления сущностей двух и более хранилищ данных, необходимо объединить концептуальные модели предметных областей (ПрО). В силу того, что одну и ту же ИС (соответственно и концептуальную модель), можно построить по-разному, так же, как и одну задачу можно решить различными способами, становится актуальной проблема поиска подобных данных для минимизации дублирования данных (с точки зрения хранения) и времени получения пользователем информации. В связи с выше изложенным, рассмотрим несколько альтернативных уровней интеграции данных: интеграция на физическом, логическом и семантическом уровне. Интеграция данных на физическом уровне представляет собой конвертирование данных в единый формат, что является наиболее простым методом. Синтаксическая или логическая интеграция данных основывается на внешнем сходстве объединяемых данных, обеспечивает возможность доступа к данным в терминах единой глобальной схемы и учитывает структурные свойства данных, которые могут находиться в различных источниках. Семантический уровень интеграции основывается на содержательном сходстве объединяемых данных. Семантическая интеграция основывается на знании и учете природы данных. Данные должны храниться вместе с метаданными, что является более сложным в реализации, но значительно увеличивает комфортность работы.

Наиболее перспективным уровнем интеграции является семантический уровень, который позволит объединить не только структурно одинаковые данные, но и несущие одинаковую смысловую нагрузку. На сегодняшний день нет универсального подхода к объединению данных на семантическом уровне. Существуют только частные решения соответствующих задач.

В большинстве существующих методов объединения БД на семантическом уровне для подтверждения корректности результата необходимо привлечение экспертов. Это является основным недостатком предложенных методов.

Проанализировав существующие решения интеграции БД, можно сделать вывод, что техническое объединение БД не решает задачу интеграции данных. Задача объединения БД на концептуальном уровне остается актуальной, поэтому авторами предлагается технология

объединения реляционных БД, как моделей ПрО, которая позволит автоматизировать создание концептуальной схемы ПрО (j+1)-го уровня, что даст возможность формировать интегрированные базы и хранилища данных ИС с экономией трудовых и временных затрат.

Анализ отраслей бизнеса показал, что технология объединения моделей предметных областей зависит от сходства ПрО. Поскольку для ПрО, описывающих абсолютно разные сферы бизнеса, объединение вероятнее всего возможно только по объектам, которые не являются характерными представителями той или иной ПрО. Например, по контрагентам, сотрудникам и т.д. То есть полное сопоставление не целесообразно. Для подобных ПрО, отображающих один и тот же вид бизнеса, необходимо сопоставлять и объединять все сущности. Для ПрО, описывающих одну и ту же область бизнеса (например, торговлю, услуги), следует сопоставлять как объекты не определяющие ПрО, так и объекты, которые могут относиться к обоим сравниваемым областям бизнеса, так называемые «пограничные» объекты.

Для определения сходства ПрО предлагается воспользоваться описанием ПрО, предложенным в [2]:

$$\text{ПрО} = \langle E, R, P \rangle,$$

где P – проблемы, решаемые в конкретной ПрО.

В работе [3] предложена технология поиска проекций одних и тех же универсальных сущностей на ПрО, в которой предлагается сопоставлять объекты на основе значений свойств экземпляров этих объектов. Алгоритмы сопоставления различаются в зависимости от типа данных конкретного свойства.

Согласно предложенной технологии объекты потенциально подобных ПрО необходимо подготовить к сопоставлению: выделить существенные свойства, основываясь на количестве информации каждого свойства и оценке экспертов; проранжировать по значимости объекты каждой сравниваемой ПрО, основываясь на количестве и степени важности связей определенного объекта с другими в этой же ПрО, и количестве значимых свойств, измеряемых определенной шкалой (порядковой, номинальной, числовой); отсортировать кортежи по значениям порядковых и номинальных свойств, соблюдая полученный ранее ранг свойств.

Далее выполняется сопоставление свойств потенциально подобных объектов по их типам, определяется коэффициент соотношения подобных свойств объекта к их общему количеству. Полученный коэффициент сравнивается с порогом, заданным экспертным путем и делается вывод о том, являются ли сравниваемые объекты подобными.

При построении общей модели ПрО предлагается включить из максимально подобных объектов максимально подобные свойства и дополнить отличающимися свойствами, упорядоченными по значимости, получив своего рода вектор свойств.

Выводы. Предложенный авторами подход не решает в полной мере задачу интеграции данных, поскольку необходимы дополнительные исследования по объединению связей между сущностями в интегрируемых БД, а также по интеграции массовых проблем, решаемых ПрО, которые подвержены объединению.

Представленная технология, в свою очередь, позволяет автоматизировать создание концептуальной схемы ПрО (j+1)-го уровня, сократив трудовые и временные затраты на интегрирование БД. А также позволяет уменьшить количество задач, решаемых с привлечением экспертов, тем самым уменьшив субъективизм принятия ими решений.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Васильева Т.П., Глава М.Г. Основные проблемы и методы интеграции баз данных // Первый независимый научный вестник.– 2015.– № 1.– С. 28–32.
2. Малахов С. В. Інформаційна технологія моделювання складноструктурованих предметних областей в системах організаційного управління (теорія та реалізація) : автореф. дис. на здобуття наук. ступеня докт. техн. наук / С. В. Малахов. – Харків : НТУ «ХПІ», 2010. – 36 с.
3. Glava, M. and Malakhov, E. (2016). Searching Similar Entities in Models of Various Subject Domains Based on the Analysis of Their Tuples. 2016 International Conference on Electronics and Information Technology (EIT), 97-100. doi: 10.1109/ICEAIT.2016.7501001.