

СЕМАНТИЧЕСКАЯ ИНТЕГРАЦИЯ КОРПОРАТИВНЫХ СИСТЕМ

Корпоративные программные комплексы: особенности и проблемы

Гетерогенный характер корпоративных программных комплексов определяется различием архитектурных принципов, лежащих в их основе (файл—сервер, клиент—сервер, интернет-сервисы и др.), а также степенью структурированности составляющих их баз и хранилищ данных. Последние могут принимать формы от строгих реляционных таблиц до коллекций мультимедийных объектов с весьма сложной, динамической структурой.

Корпоративные программные комплексы также характеризуются быстрым ростом объемов данных, в среднем удваивающим эти объемы каждые пять лет. При этом в различных информационных системах, включающих сотни основополагающих сущностей, как правило, начиная с некоторого этапа жизненного цикла (обычно на стадии концептуального проектирования или сопровождения), с высокой вероятностью могут возникать дублирования семантически близких сущностей, логические противоречия и неопределенности. Такого рода несогласованности компонентов корпоративных программных комплексов неизбежно ведут к нестабильной и непредсказуемой динамике, с математической точки зрения выражающейся в спонтанной смене состояний, а в итоге — к неустойчивому (и/или не соответствующему спецификациям) поведению и, в конечном счете, к более или менее существенным сбоям, потере работоспособности, надежности и (информационной) безопасности.

Технология семантической интеграции: модели и средства

Решение перечисленных проблем в случае существенной гетерогенности (как архитектурной, так и структурной) корпоративных программных комплексов невозможно без построения адекватного интегрированного представления системного метауровня составляющих их компонентов на основе семейства математических моделей данных, инвариантных к выявленным архитектурным и структурным особенностям [1]. При этом модели должны носить расширяемый, объектный, динамический характер и допускать неограниченное расширение посредством метауровневых переходов [2, 3]. Семейство моделей должно включать математические средства как для представления данных в составе компонентов интегрируемых корпоративных программных комплексов, так и для манипулирования данными этих компонентов (т. е. как статическую, так и динамическую составляющие). Каждый вид моделей должен быть поддержан соответствующими инструментальными средствами автоматизированного проектирования, позволяющими обеспечить индустриальный уровень конвейерной интеграции надежных, безопасных, семантически целостных корпоративных приложений и их комплексов. На этой основе и предлагается строить технологию семантической интеграции объектов данных корпоративных программных комплексов.

Подобный подход «модель — инструментарий — реализация» дает возможность за счет унифицированного теоретического обобщения обеспечить принципиально новый уровень надежности, безопасности и целостности по сравнению с существующими «методологиями» построения корпоративных программных комплексов «промышленного» типа (Microsoft MSF, IBM RUP, Oracle CDM и др.). Применение перечисленных методологий, нацеленных, прежде всего, на узконаправленные, довольно однородные корпоративные программные комплексы «моновендорного» типа, к сожалению, в случае существенной архитектурной и структурной гетерогенности не приводит к практически приемлемым уровням программных систем по направлениям масштабируемости, отказоустойчивости, надежности и, в конечном итоге, информационной безопасности. Схожие результаты наблюдаются и в случае применения



теоретически перспективных подходов [4, 5], которые неадекватно поддержаны инструментальными средствами индустриального масштаба.

Модели представления и манипулирования данными

Для обеспечения адекватного моделирования семантической интеграции гетерогенных программных комплексов разработан подход, включающий объектные модели как для представления данных, так и для организации управления ими.

Общая технологическая схема интеграции корпоративных программных комплексов обеспечивает замкнутый, двунаправленный характер их построения, а также возможность повторного проектирования (реинжиниринга) [2]. Последняя возможность весьма важна для осуществления верификации корпоративных программных комплексов, существенно повышающей уровень безопасности и надежности их функционирования.

Технологическая схема интеграции корпоративных программных комплексов содержит этапы, соответствующие формам представления информации в гетерогенных программных комплексах, взаимодействующих в глобальной сетевой среде (естественный язык, математические модели, сопряжение с инструментальными средствами, управление контентом и др.), и уровни, детализирующие эти этапы (объекты, связи, события, примеры инструментальных средств и программных систем).

Контент-ориентированный подход позволяет обобщить на базе объектных моделей понятия данных и метаданных, унифицировать управление обработкой гетерогенных объектов, а также адекватно моделировать интернет-среду, что имеет принципиальное значение для повышения уровня информационной безопасности и обеспечения индустриальной надежности корпоративных программных комплексов.

Объектный характер моделей, реализующих схему интеграции «класс—объект—значение», позволяет обеспечить преемственность с традиционным объектно-ориентированным подходом к анализу и проектированию ПО (ООАД), теоретически перспективными подходами (концептуальный метод В. Э. Вольфенгагена [1], переменные области Д. С. Скотта [6]) и развить их в направлении интернет-среды [2].

Технологическая последовательность преобразований в модели выглядит следующим образом:

- а) терм конечной последовательности (например, λ -исчисления);
- б) логический предикат (используется логика высших порядков);
- с) фрейм (графическое представление);
- д) XML-объект (описание класса на уровне встроенного инструментального средства);
- е) UML-диаграмма (схема данных инструментального средства) в репозитории корпоративных программных комплексов.

При этом модель интегрированного представления контента основана на ситуативной интерпретации в форме семантических сетей, что обеспечивает интуитивную ясность и однозначность трактовки понятий предметными экспертами при построении схемы предметной области. Таким образом, технологическая схема семантической интеграции обеспечивает надежность и безопасность программных комплексов уже на стадии проектирования их компонентов. Эргономика модели обеспечивается графическим представлением в виде фреймов. Модель управления контентом организована в виде абстрактной машины с состояниями и ролевыми соотношениями, что позволяет естественно обобщить процессы, сходные для целого ряда подобных инструментальных средств: разграничение ролей и прав доступа к данным, построение шаблонов web-страниц, редакторский цикл публикации web-страниц и др. При этом важнейшие операции по управлению контентом — персонализированное манипулирование, определение, означивание и др. — формализованы



посредством языка абстрактной машины, для которого разработаны строгий, формальный синтаксис и денотационная семантика в терминах переменных доменов (включая порядок построения объектов контента, семантические функции и предложения для этих операций) [2]. Строгость и адекватность формализации (на уровне модели и инструментального средства) гарантируют высокий уровень целостности данных, безопасности при разграничении доступа к объектам корпоративного контента, а также надежное функционирование интегрирующих этот контент и поддерживающих управление им инструментальных систем.

Технологическая последовательность преобразований в модели имеет вид:

- a) терм теории вычислений Д. С. Скотта;
- b) функция над доменами (в логике высших порядков);
- c) фрейм (графическое представление);
- d) XML-объект (шаблон web-страницы ИС управления контентом);
- e) HTML-код (код web-страницы ИС управления контентом) на портале корпоративных программных комплексов.

Интегрированные корпоративные системы: архитектура и алгоритмы

Архитектурная схема объединенного хранилища гетерогенного корпоративного контента обеспечивает унификацию за счет использования обобщенных объектных ассоциативных связей на уровне гетерогенных данных и метаданных. С другой стороны, унификация манипулирования контентом гетерогенных информационных систем основана на использовании единой метанастройки над корпоративным хранилищем данных в форме интернет-портала. При этом динамическое, сценарно-ориентированное управление контентом в рамках порталной архитектуры обеспечивается соотношениями, реализованными в форме сценариев языка программирования, изменяющих состояния абстрактной машины.

Сценарии другого рода реализуют безопасное персонализированное манипулирование контентом, поддержанное моделью многопараметрического функционала и инструментальным средством управления контентом.

Важное средство поддержания семантической целостности корпоративных программных комплексов, позволяющим обеспечить их надежную и безопасную работу, — алгоритм интеграции объектов данных в информационных системах, являющихся компонентами таких комплексов. Обобщенная последовательность шагов алгоритма семантической интеграции объектов данных имеет вид:

- a) спроектировать концептуальную модель объектов данных предметной области и построить модель объектов данных для вновь разрабатываемых компонентов корпоративного программного комплекса;
- b) произвести реинжиниринг существующих компонентов корпоративного программного комплекса до уровня концептуальной модели объектов данных предметной области и модели объектов данных;
- c) определить семантически приоритетные объекты данных для каждого из компонентов концептуальной модели предметной области и модели объектов данных в смежных областях;
- d) устранить дублирование и противоречия сущностей и связей на уровне предметной области и объектов данных из областей интеграции (в итоговых концептуальных моделях объектов данных предметной области и моделях объектов данных каждого компонента остаются лишь семантически приоритетные сущности и связи);
- e) повторить этапы (c) и (d) для концептуальной модели предметной области и модели объектов данных интегрированного корпоративного программного комплекса как объединения концептуальных моделей предметной области и моделей объектов данных для компонентов.



Алгоритм интеграции практически применим как для построения новых корпоративных программных комплексов, характеризующихся повышенной надежностью и безопасностью, так и для интеграции новых информационных систем в состав существующих комплексов. При этом предполагается, что комплексы должны либо изначально строиться на основе описанного подхода, либо адаптироваться к нему. Разработанные модели, алгоритмы и инструменты допускают возможность повторного проектирования, хотя и неизбежно влекут существенные затраты времени и средств. Поддержка полностью автоматического реинжиниринга не представляется возможной в силу неоднозначностей интерпретации семантики схожих сущностей различных информационных систем, составляющих программные комплексы.

Заключение

Применение предложенного подхода позволило реализовать повышенный уровень информационной безопасности за счет семантической интеграции в составе корпоративного программного комплекса семейства гетерогенных компонентов — модулей современных Oracle-систем финансового планирования и управления, унаследованной системы кадрового учета и слабоструктурированного мультимедиа-архива. Реализация архитектурных надстроек в форме интернет- и интранет-порталов, манипулирующих компонентами гетерогенного корпоративного программного комплекса, обеспечила ускоренное внедрение безопасных программных комплексов с гибким разграничением доступа к контенту в компаниях многопрофильной международной группы «ИТЕРА», объединяющей около 10 тысяч сотрудников.

Комплексный характер технологии (математические модели, инструментальные средства, порталное архитектурное решение) обеспечивает целостное, надежное и безопасное сопряжение с широким диапазоном современных инструментальных средств (IBM Rational, Microsoft Visual Studio .NET, Oracle Developer) разработки корпоративных программных комплексов; при этом поддерживаются стандарты UML и XML.

Функциональные преимущества подхода по сравнению с перечисленными выше аналогами — обеспечение целостности данных, безопасное манипулирование сложными, разнородными и в разной степени структурированными объектами данных, надежное объединение компонентов различной архитектуры — обусловлены ориентацией моделей и инструментальных средств на семантическую интеграцию гетерогенных порталных корпоративных программных комплексов. Сравнение важнейших макропоказателей — совокупной стоимости владения, возврата инвестиций и сроков внедрения — подтверждает предварительные качественные оценки. По результатам внедрения в МГК «ИТЕРА» превосходит передовые коммерческие аналоги по перечисленным показателям в среднем на 30–40 %, существенное улучшение достигается и в направлениях надежности и безопасности объектов данных и программного кода.

Результаты исследования, созданные на их основе программные комплексы, учебные программы и курсы внедрены в ряде корпоративных коммерческих и государственных структур, в частности в «ИТЕРА», Минпромэнерго РФ и ИПУ РАН [2].

СПИСОК ЛИТЕРАТУРЫ:

1. Вольфенгаген В. Э., Яцук В. Я. Аппликативные вычислительные системы и концептуальный метод проектирования систем знаний. М: МО СССР, 1987.
2. Зыков С. В. Проектирование корпоративных порталов. М.: МФТИ, 2005. — 258 с.
3. Зыков С. В. Технология разработки корпоративных программных комплексов — от моделей к внедрениям // Безопасность информационных технологий. 2008. № 1. С. 87–90.

