

КРАТКИЕ СООБЩЕНИЯ

УДК 339.138

**СОВРЕМЕННЫЕ ТЕХНОЛОГИИ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ
ИНФОРМАЦИОННОГО ОБМЕНА В СЛОЖНЫХ РАСПРЕДЕЛЕННЫХ
ИНФОРМАЦИОННЫХ СИСТЕМАХ**

С.С. КУЛИКОВ

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь**Поступила в редакцию 3 марта 2006*

Предложено решение проблемы повышения эффективности взаимодействия компонентов распределенных информационных систем на примере систем дистанционного обучения. Даны рекомендации по применению современных технологий повышения надежности представления данных в таких системах.

Ключевые слова: распределенные информационные системы, системы дистанционного обучения, универсальные форматы, надежность.

Введение

В настоящее время особое внимание уделяется вопросам интеграции информационных ресурсов в единое информационное пространство с целью обеспечения быстрого беспрепятственного доступа к необходимой информации и устранения избыточного дублирования данных. В рамках данного процесса расширяются требования к создаваемым информационным ресурсам, что касается и систем дистанционного обучения (СДО). В соответствии с этими требованиями многими учреждениями образования разрабатываются информационные ресурсы для создания единой информационно-образовательной среды (ЕИОС), базирующейся на принципах проекта "Semantic Web", разрабатываемого организацией "World Wide Web Consortium (W3)". В статье рассмотрены основные подходы к построению многокомпонентной распределенной информационной системы, соответствующие этим принципам, а также даны рекомендации по использованию современных технологий и языков универсального представления данных.

Формулировка проблемы

При разработке многокомпонентной распределенной информационной системы или интеграции в единую систему разрозненных приложений основной проблемой является согласование форматов представления данных и обеспечение взаимного доступа к данным. Данная проблема приводит к понижению эффективности работы таких систем и снижению надежности информационного обмена. Разработчикам приходится решать следующие задачи: выбор формата представления данных при внешнем информационном обмене; выбор транспортного протокола высокого уровня, отвечающего требованиям универсальности и надежности; обеспечение прозрачного взаимного доступа к данным. От решения перечисленных задач зависит скорость и надежность работы разрабатываемого программного обеспечения (ПО), простота его сопровождения и удобство работы с ним конечных пользователей.

Решение проблемы совместимости

Принципы, положенные в основу ЕОИС, предполагают реализацию единого подхода к структурированию, взаимосвязи и моделированию предметных областей при построении СДО уровня отдельного учебного заведения или распределенных СДО. Особое внимание уделяется вопросам совместимости форматов представления данных, в связи с чем широкое распространение при разработке таких СДО получили современные языки представления данных XML (Extensible Markup Language, расширяемый язык разметки), RDF (Resource Description Framework, структура описания ресурсов) и OWL (Web Ontology Language, язык онтологии веб). Последние два языка (RDF и OWL) предназначены для описания метаданных информационных ресурсов и моделей предметных областей при разработке автоматизированных информационных систем в сети Интернет. В силу своей специфики эти языки являются оптимальным решением при описании обучающих курсов и их составных частей.

Язык XML как средство хранения и передачи данных является высокоэффективным инструментом, сочетающим в себе такие преимущества, как гибкость описания данных, однозначность трактовки записей, универсальность представления и простоту обработки информации. Технологии обмена данными на основе XML являются оптимальными по соотношению затраты/результат при сохранении универсальности и совместимости форматов [1]. XML — это новый SGML-производный язык разметки документов (SGML — (Standard Generalized Markup Language, стандартный обобщенный язык разметки), позволяющий структурировать информацию различных типов, используя для этого произвольный набор инструкций. XML может использоваться в любых приложениях, которым нужна структурированная информация — от сложных геоинформационных систем, оперирующих большими объемами данных, до обычного ПО, использующего этот язык для описания и хранения служебной информации. Можно выделить ряд задач, связанных с созданием и обработкой структурированной информации, для решения которых оптимальным средством является XML.

В первую очередь, эта технология используется при разработке сложных (распределенных) информационных систем с большим количеством приложений, связанных потоками информации самой различной структурой. В этом случае XML-документы выполняют роль универсального формата для обмена информацией между отдельными компонентами большой системы. XML также является базовым стандартом для нового языка описания ресурсов (RDF), позволяющего решить многие проблемы, связанные с поиском информации в глобальных компьютерных сетях, обеспечением контроля за содержимым сетевых ресурсов, созданием электронных библиотек и т.д.

Язык XML позволяет описывать данные произвольного типа, а потому может использоваться для представления специализированной информации, например химических, математических, физических формул, объемных текстов, журнальных статей и т.д. Технология сериализации данных (представления в переносимом формате сложнотипизированных данных) позволяет представить в XML-формате даже такие сложные структуры, как массивы произвольных элементов, числовые данные любой длины, данные во внутренних форматах прикладного ПО и иные данные, представление которых в универсальной форме до недавнего времени было затруднено или невозможно.

XML-документы могут использоваться в качестве промежуточного формата данных в трехзвенных системах, к классу которых относятся многие СДО. Обычно схема взаимодействия между приложениями и системой управления базами данных (СУБД) зависит от конкретной СУБД и диалекта языка SQL (Structured Query Language, язык структурированных запросов), используемого для доступа к данным. Однако в том случае, если результаты запроса будут представлены в некотором универсальном текстовом формате, звено СУБД станет прозрачным для приложения, чем будет достигнуто повышение переносимости кода разрабатываемой СДО и ее универсальность.

Информация, содержащаяся в XML-документах, может изменяться, передаваться клиенту или обновляться по частям. При этом в качестве клиента может выступать компонент приложения, отдельное приложение, группа приложений, информационная система и иные объекты данного класса. Стандарт XML позволяет ссылаться на отдельные элементы докумен-

та с учетом их вложенности. Данное преимущество XML позволяет значительно снизить нагрузку на каналы передачи данных, что особенно актуально в случае передачи больших объемов информации при одновременной работе с СДО большого количества пользователей.

Поскольку XML-документ представляет собой обычный текстовый файл, в котором при помощи специальных маркеров создаются элементы данных, автоматически решаются многие проблемы, связанные с передачей по каналам связи бинарных данных.

Основным достоинством XML-документов является то, что при относительно простом способе создания и обработки (обычный текст может редактироваться любым тестовым редактором и обрабатываться стандартными XML-анализаторами), они позволяют создавать структурированную информацию, которую легко хранить, передавать и обрабатывать. Данное свойство XML делает этот язык высокоэффективным инструментом построения распределенных информационных систем, т.к. с его использованием автоматически решаются очень актуальные в прошлом проблемы представления нестандартных данных и обмена ими.

На основе XML базируется технология RDF, представляющая собой фундаментальную разработку в области семантических интернет-ресурсов. Одним из основных преимуществ RDF является информационная модель, положенная в основу данной технологии. RDF определяет простую, но в то же время мощную и гибкую модель данных, состоящую из троек (субъект, предикат, объект), где субъектом и предикатом являются URI (Uniform Resource Identifier, универсальный идентификатор ресурсов), а объектом может быть либо URI, либо литерал. Такая простая модель позволяет представлять объект и его свойства для анализа в более сложных многоуровневых системах, а также для упрощения классификации и поиска объектов. Эффективность данной информационной модели подтверждена на практике [2], и в настоящее время технология находится в стадии интенсивного развития.

RDF-схемы отличаются от XML-схем структуры документов тем, что они определяют не разрешенный синтаксис, а классы, характеристики и их взаимосвязи: они действуют напрямую на уровне информационной модели, а не на синтаксическом уровне. В масштабах глобальных компьютерных сетей ключевой технологией являются именно RDF-схемы, поскольку они позволяют системам принимать собственные решения по поводу информации, собранной в сети.

Язык OWL описывает ту же онтологию, но в более простом синтаксисе и в формате XML. Предназначение OWL заключается в том, чтобы интеллектуальные системы, воспринимающие конструкции онтологии, могли более эффективно работать с документами. Данный вопрос особенно важен при построении СДО, так как данному классу систем приходится постоянно обрабатывать большое количество разнообразной документации по самым различным темам.

Решение проблемы интеграции

По мнению аналитиков, интеграция данных является основным инструментом обеспечения функционирования масштабных систем обработки информации, к которым относятся и СДО. При планировании архитектуры таких систем специалистам приходится решать, какие технологии смогут наилучшим образом подойти к выполнению следующих обязательных задач по обработке данных: перемещению, управлению и доступу. В то же время вопрос выбора оптимальной технологии и комплекса инструментальных средств решения перечисленных задач осложняется многообразием существующих технологий интеграции данных, а также частичным перекрытием функциональных возможностей инструментов интеграции. Таким образом, разработка методологии выбора оптимальных средств и инструментов решения упомянутых выше задач позволит существенно упростить и ускорить этап анализа и проектирования информационной системы, одновременно снизив вероятность принятия неверного решения.

Наиболее простым способом выбора технологии является сравнение интеграционных подходов друг с другом, сопоставление их основных характеристик — возможностей по преобразованию данных и задержки в предоставлении данных. При решении большинства интеграционных задач данные должны быть преобразованы тем или иным способом — структурно (например, чтобы снять различия между исходной и целевой схемами для обеспечения согла-

сованности данных) или семантически (например, чтобы устранить несоответствия функциональных значений в различных системах). Технологии интеграции данных могут существенно отличаться друг от друга с точки зрения возможностей преобразования — от незначительной поддержки трансформирования или ее отсутствия (передача файлов) до широких возможностей преобразования (средства ETL (Extract, Transform and Load, извлечь, преобразовать и загрузить)). Обеспечение оптимального времени задержки в предоставлении данных приобретает особое значение, так как все большее количество информационных систем переходит к режиму работы в реальном времени.

Общая рекомендация по выбору технологии интеграции данных может быть такой: останавливаться на стандартных технологиях, которые могут быть использованы во всей организации на самых различных уровнях построения информационной системы; поскольку для решения задач интеграции данных нет универсального подхода, необходимо соотносить требования, предъявляемые к интеграционному решению, с реальными характеристиками имеющихся технологий. Частично упростить решение вопроса о выборе технологии интеграции данных позволяет архитектурный подход к решению поставленной задачи. В таблице представлен сравнительный анализ основных подходов к интеграции данных.

Таблица. Сравнительный анализ основных подходов к интеграции данных

Стратегия/подход	Организационная структура	Продолжительность развертывания	Экономия при интеграции метаданных	Необходимое аппаратное обеспечение
Перенос структур	Централизованная — основное внимание уделяется быстрому снижению дополнительных затрат	Невысокая	Высокая	Мощное
Разработка проекта с нуля	Централизованная — слияние двух одинаковых по величине разнородных информационных систем	Высокая	Высокая	Мощное
Определение единого стандарта и последующий переход к нему	Централизованная — слияние двух неодинаковых по величине разнородных информационных систем	Средняя	Средняя	Мощное
Синхронизация	Централизованная — управление основными данными	Невысокая	Средняя	Среднее
Согласованные витрины данных	Децентрализованная (при наличии единого общеорганизационного представления информации)	Средняя	Средняя	Отсутствует
Распределенные запросы	Децентрализованная — быстрое временное решение	Высокая	Минимальная	Отсутствует

При рассмотрении таблицы заметно наличие зависимости между необходимым уровнем интеграции метаданных и продолжительностью реализации подхода. Подходы, которые можно быстро внедрить — перенос структур, распределенные запросы — требуют самого низкого уровня интеграции метаданных. И наоборот, подходы, требующие наивысшего уровня интеграции метаданных — особенно разработка проекта с нуля — наиболее трудоемкие, на их реализацию часто уходит несколько лет. Помимо этого, методы отличаются друг от друга по величине экономии аппаратных средств. Перенос структур предоставляет быструю экономию средств, а подходы разработка проекта с нуля и определение единого стандарта и последующий переход к нему могут окупиться только в долгосрочной перспективе, особенно в случае, если организация не возлагает на создаваемую информационную систему коммерческих задач. Большинство других подходов на самом деле увеличивает расходы на аппаратное обеспечение, поскольку для их реализации требуется как минимум дополнительный сервер.

Таким образом, можно сформулировать несколько практических рекомендаций по выбору метода интеграции данных и аналитических структур. При рассмотрении различных методов и технологий необходимо обратить внимание на целый ряд факторов, наиболее часто встречающиеся из которых следующие:

- стратегические цели руководства (сокращение расходов или инновационные направления);

- организационная структура (централизованная или децентрализованная) и степень автономии отделов организации (или отдельных организаций) в проекте;

тип слияния или поглощения – слияние одинаковых или неодинаковых по величине однородных или разнородных информационных систем;
временные рамки для завершения проекта;
доступные материальные ресурсы – как часть текущего бюджета или как целевое финансирование проекта;
характер данных и аналитических структур – хранилища/витрины данных, операционные склады данных, системы отчетности, табличные витрины, архивы мультимедийной информации, обучающие курсы и т.д.;
стабильность операционной среды;
удовлетворенность пользователей существующими средствами доступа к данным и их обработки;
доступные вспомогательные инструменты/технологии.

Заключение

Анализ рассмотренных выше технологий и их реализация на практике позволяют прийти к выводу, что перевод распределенных информационных систем на универсальные форматы представления и описания данных позволяет выйти на новый уровень межсистемной интеграции и взаимодействия с конечными пользователями. СДО, спроектированные для работы с такими технологиями, получают возможность беспрепятственно обмениваться данными, ссылаться на обучающие курсы партнерских информационных ресурсов и, в конечном итоге, интегрироваться в единое информационное обучающее пространство.

MODERN TECHNOLOGIES OF IMPROVING THE EFFECTIVENESS OF INFORMATION INTERCHANGE IN COMPLEX DISTRIBUTED INFORMATION SYSTEMS

S.S. KULIKOV

Abstract

The approach to resolve the problem of improving the effectiveness of interaction between components of distributed information systems on the base of distant education systems is reviewed in this article. Some recommendations on modern technologies of improving the reliability of data producing are given.

Литература

1. Куликов С.С. // Дистанционное обучение — образовательная среда XXI века: Сб. науч. материалов V МНМК. Минск, БГУИР. 10–11 ноября 2005 г. Минск, 2005. С. 347–350.
2. Куликов С.С. // Инженерно-педагогическое образование: проблемы и пути развития: Сб. науч. статей МНМК. Минск, МГВРК. 7–8 октября 2004 г. Минск, 2004. С. 190–195.