

ГРНТИ 004.435:004.623:004.624

Подходы и проблемы обмена данными между информационными системами

Габбасов М.Б. – канд. физ.-мат. наук, доцент, генеральный директор Компании системных исследований "Фактор", Республика Казахстан, +77019082511,
E-mail: mars@factor.kz

Куанов Т.Д. – заведующий отделом Компании системных исследований "Фактор", Республика Казахстан, +77013006556,
E-mail: tolybay@factor.kz

Обмен данными между двумя и более информационными системами всегда являлся актуальной и сложной задачей. За последние три десятилетия данная задача прошла свой эволюционный путь от ручного способа обмена до возможности автоматического обмена данными между системами. В данной работе рассматриваются различные современные технологии для организации обмена данными между информационными системами, в том числе между OLTP- и OLAP-системами. Приведено описание технологий ETL, MDM и технологий, основанных на пространствах имен, в том числе "Синхронизатор ТОФИ", разрабатываемый в Компании системных исследований "Фактор". Подробно излагается формальное понятие "данных", используемое в технологии ТОФИ, пространство имен ТОФИ и функциональные возможности синхронизатора ТОФИ по обмену данными и метаданными. Функционально синхронизатор осуществляет сопоставление метаданных различных информационных систем на основе систем кодирования, автоматически (по расписанию) или вручную производит формирование метаданных, обмен метаданными с другими системами, формирование данных для передачи другой системе, прием и передачу набора данных из/в другой(-ую) информационной(-ую) системы(-у) и т.д. Обмен данными, основанный на пространстве имен ТОФИ, может осуществляться в автоматическом режиме на основе синхронизации метаданных. **Ключевые слова:** данные, метаданные, обмен данными, OLTP, OLAP, пространство имен.

Ақпараттық жүйелер арасында мәліметтермен алмасу жолдары мен проблемалары

Габбасов М.Б. – ф.-м.ғ.к., доцент, "Фактор" жүйелік зерттеулер компаниясының бас директоры, Қазақстан Республикасы, +77019082511, e-mail: mars@factor.kz

Куанов Т.Ж. – "Фактор" жүйелік зерттеулер компаниясының бөлім меңгерушісі, Қазақстан Республикасы, +77013006556, e-mail: tolybay@factor.kz

Екі немесе одан да көп ақпараттық жүйелердің өзара мәліметтермен алмасу процесі күрделі және қажетті есеп болып табылады. Соңғы үш онжылдықта бұл есеп қолмен енгізу әдісінен бастап жүйелердің өзара автоматтандырылған түрде мәліметтермен алмасу деңгейіне дейін өзінің эволюциялық жолын өтті. Бұл жұмыста ақпараттық жүйелер арасындағы өзара мәліметтермен алмасуды ұйымдастырудың қазіргі замандағы әртүрлі технологиялары қарастырылады, соның ішінде OLTP- мен OLAP-жүйелерінің арасындағы алмасу процесі қамтылған. ETL, MDM сияқты технологиялардың және есімдер кеңістігіне негізделген технологиялардың сипаттамалары келтірілген, соның ішінде "Фактор" жүйелік зерттеулер компаниясында жасалып жатқан "ТОФИ синхронизаторы" жайлы сөз қозғалған. ТОФИ технологиясында қолданылатын "мәліметтер" түсінігі, ТОФИ есімдер кеңістігі және ТОФИ синхронизаторының функционалды мүмкіншіліктері жан-жақты зерттелген. Функционалды аспап ретінде синхронизатор кодтау жүйелері негізінде әртүрлі ақпараттық жүйелердің метамәліметтерін сәйкестендіреді, автоматты түрде (белгіленген кесте бойынша) метамәліметтерді дайындайды, басқа ақпараттық жүйелермен метамәліметтермен алмасады, басқа жүйелерге беру үшін мәліметтерді дайындайды, басқа жүйелерден мәліметтер жиынтығын қабылдайды және басқа да жұмыстар атқарады. ТОФИ есімдер кеңістігіне негізделген мәліметтермен алмасу құбылысы метамәліметтерді сәйкестендіру арқасында автоматтандырылған түрде жүруі мүмкін.

Түйін сөздер: мәліметтер, метамәліметтер, мәліметтермен алмасу, OLTP, OLAP, есімдер кеңістігі.

Approaches and problems of data exchange between information systems

Gabbassov M.B. – Ph.D., General director of System Research Company "Factor", Republic of Kazakhstan, +77019082511, e-mail: mars@factor.kz

Kuanov T.D. – Head of Department of System Research Company "Factor", Republic of Kazakhstan, +77013006556, e-mail: tolybay@factor.kz

The exchange of data between two or more information systems has always been an urgent and complex task. Over the past three decades, this task has passed its evolutionary path from a manual mode of exchange to the possibility of automatic data exchange between systems. In this paper, various modern technologies are considered for the organization of data exchange between information systems, including between OLTP and OLAP-systems. The description of technologies ETL, MDM and technologies based on namespaces, including "The synchronizer TOFI", developed in the System Research Company "Factor". The formal notion of "data" used in TOFI technology, the TOFI namespace and the functionality of the TOFI synchronizer for data and metadata exchange are described in detail. Functionally, the synchronizer compares the metadata of various information systems based on coding systems, automatically (on schedule) or manually generates metadata, exchanges metadata with other systems, generates data for transmission to another system, receives and transmits a data set from / to another (s) Information system (s), etc. Data exchange based on the TOFI namespace can be performed in an automatic mode based on metadata synchronization.

Key words: data, metadata, data exchange, OLTP, OLAP, namespace.

1 Введение

Обмен данными между двумя и более информационными системами всегда являлся актуальной и сложной задачей. За последние три десятилетия данная задача прошла свой эволюционный путь от ручного способа обмена до возможности автоматического обмена данными между системами. В данной работе рассматриваются различные современные технологии для организации обмена данными между информационными системами. Термины "данные" и "информационная система" понимаются в смысле стандартов (ISO/IEC 2382-1:1993:1-20) и (ISO/IEC 2382:2015:1-30) соответственно, а именно: данные – поддающиеся многократной интерпретации представление информации в формализованном виде, пригодном для передачи, связи и обработки; информационная система – система, предназначенная для хранения, поиска и обработки информации, и соответствующие организационные ресурсы (человеческие, технические, финансовые т т.д.), которые обеспечивают и распространяют информацию.

2 Обзор литературы

Существуют множество классификации информационных систем, для наших целей мы поделим все информационные системы на два класса: OLTP-системы (On-Line Transaction Processing) и OLAP-системы (On-Line Analytical Processing). OLTP-системы предназначены для автоматизации операционных бизнес процессов, в которых, как правило, используются структурированные, повторяющиеся задачи обработки данных. OLAP системы, или системы складирования данных, предназначены для аналитической обработки больших объемов данных. Термин OLAP ввёл Эдгар Кодд в публикации в журнале Computerworld в 1993 году (Codd Edgar F., 1993 : 56), в которой он предложил

12 принципов аналитической обработки, по аналогии с 12-ю правилами для реляционных баз данных, сформулированными им же десятилетием ранее (Codd Edgar F., 1970 : 377-387). Главный принцип концепции складирования данных состоит в том, что к данным, сохраняемым для анализа, может быть обеспечен наиболее эффективный доступ только при условии выделения их из операционной (транзакционной) системы. Такое разделение систем привело к появлению задачи миграции данных из одного типа систем в системы другого типа.

Такая классификация информационных систем выделяет следующие возможные типы обмена данными между ними:

- 1) обмен данными между OLTP-системами;
- 2) миграция данных из OLTP-систем в OLAP-системы;
- 3) обмен данными между OLAP-системами.

3 Материалы и методы

3.1 Обмен данными между OLTP-системой и OLAP-системой

Исторически первой появилась задача миграции данных из OLTP-систем в OLAP-системы, основным инструментом для решения которой является ETL-приложение (extraction, transformation, loading – извлечение, преобразование и загрузка данных). Приложения ETL извлекают информацию из одного или нескольких источников, преобразуют ее в формат, поддерживаемый системой хранения и обработки, которая является получателем данных, а затем загружают в нее преобразованную информацию. Изначально ETL-системы использовались для переноса информации из более ранних версий различных информационных систем в новые. В настоящее время ETL-системы все более широко применяются именно для консолидации данных с целью их дальнейшего анализа. Очевидно, что поскольку ХД могут строиться на основе различных моделей данных (многомерных, реляционных, гибридных), то и процесс ETL должен разрабатываться с учетом всех особенностей используемой в ХД модели.

Независимо от особенностей построения и функционирования ETL-система должна обеспечивать выполнение трех основных этапов процесса переноса данных (ETL-процессы):

- извлечение данных. На этом этапе данные извлекаются из одного или нескольких источников и подготавливаются к преобразованию. Следует отметить, что для корректного представления данных после их загрузки в хранилище данных из источников должны извлекаться не только сами данные, но и информация, описывающая их структуру, из которой будут сформированы метаданные для хранилища;
- преобразование данных. Производятся преобразование форматов и кодировки данных, а также их обобщение и очистка;
- загрузка данных - запись преобразованных данных в соответствующую систему хранения.

Перемещение данных в процессе ETL можно разбить на последовательность процедур, представленных следующей функциональной схемой (рисунок 1):

- 1) извлечение. Данные извлекаются из источников и загружаются в промежуточную область;
- 2) поиск ошибок. Производится проверка данных на соответствие спецификациям и возможность последующей загрузки в ХД;
- 3) преобразование. Данные группируются и преобразуются к виду, соответствующему структуре ХД;
- 4) распределение. Данные распределяются на несколько потоков в соответствии с особенностями организации процесса их загрузки в ХД;
- 5) вставка. Данные загружаются в хранилище-получатель.

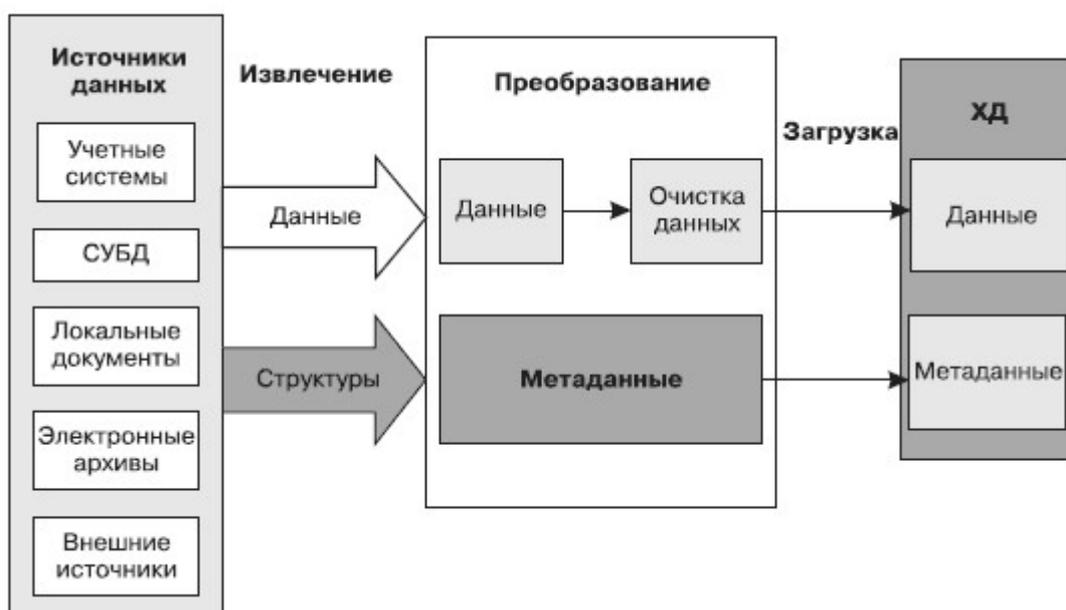


Рисунок 1 – Обобщенная схема процессов ETL.

3.2 Обмен данными между OLTP-системами

Довольно часто возникает задача обмена данными между OLTP-системами, например, системой управления кадрами и системой финансового учета. В последние годы для таких целей предлагается использовать новый класс корпоративных систем – MDM (Master Data Management – управление мастер-данными). Мастер-данные – это данные, содержащие ключевую информацию о бизнесе, в том числе о клиентах, о продуктах, о работниках, о технологиях и материалах (Черняк, 2007 : 54-62). Мастер-данные специфичны тем, что относительно редко изменяются и по своей природе не являются транзакционными. В определенных случаях мастер-данные поддерживают транзакционные процессы и операции, но в большей степени они используются для аналитической деятельности и подготовки отчетов.

Иногда для MDM используется альтернативное название – управление справочными данными (Reference Data Management, RDM). Предназначение MDM состоит в обеспечении целостного взгляда на основные компоненты бизнеса, так как такие системы обеспечивают синхронизированные справочники для обмена данными между OLTP-системами. Подлинная сложность мастер-данных заключается не в них самих, а в опи-

сывающих их метаданных. Метаданные, описывающие мастер-данные, могут претерпевать от нескольких единиц до сотен изменений в месяц, отслеживать все эти изменения в стихийно сложившейся системе MDM практически невозможно. Решение проблемы сложности состоит в создании отдельной единой среды MDM, которая будет выполнять обе функции – и сбор, и представление данных для последующей обработки средствами бизнес-аналитики. Создание такой среды может быть реализовано в трех вариантах (Черняк, 2007 : 56-60): 1) федеративный, 2) интегрированный, 3) гибридный. При федеративном подходе концентратор мастер-данных (master data hub) в соответствии с установленными правилами осуществляет асинхронный обмен данными, чтобы обеспечить согласованность мастер-данных. Основным предметом деятельности заключается в формулировании этих правил. Преимуществом этого подхода является невмешательство в действия существующих систем и сохранение контекста, а недостатком – то, что он эффективен лишь при относительно небольшом разнообразии мастер-данных и не выдерживает нагрузки при усложнении среды мастер-данных. Интегрированное решение заключается в создании автономной системы MDM, способной самостоятельно собирать изменения в мастер-данных. В таком случае все эти данные хранятся в централизованном репозитории. Данный подход обеспечивает наиболее полное и адекватное представление мастер-данных. Гибридный подход является промежуточным между первым и вторым подходами.

3.3 Обмен данными между OLAP-системами

В последнее десятилетие в информационных технологиях с развитием интернета сформировались новые подходы к автоматизации обмена данными между системами, основанные на пространствах имен, более ориентированный на обмен данными между OLAP-системами. При этом, создается пространство имен, в котором описывается некоторый формат данных и, возможно, метаданные в этом формате. Загрузку и выгрузку данных и метаданных берет на себя система, которая участвует в обмене данными, или разрабатывается специальная система, которая синхронизирует метаданные и обеспечивает автоматизированный обмен данными. Отличие от MDM систем заключается в том, что метаданные определяются пространством имен и от конкретной системы не зависят.

В Соединенных штатах Америки с 2005 года разрабатывается проект NIEM (National Information Exchange Model) – Национальная модель обмена информацией (Официальный сайт проекта NIEM). NIEM – это спонсируемая правительством США инициатива, направленная на облегчение обмена информацией между государственными и частными организациями. Первоначально акцент делался на правоохранительные органы, органы государственной безопасности и Агентство по чрезвычайным ситуациям, но область применения стандарта постоянно расширяется (Присцилла Уолмсли, 2010). Новые XML-инициативы в рамках департамента юстиции, департамента национальной безопасности и в других органах власти США предусматривают применение NIEM в качестве базовой модели данных и методологии для обеспечения взаимодействия данных и программного обеспечения, сокращения времени разработки приложений обмена информацией и многократного использования интеллектуального капитала и знаний в разных проектах.

Модель XML-данных NIEM обеспечивает блоки для описания общих объектов. Од-

нако сама модель NIEM не определяет всю информацию для обмена сообщениями. Она не содержит никаких конкретных типов сообщений или корневых элементов XML-документов. Чтобы использовать NIEM, необходимо построить документацию пакета обмена информацией NIEM, или сокращенно IEPD NIEM (NIEM Information Exchange Package Documentation). IEPD извлекает необходимые компоненты из моделей ядра NIEM и доменов, дополняет их для создания системы обмена информацией. IEPD содержит несколько артефактов: - XML-схемы, определяющие часть модели NIEM для данной системы, так называемая "схема подмножества"(subset schema);

- схема, определяющая корневой элемент системы обмена, или схема обмена (exchange schema);
- схема, определяющая расширения к модели NIEM - схема расширения (extension schema);
- документация по системе обмена, такая как диаграммы UML, примеры, описания.

Задача создателя IEPD – составить схему обмена, определив корневой элемент и базовую структуру сообщений (Присцилла Уолмсли, 2010). NIEM становится важнейшим стандартом обмена XML-информацией для госучреждений США и их информационных партнеров, хотя она не является готовой системой обмена информацией.

Если NIEM строит систему обмена данными, основываясь на готовой модели предметной области, что сильно упрощает создание такой системы, другой путь заключается в создании универсального формата данных, на основе которых создаются модели предметной области каждым субъектом самостоятельно. Тогда мы получим те же самые модели, но возникает проблема синхронизации этих моделей друг с другом. Для условий нашей страны второй путь более предпочтителен, так как у нас нет критической массы построенных моделей.

Промежуточным вариантом является инициатива SDMX (Statistical Data and Metadata eXchange), поддерживаемая Банком международных расчетов, Европейским Центральным Банком, Евростатом, Международным Валютным Фондом и иными организациями, включая статистические подразделения ООН (Официальный сайт формата обмена данными SDMX). Последняя версия стандарта была выпущена в апреле 2011 года. В 2013 году SDMX был утвержден ISO в качестве международного стандарта. Данный стандарт имеет спецификации для ряда предметных областей, таких как экономика, социальное обслуживание населения, демография и др. Использование данного стандарта предполагает централизованную архитектуру обмена статистическими данными, так как все статистические показатели и объекты закодированы специальным образом, и обмен данными происходит в утвержденных стандартом кодах. Пользователи системы обмена данными не могут добавить новые показатели или объекты самостоятельно, все должно происходить централизованно. Другим недостатком SDMX является его ориентация только на статистические данные.

Существуют множество других форматов обмена данными для автоматизации процесса обмена информацией для различных предметных областей. Приведем классификацию этих форматов согласно работе (Акаткин, 2014 : 35) в таблице 1.

Таблица 1. – Классификация моделей данных для обмена информацией.

Вид модели	Название модели
Государственное управление	Национальная модель обмена информацией NIEM. www.niem.gov
Управление воздушным движением	Модель обмена аэронавигационной информацией AIXM. www.aixm.aero
	Модель обмена метеорологической информацией WXXM. www.wxmx.aero
	Модель обмена информацией о полете FIXM. www.fixm.aero
Статистика	Открытый формат обмена статистическими данными SDMX. www.sdmx.org
Управление ЧС	Язык обмена данными о чрезвычайных ситуациях EDXL. www.oasis-open.org/committees
Сбор налогов	Стандарт автоматического обмена информацией финансового счета. www.oesd.org
Таможня	Модель данных всемирной торговой организации (ВТО). WCO DATA MODEL. www.wcoomd.org
Управление инфраструктурой ИКТ	Общая информационная модель CIM. www.xml.coverpages.org
Управление вооруженным противостоянием	Объединенная модель данных по обмену информацией для управления, контроля и консультирования НАТО (STANAG 5525) и США (JC3IEDM). www.lib.znate.ru
Подготовка государственных услуг к переводу в электронный вид	Модель информационной системы, реализующей перевод государственных услуг в электронный вид, путем организации непрерывного, связного и целостного процесса их подготовки и последующего применения полученных результатов (метаданных, моделей, процессов) для исполнения государственных услуг в автоматизированном режиме. www.kbpm.ru
Управление жизненным циклом изделия	Стандарт обмена данными модели изделия STEP. www.wikipedia.org/wiki/step

В приведенной таблице отсутствует протокол обмена финансовыми данными FIX (Financial Information eXchange) – протокол передачи данных, являющийся международным стандартом для обмена данными между участниками биржевых торгов в режиме реального времени. Протокол FIX изначально создан в 1992 г. для передачи информации о торгах акциями между компаниями Fidelity Investments и Salomon Brothers. В

настоящее время широко используется торговыми системами для обмена финансовыми данными и совершения транзакций (Официальный сайт протокола FIX). Протокол FIX поддерживается большинством крупнейших банков и электронными трейдинговыми системами, а также крупнейшими биржами мира. Отличительной чертой протокола является то, что в нем закодированы не только структура передаваемых данных, то и действия, которые получающая сторона должна выполнить.

Основными недостатками приведенных стандартов, как отмечалось выше, являются:

- обеспечение обмена только отраслевыми данными;
- поддержка только централизованной архитектуры обмена данными.

3.4 Обмен данными с помощью синхронизатора ТОФИ

Для решения проблемы обмена данными и метаданными различных форматов на основе стандарта с использованием различных архитектур организации процесса обмена данными и метаданными возникает необходимость в разработке нового единого формата обмена и информационной системы поддержки процесса обмена. Разрабатываемый формат должен удовлетворять следующим требованиям (Габбасов, 2015 : 12):

1. Для возможности передачи "больших" данных формат должен иметь компактный синтаксис;
2. Должна быть возможность передавать как все метаданные, так и части метаданных;
3. Формат должен обеспечивать многоязычную поддержку;
4. Инструментарий должен поддерживать механизм веб-сервисов и других технологических механизмов взаимодействия систем;
5. Структуры данных и метаданных должны обрабатываться одинаково, схема не должна накладывать дополнительные требования к обработке;
6. Формат должен обеспечивать распознавание органа, поддерживающего метаданные, то есть поддерживать описание ссылочных метаданных;
7. Формат должен позволять описывать свойства, не зависящие от времени;
8. Формат должен позволять описывать нечисловые свойства;
9. Формат должен поддерживать структуры, основанные на кубах.

Перечисленный набор требований является очень сложным и противоречивым для реализации. Разрабатываемая в Компании системных исследований "Фактор" концепция обмена данными основана на формализованной модели данных технологии ТОФИ, которая позволяет описывать произвольные данные в виде определенной структуры. Систему обмена данными предлагается сделать в виде системы обмена сообщениями, что совокупно с форматом данных ТОФИ позволяет выдерживать перечисленные выше требования.

Модель данных ТОФИ (Габбасов, 2013 : 14-16) дает более формальный абстрактный аппарат, для представления и обработки данных не привязываясь к определенной структуре хранения. В частности, модель данных ТОФИ может быть переведена в реляционную и другую структуру. Данные в технологии ТОФИ описываются формулой

$$(O, P, [s], [q], [T]) = V([T]), \quad (1)$$

где O (Owner) – владелец данных, P (Property) – свойство, s (Status) – статус данных, q (Author, Provider) – поставщик данных, T (Time) – период или актуальный интервал данных, V (Value) – значение данных. Необязательные элементы в формуле (1) заключены в квадратные скобки.

В базе данных информационно-аналитической системы хранятся "данные". Данные всегда являются значением некоторого свойства какого-либо объекта в некоторый период времени. Поэтому указание объекта (к которому относятся данные), свойства (определяющего тип данных), статуса и поставщика данных и временного интервала должно однозначно идентифицировать данные. В случае технологии ТОФИ в качестве объектов данных выступают объекты или отношения между объектами (элементы сущностей "Тип объектов" и "Отношения между типами объектов" соответственно). В качестве свойства выступают экземпляры сущности "свойство". Свойство характеризуется типом свойства. В технологии ТОФИ существуют 8 типов свойств: 1) свойство на основе факторов; 2) свойство на основе измерителей; 3) свойство на основе атрибутов; 4) свойство на основе типов объектов; 5) свойство на основе отношений между типами объектов; 6) свойство на основе единиц измерения; 7) комплексное свойство; 8) группа свойств. Данные являются значениями соответствующих свойств. Каждое свойство может принимать несколько значений для каждого фиксированного периода и фиксированного объекта, которые (значения) отличаются статусом и/или поставщиком данных.

Если убрать все необязательные элементы формула (1) выглядит как $(O, P) = V$, что выражает тот факт, что любые данные является значением некоторого свойства для некоторого объекта (владельца). Например, если – конкретный человек, – свойство "пол человека", то V равно одному из значений "мужской" или "женский". Заметим, что все свойства можно разделить на две группы: свойства, которые принимают значения в зависимости от конкретных периодов, и свойства, значения которых не зависят от регулярных периодов. В первом случае мы говорим, что значение свойства зависит от типа периода, а во втором случае – не зависит от типа периода. Если значение свойства зависит от типа периода, то оно меняется регулярно от одного периода к другому, например, ежемесячно или ежегодно. Если же значение свойства не зависит от типа периода, то значение свойства с течением времени может меняться, но это изменение не носит регулярный характер, а меняется случайно в любой момент времени. Следовательно, в этом случае у этого значения есть актуальный интервал жизни. Таким образом, в формуле (1) буквой T в левой части обозначен конкретный период (имеющий тип периода), а в правой части – актуальный интервал жизни значения данных (произвольный интервал времени, не привязанный к типу периода). Соответственно, если значение свойства P зависит от типа периода, то формула (1) записывается в виде $(O, P, [s], [q], T) = V$, то есть в этом случае T в левой части является обязательным атрибутом (совпадает с конкретным периодом), а в правой части отсутствует. Если же значение свойства P не зависит от типа периода, то формула (1) записывается в виде $(O, P, [s], [q]) = V()$, то

есть в левой части отсутствует, а в правой части является обязательным атрибутом (актуальный интервал значения данных). Момент времени есть интервал времени, у которого начало и конец совпадают.

Если убрать все необязательные элементы, то формула (1) записывается либо в виде $(O, P, T) = V$, либо в виде $(O, P) = V()$. Но при заданных владельце, свойстве и периоде значение данных может определяться неоднозначно. Скажем, валовый внутренний продукт (свойство) какой-либо страны (владелец) за определенный год (период) имеет несколько значений: оперативное значение и окончательное значение. Эти значения отличаются статусами данных. Статусами данных могут быть такие значения как плановое и фактическое, предварительное и окончательное, утвержденное и скорректированное и т.д. Статус данных в формуле (1) обозначен буквой s . Возможные значения статуса данных являются атрибутом свойства. В технологии ТОФИ в качестве статуса данных выступает значения факторов. Кроме статуса у значения данных могут быть поставщики данных. Поставщики данных – это субъекты, которые регистрируют или измеряют данные. У одного и того же данного могут быть несколько поставщиков. Поставщик данных в формуле (1) обозначен буквой q . Возможные поставщики данных являются атрибутом свойства. Атрибуты данных в левой части формулы (1), а именно владелец данных, свойство, статус, поставщик, период, называются **координатами данных**, а правая часть V называется **значением данных**. Значение данных может иметь различные типы, такие как число, строка, дата, фактор, объект, файл и т.д. Тип значения данных также определяется свойством. Для числовых данных у значения появляется еще одна координата, называемая единицей измерения. Тогда обобщенную формулу (1) можно написать в виде

$$(O, P, [m], [s], [q], [T]) = V([T]), \quad (2)$$

где m (Measure) – единица измерения. Как отметили выше, здесь в левой части – период, в правой части – актуальный интервал данных и обязательно присутствует в одной части формулы. Формула (1) или (2) определяет конкретное значение данных или конкретное данное. Для организации автоматического обмена данными необходимо, чтобы передающие и принимающие системы однозначно распознавали координаты данных, не теряя семантики отношений между координатами. Таким образом, задача автоматического обмена данными сводится к синхронизации координат данных между системами. Поэтому разрабатываемая система и называется "синхронизатор ТОФИ", так как главной задачей является синхронизация метаданных (координат данных) обменивающихся систем, основанная на модели данных ТОФИ. Синхронизация координат данных является сложной задачей, которая должна производиться человеком. Задачей синхронизатора ТОФИ является частичная автоматизация этого процесса. Для этого должны быть описаны структуры координат данных и системы их кодирования.

Каждая из перечисленных координат определена в базе данных в таблицах измерений. При этом каждая из них, как правило, является элементом некоторой более общей сущности, которую назовем "контейнером координаты". То есть, владелец данных, является элементом контейнера владельцев, свойство является элементом контейнера свойств, интервал времени (или период) является элементом контейнера интервалов времени (периодов). Причем один контейнер координат может быть вложен (являться элементом) в другой контейнер координат.

Мы предполагаем, что любая координата данных имеет одну из следующих структур, или, другими словами, является элементом одной из следующих структур:

1. Элемент;
2. Список;
3. Иерархия;
4. Куб.

Такие структуры присутствуют во всех информационных системах, так как объекты (владельцы данных), свойства, периоды и другие координаты являются, как описано выше, элементами более сложных структур в информационной модели. Совокупность всех элементов координат данных и их контейнеров (структур) назовем метаданными конкретной базы данных. Одной из главных функций синхронизатора является автоматизированный обмен и синхронизация метаданных. Таким образом, метаданные представляют собой описание различных структур координат данных и связей между ними. Например, владельцем данных обычно является некоторый объект, который сам является экземпляром некоторого класса объектов, при этом объекты в этом классе могут иметь структуру списка или иерархии. Тогда метаданные содержат описание класса с определенной структурой объектов и описание всех объектов этого класса.

В общем виде каждый экземпляр метаданных можно написать формально в виде $D_1(D_2(\dots D_k(E)\dots))$, где D_i – контейнеры координат вложенных друг в друга, – координата данных. Каждый контейнер координат и координата E имеет свою структуру и тип. В технологии ТОФИ контейнерами координаты "владелец данных" являются классы объектов, типы объектов, отношения между типами объектов, которые соответственно имеют структуру иерархия, иерархия и куб. Контейнерами координаты "свойство данных" являются свойства на основе измерителя и группы свойств, которые имеют структуру куб и список соответственно. Контейнерами координаты "статус данных" являются фактор, который имеет структуру список. Контейнерами координаты "поставщик данных" является класс объектов и тип объектов, которые имеют структуру список или иерархия. Контейнерами координаты "период" является тип периодов, который имеет структуру списка.

На основе концепции данных ТОФИ разработано пространство имен ТОФИ, в котором описаны формальные структуры данных (Габбасов, 2016 : 45). Пространство имен ТОФИ состоит из следующих XML-схем:

- base.xsd – хранит базовые классы;
- metadata.xsd – хранит классы для работы с метаданными;
- data.xsd – хранит классы для работы с данными;
- message.xsd – хранит классы для работы с сообщениями;

Каждый файл хранит в себе описание простых типов, глобальных элементов и составных типов. Синхронизатор ТОФИ является отдельным веб-приложением, который может быть интегрирован с платформой ТОФИ как ее модуль.

Функционально синхронизатор осуществляет сопоставление метаданных различных информационных систем на основе систем кодирования, автоматически (по расписанию) или вручную производит формирование метаданных, обмен метаданными с другими системами, синхронизацию метаданных, формирование данных для передачи другой системе, прием и передачу набора данных из/в другой(-ую) информационной(-ую) системы(-у) и т.д.

4 Результаты и обсуждения

Обмен данными с использованием синхронизатора ТОФИ основан на синхронизации метаданных обменивающихся систем. Если другие системы обмена данными, основанные на пространствах имен являются предметно-ориентированными, то пространство имен ТОФИ, в этом смысле, является универсальным. Но сложность внедрения предлагаемого подхода заключается в необходимости кодировать каждый экземпляр сущности информационной системы. Эффект от внедрения пространства имен ТОФИ может быть получен в случае принятия его в качестве стандарта в отрасли или государстве.

5 Заключение

В статье приведены обзор существующих методов обмена данными между информационными системами. Подробно описана технология обмена данными и метаданными с использованием синхронизатора ТОФИ. Сравнение его возможностей с другими технологиями позволяет сделать вывод о перспективности разрабатываемого подхода.

Работа выполнена при поддержке грантового финансирования научно-технических программ и проектов Комитетом науки МОН РК, грант № 4502/ГФ4.

Список литературы

- [1] ISO/IEC 2382-1:1993, Information technology. – Vocabulary. – Part 1: Fundamental terms. – 1993.
- [2] ISO/IEC 2382:2015 Information technology. – Vocabulary. – 2015.
- [3] *Codd, Edgar F.* Providing OLAP to User-Analysts: An IT Mandate // *Computerworld*. – 1993. – V. 27, No 30. – 26 p.
- [4] *Codd Edgar F.* A relational model of data for large shared data banks. *Comm. ACM*. – 1970. V. 13, No 6. – Pp. 377-387.
- [5] *Черняк Л.* Задачи управления мастер-данными // *Открытые системы*. – 2007. № 5. – С. 54-62.
- [6] Официальный сайт проекта NIEM [Электронный ресурс]. – 2006. – URL: www.niem.gov.
- [7] *Присцилла Уоллсли.* Создание IEPD NIEM. Моделирование системы обмена информацией NIEM [Электронный ресурс]. – 2010. – URL: www.ibm.com.
- [8] Официальный сайт формата обмена данными SDMX [Электронный ресурс]. – 2006. – URL: www.sdmx.org.
- [9] *Акаткин Ю., Дрожжин В., Коляевский В.* Стандарты моделей данных для обмена информацией как инструмент импортозамещения в стратегических информационных системах // *PC Week*. – 2014. – № 16 (871). – С. 26-52.
- [10] Официальный сайт протокола FIX [Электронный ресурс]. – 2006. – URL: <http://www.fixtradingcommunity.org>.
- [11] Разработка синхронизатора ТОФИ для обмена данными и метаданными между информационными системами: отчет о НИР (промежуточный) // АО "Нац. центр научно-техн. информ.": рук. Габбасов М.Б.; исполн: Куанов Т.Д., Абилкаева Ж.Н. и др. – Астана, 2015. – 88 стр. – № ГР 0115РК02743/

- [12] Габбасов М.Б. Онтология технологии ТОФИ // Сборник трудов по информационной безопасности. – Астана. – 2013. – С. 53-63.
- [13] Разработка синхронизатора ТОФИ для обмена данными и метаданными между информационными системами: отчет о НИР (промежуточный) // АО "Нац. центр научно-техн. информ.": рук. Габбасов М.Б.; исполн: Куанов Т.Д., Абилкаева Ж.Н. и др. – Астана, 2016. – 125 стр. – № ГР 0115РК02743.

References

- [1] ISO/IEC 2382-1:1993, Information technology. Vocabulary. Part 1: Fundamental terms (1993).
- [2] ISO/IEC 2382:2015 Information technology. Vocabulary (2015).
- [3] Codd, Edgar F. Providing OLAP to User-Analysts: An IT Mandate // Computerworld. – V. 27, No 30 (1993) : 26.
- [4] Codd Edgar F. «A relational model of data for large shared data banks», *Comm. ACM*, V. 13, No 6 (1970) : 377-387.
- [5] Chernyak L. Zadachi upravleniya master-dannyimi [Master data management tasks], *Otkryitiye sistemyi*, No 5 (2007) : 54-62.
- [6] Ofitsialnyiy sayt proekta NIEM [Official website of the project NIEM. Electronic resource], 2006. <http://www.niem.gov>.
- [7] Pristsilla Uolmsli. Sozdanie IEPD NIEM. Modelirovanie sistemyi obmena informatsiey NIEM [Create IEPD NIEM. Modeling the information exchange system NIEM. Electronic resource], 2010. <http://www.ibm.com>.
- [8] Ofitsialnyiy sayt formata obmena dannyimi SDMX [Official website of the Data Exchange Format SDMX. Electronic resource], 2006. <http://www.sdmx.org>.
- [9] Akatkin Yu., Drozhzhinov V., Konyavskiy V. «Standartyi modeley danyih dlya obmena informatsiey kak instrument importozamescheniya v strategicheskikh informatsionnykh sistemah [Standards of data models for information exchange as an instrument for import substitution in strategic information systems]», No 16 (871), (2014) : 26-52.
- [10] Ofitsialnyiy sayt protokola FIX [Official website of the protocol FIX. Electronic resource], 2006. <http://www.fixtradingcommunity.org>.
- [11] «Razrabotka sinhronizatora TOFI dlya obmena dannyimi i metadannyimi mezhdru informatsionnyimi sistemami: otchet o NIR (promezhutochnyy) [Development of the TOFI synchronizer for data and metadata exchange between information systems: Report on research work (interjacent)]», АО "Nats. tsentr nauchno-tehn. inform.": рук. Gabbasov M.B.; ispoln: Kuanov T.D., Abilkaeva Zh.N. i dr. Astana, No GR 0115RK02743 (2015).
- [12] Gabbasov M.B. «Ontologiya tehnologii TOFI [Collection of works on information security]», *Sbornik trudov po informatsionnoy bezopasnosti*. Astana, (2013).
- [13] «Razrabotka sinhronizatora TOFI dlya obmena dannyimi i metadannyimi mezhdru informatsionnyimi sistemami: otchet o NIR (promezhutochnyy) [Development of the TOFI synchronizer for data and metadata exchange between information systems: Report on research work (interjacent)]», АО "Nats. tsentr nauchno-tehn. inform.": рук. Gabbasov M.B.; ispoln: Kuanov T.D., Abilkaeva Zh.N. i dr. Astana, No GR 0115RK02743 (2016).