

Липунцов Ю.П.

Московский государственный университет имени М.В. Ломоносова, Москва, Россия

ПРИМЕНЕНИЕ ИНФОРМАЦИОННОЙ МОДЕЛИ МЕЖВЕДОМСТВЕННОГО ВЗАИМОДЕЙСТВИЯ ДЛЯ СБОРА И АНАЛИЗА ДАННЫХ ПО ФОНДОВОМУ РЫНКУ

АННОТАЦИЯ

В статье будет изложен опыт применения информационной модели межведомственного взаимодействия на основе распределенного хранилища данных, описанной в [5]. Модель была предложена при разработке Системного проекта электронного правительства РФ в 2015 г [4] и представляет собой организацию хранилища данных, реализованного в форме трехслойного ядра данных. В данном случае описывается опыт использования этой модели для сбора и анализа информации по фондовому рынку. Анализ данных по фондовому рынку предполагает наличие информации из разных источников: налоговых органов, органов статистики, национального расчетного депозитария, бирж и других источников. Для сбора данных из разных источников необходимы методика трансформации данных и приведения их к сопоставимому виду. Эти этапы выполняются на основе методики реализации модели. Описанная в статье модель обладает характеристиками гибкости, расширяемости, а также возможностью распределением ответственности и обеспечения достаточного уровня безопасности данных.

КЛЮЧЕВЫЕ СЛОВА

Хранилище данных, онтология предметной области, распределенная среда, модель предметной области, информационное моделирование.

Lipuntsov Yu.P.

APPLICATION OF INFORMATION MODEL OF INTERAGENCY DATA-EXCHANGE FOR THE AGGREGATION AND ANALYSIS OF STOCK MARKET DATA

ANNOTATION

This article will discuss the application of information model of interagency data-exchange on the basis of the distributed data warehouse, described in [5]. The model has been proposed in the E-government System project of the Russian Federation in 2015 [4]. The basis of this model is the data warehouse in the form of a three-layer core data. In this case, it describes the experience of using this model to collect and analyze information on the stock market. Analysis of stock market data suggests the availability of information from different sources: the tax authorities, statistical offices, national settlement depository, stock exchanges and other sources. To collect data from various sources are necessary method of transforming data and bring them to a compatible form. The described model has the characteristics of flexibility, extensibility, and the ability to distribute responsibilities and ensure a sufficient level of data security.

KEYWORDS

Data warehouse, domain ontology, distributed environment, a domain model, information modeling.

Введение

Во многих текстах, в том числе правительственных документах, часто поднимается тема совместного использования данных, более активное использование накопленных информационных активов. Несмотря на то, что финансовый сектор является наиболее развитым в области информационных технологий, большинство используемых систем, в которых накоплены огромные объемы данных, ориентированы на решение локальных задач, не предполагающих интеграцию с другими системами. Проблема совместного использования данных, становится более сложной с распространением мобильных устройств, новых типов данных, а также постоянного накопления и обновления данных.

В течение последних 5 лет на экономическом факультете МГУ в курсах по анализу фондового рынка используются данные собранные с помощью модели интеграции, построенные по принципу модели межведомственного взаимодействия на основе распределенного хранилища данных. Реализованная модель сбора и анализа данных фондового рынка предполагает предоставление конечному пользователю сервисов для выполнения анализа данных по собственной методике. Большинство сервисов поставки информации по фондовому рынку, такие как СПАРК, РБК Quote, Integrum, предоставляет готовые коэффициенты и мультипликаторы. При этом методика их расчета остается за кадром, а для того чтобы реализовать собственную методику расчета аналитических показателей многим аналитикам приходится поддерживать в актуальном состоянии собственные базы данных, решая на локальном уровне проблему несопоставимости данных, поступающих из разных источников. Решение проблемы приведения данных к сопоставимому виду позволит предоставлять качественные данные в формате «данные как сервис». На базе этого сервиса возможно создание компонент, которые позволят конечному пользователю разрабатывать и реализовывать собственные модели анализа данных. Прототипы таких решений отработаны на курсах «Анализ российского и международного рынка акций» читаемых в магистратуре экономического факультета МГУ. Получив основы работы с базой данных на первом занятии, студенты строят рейтинги компаний, формируют портфели на основе биржевой информации порядка 400 эмитентов, данных их финансовой отчетности в динамике, сопоставления с аналогами с зарубежных рынков. Работая с реальными данными, анализируя представительную информацию по эмитентам, студенты получают практические навыки и новое понимание теоретических основ, которые они получили на предшествующих курсах.

Основой для предоставления такого сервиса является информационная модель, которая определяет основные принципы организации обмена данными между системами. Информационная модель должна обеспечивать поставку актуальных данных о разных объектах с необходимым набором атрибутов, состав которых предопределяются методами анализа.

Представление информационной модели межведомственного взаимодействия

В Системном проекте электронного правительства РФ в разделе, посвященном использованию государственных данных и базовых государственных информационных ресурсов в электронном правительстве изложена модель, описывающая подходы и принципы поддержания расширяемого ядра общих сведений и государственных данных. Модель содержит несколько разделов, которые отличаются по характеру принадлежности к отдельным предметным областям. Ядро данных представляет собой трехслойную структуру ядра данных: общее ядро, ядро предметной области и предметно-ориентированная часть ядра. Взаимодействие между слоями осуществляется с помощью модели межуровневых связей.

Информационное взаимодействие, сводится к обмену данными, отражающих состояние базовых информационных объектов. Хранилище актуальных данных выступает в данном случае как блок данных, агрегирующий набор сведений, необходимых для обмена. Модель позволяет сопоставлять сведения об аналогичных объектах, поступающих из разных систем, отслеживать динамику их состояний. Модель данных снабжается описанием семантических связей между объектами, что позволит реализовывать семантику с использованием контекста. На базе модели возможно организовать витрину данных.

Общее ядро данных включает универсальную часть данных, которые задействуются во многих предметных областях и понимаются ими одинаково. В общее ядро данных включены данные и физических лиц, юридических лиц, описание местоположения, объекты недвижимости. Модель позволяет организовывать хранилище данных, загружаемых из разных систем и получать выборку данных по разным срезам: гражданство, место жительства, регистрации, даты актуализации и т.д.

Универсальное ядро является основой для информационного обмена предметной области. Поступающие в ядро данные проходят через обработку, в результате которой базовые сущности кодируются в соответствии онтологией предметной области. Кодификация производится таким образом, что все идентичные объекты, попадающие в систему из разных источников, получают один код. Такая система кодирования позволит просматривать отражение определенного объекта в разных системах. При загрузке в систему в записи фиксируется источник данных, время загрузки и время окончания актуальности данных.

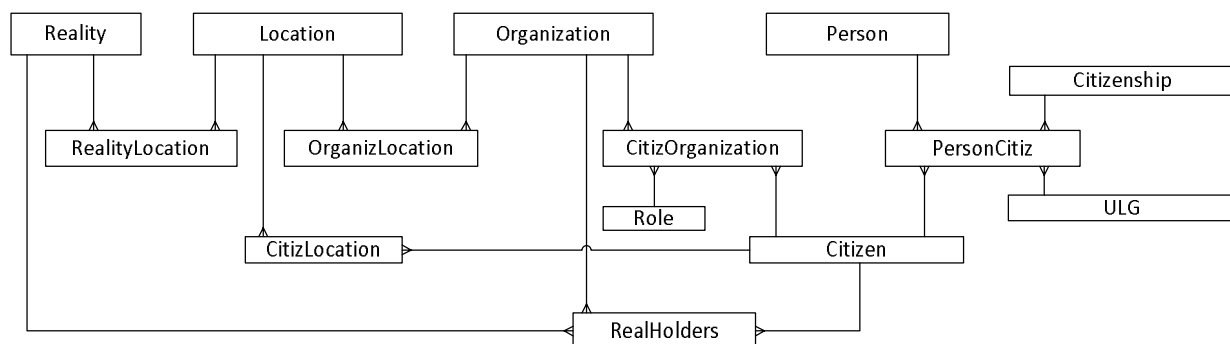


Рисунок 1. Вариант состава кодификаторов общего ядра данных

Расширением общего ядра являются *ядро предметной области*. На уровне модели предметной области описываются принципы организации деятельности в этом секторе экономики. На фондовом рынке описываются принципы регулирования, в частности процедуры выдачи лицензий, мониторинг деятельности, выполнение регламентов и прочите элементы которые позволяют на управленческом уровне описать деятельность этого сектора.

Модель предметной области включает три категории элементов: **Базовые сущности** – информационные объекты предметной области, связи между ними. **Графическое представление предметной области** – иерархическое представление объектов предметной области, которое позволяет выстроить соотношение объектов в виде модели реального мира. Дополнительными информационными ресурсами являются Классификаторы, Справочники, а также Межпредметные списки кодов.

Концептуальное описание предметной области Фондовый рынок представлено на рис.2

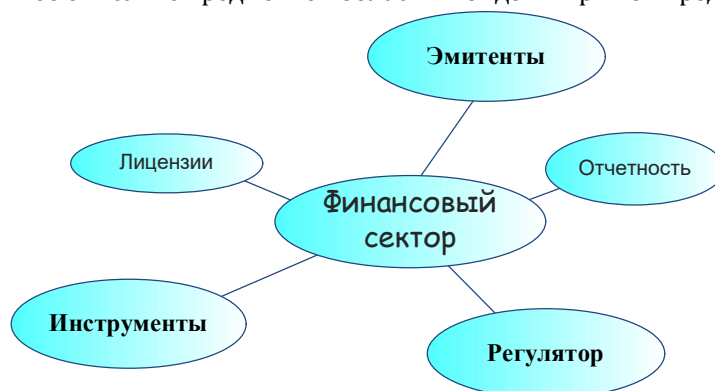


Рисунок 2. Основные сущности описания предметной области Финансовый сектор

Предметно-ориентированное ядро предполагает детальное описание. В предметной области фондовый рынок представлены такие данные как профиль эмитента, состав его финансовых инструментов, площадки на которых осуществляется торговля инструментами, данные финансовой отчетности, данные с торгов по инструментам, различные события эмитентов, такие как выплата дивидендов, и прочая важная для анализа информация.

Модель организации хранения данных предметной ориентированной части ядра, раздел Фондовый рынок

Модель включает графическое представление предметной области и модель интеграции данных.

Графическое представление предметной области

Предметна область будет представлена на концептуальном и логическом уровне. Концептуальное представление отражается в форме описания предметной области посредством концептов, их определений и отношений между ними. Поэтому сначала введем ряд концептов, которые в дальнейшем будут использованы для построения модели.

Компания — юридическое лицо; форма организации предприятия, при которой его правосубъектность отличается от правосубъектности принимающих в нём участие лиц.

Финансовый инструмент — это договор, в результате которых возникают финансовые активы у одной стороны и финансовые обязательства или долевыe инструменты у другой стороны.

Биржа — юридическое лицо, обеспечивающее регулярное функционирование организованного рынка финансовых инструментов.

Финансовая отчётность - система данных о финансовом положении компании, финансовых результатах её деятельности и изменениях в её финансовом положении; составляется на основе данных бухгалтерского учёта

Торги - конкурентная форма отбора предложений на продажу финансовых инструментов по заранее объявленным условиям, на принципах состязательности, справедливости и эффективности.

На основе концептуального представления можно построить графическое представление предметной области, которое послужит основой для составления модели данных хранилища данных (рис.3).

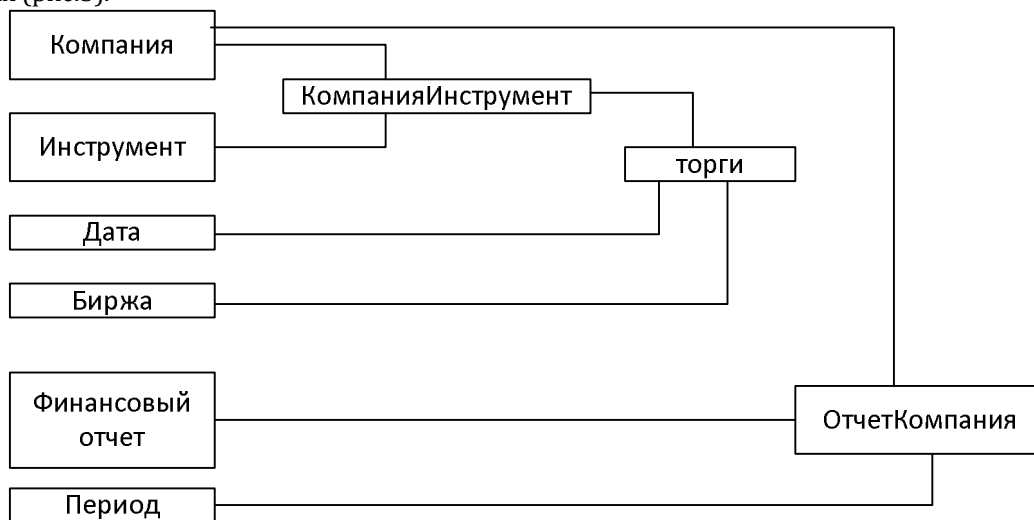


Рисунок 3 Графическое представление предметной области Фондовый рынок

На основе графического представления предметной области строится концептуальная модель фондового рынка и логическая модель данных хранилища данных. Категории сущностей модели интеграции, состав их атрибутов описаны в статье автора «Принципы создания онтологии для хранилища данных. Пример «высшее образование» [3]. Концептуальная модель фондового рынка отражена на рис 4., а логическая модель данных приведена на рис. 5.



Рисунок 4. Концептуальная модель предметно-ориентированной части Фондовый рынок

Модель включает следующие сущности: Профиль компании (Company), состав отчетности (Report), типы инструментов (Instrument), площадки (Exchange) и владельцы (Holder).

Разные категории данных поступают от разных участников информационного обмена: данные о компаниях от Министерства налогов и сборов, данные регистрации выпусков ценных бумаг от Центрального банка, финансовая отчетность от статистических ведомств, результаты биржевых торгов от бирж, состав собственников компании поставляет национальный депозитарий (рис 4.).

На рисунке 4 приведена модель данных с целевой схемой. Прежде чем загружать данные в целевую БД, данные нужно преобразовать. Преобразование данных включает следующая последовательность действий: Данные из источников загружаются в БД Extract, там организуется хранение всех доступных из внешних источников данных. Таблицы могут быть не связаны между

собой. Далее производится преобразование этих данных. Результаты преобразования записываются в БД Transform. В этой базе данных производится кодирование данных в соответствии с принятой системой кодировки. Здесь данные тоже могут быть не связаны между собой. Затем эти данные загружаются в БД Load с целевой схемой данных. Там все таблицы связаны и отвечают требованиям целостности.

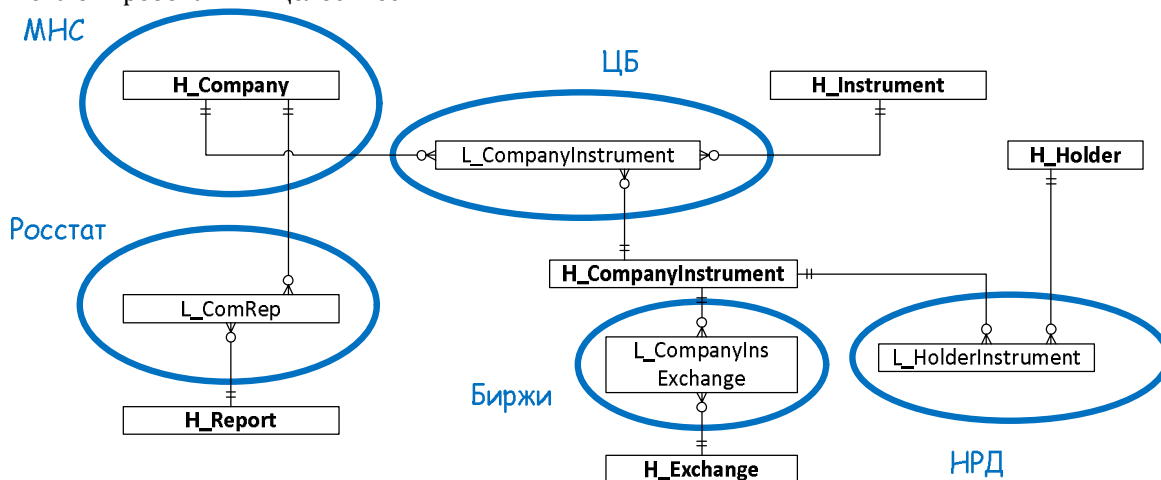


Рисунок 5. Модель интеграции данных предметной области Фондовый рынок

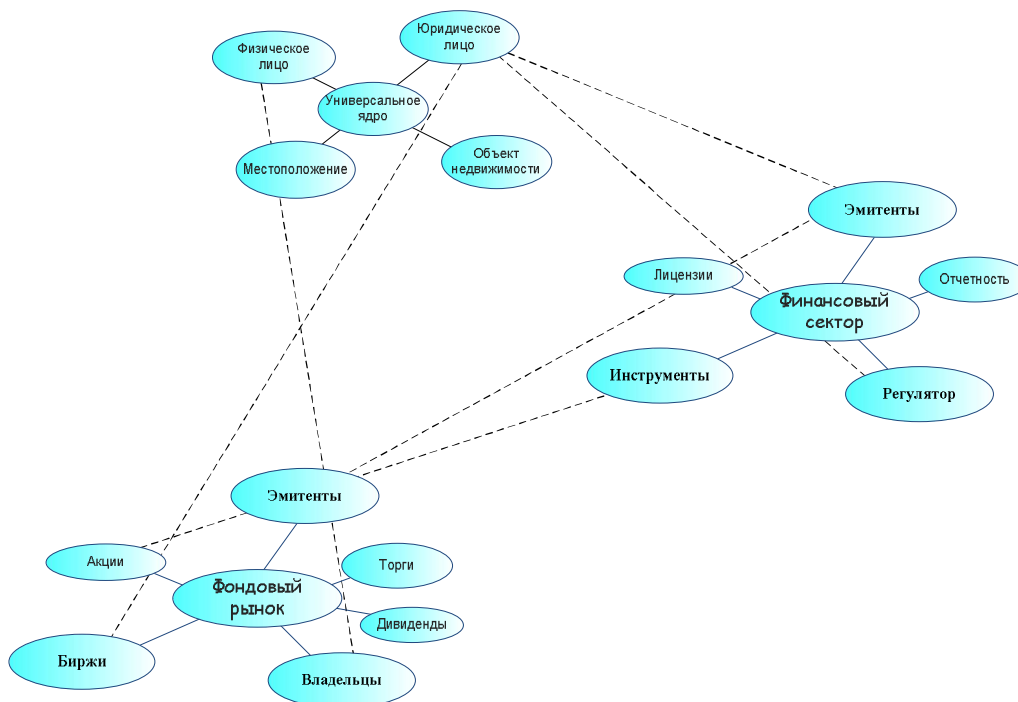


Рисунок 6. Отражение межуровневых взаимодействий между слоями ядра данных

Межуровневые взаимодействия

Важным элементом организации поставки данных из источников разного уровня ядра является согласование между отдельными моделями, как на вертикальном, так и на горизонтальном уровне. Для организации связей создается модель межуровневых взаимодействий, которая обеспечивает согласование представлений объектов в разных моделях. Вариантами согласования может быть создание механизма использования единого кодификатора объектов, либо установление соответствий между представлениями объекта в разных моделях. На рис.6 приведен вариант согласования между универсальным ядром, ядром предметной области и предметно-ориентированной частью описания.

С использованием представленной модели появляется возможность организовывать инфраструктурные элементы работы с данными. В частности, для проведения анализа по фондовому рынку необходимы данные финансовой отчетности, которые отражаются на втором

уровне. При проведении анализа по фондовому рынку важными являются натуральные показатели, такие как количество абонентов телекоммуникационных компаний, объемы выпуска автомобилей по предприятиям автомобилестроения, результаты аудита запасов нефти по нефтяным компаниям. Такие данные можно получить из предметно ориентированной части других предметных областей.

Заключение

Представленная в статье хранилище данных является для анализа фондового рынка позволяет реализовывать различные модели оценки привлекательности финансовых инструментов, включающие мультипликаторы анализа баланса, мультипликаторы анализа отчета о прибылях и убытках, мультипликаторы денежного потока, финансовые коэффициенты, и прочие показатели. Разработка и реализация собственной аналитической модели становится возможной благодаря информационной модели, позволяющей агрегировать данные из разных источников, приводить их к сопоставимому виду и выдавать конечному пользователю данные в формате «Данные как сервис». Модель является расширяемой и предполагает применение в различных предметных областях.

Литература

1. Inmon W.H. и Linstedt D. Data Architecture: A Primer for the Data Scientist, [Книга]. - [б.м.] : Morgan Kaufmann, 2014.
2. Lipuntsov Yuri P.. 2015. On the Relationship Between the Information and Analytical Components in the Shared E-Government. In Proceedings of the 2015 2nd International Conference on Electronic Governance and Open Society: Challenges in Eurasia (EGOSE '15). ACM, New York, NY, USA, 109-115. DOI=<http://dx.doi.org/10.1145/2846012.2846026>
3. Липунцов Ю. П. Принципы создания онтологии для хранилища данных. Пример: «высшее образование» // Бизнес. Образование. Право. 2016. № 3 (36). С. 72–79.
4. Системный проект электронного правительства Российской Федерации minsvyaz.ru/uploaded/files/referat-ep.pdf
5. Lipuntsov Yuri An Information model of Interagency Communication Based on Distributed Data Storage In Proceedings of the 2016 3rd International Conference on Electronic Governance and Open Society: Challenges in Eurasia (EGOSE '16). ACM, New York, NY, USA.

Reference

1. Inmon W.H. and Linstedt D. Data Architecture: A Primer for the Data Scientist, : Morgan Kaufmann, 2014.
2. Lipuntsov Yuri P. .. 2015. On the Relationship Between the Information and Analytical Components in the Shared E-Government. In Proceedings of the 2015 2nd International Conference on Electronic Governance and Open Society: Challenges in Eurasia (EGOSE '15). ACM, New York, NY, USA, 109-115. DOI = <http://dx.doi.org/10.1145/2846012.2846026>
3. Lipuntsov YP Principles for creating ontologies for data warehouse. Example: "Higher Education" // Business. Education. Right. 2016. number 3 (36). S. 72-79.
4. System of the Russian Federation, E-government System project URL minsvyaz.ru/uploaded/files/referat-ep.pdf Last seen [2016.06.19]
5. Lipuntsov Yuri An Information model of Interagency Communication Based on Distributed Data Storage In Proceedings of the 2016 3rd International Conference on Electronic Governance and Open Society: Challenges in Eurasia (EGOSE '16). ACM, New York, NY, USA.

Поступила: 10.10.2016

Об авторах:

Липунцов Юрий Павлович, доцент кафедры экономической информатики, экономический факультет МГУ имени М.В. Ломоносова, кандидат экономических наук.