

УДК 519.7:007.52

В.В. Завгородний, С.С. Щербак

Днепродзержинский государственный технический университет, Днепродзержинск

ЕДИНОЕ ИНФОРМАЦИОННОЕ ПРОСТРАНСТВО ПРОИЗВОДСТВЕННЫХ ПРЕДПРИЯТИЙ НА ОСНОВЕ СВЯЗАННЫХ ДАННЫХ

В статье разработана математическая модель единого информационного пространства производственного предприятия на основе связанных данных, получила дальнейшее развитие математическая модель распределенной информационной системы производственного предприятия, которая обеспечивает эффективное выполнение задач оперативного управления, а также предложена модель взаимодействия пользователя с информационным пространством.

Ключевые слова: *Linked Data, RDF, SPARQL, Semantic Web, Virtuoso, паттерны интеграции.*

Введение

Современное состояние рыночной экономики требует от производственных предприятий использования эффективных средств управления как производственными, так и бизнес подразделениями, что подразумевает развитие информационной инфраструктуры предприятия в сторону увеличения связности составляющих ее компонентов. Такими компонентами могут выступать информационные системы (ИС) подразделений с различным уровнем межсистемной и внутрисистемной компонентной интегрированности [1].

Сегодня информационное обеспечение производственных предприятий на постсоветском пространстве часто строится на основе стихийной архитектуры, что приводит к низкому уровню интегрированности составляющих информационных систем [2]. Кроме того, зачастую на производственных предприятиях подразделения территориально распределены, что требует использования эффективных средств коммуникации.

В последнее время в качестве эффективного средства коммуникации с низкой стоимостью развертывания используется сеть Интернет и гипертекстовые технологии Всемирной Паутины (WWW) [3]. К сожалению, документо-ориентированная природа WWW и отсутствие эффективных средств описания содержащихся в документах данных привела к развитию новых направлений и концепций, которые могут стать эффективной платформой для информационных систем производственных предприятий. Одной из таких является концепция связанных данных (англ. Linked Data). В рамках этой концепции каждому объекту реального мира или объекту предметной области, автоматизация которой производится, ставится в соответствие его виртуальный аналог, а объектные связи выражаются как набор отношений между объектами, каждый из которых связан с другими с помощью универсального лока-

тора ресурсов (англ. URL). В рамках такого подхода к моделированию предметных областей, данные представляются как открытая структура данных, модификация которой может быть осуществлена без существенного изменения программных средств их обработки, и доступ к которой может быть при необходимости ограничен. Кроме того, концепция связанных данных обеспечивает эффективную платформу для создания распределенных информационных систем, основанную на технологиях искусственного интеллекта и Семантического Веба (англ. Semantic Web), при чем ориентация этой концепции на распределенное хранение данных является сегодня одним из лучших решений для организации как межсистемного, так и внутрисистемного взаимодействия распределенных информационных систем, что в условиях территориальной распределенности подразделений предприятий является одним из лучших решений [4].

При разработке ИС на основе связанных данных, в частности на основе RDF (англ. Resource Data Framework), некоторые трудности обусловлены тем, что связанные данные предназначены для распределенного хранения и описания данных и их схем, но не предоставляют возможности описания методов обработки данных (операционного аспекта) [5], что в условиях применения на промышленных предприятиях требует создания новых моделей для учета операционной составляющей процессов, протекающих на предприятиях.

Производственные предприятия в условиях постоянного повышения скорости ведения бизнеса и недостаточного финансирования нуждаются в средствах, позволяющих повышать эффективность оперативного управления производственным предприятием на основе развитых технологий поддержки принятия решений, поэтому актуальным и целесообразным является создание информационного пространства производственного предприятия на основе связанных данных, которое могло бы повысить

связность компонент ИС производственного предприятия при сохранении их территориальной распределенности и с возможностями модификации и добавления новой функциональности без перепрограммирования модулей обработки данных. Одной из таких модификаций является необходимость постоянного повышения эффективности процессов поиска распределенных между компонентами ИС данных в связи с постоянным лавинообразным ростом количества данных, хранимых на производственных серверах, и требованием хранения разноаспектных характеристик некоторых видов данных, применяемых, например, при принятии решений уполномоченными лицами [6].

Информационное обеспечение производственных предприятий, построенное на основе стихийной архитектуры, характеризуется использованием ИС контрольно-учетного характера, отсутствием единой формальной системы процесса выработки управленческих решений, отсутствием непосредственного доступа лиц, принимающих решения, к информационным ресурсам и т.д., что для эффективного решения задач, как поддержки принятия решений, так и задач оперативного управления в целом.

Целью данной работы является разработка математической модели информационного пространства производственного предприятия на основе связанных данных и построение на ее основе архитектуры распределенной информационной системы, функциональность которой может быть увеличена путем добавления новых модулей и схем данных путем конфигурирования управляющих компонентов.

Математическая модель единого информационного пространства на основе связанных данных

Связанные данные, как элементы концепции LinkedData, математически могут быть представлены формулами (1) – (4) [7].

Основным элементом связанных данных является четырехкомпонентная структура квад (англ. Quad), определенная с помощью следующей формулы:

$$t = \langle g, s, p, o \rangle, \quad (1)$$

где t – квад; g – поименованный граф; s – субъект; p – предикат; o – объект, причем каждый компонент связанных данных с учетом распределенности их хранения должен быть представлен в виде URL (англ. Uniform Resource Locator).

Совокупность структур t , определенных формулой (1), будем считать хранилищем связанных данных:

$$T = \{t_i\}, i = \overline{(1, n)}, \quad (2)$$

где t_i – i -й квад; n – конечное число квадов в хранилище.

Квады в хранилищах группируются в контексты, причем в рамках одного контекста все квады t_i относятся к одному поименованному графу g .

Множество контекстов можно выразить с помощью следующей формулы:

$$G = \{g_j\}, j = \overline{(1, m)}, \quad (3)$$

где G – множество контекстов хранилища; g_j – j -контекст хранилища; m – количество контекстов хранилища.

С учетом формулы (3) каждому контексту согласно формуле (1) поставим в соответствие трехэлементный набор $\langle s, p, o \rangle$. Таким образом, контекст определим с помощью следующей формулы:

$$\forall g \in G : g = \langle S, P, O \rangle, \quad (4)$$

где g – контекст хранилища данных; G – множество всех контекстов; S – множество субъектов; P – множество предикатов; O – множество объектов.

Далее, если это не будет приводить к противоречию, «контекст», «поименованный граф», будем использовать как синонимы.

Производственные предприятия, построенные с использованием стихийной архитектуры, состоят из множества распределенных источников данных, каждому из которых соответствует некоторая информационная система подразделения [6], поэтому будем рассматривать информационное пространство как некий промежуточный интеграционный компонент с универсальным коммуникационным интерфейсом (рис. 1).

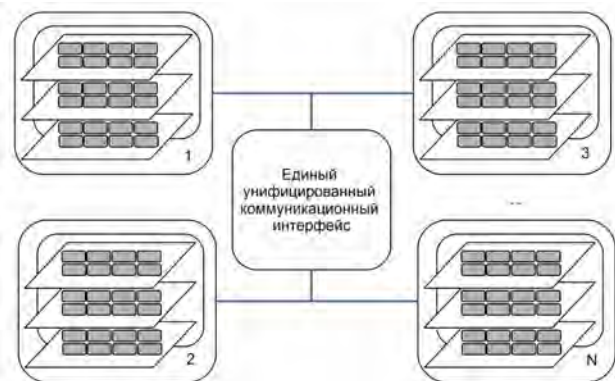


Рис. 1. Информационное пространство производственного предприятия из N -источников данных

В рамках такого подхода, пусть I – информационное пространство производственного предприятия, тогда формально его выразим как конечное множество хранилищ связанных данных G , определенных с помощью формулы (2), следующим образом:

$$I = \{G_k\}, k = \overline{(1, z)}, \quad (5)$$

где G_k – k -е хранилище связанных данных; z – общее число хранилищ связанных данных производственного предприятия.

Пусть G – множество всех контекстов. Разобьем множество всех контекстов G на два подмножества G^S и G^I , причем так что $G^S \cap G^I = \emptyset$, тогда множество триплетов, принадлежащих G^S , будут определять схему связанных данных, а G^I – собственно сами связанные данные.

Структурно-логическая схема хранилища связанных данных

Пусть $G_{t_i}^S$ – структурно-логическая схема t_i -го хранилища связанных данных, тогда определим ее как подмножество $G_{t_i}^S$ контекстов G хранилища связанных данных T , согласно следующей формулы:

$$G_{t_i}^S \subset G. \tag{6}$$

Определим структурно-логическую схему связанных данных $G_{t_i}^S$ (рис. 2), выраженных формулой (4), в терминах объектно-ориентированного проектирования (класс, подкласс, экземпляр класса, отношение между классами):

$$G_{t_i}^S = \langle C, R, P \rangle, \tag{7}$$

где $G_{t_i}^S$ – схема связанных данных; C – множество классов, причем $C \subset G_{t_i}^S$; R – множество отношений между классами ($R \subset G_{t_i}^S$); P – множество свойств классов ($P \subset G_{t_i}^S$).

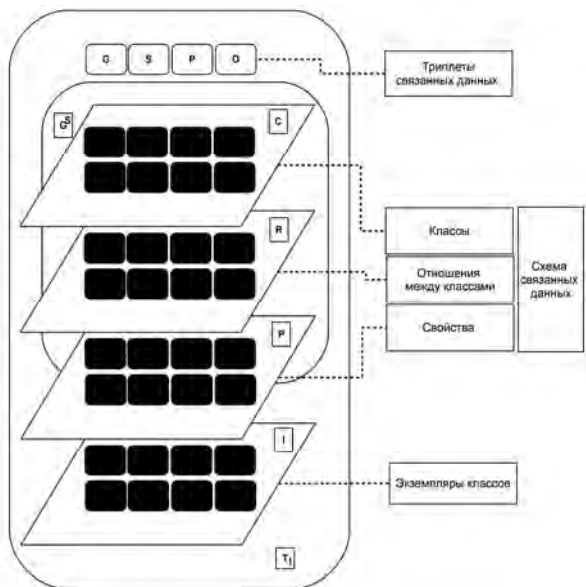


Рис. 2. Структура хранилища связанных данных

Определим множество доступных отношений между классами R для организации связей «быть подклассом» и «быть экземпляром» следующим образом:

$$R = \{is_subclass, instance_of\}, \tag{8}$$

где $is_subclass$ – отношение «быть подклассом»; $instance_of$ – отношение «быть экземпляром класса».

С учетом формулы (8) представим множество отношений R как объединение непересекающихся подмножеств R^S и R^I ($R^S \cap R^I = \emptyset$), логическая структура которых задана формулами (10) и (11):

$$R = R^S \cup R^I. \tag{9}$$

Логическая структура R^S :

$$R^S = \{ \langle c'_1 \in C : c' \text{ subclass } c \rangle, \dots, \langle c'_n \in C : c' \text{ subclass } c \rangle \}, \tag{10}$$

где c, c' – классы; C – множество классов.

Логическая структура R^I :

$$R^I = \{ \langle c'_1 \in C : c' \text{ instanceof } c \rangle, \dots, \langle c'_n \in C : c' \text{ instanceof } c \rangle \}, \tag{11}$$

где c, c' – классы; C – множество классов.

Определим множество свойств классов

$$P = \langle T, L, V \rangle, \tag{12}$$

где T – тип данных, L – язык представления, V – значения свойства.

Учитывая, что язык представления L необходим при определении только значений символьного или строкового типов свойств экземпляров класса, то для всех остальных установим нулевое значение L .

$$P = \langle T, \emptyset, V \rangle. \tag{13}$$

Для определения структуры классов значение свойств не имеют смысла, поэтому $V = \emptyset$, а выражение (13) будет иметь вид:

$$P = \langle T, \emptyset, \emptyset \rangle. \tag{14}$$

Таким образом, математическую модель информационного пространства на основе связанных данных представим формулами (1) – (14).

Модель производственного предприятия на основе единого информационного пространства

Задачи ИС тесно связаны с организацией обработки данных, циркулирующих через ее встроенные компоненты, так и получаемые с внешней среды, поэтому понятие объекта информационного пространства расширим множеством присоединенных процедур, реализующих операционную составляющую ИС. Таким образом, с учетом вышесказанного модифицируем формулу (7) следующим образом:

$$G_{t_i}^S = \langle C, R, P, W \rangle, \tag{15}$$

где $G_{t_i}^S$ – схема связанных данных; C – множество классов, причем $C \subset G_{t_i}^S$; R – множество отношений между классами, причем $R \subset G_{t_i}^S$; P – множест-

во свойств классов, причем $P \subset G_{t_i}^S$; W – множество присоединенных процедур, реализующих операционную составляющую ИС.

Кроме того, определим множество пользователей системы U и подмножество процедур W^U ($W^U \subset W$), которые будут отвечать за обеспечение интерактивного взаимодействия пользователя с ИС с помощью формул (16) – (17).

Пользователей ИС производственного предприятия определим следующим образом:

$$U = \{u_i\}, i = \overline{1, h}, \quad (16)$$

где U – множество пользователей ИС; u_i – i -й пользователь ИС.

Таким образом, взаимодействие пользователя (t_i) с ИС определим в виде параметризованной присоединенной процедуры w^{int} как процесс формирования запроса к ИП (входного воздействия), выполнения соответствующей входному запросу процедуры объекта $w_j \in W^U$ (реакция ИС на входное воздействие) и визуализации результатов выполнения запроса (w^{viz}):

$$w^{int}(q) = \left\{ \begin{array}{l} w^{err}(q) | w_j(q) = \emptyset \\ w^{viz}(q, C_q^I) | w_j(q) \neq \emptyset \end{array} \right\}, \quad (17)$$

где q – запрос к ИС; C_q^I – множество экземпляров классов, полученных в качестве результата выполнения запроса j -процедурой w ;

w^{err} – реакция ИС в случае ошибки выполнения j -процедурой w .

Таким образом, модель ИС производственного предприятия определим как расширенную модель информационного пространства, выраженную с помощью формул (1) – (14), с модификацией (16) и процедурой взаимодействия пользователя (17).

Выводы

1. Разработана математическая модель информационного пространства производственного предприятия на основе связанных данных, позволяющая устранить стихийность при разработке информационной управляющей системы.

2. Получила дальнейшее развитие математическая модель распределенной информационной системы производственного предприятия, что дает возможность реализовать функциональные задачи оперативного управления бизнес-процессами на всех уровнях на основе единого подхода к обработке данных.

3. Разработана модель взаимодействия пользователя со структурированным информационным пространством, что позволяет эффективнее решать задачи оперативного управления.

Список литературы

1. Хоп Г. Шаблоны интеграции корпоративных приложений; Пер. с англ. / Г. Хоп, Б. Вульф. – М.: ООО «И. Д. Вильямс», 2007. – 672 с.
2. Шаппелл Д. ESB – сервисная шина предприятия / Д. Шаппелл. – СПб.: БХВ-Петербург, 2008. – 368 с.
3. Грейсв М. Проектирование баз данных на основе XML; Пер. с англ. / М. Грейсв. – М.: Издательский дом «Вильямс», 2002. – 640 с.
4. Powers S. Practical RDF / S. Powes. – O'ReillyMedia: 2008. – 352 p.
5. DuCbarne B. Learning SPARQL / B. DuCbarne. – O'ReillyMedia: 2011. – 258 p.
6. Арсеньев Б.П. Интеграция распределенных баз данных / Б.П. Арсеньев, С.А. Яковлев. – СПб.: Издательство «Лань», 2001. – 464 с.
7. Удосконалення технологій доступу та обробки пов'язаних даних семантичних додатків Linked Data / І.М. Галушка, В.В. Завгородній, С.М. Солошич, С.С. Щербак // АСУ и приборы автоматизации. – X.: ХНУРЕ, 2012. – № 159. – С. 67-73.

Поступила в редколлегию 4.02.2013

Рецензент: д-р техн. наук, проф. А.П. Оксанич, Кременчугский национальный университет им. М. Остроградского, Кременчуг.

ЄДИНИЙ ІНФОРМАЦІЙНИЙ ПРОСТІР ВИРОБНИЧИХ ПІДПРИЄМСТВ НА ОСНОВІ ЗВ'ЯЗАНИХ ДАНИХ

В.В. Завгородній, С.С. Щербак

В статті розроблена математична модель єдиного інформаційного простору виробничого підприємства на основі зв'язаних даних, отримала подальший розвиток математична модель розподіленої інформаційної системи виробничого підприємства, яка забезпечує ефективне виконання задач оперативного управління, а також запропонована модель взаємодії користувача з інформаційним простором.

Ключові слова: Linked Data, RDF, SPARQL, Semantic Web, Virtuoso, шаблони інтеграції.

COMMON INFORMATIONAL SPACE ON LINKED DATA FOR MANUFACTURING ENTERPRISES

V.V. Zavgorodny, S.S. Shcherbak

The article describes mathematical model of a common informational space designed for manufacturing enterprises. This common informational space is based on linked data. Also, the mathematical model of distributed information system has been extended. Such a system has been designed for a manufacturing enterprise which effectively solves operational control tasks. The model for users' interaction with informational space is also proposed.

Keywords: Linked Data, RDF, SPARQL, Semantic Web, Virtuoso, Integration patterns