



УДК 004.822:514

ОСОБЕННОСТИ ПОСТРОЕНИЯ ИНФОРМАЦИОННЫХ ПРОСТРАНСТВ ПРОМЫШЛЕННЫХ ПРЕДПРИЯТИЙ НА ОСНОВЕ СВЯЗАННЫХ ДАННЫХ

Галушка И.Н. *, Щербак С.С. *

**Кременчугский национальный университет имени Михаила Остроградского,
г. Кременчуг, Украина*

ilona.galushka@ya.ru

sergey.shcherbak@gmail.com

В работе рассмотрены вопросы построения информационных пространств, как интеграционного процесса объединения источников данных информационных систем промышленных предприятий для формирования единого интерфейса доступа на основе объектно-ориентированного подхода. Разработана формальная модель информационного пространства и предложены соответствующие отображения в качестве интеграционных компонентов существующих источников в информационное пространство

Ключевые слова: интеграция; связанные данные; паттерны реализации; информационное пространство

Введение

В современных условиях управление данными на промышленных предприятиях (ПП) характеризуется наличием большого количества распределенных гетерогенных источников данных (ИД) зачастую не связанных едиными механизмами управления. Причиной накопления информации в таких источниках является отсутствие реализации политики автоматизации ПП по единому принципу, что приводит к тому, что на одном ПП могут функционировать несколько локальных информационных систем (ИС). Интеграция подобных источников данных является ключевым фактором для их объединения в единое информационное пространство (ИП) связанных данных предприятия (англ. Linked Enterprise Data, LED), что существенно упрощает доступ к данным, которые необходимы для лиц, принимающих решения на ПП. В рамках такого подхода информационное пространство будем рассматривать как некий промежуточный интеграционный компонент с универсальным коммуникационным интерфейсом (рис.1).

В последнее время в качестве эффективного средства коммуникации с низкой стоимостью развертывания используется сеть Интернет и гипертекстовые технологии Всемирной Паутины (WWW). Документно-ориентированная природа WWW и отсутствие эффективных средств описания содержащихся в документах данных со временем привела к развитию новых направлений и

концепций, которые могут стать эффективной платформой для информационных систем промышленных предприятий.

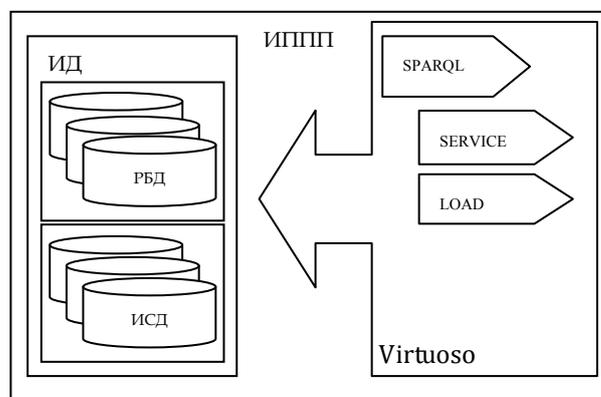


Рисунок 1 — информационное пространство на основе различных типов источников данных

Одной из таких является концепция связанных данных. 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 идентификатора ресурсов URI или его подкласса универсального локатора ресурсов 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), причем ориентация этой концепции на распределенное хранение данных является на 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.

При разработке ИС на основе связанных данных, в частности на основе RDF, 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), что в условиях применения на промышленных предприятиях требует создания новых моделей представления данных для учета операционной составляющей процессов.

Промышленные 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 дало бы возможности 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 ростом количества данных, хранящихся на производственных серверах, и требованием хранения разноаспектных характеристик некоторых видов данных, применяемых, например, при принятии решений уполномоченными лицами.

Информационное обеспечение производственных предприятий, построенное в условиях лоскутной автоматизации, строится на основе стихийной архитектуры и характеризуются использованием ИС контрольно-учетного характера, отсутствием единой формальной системы процесса выработки управленческих решений, отсутствием непосредственного доступа ЛПР к информационным ресурсам и т.д., что необходимо

обеспечить для эффективного решения задач как поддержки принятия решений, так и задач оперативного управления в целом.

Формальная модель информационного пространства промышленного предприятия

Для разработки модели информационного пространства уточним определения основных понятий, используемых в работе.

Определение 1. Связанные данные – это структурированные данные вида «поименованный граф, субъект, предикат, объект», каждый компонент которых для идентификации использует унифицированный идентификатор ресурсов URI (англ. Uniform Resource Identifier).

Замечание 1. Учитывая необходимость использования информации о местонахождении данных в работе, используется подкласс URI универсальный локатор ресурсов URL (англ. Uniform Resource Locator), обеспечивающий помимо идентификации ресурса представление информации о местонахождении ресурса.

Определение 2. Информационное пространство промышленного предприятия – это совокупность распределенных источников данных, соответствующих информационным системам предприятия, средств их сопровождения и использования, функционирующих на основе общих принципов.

Замечание 2. В связи с тем, что информационное пространство промышленного предприятия предложено создавать в соответствии с положениями концепции связанных данных, классы источников данных, используемых в работе, ограничим классом источников связанных данных, что с учетом развитых средств интеграции не сужает общность реализуемых в работе идей.

Сохраняя общность с концепцией связанных данных, определим информационное пространство промышленного предприятия согласно определению (2) как совокупность множеств источников связанных данных Src и присоединенных процедур Pr^U , которые реализуют функциональность информационного пространства ИП:

$$I = \langle Src, Pr^U \rangle, \quad (1)$$

Источники связанных данных Src представим как множество четырехкомпонентных структур – квадов (англ. Quad), которые состоят из поименованных графов (G), субъектов (S), предикатов (P) и объектов (O), причем каждый компонент квада с учетом распределенности их хранения должен быть представлен в виде URL:

$$Src = \langle G, S, P, O \rangle, \quad (2)$$

Квады в источниках связанных данных

группируются в контексты $G^c \subseteq G$ – множества квадов, сгруппированных по G . Каждому контексту g с учетом формулы (2) поставим в соответствие триплет $\langle S, P, O \rangle$:

$$\forall g \in G^c : g = \langle S, P, O \rangle, \quad (3)$$

где g – контекст источника связанных данных, G^c – множество всех контекстов, S – множество субъектов, P – множество предикатов, O – множество объектов.

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

С учетом необходимости хранения не только связанных данных в источниках, но и их схем, множество G^c разделим на два подмножества G^S и G^I , причем так, что $G^S \cap G^I = \emptyset$, тогда множество триплетов $\langle S, P, O \rangle$, принадлежащих G^S будут определять схему связанных данных, а G^I собственно сами связанные данные.

Далее опишем связанные данные в терминах, понятных пользователям ИС, а именно в терминах объектно-ориентированного проектирования.

Пусть $G_t^S \subset G^c$ – структурно-логическая схема t -го источника связанных данных (рис. 2.), тогда определим ее в терминах объектно-ориентированного проектирования (класс, подкласс, экземпляр класса, отношения между классами):

$$G_t^S = \langle C, R, F \rangle, \quad (4)$$

где G_t^S – схема связанных данных, C – множество классов, причем $C \subset G_t^S$, R – множество отношений между классами ($R \subset G_t^S$), F – множество свойств классов ($F \subset G_t^S$).

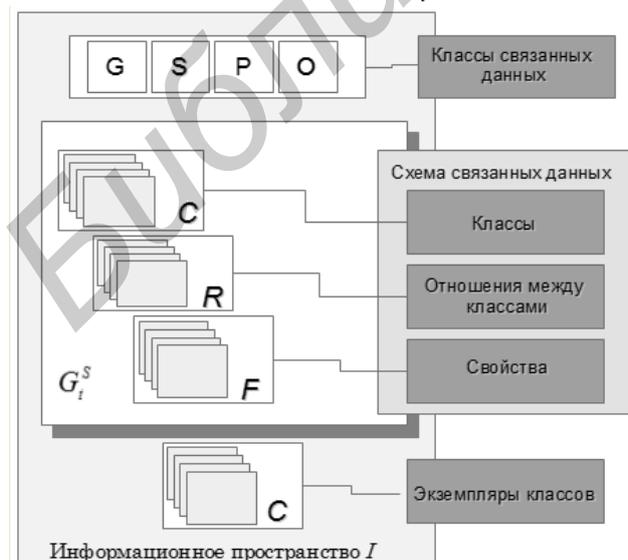


Рисунок 2 – Схема источника связанных данных

Множество свойств классов представим следующим образом:

$$F = \langle T, L, V \rangle, \quad (5)$$

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

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

$$P = \langle T, \emptyset, V \rangle, \quad (6)$$

Для определения структуры классов значение свойств не имеют смысла, поэтому $V = \emptyset$, а выражение (6) будет иметь вид:

$$P = \langle T, \emptyset, \emptyset \rangle, \quad (7)$$

Определим множество доступных отношений между классами R для организации связей «быть подклассом» и «быть экземпляром класса» и др., формально определенных следующим образом:

$$R = \{isSubclass, instanceOf, isEquivalent\}, \quad (8)$$

где *isSubclass* – отношение «быть подклассом», *instanceOf* – отношение «быть экземпляром класса», *isEquivalent* – отношение «быть эквивалентным».

Тогда с учетом формулы (8) представим множество $R = R^s \cup R^i \cup R^e$ как объединение непересекающихся подмножеств отношений «быть подклассом» R^s , «быть экземпляром класса» R^i и «быть эквивалентным» R^e ($R^s \cap R^i \cap R^e = \emptyset$), логическая структура которых задана формулами (5) – (8).

Логическая структура R^s :

$$R^s = \{ \langle c'_1 \in C : c'_1 isSubclass c \rangle, \dots, \langle c'_v \in C : c'_v isSubclass c \rangle \}, \quad (9)$$

где c, c' – классы, C – множество классов, v – общее количество классов, связанных отношением «быть подклассом».

Логическая структура R^i :

$$R^i = \{ \langle c'_1 \in C : c'_1 instanceOf c \rangle, \dots, \langle c'_w \in C : c'_w instanceOf c \rangle \}, \quad (10)$$

где c, c' – классы, C – множество классов, w – общее количество классов, связанных отношением «быть экземпляром класса».

Логическая структура R^e :

$$R^e = \{ \langle c'_1 \in C : c'_1 isEquivalent c \rangle, \dots, \langle c'_z \in C : c'_z isEquivalent c \rangle \}, \quad (11)$$

где c, c' – классы, C – множество классов, z – общее количество классов, связанных отношением «быть эквивалентным».

С учетом вышесказанного, класс $c \in C$ схемы G_t^S определим как логическую

структуру с установленными отношениями Rс другими классами, свойствами F формулы (5) и функциональностью $Pr \subset Pr^U$ следующим образом:

$$\forall c \in C^{G^S} : c = \langle R, F, Pr \rangle, \quad (12)$$

Важно отметить, что для разделения классов и их экземпляров в работе используется не только отношение «быть экземпляром класса», но и установленные значения (V) свойств F класса.

Таким образом, формальную модель информационного пространства на основе связанных данных представим формулами (1) – (12).

Интеграционный процесс составляющих компонентов информационного пространства представим как отображение (δ_1) существующих данных и схем источников (SRC) на информационное пространство (I) предприятия и формально представим следующим образом:

$$\delta_1 : \forall s \in SRC \xrightarrow{sh^1} I, \quad (13)$$

где sh^1 – интеграционная схема информационного пространства.

С учетом положений модели, программное обеспечение для обработки связанных данных, будем рассматривать как расширение функциональных возможностей информационного пространства, реализующее конкретные необходимые для пользователя процедуры. В рамках такого подхода практическая реализация изложенных в работе положений подразумевает использование протокола SPARQL и промежуточного программного обеспечения OpenLink Virtuoso или любого другого, поддерживающего коммуникацию посредством протокола SPARQL.

Для реализации подключения к источнику данных и загрузки данных в информационное пространство можно использовать следующую процедуру языка Virtuoso/PL ПП OpenLink Virtuoso:

```
create procedure connectRS ()
{
  declare graph varchar;
  for ( sparql SELECT ?graph ?s ?p ?o
  WHERE
  {
    SERVICE
    <http:// tniks.kdu.edu.ua/sparql>
    {
      graph ?graph {?s ?p ?o }
    }
  })
  do {
    sparql load ?graph into ?graph;
  }
};

```

где graph –именованный граф, s- субъект, p - предикат, o - объект.

Обращаем внимание, что вышеприведенный скрипт является реализацией обобщенного

алгоритма загрузки без учета специфических особенностей для каждого источника данных, загрузка которых осуществляется в информационное пространства.

Из вышесказанного следует, что при построении информационного пространства промышленного предприятия задача интеграции составляющих информационное пространство ресурсов требует проведение дополнительных исследований, которые будут проведены в дальнейшем.

Заключение

В работе были рассмотрены вопросы построения информационных пространств, как интеграционного процесса объединения источников данных информационных систем промышленных предприятий для формирования единого интерфейса доступа на основе объектно-ориентированного подхода.

Разработана формальная модель информационного пространства и предложены соответствующие отображения в качестве интеграционных компонентов существующих источников в информационное пространство

Библиографический список

- [Майер-Шенбергер В., 2014] Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / Майер-Шенбергер В., Кукьер К. – М.: Манн, Иванов и Фербер, 2014. – 240 с.:ил.
- [DuCarme, 2011] DuCarme B. Learning SPARQL / B. DuCarme. - O'ReillyMedia: 2011. - 258 с.
- [Powers, 2008] Powers S. Practical RDF / S. Powes. - O'Reilly Media: 2008. - 352 с.
- [Холзнер, 2004] Холзнер С. XML. Энциклопедия, 2-е издание / Стивен Холзнер. - СПб.: Питер, 2004. - 1101 с.:ил.
- [Бек, 2008] Бек К. Шаблоны реализации корпоративных приложений.: Пер. с англ. - М.:ООО "И. Д. Вильямс", 2008. – 176 с.:ил.
- [Гудсон, 2013] Гудсон Дж. Практическое руководство по доступу к данным / Д. Гудсон, Р. Стюарт. – СПб.:БВХ-Петербург, 2013. – 304с.:ил.
- [Палкин, 2013] Палкин Н.Б. Бизнес-аналитика от данных к знаниям, 2-е издание / Палкин Н.Б., Орешков В.И. – СПб.: Питер, 2013. – 704 с.:ил.

PECULIARITIES OF LINKED DATA INFORMATIONAL SPACE DEVELOPMENT FOR INDUSTRIAL ENTERPRISES

Galushka I.N. *, Shcherbak S.S. *

*KremenchukMykhailoOstrohraskyiNational University, Kremenchuk, Ukraine

ilona.galushka@ya.ru

sergey.shcherbak@gmail.com

Issues on development of informational spaces are observed as an integration process of uniting information system data sources of industrial enterprises for generation of unique access interface based on object-oriented design. Informational space models are developed with corresponding mappings (as integration components) of existing resources into the informational space