

УДК 681.3

Ю.И. Воротницкий, А.В. Пупко

ИЕРАРХИЧЕСКИЕ КОРПОРАТИВНЫЕ СИСТЕМЫ WWW-САЙТОВ

Предлагается новая объектно-ориентированная архитектура иерархических корпоративных информационных систем, базирующихся на www-технологиях.

Введение

Широкое распространение в интернет- и интранет-сетях корпоративных систем www-сайтов [1] обусловлено необходимостью создания эффективных механизмов сбора, накопления, тиражирования, отображения разнородной информации и, в конечном итоге, управления корпоративными информационными ресурсами (ИР). Для решения этих задач требуется создание развитых механизмов информационного обмена, отличных от копирования информации или построения систем гиперссылок [2] на ИР.

В настоящей работе предлагается новая архитектура корпоративных систем www-сайтов, основанная на хранении всех ИР в интегрированной базе данных. Рассматриваются новые принципы построения корпоративных иерархических систем www-сайтов, обеспечивающие использование объектно-ориентированной методологии в процессах их разработки и функционирования; архитектура информационного хранилища, обеспечивающего структурирование разнородных ИР, их объединение с методами обработки и представления, а также информационное взаимодействие ИР в соответствии с иерархическим подходом к их организации.

1. Информационное взаимодействие в иерархических системах www-сайтов

Иерархический подход к построению www-систем, предложенный в [3], позволяет адекватно отразить иерархию административно-организационной структуры и информационных потоков в большинстве корпоративных систем. Например, в территориально распределенной корпоративной интернет-системе такая иерархия может быть построена по схеме «республика \Rightarrow области \Rightarrow районы». Во внутренней интранет-сети вуза естественной является иерархическая организация образовательных ИР по схеме «учебное заведение \Rightarrow факультеты \Rightarrow кафедры \Rightarrow персональные страницы преподавателей, аспирантов, студентов». В подобных иерархических системах обеспечивается многопользовательская и многоуровневая система независимого наполнения и администрирования сайтов, ключевым вопросом при этом является обеспечение эффективного информационного взаимодействия между сайтами. Известные механизмы такого взаимодействия, заключающиеся в тиражировании (распространении) ИР (например, по протоколу RSS [4]), предлагается развить на основе объектно-ориентированного подхода с помощью следующих способов взаимодействия www-сайтов:

- тиражирования ИР и информационных сервисов (подписок, интерактивных опросов, форумов и др.) с верхних уровней иерархии (сайты-издатели) на нижележащие (сайты-подписчики);
- передачи данных с сайтов нижележащих уровней на сайты вышележащих уровней (при необходимости с дальнейшей агрегацией и статистической обработкой);
- использования при отображении информации на нижележащих уровнях иерархии методов обработки и представления информации с вышележащих уровней.

Понимая под ИР произвольный объем разнородной информации, хранимый в некотором унифицированном формате, обозначим R_i^j как i -й ИР, доступный на j -м уровне иерархии. Этот ИР может быть непосредственно сформирован на этом же j -м уровне или на более раннем k -м

($k \leq j$). В последнем случае можно говорить о наследовании содержимого информационного ресурса, что можно записать как $R_i^j \Leftarrow R_i^k$.

На каждом уровне иерархии ИР может храниться в полях баз данных, быть представлен в промежуточном структурированном виде (например, в формате XML) и в виде, готовом для отображения (формат HTML). Отображение информации на www-сайте в этом случае обеспечивается с помощью специальных программных модулей, реализующих соответствующие методы доступа к ИР с преобразованием в формат XML и отображения в формате HTML. При этом наследование ИР не предполагает простого копирования информации или размещения ссылки на нее, так как совместно с информацией на каждом уровне иерархии должны применяться соответствующие методы (процедуры) доступа $A_i^j(a_{i,q}^j)$ и отображения $P_i^j(p_{i,r}^j)$, где $a_{i,q}^j$ и $p_{i,r}^j$ – параметры этих методов. Методы, реализованные в виде программных модулей, хранятся вместе с соответствующими ИР, что соответствует принципу инкапсуляции информации и методов ее обработки. Данные методы и значения параметров также могут наследоваться: $A_i^j(a_{i,q}^j) \Leftarrow A_i^k(a_{i,q}^k)$, $P_i^j(p_{i,r}^j) \Leftarrow P_i^k(p_{i,r}^k)$. Таким образом, унаследовав информационный ресурс R_i^k , сайт j -го уровня может обработать его с помощью собственного метода отображения $P_i^j(p_{i,r}^j)$ либо метода, унаследованного на верхнем уровне, $P_i^j(p_{i,r}^j) \Leftarrow P_i^k(p_{i,r}^k)$, изменив или оставив прежними значения параметров $p_{i,r}^j$.

Отображение на www-сайте j -го уровня ресурса, сформированного на этом же уровне, выполняется в соответствии со следующим правилом

$$R_i^j \xrightarrow{A_i^j(a_{i,q}^j)} X_i^j \xrightarrow{P_i^j(p_{i,r}^j)} H_i^j. \quad (1)$$

При получении запроса в результате обработки ИР с помощью соответствующего метода доступа $A_i^j(a_{i,q}^j)$ в оперативной памяти сервера динамически генерируется промежуточное структурированное представление X_i^j , которое затем обрабатывается программной процедурой, реализующей метод отображения $P_i^j(p_{i,r}^j)$, в результате чего динамически генерируется отображаемый HTML документ H_i^j .

Механизм тиражирования ИР k -го уровня на нижележащий j -й уровень иерархии в простейшем случае может быть представлен в виде

$$(R_i^j \Leftarrow R_i^k) \xrightarrow{A_i^k(a_{i,q}^k)} X_i^k \xrightarrow{P_i^j(p_{i,r}^j)} H_i^j. \quad (2)$$

Для доступа к наследуемому ИР $R_i^j \Leftarrow R_i^k$ и преобразования в стандартный промежуточный формат формируется запрос к серверу k -го уровня и используется метод A_i^k , определенный на том же k -м уровне иерархии, что и родительский ИР. Напротив, методы отображения либо формируются на i -м уровне формулы (2), либо наследуются у родительского k -го с возможностью модификации значений отдельных параметров этого метода $p_{i,r}^j$. В последнем случае получаем

$$(R_i^j \Leftarrow R_i^k) \xrightarrow{A_i^k(a_{i,q}^k)} X_i^k \xrightarrow{P_i^j(p_{i,r}^j) \Leftarrow P_i^k(p_{i,r}^k)} H_i^j. \quad (3)$$

Наконец, процедуры сбора, агрегации, анализа и отображения ИР, формируемых на нижних уровнях иерархии, в предлагаемой архитектуре могут быть реализованы в соответствии со следующей схемой:

$$H_i^k \leftarrow \frac{P_i^k(p_{i,r}^k)}{G_i^k} \leftarrow \frac{S_i^k(s_{i,t}^k)}{X_i^j} \leftarrow \frac{A_i^j(a_{i,q}^j)}{R_i^j} \quad (4)$$

Здесь на k -м уровне системы может дополнительно формироваться промежуточное представление данных G_i^k , агрегированных от нескольких нижележащих источников, для чего в базе данных на k -м уровне хранятся процедуры $S_i^k(s_{i,t}^k)$ агрегирования и формирования промежуточного структурированного представления.

2. Архитектура информационного хранилища

Для реализации рассмотренной методологии информационного взаимодействия *www*-сайтов предлагается новая архитектура хранилища ИР, основанная на установлении отношений между ИР (и методами их обработки и представления, хранимыми совместно с ИР) путем формирования в рамках этого хранилища динамически модифицируемой иерархической системы информационных ресурсов (ИСИР). Такая иерархическая система (рис. 1) может быть построена по следующим правилам:

1. Информационная составляющая системы (ИСС) представляет собой базу данных, в которой ИР, способы их обработки и пользователи хранятся в неструктурированном или слабоструктурированном виде.

2. Для структурирования информации в ИСС используется многоуровневая древовидная структура (МДС), узлы которой связаны отношениями принадлежности и, в свою очередь, могут являться структурами низшего ранга.

3. Взаимосвязи между ИР и узлами МДС задаются в виде отношений, определяемых статически или динамически. Каждый ИР обязательно должен принадлежать одному или нескольким узлам.

4. Отношения между узлами могут динамически изменяться в процессе функционирования ИСИР. Каждому узлу МДС может быть назначен родительский узел системы, который имеет в качестве одного из родительских узлов узел, являющийся активным (тем самым запрещается переопределение узлов между иерархиями, имеющими различные нулевые уровни). Данное ограничение введено для разделения данных, способов обработки, объектов доступа и области удаленной информации.

5. На каждом уровне иерархии ИР хранятся в полях базы данных, могут быть представлены в промежуточном структурированном виде или в виде, готовом для отображения.

Можно определить пять основных типов узлов, используемых в ИСИР:

– Статический информационный узел. Отношения «ИР \Rightarrow узел» определяет владелец ИР.

– Динамический информационный узел. Отношения «ИР \Rightarrow узел» строятся на основании характеристик и реквизитов, используемых при описании ИР.

– Узел обработки информации (статический или динамический). Узел определяет отношения к способам обработки ИР, а не к самим ИР. Данный тип узла может размещаться только на нулевом уровне иерархии.

– Узел доступа (статический или динамический). Узел определяет отношения к пользователям, имеющим доступ к узлам, определенным выше. Статический узел может размещаться только на нулевом уровне иерархии. На нижних по отношению к нему уровнях иерархии могут порождаться только динамические узлы, позволяющие осуществлять необходимое группирование пользователей, задавая для них права доступа к ИР.

– Узел удаленной информации. На этот узел ссылается вся информация из вышеперечисленных узлов при её удалении. Таким образом, процесс удаления состоит в уничтожении всех отношений от ресурса к узлам и создании отношения к узлу удаленной информации.

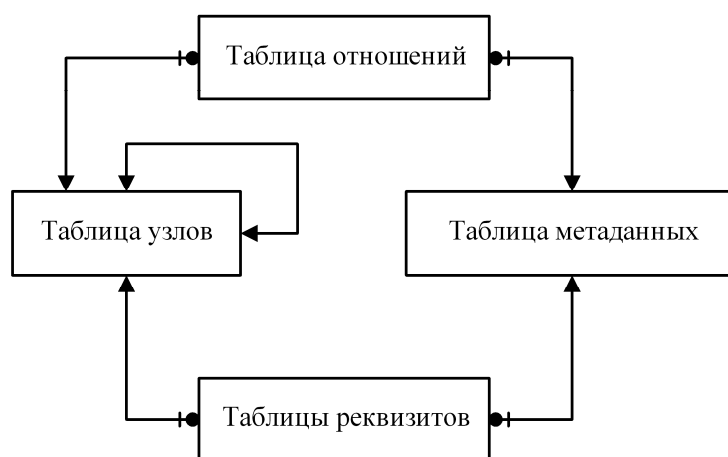


Рис. 2. Пример схемы отношений

По сути, использование предложенной многоуровневой системы обеспечивает построение гибких систем классификации ИР, позволяющих применять к одним и тем же объектам классификаторы из различных предметных областей. Особенно важно, что при создании нового классификатора не требуется изменение форматов наполнения и хранения ИР, поскольку для привязки ИР к классификатору используется создание дополнительного отношения.

Построенная по предложенной методике ИСИР позволяет осуществлять построение информационных хранилищ отдельных www-сайтов и объединять ИР множества сайтов, входящих в единую корпоративную систему. Характеристики сайтов и соответствующие ИР располагаются в общем хранилище информации, а отношения между сайтами строятся по принципу подчиненности (реальной или виртуальной) организаций друг другу.

Заключение

К основным достоинствам предложенной архитектуры корпоративных систем www-сайтов, прежде всего, относятся:

- структурирование ИР на уровне отношений полей таблиц баз данных, что облегчает администрирование и поиск ИР;
- сокращение дублирования информации путем соответствующей организации информационных потоков при реализации механизмов наследования;
- унификация методов обработки и возможность интеграции с решениями сторонних разработчиков за счет использования промежуточных структурированных представлений ИР в общепринятых форматах, например XML;
- обеспечение безопасности ИР благодаря отсутствию прямого доступа процедур отображения к полям таблиц баз данных (эти процедуры обращаются только к динамически генерируемым промежуточным представлениям).

Реализация рассматриваемой архитектуры корпоративных www-систем предполагает решение следующих ключевых задач:

- разработки механизмов исполнения наследуемых процедур;
- разработки методики формирования информационных каналов с целью передачи запросов на извлечение данных, наследуемых процедур и самих данных;
- создания эффективных методов поиска информации с учетом иерархической организации ИР.

Предлагаемая архитектура корпоративных www-систем была успешно апробирована при создании Единой системы информационно-образовательных ресурсов Минска [5–7], включающей более 250 www-сайтов.

Список литературы

1. Спирли Э. Корпоративные хранилища данных. Планирование, разработка, реализация. – М.: Мир, 2001. – 540 с.
2. Хольцшлаг М. Использование HTML и XHTML. – М.: Мир, 2003. – 430 с.
3. Воротницкий Ю.И. Об архитектуре корпоративных информационных веб-систем // Докл. НАН Беларуси. – 2004. – Т. 48. – № 6. – С. 32–34.
4. RSS 2.0 Specification // <http://blogs.law.harvard.edu/tech/rss>
5. Система информационно-образовательных интернет-ресурсов города Минска / А.В. Воронов, Ю.И. Воротницкий, М.Г. Зеков и др. // Управление информационными ресурсами: мат. II науч.-практ. конф. – Мн.: Редакционно-издательский центр Академии управления при Президенте Республики Беларусь, 2004. – С. 204–206.
6. Воротницкий Ю.И., Пупко А.В. Система управления образовательными веб-ресурсами // Информационные технологии в образовании: мат. XIV Междунар. конф.-выст. – М.: МИФИ, 2004. – С. 262–263.
7. Единое информационное образовательное пространство // <http://adm.pst.by>

Поступила 20.01.05

*Белорусский государственный университет,
Минск, Скорины, 4
email: vorotn@bsu.by
pupko@bsu.by*

Y.I. Varatnitsky, A.V. Poupko

HIERARCHIC CORPORATE SYSTEMS OF WWW-SITES

A new object-oriented architecture of hierarchic corporate intelligence systems based on www-solutions is proposed.