
Информатика и вычислительная техника

УДК 004.65:004.75

DOI: 10.22363/2312-9735-2017-25-3-295-305

Современные технологии интеграции информации из независимых источников и их применение при построении информационной системы, объединяющей транспортные расписания

А. С. Панкратов, А. Х. Пшеунов

*Кафедра информационных технологий
Российский университет дружбы народов
ул. Миклуто-Маклая, д. 6, Москва, Россия, 117198*

Рассматривается задача объединения в единое целое автобусных расписаний, предоставленных различными источниками – автобусными парками, независимыми друг от друга и, возможно, находящимися в разных регионах страны. Эта объединённая система предназначена, в частности, для удобства поиска имеющихся автобусных маршрутов между двумя заданными точками. Для решения рассматриваемой задачи предлагается использовать комбинированный подход, сочетающий преимущества технологий медиаторов и хранилищ. В статье рассматривается модель из трёх источников данных, которыми являются автобусные парки, выставляющие информацию о своих расписаниях. Предполагается, что все источники имеют сходные концептуальные схемы, но обладают своими особенностями. В частности, могут быть разные наименования таблиц и атрибутов на разных источниках и разное распределение атрибутов по таблицам. Возможно также отсутствие некоторых атрибутов на отдельных источниках. Строятся таблицы соответствия и медиатор, транслирующий пользовательские запросы на источники. При этом для идентификации нужных источников поддерживается вспомогательное хранилище небольшого объёма, которое содержит информацию об остановочных пунктах, обслуживаемых каждым из источников. Описывается технология обновления хранилища и схема исполнения пользовательского запроса, использующая информацию, содержащуюся в хранилище.

Ключевые слова: база данных, интеграция информации, мультибазовая система, медиатор, хранилище

1. Введение

С каждым годом проблема интеграции информации из различных независимых источников становится все актуальнее. Часто возникает необходимость объединить в единое целое разнородные БД. Для формального описания больших объёмов разнородных данных при проектировании интегрированных баз данных и программного обеспечения, наряду с традиционной реляционной моделью, предлагается ряд новых подходов, в том числе основанных на использовании многомерных моделей данных. Один из таких подходов, предложенный в работах С. В. Павлова, О. И. Христуло и ряда других специалистов, основывается на использовании концепции многомерных информационных объектов. Также данной проблеме особое внимание уделили в своих работах А. В. Черноусова, А. Кудинова, другие авторы. Однако на данный момент намечаются лишь контуры общего подхода к решению задачи интеграции данных из разных компьютерных систем, а основная часть публикаций по данной проблеме затрагивает лишь частные случаи, со своими индивидуальными особенностями [1–4].

Данная статья также относится к частному случаю. Рассматривается задача объединения в единое целое автобусных расписаний, предоставленных различными источниками — автобусными парками, независимыми друг от друга и, возможно, находящимися в разных регионах страны.

Статья поступила в редакцию 4 марта 2017 г.

Следует отметить, что подобные справочные системы в настоящее время существуют и ими можно пользоваться через Интернет (например, система «Яндекс.Расписания»), однако, как показывает практика, они охватывают в основном магистральные маршруты и не всегда выдают достоверную информацию. В настоящей работе делается попытка построения системы, лишённой этих изъянов и выдающей исключительно актуальную информацию.

2. Описание предметной области

В наше время существует много видов автобусных парков разного уровня и масштаба: городские, районные, региональные. При этом в каждом из них ведётся свой учёт маршрутов, имеются свои расписания и т.д. Имеется некоторое количество населённых пунктов, связанных автобусными маршрутами, для каждого маршрута имеются конечные и промежуточные пункты, в которых пассажиры могут выходить или садиться в автобус. Одни и те же пункты могут связываться разными маршрутами. Имеется расписание автобусов, которое связывает определённые пункты между собой. В расписании определяются дни недели, время отправления из начального в конечный пункт, список промежуточных пунктов и время их прохождения. В автобусных парках находится некоторое количество автобусов разных категорий, с разным количеством мест и некоторое количество водителей. На основе расписания формируются конкретные рейсы (к рейсам, определённым в расписании, прикрепляются водитель, автобус, устанавливается дата). Ведётся продажа билетов, в билете указаны номер, дата, код рейса, пункт отправления, пункт назначения и цена. Цена билета зависит от расстояния между пунктами отправления и назначения, категории автобуса, ценовой политики транспортной компании и т.д. Предлагаемая в настоящей работе интегрированная (мультибазовая) система, объединяющая изначально разрозненные расписания, предназначена для удобства поиска имеющихся автобусных маршрутов между двумя заданными точками. Пользователь вводит пункт отправления, пункт назначения и дату в соответствующие поля, система выдаёт ему расписание рейсов, удовлетворяющее критериям поиска. В расписании показаны поля: маршрут (перечислены основные населённые пункты, через которые будет проходить движение), дата, время отправления, время прибытия, категория автобуса, количество свободных мест, цена билета. Пользователь выбирает подходящий ему рейс и оплачивает билет.

3. Основные технологии интеграции данных

Существует три основных подхода к интеграции баз данных [5]:

- 1) федеративные базы данных — источники независимы, но могут сообщаться между собой для обмена информацией;
- 2) хранилища данных — данные от источников на периодической основе загружаются в централизованное хранилище, возможно, с предварительной обработкой с целью приведения их в соответствие со структурой хранилища;
- 3) медиаторы — программные компоненты, принимающие запросы от пользователей и затем направляющие их к соответствующим источникам, возможно, с предварительной трансляцией; полученные ответы от источников приводятся в соответствие со структурой медиатора, объединяются и выдаются пользователю.

Каждый из подходов обладает своими преимуществами и недостатками. Федеративные базы данных бывают эффективными при наличии малого числа источников, в противном случае возникает необходимость поддержания большого количества связей между источниками (для экспорта-импорта информации). В хранилищах существует проблема периодической загрузки большого объёма информации с соответствующими требованиями к ресурсам системы, кроме того, информация, полученная от хранилища, может быть несколько устаревшей (не вполне актуальной). В медиаторах нет этих проблем, но конкретные запросы к ним исполняются

медленнее, чем к хранилищу. Кроме того, в медиаторах необходимо поддерживать механизм идентификации необходимых источников.

4. Предлагаемая схема решения

Для решения рассматриваемой задачи интеграции междугородних автобусов предлагается использовать комбинированный подход, сочетающий преимущества медиаторов и хранилищ. Этот подход состоит в применении технологии медиаторов с участием вспомогательного хранилища небольшого объёма, которое позволяет идентифицировать источники, необходимые для исполнения запроса пользователя.

Будем предполагать, что существует несколько независимых автобусных станций, работа каждого из которых поддерживается собственной компьютерной системой с использованием какой-либо СУБД, поддерживающей формат SQL. В рамках каждого из источников эта система может сопровождать различные бизнес-процессы (создание маршрутов, добавление рейса).

Как выше отмечалось, для потенциального пассажира часто бывает важно получить все возможные варианты, удовлетворяющие его запросу. Для этого следует отправить запрос во все автобусные парки (источники), обслуживающие населённые пункты, соответствующие требованиям клиента. При этом для оптимизации процесса (быстродействие транзакции, минимизация нагрузки на сеть) важно запрашивать только те источники, которые содержат информацию по данным населённым пунктам и не запрашивать другие источники. Для этой цели предлагается создать медиатор со вспомогательным хранилищем. Медиатор принимает запрос от пользователя и переадресует его только к тем источникам, где находится интересующая его информация. Для определения того, к каким источникам следует обращаться, при медиаторе создаётся «мини-хранилище», периодически закачивающее от источников лишь краткую информацию о маршрутных пунктах, обслуживаемых данным источником (автопарком). Примерный алгоритм: при поступлении запроса медиатору происходит формирование вспомогательного запроса хранилищу, выявляющего, какие источники-автопарки обслуживают заданные маршрутные пункты. После составления списка этих автобусных парков (т. е. источников информации) исходный запрос преобразуется в серию подзапросов, каждый из которых адресуется через сеть своему источнику в терминах этого источника.

5. Проблемы интеграции информации

В мультибазовой (интегрированной) системе каждый из источников использует специфические термины и таблицы. В общем случае данные от разных источников обычно между собой не согласованы, а сами источники могут работать под управлением различных СУБД. Вследствие этого при обращении к данным нескольких источников через единую программную компоненту возникают проблемы интеграции информации, связанные с наличием или отсутствием каких-либо признаков в источниках, несоответствиями в типах представленных данных и т.д. Рассмотрим более подробно основные проблемы интеграции.

1. *Различия в наименованиях.* Различия в обозначениях данных в разных источниках, несущих одну и ту же нагрузку на концептуальном уровне, т.е. когда одни и те же сущности (или поля) называются разными именами. Например, сущность «Туре» (тип автобуса) может обозначаться как «TypeBus», «Тип», «Категория», поле «Driver» (водитель) как «Voditel», «Водитель» и т.п.

2. *Различия в типах данных.* Например, номера маршрутов в одном источнике хранятся в виде символьных строк переменной длины, в другом – как строки постоянной (причём, возможно разной) длины, в третьем – в формате целочисленных значений.

3. *Различия в множестве допустимых значений.* Один и тот же признак в разных источниках может определяться разными константами. Например, прямой

рейс на одном источнике может обозначаться строкой «пр», а на другом строкой «прямой».

4. *Отсутствующие значения.* В каждом конкретном источнике может не быть такой информации, которая наличествует в других. Например, на некоторых источниках может иметься информация о том, к какому региону и району принадлежит населённый пункт, а на другом нет.

Самый распространённый метод решения проблем интеграции — это поддержка таблиц соответствия, описывающих взаимосвязь элементов данных из различных источников. При интеграции данных с помощью медиатора, представляющего, по сути, виртуальную базу данных (представление для пользователя), эта виртуальная база имеет свои характеристики (реквизиты), которые также заносятся в общую таблицу соответствия.

6. Требования к интегрирующей системе

Помимо традиционной работы с каждой из баз данных-источников на локальном уровне требуется иметь возможность исполнять следующий набор дополнительных транзакций:

– для источников:

- 1) создавать новый маршрут с заполнением необходимых атрибутов;
- 2) возможность обновлять данные, когда это необходимо;
- 3) возможность удалять данные при потере актуальности;

– для медиатора:

- 1) соответствующие подзапросы направлять не ко всем источникам, а только к тем, где имеется запрашиваемая информация; для этой цели создаётся небольшое по объёму хранилище данных, периодически закачивающее от источников информацию обо всех новых маршрутах и рейсах;
- 2) подключать к системе новые источники информации.

Требуется создать рабочие интерфейсы для пользователя и администратора, последний должен иметь возможность отслеживать состояние хранилища и процесс загрузки данных.

7. Описание рабочей модели

Рассмотрим модель из трёх источников данных, которыми являются автобусные парки, выставляющие информацию о своих расписаниях. Будем их условно обозначать DB1, DB2 и DB3. Предполагается, что все источники имеют сходные концептуальные схемы, каждый содержит информацию о своих населённых пунктах, автобусах, маршрутах и рейсах, но могут быть разные наименования таблиц и атрибутов на разных источниках. При этом источник DB2 обслуживает населённые пункты в пределах одного района, источник DB1 — в пределах одного региона (но, возможно, в разных районах), маршруты, относящиеся к источнику DB3, могут заходить в разные регионы.

Выпишем структуру реляционных таблиц на источниках данных в виде: **Таблицы**(поля). Ключевые поля подчёркнуты.

DB1:

Drivers(tab_no, Name);

Buses(Id, reg_num, Id_type);

Bus Type(Id, type, seats, speed);

Points(Id, Name, District);

Routes(Route_no, Id_start_point, Id_end_point);

Waypoints(Route_no, Id_point, Distance_from_start, Time_from_start);

Trips(Id, Route_no, start_time, finish_time, frw/back, week_day);

Sendings(Id, Trip_no, Driver_Id, Bus_Id).

DB2:**Водители**(табельный_номер, ФИО);**Автобусы**(Id, Гос-Номер, вместимость, скорость);**Останов пункты**(Id, Название);**Маршруты**(Номер маршрута, Id_нач_пункт, Id_кон_пункт);**Маршрутные Пункты**(Номер маршрута, Id_пункта, Расст_от_нач_пункта,
Время_движ_от_нач_пункта);**Рейсы**(Id, Номер маршрута, время отправки, время прибытия, Пря-
мой/Обратный, День_недели);**ОтправкаРейсов**(Id, Рейс_Id, Водитель_Id, Автобус_Id).**DB3:****Водители**(таб_номер, ФИО, год_рождения, Id_автоколонна);**Автобусы**(Id, Гос-Номер, год выпуска, Id_тип, автоколонна_номер);**Автоколонны**(Номер, Начальник);**Типы автобусов**(Id, тип, макс_вместимость, сред_скорость)**Ост_пункты**(Id, Название, Район, код_региона);**Маршруты**(Номер маршрута, Id_нач_пункт, Id_кон_пункт);**Маршрутные Пункты**(Маршрут_номер, Id_пункта, Расст_от_старт_точки,
Время_движ_от_нач);**Рейсы**(Id, Номер маршрута, время отправки, время прибытия, Пря-
мой_или_Обратный, День_недели);**ОбеспечениеРейсов**(Id, Рейс_Id, Водитель_Id, Автобус_Id).**Структура медиатора:****Автобусы**(Id, Гос-Номер, вместимость, скорость);**Останов пункты**(Id, Название, Район, код_региона);**Маршруты**(Номер маршрута, Id_нач_пункт, Id_кон_пункт);**Маршрутные пункты**(Номер маршрута, Id_пункта, Расст_от_нач_пункта,
Время_от_нач);**Рейсы**(Id, Номер маршрута, время отправки, время прибытия, Пря-
мой/Обратный, День_недели);**ОбеспечениеРейсов**(Id, Рейс_Id, Автобус_Id).

Примечание. На всех источниках предполагается, что у каждого маршрута есть своя «точка базирования» (начальный пункт), от которой начинается прямой рейс. Обратный рейс начинается от конечного пункта.

8. Таблицы соответствий медиатора

В рассматриваемой рабочей модели каждой таблице из структуры медиатора соответствует, как правило, по одной таблице каждого из источников. Исключение составляет следующая ситуация: таблице «Автобусы» медиатора на источниках DB1 и DB3 соответствуют по две таблицы, соответственно, «Buses», «Bus Type» и «Автобусы», «Типы автобусов». Кроме того, на источнике DB1 отсутствует информация о регионах (предполагается, что для всех маршрутов регион один и тот же), а на источнике DB2 по той же причине отсутствует информация о районах и регионах. В связи с этим представляется целесообразным поддержание на медиаторе одной общей таблицы соответствий медиатору для всех источников (табл. 1), а также таблицы соответствий источников районам, регионам и сетевым адресам источников (табл. 2).

Таблица 1

Tabl_Corr: соответствие источников медиатору

Медиатор. таблица	Медиатор. атрибут	DB1. таблица	DB1. атрибут	DB2. таблица	DB2. атрибут	DB3. таблица	DB3. атрибут
Автобусы	Id	Buses	Id	Автобусы	Id	Автобусы	Id
Автобусы	Гос-Номер	Buses	reg_num	Автобусы	Гос-Номер	Автобусы	Гос-Номер
Автобусы	вместимость	Bus Type	seat	Автобусы	вместимость	Типы автобусов	макс_вместимость
Автобусы	скорость	Bus Type	speed	Автобусы	скорость	Типы автобусов	сред_скорость
Останов пункты	Id	Points	Id	Останов пункты	Id	Ост_пункты	Id
Останов пункты	Название	Points	Name	Останов пункты	Название	Ост_пункты	Название
Останов пункты	Район	Points	District	NULL	NULL	Ост_пункты	Район
Останов пункты	код_региона	NULL	NULL	NULL	NULL	Ост_пункты	код_региона
Маршруты	Номер маршрута	Routes	Route_no	Маршруты	Номер маршрута	Маршруты	Номер маршрута
Маршруты	Id_нач_пункт	Routes	Id_start_point	Маршруты	Id_нач_пункт	Маршруты	Id_нач_пункт
Маршруты	Id_кон_пункт	Routes	Id_end_point	Маршруты	Id_кон_пункт	Маршруты	Id_кон_пункт
Маршрутные Пункты	Номер маршрута	Waypoints	Route_no	Маршрутные Пункты	Номер маршрута	Маршрутные Пункты	Маршрут_номер
Маршрутные Пункты	Id_пункта	Waypoints	Id_point	Маршрутные Пункты	Id_пункта	Маршрутные Пункты	Id_пункта
Маршрутные Пункты	Расст_от_нач_пункта	Waypoints	Distance_from_start	Маршрутные Пункты	Расст_от_нач_пункта	Маршрутные Пункты	Расст_от_старт_точки

Маршрутные Пункты	Время_от_нач	Waypoints	Time_from_start	Маршрутные Пункты	Время_движ_от_пункта	Маршрутные Пункты	Время_движ_от_нач
Рейсы	Id	Trips	Id	Рейсы	Id	Рейсы	Id
Рейсы	Номер маршрута	Trips	Route_no	Рейсы	Номер маршрута	Рейсы	Номер маршрута
Рейсы	время отправки	Trips	start_time	Рейсы	время отправки	Рейсы	время отправки
Рейсы	время прибытия	Trips	finish_time	Рейсы	время прибытия	Рейсы	время прибытия
Рейсы	Прямой/ Обратный	Trips	frw/back	Рейсы	Прямой/ Обратный	Рейсы	Прямой или Обратный
Рейсы	День_недели	Trips	week_day	Рейсы	День_недели	Рейсы	День_недели
Обеспечение Рейсов	Id	Sendings	Id	Отправка Рейсов	Id	Обеспечение Рейсов	Id
Обеспечение Рейсов	Рейс_Id	Sendings	Trip_no	Отправка Рейсов	Рейс_Id	Обеспечение Рейсов	Рейс_Id
Обеспечение Рейсов	Автобус_Id	Sendings	Bus_Id	Отправка Рейсов	Автобус_Id	Обеспечение Рейсов	Автобус_Id

Таблица 2

Source_Regions: соответствие источников районам, регионам и сетевым адресам источников

IdSource	Адрес в сети	Район	Регион
DB1	...	NULL	...
DB2
DB3	...	NULL	NULL

9. Подключение нового источника

Эту процедуру выполняет специальный человек. Рассмотрим для примера новый источник DB4. Процедура подключения состоит в следующем: в таблице соответствия «Tabl_Corr» заводятся новые колонки, а в таблице «Source_Regions» – новая строка, соответствующие вновь подключённому источнику. Производится соответствующее обновление хранилища при медиаторе: оно дополняется маршрутными пунктами, обслуживаемыми новым источником. Вся дальнейшая работа мультибазовой системы (запросы, обновление данных в хранилище) производится уже с учётом нового источника.

10. Структура вспомогательного хранилища и его обновление

Хранилище состоит из одной таблицы «ПунктыНаМаршрутах» с полями: IdSource, IdПункта, ПунктНазвание, Район, Код_региона. Первичным ключом в ней является пара: IdSource, IdПункта.

Если на источнике с номером n происходит добавление, удаление, переименование пункта как места остановки или изменение административных границ, то на медиатор подаётся специальный сигнал, по которому запускается транзакция, переносящая все обновления источника на таблицу-хранилище «ПунктыНаМаршрутах». Обновляются строки с IdSource= n . В случае, когда на источнике отсутствует информация о регионах и районах, соответствующие значения берутся из таблицы Source_Regions.

11. Схема исполнения запроса на медиаторе

Рассмотрим для примера наиболее популярный запрос: показать рейсы всех маршрутов, следующих из Пункта1 в Пункт2, со временем отправления из Пункта1. Предусматривается возможность фильтрации запроса по дням недели и времени отправления. Время отправления от Пункта1 вычисляется как сумма времени отправки от стартового пункта и времени движения от стартового пункта до Пункта1. SQL-код запроса в этом случае может выглядеть следующим образом (для прямого рейса):

```
SELECT Маршруты.[Номер маршрута] AS Маршрут,
       ОП1.Название AS [Пункт отправления],
       ОП2.Название AS [Пункт назначения],
       Рейсы.[время отправки]+МП1.Расст_от_нач_пункта AS [Время отправления]
FROM Маршруты, Рейсы,
```



```

[Останов пункты] AS ОП1, [Останов пункты] AS ОП2,
[Маршрутные пункты] AS МП1, [Маршрутные пункты] AS МП2

WHERE МП1.Id\_пункта=Id\_Пункта1 AND МП2.Id\_пункта=Id\_Пункта2 AND
МП1.[Номер маршрута]=МП2.[Номер маршрута] AND
МП1.Расст\_от\_нач\_пункта < МП2.Расст\_от\_нач\_пункта AND
Маршруты.[Номер маршрута]=Рейсы.[Номер маршрута] AND
Маршруты.[Номер маршрута]=МП1.[Номер маршрута] AND
Маршруты.Id\_нач\_пункт=ОП1.Id AND
Маршруты.Id\_кон\_пункт=ОП2.Id

```

Предварительно должны быть найдены `Id_Пункта1` и `Id_Пункта2` в таблице «Останов пункты». Технология может быть, например, такой: при запуске запроса пользователь задаёт названия каждого из пунктов, при этом по мере набора ему выплывают подсказки в виде списка, из которого щелчком мыши надо выбрать нужный пункт. При этом щелчке система автоматически выбирает из таблицы «Останов пункты» соответствующий `Id`, который и передается в запрос.

Запрос может также учитывать и другие введённые параметры о маршруте и рейсе (например, диапазон времени отправки, день отправки, тип и гос. номер автобуса), требующие в некоторых случаях присоединения таблиц «ОбеспечениеРейсов» и «Автобусы».

Для обратного рейса имеет место аналогичная формула, только вместо времени движения от начального пункта маршрута до `Пункта1` потребуется предварительно вычислить время движения от конечного пункта до `Пункта1`, по имеющимся данным таблицы «Маршрутные пункты».

Опишем технологию исполнения данного запроса.

Пользователь мультибазовой системы вводит в форму-интерфейс критерии поиска: пункт отправки, пункт прибытия, время отправки, время прибытия. На первом этапе запрос адресуется таблице «ПунктыНаМаршрутах» хранилища. Отсеиваются источники, на которых не обслуживают одновременно `Пункт1` и `Пункт2`:

```

SELECT Source
FROM ПунктыНаМаршрутах
WHERE <условия соответствия данным Пункта2>
AND Source IN (SELECT Source
FROM ПунктыНаМаршрутах
WHERE <условия соответствия данным Пункта1>)

```

Результаты записываются во временную таблицу «Sources».

На втором этапе происходит обращение с исходным запросом к каждому из найденных на первом этапе источникам (по сетевым адресам, указанным в таблице «Source_Regions»). При этом для каждого источника SQL-код запроса преобразуется к виду, который данный источник способен воспринять. Преобразование запроса производится с помощью таблиц соответствия.

На третьем этапе медиатор получает ответы от источников и производит их обратное преобразование к структуре медиатора. После этого каждый ответ дополняется полем с информацией об источнике (автопарке), все ответы объединяются и выдаются пользователю.

12. Особые ситуации

Итак, описана общая схема построения мультибазовой системы, интегрирующей информацию из автобусных парков. Однако при каждой конкретной реализации могут возникать особые ситуации, требующие отдельного изучения. Рассмотрим для примера две из них.

Ситуация первая. Неоднозначная идентификация информации. Не все источники могут поддерживать атрибуты, однозначно определяющие маршрут. Например, на разных источниках номера маршрутов могут совпадать, а пути их следования могут оказаться совершенно разными, и как следствие возникает конфликт именования на уровне медиатора. Он может разрешиться, если мы в качестве первичного ключа для маршрута (на уровне медиатора) возьмём пару: номер маршрута, номер перевозчика (источника).

Ситуация вторая. В некоторых регионах (например, в Архангельской области или в Якутии) у населённых пунктов могут встречаться двойные (а иногда и тройные) названия. Например, одно официальное (обычно по наименованию почтового отделения), а другое — как называют это место сами жители. Чтобы система могла принимать все варианты наименований, можно поддерживать при медиаторе дополнительные таблицы, содержащие список населённых мест (с указанием соответствия остановочным пунктам) и вариантами их названий.

13. Заключение

Предложен и описан алгоритм решения задачи интеграции автобусных парков, использующий технологию медиаторов с участием вспомогательного хранилища небольшого объёма, с помощью которого идентифицируются необходимые источники данных. Этот подход, позволяет использовать в сочетании преимущества медиаторов и хранилищ. Описана рабочая модель интегрирующей системы, состоящей из трёх источников данных с различной структурой. Проиллюстрирована технология исполнения запроса к медиатору. Прописана технология подключения нового источника. Представляется, что предложенная схема может быть применима и в других задачах, связанных с интеграцией информации. Для рассматриваемой задачи дальнейшим развитием мог бы служить выбор оптимального маршрута на основе подхода работы [6].

Литература

1. Воробьёва М. С. Построение модели интеграции данных в информационно-управляющих системах // Модернизация образования в условиях глобализации: Круглый стол «Образование через науку и инновации», 14–15 сентября 2005 г. / под ред. В. Н. Кутрунов. — 2005. — С. 26–28.
2. Торшин Д. В. Организация единого интегрированного пространства на основе универсального формата обмена данными // Научно-технические ведомости СПбГПУ, серия «Информатика. Телекоммуникации. Управление». — 2009. — № 2(71). — С. 26–32.
3. Панкратов А. С. Технология интеграции разнородных баз данных на примере объединения электронных медицинских карт // Искусственный интеллект и принятие решений. — 2011. — № 4. — С. 60–67.
4. Валуев А. М., Панкратов А. С. Современные технологии интеграции информации из независимых баз данных и возможности их применения в задачах планирования и управления // Горный информационно-аналитический бюллетень (научно-технический журнал). — 2013. — № 7. — С. 170–174.
5. Гарсия-Моллина Г., Ульман Д., Уидом Д. Системы баз данных. Полный курс. — Москва: Вильямс, 2003. — 1088 с.
6. Валуев А. М., Якушнов А. А. Задача выбора рационального маршрута с использованием общественного транспорта // Управление развитием крупномасштабных систем: Материалы Шестой международной конференции. — Т. II. — Москва: 2012. — С. 31–33.

UDC 004.65:004.75

DOI: 10.22363/2312-9735-2017-25-3-295-305

Modern Technologies of Information Integration from Independent Sources and their Application in the Construction of an Information System that Combines Transport Timetables

A. S. Pankratov, A. Kh. Psheunov

*Information Technologies Department
Peoples' Friendship University of Russia (RUDN University)
6 Miklukho-Maklaya St., Moscow, 117198, Russian Federation*

The problem of constructing an integrating system combining bus timetables of different bus depots is considered. We suppose that these bus depots are independent and, possibly, located at different regions of the country. The purpose of this integrated system is, in particular, the easement of finding available bus routes between two given points. To solve this problem, it is proposed to use an approach combining the advantages of mediator and data repository technologies. The article considers a model of three data sources, which are bus depots that display information about their timetables. It is assumed that all sources have similar conceptual schemes, but they have their own specific characteristics. In particular, there may be different names of tables and attributes in different sources and different distribution of attributes throw tables. Also at some sources may be an absence of certain attributes. We construct correspondence tables and a mediator that translate user queries to the sources. To identify the necessary sources, a small auxiliary repository is maintained that contains information about the stop points served by each of the sources. We describe the technology for updating the repository and the executing strategy for the user query, using the information contained in the repository.

Key words and phrases: database, integration of the information, multidatabase system, mediator, data warehouse

References

1. M. S. Vorobyova, Development of the Model of Data Integration in Information and Control Systems, in: V. N. Kuturnov (Ed.), Modernization of education in the context of globalization: Round-table discussion "Education through Science and Innovation", 2005, sept. 14–15, 2005, pp. 26–28, in Russian.
2. D. V. Torshin, Organization of an United, Integrated Space Based on Data Exchange Universal Format, Scientific and Technical Sheets of St-Petersburg Polytechnic University, Series "Information Sciences. Telecommunications. Management" (2(71)) (2009) 26–32, in Russian.
3. A. S. Pankratov, Integration Technology of Heterogeneous Databases Exemplified by Consolidation of Electronic Medical Cards, Artificial Intelligence and Decision Making (4) (2011) 60–67, in Russian.
4. A. M. Valuev, A. S. Pankratov, Modern Technologies for Information Integration from Independed Databases and Ways for Application in the Problems of Planning and Management, Mining Informational and Analytical Bulletin (scientific and technical journal) (7) (2013) 170–174, in Russian.
5. H. Garsia-Molina, J. Ullman, J. Widom, Database Systems: the Complete Book, Williams, 2003.
6. A. M. Valuev, A. A. Yakunkhov, The Task of Choosing the Rational Route using Public Transport, in: Managing the Development of Large-Scale Systems: Proceedings of the Sixth International Conference, Vol. II, Moscow, 2012, pp. 31–33, in Russian.