

Технология сбора и анализа реляционных данных¹

Цель исследования. В научно-образовательных учреждениях для оперативного сбора информации и управления их научно-организационной деятельностью традиционно используется электронная почта для получения электронных таблиц Microsoft Excel или документов Microsoft Word. Недостатками этого подхода можно отнести отсутствие контроля корректности ввода данных, сложность обработки полученной информации из-за не реляционной модели данных и т.д. Существуют Интернет сервисы, позволяющие организовать сбор данных в реляционном виде. Недостатками этих систем являются: отсутствие работы с таблицами справочниками; ограниченный набор элементов управления ввода данных; ограниченные возможности управления работой формы ввода; большая часть систем является условно бесплатными и т.д. Таким образом, можно отметить актуальность разработки Интернет технологии сбора и анализа данных, ориентированная на обычного пользователя, которая должна позволять оперативно определять модель собираемых данных, автоматически реализовывать ввод данных в соответствии с этой моделью и проводить контроль вводимых данных.

Материалы и методы. В рамках статьи описывается разработанная и апробированная технология оперативного сбора и анализа данных, использующая систему ввода и редактирования реляционных данных «Фарамант». Работа системы «Фарамант» основывается на модели документа, которая включает три компонента: описание структуры данных; отображение данных; логика работы формы. Все этапы технологии выполняются пользователем с помощью браузера. Основным этапом предложенной технологии является определение модели данных в виде совокупности реля-

ционных таблиц. Для создания таблицы в рамках системы нужно определить название и список полей. Для каждого поля необходимо указать его название и используемый элемент управления для ввода данных и логику его работы. Элементы управления позволяют организовать корректный ввод данных в зависимости от типа данных. Для созданной модели система «Фарамант» автоматически создает форму заполнения, используя которую пользователи вносят информацию. Для изменения вида формы используется шаблон формы. Полученные данные можно просмотреть постранично в виде таблицы. К строкам таблицы применяются различные фильтры. Для обобщения информации реализован механизм группировки данных, который позволяет получить данные о количестве записей, максимальных, минимальных, средних значениях по различным группам записей.

Результаты. Данная технология апробирована на проведении мониторинга требований к услугам дополнительного профессионального образования и определения образовательных потребностей педагогических и руководящих работников образовательных организаций Иркутской области. В опросе приняли участие 2780 респондентов из 36 муниципальных образований. На создание модели данных затрачено несколько часов. Опрос проводился в течение месяца.

Заключение. Предложенная технология позволяет в короткие сроки провести сбор информации в реляционном виде, а затем ее анализ без необходимости программирования с возможностью гибкого задания логики работы формы.

Ключевые слова: WEB сервисы, Базы данных, дескриптивные спецификации.

Helen N. Fedorova¹, Gennadiy M. Ruzhnikov², Roman K. Fedorov³

¹GAU DPO «Institute of education development of the Irkutsk region», Irkutsk, Russia

²Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences, Irkutsk, Russia

³The Irkutsk scientific center of Siberian Branch of Russian Academy of Sciences, Irkutsk, Russia

Technology for collecting and analyzing relational data

Purpose of the study. The scientific and educational organizations use traditionally e-mail with Microsoft Excel spreadsheets and Microsoft Word documents for operational data collection. The disadvantages of this approach include the lack of control of the correctness of the data input, the complexity of processing the information received due to non-relational data model, etc. There are online services that enable to organize the collection of data in a relational form. The disadvantages of these systems are: the absence of thesaurus support; a limited set of elements of data input control; the limited control the operation of the input form; most of the systems is shareware, etc. Thus, it is required the development of Internet data collection and analysis technology, which should allow to identify quickly model the data collected and automatically implement data collection in accordance with this model. *Materials and methods.* The article describes the technology developed

and tested for operational data collection and analysis using "Faramant" system. System operation "Faramant" is based on a model document, which includes three components: description of the data structure; visualization; logic of form work. All stages of the technology are performed by the user using the browser. The main stage of the proposed technology is the definition of the data model as a set of relational tables. To create a table within the system it's required to determine the name and a list of fields. For each field, you must specify its name and use the control to the data input and logic of his work. Controls are used to organize the correct input data depending on the data type. Based on a model system "Faramant" automatically creates a filling form, using which users can enter information. To change the form visualization, you can use the form template. The data can be viewed page by page in a table. For table rows, you can apply different filters. To summarize the information there is a

¹ Работа выполнена при поддержке центра коллективного пользования ИИВС ИРНОК, программы ИНЦ СО РАН и грантов РФФИ, № 14-07-00166, 16-07-00411-а, 16-37-00110 мол_а, 16-57-4034 Монг_а.

mechanism of data grouping, which provides general data of the number of entries, maximum, minimum, average values for different groups of records.

Results. This technology has been tested in the monitoring requirements of the services of additional professional education and the definition of the educational needs of teachers and executives of educational organizations of the Irkutsk region. The survey has involved 2,780 respondents in 36

municipalities. Creating the data model took several hours. The survey was conducted during the month.

Conclusion. The proposed technology allows a short time to collect the information in relational form, and then analyze it without the need for programming with flexible assignment of the operating logic for form.

Keywords: Web services, Databases, descriptive specifications.

Введение

В научно-образовательных учреждениях для оперативного сбора информации и управления их научно-организационной деятельностью традиционно используется электронная почта для получения электронных таблиц Microsoft Excel или документов Microsoft Word. Недостатками этого подхода можно отнести отсутствие контроля корректности ввода данных, сложность обработки полученной информации из-за не реляционной модели данных и т.д. Известно, что использование реляционной модели сбора данных позволяет применять готовые функции агрегации, фильтрации и отображения данных, что значительно упрощает их анализ. Существуют Интернет сервисы, позволяющие организовать сбор данных в реляционном виде. Например, Google forms [1], SurveyMonkey [2], Survio [3] и т.д. Недостатками этих систем являются: отсутствие работы с таблицами справочниками; ограниченный набор элементов управления ввода данных; ограниченные возможности управления работой формы ввода; большая часть систем является условно бесплатными и т.д. Кроме того также существуют нормативные документы, ограничивающие использование зарубежных систем в органах государственной власти и местного самоуправления.

Также для решения задачи сбора данных можно создавать специализированные информационные системы (ИС) с использованием модельно-ориентированного подхода (Model Driven Architecture) [4], который основан на построении абстрактной метамодели управления и обмена метаданными (моделями) и задании способов ее трансформации в поддерживаемые технологии программирования.

Применение MDA, в свою очередь, требует соответствующей квалификации пользователей, знания языков программирования и моделирования. Одной из альтернатив к использованию подхода MDA к разработке ИС является интерпретация модели непосредственно в пользовательские интерфейсы [5,6], где модель задается на некотором текстовом языке, знание которого необходимо.

Таким образом, можно отметить актуальность разработки технологии сбора и анализа данных, которая должна оперативно определять модель собираемых данных, автоматически реализовывать ввод данных в соответствии с этой моделью и проводить контроль вводимых данных. Данная технология должна быть ориентирована на обычного пользователя.

В Институте динамики систем и теории управления Сибирского отделения РАН (ИДСТУ СО РАН) совместно с Институтом развития образования Иркутской области (ИРО) разработана и апробирована технология сбора и анализа данных, использующая систему ввода и редактирования реляционных данных «Фарамант». Работа системы «Фарамант» основывается на модели документа (см. рис. 1), которая включает три компонента: описание структуры данных; отображение данных; логика работы формы.

На основе модели документа создается таблица в СУБД PostgreSQL, формируется форма ввода и отображение данных в виде таблицы. Все этапы технологии выполняются пользователем с помощью браузера.

Технология включает следующие этапы:

- 1) определение модели данных;
- 2) импорт данных в созданные таблицы;
- 3) разработка форм ввода, определение логики;
- 4) анализ введенных данных.

Разработанная технология апробирована на создании ИС сбора и мониторинга заявок на услуги дополнительного профессионального образования и определении потребностей образовательных педагогических и руководящих работников образовательных организаций Иркутской области для составления план-графика образовательных услуг ИРО. Требовалось создать перечень востребованности обучения по дополнительным профессиональным программам повышения квалификации и профессиональной переподготовки, времени проведения, организациях респондентов, форме обучения и т.д. Рассмотрим подробнее этапы технологии, с использованием системы ввода и редактирования реляционных данных «Фарамант».

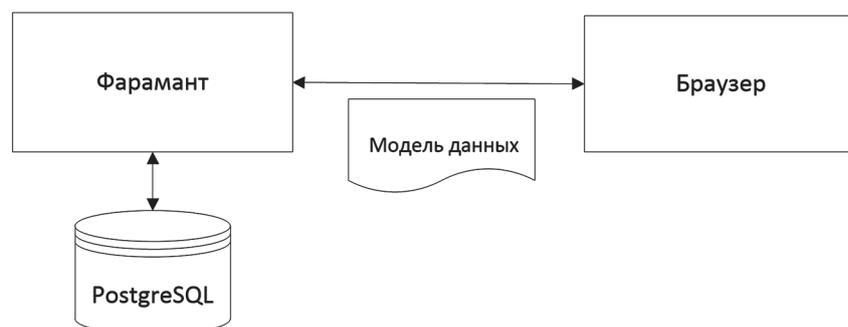


Рис. 1. Архитектура работы системы «Фарамант»

Определение модели данных

Первым этапом предложенной технологии является определение модели данных в виде совокупности реляционных таблиц. Первоначально требуется спроектировать базу данных в третьей нормальной форме. Для создания таблицы в рамках системы нужно определить название и список полей (см. рис. 2).

Для каждого поля необходимо указать его название и используемый элемент управления (ввода). За каждым элементом управления закреплен определенный тип атрибута базы данных. Элементы управления позволяют организовать корректный ввод данных в зависимости от типа данных. Например, при выборе элемента управления «Дата» пользователь может вводить дату с помощью календаря, при выборе «Текст» создается многострочное поле ввода. При проектировании базы данных часто производится создание таблиц справочников. Для ввода атрибутов, ссылающихся на таблицы без иерархической зависимости, применяется элемент управления «classify». Данный элемент управления позволяет использовать в качестве справочника любую таблицу, зарегистрированную в каталоге. Например, в представленной задаче «classify» используется для выбора формы обучения.

После создания таблицы определяются права доступа. Первоначально к таблице имеет доступ только владелец. Имеется три группы пользователей: аноним, любой зарегистрированный в системе пользователь и конкретные пользователи. Для каждой группы можно разрешить следующие операции: просмотр своих данных, просмотр всех записей, редактирование своих записей, редактирование всех записей.

Загрузка данных

До ввода данных таблицы справочники должны быть заранее заполнены в системе. В рассматриваемой задаче это категории слушателей, дополнительные профессиональные программы, реестр

Рис. 2. Создание таблицы

образовательных организаций и т.д. Реестр организаций составляет более тысячи записей. Поэтому ввод в ручном режиме трудоемок. Разработан специальный модуль для загрузки данных в таблицы в формате csv. Формат csv поддерживается разными системами, в том числе Microsoft Excel. Модуль производит разбор текстового файла в формате csv, извлекает названия атрибутов и данные. Пользователь может указать кодировку данных и используемый разделитель. Далее пользователь должен сопоставить атрибуты результирующей таблицы и импортируемой таблицы из текстового файла. В процессе загрузки можно обработать импортируемые значения. Пользователь

должен выбрать функцию конвертации для каждого атрибута. На текущий момент реализованы следующие функции:

1) Copy – производит копирование без обработки значения из указанного атрибута импортируемой таблицы в атрибут результирующей таблицы. Если атрибут результирующей таблицы является ссылкой на таблицу справочник, то производится поиск соответствующего значения в таблице справочник и берется значение первичного ключа;

2) CopyDate – производит конвертацию даты из текстового формата в формат базы данных. Производится распознавание используемого формата даты импор-

Рис. 3. Шаблон формы ввода анкеты формируется в WYSIWYG редакторе

тируемого атрибута. Если формат не распознал, то пользователь может указать его вручную;

3) `CopyValue` – используется для того чтобы при загрузке в указанном атрибуте всех импортируемых записей было присвоено значение-константа;

4) `CopyPoint` – используется для определения атрибута типа «Точка» (геокодирование). На входе метода указывается два атрибута, содержащие широту и долготу точки.

Разработка форм ввода

После создания таблицы система «Фарамант» автоматически создает форму заполнения. Для ввода каждого атрибута используются указанные элементы управления. Элементы управления отображаются в порядке следования атрибутов в таблице. Для изменения вида формы (дизайна) требуется создать шаблон формы. Шаблон формы – это HTML код с указанием положения вставки элементов управления для ввода атрибутов. Создается шаблон в специальном редакторе или блокноте (см. рис. 3).

Указание положения элементов управления атрибутов осуществляется с помощью специальных тегов вида «`#fieldname#`», где `fieldname` это имя атрибута в БД, либо с помощью HTML контейнеров `<div>`, у которых идентификатор совпадает с именем атрибута. Применение контейнеров `<div>` позволяет управлять отображением заголовков полей.

Для упрощения заполнения формы ввод данных должен подчиняться определенной логике. Так возможные значения атрибутов могут зависеть от других. Задание логики работы формы осуществляется с помощью задания зависимостей между атрибутами и управлением видимостью элементов управления атрибутов. В нашей задаче пользователь должен указать организацию, в которой он работает. Учитывая количество организаций, эффективнее чтобы пользователь указал муниципальное образование и организационную форму организации, и только



Рис. 4. Фильтрация данных. Овалами выделены значения, по которым производится фильтрация соответствующих атрибутов



Рис. 5. Условие видимости элемента управления атрибута

затем выбрал организацию из отфильтрованного списка. Реализует зависимость элемент управления «classify». Для организации фильтрации текущего элемента управления «classify» необходимо указать, по каким атрибутам таблицы справочника производить фильтрацию, и какие значения используются для фильтрации (см. рис. 4).

Если в качестве значения указывается «`#fieldname`», то значение для фильтрации автоматически берется из соответствующего атрибута. На Рис. 4 представлен атрибут, определяющий организацию. Для этого атрибута используется таблица справочник «Реестр организаций». Фильтрация производится по атрибутам «Муниципалитет» и «orgform». Соответственно в выпадающем списке будут отображаться только из определенного муниципального образования и определенной организационной формы. Их идентификаторы автоматически берутся из соответствующих элементов управления. При изменении муниципального образования список организаций автоматически обновляется. Пока значения фильтрации не определены, т.е. пользователь не указал значения атрибутов, элемент управления не доступен для ввода.

Следующей возможностью управления логикой работы формы является управление видимостью элементов управления атрибутов. Например, при отсутствии соот-

ветствующей программы, пользователь должен указать желаемое название программы обучения. Для хранения программы обучения создан атрибут «`program_name`», используется для него элемент управления «Текст». В свойстве элемента управления можно задать условие видимости в поле «`Visibility`» (см. рис. 5).

В примере указано условие: `section==-1 || program==-1 || retrain==-1`, что означает, что элемент управления отображается, если хотя бы в одном из атрибутов `section`, `program`, `retrain` указано «Другое».

Условие выражается в соответствии с синтаксисом языка JavaScript. Может содержать только имена атрибутов, константы и специальные символы (например `=`). При изменении значения любого атрибута производится проверка условия видимости. Если в шаблоне формы для атрибута задан `<div>`, то в этом случае можно скрыть атрибут вместе заголовком.

Анализ данных

Полученные данные можно просмотреть постранично в виде таблицы. К таблице применяются различные фильтры, в соответствии с которыми отображаются строки таблицы. Для строковых атрибутов можно определить ус-

Муниципалитет	Форма	Наименование	Адрес
----		МБОУ	
МБОУ "Морозовская ООШ"	Общеобразовательная	МБОУ	669318, Иркутская
МБОУ г. Иркутска СОШ № 27	Общеобразовательная	МБОУ г. Иркутска	664046, г. Иркутск,
МБОУ ЦО "Альянс"	Общеобразовательная	МБОУ ЦО	п. Харик, ул. Ленина,
МБОУ Баяндаевская СОШ	Общеобразовательная	МБОУ	С. Баяндай,
МБОУ Аларская СОШ	Общеобразовательная	МБОУ Аларская	669450, обл.
МБОУ Александровская СОШ	Общеобразовательная	МБОУ	669457, обл.
МБОУ Алятская СОШ	Общеобразовательная	МБОУ Алятская	669473, обл.
МБОУ Ангарская СОШ	Общеобразовательная	МБОУ Ангарская	669459, обл.
МБОУ Бахтайская СОШ	Общеобразовательная	МБОУ Бахтайская	669468, обл.
МБОУ Гаханская СОШ	Общеобразовательная	МБОУ Гаханская	с.Бадагуй, Урбаева,

Рис. 6. Фильтрация данных

Муниципалитет	Форма	Наименование	Адрес
count	group	count	count
347	Общеобразовательная	347	347
2	Организация	2	2
5	Дошкольная	5	5

Рис. 7. Формирование отчета по количеству организаций по юридической форме

ловие, при котором отображаются записи, значение выбранного атрибута содержит указанную подстроку. Для атрибутов типа Date указывается диапазон дат, при этом будут отображаться все записи со значениями дат из указанного диапазона. Для числовых атрибутов указывается условие с использованием разных знаков сравнения. Для атрибутов, ссылающихся на таблицы справочники, производится фильтрацию, используя значения из таблицы справочника. Если заданы условия для нескольких атрибутов, то они все объединяются в запросе по логическому «И». Для атрибута можно определить несколько значений, по которому производится фильтрация. Все указанные значе-

ния формируются в запросе по логическому «ИЛИ» (см. рис. 6).

Для обобщения информации реализован механизм группировки данных, который позволяет получить данные о количестве записей, максимальных, минимальных, средних значениях по различным группам записей. В пользовательском интерфейсе для группировки записей необходимо выбрать поля, по оставшимся полям требуется выбрать групповую функцию (max, min, avg, count). Группировка записей производится для всех типов атрибутов по равенству их значений. Для полей типа «Date» можно группировать записи по неделям, месяцам, кварталам, годам и десятилетиям. При обобщении информации использу-

ются заданные пользователем фильтры. Данные таблицы выгружаются в формате CSV (см. рис. 7).

Апробация

Данная технология апробирована на проведении мониторинга требований к услугам дополнительного профессионального образования (ДПО) и определения образовательных потребностей педагогических и руководящих работников образовательных организаций Иркутской области. В опросе приняли участие 2780 респондентов из 36 муниципальных образований (МО).

Заключение

Предложенная технология позволяет в короткие сроки провести сбор информации в реляционном виде, а затем ее анализ. Недостатком технологии является то, что пользователь ограничен вводом только одной записи в таблицу, но проведенная апробация подтверждает ее работоспособность и функциональность. Достоинствами разработанной технологии являются:

- наличие широкого набора элементов управления, позволяющих упростить ввод различных типов данных с применением динамической загрузки данных (AJAX);
- возможность задания логики с помощью определения зависимостей между атрибутами и управления видимостью;
- гибкий механизм задания дизайна формы;
- наличие методов загрузки и выгрузки данных;
- наличие развитых методов фильтрации и обобщения данных;
- высокая скорость разработки форм без необходимости программирования.

Литература

1. Google forms https://www.google.com/intl/ru_ru/forms/about/ (дата обращения 09.06.2016).
2. SurveyMonkey <https://ru.surveymonkey.com/> (дата обращения 09.06.2016).
3. Survio <http://www.survio.com/ru/> (дата обращения 09.06.2016).
4. Frankel D. Model Driven Architecture: Applying MDA to Enterprise Computing. Willey, 2003. – P. 352.

References

1. Google forms https://www.google.com/intl/ru_ru/forms/about/ (Accessed 09.06.2016).
2. SurveyMonkey <https://ru.surveymonkey.com/> (Accessed 09.06.2016).
3. Survio <http://www.survio.com/ru/> (Accessed 09.06.2016).
4. Frankel D. Model Driven Architecture: Applying MDA to Enterprise Computing. Willey, 2003. – 352 P.

5. Фереферов Е.С., Бычков И.В., Хмельнов А.Е. Технология разработки приложений баз данных на основе декларативных спецификаций // Вычислительные Технологии. – 2014. – Т. 19, № 5. – С. 85–100.

6. Инфраструктура информационных ресурсов и технологии создания информационно-аналитических систем территориального управления / Бычков И. В., Ружников Г. М., Хмельнов А. Е., Фёдоров Р. К., Парамонов В. В., Шигаров А. О., Фереферов Е. С., Гаченко А. С., Михайлов А. А., Маджара Т. И., Шумилов А. С., Авраменко Ю. В. : Издательство СО РАН, 2016. – 242 с.

5. Fereferov E.S., Bychkov I.V., Khmel'nov A.E. Tekhnologiya razrabotki prilozhenii baz dannykh na osnove deklarativnykh spetsifikatsii // Vychislitel'nye Tekhnologii. – 2014. – T. 19, № 5. – Pp. 85–100. (in Russ.)

6. Infrastruktura informatsionnykh resursov i tekhnologii sozdaniya informatsiono-analiticheskikh sistem territorial'nogo upravleniya / Bychkov I.V., Ruzhnikov G.M., Khmel'nov A.E., Fedorov R.K., Paramonov V.V., Shigarov A.O., Fereferov E.S., Gachenko A.S., Mikhailov A.A., Madzhara T.I., Shumilov A.S., Avramenko Yu.V. : Izdatel'stvo SO RAN, 2016. –242 P. (in Russ.)

Сведения об авторах

Елена Николаевна Фёдорова,

начальник учебной части

ГАУ ДПО «Институт развития образования Иркутской области», Иркутск, Россия

Эл. почта: e.fedorova@iro38.ru

Тел.: (3952) 500-904, доб. 230

Геннадий Михайлович Ружников,

заведующий отделением

Институт динамики систем и теории управления

имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Эл. почта: rugnikov@icc.ru

Тел.: (3952) 453-006

Роман Константинович Фёдоров,

ведущий научный сотрудник

Иркутский научный центр Сибирского отделения

Российской академии наук, Иркутск, Россия

Эл. почта: fedorov@icc.ru

Тел.: (3952) 453-108

Information about the authors

Elena N. Fedorova,

Head of Studies

Institute of education development of the Irkutsk region, Irkutsk, Russia

E-mail: e.fedorova@iro38.ru

Tel.: (3952) 500-904, additional 230

Gennadiy M. Ruzhnikov,

Head of department

Matrosov Institute for System Dynamics and Control Theory of

Siberian Branch of Russian Academy of Sciences,

Irkutsk, Russia

E-mail: rugnikov@icc.ru

Tel.: (3952) 453-006

Roman K. Fedorov,

Leading Researcher

The Irkutsk scientific center of Siberian Branch of Russian

Academy of Sciences, Irkutsk, Russia

E-mail: fedorov@icc.ru

Tel.: (3952) 453-108