
РАБОТА С АНАЛИТИЧЕСКИМИ ОТЧЁТАМИ В ИССЛЕДОВАТЕЛЬСКОМ ПОРТАЛЕ “ИННОВАЦИОННОЕ РАЗВИТИЕ РЕГИОНОВ”

Павел Мальцев

Аннотация: Рассмотрена архитектура и базовые концепции подсистемы представления отчётов в программном комплексе ViP. Программный комплекс ViP предназначен для многомерного анализа данных, получаемых из гетерогенных источников. Он позволяет упростить разработку приложений Business Intelligence. Программный комплекс ViP включает набор готовых компонентов, которые могут быть использованы при разработке приложений. В частности, в комплексе ViP реализована подсистема представления отчётов. В данной работе приводится описание этой подсистемы и пример её использования при разработке исследовательского портала “Инновационное развитие регионов”.

Keywords: Business Intelligence, BI, бизнес-анализ, OLAP, Reporting, системы поддержки принятия решений, DSS, информационно-аналитические системы.

ACM Classification Keywords: H.4 Information Systems Applications: H.4.2 Types of Systems – Decision support (e.g., MIS).

Conference: The paper is selected from Seventh International Conference on Information Research and Applications – i.Tech 2009, Varna, Bulgaria, June-July 2009

Введение

Проект создания портала «Инновационного развития регионов» решает задачу разработки информационно-аналитической системы, реализующей сбор, хранение, предоставление и анализ данных об инновационной активности регионов. При реализации данного проекта одно из важнейших мест занимает задача визуализации данных. Эта задача заключается в представлении данных в удобной для восприятия пользователем форме, позволяющей «погрузиться» в данные, работать с их визуальным представлением, понять их суть и выявить существующие закономерности, сделать выводы и напрямую взаимодействовать с данными.

В данной статье приводится описание подхода к решению задачи визуализации данных на основе программного комплекса ViP (информацию о комплексе ViP можно получить в [1]).

Архитектура подсистемы представления отчётов в системе ViP

Отчётом будем называть документ, содержание которого динамически определяется на основе данных, содержащихся в базе данных. На рис. 1 схематично представлен путь данных из базы до того, как их увидит пользователь на экране в виде отчёта. Поскольку комплекс полностью ориентирован на работу с *многомерными данными*, данные для отчётов выбираются из *гиперкубов*. Все доступные гиперкубы в системе ViP хранятся в каталоге показателей. Каталог показателей полностью открыт, что позволяет разработчикам и пользователям использовать все доступные гиперкубы.

Использование гиперкубов в качестве источников данных для отчётов в системе ViP обусловлено рядом преимуществ:

- Многомерная модель в наибольшей степени удовлетворяет потребностям Business Intelligence благодаря своей возможности *получать агрегированные данные по различным группировкам* [1].

- Реализация гиперкубов в комплексе ViP предполагает наличие семантического уровня, который скрывает внутреннюю структуру куба и позволяет пользователю оперировать данными в терминах предметной области.
- Гиперкубы в системе ViP могут содержать вычисляемые показатели; кроме того, менеджер математических моделей комплекса реализует возможность построения вычисляемых кубов, что расширяет возможности системы.

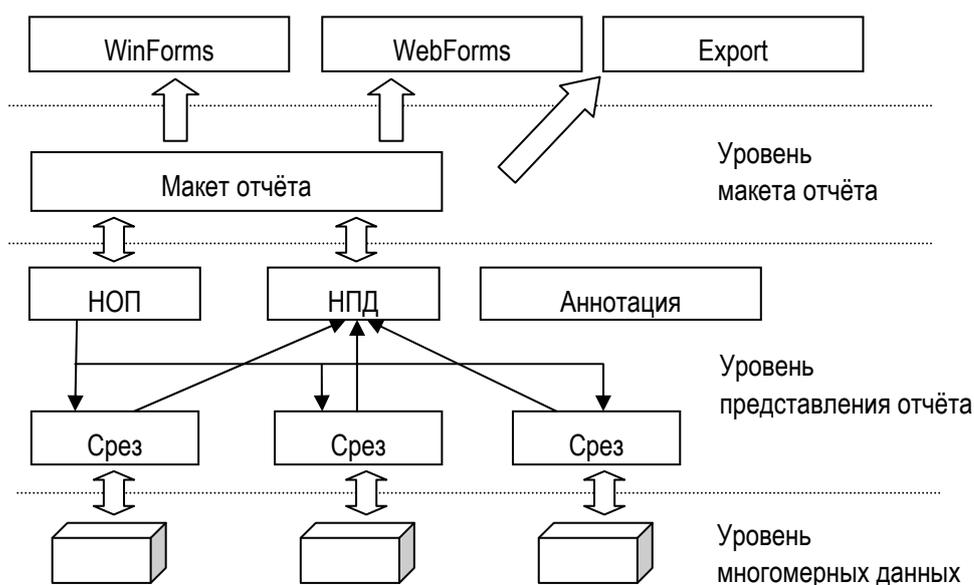


Рис. 1. Архитектура отчёта в подсистеме представления отчётов

Для каждого гиперкуба, участвующего в построении отчёта, строится срез. Срез куба представляет собой объект репозитория, который определяет набор фиксированных измерений с установленными отметками, а для каждого открытого (не фиксированного) измерения определяются ограничения на отметку (возможность множественной отметки и разрешённые отметки). Другими словами можно сказать, что срез – это параметризованный запрос к многомерной базе данных.

В системе ViP все отчёты группируются по каталогам отчётов. В каталогах отчётов каждый отчёт сопровождается метаданными (аннотацией) которые позволяют разработчику упростить реализацию поиска отчётов и разграничение прав пользователям на отчёты.

Все отчёты в системе ViP имеют многоуровневую структуру, каждый уровень которой соответствует определённой степени абстракции. Самым первым является уровень представления отчёта (рис. 1).

На уровне представления отчёта определяются:

- Срезы гиперкубов.
- Набор открытых параметров отчёта (НОП). Данный набор предназначен для задания отметок всем открытым измерениям всех срезов, участвующих в построении отчёта, кроме того, для каждого открытого параметра определяется значение по умолчанию.
- Основная задача, стоящая перед отчётами, это визуализация данных. Поэтому на уровне представления отчёта строится набор представлений данных (НПД). Для каждого представления данных определяются срезы и способ визуализации (таблица, диаграмма, график карта и т.п.), хотя конкретные свойства каждого способа визуализации не определяются. Библиотека средств

визуализации программного комплекса ViP открыта для расширения и может быть дополнена новыми элементами. Имеется возможность отобразить данные нескольких срезов в одном элементе визуализации, но для этого необходимо, чтобы срезы гиперкубов, на основе которых построены срезы, имели общие измерения.

- Для каждого отчёта на уровне представления задаётся *аннотация*, которая призвана помочь найти нужный отчёт в каталоге отчётов. Каждый разработчик может использовать в своём приложении несколько каталогов отчётов и в каждом каталоге определить собственную структуру аннотации, но внутри одного каталога отчёты имеют одну и ту же структуру аннотации.

Таким образом, на уровне представления отчёта определяется, что будет отображать отчёт.

На следующем уровне на основе представления строится *макет отчёта*.

Представление отчёта является самостоятельным объектом репозитория, что даёт возможность построения нескольких макетов на основе одного представления отчёта. На уровне макета отчётов определяются конкретные *свойства средств визуализации данных*. Кроме того, определяются дополнительные объекты, такие как надписи и графические изображения. Для каждого открытого параметра определяется элемент управления, который будет задавать значение данного параметра, его свойства размеры и размещение.

Таким образом, на уровне макета отчёта определяется, как будет выглядеть отчёт. Макет отчёта, как и представление отчёта, является объектом репозитория.

Для каждого макета возможны разные способы представления: отчёт может быть отображён в Web-приложении, Windows-приложении или экспортирован в документ PDF, за то, как отобразить отчёт, отвечает презентационный уровень.

Отдельного модуля, отвечающего за весь жизненный цикл построения отчётов, в программном комплексе нет, вместо этого *в разных модулях комплекса реализованы отдельные части подсистемы представления отчётов*.

Прежде всего, *менеджер метаданных* комплекса [1] реализует сервисы по работе с объектами, участвующими в построении отчётов: срезы, параметры отчётов, представления данных, аннотации, представления отчётов, средства визуализации, макеты отчётов. Разработчику предоставляется программный интерфейс для манипулирования данными объектами.

Менеджер сервера реализует функциональность по выполнению запросов на построение, создание и манипулирование отчётами.

В библиотеке готовых компонентов реализованы элементы управления как для различных средств визуализации, так и для визуализации отчётов в целом.

Навигация по данным портала

Основным источником данных в портале является *каталог показателей*. Каталог показателей представляет собой базу гиперкубов (термин «гиперкуб» в пользовательском интерфейсе не используется, вместо этого используется понятный пользователю термин «показатель»). При работе с порталом пользователь имеет доступ к каталогу показателей портала и способен загружать и просматривать данные по интересующим его показателям. В портале реализована концепция *пользовательского портфеля показателей*. Пользовательский портфель показателей – это коллекция ссылок на показатели из глобального каталога, формируемая самим пользователем: в портфель пользователь помещает ссылки на показатели, с которыми регулярно работает, что помогает облегчить его работу, сократив время на поиск.

Каждый показатель в каталоге сопровождается информацией об источниках данных, на основе которых был сформирован этот показатель, дате и времени формирования показателя, дате и времени внесения последних обновлений; если показатель является регулярно обновляемым, то для показателя устанавливается экономический календарь, в котором отражается периодичность обновлений данных и планируемая дата следующего обновления, а пользователь имеет возможность подписаться на рассылку, в которой бы отражалась информация об обновлениях интересующих его показателей.

Стандартным средством отображения данных показателя является *сводная таблица*. Сводная таблица является самым универсальным средством представления многомерных данных благодаря своей возможности выполнения всех типов запросов к многомерным данным, таким как:

1. Запрос "slice-and-dice" – сокращение гиперкуба путём фиксации части измерений.
2. Запрос "drill-down" – получение более детальных данных путём установки отметок части измерений на более низкий уровень.
3. Запрос "roll-up" – получение агрегированных данных путём установки отметок части измерений на более высокий уровень.
4. Запрос "rotating" – получение новой группировки куба, представляющей данные в удобной для пользователя форме.

Несомненным достоинством сводной таблицы как средства визуализации является универсальность и гибкость. С другой стороны, использование сводной таблицы требует от пользователя определённого уровня подготовки, поэтому большинство пользователей предпочтут работать с готовыми отчётами.

Работа с отчётами в портале

В портале пользователю доступно сразу несколько каталогов отчётов:

1. *Глобальный* каталог отчётов, который формируется аналитиками портала и доступен всем пользователям.
2. *Личный* каталог отчётов формируется пользователем. Отчёты в данном каталоге разрабатываются самим пользователем на основе данных из глобального каталога показателей. Данный каталог доступен только для самого пользователя.
3. *Публичные* каталоги пользовательских отчётов формируются, когда часть своих отчётов пользователь может опубликовать для других пользователей портала. В этом случае пользователем формируется публичный каталог отчётов, который может быть доступен для других пользователей. Кроме того, отчёт после соответствующей проверки и при желании пользователя может быть помещён в глобальный каталог отчётов.

Отчёт в портале представляет собой документ, состоящий из следующих элементов:

1. *Статические элементы*, т.е. элементы, содержимое которых не зависит от данных. К таким элементам относятся картинки, надписи и заголовки.
2. *Элементы визуализации данных* – элементы, содержимое которых зависит от данных из каталога показателей. К таким элементам относятся: таблицы, графики, карты, диаграммы. При размещении элементов визуализации для каждого элемента указывается показатель, данные которого должен отображать элемент. Часть элементов визуализации способны отображать данные сразу нескольких показателей, в этом случае указывается группа визуализируемых показателей. Для каждого показателя фиксируются отметки измерений, на основе которых будет сформирован срез гиперкуба.

3. При размещении элементов визуализации для части измерений показателей отметки могут быть не заданы, для таких измерений в отчёте размещаются *элементы управления, задающие данные отметки*. На этапе проектирования отчёта пользователем определяется привязка данных элементов к конкретным измерениям конкретных срезов. В качестве подобных элементов управления, наряду с традиционными списками и переключателями могут выступать и другие элементы визуализации, например карта, отображающая данные некоторого показателя, может фиксировать нажатия пользователем мыши в своей области для получения более детальной информации по выбранному на карте региону.

Поддержка отчётами открытых параметров делает их интерактивными, наглядными и позволяет избавить часть пользователей от использования сложных сводных таблиц.

Заключение

В данной статье описаны архитектура и базовые концепции подсистемы представления отчётов в комплексе BiP. Данный комплекс предлагается использовать в качестве подсистемы при разработке исследовательского портала «Инновационное развитие регионов». Существуют различные методики оценки инновационной активности, развития регионов. Разрабатываемый портал должен помочь исследователям изучить и сравнить эти методики, попробовать применить их на практике. Решение этой задачи требует реализации гибких средств настройки, предоставления пользователям возможности работы в терминах их предметной области. Данные для анализа собираются из различных источников, что делает задачу их анализа и визуализации еще более важной и сложной. Описанные компоненты интегрируются со средствами поиска, классификации и каталогизации документов, их анализа с целью извлечения полезной информации.

Благодарности

Работа выполнена при финансовой поддержке РГНФ (проект № 09-02-00373В/И).

Библиографический список

- [1] Мальцев П. Моделирование многомерных данных в системе METAS BI-PLATFORM // Advanced Studies in Software and Knowledge Engineering: International Book Series / Sofia, 2008. Pp. 173-180.

Сведения об авторе

Павел Мальцев – Пермский государственный университет, аспирант кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, 614990, ул. Букирева, 15;
e-mail: pavel_maltsev@mail.ru