

**Комплекс «Библиотечная статистика»
компании «Открытые библиотечные системы» –
новые технологии для сбора и обработки данных**

**Library Statistics Comprehensive System
of the «Open Library Systems» Company –
New Data Retrieval and Processing Technologies**

А. В. Кедрин

*Санкт-Петербургский государственный политехнический университет,
институт корпоративных библиотечно-информационных систем, Санкт-Петербург, Россия*

К. В. Поникаровский

Компания «Открытые библиотечные системы», Санкт-Петербург, Россия

Andrey Kedrin

*St. Petersburg State Polytechnics University, Institute of Corporate Library and Information Systems,
St. Petersburg, Russia*

Konstantin Ponikarovsky

«Open Library Systems» Company, St. Petersburg, Russia

Рассматриваются задачи, решаемые программным комплексом «Библиотечная статистика» компании «Открытые библиотечные системы». Проведено сравнение с решениями других компаний в данной области. Раскрыты основные особенности комплекса – независимость от используемых в библиотеке автоматизированных средств (АБИС), распределённые возможности комплекса, использование новых технологий (MOLAP, аналитические базы данных). Вся работа с отчётами, включая процесс создания и редактирования, ведётся непосредственно в программе MS Excel без использования дополнительных программ.

По мере автоматизации библиотечных процессов появляется возможность получать объективные данные о работе различных сотрудников, отделов и библиотеки в целом по широкому кругу показателей. Такие детальные сведения позволяют не только сформировать отчеты для вышестоящих организаций, но и обеспечить оперативное и стратегическое управление библиотекой, они могут быть применены во многих сферах деятельности библиотеки.

По мнению авторов, одно из основных требований к автоматическому сбору статистики – накопление данных по максимально возможному количеству показателей. Возможно, большое количество показателей может показаться избыточным для традиционных годовых отчётов, однако без такого числа показателей, статистические данные не могли бы быть применимы в области аналитики и принятия решений. Например, руководитель отдела, имея все необходимые данные о работе своих сотрудников, может перераспределять нагрузку работников, а также учитывать эти данные при выдаче премий. Другой вариант – предоставление информации кафедрам и факультетам о предоставлении услуг их студентам. Аналитика поможет скоординировать развитие библиотеки в целом, ведь теперь возможно получение реальных данных для проведения многоаспектного анализа.

Основные требования к автоматизированному средству сбора статистики с учётом возможности использования его в аналитической сфере – очень высокая скорость обработки данных, а также сбор большого количества показателей. Кроме того, к автоматизированной системе сбора статистики должны предъявляться требования не только по накоплению данных по множеству критериев, но так же и по максимально удобному представлению этих данных конечному пользователю. Также желательна возможность не перестраивать отчёт с нуля, а получать новые данные, лишь уточнив или расширив какие-то из рассматриваемых показателей. Соответственно время создания, изменения и уточнения отчётов становится важным ресурсом. Конечный пользователь автоматизированного средства сбора статистики должен не только оперативно получать статистические

данные, но и получать их в максимально удобном виде, чтобы не тратить время на преобразование отчётов из одного вида в другой (худший вариант – это копирование всех получаемых данных из одной программы в другую).

Состав современных АБИС всё чаще пополняется так называемым АРМом Статистики. Однако анализ таких продуктов показал, что их функциональность зачастую не удовлетворяет сформулированным выше требованиям. Обычно АРМ Статистики предлагает набор шаблонных отчётов по книговыдаче, читателям, записям, фондам и заказам. При таком подходе набор отчётов заранее предопределён и может быть расширен только разработчиками АБИС.

Новые свойства комплекса «Библиотечная статистика» компании «Открытые библиотечные системы» обеспечиваются использованием новых подходов при его проектировании и реализации. Аналитические базы данных позволяют варьировать запросы пользователей в очень широком диапазоне без изменения программного кода. Кроме того, аналитические базы данных организованы особым образом, позволяющим выполнять обработку значительных массивов данных за весьма непродолжительное время – секунды. Всё это стало возможно благодаря MOLAP технологии, которая используется в комплексе «Библиотечная статистика».

MOLAP (Multidimensional Online Analytical Processing, аналитическая обработка в реальном времени со многими измерениями) — технология обработки информации, включающая составление и динамическую публикацию отчётов и документов, используется для быстрой обработки сложных запросов к базе данных. MOLAP делает мгновенный снимок реляционной БД и отображает ее в пространственную модель. Заявленное время обработки запросов в MOLAP составляет около 1 % от аналогичных запросов к реляционной БД. Реляционные БД хранят сущности в отдельных «плоских» таблицах. Эта структура удобна, но сложные многотабличные запросы в ней выполняются относительно медленно. Более адекватной моделью для запросов, а не для изменения, является организация данных в виде гиперкубов.

Гиперкуб – это MOLAP-структура, созданная из накапливаемых данных. Гиперкуб создаётся из соединения таблиц с применением схемы звезды. В центре «звезды» находится таблица фактов или параметров, по которым делаются запросы. Множественные таблицы с измерениями присоединены к таблице фактов. Эти таблицы показывают, как могут анализироваться агрегированные реляционные данные. Количество возможных агрегирований определяется количеством способов, которыми первоначальные данные могут быть иерархически отображены. Например, показатели книговыдачи могут быть сгруппированы по библиотекам, отделам библиотек и сотрудникам библиотек, таким образом, 50 сотрудников, 8 отделов и 2 библиотеки составят 3 уровня иерархии с 60 членами.

MOLAP-куб потенциально содержит всю информацию, которая может потребоваться для ответов на произвольные запросы. Итак, главные достоинства MOLAP технологии:

- принципиально новый подход к хранению данных в гиперкубах, позволяет делать так называемые срезы данных по любым заложенным в гиперкуб измерениям, то есть существует возможность осуществления любых «срезов» и «углублений» в определённой структуре БД;
- самое быстрое построение отчётов в сравнении как с реляционными базами данными так и с другими вариантами OLAP подхода;
- слабое влияние объёма аналитической базы данных на скорость построения отчётов.

В программном комплексе «Библиотечная статистика» компании «открытые библиотечные системы» обеспечивается быстрая обработка данных за счет использования гиперкубов, а также гарантируется удобство и простота представления данных конечному пользователю. Вся работа по созданию и редактированию отчётов ведётся непосредственно в программе Microsoft Excel. При этом никаких как таковых запросов пользователь не пишет, а работает только в графическом интерфейсе MS Excel, без использования дополнительных программных продуктов. По сути вся его работа заключается только в придании внешнего облика будущего отчёта.

Важная особенность комплекса «Библиотечная статистика» состоит в том, что сохранение созданных в Microsoft Excel отчётов позволяет в будущем не пересоздавать их, а всего лишь открыть. При этом вся преданная ранее визуальная составляющая сохранится, а сами данные автоматически обновятся. Например, если Вы подготовили в начале сентября отчёт по каким-либо показателям за текущий месяц, то, открыв этот документ в конце сентября, Вы увидите, что данные автоматически

обновились. Кроме этого вы можете изменить месяц на октябрь, таким образом, получив новые данные не пересоздавая отчёт, а лишь изменив значение одного из показателей.

Программный комплекс меняет представление о построении статистических отчётов, а именно:

- Нет никаких предустановленных шаблонных отчётов, обычно предлагаемых разработчиками, выйти за рамки которых нельзя. В комплексе **можно придать любую функциональность и оформление отчёту**. За функциональность отвечает MOLAP технология, а в оформлении поможет программа MS Excel;

- **Отчёты любой сложности строятся за считанные секунды**, при этом напрямую во время построения отчётов никакие данные из АБИС не берутся;

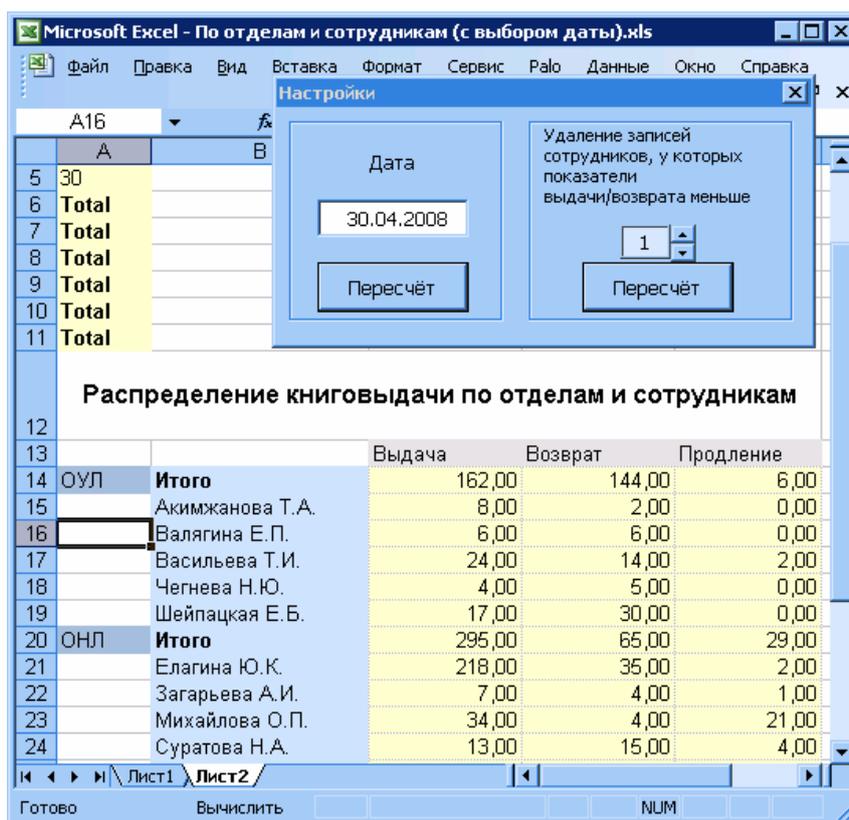
- **Сбор статистики полностью автоматизирован**. Комплекс устанавливается и настраивается один раз, а статистика собирается постоянно без участия человека. Обновление базы статистики настраивается – например, Вы можете указать ночное время для пополнения базы статистики, при этом минимизируя нагрузку на АБИС. Новые данные извлекаются в течение нескольких минут и хранятся в не зависящем от АБИС хранилище по технологии гиперкубов;

- Комплекс является распределённой системой. **Одновременно с базой статистики может работать любое количество пользователей**. При этом каждый из них работает в Excel на своём компьютере, условие здесь одно – все компьютеры должны быть связаны по сети с сервером;

- **В одном отчёте отображается до шести различных параметров**. При этом, например в статистике по книговыдаче, используется не одно, а несколько временных измерений – отчёты можно строить за любой временной интервал (от года до пяти минут).

На рисунках ниже показано использование функциональности Microsoft Excel в созданных в этой же программе статистических отчётах библиотеки.





В комплексе «Библиотечная статистика» реализовано построение гиперкуба по книговыдаче, имеющего параметры, по которым может производиться «срез» данных. В первую очередь, это основные показатели по книговыдаче: выдача; возврат; продление литературы; посещение. Отметим, что в комплексе предложен оригинальный настраиваемый алгоритм учета посещений.

Также в базу статистики введены такие измерения (параметры), как отделы библиотеки; сотрудники библиотеки; тип литературы; контингент читателей; место выдачи. К числу временных измерений относится: год; месяц; число; неделя; день недели; час; пятиминутный интервал.

Кроме этого строится гиперкуб со статистикой по читателям. Он содержит следующие показатели: год регистрации; тип читателя (студент, аспирант, ...); возрастные группы; гражданство. Помимо этого для студентов: факультет; группа; направление; специальность; форма обучения (очная, заочная, ...); вид обучения (бюджет, контракт, ...)

Важно отметить, что комплекс «Библиотечная статистика» в данный момент совместим с АБИС «Руслан», откуда берутся данные для накопления в аналитической базе данных. Соответственно, может быть осуществлена доработка комплекса для других АБИС или любых других источников, из которых можно извлечь необходимые для построения гиперкуба данные.

В качестве хранилища под аналитическую базу данных используется Palo Server (Jedox GmbH), который имеет бесплатную лицензию (GNU, General Public License).

Комплекс «Библиотечная статистика» не стоит на месте и постоянно развивается. Увеличивается как количество показателей в существующих аналитических базах, так создаются новые базы для охвата большего числа видов и форм работы современной библиотеки. На сайте компании <http://obs.ruslan.ru> можно получить актуальную информацию и развитии комплекса. А также посмотреть ролик, демонстрирующий его использование в библиотеке.