

От библиотечной статистики к системам анализа эффективности

*Поникаровский Константин Викторович, программист,
Институт корпоративных библиотечно-информационных
систем Санкт-Петербургского государственного поли-
технического университета*

В докладе рассматриваются задачи, решаемые программным комплексом «Библиотечная статистика». Обоснована необходимость применения специального программного обеспечения для автоматизации сбора статистики работы библиотеки по широкому кругу показателей. Раскрыты основные особенности комплекса. Приведены примеры того, как можно проводить анализ эффективности работы сотрудников, отделов и библиотеки в целом на основе статистических данных и MOLAP-технологии.

По мере автоматизации библиотечных процессов появляется возможность получать объективные данные о работе различных сотрудников, отделов и библиотеки в целом по широкому кругу показателей. Как правило, программы библиотечной статистики настраиваются на набор заранее сформированных запросов, позволяющих получить типовые показатели работы библиотеки. Однако при желании сформировать иной запрос, получить новые показатели или же узнать зависимости показателей от новых параметров необходимо обращаться к программистам-разработчикам для доработки программы.

Современные технологии баз данных предлагают новое решение – аналитические системы, позволяющие 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 (Multidimensional Online Analytical Processing) технологии, обеспечивающей составление и динамическую публикацию отчётов и документов. Причина использования MOLAP для обработки запросов — это скорость. MOLAP с заданной периодичностью делает мгновенные снимки АБИС, фиксируя текущие состояния процессов в системе, и формирует пространственную модель (так называемый гиперкуб) для быстрой обработки последующих запросов. Заявленное время обработки запросов в MOLAP составляет около 1 % от аналогичных запросов к реляционным БД, как правило, используемым в АБИС. MOLAP-куб содержит в себе базовые данные и информацию об измерениях (агрегатах) и потенциально включает всю информацию, которая может потребоваться для ответов на любые запросы.

Итак, главные достоинства MOLAP технологии:

- принципиально новый подход к хранению данных в гиперкубах, позволяющий делать так называемые срезы данных по любым заложенным в

гиперкуб измерениям, то есть существует возможность осуществления любых «срезов» и «детализаций» в определённой структуре БД;

- самое быстрое построение отчётов в сравнении как с реляционными базами данными, так и с другими вариантами хранения данных в АБИС;
- слабое влияние объёма аналитической базы данных на скорость построения отчётов.

Да, современные технологии дают принципиально новые возможности. Однако существуют препятствия для их эффективного использования. Первая проблема связана с продвижением таких технологий в повседневную практику - нужны удобные и понятные для конечного пользователя программные разработки. Вторая состоит в непонимании преимуществ новых подходов и в нежелании в них разобраться, научиться ими пользоваться. Третья – в высокой стоимости средств поддержки MOLAP технологий.

Программный комплекс "Библиотечная статистика" служит для автоматического сбора статистики о процессах в библиотеке, в которой функционирует АБИС "Руслан", а также для быстрого и, главное, удобного представления результатов конечному пользователю. Скорость обеспечивается новыми технологиями, а удобство - новыми решениями. Вся работа по созданию и редактированию отчётов ведётся непосредственно в программе MS Excel. Большое внимание уделено руководствам конечного пользователя, в которых рассказывается не только о принципах работы с комплексом, но также и о преимуществах самой технологии MOLAP, и даются советы в построении собственных отчётов для более полного использования возможностей комплекса. В качестве хранилища для аналитической базы данных используется Palo Server (Jedox GmbH), который имеет бесплатную лицензию (GNU, General Public License).

На данный момент в комплексе реализован сбор статистики следующих видов:

- по книговыдаче (основные показатели: выдача, возврат, продление, посещение, а также более 10 дополнительных измерений, включая сотрудников, отделы, тип литературы, тип читателя и др.);
- по читателям (основные показатели: количество зарегистрированных читателей, количество реально обслуженных читателей, количество читателей-должников и количество задолженных экземпляров, а также более 10 дополнительных измерений, включая категорию читателей, организацию читателей, возраст читателя, факультет и академическую группу читателя-студента и др.);
- по фонду (основные показатели: количество поступивших/списанных документов, количество штрих-кодированных документов, а также более 6 дополнительных измерений, включая тип документа, язык документа, сиглу хранения, счёт, базу хранения и др.);
- по счетам (основные показатели: количество записей КСУ, единицы внебаланса/набалансе, сумма внебаланса/набалансе, а также ряд дополнительных измерений, включая тип счёта, источник и сопроводительный документ)

- по заказам (основные показатели: количество единиц и сумма, а также ряд дополнительных измерений, включая источник финансирования, поставщика и заказчика).

Таким образом, комплекс является средством быстрой и точной подготовки отчетов. Кроме того, в него заложены механизмы для внутреннего анализа библиотечных процессов, мониторинга, краткосрочного и среднесрочного планирования.

Для этих целей при построении гиперкубов данных со статистикой использовалось максимально возможное количество измерений - по максимуму использовались данные баз АБИС "Руслан". Лишь некоторая часть из этих измерений нужна для традиционных отчетов библиотеки. Использование всей полноты этих измерений позволяет перейти на новый уровень - узнать зависимости, которые раньше никогда не были доступны. Эти новые данные могут повлиять на процессы, протекающие в библиотеке, через принимаемые на их основе управленческие решения. С помощью этих данных можно легко производить анализ ситуации и делать прогнозы на будущее. И все эти аналитические и прогностические выводы будут иметь под собой объективное обоснование.

Вот несколько примеров, показывающие возможности комплекса для анализа и прогнозирования.

1. Составление расписания отделов обслуживания с учетом типичных «пиков» и «спадов».

Это возможно благодаря введению в гиперкуб статистики по книговыдаче нескольких временных измерений, в том числе измерения "часы" и "минуты". На рис.1 показан скриншот с зависимостью количества операций (выдач, возвратов, продлений) для одного из отделов обслуживания библиотеки в зависимости от времени суток. График построен в программе MS Excel по данным, полученным от комплекса Библиотечная статистика.



Рис.1 График загрузки отдела обслуживания в течение рабочего дня

2. Определение занятости и нагрузки сотрудников отделов обслуживания.

Это возможно благодаря введению измерений с отделами обслуживания и сотрудниками библиотеки в гиперкуб статистики по книговыдаче. Детализируя

именно эти измерения (не трогая остальных) можно получить результат, который показан на рис.2 - зависимость количества операций (выдач, возвратов, продлений) для одного из отделов обслуживания по каждому сотруднику этого отдела (фамилии сотрудников и приведенные показатели являются вымышленными). Временной диапазон, за который можно получить эти данные, можно варьировать.

Распределение книговыдачи по отделам и сотрудникам (за один день)

		Выдача	Возврат
ОНЛ	Итого	118	4
	Михайлова А.И.	12	2
	Козлова Т.В.	18	7
	Сидорова А.Н.	14	2
	Романова О.П.	31	4
	Соколова Ю.К.	24	26

ОУЛ	Итого	280	97
	Петренко Е.В.	5	3
	Макарова М.В.	12	14
	Семенова Е.Б.	46	24

НЧЗ	Итого	5	4
	Петрова О.С.	2	2
	Иванова О.А.	1	0

3. Подготовки отчетов кафедрам и факультетам о предоставлении услуг их студентам, а так же их задолженностях.

Это возможно благодаря введению в базы статистики по книговыдаче и по должникам таких измерений как организация читателя (для студента - это факультет) и академическая группа читателя-студента. Срез по этим измерениям показан на рисунке.

Количество должников библиотеки с распределением по факультетам

	Количество должников	Количество экземпляров
Итого	3 163	10 326
Факультет	3 111	10 218
Факультет Гуманитарный	137	200
Факультет Инженерно-строительный	469	1 510
Факультет Инноватики	47	148
...

4. Определение контингента фактически обслуженных читателей за тот или иной период (в отличие от просто зарегистрированных в библиотеке), а также определение интересов читателей по типам литературы для корректировки профиля комплектования.

Для этого используются различные измерения, характеризующие как читателя (категория, возрастная группа, вид и форма обучения и др.), так и документ (тип литературы согласно формату RUSMARC, язык документа, страна издания). Это продемонстрировано на рисунке ниже.

Читатели-студенты библиотеки с распределением по стране и форме обучения

			Выбывшие	Текущие
Студент	Белоруссия	заочное	1	6
		очное	3	138
		очно-заочное	0	7
	Китай	заочное	0	0
		очное	16	279
		очно-заочное	0	0
	Россия	заочное	273	1 767
		очное	1 750	17 446
		очно-заочное	299	2 449

5. Формирование подписки с учетом реальной спрашиваемости журналов, оценка целесообразность затрат на те или иные дорогостоящие книги.

В комплекс "Библиотечная статистика" включена дополнительная программа, которая строит отчёты по запросу пользователя "на лету". Программа позволяет проследить историю документа - кем и когда он был выдан, количество задолженностей и другие показатели.

Вид результирующих отчётов на примере журнала "Наука и жизнь" приводится ниже.

Вначале приводится описание искомых документов. Указывается поисковый атрибут (ключ записи, инвентарный номер или штрих-код документа), затем описание из найденной по поисковому атрибуту записи и информация об инвентаризации.

Отчёт за период		
с 01.12.2008	по 01.06.2009	
Описание документов		
P32852	Ключ: RU\SPSTU\ser\258899 Наука и жизнь: ежемесячный научно-популярный журнал . №4 - М.: Пресса, - 1935- №4 - , - 2009	Поступило всего: 1 Списано всего: 0 Штрихкодировано всего: 1 Штрихкодированных списано: 0
...

Далее приводятся данные текущего момента - информация об инвентаризации по сиглам хранения, а так же информация о количестве выданных экземпляров. Если поиск производился по инвентарному номеру, то будет отображена информация только для этого инвентарного номера, а не для всей записи в целом. Аналогично для штрих-кодов.

Данные текущего момента

	ОЧЗ	ОНЛ
Всего поступило	110	59
Всего списано	0	0
Всего на учёте	110	59
Штрихкодированных всего	110	59
Штрихкодированных списано	0	0
Штрихкодированных на учёте	110	59
Выдано	0	0

Далее следует таблица с историей книговыдачи. Таблица может не включаться в отчёт (для этого существует специальная настройка на главной форме приложения). Состав столбцов этой таблицы также может быть настроен пользователем (пользователь может выбрать, какие теги из баз книговыдачи и баз читателей он хотел бы видеть в этой таблице).

История книговыдачи					
Дата/время операции (3)	Дата выдачи (142)	Дата возврата (143)	Ключ записи (145)	Отдел хранения книги (147)	Категория читателя (200)
20080114110558	20080114	20080710	RU\SPSTU\ser\196689	ОУЛ	Студент
20080206153610	20080206	20080710	RU\SPSTU\ser\196689	ОУЛ	Студент
20080207160857	20080207	20080710	RU\SPSTU\ser\196689	ОУЛ	Студент
20081111115718	20081111	20090202	RU\SPSTU\ser\196689	ОУЛ	Студент
....

Далее идут обобщенные показатели по книговыдаче документа. Здесь содержится информация о количестве выдач документа за указанный пользователем период перед созданием отчёта. Вторая цифра во всех строках показывает количество не сданных на конечную дату периода (установленную пользователем) документов из количества выданных за этот период (первая цифра во всех строках). При этом цифры приводятся по отделам выдачи, по категориям читателей, по организациям читателей (в т.ч. факультетов), с указанием курсов и академических групп для читателей-студентов. После этой таблицы приводятся аналогичные таблицы по каждому отделу. В конце такой таблицы приводится информация о должниках: перечисляются штрих-коды читателей, у которых на конечную дату указанного пользователем периода, документы уже были просрочены.

Использование комплекса также дает и дополнительные возможности, не являющиеся основной целью создания комплекса. Так, происходит автоматическое выявление ошибок электронного каталога. Во время обработки баз данных комплекс "Библиотечная статистика" создаёт лог-файлы с указанием встретившихся неточностей и ошибок, например, ошибки при заполнении организации читателя или ошибки заполнения полей RUSMARC-записей, относящихся к инвентаризации (неправильно указана дата счёта или количество экземпляров - поля 989/999 подполя 'v' и 'e'). Так как базы данных АБИС "Руслан" в первую

очередь используются для машинной обработки, то наличие таких неточностей должно быть сведено к нулю. Исправление подобных ошибок, в свою очередь, положительно влияет на точность статистических данных.

Важно отметить, что комплекс "Библиотечная статистика" в данный момент совместим с АБИС "Руслан". Соответственно, может быть осуществлена доработка комплекса для других АБИС или любых других источников, из которых можно извлечь необходимые для построения гиперкуба данные.

Комплекс "Библиотечная статистика" не стоит на месте и постоянно развивается. Ведётся работа по улучшению существующих и построению новых гиперкубов данных со статистикой. Ознакомительную демо-версию комплекса (с минимальными функциональными ограничениями), более подробное описание, а также подробную документацию можно загрузить с сайта ООО «Открытые библиотечные системы» <http://obs.ruslan.ru/>.