

**Проблемы репликации данных
в информационных системах библиотек**
Data Replication Problems in Library Information Systems

**Проблеми реплікації даних
в інформаційних системах бібліотек**

Халабія Р. Ф.

*Московский государственный университет приборостроения и информатики,
Москва, Россия*

Халабія М. Л.

Московский государственный университет культуры и искусств, Москва, Россия

Rustam Khalabia

Moscow State University of Instrument-Making and Informatics, Moscow, Russia

Mariya Khalabia

Moscow State University of Culture and Arts, Moscow, Russia

Халабія Р. Ф.

*Московський державний університет приладобудування та інформатики,
Москва, Росія*

Халабія М. Л.

Московський державний університет культури та мистецтв, Москва, Росія

Одной из задач автоматизации крупных библиотек, имеющих территориально-распределенную структуру, является оперативный обмен данными между библиотеками-филиалами. Такими данными могут быть как справочники, используемые в автоматизированных системах (например, кодификаторы, шифраторы), так и результаты деятельности библиотек (базы данных, электронные каталоги, статистические данные). Все эти сведения необходимы библиотекам-филиалам для текущей повседневной работы, и для анализа своей деятельности, по завершении отчетных периодов. Обеспечить такой обмен, призвана система репликации, которая объединяет разрозненные данные филиалов-библиотек в единую репликационную сеть. Кроме того, в докладе изложены основные проблемы репликации данных в информационных системах библиотек.

Among the automation tasks of the major libraries having a regionally ramified structure is efficient data exchange between branch libraries. These data include reference-books used within automated systems (codifiers, scramblers) and results of library activities (databases, electronic catalogs, statistic data). These data are essential for the routine work of branch libraries and for the analysis of the activity upon completion of reporting periods. To ensure such exchange a replication system has been introduced which unites the scattered data of branch libraries into a joint replication network. Additionally, the report describes the most serious replication problems in library information systems.

Одним із завдань автоматизації великих бібліотек, які мають територіально-розподілену структуру, є оперативний обмін даними між бібліотеками-філіалами. Такими даними можуть бути як довідники, що використовуються в автоматизованих системах (наприклад, кодифікатори, шифратори), так і результати діяльності бібліотек (бази даних, електронні каталоги, статистичні дані). Всі ці відомості необхідні бібліотекам-філіалам для поточної повсякденної роботи, і для аналізу своєї діяльності, по закінченню звітних періодів. Забезпечити такий обмін, покликана система реплікації, яка об'єднує розрізнені дані філіалів-бібліотек у єдину реплікаційну мережу. Крім того, у доповіді викладено основні проблеми реплікації даних в інформаційних системах бібліотек.

Современная библиотека является сложной и территориально-распределенной структурой, имеющей множество филиалов, между которыми происходит непрерывный обмен информационными потоками. Под информационным потоком будем понимать множество данных: таких как полнотекстовые, фактографические базы данных, библиографические записи, статистические данные о деятельности крупных библиотечных систем.

При обмене данными между филиалами возникают следующие проблемы: потеря, дублирование записей, потеря связи с центральной базой данных, низкая производительность системы, а также необходимость приобретения дорогостоящего оборудования. Далеко не все библиотеки-филиалы могут обмениваться данными в режиме «on-line» – некоторые удаленные филиалы не имеют прямого доступа к общей базе данных. Это зачастую создает определенные проблемы: обмен информацией производится при помощи электронных носителей, либо в пакетном режиме (передача файлами) по электронной почте с их последующей конвертацией.

Важным вопросом для распределенных систем библиотек является репликация данных. Под репликацией понимают процесс приведения данных электронных таблиц двух баз данных (БД) в идентичное состояние. Данные обычно реплицируются для повышения надежности и увеличения производительности. Одна из основных проблем при этом – сохранение непротиворечивости реплик. В поддержании непротиворечивости реплик играют роль два более или менее независимых аспекта. Первый из них – это реальное распространение обновлений, куда входят вопросы размещения реплики того, как по ним расходятся обновления. Второй аспект касается поддержания непротиворечивости реплик. В большинстве случаев при репликации необходим жесткий вариант непротиворечивости [1, 4]. Говоря формально, это означает, что обновления должны расходиться по репликам более или менее быстро.

Таким образом, проблема дублирования библиографических записей, обновленные филиалами или головной библиотекой, тесно связана с таблицами уникальных идентификаторов или с проблемой низкого качества передачи данных по каналам связи.

Каждая библиографическая запись, которая находится в электронном каталоге OPAC (Open Public Access Catalogue) содержит уникальный идентификатор вида [3]:

BY-CNB-IN-XXX-nnn,

где BY – страна, создатель информационного продукта
CNB – учреждение, создающее библиографическую запись,
IN – обозначение сериального издания на иностранном языке,
XXX – имя базы данных, где запись была впервые создана,
nnn – порядковый номер записи в таблице электронного каталога, формируемый системой управления базой данных (СУБД) типа Oracle или DBase.

Однако такое решение является хорошо масштабируемым и позволяет заложить в идентификатор записи (ID) много важной дополнительной информации. Но у него тоже имеются свои недостатки. Во-первых, некоторая избыточность информации. Во-вторых, более сложный механизм генерации ID. На наш взгляд, процесс репликации ID в электронных каталогах должен быть ограничен форматом описания библиографических данных MARC и протоколом Z 39.50 или форматом описания метаданных Дублинского ядра (Dublin Core) и протоколом, что позволит решить проблему синхронизации ID и хоть как-то сократить дублирование записей при репликации данных.

Проблема низкого качества передачи данных существует очень давно. По той причине, что многие библиотеки не обладают достаточными материальными средствами для закупки надежного оборудования и каналов связи.

Еще одной не менее важной проблемой репликации данных является удаление записей из справочников. Данный конфликт, вытекает сам собой из самого понятия распределенной базы данных (РБД). Если в некоторый момент времени происходит удаление какого-либо объекта в одной из частей БД, то другая БД об этом удалении ничего не знает. В этом и состоит природа этой проблемы.

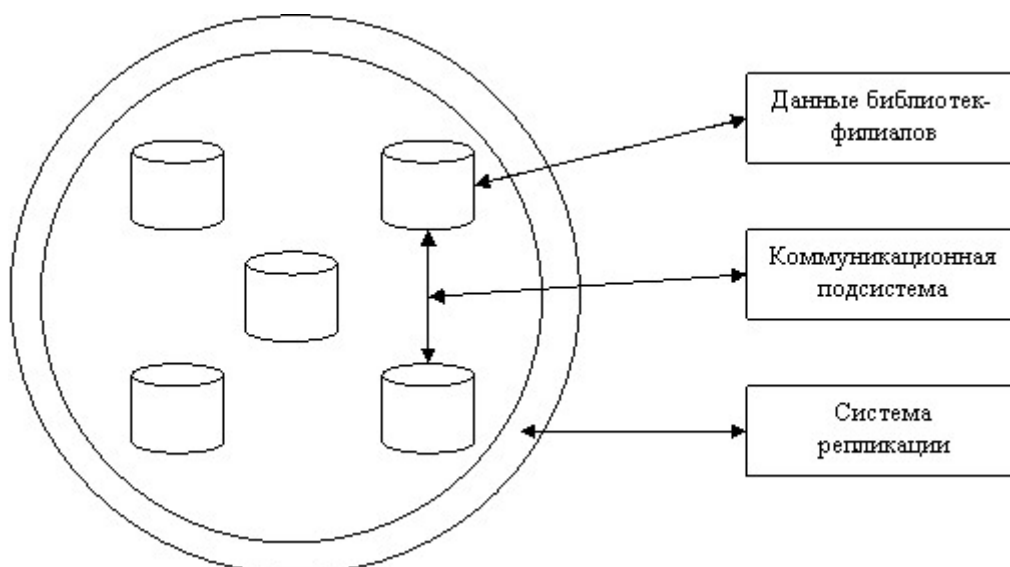


Рис. 1. Схема репликационной сети крупной библиотечной системы

Существующие проблемы можно решить при помощи системы репликации данных. Система репликации данных предназначена для обеспечения оперативного обмена между удаленными филиалами в режиме «off-line» с наименьшими временными и трудовыми затратами. По сути это система управления централизованными и распределенными данными, совместно используемыми территориально разделенными подразделениями в одну **репликационную сеть** (см. рис.1) [2]. Данная сеть имеет 3 составляющие:

1. Данные библиотек-филиалов;
2. Коммуникационная инфраструктура;
3. Система с единым центром репликации.

По принципу размещения все данные можно разделить на:

- Базы данных, размещенные в библиотеках-филиалах;
- Центральную базу данных она находится в Центральной библиотеке;
- Справочники (коды стран, коды языков, таблицы УДК или ББК).

Это означает, что весь обмен информацией между филиалами контролируется одним **центром репликации** (см. рис. 2) [2]. Все данные, которыми обмениваются библиотеки-участники корпоративной сети, высылаются сначала в центр, а потом он в зависимости от типа информации – либо использует ее сам, либо рассылает филиалам, которым она предназначена. Таким образом, все филиалы, обменивающиеся информацией, объединяются в репликационную сеть с общим центром репликации. Физически это выполняется посредством каналов связи любого типа (телефонных или компьютерных сетей, протоколов, например Z39.50).

Следует отметить, что все информационные потоки проходят через центр репликации, в котором хранятся копии баз данных, участвующие в данном информационном обмене.

Все вышеперечисленные проблемы репликации возникают и в корпоративных сетях библиотек. Центр репликации позволяет решить вопросы потери данных при их передачи, сократить дублирование записей и сделать технологию описания библиографических записей еще более корпоративной.

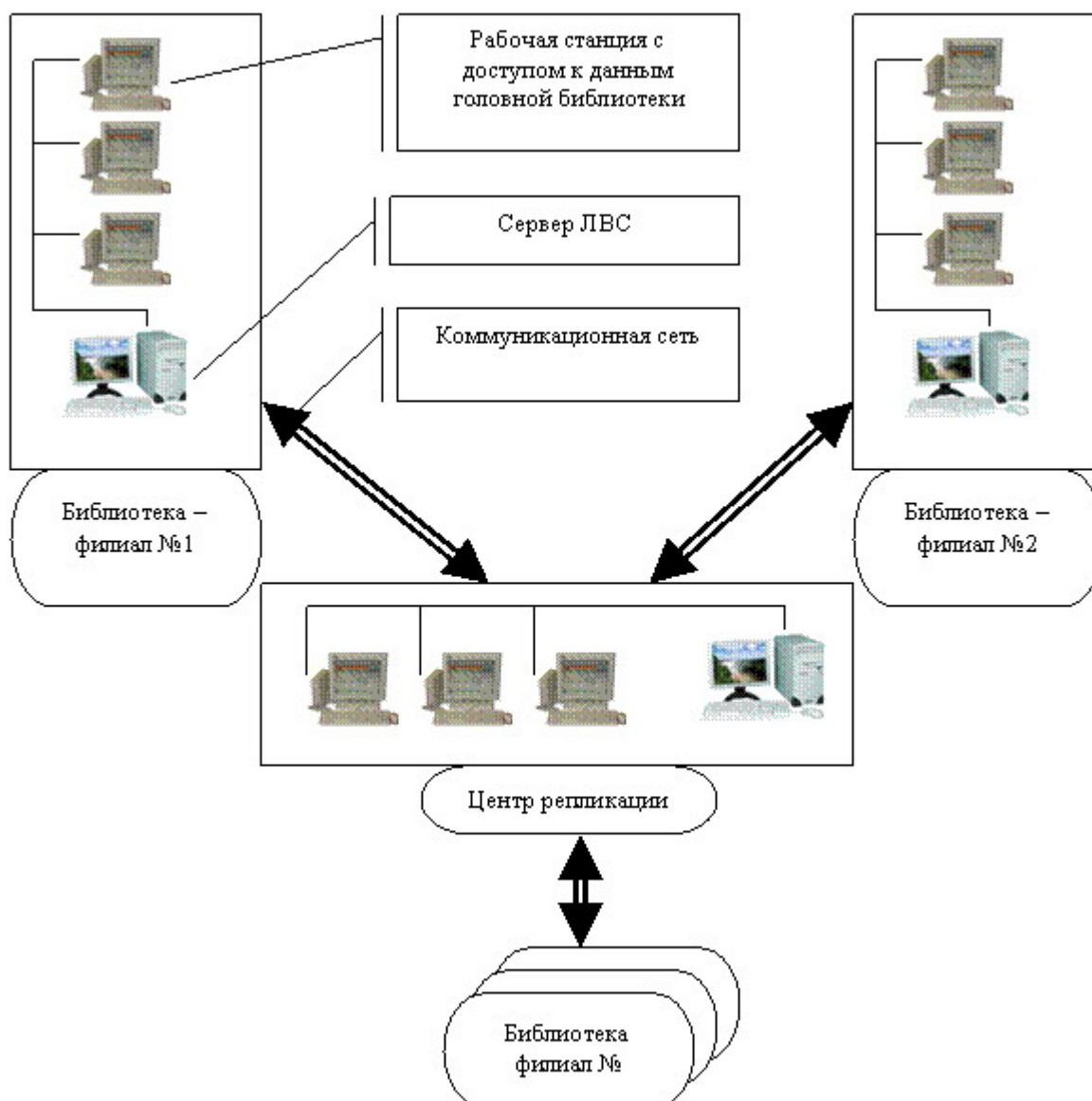


Рис. 2. Структура репликационной сети

Литература

1. Евдокимов А. Репликация базы данных [электронный ресурс]
2. Петров Е. Система репликации данных // RS-Club: информационно-аналит. изд. – 1997. – №4. – с. 56-59.
3. Сапун М.Л. (Халабия М.Л.) Особенности библиографического описания сериальных документов в коммуникативном формате BELMARC // Информационное обеспечение науки Беларуси: к 80-летию со дня основания ЦНБ им. Я. Коласа НАН Беларуси: Сб. науч. ст. Отв. ред. Н.Ю. Березкина; НАН Беларуси. ЦНБ им. Я.Коласа. – Мн.: Красико-Принт, 2004. – 176 с.
4. Таненбаум Э., ван Стеен М. Распределенные системы: принципы и парадигмы / Э. Таненбаум, М. ван Стеен. – Спб.: Питер, 2003. – 876 с.