

Изучение целостности записей корпоративных АБИС

Integrity of bibliographic records for distributed library system

А.С. Карауш, Д.Ю. Копытков
Муниципальная информационная библиотечная система г.Томска

Alexander S. Karaush, Dmitry Yu. Kopytkov
Tomsk Municipal Information Library System

В рамках проекта «Relication» были проведены работы по изучению целостности данных при эксплуатации корпоративной автоматизированной библиотечной информационной системы (АБИС). На примере использования в системе из пяти библиотек дается перечень и анализ выявленных потерь данных. При проведении экспериментов для тиражирования (репликации) использовалось программное обеспечение «Relication» (Муниципальная информационная библиотечная система, г.Томск), для задач каталогизации локальной автоматизированной библиотечной информационной системы использовалась Система «ИРБИС» (ГПНТБ России). В качестве основного идентифицирующего признака библиографической записи для тиражирования применялся алгоритмический код.

Корпоративные автоматизированные библиотечные информационные системы (АБИС) работают в ряде библиотечных систем и корпораций. Построены такие системы на основе различных моделей тиражирования электронных каталогов (ЭК) или данных [1-3]. Причем, выбор модели зависит, как правило, не от структуры библиотеки, а от используемого программного обеспечения для создания ЭК. При эксплуатации корпоративных АБИС в реальных условиях происходят разные конфликтные ситуации, влияющие на целостность данных ЭК. Под целостностью данных понимается состояние хранимых данных, характеризующееся отсутствием нарушений спецификаций схемы хранения, а также физических разрушений данных на носителях [1].

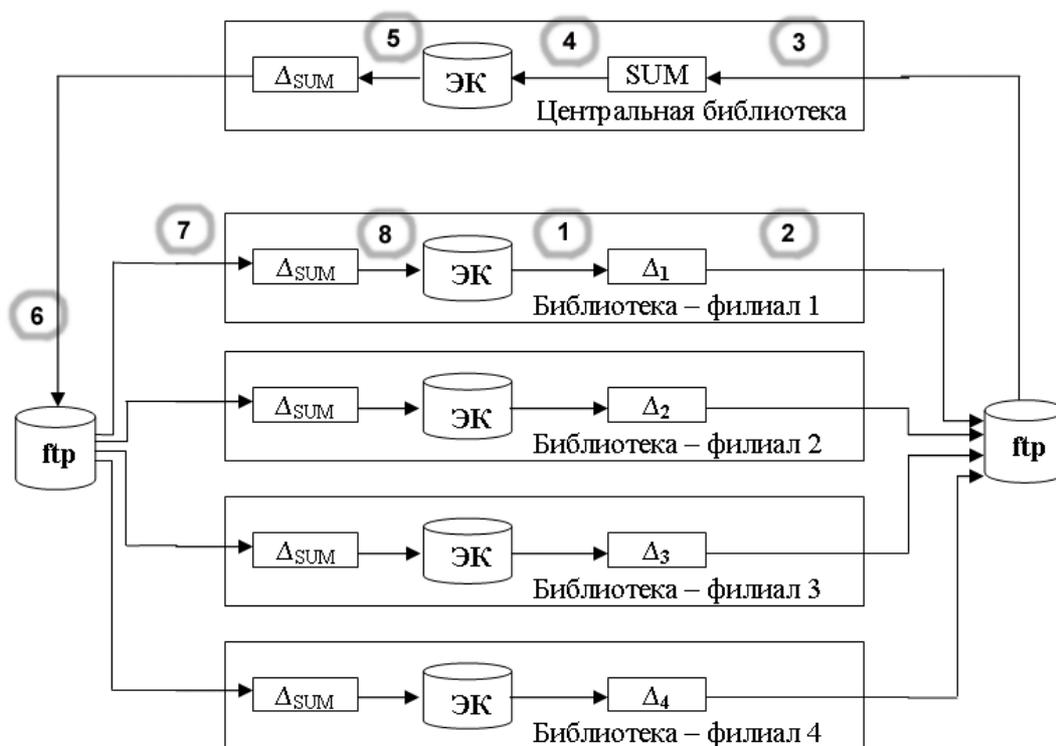


Рисунок 1. Структурная схема используемой корпоративной АБИС

Далее рассматривается корпоративная АБИС для объединения библиотек, созданная на основе модели тиражирования с использованием алгоритмических кодов

(алкодов) библиографических записей [4], состоящая из центральной библиотеки и четырех библиотек-филиалов. Структурная схема такой корпоративной АБИС представлена на рисунке 1.

Согласно этой схемы (рис.1) все действия и алгоритмы по обеспечению целостности данных ЭК выполняются в центральной библиотеке. Обмен данными с библиотеками-филиалами производится по информационным каналам связи, во время исследования - модемным линиям. Это позволило гарантировать «худшие» условия работы исследуемой корпоративной АБИС. Было принято условие, что актуальность данных в корпоративной АБИС составляет 24 часа, т.е. изменения, сделанные в локальном ЭК любой библиотеки - филиала должны быть доступны во всех остальных локальных ЭК через сутки. Временная диаграмма работы процесса тиражирования ЭК корпоративной АБИС показана на рисунке 2. Представленная модель является частью общей модели распределенного создания и тиражирования локальных ЭК библиотек, которая работает в Муниципальной информационной библиотечной системе города Томска [5].

В каждой библиотеке (рисунок 1) установлена локальная копия имеющейся АБИС. В течение каждого дня в ЭК локальной АБИС производится создание новых, удаление или изменения записей.

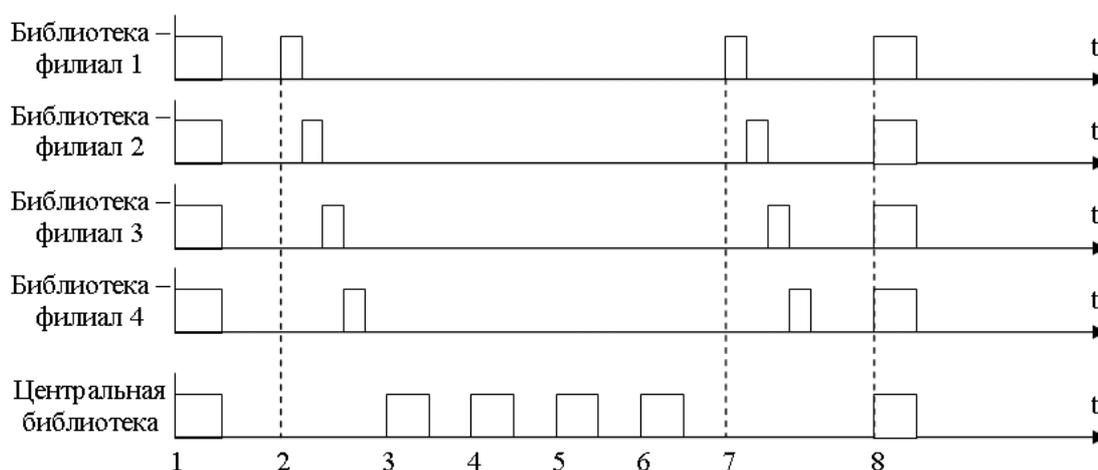


Рисунок 2. Временная диаграмма работы исследуемой корпоративной АБИС

В период (1...2) в каждой библиотеке запускается программа, которая создает файл отчета изменений локального ЭК этой библиотеки $\Delta_1... \Delta_4$ с учетом локальных ограничений. В период времени (2...3) каждая библиотека передает по информационным каналам связи полученные файлы изменений. На рисунке 2 показано использование одного канала связи (один входной модем), по которому последовательно передаются файлы на ftp-сервер. После того, как все файлы локальных изменений поступили с ftp-сервера центральной библиотеки, на сумматор «SUM» в период времени (3...4) запускается программа суммирования данных из библиотек-филиалов. Затем эти данные применяются на ЭК центральной библиотеки в период времени (4...5), после чего создается суммарный (сводный) файл Δ_{SUM} (5...6), который передается на ftp-сервер (6...7). После этого библиотеки последовательно (7...8) получают файл суммарных изменений ЭК Δ_{SUM} за последний период времени для всех библиотек. После того, как файл Δ_{SUM} доставлен в каждую библиотеку (момент 8) происходит применение содержащихся в нем инструкций на изменение согласованного снимка ЭК от момента времени последнего согласованного состояния с другими ЭК системы (утро предыдущего дня).

Таким образом, на утро в каждой библиотеке будет:

- равенство ЭК локальных АБИС- согласованное состояние АБИС, когда множество записей каждого ЭК тождественно равно множеству любого другого ЭК корпоративной системы;
- ЭК библиотек будет суммарная информация об изменениях библиографических записей, сделанных за последний день во всех библиотеках.

Далее рассмотрены причины и ошибки, которые могут привести к срыву нормальной работы корпоративной АБИС и срыву процесса тиражирования ЭК, в частности.

1. **Отсутствие информационного канала связи** в периоды времени (2...3) и (7...8). При отсутствии возможности передачи или приема файлов локальных или суммарных изменений Δ_{SUM} принимается решение о прекращении всех операций тиражирования и продолжении работы локальной АБИС в рассогласованном состоянии. При этом построение файла локальных изменений ЭК будет проводиться относительно последнего снимка ЭК в согласованном состоянии. И при первом удачном соединении будут загружены файлы изменений Δ_{SUM} от тех дней, когда информационный канал отсутствовал. Затем будет последовательно проведено изменение локального рассогласованного ЭК в соответствии с файлами изменений Δ_{SUM} .
2. **Ошибки при построении файлов локальных изменений ЭК $\Delta_1... \Delta_4$.** Можно определить несколько типов таких ошибок:
 - а. **Ошибки применения локальных ограничений.** Эти ошибки возникают при анализе записей, для которых был нарушен принцип уникальности алгоритмического кода записи, т.е. в ЭК встречаются две записи, имеющие одинаковый алкод и система тиражирования не может применять локальные ограничения [6].
 - б. **Пустые записи.** В процессе работы происходит сохранение некоторых записей без библиографической информации, для такой записи невозможно построить алкод. Как правило, это «пустые» записи, содержащие в полях данных только значения умолчаний (шаблона) полей. В используемой схеме при построении файлов локальных изменений $\Delta_1... \Delta_4$ «пустые» записи удаляются и в процессе тиражирования не участвуют.
3. **Ошибки передачи данных $\Delta_1... \Delta_4$,** из библиотеки-филиала в центральную библиотеку. Возникают, когда данные не могут быть переданы по техническим причинам или в момент, когда был произведен прием суммарного файла изменений Δ_{SUM} , но в этом файле отсутствуют сведения о том, что при его создании были использованы инструкции отправленного локального файла изменений $\Delta_1... \Delta_4$. В таком случае принимается решение о прекращении процесса тиражирования и ЭК локальной АБИС находится в несогласованном состоянии до следующей попытки тиражирования.
4. **Ошибки приема суммарного файла изменений Δ_{SUM}** в какой-либо библиотеке-филиале. Возникают в момент приема файла по техническим причинам или по причинам несогласования времени (6...7), когда для создания суммарного файла потребовалось больше времени, чем было отведено. При этом библиотека-филиал делает запрос на загрузку файла изменений Δ_{SUM} , но запрошенный файл на ftp-сервере отсутствует.
5. **Невозможность работы библиотеки-филиала в структуре распределенной АБИС.** Подобные ошибки можно отнести скорее к организационным и ошибкам «человеческого фактора», нежели к техническим. Самые распространенные из таких ошибок:
 - а. **Невозможность работы сервера библиотеки-филиала в круглосуточном режиме** по причине, в частности, пожарной безопасности. В этом случае программа настраивалась таким образом,

что компьютер автоматически включается от встроенного в BIOS будильника, а после процесса тиражирования все приложения закрываются, и компьютер выключается.

- b. **Сервер библиотеки-филиала вечером был непреднамеренно выключен.** ЭК в этом случае находится в несогласованном состоянии до следующего момента тиражирования.
- c. **Частые случаи работы с ЭК в несогласованном состоянии.** При накоплении статистики работы программы тиражирования и анализе полученных данных можно будет говорить об относительной величине пропусков тиражирования, при которой возможна стабильная работа системы. На данном этапе для каждой библиотеки-филиала принята пороговая величина 30% неудачных попыток тиражирования за последние 10 дней, при превышении которой система извещает администратора о проблемах в работе библиотеки-филиала.

Представленные выше ситуации, приводящие к срыву тиражирования как единичного процесса, на практике не представляют серьезных проблем для работы корпоративной АБИС. Однако при появлении частых сбоев в работе системы тиражирования может привести к появлению конфликтных ситуаций в библиографических записях. Конфликтные ситуации, возникающие при работе ЭК в несогласованном состоянии можно определить:

1. **Конфликты при запаздывании данных.** Возникают, если в какой-либо библиотеке несколько дней ЭК находился в несогласованном состоянии, и некоторая запись подверглась изменению. В то же время в другой библиотеке та же запись была аналогичным образом изменена, причем изменения вступили в силу при первом удачном процессе тиражирования. При использовании алкодов в этом случае будет невозможно найти повторение поля, которое требуется изменить, поскольку оно в согласованной базе уже было изменено ранее. Решение в таком случае принимается по последнему значению технологического поля, в котором содержится дата последнего удачного обновления этой библиографической записи.
2. **Конфликты при нарушении принципа уникальности алгоритмического кода.** При возникновении нарушения уникальности алкода, система автоматически производит полное сравнение записей на этапе создания суммарного файла изменений Δ_{SUM} , и редактирует технологические поля, участвующие в построении алкода, чтобы вернуться к состоянию уникальности алкода.
3. **Конфликты, связанные с другими ЭК, используемыми в локальных АБИС библиотек.** Подобные конфликты могут возникнуть при использовании системы книговыдачи на основе локальных баз данных читателей или при использовании авторитетных баз данных, установленных только в некоторых библиотеках и использующие данные номера записи для организации ссылок на записи, находящиеся в разных базах данных. Такие конфликты не будут влиять на работу системы тиражирования, однако нормальное функционирование локальной АБИС будет нарушено. Решение этой проблемы видится в запуске системы тиражирования для всех баз данных системы библиотек.
4. **Конфликты, возникающие при использовании глобального удаления записей или глобальной корректировки.** Подобные конфликты связаны скорее с психологической составляющей, нежели с устойчивостью системы [6]. На практике возникают ситуации, когда в библиотеке-филиале по причине недостаточной квалификации сотрудников или несанкционированного доступа произошло глобальное удаление или изменение данных. При этом ограничения на изменения данных системы тиражирования для этой библиотеки не позволят ошибочной информации попасть в суммарный файл изменений Δ_{SUM} . Но в этом

случае требуется уведомление лица, ответственного за ведение ЭК, об имевшем место превышении некоторого порогового значения количества удаленных или измененных библиографических записей для конкретной библиотеки. Это делается в автоматическом режиме по электронной почте.

5. **Конфликты, возникающие при «перемешивании» библиографических записей** в локальной АБИС. ЭК отличаются последовательным расположением записей. При удалении библиографических записей появляется пустая запись, которая затем заполняется новыми данными. Существуют технологии, которые позволяют провести изменение места расположения записей в ЭК. Иногда это необходимо для задач ускорения последовательного поиска. При этом сравнение записей по месту расположения для построения локального файла изменений АБИС в библиотеке приведет к непредсказуемым результатам. Использование же алгоритмических кодов позволит осуществить тиражирование ЭК даже при «перемешивании» записей в локальном ЭК. При этом увеличится время на создание файла локальных изменений $\Delta_1 \dots \Delta_4$ за тот день, когда было проведено «перемешивание».
6. **Конфликты, возникающие при локальном восстановлении данных** в ЭК. Надежность технических и программных средств конечна. Возникают ситуации, приводящие к потере всего массива информации в какой-либо библиотеке или нескольких библиотеках одновременно. Для восстановления нормальной работы требуется использовать последнюю согласованную версию ЭК любой другой библиотеки.
7. **Конфликты приоритета изменений данных.** Возникают в случае изменения одной и той же библиографической записи в разных библиотеках в один день. Анализ измененных полей библиографической записи позволяет просуммировать изменения полей и избежать конфликтов на уровне полей и их повторений. Однако может возникнуть ситуация, когда одно и то же поле в нескольких библиотеках изменено на разные значения. Для решения таких конфликтов используется принцип ранжирования библиотек [6]. Ранг конкретной библиотеке присваивается ответственным лицом на основе данных о квалификации сотрудников. В результате будут оставлены только изменения конфликтных полей, выполненных в библиотеке с высшим рангом.
8. **Конфликты блокировок записей локальных АБИС.** Для работы в сетевом режиме имеющиеся на рынке АБИС используют технологии блокировок редактируемых записей. Заключается это в следующем: пока с записью в ЭК работает кто-то, никто другой не может проводить в ней изменения. На этапе эксплуатации АБИС возникают аварийные завершения работы клиентских приложений с заблокированными записями в ЭК в момент работы системы тиражирования. Для устранения подобных конфликтов при запуске системы тиражирования производится принудительная разблокировка всех библиографических записей в ЭК.
9. **Конфликты «промежуточного уровня».** Самый неизученный на сегодняшний день класс конфликтов. Возникают при использовании системы тиражирования совместно с ПО «промежуточного слоя», например шлюзами. Решение подобных конфликтов видится в четком согласовании всех возможных значений данных, используемых в распределенной системе.

Анализ проявившихся конфликтных ситуаций целостности библиографических данных на этапе эксплуатации корпоративной системы в МИБС Томска с ноября 2002 года, позволяет сделать вывод о возможности существования подобных корпоративных систем, обеспечивающих нормальное функционирование локальных АБИС библиотек.

Литература

1. **Когаловский М.Р.** Энциклопедия технологий баз данных. – М.: Финансы и статистика, 2002. – 800 с.
2. **Коннолли Т., Бегг, К., Страчан, А.** Базы данных: проектирование, реализация и сопровождение. Теория и практика, 2-е изд.: Пер. с англ. – М.: Издательский дом «Вильямс», 2001. – 1120 с.: ил. – Парал. тит. англ.
3. **Таненбаум Э.** Распределенные системы: Принципы и парадигмы / Э. Таненбаум, М. Ван Стеен; В. Горбунков (пер. с англ.). – СПб.: Питер, 2003. - 876 с.
4. **Карауш А.С.** Модель тиражирования библиографических баз данных с использованием алгоритмических кодов записей // «VIII Конф. по электронным публикациям «EL-PUB2003»: Сб. тез. докл. – Новосибирск: ИВТ СО РАН, 2003. – С. 14-15.
5. **Карауш А.С.** Программное обеспечение для автоматической синхронизации баз данных системы ИРБИС / А.С. Карауш, Д.Ю. Копытков // Науч. и техн. б-ки - 2003. – № 10. - С. 88-91.
6. **Карауш А.С.** Вопросы обеспечения ответственности при распределенном ведении электронного каталога // «Информационные технологии, компьютерные системы и издательская продукция для библиотек»: Доклады и тез. докладов. – М.: ГПНТБ России, 2003. – С. 117-120.