

Тиражирование электронных каталогов библиотек

Копытков Д.Ю., Цой К.В., Карауш А.С., Кравчук С.С.

- (1, 2) *Томский университет систем управления и радиоэлектроники, г.Томск
Кафедра радиоэлектроники и защиты информации, 5 курс, гр.180*
 - (1) *E-mail: ckv@library.tomsk.ru*
 - (2) *E-mail: duk@library.tomsk.ru*
- (3) *Томский университет систем управления и радиоэлектроники, г.Томск
Кафедра радиоэлектроники и защиты информации, ст. преподаватель*
E-mail: ask@library.tomsk.ru
- (4) *Томский университет систем управления и радиоэлектроники, г.Томск
Кафедра радиоэлектроники и защиты информации, 4 курс, гр.1А1*
E-mail: kss@library.tomsk.ru

Рассмотрена схема тиражирования электронных каталогов библиотек на примере Муниципальной информационной библиотечной системы г.Томска. Решены задачи репликации, сохранности и безопасности как определяемых и передаваемых данных, так и технологических данных, требующихся для объединения территориально распределенных подразделений организации в единую информационную систему. Описана схема синхронизации (тиражирования, репликации) баз данных структуры CDS/ISIS, рассмотрено созданное на основе разработанной и представленной модели тиражирования программное обеспечение Relication.

Муниципальная информационная библиотечная система (МИБС) г. Томска имеет распределенную структуру, в состав которой входят 19 библиотек и 6 отделов. В восьми библиотеках в настоящее время имеются локальные вычислительные сети, которые составляют единое информационное пространство МИБС. Подключение к сети Интернет библиотек-филиалов осуществляется с помощью модемных соединений или с использованием Томского сегмента сети Интернет. Основным продуктом, создаваемым и используемым МИБС для основной деятельности, является единый электронный каталог (ЭК) всех документов, поступающих во все библиотеки с информацией обо всех возможных состояниях этого документа и фиксированием связей, которые могут быть документом получены в процессе «жизни» (пути документа), например хранение всех состояний связи «книга-читатель».

Распространенные и описанные способы объединения библиотек в единое информационное пространство на основе тиражирования данных не могут быть использованы. Это объясняется следующими особенностями, которые не позволяют полноценно функционировать технологическому процессу распределенного создания и редактирования электронного каталога в процессе работы систем каталогизации и книговыдачи в каждой из библиотек-филиалов, а именно:

- 1) количество библиотек способных подключиться для доступа в Интернет больше входных линий, вследствие чего, синхронизация каких-либо данных в синхронном режиме не представляется возможным;
- 2) при передаче большого объема информации по модемным соединениям, существует большая вероятность разрыва соединения;
- 3) желаемый объем передачи данных между библиотеками превышает пропускную способность линий (максимальная скорость передачи данных, в данном случае, может достигать 33 KBit). В нашем случае, размер электронного каталога превышает 600 Мб, а синхронизовать его желательно не реже раза в день;
- 4) нет технической возможности организации скоростной связи с библиотеками-филиалами, расположенными в отдаленных районах города;
- 5) отсутствие штатных технологий тиражирования и репликации данных библиографических записей;

б) особенности применения специальных моделей и схем данных, принятых в библиотечном деле и зафиксированных в виде системы государственных стандартов по библиотечному и издательскому делу.

Вследствие выше описанных особенностей возникла потребность в новых алгоритмах синхронизации и репликации библиографических данных в асинхронном режиме [1-10].

На рисунке 1 представлена используемая схема синхронизации единого электронного каталога в асинхронном режиме.

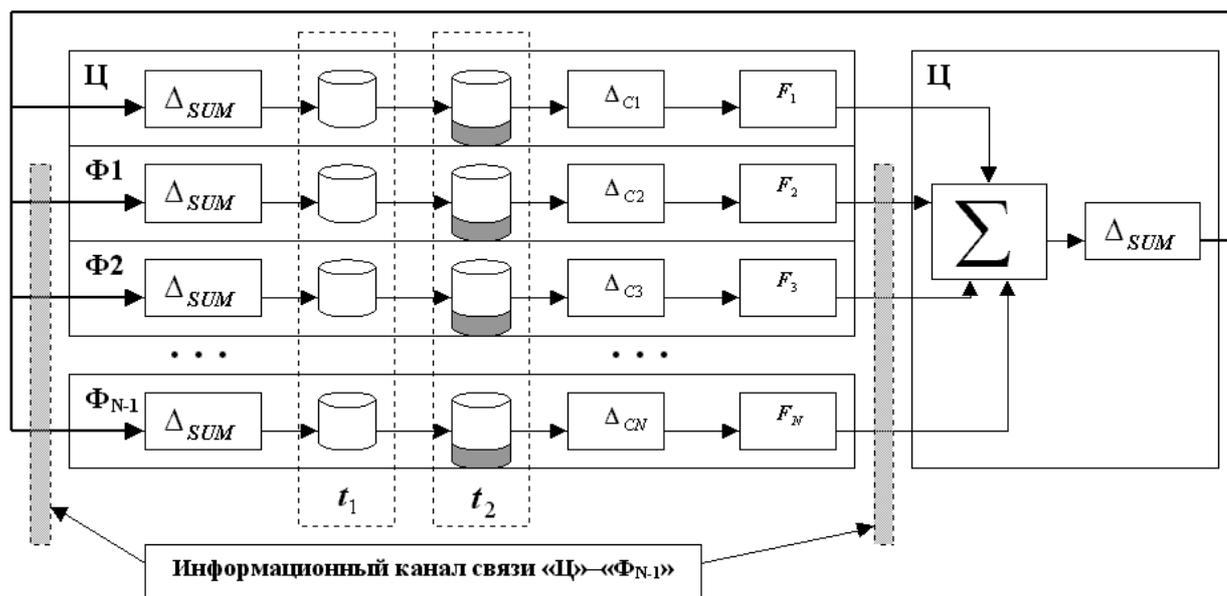


Рисунок 1. Схема тиражирования данных

Принцип работы схемы, представленный на рисунке 1 следующий: в каждой библиотеке (Ц, Ф1, Ф2, ..., ФN-1) установлена локальная копия единого ЭК. В момент времени t_1 ЭК всех библиотек равны. В течение рабочего дня в каждом локальном ЭК производится создание новых, удаление или изменения записей. В конце рабочего дня t_2 по файлу отчета СУБД, если такая возможность имеется, или методом вычитания содержимого библиографических записей в момент времени t_1 строится файл изменений локального ЭК библиотеки Δ_{CN} , который затем обрабатывается с учетом локальных ограничений на изменения ЭК конкретной библиотеки F_N .

Далее, отредактированные с учетом ограничений, из всех библиотек системы файлы изменений локальных ЭК поступают на интеллектуальный сумматор Σ , установленный в центральной библиотеке (Ц), который производит построение единого файла изменений единого ЭК Δ_{SUM} за период времени $t_2 - t_1$ для всех библиотек. После этого файл Δ_{SUM} поступает по каналам связи в каждую библиотеку (Ц, Ф1, Ф2, ..., ФN-1), где в соответствии с его содержимым производится изменение локального ЭК в состоянии на момент времени t_1 (утро предыдущего рабочего дня). После этого осуществляется подмена ЭК локальной АБИС в каждой библиотеке. Таким образом, на утро следующего рабочего дня ($t_1 + 24$ часа) в каждой библиотеке будет:

- равенство локальных ЭК библиотек;
- в локальных ЭК будет суммарная информация об изменениях библиографических записей с учетом ограничений на изменение, сделанных за период $t_2 - t_1$ во всех библиотеках системы.

Особенностью работы представленной модели корпоративной АБИС являются:

- наличие «мертвого» периода времени, который не позволяет работать с локальным ЭК в библиотеке;
- минимизация передаваемой по каналам связи информации об изменениях в ЭК библиотек (Φ , Φ_1 , $\Phi_2, \dots, \Phi_{N-1}$), так как передается только значение добавленных записей, измененных полей в записях и алгоритмические коды удаленных записей;
- необходимость использования функций синхронизации отсчетов времени t_1 и t_2 в библиотеках;
- возможность определения любого уровня доступа к записям локального ЭК в библиотеке в период времени $t_2 - t_1$, поскольку переданы будут только разрешенные для этой библиотеки изменения, а все остальные (неразрешенные) изменения будут утеряны в момент подмены ЭК.

Особенностью работы представленной схемы синхронизации является использование механизма алгоритмических кодов библиографических записей [5, 9], который позволяет для каждой записи создавать идентификатор с ограниченной длиной, который в свою очередь позволяет работать не с отдельными записями, измененными в библиотеках, а с данными полей и их повторений каждой конкретной записи, используя технологии ранжирования веса библиотек на уровне полей для принятия решения о правильности изменений данных [6-8].

Руководствуясь вышеприведенной логикой, было создано программное обеспечение «Relication» [1-4, 10], которое способно выделять изменения, сделанные в течение дня в локальных копиях ЭК библиотек, с использованием системы CDS/ISIS, пересылать эту информацию на корпоративный сервер организации, затем суммировать изменения во всех филиалах и готовить специальные порции информации (дельты) для дополнения баз данных библиотек-филиалов. При этом информация об изменениях в базах данных передается как из каждого филиала в центр, так и из центра в филиалы. Таким образом, на начало рабочего дня библиотека-филиал имеет базу данных из центра и всех филиалов с информацией от предыдущего дня. При этом вся работа во всех библиотеках идет с одной базой данных системы.

Основные характеристики ПО «Relication»:

1. Создание в конкретные промежутки времени отчета об изменениях в базах данных CDS/ISIS.
2. Пересылка информации об изменениях на центральный сервер организации.
3. Передача по сети Интернет только изменений полей в базах данных. Это позволяет максимально сократить расходы на оплату доступа в Интернет.
4. Использование модемного соединения или локальной сети для передачи файлов изменений.
5. Суммирование на головном (центральном) сервере библиотечной системы информации от каждого филиала (максимум 255 филиалов).
6. Передача суммарной информации обо всех измененных записях в библиотеках системы в каждый филиал.
7. Резервирование баз данных с использованием настраиваемых временных схем.
8. Автоматическая система дозагрузки изменений каталога при проблемах со связью.
9. Настраиваемая система передачи прав на добавление, редактирование полей и подполей в филиалах.
10. Автоматическое оповещение администратора баз данных по электронной почте при отсутствии связи или проблемах с обновлением.

Технические требования и особенности ПО «Relication»:

1. Требования к аппаратуре: Процессор- Pentium-166, ОЗУ- 32Mb.
2. Программа может работать как «сервис» на компьютерах библиотек-филиалов.
3. Программа работает в среде Windows 9x/NT/2000/XP.

4. Все настроечные файлы имеют xml-формат.

Использование программы «Relication» совместно позволяет создать в удаленных библиотеках-филиалах места доступа к электронному каталогу всей системы библиотек с автоматическим ежедневным обновлением без технического обслуживания. При этом возможно подключение отдельных компьютеров с ежедневной актуализацией электронного каталога, установленных не только в библиотеках для возможности поиска. При использовании простейшей технологической схемы в библиотеке, где будет только доступ к электронным каталогам, в библиотеке необходим компьютер, телефон и модем.

Литература

1. **Карауш А.С., Копытков Д.Ю.** Программное обеспечение для автоматической синхронизации баз данных системы ИРБИС // Науч. и техн. б-ки - 2003. – № 10. - С. 88-91.
2. **Карауш А.С., Копытков Д.Ю.** «IsisCompare» - программа для сравнения Электронных каталогов библиотек. - М.: ВНТИЦ, 2003. – № 50200300829.
3. **Карауш А.С., Копытков Д.Ю.** «IsisUtil» - программа для проверки и коррекции Электронных каталогов библиотек. - М.: ВНТИЦ, 2003. – № 50200300830.
4. **Карауш А.С., Копытков Д.Ю.** «Relication» - тиражирование библиографических баз данных. - М.: ВНТИЦ, 2003. – № 50200300828.
5. **Карауш А.С.** Модель тиражирования библиографических баз данных с использованием алгоритмических кодов записей // «VIII Международная конференция по электронным публикациям «EL-PUB2003»: Сб. тез. докл. – Новосибирск: ИВТ СО РАН, 2003. – С. 14-15.
6. **Карауш А.С.** Вопросы обеспечения обратной связи при работе с библиографическими записями // «Информационные технологии, компьютерные системы и издательская продукция для библиотек»: Доклады и тез. докладов. – М.: ГПНТБ России, 2003. – С. 113-116.
7. **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.
8. **Карауш А.С.** Вопросы целостности записей корпоративных АБИС // IX рабочее совещание по электронным публикациям "EL-PUB2004": Сб. тез. докл. – Новосибирск, 2004.
9. **Карауш А.С.** Модель корпоративного создания и тиражирования электронных каталогов библиотек: Дис...канд. техн. наук. - Новосибирск: Б.и., 2004. - 157 л.- Библиогр.: с. 112-133.
10. **Карауш А.С., Копытков Д.Ю., Кравчук С.С.** Проект "Relication": состояние и перспективы // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: Докл. и тез. докл. междунар. конференции. - М., 2004.