

Модели тиражирования библиографических баз данных

Карауш А.С.

Предложены модели для создания распределенных автоматизированных библиотечных информационных систем (АБИС) на основе использования процесса тиражирования (репликации) библиографических баз данных. Выполнено описание и анализ моделей тиражирования. Введено понятие алгоритмического кода для задач тиражирования и предложена новая модель тиражирования библиографических баз данных с использованием алгоритмического кода библиографической записи. Описание и расчет представленных моделей выполнены с использованием элементов дискретного анализа и комбинаторики. Новая модель тиражирования позволяет создавать распределенные АБИС в системах библиотек без использования круглосуточного канала связи, обеспечив при этом минимальный объем передаваемых данных.

Распределенные автоматизированные информационные библиотечные системы (АБИС) за последние несколько лет получили активное развитие на российском рынке информационных систем. В каждой библиотечной корпорации, существующей в рамках ассоциации АРБИКОН, имеется собственная технология создания и поддержки работоспособности распределенных или сводных электронных каталогов [2, 11, 17].

Задача распределения и территориального разнесения и тиражирования баз данных появляется в библиотеках не случайно. Каждая библиотека состоит в какой-либо системе или корпорации, а то и в нескольких. На уровне вузов это связка библиотека - филиалы - кабинеты курсового проектирования – методические кабинеты, на уровне научных или централизованных библиотечных систем – это связка центральная библиотека – филиал, где необходимо использование электронного каталога центральной библиотеки [13]. Не всегда для решения этих задач можно использовать информационные каналы связи между центральной библиотекой и филиалом, например, канал Интернет, по причинам:

1. постоянный канал связи организовать невозможно по экономическим или иным соображениям;
2. канал связи имеет недостаточную пропускную способность для комфортной работы в библиотеке-филиале;
3. защиты информационных потоков и баз данных.

В этих случаях, как правило, создается распределенная АБИС библиографических баз данных с тиражированием электронного каталога центральной библиотеки в филиалах для задач поиска и распределенного заказа литературы. Кроме этого распределение данных электронного каталога позволяет использовать достоинства подобных систем [5, 6, 8, 9], а именно:

1. **Отражение структуры организации.** Крупные библиотеки имеют, как правило, множество филиалов и отделений, которые могут находиться в разных концах города, страны и даже за ее пределами.
2. **Разделяемость ресурсов и локальная автономность.** Библиографические данные могут быть помещены на тот сервер системы библиотек, на котором зарегистрированы пользователи, которые их чаще всего используют.
3. **Повышение доступности данных.** В централизованных системах и при использовании сводного электронного каталога библиотек отказ центрального сервера вызывает прекращение функционирования всей системы. Однако отказ одного из серверов распределенной базы данных или линии связи между серверами сделает недоступным лишь некоторые базы данных, тогда как вся система в целом сохранит свою работоспособность.
4. **Повышение надежности.** Если организована репликация данных средствами СУБД, в результате чего библиографические записи и их копии будут размещены на более чем одном сервере, то отказ отдельного сервера или линии связи между серверами не приведет к недоступности данных в системе.

5. **Повышение производительности.** Использование распределенной библиографической базы данных может способствовать повышению скорости доступа к данным по сравнению с доступом к удаленной централизованной базе данных.
6. **Экономические выгоды.** В настоящее время считается общепринятым положение, согласно которому намного дешевле собрать из небольших компьютеров систему, мощность которой будет эквивалентна мощности одного большого компьютера, кроме того, гораздо дешевле добавить в сеть новые рабочие станции, чем модернизировать систему с мощным сервером.
7. **Модульность системы.** В распределенной среде расширение существующей системы осуществляется намного проще. Добавление в сеть нового сервера не оказывает влияния на функционирование уже существующих.

Необходимо также описать сложности, с которыми сталкиваются библиотеки при использовании распределенных АБИС:

1. **Повышение сложности.** Распределенные библиографические базы данных, способные скрыть от конечных пользователей распределенную природу данных и обеспечить необходимый уровень производительности, надежности и доступности, являются более сложными программными комплексами, чем централизованные базы данных. Тот факт, что данные могут подвергаться репликации, также добавляет дополнительный уровень сложности в программное обеспечение.
2. **Увеличение стоимости.** Увеличение сложности предопределяет увеличение затрат на приобретение и сопровождение распределенной библиографической базы данных (по сравнению с обычными централизованными базами данных).
3. **Проблемы защиты.** В централизованных системах доступ к данным легко контролируется. Однако в распределенных системах потребуются организовать контроль доступа не только к данным, реплицируемым на несколько различных серверов, но и защиту сетевых соединений.
4. **Усложнение контроля за целостностью данных.** Целостность библиографической базы данных означает корректность и согласованность сохраняемых в ней записей. Требования обеспечения целостности обычно формулируются в виде некоторых ограничений, выполнение которых будет гарантировать защиту информации в базе данных от разрушения.
5. **Отсутствие стандартов.** Функционирование распределенных баз данных зависит от качества используемых каналов связи. В последнее время стали появляться стандарты на каналы связи и протоколы доступа к данным. Кроме того, не существует инструментальных средств и методологий, способных помочь пользователям в преобразовании централизованных систем баз данных в распределенные.
6. **Недостаток опыта.** В настоящее время в России эксплуатируются несколько работающих моделей распределенных библиографических баз данных. Это позволило уточнить требования к используемым протоколам и установить круг основных проблем.
7. **Усложнение процедуры разработки базы данных.** Разработка распределенных баз данных, помимо обычных трудностей, связанных с процессом проектирования централизованных баз данных, требует принятия решения о фрагментации данных, распределении фрагментов по отдельным серверам и создании процедур репликации данных.

Для поддержания библиографических баз данных (электронных каталогов библиотек) в актуальном состоянии используются различные модели тиражирования, или репликации, если использовать англоязычную терминологию процесса [8, 12].

Для описания процесса тиражирования с точки зрения построения модели в дальнейшем используется комбинаторное представление библиографических баз данных, как множеств библиографических записей. Использование моделей дискретного анализа, комбинаторики и теории множеств позволяет логично описать процессы тиражирования в достаточно простом и понятном виде.

На примере использования различных способов тиражирования далее представлены и проанализированы четыре модели. Каждая модель имеет свои плюсы и минусы, используется в библиотечной практике и позволяет создать работоспособную распределенную АБИС.

Для описания моделей тиражирования вводится несколько ключевых понятий, так библиографическая база данных определяется как множество библиографических записей [1]:

$$C = \{B_k \mid k \in N\}, \quad (1)$$

где $N = \{0, \dots\}$ - множество записей в библиографической базе данных.

В свою очередь, библиографическая запись, также определяется как множества полей, входящих в конкретную библиографическую запись:

$$B = \{H_j \mid j \in N\}, \quad (2)$$

где $N = \{0, \dots\}$ - множество полей в каждой библиографической записи,

причем, каждое поле представляет собой также множество H для повторений этого поля:

$$H = \{A_i \mid i \in N\}, \quad (3)$$

где A - поле библиографической записи, которое представляет также множество атомарных текстовых строк, называемых подполями, в которые заносится непосредственно информация, касающаяся библиографического описания информационного источника.

$$A = \{a_i \mid i \in S\}, \quad (4)$$

где $S = \{0, \dots, 9, a, \dots, z\}$ -

множество подполей в повторении поля библиографической записи, которые могут принимать значения меток букв латинского алфавита и одиночных цифровых символов.

Основная задача состоит в определении наиболее приемлемого решения для синхронизации изменений библиографических данных, выполненных в течение некоторого промежутка времени $(t_2 - t_1)$ и $(t_2 > t_1)$ в центральной библиотеке с тем, чтобы наилучшим образом привести базу данных в библиотеке-филиале в состояние равенства с базой данных центральной библиотеки.

Итак, пусть, C_{t_1} – база данных C в момент времени t_1 ,

C_{t_2} – база данных C в момент времени t_2 , при условии $t_2 > t_1$.

Самый простой способ тиражирования библиографических баз данных представляет собой замену файлов баз данных в библиотеке-филиале на такие же файлы, доставленные из центральной библиотеки [10, 15]. Данная модель тиражирования является самой простой для создания и настройки. Для нее можно определить основные достоинства и недостатки.

Достоинства.

1. **Простота.** При использовании данной модели необходимо определить моменты времени для тиражирования библиографических баз данных и создать систему, которая позволит доставлять актуальные копии базы данных в библиотеки-филиалы.
2. **Экономичность.** При передаче информации на перезаписываемом компакт диске (CD-RW или DVD-RAM) курьером стоимость передачи одного бита составляет минимальное значение среди всех остальных возможных способах тиражирования баз данных при том же объеме передаваемых данных.

3. **Надежность.** Высокий уровень надежности здесь достигается многократным копированием библиографической базы данных в несколько территориально распределенных библиотек-филиалов, что позволяет в случае утраты базы данных в какой-либо библиотеке восстановить последнюю версию файлов баз данных.

Недостатки.

1. **Невозможность распределенной работы с базой данных.** Заключается это, прежде всего в том, что если в какой-либо библиотеке-филиале сделали редактирование или добавление каких-либо полей в базу данных, то при очередном тиражировании они будут утрачены.
2. **Недостаточная актуальность данных.** На практике обновления файлов библиографических баз данных в системах библиотек происходит не чаще двух раз в месяц, что явно недостаточно для передачи данных о вновь поступивших источниках в фонды системы библиотек.

Таким образом, данная модель позволяет работать с тиражируемыми библиографическими базами данных системе библиотек, но не позволяет создавать распределенные информационные библиотечные системы, которые бы обеспечивали приемлемый для пользователя (1 день) период актуальности библиографических данных.

Для уменьшения объема данных, передаваемых в момент тиражирования, необходимо использовать другие модели, которые строятся на утверждении того, что при достаточно большом объеме библиографической базы данных объем выполненных изменений группой каталогизаторов будет конечным и несравненно меньшим, чем весь объем базы данных. Основная задача для работы таких моделей синхронизации определяется необходимостью использования алгоритмов, которые бы позволили определить измененные записи в библиографической базе данных за период времени.

Таким образом, для имеющихся баз данных C_{t1} и C_{t2} можно определить множества C_{const} и C_{Δ} , такие что:

$$C_{const} = C_{t1} \cap C_{t2} \quad (6)$$

$$C_{\Delta} = C_{t1} \Delta C_{t2} \quad (7)$$

Т.е. операция пересечения для множеств C_{t1} и C_{t2} даст множество C_{const} , состоящее из библиографических записей, которые не изменялись в течение промежутка времени $t_2 - t_1$. Напротив, операция симметричной разности множеств C_{t1} и C_{t2} даст множество C_{Δ} , состоящее из записей B_k , которые были отредактированы, удалены или созданы заново в течение времени $t_2 - t_1$ и входят в одно из множеств C_{t1} или C_{t2} . При этом выполняется предположение, что если библиографическая запись подверглась редактированию, то она представляет собой новое значение последовательности C_{t2} и с предыдущим своим состоянием в последовательности C_{t1} связь теряет.

Основываясь на вышеприведенных утверждениях можно определить множества:

$$C_{t1} \cap C_{\Delta} = C_{\Delta1} \quad (8)$$

$$C_{t2} \cap C_{\Delta} = C_{\Delta2} \quad (9)$$

$$C_{\Delta1} \cup C_{\Delta2} = C_{\Delta}, \quad (10)$$

где $C_{\Delta1}$ - множество библиографических записей, которые входят во множество C_{t1} , т.е. записи, которые были изменены или удалены во множестве C_{t1} за время $t_2 - t_1$;

$C_{\Delta2}$ - множество записей, которые были изменены или добавлены в базу данных C_{t1} за время $t_2 - t_1$, но не входят во множество C_{t1} .

Используя вышеприведенные рассуждения и формулы (1)-(10), можно создать модель тиражирования библиографической базы данных, измененной в центральной

библиотеке за период времени $t_2 - t_1$. Эта модель позволяет обеспечить существования «зеркального» электронного каталога в библиотеках-филиалах (рис.1). Объем передаваемой информации по каналам передачи данных будет ограничен только множеством библиографических записей C_{Δ} , которое состоит из измененных, удаленных или добавленных записей. Такая модель уже существует и работает при тиражировании систем правовых баз данных.

Алгоритм, основанный на использовании модели тиражирования методом передачи изменений базы данных, содержит следующие операции:

- сравнение библиографических записей в базах данных C_{t_1} и C_{t_2} ;
- определение записей, которые были удалены, отредактированы или добавлены за период времени $t_2 - t_1$, согласно (7)-(9);
- передача по каналам связи множеств библиографических записей $C_{\Delta 1}$ и $C_{\Delta 2}$;
- удаление в базе данных библиотеки-филиала записей, входящих во множество $C_{\Delta 1}$;
- добавление в базу данных библиотеки-филиала записей, входящих во множество $C_{\Delta 2}$.

Использование данного алгоритма позволяет создать механизм тиражирования электронного каталога центральной библиотеки в автоматическом режиме. При этом объем информации, передаваемой по каналам связи, будет равен суммарному объему удаленных, добавленных записей и измененных записей базы данных, т.е. суммарному объему множеств $C_{\Delta 1}$ и $C_{\Delta 2}$.

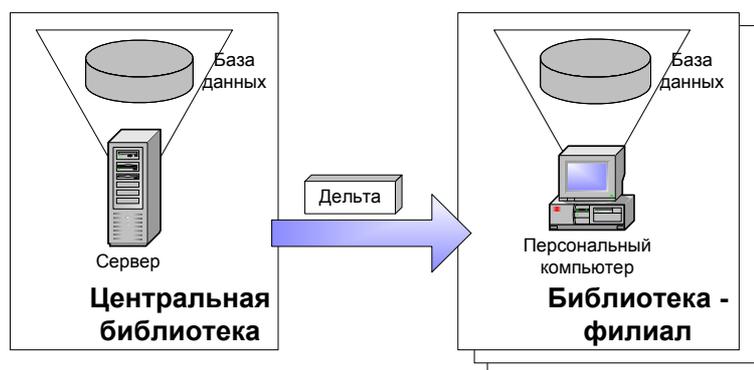


Схема работы алгоритма для тиражирования библиографических баз данных методом передачи изменений базы данных (дельта).

Рис. 1.

Для модели тиражирования библиографических баз данных методом передачи измененных записей можно представить основные достоинства и недостатки.

Достоинства:

1. **Относительная простота.** Алгоритм может работать на всех моделях построения баз данных, не вдаваясь в подробности состава библиографической записи.
2. **Уменьшение сетевого трафика.** По каналам связи будут переданы только удаленные, измененные или добавленные библиографические записи в базе данных за период $t_2 - t_1$.
3. **Возможность работы в асинхронном режиме и независимость синхронизации от времени.** При определении периода времени $t_2 - t_1$, нет необходимости использовать какие-либо схемы. Полученное множество

C_{Δ} будет однозначно определять изменения базы данных от значений t_1 или t_2 .

Недостатки:

1. **Невозможность редактирования базы данных в филиале.** В данной модели отсутствует возможность вносить изменения в базу данных, находящуюся в библиотеке-филиале.
2. **Требуются большие вычислительные мощности.** Особенностью библиографических баз данных является непостоянство идентификатора библиографической записи. За время $t_2 - t_1$ записи могут быть «перемешаны» в базе данных. Поэтому, алгоритм должен сравнивать каждую библиографическую запись в базе данных C_{t_1} с каждой записью в базе данных C_{t_2} .

При рассмотрении достоинств и недостатков данной модели следует обратить внимание на то, что ее использование оправдано в централизованных библиотечных системах, где ставится задача тиражирования библиографических баз данных в библиотеках-филиалах только для задач поиска и заказа, но не для редактирования и создания новых записей. Применение же данной модели и алгоритмов на ее основе для построения распределенных автоматизированных библиотечных систем для каталогизации ресурсов потребует создания схемы (рис.2), где каждая библиотека имеет собственную базу данных, которая тиражируется в автоматическом режиме в любую другую библиотеку [10]. Модель создания распределенной АБИС, показанная на рис.2, можно определить как модель распределенного создания и тиражирования библиографической базы данных, на основе копирования записей из распределенных тиражируемых библиографических баз данных.

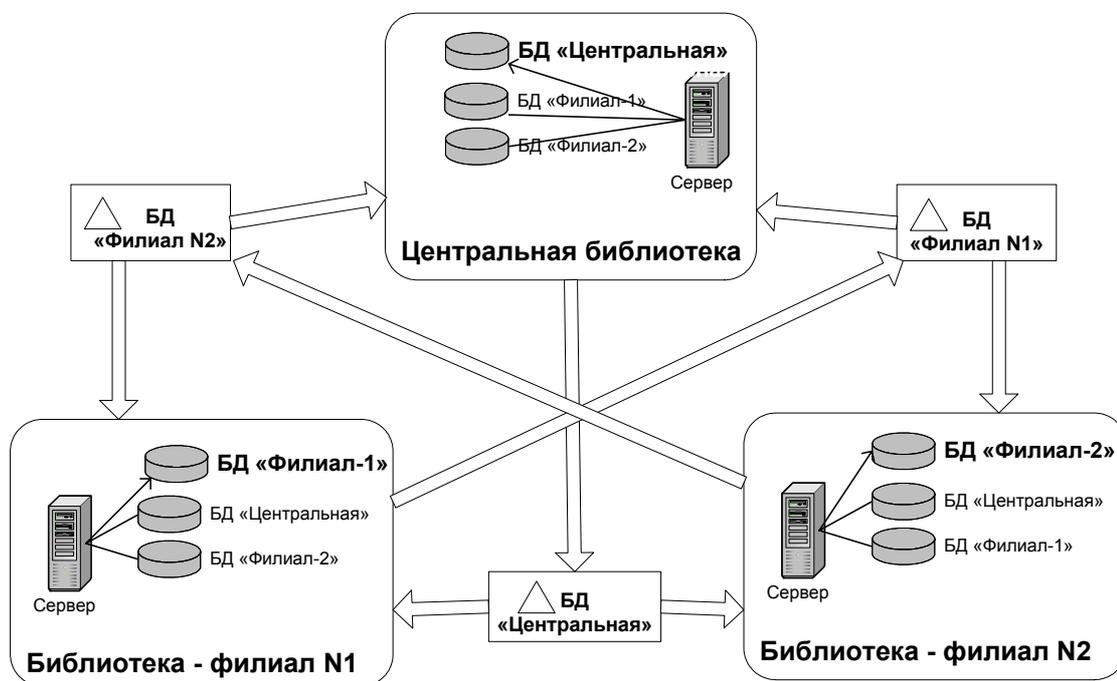


Схема распределенной автоматизированной библиотечной информационной системы. Модель распределенного создания и тиражирования библиографической базы данных, на основе копирования записей из распределенных тиражируемых библиографических баз данных.

Рис.2.

На рис.2. представлена схема распределенного создания библиографических баз данных системы библиотек с возможностью тиражирования баз данных библиотек системы на устройства хранения друг друга. Условные обозначения на рис.2.:

- **БД «Центральная»** - библиографическая база данных, создаваемая в центральной библиотеке, выполняет задачи сводного электронного каталога системы библиотек для задач поиска и заказа информационных источников.
- **БД «Филиал N1», БД «Филиал N2»** - библиографические базы данных, создаваемые и редактируемые соответственно в библиотеке-филиале N1, библиотеке-филиале N2.
- **ДБД «Центральная»** - множество удаленных, измененных или добавленных записей C_{Δ} для библиографической базы данных центральной библиотеки, созданное за определенное время;
- **ДБД «Филиал-1»** - множество удаленных, измененных или добавленных записей C_{Δ} для базы библиографической базы данных библиотеки-филиала 1;
- **ДБД «Филиал-2»** - множество удаленных, измененных или добавленных записей C_{Δ} для базы библиографической базы данных библиотеки-филиала 2.

Последовательность действий по наполнению сводной библиографической базы данных системы библиотек можно в этом случае определить как:

1. создание библиографической записи на информационный источник в библиотеке-филиале;
2. тиражирование базы данных библиотеки-филиала, где создана запись на сервер центральной библиотеки;
3. проверка принятой библиографической записи и ее коррекция в соответствии с требованиями корпоративной каталогизации.
4. копирование записи из базы данных библиотеки-филиала в базу данных центральной библиотеки (сводный электронный каталог).
5. тиражирование базы данных центральной библиотеки в библиотеки-филиалы.

Время с момента создания библиографического описания в библиотеке-филиале и до момента, когда пользователю будет доступен источник для поиска в тиражируемой базе центральной библиотеки из библиотеки-филиала, представляет сумму времени, требующегося на выполнение вышеприведенной последовательности действий. При этом осуществить выполнение нескольких действий, определенных пунктами 1-5 параллельно, не представляется возможным. Для представленной на рис.2. модели распределенного создания и тиражирования библиографической базы данных, на основе копирования записей из распределенных тиражируемых библиографических баз данных можно определить достоинства и недостатки.

Достоинства.

1. **Простота реализации.** Использование данной модели позволяет создавать распределенные системы в рамках нескольких библиотек, используя организационную структуру библиотечных систем.
2. **Качество библиографических записей.** Качество гарантировано двойным уровнем контроля, который состоит из выходного контроля ответственного за создания базы данных библиотеки-филиала и входного контроля отдела баз данных (каталогизации) центральной библиотеки.
3. **Дешевизна использования.** При наличии в библиотеке-филиале компьютера с каналом связи возможно использование данной модели. Основные настройки программного обеспечения АБИС тиражируются во всех библиотеках. И если во всех библиотеках системы используют одни и

те же правила каталогизации информационных источников, то данная модель позволяет исключить затраты, связанные с содержанием обслуживающего персонала информационных систем в библиотеках-филиалах, и при этом осуществлять контроль за функционированием информационной системы.

4. **Уменьшение сетевого трафика.** Для нормального функционирования системы достаточно процесса передачи изменений библиографических баз данных системы библиотек между собой. При этом передаются только множества библиографических записей C_{Δ} каждой из библиотек системы.
5. **Распределенное резервирование баз данных.** Резервирование данных библиографических баз данных – весьма актуальная проблема для библиотек. При работе данной модели при самом плохом случае потери составят только библиографические записи созданные в этой библиотеке с момента последнего C_{Δ} тиражирования данных.

Недостатки:

1. **Затраты на проверку данных.** Положительной стороной данной модели было наличие двойного контроля качества библиографических записей, однако это влечет затраты времени персонала библиотек.
2. **Невозможность использования единой библиографической базы данных.** Особенность данной модели не предоставляет возможностей для работы во всех библиотеках системы в единой базе данных.

Совокупность положительных сторон и недостатков данной модели обеспечило ее популярность. Многие библиотечные системы используют подобные модели интуитивно, полагаясь на положительные стороны, отличающие данную модель, однако улучшение производственных показателей своей работы напрямую не связывают с использованием той или иной модели создания библиографических баз данных.

Дальнейшее улучшение и усовершенствование моделей создания и тиражирования библиографических баз данных возможно при использовании ключей библиографических записей. Подобная практика разработки ключей для библиографических записей существует для определения дублетных записей, а также для кодирования записей на основе имеющейся в ее полях информации.

Как было показано ранее на основе (8)-(10), можно определить множества записей, которые были изменены, удалены или добавлены в базы данных C_{t1} и C_{t2} .

На этом анализ данных, подвергшихся редактированию вести невозможно без определения алгоритмического кода (АЛКОДа) для библиографической записи B_k .

Алгоритмический код (АЛКОД) - идентификатор издания, который создается по определенным правилам (алгоритму) и однозначно идентифицирует конкретное издание (источник), позволяет хранить информацию об издании в удобном, компактном виде, а также позволяет осуществлять обслуживание пользователей в автоматизированном режиме.

Пусть G - множество алгоритмических кодов для базы данных C , которое может быть получено с использованием функции преобразования f для каждой записи B_k множества C :

$$G = \{X_k \mid k \in N\}, \quad (11)$$

где X_k - строка алгоритмического кода для конкретной библиографической записи в базе данных C .

$$X_k = f(B_k) \quad (12)$$

В результате выполнения функции f построения АЛКОДа для каждой библиографической записи B_k , будет построена строка X_k , состоящая из символов, находящихся в полях A_{ik} этой записи.

Задача создания АЛКОДа сводится к условию получения одинаковых значений X_k для библиографических записей одного и того же ресурса, описанного разными каталогизаторами, но в то же время, к получению различных значений X_k для различных библиографических ресурсов. На практике, строка АЛКОДа - свертка библиографической записи, содержащая элементы обязательных полей библиографического описания.

В дальнейшем, операции по сравнению множеств библиографических записей C_{t_1} и C_{t_2} сводятся не только к получению множеств C_{const} и C_{Δ} , что определяют состояние библиографической базы данных на период времени t_2 , но и к анализу множества алгоритмических кодов G для записей, входящих во множество C_{Δ} .

Для дальнейших рассуждений требуется ввести множества:

$G_{\Delta 1}$ - множество алгоритмических кодов для библиографических записей $C_{\Delta 1}$;

$G_{\Delta 2}$ - множество алгоритмических кодов для библиографических записей $C_{\Delta 2}$;

Такие, что:

$$G_{\Delta 1} = f(C_{\Delta 1}) \quad (13)$$

$$G_{\Delta 2} = f(C_{\Delta 2}) \quad (14)$$

Библиографическая запись считается отредактированной, если изменения данных, сделаны в полях, которые не участвуют в создании строки АЛКОДа для соответствующей записи. В противном случае запись считается удаленной за период времени $t_2 - t_1$ и вновь созданной с другими данными в полях, участвующих в построении АЛКОДа. Таким образом, множество АЛКОДов для измененных библиографических записей, при условии что значения АЛКОДа не изменялось, можно определить, как G_{const} , получаемое пересечением множеств $G_{\Delta 1}$ и $G_{\Delta 2}$:

$$G_{const} = G_{\Delta 1} \cap G_{\Delta 2} \quad (15)$$

При этом возможно определение АЛКОДов для библиографических записей, которые были удалены G_{del} за время $t_2 - t_1$:

$$G_{del} = G_{\Delta 1} \setminus G_{const}, \quad (16)$$

причем

$$G_{\Delta 1} = G_{del} \cup G_{const} \quad (17)$$

Множество АЛКОДов для библиографических записей G_{Δ} , которые были созданы за время $t_2 - t_1$:

$$G_{add} = G_{\Delta 2} \setminus G_{const}, \quad (18)$$

причем

$$G_{\Delta 2} = G_{add} \cup G_{const} \quad (19)$$

Для дальнейших исследований потребуется:

$$G_{const_{t_1}} = G_{\Delta 1} \setminus G_{del} \quad (20)$$

$$G_{const_{t_2}} = G_{\Delta 2} \setminus G_{add}, \quad (21)$$

причем должно выполняться:

$$G_{const_{t_1}} = G_{const_{t_2}} = G_{const}, \quad (22)$$

где $G_{const_{t_1}}$ - множество АЛКОДов, входящих во множество G_{Δ_1} и определяющих измененные библиографические записи во множестве C_{Δ_1} ;

$G_{const_{t_2}}$ - множество АЛКОДов, входящих во множество G_{Δ_2} и определяющих измененные библиографические записи во множестве C_{Δ_2} ;

Следует также определить действия, которые могут быть выполнены с данными, находящимися в полях библиографической записи:

- удаление повторения поля;
- добавление повторения поля;
- изменение данных в повторении поля.

Для удобства будем считать, что изменение данных в повторении поля есть удаление повторения поля и добавления нового повторения поля.

Таким образом, для определения данных, необходимых для новой модели тиражирования требуется знание того, какие поля были удалены или (и) добавлены в библиографических записях за время $t_2 - t_1$. Удаленные и добавленные поля в библиографических записях, измененных целиком, определены ранее из (16) и (18). Для определения списка измененных полей в библиографических записях имеющих одинаковое значение АЛКОДа, и входящих во множества G_{Δ_1} и G_{Δ_2} , потребуется провести сравнение полей и их повторений на равенство.

Для каждого АЛКОДа множества $G_{const_{t_1}}$, ищется равный ему во множестве $G_{const_{t_2}}$:

$$G_{const_{t_1}} = \{X_{const_k} \mid k \in N\}, \quad (23)$$

$$G_{const_{t_2}} = \{X_{const_p} \mid p \in N\}, \quad (24)$$

$$N = \{0, \dots\},$$

такой что:

$$X_{const_k} = X_{const_p}, \quad (25)$$

при этом для каждой пары k и p производится вычитание множеств полей библиографической записи (26), (27) с равным значением АЛКОДа, такие что:

H_{t_1} - множество полей библиографической записи на момент времени t_1 ;

H_{t_2} - множество полей библиографической записи на момент времени t_2 ;

В результате, для каждой библиографической записи, входящей во множество C_{Δ} , такой, что ее АЛКОД принадлежит множеству G_{const} , будет определено множество полей, которые были удалены H_{del} или добавлены H_{add} за время $t_2 - t_1$:

$$H_{del} = H_{t_1} \setminus H_{t_2}, \quad (26)$$

$$H_{add} = H_{t_2} \setminus H_{t_1} \quad (27)$$

Можно предложить новую модель тиражирования библиографических данных, со схемой, показанной на рис.2., но основанную на использовании АЛКОДов библиографических записей.

Последовательность действий, выполняемых при тиражировании библиографических баз данных с использованием АЛКОДов, представляет:

1. сравнение библиографических записей в базах данных C_{t_1} и C_{t_2} , причем сравнению подлежат каждая запись в базе C_{t_1} с каждой записью в базе C_{t_2} ;

2. определение множеств записей $C_{\Delta t_1}$ и $C_{\Delta t_2}$, которые были удалены, отредактированы или добавлены, согласно (7)-(9);
3. построение АЛКОДов для множеств записей $C_{\Delta t_1}$ и $C_{\Delta t_2}$ согласно (11)-(14);
4. определение множеств G_{del} и G_{add} согласно (16) и (18);
5. определение множеств $G_{const_{t_1}}$, $G_{const_{t_2}}$ (20), (21);
6. определение множеств удаленных H_{del} и добавленных полей H_{add} за время $t_2 - t_1$ в библиографических записях баз данных, согласно (23)-(27).
7. передача по каналам связи:
 - a. множества библиографических записей, которые имеют АЛКОДы, принадлежащие множеству G_{add} , для добавления этих записей в синхронизируемую базу данных филиала;
 - b. множества АЛКОДов G_{del} , для удаления записей с равными АДКОДами, при этом вся запись не передается;
 - c. множеств H_{del} и H_{add} с привязкой к значению АЛКОДа для каждой записи библиографической базы данных.
8. изменение синхронизируемой базы данных в филиале в соответствии со значением переданных множеств записей и АЛКОДов.

Можно определить достоинства и недостатки для модели тиражирования библиографических баз данных с использованием АЛКОДа.

Достоинства:

1. **Минимальный объем данных, передаваемых по каналам связи.** Действительно, по каналам передаются только изменения баз данных, причем если в библиографической записи проводилась редакция только одного поля, то только значение этого поля до и после редакции будет передано.
2. **Последовательность расположения записей в базах данных не важна.** В этом случае, если произведена реорганизация библиографической базы данных в каком-либо филиале, то это никак не отразится на работе модели тиражирования. Таким образом, записи в центральной или синхронизируемой базе данных за время между моментами синхронизации могут быть «перемешаны».
3. **Возможность создания алгоритма двухстороннего тиражирования данных,** который позволяет производить изменение тиражируемой библиографической базы данных одновременно в двух библиотеках.

Недостатки:

1. **Сложность алгоритма построения АЛКОДа.** Действительно, основным условием работы данной модели является уникальность АЛКОДа для любой записи, входящей в любую базу данных. Для соблюдения этого условия приходится использовать дополнительные проверки на этапе ввода библиографического описания, когда АЛКОД может строиться автоматически.
2. **Возможность конфликтных ситуаций.** Такое может произойти при одновременном (в период времени $t_2 - t_1$, между тиражированием) изменении в разных библиотеках одной и той же записи.
3. **Дополнительные вычислительные ресурсы.** Сравнение библиографических записей попарно, находящихся в базах данных C_{t_1} и C_{t_2} требует дополнительных вычислительных ресурсов.

При анализе недостатков последней рассмотренной модели можно сказать, что конфликтные ситуации, возникающие в процессе тиражирования, можно существенным

образом уменьшить, если ввести алгоритмы дополнительного анализа данных в полях библиографической записи. Затраты вычислительных ресурсов можно существенно сократить, если вместо сравнения библиографических записей целиком использовать для сравнения их АЛКОДы. И при утверждении, что в течение небольшого отрезка времени в библиографической базе данных изменяется незначительное количество записей, относительно всего объема базы данных, возможно упрощение алгоритмов сравнения записей.

Выводы

Основные причины перехода библиотек от централизованного процесса создания и предоставления своего электронного каталога к распределенному, по видимому, кроются не столько в улучшении качества обслуживания, сколько в создании более эффективных схем управления и организации информационных и производственных ресурсов. Описанные в статье модели тиражирования библиографических баз данных нашли и находят свое применение в распределенных АБИС. На основе этих моделей возможно построение эффективной схемы предоставления электронного каталога библиотеки или системы библиотек, а также подготовка структуры организации для введения новых форм управления фондами и персоналом.

Список используемой литературы

1. UNIMARC Manual. Руководство по UNIMARC / Пер. на рус. яз. коллектива под рук. А.И. Земскова, Я.Л. Шрайберга. М.: ГПНТБ России, 1992. 319 с.
2. АБИС “Руслан” – от локальной автоматизации к корпоративным проектам/ Баранов В. Л., Сова Д. Н.// Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: Материалы конф. – М., 2002. – Т.2.
3. Асинхронное тиражирование данных в гетерогенных средах / Калиниченко Б.О. // Системы Управления Базами Данных – 1996. – N 3. - 118-124
4. Гайдамакин Н.А. Автоматизированные информационные системы, базы и банки данных. Вводный курс: Учебное пособие.- М.: Гелиос АРВ, 2002. – 368 с., ил.
5. Гома Х. UML. Проектирование систем реального времени, параллельных и распределенных приложений.: Пер. с англ. – М.: ДМК Пресс, 2002. – 704 с.: ил. – (Серия «Объектно-ориентированные технологии в программировании»).
6. Дейт, К., Дж. Введение в системы баз данных, 7-е издание.: Пер. с англ. – М.: Издательский дом «Вильямс», 2001. – 1072 с.: ил. – Парал. тит. англ.
7. Информационная система сопровождения жизненного цикла разработки распределенных систем обработки данных / Григорьев Ю.А. // Вестн. МГТУ. Сер. Приборостр. - 1999. - N 2. - 37-45, 127 - ISSN 0236-3933
8. Когаловский М.Р. Энциклопедия технологий баз данных. – М.: Финансы и статистика, 2002. – 800 с.: ил. - ISBN 5-279-02276-4
9. Коннолли, Томас, Бегг, Каролин, Страчан, Анна 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.
10. Модели построения и функционирования корпоративной информационной сети муниципальных библиотек/ Карауш А.С., Левицкая Л.В.// Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: Материалы конф. – М., 2002. – Т.2. – С. 912-914.
11. Опыт создания региональной распределенной информационно-поисковой системы учреждений науки и образования Северо-Запада России / Племнек А.И., Дорофеев С.Т., Купреенко С.В., Баранов В.Л., Миронов К.Б., Волжев И.А., Стогов Р.В., Агаджанова А.Н. // Пробл. информатиз. - 1996. - N 4. - 44-47

12. Практическая репликация / Луковенко А., Фаритов А. // Открытые системы – 2001. – N12
13. Программное обеспечение для автоматической синхронизации баз данных системы "ИРБИС" / Карауш А.С., Копытков Д.Ю. // 10 -я Международная конференция "Крым 2003" Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества. ", Судак, 8-16 июня, 2003 - М.: Изд-во ГПНТБ России, 2003. - Т. 2.
14. Распределенные библиотечные системы / Баранов В.Л., Племнек А.И., Соколова Н.В. // Науч. и техн. б-ки - 1997. - N 12. - 11-30 - ISSN 0130-9765
15. Репликация данных как управленческая задача: подходы к решению/ Максименко Ю. // ВУТЕ/Россия – 2001. – N2 – 4.
16. Ролланд, Фред, Д. Основные концепции баз данных. : Пер. с англ. – М.: Издательский дом «Вильямс», 2002. – 256 с.: ил. – Парал. тит. англ.
17. Система обеспечения контроля и учета доступа из Интернет к распределенным библиографическим базам данных / Баженов С.Р., Баженов И.С., Мазов Н.А., Малицкий Н.А. // 6-ая Междунар. конф. 'Б-ки и ассоц. в меняющ. мире: нов. технол. и нов. формы сотрудничества', Судак и др., 5-13 июня, 1999: "Крым 99": Матер. конф. - Симферополь: Таврида, 1999. - Т. 1 - 226-228