

АЛЕКСАНДР КАРАУШ,  
 ДМИТРИЙ КОПЫТКОВ,  
 АЛЕКСАНДР МАКАРЕВИЧ

## Утилиты для проверки и коррекции электронных каталогов

**И**СПОЛЬЗОВАНИЕ систем автоматизации в повседневной деятельности обеспечивает повышение производительности труда, но сопряжено с непредвиденными проблемами, вызванными ошибками данных или отказами системы. **Типы непредвиденных проблем в работе информационной системы** могут быть следующими:

- 1) недопустимые простои критически важных сервисов (например, программ опрелделения задолженных документов в филиалах библиотеки для книговыдачи во время рабочего дня);
- 2) прекращение сопровождения критически важных приложений по причине распада внутренней команды разработчиков или фирмы-поставщика;
- 3) масштабное поражение корпоративной сети компьютерным вирусом;
- 4) необходимость срочной легализации применяемого программного обеспечения (ПО) вследствие возникновения правовых проблем и пр.

«Среда обитания» информационной системы слабо поддается контролю и прогнозированию. Особенно это касается вопросов правового и финансового регулирования. **К рискам и проблемам внешней среды**, которые зачастую невозможно контролировать, относятся:

- 1) технические проблемы, не предусмотренные в рамках регулярных процедур планирования и управления информационными технологиями (ИТ), например проблема 2000 года;
- 2) изменение технической политики производителей оборудования и ПО;
- 3) изменение лицензионной политики производителей ПО.

Данные проблемы находятся вне зоны контроля отделов информационных технологий библиотек.

Для решения непредвиденных проблем ведется работа над соответствующими проектами. Подобный проект предполагает медленное изменение инфраструктуры отдела ИТ или всей библиотеки в целом. С этой целью может потребоваться замена (причем срочная) оборудования или ПО, явившегося причиной проблемы.

Как свидетельствует опыт авторов, **наиболее частыми являются проблемы, связанные с отказом компьютерного оборудования и системного ПО**. Однако нередки случаи, когда наряду с вышеперечисленными проблемами **появляются потери данных в АБИС, в том числе в электронном каталоге (ЭК) библиотеки**. Можно предложить следующий перечень встречающихся **проблем состояния баз данных (БД)**, произошедших по причине отказа в работе оборудования или операционных систем:

- 1) физическое уничтожение данных;
- 2) потеря целостности структуры данных в среде СУБД;
- 3) физические нарушения структуры файлов;
- 4) потеря одного или нескольких значимых файлов СУБД, при которой работа АБИС становится невозможной.

Что касается проблемы **физического уничтожения данных**, где самой распространенной причиной является **отказ жестких дисков или их физическое разрушение**, то можно сказать, что ничего сопоставимого по эффективности резервированию данных человечество еще не придумало. По мнению авторов, оптимальный период резервного копирования данных информационных систем составляет 24 часа. Это позволяет свести основные затраты от случайной потери файлов БД к минимуму, за счет того, что потеря данных возможна только в ретроспективе одного дня. Имеются случаи, когда отказ

**При работе автоматизированных библиотечно-информационных систем существует возможность частичной или даже полной потери данных. Программное обеспечение по поиску, локализации и исправлению частых и случайных ошибок каталогизации, разрабатываемое в Муниципальной информационной библиотечной системе г. Томска, позволяет улучшить качество электронных каталогов и баз данных библиотек.**

*Александр Сергеевич Карауш,  
 заместитель директора Муниципальной  
 информационной библиотечной системы,  
 Томск*

*Дмитрий Юрьевич Копытков,  
 программист-технолог Отдела новых  
 информационных технологий Муниципальной  
 информационной библиотечной системы,  
 Томск*

*Александр Сергеевич Макаревич,  
 программист-технолог Муниципальной  
 информационной библиотечной системы,  
 Томск*

оборудования происходит «постепенно», вследствие чего невозможно восстановить данные по копии предыдущего дня. Тогда для нормальной работы системы резервирования следует применять сложные схемы, где объем места, необходимый для хранения копий баз данных, должен в 15-50 раз превышать объем самой базы данных. Это необходимо помнить, приобретая оборудование, чтобы обезопасить организацию от полной или существенной потери данных.

**Потеря целостности данных для локальных АБИС представляет собой нарушение организации данных в одном или нескольких файлах СУБД, в результате чего становится невозможным поиск или показ результатов поиска.** В Системе «ИРБИС32» (CDS/ISIS) такие ситуации возникают при ошибках уровня таблиц размещения файлов на жестком диске и проявляются в невозможности показа данных словарей или появлении ошибок работы АРМов. Решением данной проблемы является выполнение последовательности команд АРМа «Администратор»: «Диагностика файла документов» и «Создание словарей заново».

При проблемах, связанных с **физическим нарушением в структуре файла**, например потерей данных блока или кластера на жестком диске, возможны случаи потерь диапазона записей ЭК, которые:

- 1) не поддаются какой-либо диагностике с точки зрения наличия /отсутствия данных и их содержимого;
- 2) не поддаются анализу по возможной локализации «проблемных» записей;
- 3) не подлежат восстановлению системными утилитами или утилитами, поставляемыми разработчиками ПО.

При появлении подобных проблем **все действия, связанные с доступом к БД, следует производить только после резервирования состояния всех файлов системы.** Практика показала, что желательно делать резервирование файлов СУБД средствами операционной системы с монопольным доступом к файлам, чтобы уменьшить влияние возможных ошибок на физическом уровне размещения данных в файлах.

Восстановление данных при нарушениях структуры файлов может потребовать проверки на дополнительные условия после восстановления записей для утверждения того факта, что в процессе восстановления не были потеряны значимые блоки информации. Например, при сбое системы могут возникнуть случаи, когда записи потеряны, а вместо них в БД присутствует такое же количест-

во хаотично полученных кодов, не содержащих никакой информации. Простейшим случаем таких проблем является поиск в каждой записи заданного ключевого слова (набора слов с условием). Например, создание запроса, который ищет в полях рабочих листов только разрешенные имена.

**Потеря файлов** - это частный случай потери данных (далее будет рассматриваться только случай потери файлов, не несущих значимой информации). Для Системы «ИРБИС32» критическим является только один файл с расширением MST (master-file), имя которого совпадает с именем конкретной базы данных. При потерях других файлов Система «ИРБИС32» полностью сохранит записи и находящуюся в них информацию. Однако для восстановления работоспособности потребуются произвести манипуляции, связанные с восстановлением файлов перекрестных ссылок (расширение XRF) и созданием файлов словарей. Наиболее частой причиной потери данных файла перекрестных ссылок (XRF) являются случаи переполнения емкости жесткого диска или попытки сохранения записи, превосходящей физически допустимый объем, который разрешен для Системы «ИРБИС32» (32 Кбайта). Восстановление XRF-файла может быть проведено при помощи утилиты MSTRCV (Master Receive), поставляемой UNESCO, которая позволяет восстановить XRF-файл по структуре MST-файла. После этого для восстановления работоспособности Системы «ИРБИС32» требуется заново создать словари.

В ряде случаев возможны нарушения работоспособности Системы «ИРБИС32», которые не поддаются какой-либо систематизации и анализу. Авторами используется следующая **последовательность действий в АРМ «Администратор» для восстановления работоспособности:**

1. Копирование файлов базы данных Системы «ИРБИС32» с целью повторных попыток восстановления.
2. Проверка записей в БД («Диагностика файла документов»).
3. Копирование файла документов (для задач дополнительной проверки структуры данных).
4. Экспорт всех записей базы данных в текстовый формат без использования переформатирования. Использование текстового формата передачи позволяет сохранить поля с технологическими метками более 999, которые все чаще используются работниками отделов автоматизации для настройки Системы «ИРБИС32».

5. Опустошение БД.

6. Импорт ранее экспортированных записей из текстового формата. При этом следует помнить, что вследствие исчезновения логически и физически удаленных записей в базе данных произойдет изменение номеров очередности записей, т. е. база данных «сожмется».

7. Создание заново словарей базы данных.

Осуществление последовательности действий по восстановлению работоспособности Системы «ИРБИС32» занимает около 1,5-2 часов для базы данных объемом 100-120 тысяч записей при работе непосредственно на сервере (Pentium-IV 2GHz), но при работе на удаленной машине со скоростью сети 100 Мбит на это уйдет уже не менее 6 часов. Данная последовательность действий может входить в плановое обслуживание Системы «ИРБИС32» и выполняться не реже 1 раза в 2 месяца.

Следует обратить внимание, что вышеперечисленные операции по восстановлению данных оказались успешными во всех случаях потерь данных, кроме одного, когда потребовалась смена отказавших магнитных головок жесткого диска. Стоимость подобной операции, не дающей стопроцентной гарантии восстановления потерянной информации, составляет от 6 до 25 тысяч рублей (без стоимости заменяемых деталей), что существенно выше, чем затраты на ежедневное резервирование данных.

**Одной из важнейших проблем использования информационных систем является наличие разного рода ошибок в данных.** Теория разработки и поддержки ПО уделяет недостаточное внимание вопросам исправления и корректировки ошибок в БД и массивах информационных систем. Не секрет, что просчитать все ошибки при вводе их в систему и придумать алгоритмы проверки и коррекции невозможно. При интенсивной работе с АБИС возникает множество ошибочных данных разного уровня, которые могут повлиять не только на выполнение конкретной функции, но на работу всей информационной системы.

В общем случае ошибки работы АБИС включают:

**1. Модульные ошибки.** Встречаются при неправильном согласовании модулей АБИС. Оказывают влияние на взаимодействие отдельных блоков, а также на взаимодействие с другими системами при обмене данными. Например, неправильное написание программ переноса записей из одной подсистемы

мы в другую или ошибки конвертирования. В этом случае возникают ошибки систематического характера, которые проявляются по ряду признаков. Такие ошибки присутствуют во всех или во многих записях системы.

**2. Функциональные ошибки.** Появляются при неправильном описании или понимании технологических этапов, цепочек. Оказывают влияние на работоспособность системы, особенно в логическом плане при анализе данных. Например, при неправильном понимании технологического процесса возникают ошибки заполнения полей записей и настройки алгоритмов работы. Такие ошибки также носят систематический характер и присутствуют во всех или многих записях.

**3. Дополнительные ошибки.** Возникают при автоматическом генерировании данных информационной системы с ошибочными начальными данными. Оказывают влияние на все документы, «родителями» которых являлись ошибочные документы. Например, неправильная генерация аналитических описаний при ошибочном описании номера периодического издания.

**4. Ошибки, приводящие к работе системы в экстремальных ситуациях.** Данный тип ошибок возникает при появлении в системе записей и данных, близких или превосходящих физические ограничения или целостные ограничения данных. Например, физические ограничения на объем записи 99 Кбайт и 999 (количество) меток полей при передаче библиографических описаний документов в формате ISO-2709, на котором основаны все форматы семейства MARC.

**5. Ошибки, возникающие при дополнительной нагрузке на систему.** Автоматизированная система может изменять данные в соответствии со своим состоянием. Возможны случаи появления ошибочных данных при физических нарушениях аппаратных и программных средств, а также при работе всего программно-аппаратного комплекса на предельной мощности.

**6. Ошибки, возникающие при изменении производительности.** Такие ошибки появляются при неравномерной или импульсной нагрузке и определяются человеческим фактором, а именно тем, что людям свойственно забывать усвоенные знания и навыки, находясь некоторое время «без дела».

Поиск ошибок в любой системе следует начинать с анализа их возможного появления. Основные варианты поиска и обнаружения ошибок определяются как:

### 1. Поиск ошибок по запросу:

- автоматический запрос - использование типовых запросов на частые ошибки;
- случайные запросы - использование изменяемых запросов для поиска нестандартных ошибок.

**2. Использование обратной связи** для нахождения ошибок. Наиболее частый случай - исправление ошибок на основании отзывов пользователей системы.

После нахождения и локализации ошибки ее следует исправить, а также найти и «наказать» виновных. Сам процесс исправления ошибок может вестись разными способами:

**1. Автоматическое исправление** часто встречающихся ошибок. Ошибки в данном случае исправляет сама АБИС или внешнее ПО.

**2. Ручное исправление** ошибок. По мере нахождения ошибок о них сообщается оператору или библиотекарю, который их исправляет.

**3. Полуавтоматическое исправление** ошибок. При нахождении некоторого количества одинаковых ошибок в системе имеет возможность исправить их пакетным способом.

**4. Исправление ошибок внешними пользователями** системы. В этом случае может быть организован доступ пользователей для организации исправления ошибочных данных.

Программное обеспечение для анализа и коррекции данных ЭК библиотек - IsisUtil разрабатывается в Муниципальной информационной библиотечной системе г. Томска с 2001 г. в рамках проекта Relication. Основные характеристики версии 5.20, которая находится в свободном доступе и используется работниками многих библиотек, ранее описаны в работах авторов этой статьи."

В настоящее время разработана новая версия программы IsisUtil, включающая в себя блоки для поиска и анализа данных ЭК и БД:

1. Анализ дублетности записей на основе построения изменяемого по запросу алгоритмического кода (ключа, идентификатора записи).

2. Поиск истории и возможность отката записи в предыдущие состояния, в том числе возможность группового отката результатов последней ошибочной (неудачной) пакетной (глобальной) корректировки записей ЭК.

3. Редактирование записей без использования шаблонов и алгоритмов проверки данных, предусмотренной АБИС.

4. Работа с множествами записей (использование дополнительных возможностей сортировки, операции с результатами запросов, возвращаемыми значениями, сортировка и выгрузка результирующих множеств в нужной последовательности для использования в других АБИС).

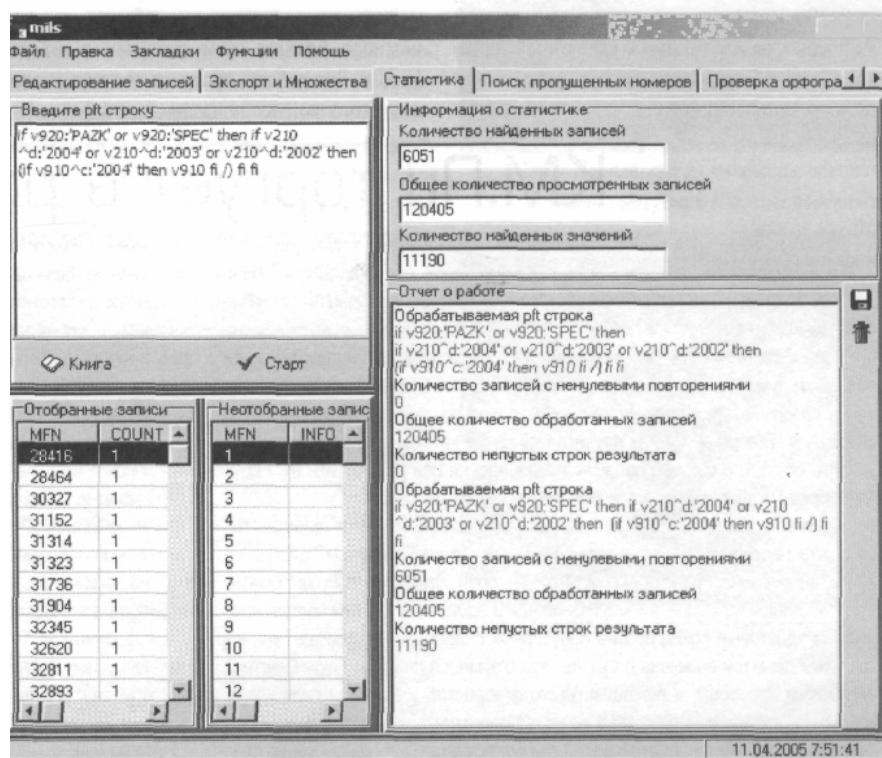


Рис. 1. Вид интерфейса ПО IsisUtil при статистических подсчетах значений полей в ЭК

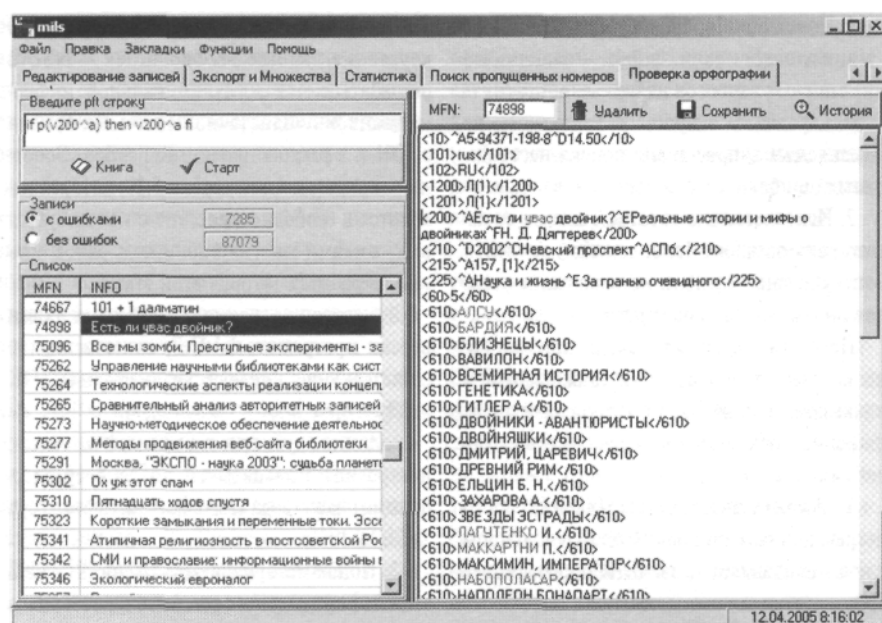


Рис. 2. Вид интерфейса программы IsisUtil при работе по пакетной орфографии баз данных и электронных каталогов

5. Статистический анализ данных с учетом особенностей структуры данных ISO-2709 и наличия повторяющихся и подполей.

6. Анализ последовательностей данных, имеющих в записях.

7. Анализ орфографии в заданном наборе полей и подполей, а также в результатах запроса к ЭК.

8. Восстановление записей базы данных при наличии физических повреждений на жестком диске компьютера.

9. Возможность загрузки обновлений запросов к базам данных с удаленных серверов, а также сортировка и управление количеством запросов.

Относительно ранее описанных версий программа была подвергнута серьезным изменениям, в нее добавлено множество полезных функций. В качестве примера работы ПО IsisUtil со статистикой можно остановиться на значении обновляемости фонда книгами последних трех лет издания, поступивших в библиотеку в прошлом году. Для этого нужно провести подсчет всех повторений полей экземпляров при заданных условиях. Конечно, средствами АБИС такой запрос выполнить можно, но для этого потребуется времени намного больше, чем при использовании IsisUtil. Вид работы программы при выполнении этого запроса показан на рис. 1.

Еще одна функция, на которую стоит обратить внимание, - пакетная проверка орфографии, реализованная в виде анализа и обработки разделительных тегов подполей возвращаемых значений запросов и проверки полученных строк на правильность написания для каждой записи базы данных. Пример работы программы показан на рис. 2.

Ознакомительная версия ПО IsisUtil доступна для скачивания с файлового сервера Муниципальной информационной библиотечной системы города Томска по адресу: [ftp://library.tomsk.ra/pub/isis/isis\\_tomsk/isisutil/](ftp://library.tomsk.ra/pub/isis/isis_tomsk/isisutil/). Данная версия имеет ограничение на 500 анализируемых записей, других функциональных ограничений нет.

Учебники учат, что ошибки и проблемы нужно предотвращать на этапе их возможного возникновения. Но никто не пытается описывать модели поведения и исправления сделанных ошибок. В данной статье авторы не претендуют на описание модели исправления ошибок. Однако разработанные новые возможности программного обеспечения по поиску, локализации и исправлению частых и случайных ошибок каталогизации в конечном счете позволяют улучшить качество электронного каталога и баз данных, от состояния которых зависит работа библиотеки.

Карауш А.С., Копытков Д. Ю.  
Программное обеспечение корректора электронного каталога системы «Ирбис» // Науч. и техн. б-ки. - 2003. - №10. - С. 83-87

Карауш А.С., Копытков Д. Ю. IsisUtil – программа для проверки и коррекции Электронных каталогов библиотек.- М.: ВНИИЦ, 2003.- №50200300830

Карауш А.С., Копытков Д. Ю., Кравчук С. С. Проект Relication: состояние и перспективы // Науч. и техн. б-ки. - 2005. - №1. - С. 76-82

Библиотечное дело. - 2005.- №6(30).- С.21-24