

Инструменты поиска и исправления ошибок в электронных каталогах библиотек

Карауш А. С., Копытков Д. Ю., Макаревич А. С.

Муниципальная информационная библиотечная система, Томск, Россия

Рассмотрены инструменты поиска ошибок в электронных каталогах и базах данных автоматизированных библиотечно-информационных систем. Приведены примеры работы по исправлению. Описано программное обеспечение IsisUtil, предназначенное для анализа, локализации, поиска и коррекции баз данных и электронных каталогов. Описаны новые возможности IsisUtil по пакетному анализу данных.

Теория разработки и поддержки программного обеспечения недостаточное внимание уделяет вопросам исправления и корректировке ошибочных значений в базах данных и массивах информационных систем. Не секрет, что просчитать все ошибки при вводе их в систему и придумать алгоритмы проверки и коррекции данных невозможно. При интенсивной работе с информационной системой возникает множество ошибочных данных разного уровня, которые могут повлиять не только на конкретную транзакцию, но и на всю информационную систему.

В общем случае ошибки работы автоматизированной системы можно определить:

1. **Модульные ошибки.** Встречаются при неправильном согласовании модулей автоматизированных систем. Оказывают влияние на взаимодействие отдельных блоков, а также на взаимодействие с другими аналогичными системами при обмене данными. Например, неправильное написание программы переноса записей из одной подсистемы 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.
3. **Дополнительные ошибки.** Представляют собой ошибки, возникающие 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 определены «человеческим фактором», а именно тем, что люди забывают знания и навыки при нахождении некоторое время «без дела».

Поиск ошибок в любой системе следует начинать с возможного анализа их появления. Основные варианты поиска и обнаружения ошибок можно определить следующим образом:

1. Поиск ошибок по запросу библиотекарей или «автоматизаторов»:
 - a. Автоматический запрос – использование типовых запросов на частые ошибки;
 - b. Случайные запросы – использование изменяемых запросов для поиска нестандартных ошибок;
2. Использование обратной связи для нахождения ошибок. Наиболее частый случай – исправление ошибок согласно отзывам 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 программное обеспечение для автоматизации или внешнее программное обеспечение.
2. Ручное исправление ошибок. По мере нахождения ошибок о них сообщается оператору или библиотекарю, который их исправляет.
3. Полуавтоматическое исправление ошибок. В данном случае при нахождении некоторого количества одинаковых ошибок в системе имеется 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 электронных каталогов библиотек – IsisUtil разрабатывается в Муниципальной информационной библиотечной системе города Томска с 2001 года в рамках проекта Relication. Основные характеристики версии 5.20 описаны в [1-3], которая находится в свободном доступе и используется работниками многих библиотек. В настоящее время разработана новая версия программы IsisUtil, включающая в себя основные блоки для поиска и анализа данных электронных каталогов и баз данных:

1. Анализ на дублетность данных и 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 электронного каталога.
3. Редактирование записей в «прямом доступе» к базе данных без использования шаблонов и алгоритмов проверки данных.
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 системах.
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 базе данных.
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.

Программа, относительно ранее описанных версий [1-3] была подвергнута серьезным изменениям, в нее добавлено множество полезных функций по работе с электронными каталогами и базами данных, не описанных в рамках данной публикации.

В современных условиях развития АБИС и методов анализа данных электронных каталогов в них не всегда имеется возможность получить значения для отчетов. Например – обновляемость фонда. Запрос определит количество экземпляров книг, последних трех лет издания, поступивших в библиотеку в последнем году, и для базы данных CDS/ISIS (ИРБИС32) будет иметь вид:

```
if v920:'PAZK' or v920:'SPEC' then if v210^d:'2004' or v210^d:'2003' or v210^d:'2002' then (if v910^c:'2004' then v910 fi /) fi fi
```

Потребуется возможность подсчета всех повторений поля 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 рисунке 1.

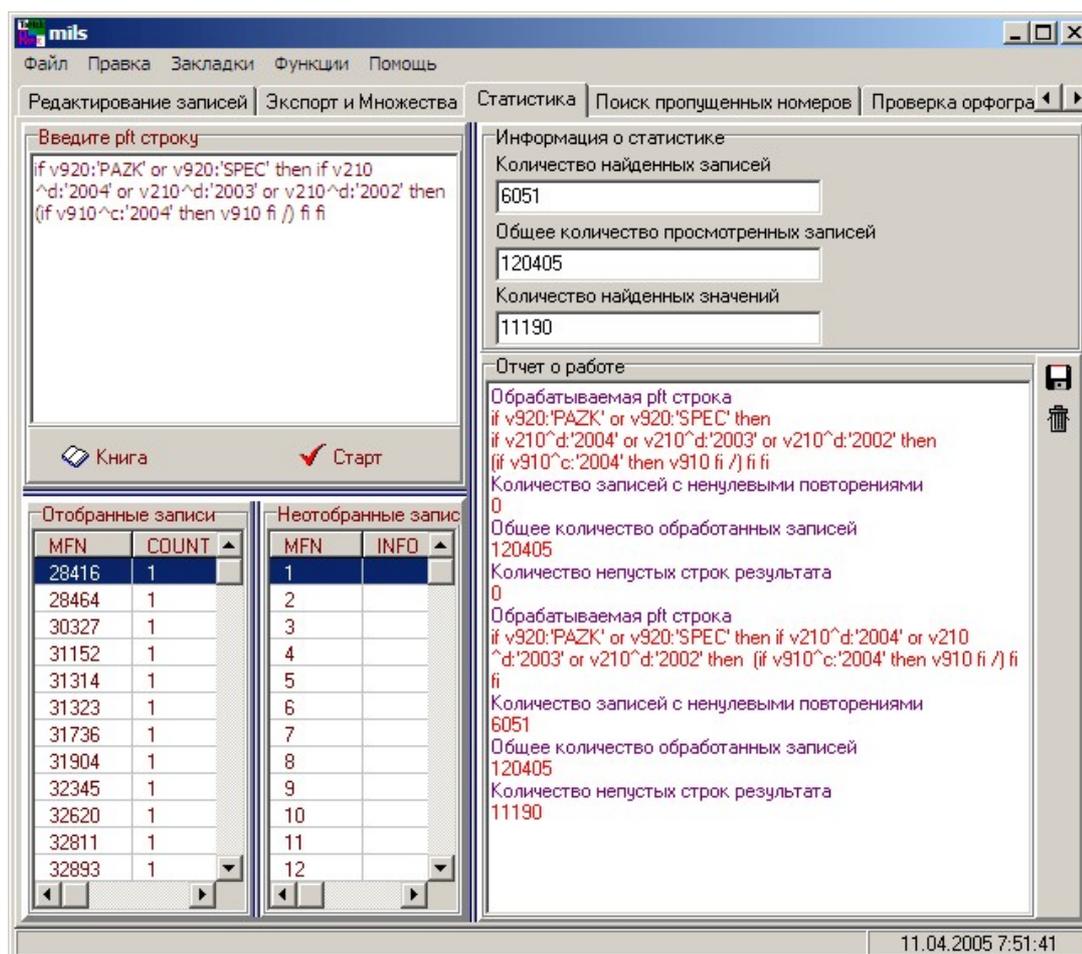


Рисунок 1. Вид интерфейса программы IsisUtil при статистических подсчетах значений полей в электронных каталогах библиотек.

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

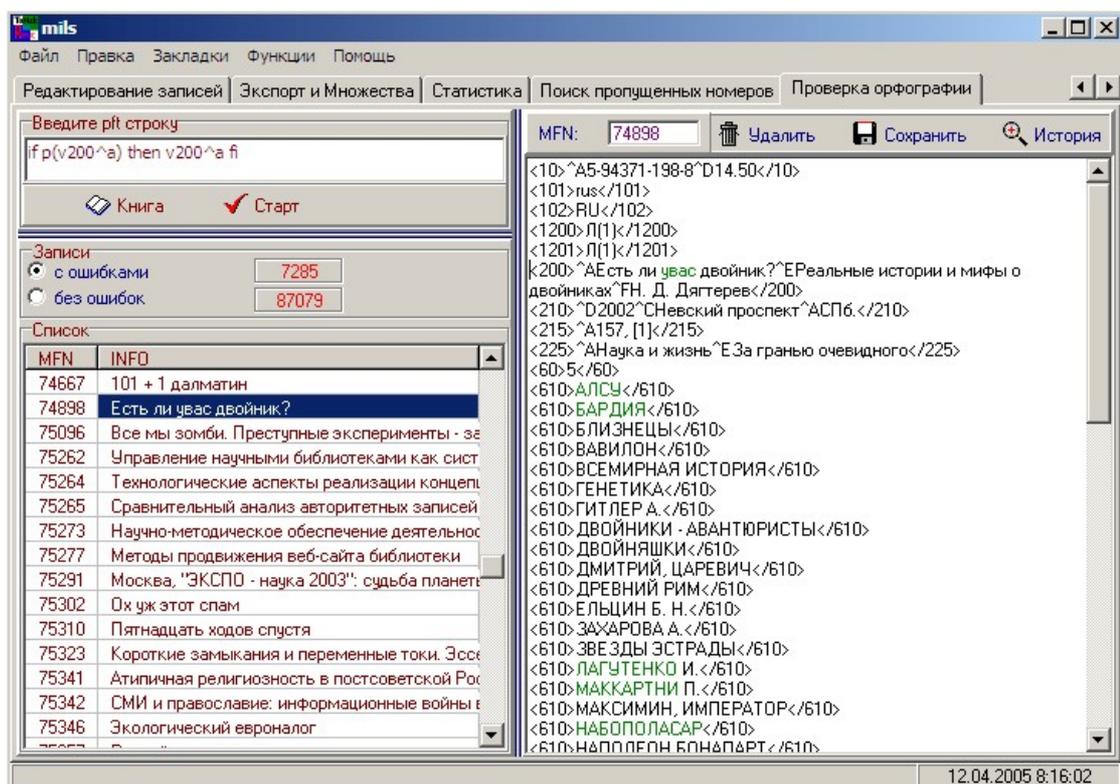


Рисунок 2. Вид интерфейса программы IsisUtil при работе по пакетной проверке орфографии баз данных и электронных каталогов.

Ознакомительная версия программы IsisUtil будет доступна для скачивания с файлового сервера Муниципальной информационной библиотечной системы г.Томска (ftp://library.tomsk.ru/pub/isis/isis_tomsk/isisutil/). Данная версия имеет ограничение на 500 анализируемых записей, других ограничений нет.

Заключение

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

Литература:

1. **Карауш А.С.** Программное обеспечение корректора электронного каталога системы ИРБИС / А.С. Карауш, Д.Ю. Копытков // Науч. и техн. б-ки - 2003. - № 10. - С. 83-87.
2. **Карауш А.С.** "IsisUtil" - программа для проверки и коррекции Электронных каталогов библиотек / А.С. Карауш, Д.Ю. Копытков. - М.: ВНИИЦ, 2003. - № 50200300830.
3. Проект Relication: состояние и перспективы / А.С. Карауш, Д.Ю. Копытков, С.С. Кравчук // Науч. и техн. б-ки - 2005. - № 1. - С. 76-82.