

ТЕХНОЛОГИЯ ПРОИЗВОДСТВА. ПРИБОРЫ И УСТРОЙСТВА

УДК 004.9

Ю.А. Радионова

ХРАНЕНИЕ ИСТОРИИ ИЗМЕНЕНИЯ ДАННЫХ В АРХИВАХ ТЕХНИЧЕСКОЙ ДОКУМЕНТАЦИИ

Радионова Юлия Александровна, окончила механико-математический факультет Ульяновского государственного университета, аспирант кафедры «Информационные системы» Ульяновского государственного технического университета. Ведущий инженер-программист ФНПЦ ОАО «НПО «Марс». Сфера научных интересов – автоматизированные системы документооборота, организация хранилищ технической документации. [e-mail: julia-owl@mail.ru].

Аннотация

В предлагаемой работе речь идет о хранении истории изменений, основанной на информационно-поисковых характеристиках документов. Материал основан на структуре архива электронной документации и структуре электронной картотеки бумажной документации, реализованных в отделе технической документации ФНПЦ ОАО «НПО «Марс». Рассмотрены три возможных способа хранения истории изменений с описанием структуры таблиц базы данных для каждого.

Ключевые слова: архив, техническая документация, история изменений, система управления базами данных (СУБД), электронная картотека.

Abstract

The present article deals with storage of change history on basis of information retrieval features of documents. The article is based on the structure of an archive of e-documents and the structure of an e-index of paper documents, which are implemented in the Department of Technical Documents of Federal Research-and-Production Center Open Joint-Stock Company 'Research-and-Production Association 'Mars'. The article also considers three possible ways for storage of change history including description of structure of database tables for each of them.

Key words: archive, technical documents, change history, DBMS, e-index.

ВВЕДЕНИЕ

В архиве технической документации научно-производственного объединения ведется учет конструкторской, программной, организационно-нормативной документации. Для организации учета документов разработан набор реквизитов. В архивохранилище функционирует информационно-поисковая система, организованная в виде электронной картотеки.

Все изменения, проводимые с документацией, фиксируются в виде отдельных записей в картотеке — для более быстрого поиска по архиву и упрощения составления информационных запросов.

Одной из проблем архива технической документации является необходимость резервного дублирования информации на отчуждаемых носителях. Если по требованиям ГОСТ электронные документы должны содержаться в архиве в

двух экземплярах (подлинник и контрольная копия, причем эти экземпляры в худшем случае должны храниться в разных помещениях одного здания, а в лучшем должны быть разнесены в разные здания), то того же самого необходимо требовать и для информации электронных картотек.

Первый вопрос, возникающий при постановке подобной задачи, — насколько целесообразно сохранять информацию электронной картотеки полностью и в том же формате (например, формат MS SQL Server). Или, наоборот, хранить информацию в наиболее простом формате, гарантирующем, что в любом месте и в любое время найдется ПЭВМ с конфигурацией и программным обеспечением, с помощью которого информация может быть прочитана и обработана.

Также необходимо предусмотреть возможность восстановления и замены информации в функционирующей картотеке за определенный период времени и с гарантией корректности проводимой замены.

Особенности ведения истории изменений для архива технической документации:

- недопустимость удаления данных, бывших когда-либо актуальными;
- в качестве первоначальных данных о документе можно идентифицировать запись, содержащую информацию о наименьшем номере изменения (корректировки) документа;
- актуальными данными считается запись с максимальным номером изменения;
- изменение записи может затрагивать любое количество полей (реквизитов) записи;
- нет необходимости связывать актуальную запись с замененными с помощью дополнительных иерархий, так как любая запись из данной цепочки корректировок будет содержать последовательно изменяемое значение некоторого ключевого поля. То есть, для любой записи (если это не первая) мы можем точно указать предыдущую. Так же для любой записи (если она не является последней) мы можем указать последующую;
- у документа всегда есть неизменяемый реквизит, идентифицирующий документ — десятичный номер.

В качестве исходных данных имеем электронные картотеки нескольких архивов отдела технической документации ФНПЦ ОАО «НПО «Марс», различающихся структурой и правилами учета хранящейся документации. В соответствии с видами архивов различаются структуры (реквизиты) и принципы функционирования картотек.

Поставлена задача:

- определить оптимальный метод хранения истории корректировок документов;
- разработать метод сохранения истории изменений на отчуждаемый носитель;
- определить условия, которым должна удовлетворять сохраняемая информация (что имен-

но надо сохранить);

- определить формат хранения данных;
- проанализировать, насколько данный метод подходит для проверки корректности хранимых данных и их восстановления в случае повреждения.

Структура данных электронной картотеки архива электронной документации

В картотеку заносится реквизитная информация электронного документа, которую можно разбить на следующие группы.

1. Учетная:

- десятичный номер;
- инвентарный номер;
- ФИО ответственного архивариуса;
- первичное применение;
- наличие литеры;
- изделие, в схеме которого закреплен документ.

2. Авторская:

- наименование;
- структурное подразделение, ответственное за разработку документа;
- автор, ответственный за оформление документа.

3. Коррекционная:

- номер извещения;
- номер изменения (версии);
- дата регистрации текущей версии в архиве;
- количество и форматы листов документа (в случае наличия визуального представления документа).

4. Контрольно-файловая:

- идентификаторы файлов документа;
- объем файлов документа;
- контрольные суммы файлов;
- ссылки на местоположение файлов подлинника и контрольной копии в системе архивохранилища.

Прочие реквизиты, такие как примечание, ссылки на сопутствующие документы и другие в рамках решения поставленной задачи можно считать несущественными.

Принцип организации картотеки архива можно сформулировать как «каждое изменение отдельной строкой». Это означает, что при постановке на учет новой версии документа в картотеке заводится новая строка, заполняемая реквизитами новой версии, а строка предыдущей версии получает статус аннулированной, но сохраняется в картотеке.

Основным минусом подобного способа хранения информации является многократное дублирование данных, так как, если реквизит не меняется от версии к версии, он все равно записывается в картотеку столько раз, сколько версий документа было поставлено на учет.

Однако использование для хранения данных одной таблицы существенно упрощает формулировки и ускоряет выполнение запросов.

СТРУКТУРА ДАННЫХ ЭЛЕКТРОННОЙ КАРТОТЕКИ АРХИВА БУМАЖНОЙ ДОКУМЕНТАЦИИ

Картотека архива бумажной документации содержит реквизиты, подобные картотеке архива электронной документации (за исключением контрольно-файловых), но имеет принципиально отличающуюся структуру. Определяющая документ группа реквизитов содержится в главной таблице картотеки (podlinnik_db):

- десятичный номер;
- первоначальный инвентарный номер;
- наименование;
- автор;
- количество и форматы листов документа.

Две подчиненные таблицы (izmenenia_db, primen_db) содержат информацию о проведенных извещениях и изменениях первичной применимости документа.

В таблице извещений при каждой корректировке документа сохраняются номер извещения, номер изменения, новый инвентарный номер (если документ заменяется полностью), данные о присвоенной литературе.

В таблице применимости запись добавляется в том случае, если изменяется первичная применимость документа. В новую строку заносится информация о новой применимости.

При подобной организации картотеки практически полностью исключается дублирование информации, а также появляется возможность получить информацию о наличии документа без привлечения подчиненных таблиц изменений. Однако появляются свои минусы.

Во-первых, хранение информации в нескольких разных таблицах усложняет формулировки запросов с выборкой полных данных о корректировках документов и увеличивает время их выполнения.

Во-вторых, в каждой из подчиненных таблиц приходится хранить поле, связывающее ее с главной таблицей.

В-третьих, — и это можно считать самым серьезным недостатком данного способа — не достигается универсальность хранения измененных данных. Полностью сохраняются данные лишь об извещениях, инвентарных номерах и первичной применимости, а в случае корректировки данных основной таблицы (например, наименования или автора) предыдущие значения реквизитов будут утрачены.

СТРУКТУРА УНИВЕРСАЛЬНОГО СПОСОБА ХРАНЕНИЯ ИСТОРИИ ИЗМЕНЕНИЙ

Для более универсального хранения данных корректировок с минимальным дублированием информации можно предложить структуру картотеки, состоящую из двух таблиц: таблицы для хранения полного набора реквизитов актуальной версии документа и таблицы для хранения истории изменения, имеющей следующую структуру:

- идентификатор документа (по данному полю идет связь с основной таблицей);
- дата проведения изменения;
- сотрудник архива, ответственный за проведение корректировки;
- наименование измененного реквизита;
- замененное значение реквизита.

Для хранения замененного значения реквизита потребуется несколько столбцов — по числу типов данных, используемых в основной таблице. Так как в качестве данных используются только строковый и целочисленный типы (дата уже заложена в структуре таблицы хранения истории изменений), то подобная организация хранения истории получается достаточно выгодной с точки зрения объема базы данных. Для формирования запросов используются две таблицы (в отличие от трех таблиц картотеки второго типа).

Наиболее существенным недостатком подобной организации картотеки можно назвать необходимость формирования полных данных о какой-либо корректировке по нескольким строкам таблицы хранения истории изменений — чем больше реквизитов документа было затронуто текущей корректировкой, тем больше записей будет занесено в таблицу.

Как вариант подобного способа хранения данных можно предложить организацию хранения истории в виде нескольких таблиц, каждая из которых содержит историю изменений целой группы реквизитов. В данном случае реквизиты картотеки разбиваются на группы так, чтобы к каждой группе принадлежали реквизиты, изменение которых с большой степенью вероятности происходит одновременно.

ВОЗМОЖНЫЕ СПОСОБЫ ХРАНЕНИЯ КОПИИ ИНФОРМАЦИИ ЭЛЕКТРОННОЙ КАРТОТЕКИ

В рамках данного пункта необходимо провести выбор формата хранения данных копии и метода (критерия) выбора сохраняемой информации.

Контрольную копию информации электронной картотеки можно сделать несколькими способами. Если рассматривать формат данных, то можно предложить три варианта:

- хранить информацию в формате той СУБД, в среде которой функционирует картотека;
- хранить информацию в файлах какого-либо другого формата, который можно прочитать и обработать на ПЭВМ общепринятой конфигурации;
- хранить информацию в виде твердой копии, то есть в виде бумажных распечаток.

Что касается полноты сохраняемой информации, то наиболее подходящими являются следующие три варианта:

- сохранять информацию картотеки каждый раз в полном объеме и хранить каждую копию отдельно;

- сохранять только последнюю полную копию базы данных;
- сохранять определенные выборки информации, удовлетворяющие заранее заданному критерию (например, за определенный промежуток времени).

Рассмотрим формат сохранения данных. Главным недостатком хранения данных в формате той СУБД, в среде которой функционирует информационно-поисковая система архивохранилища, является необходимость наличия администратора базы данных, имеющего достаточно навыков для работы с СУБД, а также для восстановления информации какой-либо копии базы данных.

Кроме этого, для извлечения информации обязательно наличие ПЭВМ определенной конфигурации с составом программного обеспечения, поддерживающего формат СУБД.

Более универсальным способом представляется хранение данных либо в файлах общепринятого формата, либо в виде твердых копий.

Бумажный вариант может показаться менее приемлемым в среде электронных картотек, однако есть документы (например, инвентарные книги), для которых необходима твердая копия, зарегистрированная в структурном подразделении предприятия, ответственном за учет подобных документов. Наличие только электронных копий требует функционирования в информационно-поисковой системе архивохранилища системы электронно-цифровой подписи.

Таким образом, наиболее целесообразным представляется реализация возможности сохранения данных в формате текстового файла с дополнительной функцией распечатки.

Самый простой способ сохранения контрольной копии электронной картотеки архива — сохранять с определенной периодичностью резервную копию базы данных картотеки. Можно сохранять лишь последнюю актуальную версию или сохранять каждую версию (копию) отдельно.

Наиболее полное сохранение не только актуальных, но и замененных данных достигается путем хранения каждой копии базы данных, однако в этом случае хранится слишком много повторяющихся данных и выбрать информацию, бывшую актуальной в какой-либо определенный момент времени, будет достаточно сложно.

Следует также рассматривать возможность повреждения данных в какой-либо произвольный момент времени. В случае сохранения полной копии базы данных выявить этот момент не менее сложно — придется разработать алгоритм сравнения каждой копии базы данных с предыдущими копиями на выполнение критериев корректности информации.

Если выбран вариант сохранения только последней полной копии базы данных, то выявить момент повреждения информации будет практически невозможно.

Из вышеизложенного можно сделать вывод, что наиболее целесообразным в принятых условиях является сохранение определенного набора данных, удовлетворяющего заранее заданному критерию.

ВЫБОР СТРУКТУРЫ СОХРАНЯЕМОЙ ИНФОРМАЦИИ

В рамках данного пункта необходимо определить набор сохраняемых реквизитов картотеки и критерий для выбора сохраняемых данных.

Для картотеки архива электронных документов (первый вариант) сохраняем полный набор реквизитов, то есть всю строку основной таблицы картотеки. Для определения критерия выбора данных необходимо учитывать дату приема и дату аннулирования версии документа.

```
select * from MainTable
where ((data_receive >= data_begin)
and (data_receive <= data_end)) or ((data_annul >= data_begin)
and (data_annul <= data_end))
```

Полученную выборку можно сохранить в текстовый файл или файл формата MS Word, а также распечатать.

Необходимо проанализировать полноту информации и отсутствие дублирующей информации в полученной выборке. Так как для любой записи в таблице реквизит «дата приема» заполнен, то каждая запись гарантированно попадет в выборку того или иного периода при условии постоянного периодического сохранения истории изменений. Что касается наличия дублирующей информации, то в каждой отдельной записи присутствует некоторое количество реквизитов, оставшихся неизменными, то есть одно и то же значение реквизита будет присутствовать в выборках нескольких периодов. Количество повторений одного и того же значения реквизита можно оценить следующим образом:

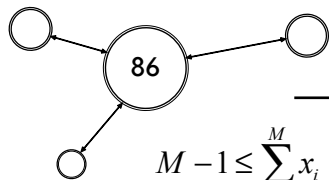
пусть N — количество реквизитов документа;
 M — количество версий документа, зарегистрированных в картотеке.

Тогда (x_1, \dots, x_m) — вектор, каждый элемент которого x_i представляет собой количество реквизитов версии i , оставшихся неизменными по сравнению с версией $i-1$,

$x_1 = 0, 1 \leq x_i < N$ для $i > 1$. Один реквизит всегда остается неизменным (децимальный номер). Принимая во внимание правила хранения и учета реквизитов в картотеке, можно оценить верхнюю границу x_i :

$$x_i \leq N - 8,$$

так как реквизиты коррекционной и контрольно-файловой групп, а также инвентарный номер меняются всегда. Таким образом, получаем, что количество повторяющихся значений реквизитов для одного документа лежит в пределах:



$$M - 1 \leq \sum_{i=1}^M x_i \leq M(N - 8).$$

Если в качестве примера взять данные картотеки архива машинных носителей ФНПЦ ОАО «НПО «Марс», то получим, что на 54903 строки электронной картотеки приходится 23143 документа, в среднем 1,37 изменений на один документ и общее количество повторов реквизитов — 80231 учетных и 62380 авторских.

Таким образом, получаем в среднем 4,5 повторов реквизитов на одно изменение одного документа.

Особой группой являются записи в картотеке, не заменяемые новой версией, а приобретающие дополнения в тексте некоторых реквизитов (в качестве примера можно привести присвоение литеры документам, не имеющим визуального представления, — в данном случае добавляются данные в реквизиты коррекционной группы текущей записи, остальные реквизиты остаются неизменными). Подобные изменения не попадут в выборку по дате приема или аннулирования, и в результате в истории будет сохранена только самая первая версия.

В картотеке второго типа (архив бумажных документов) сохраняем полностью реквизиты основной таблицы и двух вспомогательных таблиц для тех документов, дата приема или корректировки которых принадлежит выбранному интервалу.

```
// корректировки, прошедшие за указанный период
select * from izmenenia_db
where (data >= data_begin) and (data <= data_end)

// документы, принятые в указанный период
select * from podlinnik_db
where (data >= data_begin) and (data <= data_end)

// документы, у которых в указанный период корректировалась
// первичная применяемость
select * from primen_db
where (data >= data_begin) and (data <= data_end)
```

С принятием подобного критерия выбора сохраняемой информации мы теряем информацию, которая изменяется в основной таблице при корректировках извещений (наименование, автор, форматы листов). Но в данном случае в сохраняемой копии полностью отсутствует дублирование информации.

Возможен еще вариант сохранения с каждым проведенным извещением еще и информации основной таблицы.

```
select * from podlinnik_db, izmenenia_db
where (podlinnik_db.dec_number =
izmenenia_db.dec_number)
```

```
and (izmenenia_db.data >= data_begin)
and (izmenenia_db.data <= data_end)
```

В данном случае полноту сохраняемой истории можно считать 100 %, так как любая корректировка документа сопровождается выпуском извещения. Отсутствует проблема, возникающая в картотеке первого типа в том случае, когда текущая корректировка не сохраняется в картотеке отдельной строкой. Если оценивать количество дублируемых данных, сохраняемых в истории изменений, то получаем, что в каждой строке повторяются значения двух реквизитов, изменения которых достаточно редки (наименование, формат документа), и значения трех реквизитов, которые никогда не изменяются (десимальный номер, первоначальный инвентарный номер, дата приема в архив первой версии документа).

Для картотеки третьего типа сохраняем строки таблицы изменений с датой корректировки, принадлежащей указанному интервалу:

```
select * from izmenenia
where (data >= data_begin) and (data <= data_end)
```

Заменяем данные хранимой основной таблицы на данные основной таблицы картотеки.

```
drop table history_ActualData
select * into history_ActualData from ActualData
```

В данном случае имеем 100 % полноту сохраняемых данных. В качестве недостатка данного способа можно указать необходимость полностью перезаписывать таблицу актуальных данных при каждом сохранении, что может занять достаточно много времени. Также при достаточно большом количестве заменяемых реквизитов для одной корректировки документа имеем множество повторений полей «Архивариус», «Дата изменения», «Идентификатор документа».

АЛГОРИТМ СОХРАНЕНИЯ ИСТОРИИ ИЗМЕНЕНИЙ

Таким образом, для хранения истории изменений данных принимается третий способ как наиболее оптимальный. Принимая во внимание сложившуюся структуру картотек архива, необходимо разработать алгоритм формирования структуры данных для хранимой истории изменений из существующих структур.

Для хранения истории изменений определяем структуру двух таблиц: актуальных данных (ActualData) и истории изменений (HistoryData). В качестве периода сохранения истории изменений выбираем временной интервал [data_begin... data_end]¹.

Таблица актуальных данных для хранения

1 Для гарантированного сохранения полной истории изменений временной интервал можно определять автоматически: начало — дата, следующая за датой последнего сохраненного изменения; окончание — текущая дата.

истории изменений картотеки электронной документации будет иметь структуру, аналогичную структуре таблицы самой картотеки. Принимая условие автоматического определения временного интервала, заполняем таблицу данными, актуальными в настоящий момент времени.

```
drop table ActualData
```

```
select * into ActualData
```

```
from MainTable where data_annul is NULL
```

Чтобы исключить большое количество повторяющейся информации, воспользуемся заранее заданной структурой и свойствами электронной картотеки и составим две таблицы — для хранения истории изменений часто и редко меняющихся реквизитов.

Таким образом, для хранения истории изменений получаем следующие таблицы:

1. HistoryDataOften (часто или постоянно меняющиеся реквизиты):

- десятичный номер (идентификатор документа);
- идентификаторы, размеры, контрольные суммы файлов;
- инвентарный номер;
- номер извещения и изменения (версии);
- местоположение файлов документа в системе архивохранилища;
- архивариус;
- дата проведения изменения;

2. HistoryDataSeldom (редко меняющиеся реквизиты):

- десятичный номер (идентификатор документа);
- название измененного реквизита;
- замененное значение реквизита;
- архивариус;
- дата проведения изменения.

Заполняем таблицы данными за выбранный временной интервал.

```
insert into HistoryDataOften ([list_of_fields]2) values
```

```
(select [list_of_fields] from MainTable where (data_receive>=data_begin) and (data_receive<data_end)
```

```
and (not (data_annul is NULL)))
```

Для таблицы редко изменяющихся реквизитов заполнение проводится для каждого реквизита отдельно по следующему алгоритму:

- выбрать все строки для данного документа из исходной таблицы картотеки в порядке их регистрации;

- для каждой строки сравнить значение реквизита F1 с его значением в следующей строке F2;

- если значение реквизита изменилось, то добавить в таблицу хранения истории изменений значение F1 с датой регистрации значения F2 в

картотеке.

Таким образом, за счет некоторого усложнения структуры хранимых таблиц обеспечивается минимальное дублирование информации. Для выбора полной информации по какой-либо корректировке документа необходимо объединить в запросе три таблицы, что вполне приемлемо по сложности запроса и скорости его обработки. Данный способ хранения истории изменений снимает проблему сохранения актуальных данных для строк таблицы, изменение которых состоит только в дополнении реквизитов текущей строки, без добавления новой строки.

Для хранения истории изменения картотеки бумажного архива задействуем две таблицы: актуальных данных (ActualData) и истории изменений (HistoryData). Должны выполняться следующие правила сохранения данных:

- связь между таблицами осуществляется по уникальному коду документа (так как в таблице основных данных картотеки каждый отдельный документ представлен одной строкой);

- данные последних корректировок таблицы изменений картотеки сохраняются в таблице ActualData как реквизиты текущей версии документа;

- каждая строка таблицы изменений картотеки сохраняется отдельной строкой в HistoryData;

- если при сохранении новых актуальных данных основной таблицы картотеки значение какого-либо реквизита изменилось, то замененное значение сохраняется в HistoryData.

Подобный способ хранения истории изменений практически снимает проблему потери замененных значений реквизитов основной таблицы картотеки, так как при достаточно частом сохранении истории замененные значения фиксируются.

Возможность использования сохраненной истории изменений для коррекции поврежденных данных картотек

Повреждение информации электронных картотек может произойти как по причине операторских ошибок, так и при аппаратных сбоях, приводящих к повреждению структур и содержания таблиц базы данных. Резервное копирование базы данных частично решает подобные проблемы, однако существует вероятность, что при достаточно частом периодичном резервном копировании сохранится копия базы данных с ошибками, которые не были выявлены сразу, но могут повлиять на дальнейшую работу с базой данных.

Решение проблемы обнаружения и коррекции ошибок может быть упрощено с помощью хранимой истории изменений данных. Для этого, наряду с резервным копированием базы данных стандартными средствами СУБД, необходимо

² list_of_fields — полный список полей таблицы HistoryDataOften.

проводить проверку корректности данных, используя сохраненные изменения данных за прошедшие периоды.

Для проверки корректности данных картотеки, измененных в течение интервала времени [D1;D2], необходимо выбрать все записи из таблиц хранения истории изменений, датированные этим периодом, и сформировать из них таблицу, каждая строка которой будет содержать полный набор измененных реквизитов. Если он не является полным, то необходимо выбрать значения недостающих реквизитов из предыдущих периодов истории хранения. Таким образом, будет составлена таблица, каждая строка которой содержит полный набор реквизитов документа. Сравнивая данные этой таблицы с текущей версией таблицы в рабочей базе данных, можно определить, насколько корректны текущие данные за указанный период. В случае утери информации за какой-либо период в рабочей базе данных можно занести сформированные из истории изменений данные взамен утерянных.

Выводы

Таким образом, проведено сравнение трех видов структур для электронных картотек хранения корректировок в архивохранилище технической

документации. Определена структура хранения истории изменений, позволяющая использовать для текущей работы сложившуюся систему картотек без изменений принципов ее организации и условий работы. Определен алгоритм отбора и сохранения данных для последующего формирования отчета и сохранения его на отчуждаемый носитель. Проведен анализ возможности использования сохраненной истории изменений для выявления и коррекции ошибок в рабочей версии базы данных электронных картотек.

СПИСОК ЛИТЕРАТУРЫ

1. Сидельников В. И. Метод представления истории изменения данных / В. И. Сидельников // Информационные технологии. — 2006. — № 7. — С. 37–48.
2. Радионова Ю. А. Проектирование структуры данных архива электронных документов НПО с использованием среды MS SQL Server 2000 / Ю. А. Радионова, В. Г. Тронин // Автоматизация процессов управления. — 2008. — № 2 (12). — С. 119–123.
3. ГОСТ 2.501-88. ЕСКД. Правила учета и хранения. — Введ. 1989-01-01. — М. : Изд-во стандартов, 2003.