

Моделирование автоматического исправления ошибок в технологии информационных систем

Разработана модель автоматического исправления ошибок при обработке документов табличного вида в рамках информационных систем вуза, научных организаций и др. Разработана соответствующая программа ЭВМ. Программа исключают необходимость ручных операций при корректировке ошибок в технологии обработки табличных документов. Результаты компьютерных экспериментов подтвердили адекватность модели. Выполнено измерение и показана технико-экономическая эффективность модели.

Ключевые слова: моделирование, алгоритм, программа, компьютер, коррекция ошибок, матрица документа.

MODELLING OF AUTOMATIC CORRECTION OF MISTAKES IN TECHNOLOGY OF INFORMATION SYSTEMS

The model of automatic correction of mistakes is developed at processing documents of a tabulated kind within the framework of information systems of high school, the scientific organizations, etc. The corresponding program of the computer is developed. The program exclude necessity of manual operations at updating mistakes in technology of processing of tabulated documents. Results of computer experiments have confirmed adequacy of model. Measurement is executed and technical and economic efficiency of model is shown.

Keywords: modelling, algorithm, the program, a computer, correction of mistakes, a matrix of the document.

Введение

В улучшении качества информационных систем (ИС) особое место занимают задачи устранения множества дефектов обработки данных [1]. Такими дефектами могут быть искажение значения цифрового показателя, пропуск значения показателя, транспозиция («наполнение») одного значения на другое и др. Процедуры обнаружения, идентификации класса и адреса дефекта и последующего исправления дефектов требуют значительных временных, трудовых и финансовых ресурсов. Наиболее эффективными средствами коррекции дефектов представляются программные модули, ориентированные на определение модификаций и адресов лексических, синтаксических, логических и арифметических ошибок в данных. Значительный эффект методы программного контроля дают в системах обработки информации

цифрового содержания – учетных, отчетных, статистических, плановых, параметрических, где искажение даже одной цифры, например, в финансовых, научных, военных ИС, может привести в некоторых случаях к серьезным негативным последствиям. Однако эти программные средства не могут выполнять автоматически операции вычисления достоверного значения показателя и последующей замены ошибочного значения показателя на его достоверное значение. Эта трудоемкая часть операций выполняется вручную, иногда с оформлением соответствующего корректировочного бланка. При этом не устраняется вероятность внесения дополнительной ошибки при вводе данных уже самого корректировочного бланка.

Вместе с тем следует отметить, что в соответствии с принципами самоорганизации и адаптивности качество ИС будет улучшаться, если в технологии обработки дан-

ных ИС структура технологических операций будет меняться в сторону расширения автоматических, программных процедур. Для того, чтобы заменить в нашем случае ручные операции на программные необходимо разработать модель автоматического обнаружения и исправления ошибок в технологии ИС, применяемых в задачах вуза, научных учреждений и др. [2].

1. Требования, предъявляемые к модели автоматического исправления ошибок в данных

С целью определения основных требований к методам и средствам повышения уровня достоверности обратимся к технологии обработки данных ИС. Почти каждый этап обработки сопровождается выполнением операций контроля данных, в которых значительный объем при-



Георгий Николаевич Исаев,
к.т.н., доцент кафедры Сервисного
инжиниринга
Тел.: 8 (915) 394-74-29
Эл. почта: georg.isaev@mail.ru,
Российский государственный
университет туризма и сервиса
<http://www.rguts.ru>

George N. Isaev,
cand.tech.sci., the senior lecturer of
faculty of Service engineering
Tel.: 8 (915) 394-74-29
E-mail: georg.isaev@mail.ru
The Russian state university of tourism
and service
<http://www.rguts.ru>

ходится на контроль достоверности и полноты сведений в обрабатываемых документах. Особо тщательно должна проверяться производная документация перед выдачей ее абонентам. Неадекватность (недостоверность и неполнота) сведений в документации влечет соответственно снижение эффективности принимаемых решений. Иногда это обуславливает повторную обработку пакета первичных документов, что увеличивает стоимость обработки информации, снижает уровень своевременности, ухудшает качество результатной информации. Таким образом, эффективность обработки в значительной мере зависит от операций контроля, предшествующих этапу обработки документов в ЭВМ.

Таким образом, алгоритм и программа автоматического обнаружения ошибок и восстановления достоверности значений показателей документов должны удовлетворять следующим требованиям:

- повышение уровня достоверности, полноты и своевременности информации;
- снижение объемов временных, трудовых, материальных и финансовых ресурсов, используемых в технологии обработки данных;
- адаптация к сравнительно широкому классу обрабатываемых форматов табличных документов;
- дружественный интерфейс;
- возможность применения в других технологиях обработки данных;
- реализация максимального состава функций лексического, синтаксического, логического и арифметического контроля при условии сравнительно минимального физического объема программного модуля.

2. Определение модели табличного документа

Моделирование процесса автоматического обнаружения ошибок и восстановления достоверности значений показателей в табличных документах вызывает необходимость анализа структуры и свойств указанных документов. Модель табличного документа можно представить в виде матрицы (рис. 1).

Модель табличного документа можно отобразить четверкой

$$Q = \langle Q^k, A, B, D \rangle$$

где Q^k – матрица документа $Q^k = \|q_{ij}\|$, $i = 1, m, j = 1, n$;

q_{ij} – реквизиты–основания (числа), отражающие количественное состояние объектов;

A – кортеж реквизитов-признаков (наименование строк таблицы), отражающий качественные стороны состояния объектов;

B – кортеж реквизитов-признаков, (наименования столбцов таблицы), отражающий качественные стороны состояния объектов;

D – кортеж реквизитов-признаков, отражающий качественные стороны состояния объектов общего уровня и относящиеся как к A , так и к B ;

$q_{i, n+1}, q_{m+1, j}$ – реквизиты-основания типа «итого», «всего» или контрольные суммы соответственно по строкам и столбцам, отражающие количественное состояние объектов.

Рассматриваемая модель табличного документа в определенном допущении может быть представлена в синтаксическом отношении как кодовый ансамбль. В этом ансамбле информационные группы отображаются совокупностью значений показателей по документо-строкам и (или) документо-графам, а также контрольными суммами и (или) значениями показателей типа «всего» и «итого». В данном случае контрольные суммы обла-

| $D = \langle d_1, d_2, \dots, d_s, \dots, d_k \rangle$ | | | | | | | | |
|--|--------------|--------------|----------|--------------|----------|--------------|----------------|--------------|
| B | b_1 | b_2 | ... | b_j | ... | b_n | b_{n+1} | |
| A | a_1 | q_{11} | q_{12} | ... | q_{1j} | ... | q_{1n} | $q_{1, n+1}$ |
| a_2 | q_{21} | q_{22} | ... | q_{2j} | ... | q_{2n} | $q_{2, n+1}$ | |
| ... | ... | ... | ... | ... | ... | ... | ... | |
| a_i | q_{i1} | q_{i2} | ... | q_{ij} | ... | q_{in} | $q_{i, n+1}$ | |
| ... | ... | ... | ... | ... | ... | ... | ... | |
| a_m | q_{m1} | q_{m2} | ... | q_{mj} | ... | q_{mn} | $q_{m, n+1}$ | |
| a_{m+1} | $q_{m+1, 1}$ | $q_{m+1, 2}$ | ... | $q_{m+1, j}$ | ... | $q_{m+1, n}$ | $q_{m+1, n+1}$ | |

Рис. 1. Матрица документа табличного вида

дают своеобразными свойствами синдромов, то есть опознавателей ошибок.

Указанные связи между значениями элементов матрицы документа обеспечивают потенциальную возможность автоматического обнаружения ошибок и их исправление без ручных операций.

3. Построение алгоритма автоматического исправления ошибок

С учетом выявленных выше свойств модели (рис. 1) построим модель алгоритма автоматического восстановления достоверности показателей документов табличного вида.

Исходя из анализа свойств реквизитов-оснований, наблюдается взаимосвязь элементов типа арифметического баланса

$$q_{i,n+1} = \sum_{j=1}^n q_{ij} \quad (1)$$

$$q_{m+1,j} = \sum_{i=1}^m q_{ij} \quad (2)$$

При условии внесения ошибки в какой-либо элемент q_{ij} на этапах обработки нарушаются условия соотношений типа (1), (2). С целью автоматического обнаружения ошибок и их исправления при вводе в ЭВМ указанные соотношения проверяются программно. Сначала проверяется равенство

$$\sum_{i=1}^m q_{i,n+1} = \sum_{j=1}^n q_{m+1,j} \quad (3)$$

Если равенство не соблюдается, то на принтер или дисплей, в рамках протокола ввода документов в ЭВМ, выдается сообщение об отсутствии равенства указанного типа и идентификатор документа. Если же равенство (3) соблюдается, то проверяется далее условие

$$\sum_{j=1}^n q_{ij} = q_{i,n+1} \quad (4)$$

если в i -й строке равенство не выполняется,

$$\sum_{j=1}^n q_{ij} \neq q_{i,n+1} \quad (5)$$

то производится замена ошибочной строки \bar{i} на строку с элементами

$$\bar{q}_{\bar{i}j} = q_{m+1,j} - \sum_{i \neq \bar{i}} q_{ij} \quad (6)$$

после чего выдается сообщение на принтер об ошибке и ее исправлении с указанием индекса документа, а также значение замененного ошибочного реквизита-основания и заменяющего достоверного реквизита-основания.

Если же нарушение условия (4) более чем в одной строке, то для столбцов матрицы проверяется условие

$$\sum_{i=1}^m q_{ij} = q_{m+1,j} \quad (7)$$

Если же нарушение условия имеет место в одном столбце

$$\sum_{j=1}^m q_{i\bar{j}} \neq q_{m+1,\bar{j}} \quad (8)$$

то заменяется столбец \bar{j} на столбец с элементами

$$\bar{q}_{i\bar{j}} = q_{m+1,\bar{j}} - \sum_{j \neq \bar{j}} q_{ij} \quad (9)$$

и выдается сообщение на принтер об ошибке и её исправлении.

Если же нарушение более чем в одном столбце, то на дисплей или принтер выдается сообщение об ошибках с обозначением модификаций ошибок и их адресов. Заметим, что при условии какой-либо ошибки, например, транспозиции (перестановки) q_{ij} , $q_{i,j+1}$ нарушается условие (1). Это и идентифицируется как ошибка относительно $q_{m+1,j}$ и $q_{m+1,j+1}$. Обнаружение выполняется не только в случаях транспозиции, но и других различных искажений лексического, синтаксического, логического и арифметического свойства по набору q_{ij} каждой отдельной строки и (или) столбца матрицы документа. Таким образом, алгоритм позволяет программное исправление однократных и обнаружение многократных ошибок относительно

но строки и (или) столбца матрицы контролируемого документа. Следует отметить, что вероятность двойных ошибок в документо-графе и (или) в документо-строке мала, что практически все ошибки справляются автоматически.

При практическом применении данного метода следует учитывать два случая. В первом случае в формате первичного документа, например, форма статистического отчета обычно присутствуют реквизиты типа «всего», «итого» как по строкам, так и по столбцам. В этом случае никаких дополнительных трудозатрат не требуется. Во втором случае указанные реквизиты отсутствуют и тогда целесообразно ввести в таблицу документа реквизиты типа «всего» и подсчитать контрольные суммы.

4. Структура и алгоритм программы автоматического восстановления достоверности данных

Одним из принципиальных вопросов является разработка структуры и алгоритма программы контроля данных. Программа разработана средствами языка Assembler. Структура программы представлена на уровне блоков (рис. 2). В блоках с 6 по 10 происходит обработка ошибок по их модификациям, указанным внутри блоков.

Блок сопряжения с системой подготовки данных (СПД) информационной системы осуществляет интерфейс между программой и СПД. В режиме открытия блок сопряжения сохраняет адрес модуля печати СПД, в дальнейшем печать



Рис. 2. Структура программы «Автоматическое обнаружение ошибок и восстановление достоверности показателей документов табличного вида»

диагностической информации осуществляется через СПД. В режиме закрытия выполняется завершение работы программы. В рабочем режиме выполняется анализ описания свойств данных входных документов (ОСД), очистка рабочих полей.

Блок контроля документов (БКД) служит для выработки пятиразрядного кода завершения, в дальнейшем код завершения используется как номер подпрограммы обслуживания, максимальное значение которого может быть 25, то есть 32. Блок контроля документов получает от СПД адрес документа и число строк в этом документе. Графа с номером строки в контроле не участвует и используется при выдаче диагностических сообщений. Остальные графы образуют матрицу, в которой графа «КС» содержит контрольные суммы по каждой строке.

При разработке алгоритма учитывались требования, которые были определены в разделе 3. Контроль матрицы документа выполняется в соответствии с алгоритмом программы (рис. 3). В программе выделяется 16 рабочих ячеек, в которых происходит накопление ошибки, выдается сообщение: «В документе XXXXX УУ обнаружено более одной ошибки».

Подпрограмма печати адресов получает управление каждый раз, когда в документе обнаружено более одной ошибки. При этом каждая строка с ошибкой сопровождается сообщением: «Строка XXXXXX содержит ошибку» и соответственно каждая графа с ошибкой сопровождается сообщением «графа XX содержит ошибку». Подпрограмма СС06 получает управление, когда вычисленная контрольная сумма по строке 01 равна вычисленной контрольной сумме по графе КС и не равна заданной контрольной сумме, то есть оператор допустил ошибку в подсчете контрольной суммы. Подпрограмма заменяет ошибочное число на вычисленное достоверное и выдает сообщение «строка 01, графа КС значение XXXX скорректировано на УУУУ».

Подпрограмма ССII получает управление при условии ошибок в контрольной сумме в графе ХХ. Ошибочная контрольная сумма за-

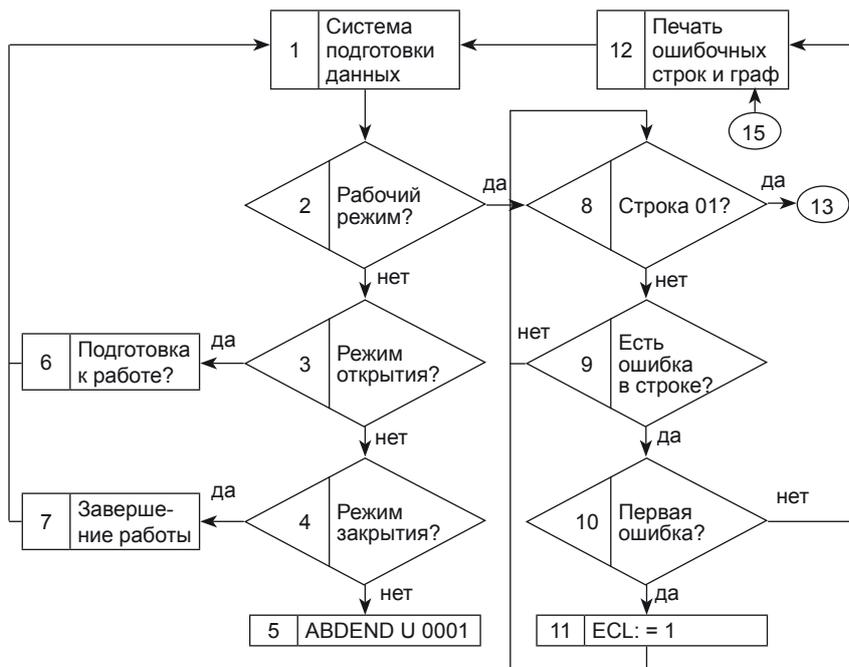


Рис. 3. Блок-схема алгоритма программы «Автоматическое восстановление достоверности значений показателей документов табличного вида» (начало)

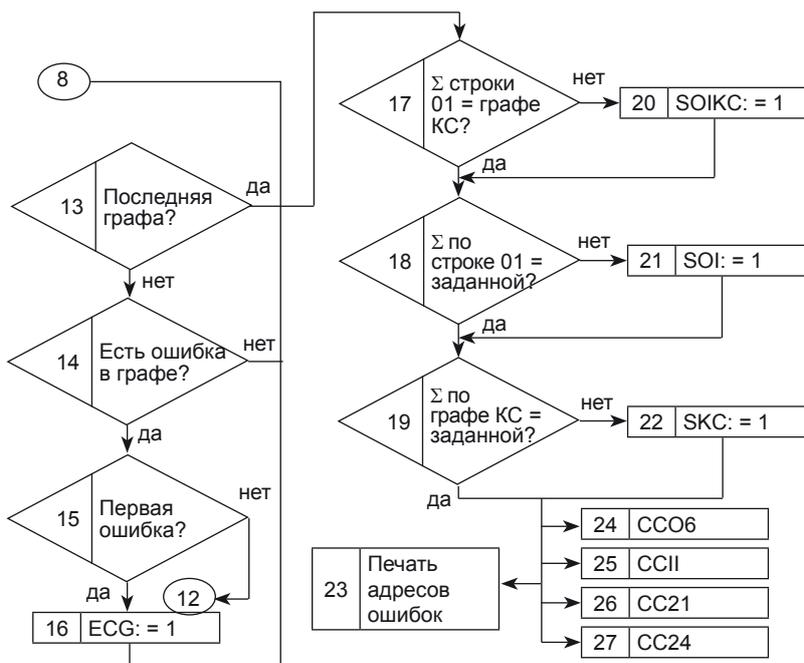


Рис. 3. Блок-схема алгоритма программы «Автоматическое восстановление достоверности значений показателей документов табличного вида» (окончание)

меняется на правильно вычисленную, и выдается сообщение «строка 01, графа ХХ, значение XXXX скорректировано на УУУУ».

Подпрограмма СС2I получает управление, если есть ошибка в контрольной сумме одной из строк. Производится замена ошибочного реквизита-основания на вычисленное достоверное с последующей выдачей сообщения «строка XXXXXX, графа КС, зна-

чение XXXX скорректировано на УУУУ».

Подпрограмма СС24 получает управление при наличии ошибки в одной из строк и в одной из граф. Ошибочное число, расположенное на пересечении ошибочной строки с ошибочной графой заменяется на вычисленное достоверное. При этом выдается сообщение «строка XXXXXX, графа УУ, значение XXXX скорректировано на УУУУ».

Блок переходов осуществляет передачу управления на соответствующую подпрограмму обслуживания, используя код завершения в байте FLAG. В случае, если для какого-либо кода завершения нет подпрограммы обслуживания, управление передается на печать кода завершения (ПКЗ).

Подпрограмма печати кода завершения получает управление, если для кода завершения, выработанного блоком БКД, нет соответствующей подпрограммы обработки. При этом полученный код распечатывается в двоичном виде «Некорректируемая ошибка с кодом XXXXX», где:

IXXXX-ECL – признак ошибки в строке,

XIXXX-ECG – признак ошибки в графе,

XXIXX-SKC – ошибочная контрольная сумма по графе KC,

XXXIX-SOI – ошибочная контрольная сумма по строке OI,

XXXXI-SOIKC – сумма по графе KC не равна сумме по строке OI.

Программа не накладывает ограничений на способы ввода документов в ЭВМ. Документы в ЭВМ могут быть введены через магнитные носители – диски, ленты, каналы передачи данных, сканирующие устройства в зависимости от конкретных условий обработки и характера решаемых задач [3].

5. Экспериментальная проверка адекватности модели

В целях экспериментального исследования, проверки работоспособности и оценки эффективности алгоритма анализу подвергалась программа, реализующая указанный алгоритм. В качестве экспери-

ментального материала привлечено 20 документов, в каждом из которых содержалось 7 таблиц, то есть 140 таблиц. Для проверки программы в таблицы были внесены ситуационные ошибки. Объем и модификации ошибок определялись с учетом необходимости проверки максимального набора вероятных типов ошибок и полного объема функциональных свойств программы автоматического восстановления данных.

Эксперименты показали, что общее время работы программы, как с включением программы в СПД, так и без включения, одинаково и составило 2 минуты на обработку 20 отчетов. Однако время работы процессора ЭВМ по реализации операций входного контроля отчетов не равнозначно относительно указанных вариантов контроля. Так, например, время работы процессора с включением программы в СПД по данным протокола ввода и контроля документов составило 38,02 сек. Процессорное время контроля без применения рассматриваемой программы (только средствами СПД) составило 44,98 сек. Таким образом, применение программы сократило время работы процессора на 6,96 сек., то есть на 15,5 %. Вместе с тем, взаимодействуя с СПД, программа освобождает от необходимости объемной распечатки протокола диагностики ошибок. В любом случае посредством рассматриваемой программы на принтер выдаются более краткие и вместе с тем более информативные сообщения об адресе, исправлении и модификации ошибок. Так, например, распечатка диагностики ошибок по протоколу ввода и контроля 20 отчетов при условии включения рассматриваемой программы в СПД заняла 8 листов

бумаги формата 207x210 мм. Распечатка диагностики указанных отчетов без включения программы, то есть только средствами СПД, заняла 14 листов бумаги указанного формата. Таким образом, применение программы уменьшает расход бумаги при выполнении этапа ввода и контроля документации ориентировочно до 40%.

Заключение

1. Одним из способов минимизации дефектов обработки и улучшения качества ИС является метод автоматической коррекции ошибок в значениях цифровых показателей документов табличной структуры ИС различного класса и назначения.

2. На основе модели табличного документа определяются взаимосвязи между реквизитами-основаниями. По результатам определения свойств модели разрабатываются структура и алгоритм реализации программы автоматической коррекции ошибок в данных.

3. Программа выполняет обнаружение ошибки в значении реквизита-основания, определяет его истинное (достоверное) значение, а затем заменяет ошибочное значение на вычисленное достоверное значение реквизита-основания.

4. Программа может определять значения пропущенных в строке (графе) реквизитов-оснований и восстанавливать их значения, исключая при этом необходимость выполнения ручных операций.

5. Проведенные эксперименты показали адекватность модели, функциональную и экономическую эффективность программы автоматической коррекции ошибок в данных.

Литература

1. Дружинин Г.В., Сергеева И.В. Качество информации. – М.: Радио и связь, 1990. – 172 с.
2. Исаев Г.Н. Управление качеством информационных систем: Теоретико-методологические основания: монография. – М.: Наука, 2011. – 279 с.
3. Исаев Г.Н. Устройство для автоматического определения и коррекции ошибок в табличных документах. Патент RU № 52221 U1. Бюллетень Федеральной службы по интеллектуальной собственности, патентам и товарным знакам, 2006, № 7 от 10.03.2006.

References

1. Druzhinin G.V., Sergeeva I.V. Kachestvo informacii. – M.: Radio i svjaz', 1990. – 172 s.
2. Isaev G.N. Upravlenie kachestvom informacionnyh sistem: Teoretiko-metodologicheskie osnovaniya: monografija. – M.: Nauka, 2011. – 279 s.
3. Isaev G.N. Ustrojstvo dlja avtomaticheskogo opredelenija i korrekcii oshibok v tablichnyh dokumentah. Patent RU № 52221 U1. Bjulleten' Federal'noj sluzhby po intellektual'noj sobstvennosti, patentam i tovarnym znakam, 2006, № 7 ot 10.03.2006.