

УДК 004.624

С.Ю. Марулин

ДОКУМЕНТО–ОРИЕНТИРОВАННАЯ АВТОМАТИЗАЦИЯ ПРОЦЕССА ЗАПОЛНЕНИЯ БАЗЫ ДАННЫХ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Аннотация. Рассматриваются различные способы ведения баз данных. Описаны подходы, позволяющие сократить время на перенос содержимого бумажных копий документов в базу данных информационной системы. В основе данных подходов лежит использование в качестве промежуточного звена электронного документа либо его отображения в виде экранной формы. На примерах показана эффективность использования подходов, и как следствие, сокращение времени на достижение поставленной цели.

Ключевые слова: электронный документ, экранная форма, база данных, информационная система.

S.U. Marulin

DOCUMENT–ORIENTED DATABASE AUTOMATION OF THE PROCESS FILLING DATA INFORMATION SYSTEM

Abstract. The article deals with different methods of data base maintenance. Approaches allowing to reduce time of transferring the content of hard copy documents into the data base of the information system are considered. The essence of the approaches lies in the use of an electronic document or its screen mapping as an intermediate link. Efficiency of the given methods and, consequently, reduced time for reaching the set goal are illustrated on examples.

Keywords: e–document, screen form, database, information system.

С.Ю. Марулін

ДОКУМЕНТНО–ОРИЄНТОВАНА АВТОМАТИЗАЦІЯ ПРОЦЕСУ ЗАПОВНЕННЯ БАЗИ ДАНИХ ІНФОРМАЦІЙНОЇ СИСТЕМИ

Анотація Розглядаються різні способи ведення баз даних. Описано підходи, що дозволяють скоротити час на перенесення вмісту паперових копій документів в базу даних інформаційної системи. В основі цих підходів лежить використання в якості посередника електронного документа або його відображення у вигляді екранної форми. На прикладі показано ефективність використання даних підходів, і як наслідок скорочення часу на досягнення поставленої мети.

Ключові слова: електронний документ, екранна форма, база даних, інформаційна система.

Введение. Традиционные способы ведения документооборота, использующие как промежуточное звено электронный документ (ЭД), сталкиваются со значительной проблемой – переносом содержимого множества бумажных документов в базу данных (БД) информационной системы.

Существует основополагающее правило внедрения и использования автоматизированной ИС (АИС) на предприятии – «Одна точка сбора», «Данные собираются там, где они появляются» [1]. Использование этих правил позволяет избежать затрат на необоснованное дублирование информации и, что важнее, – потерь от ошибок учёта, неизбежно возникающих при дублировании точек ввода. Внедрение АИС на предприятии приводит к необходимости оснащения всех точек ввода информации автоматизированными рабочими местами.

Но подобные решения могут иметь ограничения, связанные с недостаточно развитой компьютерной сетью и требованиями по безопасности в процессах ввода данных в БД. Поэтому одним из решений с учетом указанных ограничений является использование промежуточных структур данных, редактируемых на основе автономных офисных пакетов.

В работе [5] предложена технология автоматизированного ведения БД на основе ЭД, отдельные элементы которой внедрены в систему «Электронный деканат». Опыт эксплуатации показал, что для разных видов бумажных документов с использованием различных способов внесения данных (рис. 1) в БД информационной системы трудозатраты различны.

Поэтому целью работы стало сокращение времени на выполнение процесса переноса содержимого множества бумажных документов в базу данных информационной системы с использованием разных способов.

© Марулин С.Ю., 2012

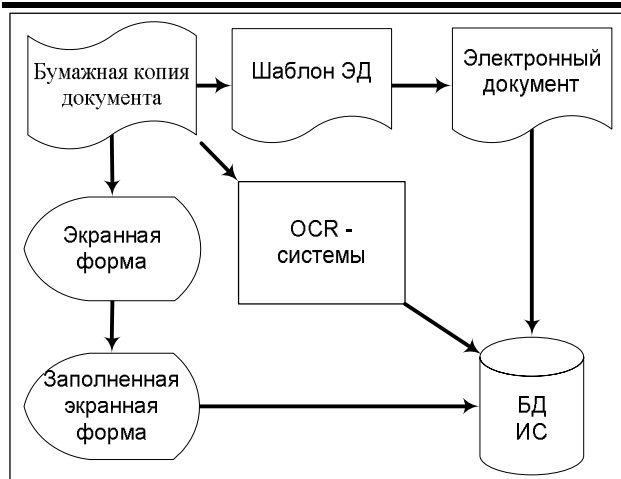


Рис. 1. Способы ведения БД

Математическая постановка задачи оптимизации процессов ведения БД на основе ЭД. Любое предприятие или организация сталкивается с проблемой оптимизации рабочих процессов. Выбор оптимального решения при достижении определенной цели позволяет сэкономить не только время, но и как известно – деньги. Зачастую на пути к решению приходится сталкиваться с определенными трудностями и ограничениям, которые, несомненно, должны быть учтены. Поэтому выбор оптимального пути решения в условиях определенных ограничений является одной из важных задач.

В общем виде задачу оптимизации можно представить в следующем виде [1]:

$$\left\{ \begin{array}{l} ЦФ : F = f(x_j) \rightarrow \max // \min // value \\ ОГР : g_i(x_j) \leq b_i \\ ГРУ : d_j \leq x_j \leq D_j \\ i = 1..m; j = 1..n \end{array} \right\},$$

где ЦФ – целевая функция, или критерий оптимизации, показывающий, в каком смысле решение должно быть оптимальным; ОГР – ограничения: односторонние $g_i(x_i) \leq b_i$ или двухстороннее $a_i \leq g_i(x_i) \leq b_i$; ГРУ – граничные условия, показывающие, в каких пределах могут быть значения искомых переменных в оптимальном решении.

В данной работе рассматривается частная задача подсчета времени на ручные операции ввода данных в условиях ограничений по типу документа, типу метода переноса и наличия сети доступа к БД информационной системы.

Опишем исходные данные (ИД) постав-

ленной задачи в виде тройки

$$\langle U, D, N \rangle,$$

где U – множество пользователей; D – множество документов; $N \subset U$ – множество пользователей, имеющих доступ к сети.

Целевая функция в данном случае стремится к минимальным затратам времени на перенос бумажных документов в БД информационной системы

$$F = f(D, U, N) \rightarrow \min.$$

В качестве ограничений выступает

вектор $\vec{C} = \langle c_1, \dots, c_i, \dots, c_n \rangle$, где $c_i = \begin{cases} 0, & c_i \in N \\ 1, & c_i \notin N \end{cases}$.

Описание последовательности действий по достижению цели с использованием ЭД. На рис. 2,а представлена последовательность действий по достижению поставленной цели с использованием автоматизированного подхода на основе шаблонов – метод А. На рис. 2,б представлена последовательность действий по достижению поставленной цели с использованием экранных форм (ЭФ) – метод В.

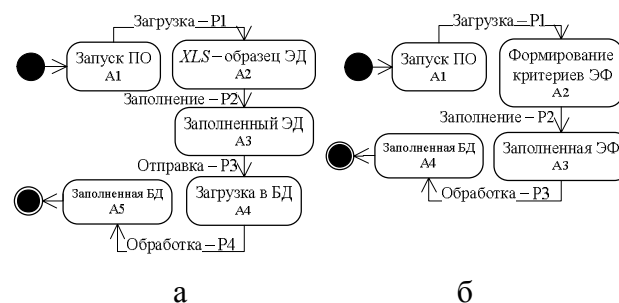


Рис. 2. Диаграмма состояний:
а – метод А; б – метод В

Каждый метод определяет состояние A , а сопровождает A действие P . Для каждого метода определены свои действия и процессы.

Для метода А: $A1$ – запуск ПО, $A2$ – XLS – образец ЭД, $A3$ – ЭД, $A4$ – загрузка в БД, $P1$ – загрузка, $P2$ – заполнение, $P3$ – отправка, $P4$ – обработка.

Для метода В: $A1$ – запуск ПО, $A2$ – формирование критериев ЭФ, $A3$ – заполненная ЭФ, $P1$ – загрузка, $P2$ – заполнение, $P3$ – обработка.

Описание возможных сценариев. Для каждого из представленных методов существует четыре сценария использования (рис. 3).

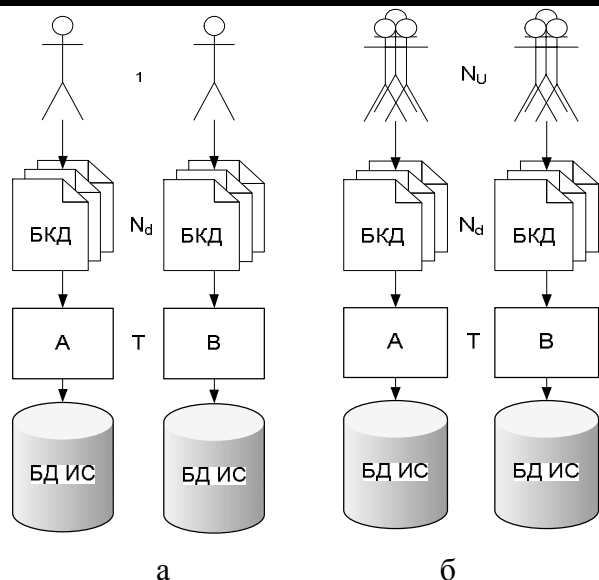


Рис. 3. Сценарии использования:
а – метод А; б – метод В

На рис. 3,а один пользователь выполняет операции по внесению данных из множества ЭД в БД информационной системы с использованием метода А и В. На рис. 3,б множество пользователей выполняет операции по внесению данных из множества ЭД в БД информационной системы с использованием метода А и В. По каждому из четырех сценариев можно подсчитать общее время на перенос информации

для 1-го:
$$T_{общ} = \sum_{i=1}^{N_d} T_{A_i},$$

для 2-го:
$$T_{общ} = \sum_{i=1}^{N_d} T_{B_i},$$

для 3-го:
$$T_{общ} = \frac{\sum_{i=1}^{N_d} T_{A_i}}{N_u},$$

для 4-го:
$$T_{общ} = \frac{\sum_{i=1}^{N_d} T_{B_i}}{N_u},$$

где N_d – множество электронных документов d ; $T_{общ}$ – время на перенос N_d в БД информационной системы; T_A – суммарное время на выполнение операции P_i с использованием метода А; T_B – суммарное время на выполнение операции P_i с использованием метода В; N_u – множество пользователей u ; P_i – i -й процесс в цепочке процессов по переносу данных.

Для метода А суммарное время рассчитывается по формуле

$$T_A = T_{P_1} + T_{P_2} + T_{P_3} + T_{P_4}, \quad (1)$$

Для метода В суммарное время

$$T_B = T_{P_1} + T_{P_2} + T_{P_3}, \quad (2)$$

Наиболее трудоемкий процесс – процесс P_2 , который подразумевает выполнение ручных операций по вводу данных, а время, потраченное на его реализацию, зависит от типа вносимых данных и выбора метода. Таким образом, отдельно можно посчитать время по процессу P_2 .

Если обозначить $A(T_{P_2})_{string}$ как время на выполнение операции P_2 с использованием метода А при условии, что тип вносимых данных $string$, тогда

$$A(T_{P_2})_{string} = \sum_{i=1}^{N_s} S_i \cdot length * t_{cp.string}, \quad (3)$$

где N_s – множество строк S ; $S_i \cdot length$ – длина (в символах) i -й строки S ; $t_{cp.string}$ – среднее время на ввод одного строкового символа.

Если обозначить $B(T_{P_2})_{string}$ как время на выполнение операции P_2 с использованием метода В при условии, что тип вносимых данных $string$, тогда

$$B(T_{P_2})_{string} = \sum_{i=1}^{N_s} S_i * t_{f.e.}, \quad (4)$$

где S_i – искомая строка $S=1$; $t_{f.e.}$ – среднее время на поиск S_i строки.

Система визуального восприятия человека может «сфотографировать» полную страницу печатного текста за одну двадцать пятую долю секунды, а скорость визуального сканирования текста с целью поиска необходимого элемента зависит от знания структуры текста [4], поэтому $t_{f.e.} = 2$ с.

Обозначим $A(T_{P_2})_{int}$ время на выполнение операции P_2 с использованием метода А при условии, что тип вносимых данных $integer$ тогда

$$A(T_{P_2})_{int} = \sum_{i=1}^{N_s} S_i \cdot length * t_{cp.int}, \quad (5)$$

где N_s – множество строк S ; $S_i \cdot length$ – длина (в символах) i -й строки S ; $t_{cp.int}$ – среднее время на ввод одной цифры.

Обозначим $B(T_{P_2})_{int}$ время на выполнение операции P_2 с использованием метода В при условии, что тип вносимых данных $integer$, тогда

$$B(T_{P2})_{int} = \sum_{i=1}^{N_s} S_i * t_{cp.int} , \quad (6)$$

где S_i – искомая строка $S=I$, $t_{cp.int}$ – среднее время на ввод одной цифры.

Общая трудоемкость (*labor input*) процесса переноса данных из ЭД в БД информационной системы с использованием как метода A , так и метода B представляется как:

$$lab = \sum_{i=1}^{N_s} (S_i.length * t_{cp.string} * c_i + S_i * t_{f.e.} * (1 - c_i)) * k^t + (S_i.length * t_{cp.int} * c_i + S_i * t_{cp.int} * (1 - c_i)) * (1 - k^t)$$

$$k^t = \begin{cases} 1, npu T = string \\ 0, npu T = integer \end{cases} ,$$

где k^t – коэффициент, определяющий тип входного документа.

В нормативном документе о “Единых нормах времени (выработки) на машинописные работы” [3] определены нормы времени на набор текста на персональном компьютере в зависимости от сложности оригинала, с которого набирается материал, размеров интервалов и количества граф (колонок), предусмотренных оригиналом.

По сложности оригиналы делятся на три группы: I – печатный или разборчивый рукописный оригинал; II – печатный или рукописный оригинал с поправками и вставками, затрудняющими прочтение, до 50 % текста; III – неразборчивый рукописный оригинал с большим количеством поправок, превышающих 50 % текста, мелкий типографский или компьютерный шрифт.

В зависимости от уровня сложности учитываются нормы времени на набор текста на типовых бланках (формах, таблицах).

1. Нормы времени на 100 печатных знаков

Группа сложности оригинала	Норма времени мин. (НВ)
I	1,80
II	2,20
III	2,70

Учитывая эти нормы времени, рассчитываем $t_{cp.string}$ и $t_{cp.int}$ (среднее время на ввод одного буквенного или цифрового символа), с

$$t_{cp.string} = t_{cp.int} = \frac{HB_N * 60}{100} ,$$

где HB_N принимает значения: для I группы сложности $\approx 1,08$; для II группы $\approx 1,32$; для III группы $\approx 1,62$.

На основании описанных формул предложена методика сокращения трудоемкости процесса переноса содержимого бумажных копий документов в базу данных информационной системы, которая включает следующие пять этапов:

- подготовку ИД;
- определение типа вносимых данных;
- определение уровня сложности исходного документа;
- определение метода ввода данных;
- расчет времени процесса переноса данных по выбранным характеристикам.

Апробация. В качестве входных данных для предложенного подхода рассмотрим два документа учебного процесса с двумя разными типами вносимых данных: «Закрепление кураторов и старост», «Итоговая ведомость».

Документ «Закрепление кураторов и старост».

Тип данных: *String*.

Уровень сложности: I.

Метод ввода: A .

Расчет: по формуле 1 получим общее время T_A . Суммарное время T_{P1} , T_{P3} , $T_{P4} \approx \approx 180$ сек. По формуле (3) рассчитаем T_{P2} :

$$A(T_{P2})_{string} = \sum_{i=1}^{30} S_i.length * 1.8 = 1154 \text{ сек,}$$

$$T_A = 180 + 1154 = 1334 \text{ сек} = 22.2 \text{ мин.}$$

Документ «Закрепление кураторов и старост».

Тип данных: *String*.

Уровень сложности: I.

Метод ввода: B .

Расчет: по формуле (2) получим общее время T_B . Суммарное время T_{P1} , $T_{P3} \approx 60$ с. По формуле 4 находим T_{P2} :

$$B(T_{P2})_{string} = \sum_{i=1}^{30} S_i * 2 = 60 \text{ с,}$$

$$T_B = 60 + 60 = 120 \text{ с} = 2 \text{ мин.}$$

Документ «Итоговая ведомость».

Тип данных: *int*.

Уровень сложности: I.

Метод ввода: A .

Расчет: по формуле (1) получим общее время T_A . Суммарное время T_{P1} , T_{P3} , $T_{P4} \approx 180$ с. По формуле (3) рассчитываем T_{P2} :

$$A(T_{P2})_{int} = \sum_{i=1}^{25} S_i.length * t_{cp.int} = 90 \text{ с,}$$

$$T_A=180+90=270 \text{ с} = 4.5 \text{ мин.}$$

Документ «Итоговая ведомость».

Тип данных: *int*.

Уровень сложности: I.

Метод ввода: *B*.

Расчет: по формуле (2) находим общее время T_B . Суммарное время T_{P1} , $T_{P3} \approx 60$ с. По формуле (5) определяем T_{P2} :

$$B(T_{P2})_{int} = \sum_{i=1}^{25} S_i * t_{cp.int} = 50 \text{ с,}$$

$$T_B = 60+50=110 \text{ с} = 1.8 \text{ мин.}$$

Результаты экспериментов сведены в таблицу и показываю во сколь раз снижается время на перенос данных в БД в сравнении метода А и В:

$$((T_A/T_B)_{string} + (T_A/T_B)_{int})/2 = 7 \text{ раз.}$$

2. Время на перенос данных в БД, мин

Название документа	Метод ввода	Тип данных	
		<i>String</i>	<i>int</i>
Закрепление кураторов и старост	A	22.2	---
	B	2	---
Итоговая ведомость	A	---	4.5
	B	---	1.8

Для демонстрации работы метода *A* и *B* была разработана программа «Электронный деканат». Программа позволяет управлять ЭД и процессами, сопровождающими их, с использованием метода *A* и *B*. Для метода *A* она позволяет сгенерировать шаблон ЭД. Пользователь вручную заполняет шаблон и отправляет его на автоматическую обработку.

Для каждого вида ЭД предусмотрена экранная форма (метод *B*), которую можно заполнять в режиме реального времени и записи сразу попадают в БД информационной системы. Расчетные данные были подтверждены экспериментами, проводимыми на кафедре системного программного обеспечения Одесского национального политехнического университета.

Выводы. Представленная методика позволяет определить наиболее рациональный способ заполнения БД с использованием либо электронного документа, либо экранных форм представления электронного документа. Выбор метода определяет тип данных и уровень сложность исходного документа. Предложена формула расчета времени опе-

рации ручного ввода символов, а также расчета времени операции выбора необходимого значения из списка. На примере показана эффективность использования экранных форм для ведения БД. Данный подход позволяет в семь раз снизить время на перенос данных из бумажных документов в БД информационной системы.

Список использованной литературы

1. Гурман В. И. Абстрактные задачи оптимизации и улучшения / В. И. Гурман // Программные системы: теория и приложения. – М.: 2011, № 5 (21). – 29 с.

2. Маглинец Ю.А. Анализ требований к автоматизированным информационным системам: Учебное пособие / Ю.А. Маглинец – М.: Интернет–Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2008. – 200 с.

3. Постановление Минтруда России от 26 марта 2002 г. № 23. Об утверждении норм времени на работы по документационному обеспечению управленческих структур федеральных органов исполнительной власти.

4. Развитие продвинутых навыков скольжения и сканирования при быстром чтении (29.01.2012.). WEB–сайт (Электрон. ресурс) / Способ доступа: URL: <http://www.prosvetlenie.org/razum/16/index.html>. – Загол. с экрана.

5. Blazhko A. A. Data Exchange Technology Between Electronic Documents and Relation Databases / A. A. Blazhko, S. Marulin, V.Kalashnikova // Procs of 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), September 15–17, 2011, Prague, Czech Republic. – P. 624–628

Получено 08.02.2012

References

1. Gurman V.I. Abstract of the problem of optimization and improvement / Software systems: theory and applications – Moscow: 2011, № 5 (21). – 29 p. [in Russian].

2. Maglinets Y. A. Analysis of the requirements for automated information systems: the manual – Moscow: Internet University of Information Technology; BINOM. Laboratory of Knowledge, 2008. – 200 p. [in Russian].

3. Ministry of Labor of Russia from March 26, 2002 № 23 «On approval of rules of time to work on the documentation ensure administrative structures federal executive authority» [in Russian].

4. Development of advanced skills, slip and scan with rapid reading (29.01.2012). WEB-site (Electronic resource) / access: URL: <http://www.prosvetlenie.org/razum/16/index.html>. – Headers from the screen.

5. Blazhko A. A., Marulin S., Y. Kalashnikova V. Data Exchange Technology Between Electronic Documents and Relation Databases / Procs of 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), September 15–17, 2011, Prague, Czech Republic. – P. 624–628 [in English].



Марулин
Станислав Юрьевич,
аспирант каф. системного
программного обеспечения
Одесск. нац. политехн. ун-
та,
e-mail: StasFoot@mail.ru