

УДК 004.42

Е.А.Орлов (5 курс, каф. КИТвП), С.Г.Попов, асс.

РАЗРАБОТКА ПОДСИСТЕМЫ АВТОМАТИЗИРОВАННОЙ ГЕНЕРАЦИИ СТРУКТУРЫ БАЗЫ ДАННЫХ

На сегодняшний день наблюдается явный переход от бумажного к электронному документообороту. Электронный документооборот характеризуется высокой степенью формализма и высокими скоростями обмена документами. Как положительный эффект, возрастает объем обрабатываемых данных и появляется возможность удаленного доступа к электронному документу.

Однако существенной проблемой при создании систем электронного документооборота является разработка хранилища данных для экземпляров документов. На сегодняшний день существует два подхода к решению этой проблемы: классический программистский и автоматизированный. При первом подходе аналитик разрабатывает схемы хранения, а потом для них программируются конкретные процедуры обработки и управления. При втором подходе схема хранилища данных генерируется автоматизированно по предоставленным пользователем описаниям первичных документов; в этом случае аналитик освобождается от рутинных операций проектирования. Именно исследованию вопроса эффективной реализации второго подхода и посвящена данная работа.

В вычислительной машине можно построить иерархию структур данных. Можно выделить уровни аппаратуры, драйвера, операционной системы, внешних программ и прикладной программы. На всех уровнях, кроме последнего, данные уже структурированы, и имеются процедуры для их контроля и управления. Разработка структур данных на уровне прикладной программы ложится на пользователя и может быть решена неоднозначно.

Под содержательной обработкой данных прикладной программой будем понимать такое их агрегатное преобразование, которое обеспечивает удовлетворяющий пользователя результат.

Пусть имеется некоторый блок данных D объемом $V(D)$ и программа P , обрабатывающая эти данные. При работе с блоком D программа содержательно обрабатывает только часть данных объемом $P(D)$. Тогда коэффициентом структурированности блока данных D при обработке программой P является: $Str(D, P) = P(D)/V(D)$.

Опираясь на эти определения, можно разделить технологии организации хранения и обработки данных на те, которые используют жестко формализованные данные ($str \sim 1$), и на те, в которых данные слабо структурированы ($str \ll 1$).

При первом подходе эффективно и просто реализуются алгоритмы управления данными, однако разработка схемы хранения сложна и требует высокой квалификации аналитика в области разработки схем баз данных (пример: 1С).

Второй подход характеризуется неэффективностью и сложностью обработки данных при простоте добавления новых данных (Lotus Notes).

Таким образом, нам кажется целесообразным разработать технологию генерации схемы базы данных по описаниям первичных документов, удовлетворяющую следующим требованиям:

- независимость от предметной области;
- автоматизированное создание хранилища данных;
- ввод данных посредством диалога с пользователем;
- настройка параметров хранилища данных пользователями.

В ходе выполнения работы были получены следующие результаты:

- проанализировано понятие структурированности данных на различных уровнях иерархии структур ЭВМ; выделены признаки внешних и внутренних, по отношению к прикладной программе структур;
- определено понятие содержательной обработки и введено понятие коэффициента структурированности как ее количественной меры;
- проведено исследование первичных документов, и сделан вывод о распространенности табличной формы как наиболее характерной для их представления;
- предложена схема эффективного хранения описаний первичных документов, представленных в табличной форме;
- разработана структурно – функциональная схема подсистемы, отвечающей за создание хранилищ документов, состоящая из следующих компонент – интерфейс с пользователем, генерация структуры БД генерация инструкций для создания БД во внешней СУБД;
- в качестве интерфейса пользователя предложен полноэкранный графический интерфейс, состоящий из меню и форм;
- в ходе разработки функций модуля «Генерации структуры БД» выдвинуты два критерия оптимизации – минимизация необходимого дискового пространства и минимизация времени доступа к данным

Дальнейшая работа будет заключаться в более тщательном исследовании критериев оптимизации и разработке алгоритмов, реализующих генерацию схемы СУБД по соответствующему критерию. Доказательством конкурентоспособности разрабатываемой технологии послужит макет программного обеспечения.