

РАЗРАБОТКА ХРАНИЛИЩА ДАННЫХ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ ДЛЯ ВРАЧА-ТРАВМАТОЛОГА

А. В. Бушманов, Ю. С. Пчелинова

Амурский Государственный университет

Поступила в редакцию 10.10.2009

Аннотация. В статье рассматривается способ организации хранения данных в системе поддержки принятия решения, который проектируется на основе теоретико-множественного подхода с изучением информационной потребности пользователя.

Ключевые слова: система поддержки принятия решения, хранилище данных.

Annotation. This article tells about the method of data storage organization in decision support system, which is engineered on basis of theory of multiple approach and user's informational demand.

The keywords: decision support system, data warehouse.

ВВЕДЕНИЕ

Автоматизированные медицинские системы широко применяются во многих учреждениях здравоохранения. Наибольший интерес для практической медицины представляют системы поддержки принятия решения (СППР). При организации СППР используются хранилища данных, которые выполняют функции предварительной подготовки и хранения данных, полученных из различных источников [1]. Хранилища данных служат исключительно информационным целям, и данные в них доступны пользователям только в режиме чтения. Информация в этих системах характеризуется стабильностью, т. е. однажды загруженные данные теоретически никогда не меняются. По отношению к ним возможны только две операции: начальная загрузка и доступ.

Известные медицинские экспертные системы, такие как МОДИС, КОНСУЛЬТАНТ-2, ЛЕДИ-Z, «ОБЕЗБОЛИВАНИЕ», «ХИРУРГ», «ДИАЛОГ-Д», «АСПОН-Д», MYCIN, CASNET, INTERNIST, PIP, IRIS, EXPERT, PUFF, HODGKINS, HEADMED и др., позволяют проводить консультации при постановке диагноза и лечении, поддерживать ход врачебных рассуждений, дополняя и восстанавливая медицинские знания, моделировать патологические процессы, формировать клинические картины заболеваний. Некоторые системы обладают

способностью улучшать свое качество в процессе функционирования. Однако следует отметить, что при всех своих достоинствах, существующие системы, даже в рамках одной предметной области-медицины, направлены на решение своих специфических задач в узкой области знаний. Сильная проблемная зависимость систем типа MYCIN определяется ее проблемно-зависимой управляющей структурой [2].

1. ПОСТАНОВКА ЗАДАЧИ

В основе современного подхода к информационному обеспечению

СППР лежит идея интегрированного хранилища данных, обеспечивающего единый логический взгляд и доступ к информации, находящейся в нескольких внешних источниках. При этом существенно, что данные в хранилище имеют исторический характер, то есть обеспечивается интеграция не только разнородных источников, но и архивных данных, возникающих в процессе функционирования той или иной системы [4].

В процессе принятия решения о способе лечения травмы врачу-травматологу необходимо знать следующее:

- вид травмы (перелом, вывих), степень сложности травмы,
- анамнестические данные (пол, возраст, вес пациента);
- сопутствующие заболевания, которые могут наложить ограничения при выборе способа лечения;

— существующие способы лечения травм (фиксирующие устройства для оперативного и консервативного лечения).

В соответствии с необходимыми знаниями для принятия решения в хранилище данных загружается информация из нескольких баз данных: БД «Перелом», БД «Фиксирующее устройство», БД «Ограничение», БД «Анатомия» и др.

В настоящем исследовании обоснованность применения технологии хранилища данных объясняется разнородностью исходных данных. Разнородность заключается как в форме представления: графические данные (рентгенограммы, схемы, чертежи), электронные таблицы, фалы, полученные в результате экспорта из учетных систем, так и в структуре самих данных.

Таким образом, для достижения поставленной цели необходимо решить следующие задачи [3]:

— определить информационные потребности пользователя — врача-травматолога создаваемой системы, по отношению к данным, накапливаемым в БД оперативных систем источников данных;

— изучить структуру локальных баз данных;

— выделить для каждой такой БД подмножества данных, необходимых для загрузки в хранилище;

— интегрировать локальные подмножества данных и разработать общую согласованную схему хранилища.

Целью исследования является разработка хранилища данных системы поддержки принятия решения (СППР) с использованием теоретико-множественного подхода, ориентированного на информационную потребность пользователя — врача-травматолога.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

При построении системы на базе хранилища наиболее важным является этап определения информационной потребности пользователя. С целью формализации описания информационной потребности, а также оптимизации структуры хранилища, используется математическая модель хранилища данных, основанная на теоретико-множественном подходе к описанию информационных потоков [4]. Теоретико-множественный подход использует понятия: факты, измерения, аналитические срезы фактических

данных. Аналитические срезы фактических данных — это представление данных по нескольким измерениям одновременно.

Теоретико-множественное представление фактических данных базируется на представлении фактов в виде множества измерений одного и того же факта, а представление информационных потоков — в виде набора различных фактов (фактических данных в различных аналитических разрезах, составленных на базе совокупности измерений). Представление фактических данных в виде аналитического среза описывается формулой:

$$F_0 = \{f_1, \dots, f_k, \dots, f_n\}, \quad (1)$$

где $f_k \in F_0$ для $\forall k$, $|F_0| = n_0$, $|U| = U$, F_0 — универсальный массив, содержащий все аналитические срезы фактических данных, n_0 — мощность множества F_0 , т.е. общее количество аналитических срезов в информационном массиве, U — универсальный набор представлений фактов, т.е. количество измерений, по которым могут быть представлены факты.

Соотношение (1) образует теоретико-множественное описание информационных потоков. Дополняет определение (1) линейное представление теоретико-множественного образа аналитического среза фактических данных:

$$f_k = \{d_{1k}, \dots, d_{jk}, \dots, d_{uk}\}, \quad (2)$$

где $d_{jk} = \begin{cases} 1, & \text{измерение входит} \\ \text{в } k\text{-й аналитический срез;} \\ 0, & \text{в противном случае.} \end{cases}$

Вектор аналитических срезов F , включает процессы получения множества аналитических срезов F_0 и преобразования (индексирования) каждого аналитического среза в образ с созданием массива образов аналитических срезов.

Теоретико-множественный подход предполагает наличие вектора запросов Q , сформированного на основе изучения частоты запрашиваемой врачом информации, то есть на основе его информационной потребности, включает множество запросов Q_0 (4):

$$Q_0 = (a_{11}, \dots, a_{j1}, \dots, a_{jm0}, \dots, a_{U1}, \dots, a_{Um0}), \quad (4)$$

где a_{11}, \dots, a_{11} — поля таблиц, по которым производится выборка данных.

Для определения, насколько результат запроса соответствует информационной потребности, вычисляется критерий его соответствия

информационной потребности, то есть формальная оценка. Критерий соответствия информационной потребности рассчитывается по формуле:

$$K = \sum_{i=1}^n \beta_i k_i^c, \quad (5)$$

где β — мера формальной релевантности, или мера близости образов аналитического среза и образов пользовательского запроса; k^c — пороговое значение меры близости, при превышении которого аналитический срез остается формально релевантным соответствующему запросу.

Сопоставление информационной потребности и возможности предоставления аналитической информации в необходимом виде состоит в построении матрицы соответствия или подобия:

$$\beta = \begin{pmatrix} \beta_{11} & \dots & \beta_{1k} & \dots & \beta_{1n_0} \\ \dots & \dots & \dots & \dots & \dots \\ \beta_{m1} & \dots & \beta_{mk} & \dots & \beta_{mn_0} \\ \dots & \dots & \dots & \dots & \dots \\ \beta_{m_01} & \dots & \beta_{m_0k} & \dots & \beta_{m_0n_0} \end{pmatrix}, \quad (6)$$

где элементы β_{mk} матрицы β определяется по формуле:

$$\beta^* = \frac{\sum_{i=1}^U a_{im} \cdot d_{ik}}{\sqrt{\left(\left(\sum_{i=1}^U a_{im}^2 \right) \cdot \left(\sum_{i=1}^U d_{im}^2 \right) \right)}}. \quad (7)$$

Элементы матрицы β представляют собой множества формально релевантных m -му запросу аналитических срезов F_m^C , таких, что для всех m : $f_k \subset F_m^C \Leftrightarrow \beta_{mk} = \beta(q_m, f_k) > k^c$.

Формальным релевантным множеством является множество элементов, относящихся к определенному классу, относительно которого производится выборка. В рассматриваемом примере, формально релевантным множеством F_m^C является набор всех фиксирующих устройств, которые подходят для лечения конкретного вида перелома кости.

Задание множества F_m^C оворит о качестве обеспечение информационной потребности q_m , если его рассматривать совместно с множеством истинно релевантных аналитических срезов $F_m^И$, поскольку в общем случае:

$$F_m^И \neq F_m^C. \quad (8)$$

Истинно релевантным множеством является множество элементов, входящих в группу

формально релевантного множества, то есть $F_m^C \subset F_m^И$, но является более точным, соответствует критериям отбора, которые задает врач.

Таким образом, результат выполнения запроса будет однозначно соответствовать информационной потребности пользователя, в том случае когда:

$$F_m^И = F_m^C. \quad (9)$$

Степень релевантности образов аналитического среза по отношению к информационному запросу определяется человеком и зависит от его индивидуальных особенностей.

3. ЛОГИЧЕСКАЯ МОДЕЛЬ ХРАНИЛИЩА ДАННЫХ

Способ проектирования хранилища данных на основе теоретико-множественного подхода относится к многомерному моделированию. В многомерном моделировании выделяются таблицы измерений и таблицы фактов. Таблицы измерений содержат основные данные (пол, возраст, вес), а таблицы фактов содержат данные транзакций (перелом, кость, фиксирующие устройства).

Запросы поддержки решения следуют в такой последовательности: запрос выбирает несколько нужных мер из таблицы фактов, объединяет строки фактов с одним или несколькими измерениями вместе с суррогатными ключами, размещает предикаты фильтров в бизнес-столбцы таблиц измерений, группирует по одной или нескольким бизнес-столбцам и собирает меры, полученные из таблицы фактов. Кластеризованный индекс таблицы фактов использует несколько столбцов суррогатных ключей измерений (столбцы внешнего ключа) в качестве ключей индекса. Столбцы, используемые чаще всего, должны появиться в списке ключей индекса. Таким образом, получаются хорошие пути доступа для наиболее частых запросов в рабочей нагрузке.

К таблицам измерений, нужно создавать индексы для каждой таблицы измерения. Это некластеризованный индекс ограничения первичного ключа на столбце суррогатного ключа измерения и кластеризованный индекс столбцов бизнес-ключа объекта измерения.

Для принятия решения лицо, принимающее решение анализирует большой объем информации о схожих случаях травм, способах лечения этих травм и результатах лечения.

В соответствии с выбранным методом построения хранилища данных «звезда», в качестве фактических выступают следующие данные: перелом, фиксирующее устройство и ограничение. С целью конкретизации этих данных и добавления им аналитической составляющей, данные могут быть представлены по следующим аналитическим измерениям: смещение отломков, ограничения по состоянию здоровья.

Сначала загружаются справочники или таблицы измерений, а для того чтобы избавиться от часто встречающихся в базах данных транзакционных систем составных ключей и

для унификации ключевых полей во всем хранилище, таблицы дополняются суррогатными ключами, независимо от наличия ключевых полей в источниках. Вслед за справочниками в хранилище загружаются факты — данные, по которым будет проводиться анализ и составление отчетов. Фактические данные из источников дополняются при преобразовании набором суррогатных ключей связанных с ними измерений. Преобразования включают предварительные расчеты, суммирование и группировку данных. Структура хранилища данных представлена на рисунке 1.



Рис. 1. Структура хранилища данных

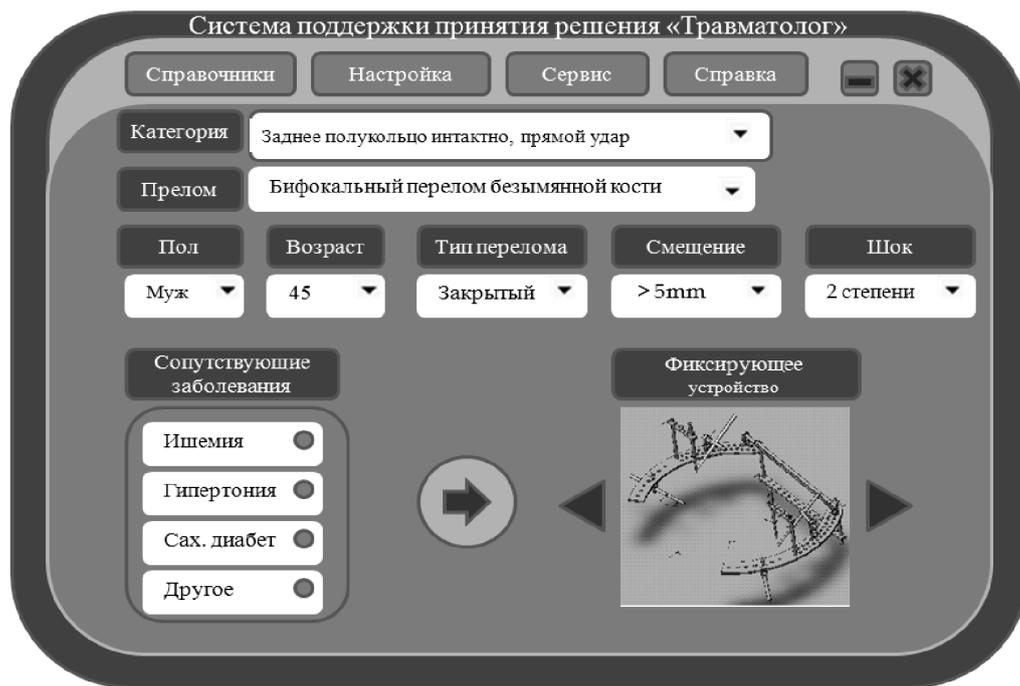


Рис. 2. Главная форма СППР для врача-травматолога

4. ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ СППР ДЛЯ ВРАЧА ТРАВМАТОЛОГА

Система поддержки принятия решения структурно состоит из двух частей — клиентской и серверной. Клиентское приложение может быть запущено на нескольких стационарных персональных компьютерах. Главная форма программы, представленная на рисунке 2, дает возможность для непосредственной работы врача — травматолога, т.е. принятие решения. В окне главной формы задается все характеристики травмы, а так же анамнестические данные пациента с указанием сопутствующих заболеваний, которые могут послужить ограничением при выборе метода лечения. После введения известных врачу — травматологу данных о переломе, СППР выдает один из вариантов лечения, который представлен на картинке (рис.2). Нажимая по стрелкам «лево» или «право» можно просмотреть и другие способы лечения перелома, если таковые имеются для данного случая.

РЕЗУЛЬТАТЫ РАБОТЫ

Разработанное хранилище данных применяется врачом-травматологом как помощник в процессе принятия решения. Система поддержки принятия решения врача-травматолога построена по технологии хранилищ данных. Технология хранилища данных позволила ин-

тегрировать локальные подмножества данных и объединить их в согласованную структуру. Схема хранилища обеспечивает возможность извлечение данных в необходимых пользователю срезах для их дальнейшего представления в аналитическом приложении. На основе полученных результатов врач определяет мероприятия по лечению травмы в период реабилитации. Использование теоретико-множественного подхода при проектировании хранилища данных позволяет быстро и эффективно представлять данные специалисту, что сокращает время затрачиваемое на поиск и принятие решения. Эта технология может быть использована при проектировании систем другой направленности, например, в экономике, планировании и т.п.

СПИСОК ЛИТЕРАТУРЫ

1. Львов В. Создание систем поддержки принятия решений на основе хранилищ данных // Системы управления базами данных. № 3. — 1997.
2. Rogers W. Computer — Aided Medical Diagnosis: Literature Preview. International Journal of Biomedical Computing, № 10, — 1979.
3. Повышение производительности хранилищ данных. 1 // ComputerWeek-Moscow. 1996. — № 32. — С. 28.
4. Александрин А.М. Разработка и реализация методов и моделей информационной системы поддержки принятия решений на уровне предприятия.

Разработка хранилища данных системы поддержки принятия решения для врача-травматолога

Бушманов А. В. — к.т.н., доцент, Амурский государственный университет.

Пчелинова Ю. С. — аспирант кафедры информационных и управляющих систем, Амурский государственный университет.

Bushmanov A. V. — technical scientific candidate, docent, Amur State University, Тел. 8 (4162) 39-46-53 . E-mail: bush29@mail.ru

Pchelinova Y.S. — post-graduate student of the chair of the informational and controlling systems. Тел. 8-914-553-18-07. E-mail: pchelik84@mail.ru