

ОБЪЕКТНО-ОРИЕНТИРОВАННЫЕ БАЗЫ ДАННЫХ КАК ОСНОВА ФОРМИРОВАНИЯ БАЗ МЕДИКО-БИОЛОГИЧЕСКИХ ЗНАНИЙ. ИНФОРМАЦИОННАЯ СИСТЕМА «СЕМЕЙНАЯ МЕДИЦИНА»

Описывается разработанная автором информационная медицинская система «Семейная медицина» и ее возможное применение при обучении студентов специальности «Лечебное дело» (специализация «Семейный врач») и в повседневной практической работе семейного врача.

Классические реляционные системы управления базами данных (БД), несмотря на их большую распространенность, являются ограниченными. Они идеально подходят для таких традиционных приложений, как всевозможные системы учета, но их применение в интеллектуальных системах обучения и других системах, основанных на знаниях, часто является весьма затруднительным. Это прежде всего связано с примитивностью структур, лежащих в основе реляционной модели данных. Теоретически они достаточны для представления данных любой предметной области. Однако в нетрадиционных приложениях в базе данных появляются сотни, если не тысячи таблиц, над которыми постоянно выполняются технически дорогостоящие операции, необходимые для воссоздания сложных структур данных, присущих предметной области, что ведет к невозможности практической реализации подобных систем.

Другим серьезным ограничением реляционных систем являются их относительно слабые возможности по части представления семантики приложения. При попытке проектирования баз данных ограниченность реляционной модели проявляется в следующих аспектах.

Модель не предоставляет достаточных средств для представления смысла данных, в связи с этим семантика реальной предметной области должна формироваться независимым от модели способом и после разработки приложения оставаться на «бумаге» и не учитываться при работе приложения.

Трудно, а в некоторых случаях практически невозможно моделировать предметную область на основе плоских таблиц. Попытки проектирования зачастую требуют определенного «насилия» проектировщика – специалиста в конкретной области знаний – над собой, чтобы описать предметную область в виде одной (возможно, даже ненормализованной) таблицы.

Проектирование начинается с выделения некоторых существенных для приложения объектов предметной области («сущностей»), выявления связей между этими сущностями и происходит на основе учета различных зависимостей, но при этом реляционная модель данных не предоставляет каких-либо средств для представления этих зависимостей и не предлагает какого-либо аппарата для разделения сущностей и связей.

Указанные выше недостатки реляционной модели данных и реляционных систем управления базами данных (СУБД) в контексте конкретной предметной области – медико-биологических знаний – приводят, помимо указанных, к дополнительным проблемам [1].

Как правило, при описании различных заболеваний используется следующая схема. Описывается группа родственных болезней, приводятся общие для них кли-

нические проявления, далее по тексту может идти дифференциальная диагностика между ними, затем общие и частные методы лечения каждого заболевания, осложнения и дифференциальная диагностика с другими, неродственными данным заболеваниями. Отдельные пункты этого описания могут присутствовать в полном объеме или вообще отсутствовать, может меняться и порядок их следования. В каждом конкретном случае такой порядок вещей определяется принципом целесообразности. Такой подход в описании болезней используется десятилетия и доказал свою эффективность. Попытка разложить его в виде структуры реляционных баз данных приводит к тому, что:

- во-первых, теряется целостность картины данных (в нашем случае – заболевания);

- во-вторых, как попытка компенсировать первый недостаток появляется излишнее дублирование информации;

- и в-третьих, как следствие первых двух указанных недостатков у пользователя – врача – всегда будет возникать ощущение неполноты полученной информации, что не улучшит качество его работы.

Применение реляционных СУБД заставляет как составителя, так и пользователя информационной системы использовать непривычную для него систему мышления, требующую уже на начальном этапе дробления проблемы на отдельные составляющие для распределения ее по жесткой структуре реляционной базы данных. В то же время логика профессионального мышления специалистов-врачей, формируемая за годы учебы и работы, где важную роль играют опыт и интуиция, далеко не всегда «вписывается» в логику работы реляционных СУБД, а зачастую вообще приводит, как показала практика обучения компьютерным технологиям в рамках факультета усовершенствования врачей, к неприятию данных программных средств.

Учитывая эти ограничения и недостатки реляционных систем, проводятся многочисленные исследования в области баз данных, основанные на идеях, выходящих за пределы реляционной модели. Спектр этих исследований достаточно широк. Но, принимая во внимание особенности медико-биологических знаний с точки зрения формирования баз знаний [2], одним из самых перспективных направлений, по нашему мнению, являются объектно-ориентированные базы данных (ООБД). Несмотря на то, что первые публикации по ООБД появились еще в середине 1980-х, наиболее активно это направление развивается в последнее время. С каждым годом увеличивается число публикаций и реализованных коммерческих и экспериментальных систем.

ООБД возникли не на пустом месте и обязаны своим появлением прежде всего потребностями практики: необходимостью разработки и внедрения информационных систем, для которых технология предшествующих реляционных систем не была вполне удовлетворительной. Соответствующий научный и технологический базис для их развития обеспечивают как предыдущие работы в области баз данных, так и давно развивающиеся направления языков программирования с абстрактными типами данных и объектно-ориентированных языков программирования.

При большом количестве публикаций, экспериментальных проектов (и даже коммерческих систем) на сегодняшний день отсутствует общепринятая объектно-ориентированная модель данных, хотя и имеются отдельные достаточно полно разработанные модели (например, проект ORION фирмы MCC (США), СУБД O2 фирмы Altair (Франция), система Jasmine фирмы Fujitsu (Япония) [3]). Существуют также и более конкретные технологические проблемы, связанные с разработкой декларативных языков запросов, выполнением и оптимизацией запросов, формулированием и поддержанием ограничений целостности, синхронизацией доступа и управлением транзакциями и т.д. Современная ситуация с ООБД напоминает ситуацию с реляционными системами середины 1970-х гг. [4].

Перед тем как перейти к описанию реализации ООБД, которую мы используем в своих разработках, рассмотрим те концепции, которых мы придерживались в нашей работе.

В классической постановке объектно-ориентированный подход основывается на таких понятиях, как объект и идентификатор объекта; атрибуты и методы; классы; иерархия и наследование классов [5].

Объект

Сущности реального мира в объектно-ориентированных системах представляются в виде объектов. Любой объект при своем создании получает генерируемый системой уникальный идентификатор, который связан с объектом все время его существования и не меняется при изменении состояния объекта.

Каждый объект имеет состояние и поведение. Состояние объекта – набор значений его атрибутов. Поведение объекта – набор методов, оперирующих над состоянием объекта (других объектов). Значением атрибута объекта может быть также некоторый объект или множество объектов. Состояние и поведение объекта инкапсулированы в объекте. Взаимодействие объектов осуществляется посредством передачи сообщений и выполнения соответствующих методов (генерации и обработки событий в терминах событийно-управляемого программирования).

Атрибуты и методы, классы

Множество объектов с одним и тем же набором атрибутов и методов образует класс объектов. Допускается наличие примитивных предопределенных классов, а также абстрактных классов. Класс, объекты которого могут служить значениями атрибута объектов другого класса, называется доменом этого атрибута.

Допускается порождение нового класса на основе уже существующего класса – наследование. Новый класс, на-

зываемый подклассом (потомком) существующего класса (суперкласса, родителя), наследует все атрибуты и методы суперкласса, но не всегда имеет непосредственный доступ к ним. В подклассе, кроме того, могут быть определены дополнительные атрибуты и методы.

Различаются случаи простого и множественного наследования. В первом случае подкласс может определяться только на основе одного суперкласса, во втором случае суперклассов может быть несколько. Если в языке или системе поддерживается единичное наследование классов, набор классов образует древовидную иерархию. При поддержании множественного наследования классы связаны в ориентированный граф с корнем, называемый решеткой классов. Объект подкласса считается принадлежащим любому суперклассу этого класса. Допускается также возможность переопределения атрибутов и методов суперкласса в подклассе (перегрузки методов).

Как отмечалось выше, на данный момент нет общепринятой объектно-ориентированной модели данных. Тем не менее в качестве формальной основы ООБД можно предложить следующую конструкцию [6]:

1. Два уровня моделирования объектов: нижний (структурный) и верхний (поведенческий).

2. База данных – набор элементов данных, связанных отношениями «входит в класс» или «является атрибутом». Таким образом, БД может рассматриваться как ориентированный граф.

3. Разделение схемы БД и самой БД. В качестве первичных концепций схемного уровня ООБД выступают типы и классы.

4. Наличие «метасхемы», которая играет для ООБД такую же роль, какую играет структурная часть реляционной модели данных для схем реляционных БД.

Предлагаемая нами и приведенная ниже реализация ООБД во многом определялась рассмотренными ранее концепциями, но учитывая то, что она создавалась исходя из потребности разработки информационно-справочных медико-биологических систем, имеет свои существенные отличия, которые определяются характером медико-биологических знаний.

Метасхема СУБД

Существуют группы объектов, решающие следующие задачи:

А) Доступ к данным согласно их внутренней структуре, имеющей вид ориентированного графа.

Б) Доступ к данным в соответствии с критериями, отличными от первой группы.

В) Манипуляция с данными.

Г) Дополнительные функции.

Группы А, Б и В имеют общую решетку классов, состоящую из трех объектов:

1) TBase – реализует набор основных атрибутов и поведение объектов;

2) TContainer – реализует методы хранения других объектов и навигацию по «дереву» или «списку» объектов;

3) TFrame – реализует нижний уровень доступа к данным.

Объекты группы А представлены двумя типами: объекты, обеспечивающие конструкцию графа, – Tnode (пред-

ки: TBase и TContainer), и объекты, обеспечивающие переход к нужным данным, – TItem (предок TBase).

Объекты группы Б представлены тремя типами: объекты, имеющие методы доступа к данным, – TSwitchFrame (предки: TFrame и TContainer); объекты-переключатели, обеспечивающие активизацию объектов первого типа – TSwitch (предки: TBase и TContainer) и объект, владеющий объектами-переключателями – Tswitch-Container (предок TContainer). Данная группа объектов призвана обеспечить реализацию «логики» доступа к данным, отличной от иерархической.

Объекты группы В представлены двумя типами: объекты, отвечающие за доступ к данным на носителе, за отображение данных на устройстве вывода и т.д. – TIssueFrame (предки: TFrame и Tcontainer); объекты, владеющие первым типом объектов, – TIssue-Container (предок TContainer).

Объекты группы Г представлены четырьмя типами объектов.

Первый тип этих объектов отвечает за реализацию «гиперпереходов» – гипертекст, гиперграфика («hot-spot»), гипермедиа – THyper. Второй тип – объекты, отвечающие за реализацию сносок (например, дополнительная информация, появляющаяся в всплывающих окнах), – TFootnoteFrame (предки: TFrame и TContainer). Третий тип – объекты, отвечающие за взаимодействие системы с внешним окружением (например, для запуска параллельно работающих программ), – TOuter. И четвертый тип – объект, владеющий тремя первыми типами, – TService (предок TContainer).

Данное описание объектов построено по принципу наследования. Если говорить о принципах владения и ссылок, то соотношение объектов выглядит следующим образом.

TNode может владеть объектами типа TNode и TItem.

TItem может иметь ссылки на объекты типа TNode, TItem, TSwitchFrame, TIssueFrame, может владеть объектом типа TService.

TSwitchFrame может иметь ссылки на объекты типа TIssueFrame, TNode, TItem, TSwitchFrame, может владеть объектом типа TService.

TSwitch может владеть объектами типа TSwitch-Frame.

TswitchContainer владеет объектами типа TSwitch.

TIssueFrame может иметь ссылки на объекты типа TNode, TItem, TSwitchFrame, TIssueFrame, может владеть объектом типа TService.

TissueContainer владеет объектами типа TIssueFrame.

TFootnoteFrame может иметь ссылки на объекты типа TNode, TItem, TSwitchFrame, TIssueFrame, может владеть объектами типа THyper, TOuter.

TService может владеть объектами типа THyper, TFootnoteFrame, TOuter.

Данная схема была нами реализована при разработке и создании информационной медицинской системы «Семейная медицина» [7] («Family Medicine Standard») (руководитель проекта – Я.С. Пеккер, зав. кафедрой биомедикибернетики Сибирского государственного медицинского университета; научные консультанты – академик РАМН М.А. Медведев (г. Томск) и академик

РАМН А.Ф. Краснов (г.Самара); основной разработчик – автор). Учитывая, что эта система изначально была нацелена на практическое и коммерческое использование, пришлось пойти на некоторое упрощение вышеприведенной схемы («цена/качество»). Для генерации подобных программных средств нами была создана программная среда HyperObject System (экспериментальный вариант). В этом программном обеспечении применяются полностью визуальные средства разработки, что позволяет избежать непосредственного программирования при создании информационных систем. Собственно языком программирования в этом проекте является язык C++. Ниже приведем краткое описание программного продукта – ИМС «Семейная медицина».

ИМС содержит информацию, необходимую для работы врача общей практики. В ней изложены экономические, организационные и правовые аспекты деятельности семейного врача. Представлена также оригинальная педагогическая концепция подготовки и специализации семейного врача академика РАМН А.Ф. Краснова. Даются сведения по генетике, иммунологии, клинической химии, фармакологии, общей и медицинской культуре. Большое внимание уделяется профилактике болезней в семье, процедурам и манипуляциям в практике семейного врача. Приведены исчерпывающие данные по заболеваниям, синдромам и неотложным состояниям, с которыми сталкивается семейный врач в своей повседневной работе. Изложение ведется в соответствии с квалификационной характеристикой врача по специальности «Лечебное дело» со специализацией «Семейный врач», приказами Министерства здравоохранения и медицинской промышленности и другими нормативно-правовыми документами РФ согласно утвержденным программам.

Отметим также оригинальный графический интерфейс программы, подробные оглавление и предметный указатель, быстрое действие, масштабирование изображения при печати, использование встроенных алгоритмов сжатия графических образов и, наконец, предельную легкость в освоении и использовании данного продукта.

Среди особенностей технической реализации ИМС по семейной медицине выделим еще независимость от локализации операционной системы компьютера, т.е. система будет выводить информацию из базы данных на дисплей или на принтер на русском (или на другом) языке, а также любое шрифтовое оформление независимо от того, поддерживает операционная система на данном компьютере русский или любой другой язык или не поддерживает, установлены на нем необходимые шрифты или нет.

Предложенная нами информационная медицинская система награждена дипломом победителя конкурса экспонатов «Сибирские Афины» на межрегиональной выставке-ярмарке «Медицина. Здравоохранение. Фармацевтика» (Томск, 1998 г.).

Объем статьи, к сожалению, не позволяет дать развернутую картину вышеприведенной модели данных для ООБД. Что касается собственно данных, то они представляют собой по сути полиморфные объекты. Такое представление, учитывая мультимедийные воз-

возможности современных компьютеров и их применение в обучении, по-видимому, является необходимым. Для повышения производительности системы в целом структурный вид данных объектов на носителе практически совпадает с их образом в оперативной памяти. Можно показать, что описанная выше модель данных включает и реляционную модель данных, и

представление данных в линейном виде, подобном представлению материала в печатных изданиях (книги, статьи, методические разработки и т.п.).

В заключение отметим, что достаточно подробный анализ состояния исследований в области объектно-ориентированных баз данных и перспективы развития этого направления приведены в обзоре [8].

ЛИТЕРАТУРА

1. Корзилов С.В., Пеккер Я.С. Особенности структуризации баз медико-биологических знаний // Сборник трудов научной конференции «Медико-биологические аспекты нейрогуморальной регуляции», посвященной 35-летию ЦНИЛ. Томск, 1997. С. 355.
2. Корзилов С.В., Пеккер Я.С. Особенности компьютерной реализации информационных систем для баз данных по медико-биологическим знаниям. Информационная медицинская система для подготовки врачей общей практики // Сборник трудов 5-й международной научно-методической конференции «Новые информационные технологии в университетском образовании». Новосибирск, 1998. С. 49.
3. Арефьев О. JASMINE: объектно-ориентированная мультимедийная СУБД // Материалы технической конференции «Корпоративные базы данных». М., 1997.
4. Atkinson Malcolm, Bansilhon Francois, DeWitt David, Dittrich Klaus, Maier David, Zdonik Stanley. The Object-Oriented Database System Manifesto // 1st Int. Conf. Deductive and Object-Oriented Databases. Dec. 4-6, 1989. Kyoto, Japan, 1989.
5. Won Kim. Object-Oriented Databases: Definition and Research Directions // IEEE Trans. Data and Knowledge Eng. 2. 1990. № 3. P. 327-341.
6. Beer Catriel. A Formal Approach to Object-Oriented Databases // Data and Knowledge Eng. 1990. № 5. P. 353-382.
7. Корзилов С.В., Пеккер Я.С. Структуризация баз медико-биологических знаний. Информационная медицинская система для врачей общей практики // Сборник трудов межрегиональной научно-практической конференции «Современные технологии в медицине». Томск, 1998 (в печати).
8. Кузнецов С. Новые возможности и тенденции развития средств управления базами данных и разработки информационных систем // Материалы технической конференции «Корпоративные базы данных». М., 1997 г.

Статья представлена кафедрой биологической и медицинской кибернетики Сибирского государственного медицинского университета, поступила в рабочую научную редакционную группу «Проблемы компьютеризации» 20 ноября 1999 г.