

УДК 519.712.2

С.Н. Щеглов

СОВРЕМЕННЫЕ ТЕХНОЛОГИИ ПОСТРОЕНИЯ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ*

Введение. Особое внимание в современных интеллектуальных информационных системах уделяется различным областям человеческой деятельности, где накоплено много данных. Но данные, сами по себе, не имеют практической ценности, поскольку они нуждаются в серьезной обработке. Пользователи прикладных систем в настоящее время предъявляют серьезные требования к результатам исследования данных, которые не могут быть получены чисто статистической обработкой. Их интересуют заключения, помогающие в изучении комплексных задач и в принятии важных решений [1]. Интеллектуальные системы, такие как экспертные системы, системы обработки естественного языка и нейронные вычислительные системы увеличивают производительность и облегчают выполнение сложных задач. Они также обеспечивают поддержку, когда информационный поток неполный или «нечеткий» [2]. Интеллектуальные системы могут использоваться индивидуально, но во многих случаях они интегрированы между собой и с другими информационными системами. Главной характеристикой информационных систем поддержки решений (ИСПР) является наличие в них, по крайней мере, одной модели. Основной идеей является осуществление анализа системы на модели реальной действительности.

Общая архитектура информационно-аналитической системы. Особое внимание в современных интеллектуальных информационных системах уделяется разработке и использованию новых гибридных технологий позволяющих эффективно работать с нечеткой или неполной входной информацией в процессе выработки и принятия решений по различным текущим вопросам. К таким технологиям можно отнести нечеткие генетические и эволюционные алгоритмы, нейросетевые модели и алгоритмы, методы интеллектуального извлечения данных, методы кластерного анализа данных.

Современный уровень развития аппаратных и программных средств с некоторой пор сделал возможным повсеместное введение баз данных оперативной информации на разных уровнях управления. В процессе своей деятельности промышленные предприятия, корпорации, ведомственные структуры, органы государственной власти и управления накопили большие объемы данных. Они хранят в себе большие потенциальные возможности по извлечению полезной аналитической информации, на основе которой можно выявлять скрытые тенденции, строить стратегию развития, находить новые решения.

Полная структура информационно-аналитической системы [3, 4], построенной на основе хранилища данных, показана на рис. 1. В конкретных реализациях отдельные компоненты этой схемы могут отсутствовать.

Данная архитектура является общепринятым пониманием места и роли различных технологий обработки данных.

* Работа выполнена при поддержке: РФФИ (гранты № 08-01-00473, № 06-01-00272), РНП 2.1.2.3193, РНП 2.1.2.2238, г/б № Т.1.04.01.

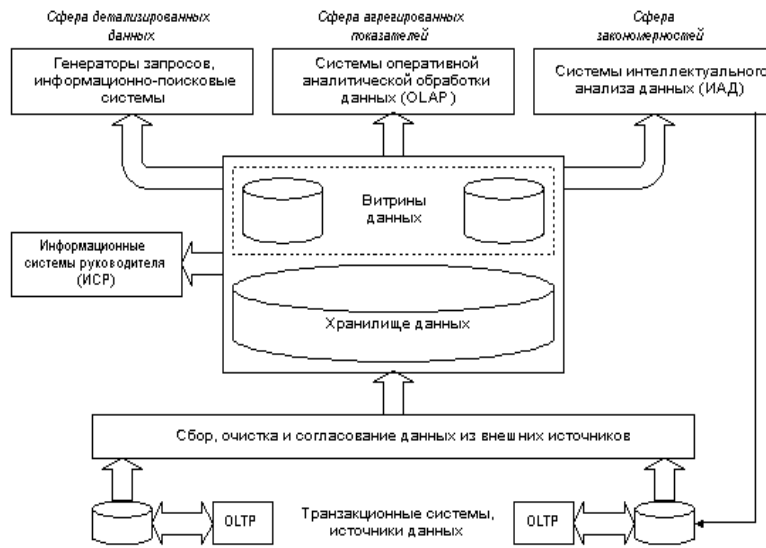


Рис. 1. Общая архитектура информационно-аналитической системы (ИАС)

Информационно-аналитические системы, создаваемые для использования лицами, принимающими решения, являются простыми в применении, но ограниченными по функциональным возможностям. Такие статические системы называются Информационными системами руководителя (ИСР), или Executive Information Systems (EIS). Они содержат в себе predetermined множества запросов и, будучи достаточными для повседневного обзора, неспособны ответить на все вопросы к имеющимся данным, которые могут возникнуть при принятии решений. Время ожидания в данном случае достаточно велико, что не всегда приемлемо. Таким образом, внешняя простота статических СППР приводит к потере функциональной гибкости.

Динамические СППР (DSS) ориентированы на обработку нерегламентированных запросов аналитиков к данным [4]. Работа аналитиков с этими системами заключается в интерактивной последовательности формирования запросов и изучения их результатов.

Динамические СППР могут действовать не только в области оперативной аналитической обработки (OLAP), но и осуществлять поддержку принятия управленческих решений на основе накопленных данных. Такой подход может использоваться в следующих областях:

1. Область детализированных данных. Это область действия большинства систем, нацеленных на поиск информации. В большинстве случаев реляционные СУБД достаточно хорошо справляются с возникающими здесь задачами. Общеизвестным стандартом языка манипулирования реляционными данными является SQL. Информационно-поисковые системы, обеспечивающие интерфейс конечного пользователя в задачах поиска детализированной информации, могут использоваться в качестве надстроек, как над отдельными базами данных транзакционных систем, так и над общим хранилищем данных.

2. Область агрегированных показателей. Комплексный взгляд на собранную в хранилище данных информацию, ее обобщение, агрегация, и многомерный анализ являются задачами систем оперативной аналитической обработки данных (OLAP). В этой области используется ориентированный подход на специальные

многомерные СУБД, либо применяются реляционные технологии. Во втором случае заранее агрегированные данные могут собираться в БД звездообразного вида. Агрегация информации производится одновременно в процессе сканирования детализированных таблиц реляционной БД.

3. Область закономерностей. Интеллектуальная обработка производится методами интеллектуального анализа данных (ИАД, Data Mining), главными задачами которых являются поиск функциональных и логических закономерностей в накопленной информации, построение моделей и правил, которые объясняют найденные аномалии и/или прогнозируют развитие некоторых процессов.

В настоящее время оформился ряд новых концепций хранения и анализа данных [4-6]:

1. Хранилища данных, или Склады данных (Data Warehouse).
2. Оперативная аналитическая обработка (On-Line Analytical Processing, OLAP).
3. Интеллектуальный анализ данных – ИАД (Data Mining).

Технологии OLAP тесно связаны с технологиями построения Data Warehouse и методами интеллектуальной обработки – Data Mining. Поэтому наилучшим вариантом является комплексный подход к их внедрению.

Концепция хранилищ данных (Data Warehouse). Для того чтобы существующие хранилища данных способствовали принятию управленческих решений, информация должна быть представлена аналитику в нужной форме, то есть он должен иметь развитые инструменты доступа к данным хранилища и их обработки.

В основе концепции Хранилищ Данных лежат две основополагающие идеи:

1. Интеграция ранее разъединенных детализированных данных (исторические архивы, данные из традиционных СОД, данные из внешних источников) в едином Хранилище Данных, их согласование и возможно агрегация.

2. Разделение наборов данных используемых для операционной обработки и данных используемых для решения задач анализа.

Рассмотрим структурную схему СППР на основе двухуровневого хранилища данных. Двухуровневое хранилище данных (рис. 2) строится централизованно для предоставления информации в рамках компании.

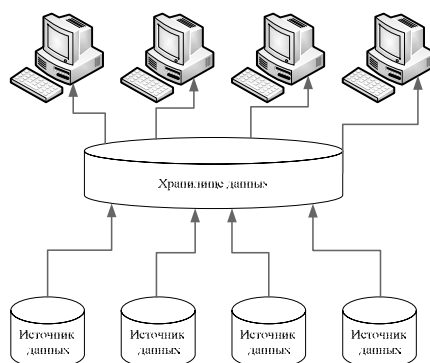


Рис. 2. Двухуровневое хранилище данных

Для поддержки такой архитектуры необходима выделенная команда профессионалов в области хранилищ данных. Это означает, что вся организация должна согласовать все определения и процессы преобразования данных.

Преимущества:

- ◆ Данные хранятся в единственном экземпляре.
- ◆ Минимальные затраты на хранение данных.
- ◆ Отсутствуют проблемы, связанные с синхронизацией нескольких копий данных.
- ◆ Данные консолидируются на уровне предприятия, что позволяет иметь единую картину бизнеса.

Недостатки:

- ◆ Данные не структурируются для поддержки потребностей отдельных пользователей или групп пользователей.
- ◆ Возможны проблемы с производительностью системы.
- ◆ Возможны трудности с разграничением прав пользователей на доступ к данным.

Кроме единого справочника метаданных, средств выгрузки, агрегации и согласования данных, концепция Хранилищ Данных подразумевает: **интегрированность, неизменчивость, поддержку хронологии и согласованность данных**. И если, два первых свойства (*интегрированность и неизменчивость*) влияют на режимы анализа данных, то последние два (*поддержка хронологии и согласованность*) существенно сужают список решаемых аналитических задач.

Без поддержки хронологии (наличия исторических данных) нельзя говорить о решении задач прогнозирования и анализа тенденций. Но наиболее критичными и болезненными, оказываются вопросы, связанные с согласованием данных.

Основным требованием аналитика, является не столько оперативность, сколько достоверность ответа. Но достоверность, в конечном счете, и определяется согласованностью. Пока не проведена работа по взаимному согласованию значений данных из различных источников, сложно говорить об их достоверности.

Концепция Витрин Данных (Data Mart). Концепция Витрин Данных имеет ряд несомненных достоинств:

- ◆ Аналитики видят и работают только с теми данными, которые им реально нужны.
- ◆ Целевая БД Витрины Данных, максимально приближена к конечному пользователю.
- ◆ Витрины Данных обычно содержат тематические подмножества заранее агрегированных данных, их проще проектировать и настраивать.
- ◆ Для реализации Витрин Данных не требуются высоко мощная вычислительная техника.

Но, концепция Витрин Данных имеет и очень серьезные недостатки. По существу, здесь предполагается реализация территориально распределённой информационной системы с мало контролируемой избыточностью, но не предлагается способов, как обеспечить целостность и непротиворечивость хранимых в ней данных.

Возможная структурная схема СППР с использованием независимых витрин данных представлена на рис. 3. Независимые витрины данных часто появляются в крупных организациях с большим количеством независимых подразделений, имеющих свои собственные отделы информационных технологий.

В настоящее время применяется объединенная концепция Хранилищ Данных и Витрин Данных, в которой предлагается использовать Хранилище Данных в качестве единого интегрированного источника данных для Витрин Данных.

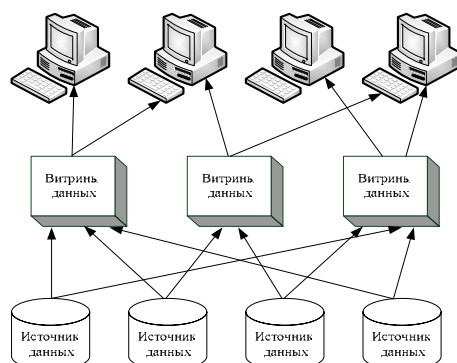


Рис. 3. Независимые витрины данных

В таком случае возможно следующее многоуровневое решение:

- ◆ *первый уровень* – общекорпоративная БД на основе РСУБД с нормализованной или слабо нормализованной схемой (детализированные данные);
- ◆ *второй уровень* – БД уровня подразделения (или конечного пользователя), реализуемые на основе МСУБД (агрегированные данные);
- ◆ *третий уровень* – рабочие места конечных пользователей, на которых непосредственно установлен аналитический инструментарий;
- ◆ постепенно становится стандартом де-факто, позволяя наиболее полно реализовать и использовать достоинства каждого из подходов;
- ◆ компактное хранение детализированных данных и поддержка очень больших БД, обеспечиваемые реляционными СУБД;
- ◆ простота настройки и хорошие времена отклика, при работе с агрегированными данными, обеспечиваемые многомерными СУБД.

Реляционная форма представления данных, используемая в центральной общекорпоративной БД, обеспечивает наиболее компактный способ хранения данных. Современные реляционные СУБД уже умеют работать даже с терабайтными базами. И, хотя такая центральная система, обычно, не сможет обеспечить оперативного режима обработки аналитических запросов, при использовании новых способов индексации и хранения данных, а так же частичной денормализации таблиц, время обработки заранее регламентированных запросов (а в качестве таких, можно рассматривать и регламентированные процедуры выгрузки данных в многомерные БД) оказывается вполне приемлемым.

В свою очередь, использование многомерных СУБД в узлах нижнего уровня обеспечивает минимальные времена обработки и ответа на нерегламентированные запросы пользователя. Кроме того, в некоторых многомерных СУБД имеется возможность хранить данные как на постоянной основе (непосредственно в многомерной БД), так и динамически (на время сеанса) загрузить данные из реляционных БД (на основе регламентированных запросов).

Таким образом, имеется возможность хранить на постоянной основе, только те данные, которые наиболее часто запрашиваются в данном узле. Для всех остальных, хранятся только описания их структуры и программы их выгрузки из центральной БД. И хотя, при первичном обращении к таким виртуальным данным, время отклика может оказаться достаточно продолжительным, такое решение обеспечивает высокую гибкость и требует более дешевых аппаратных средств

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Городецкий В.И., Самойлов В.В., Малов А.О.* Современное состояние технологий извлечения знаний из баз и хранилищ данных // *Новости искусственного интеллекта*, №3, 2002. – С. 3-13.
2. *Башмаков А.И., Башмаков И.А.* Интеллектуальные информационные технологии. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2005.
3. *Гаврилова Т.А., Хорошевский В.Ф.* Базы знаний интеллектуальных систем. – СПб.: Питер, 2000.
4. *Christopher J. Date.* The Database Relational Model: A Retrospective Review and Analysis: A Historical Account and Assessment of E. F. Codd's Contribution to the Field of Database Technology. Addison Wesley Longman, 2000.
5. *Спулли Э.* "Корпоративные хранилища данных. Планирование, разработка и реализация. Т.1". Издательство: Вильямс 2001. ISBN 5-8459-0191-X.
6. *Devlin B.,* "Data warehouse: from architecture to implementation". Addison Wesley Longman, Inc. 1997. ISBN 0-201-96425-2.

УДК 321.3

П.В. Сороколетов

**ПРИНЦИПЫ И НЕЧЕТКИЕ АЛГОРИТМЫ АНАЛИЗА МОДЕЛЕЙ
ПРИНЯТИЯ РЕШЕНИЙ***

Введение. В настоящее время современные интеллектуальные и производственные системы характеризуются отсутствием всей или частичной необходимой информации, что делает некорректным их эффективное функционирование. Это приводит к невозможности и затруднению практического применения существующих систем и методов поддержки принятия решений [1, 2].

При решении комплексных задач энергетики, проектирования, системного анализа, теории игр, где существует понятие «проклятия размерности» анализируемой информации, ограничений и граничных условий необходимо сформулировать постановку и условия задачи, а также понятие решения. Тогда исходная задача требует решения, если имеются некие условия ее получения. Получение решения осуществляется с помощью комплекса методов и алгоритмов. Если существует некий метод, позволяющий на множестве параметров сопоставлять некоторые величины между собой, которые приводят к оптимальному использованию ресурсов, то говорят, что получено оптимальное решение. Оно может представлять собой некое множество значений, функций, правил или методов, приводящих к преобразованию исходных величин и удовлетворению условий задачи [3, 4].

Использование нечетких исходных данных является одним из способов расширения сферы применения формальных методов, в частности, теории ожидаемой полезности, однако этот путь предполагает наличие фактически той же информации, только представленной в другом виде. В работе рассматриваются возможные подходы к анализу решений на основе не только нечеткой, но и ограниченной информации, недостаточной для расчета оценок ожидаемой полезности.

Принятием решений считают множество альтернатив в условиях определенности, позволяющих получать, однозначные, непротиворечивые, корректные решения

* Работа выполнена при поддержке: РФФИ (гранты № 08-01-00473, № 06-01-00272), РНП 2.1.2.3193, РНП 2.1.2.2238, г/б № Т.1.04.01.