

Список литературы

1. РД РФ «Безопасность информационных технологий. Критерии оценки безопасности информационных технологий». — Гостехкомиссия, 2003.
2. РД РФ «Безопасность информационных технологий. Типовые модели угроз безопасности информационных технологий автоматизированных систем критического применения» — 2-я ред. — Гостехкомиссия, 2003.
3. ФЗ РФ «О техническом регулировании». № 184 от 27.12.2002 г.
4. ФЗ РФ «О коммерческой тайне» № 98 от 29.07.2004 г. (в ред. Закона РФ № 9-ФЗ от 02.02.2006 г., с изм., внесенными Федеральным законом № 231-ФЗ от 18.12.2006 г.).
5. ФЗ РФ «О государственной тайне» № 5 485-1 от 21.07.1993 г. (в ред. Федеральных законов № 131-ФЗ от 06.10.1997 г., № 86-ФЗ от 30.06.2003 г., № 153-ФЗ от 11.11.2003 г., № 58-ФЗ от 29.06.2004 г., № 122-ФЗ от 22.08.2004 г., с изм., внесенными Постановлением Конституционного Суда РФ № 8-П от 27.03.1996 г., определениями Конституционного Суда РФ № 293-О от 10.11.2002 г., № 314-О от 10.11.2002 г.).

И. К. Фомина,

канд. техн. наук, доц., СПГУВК

**РЕШЕНИЕ ПРОБЛЕМ, СВЯЗАННЫХ СО СЛОЖНОЙ СИСТЕМОТЕХНИЧЕСКОЙ
СТРУКТУРОЙ ПРЕДМЕТНЫХ ОБЛАСТЕЙ ПРИ ИНТЕЛЛЕКТУАЛЬНОМ
АНАЛИЗЕ ДАННЫХ**

**ADDRESSING THE CHALLENGES ASSOCIATED WITH COMPLEX SYSTEM
OF TECHNICAL SUBJECT AREAS, WITH A SMART DATA ANALYSIS**

В данной статье рассмотрены вопросы структуризации предметной области, подготовки и представления формальных данных в разрезе прогнозных, экспертно-ориентированных, символично-статистических моделей. Анализ особенностей современных технологий интеллектуального анализа данных. Раскрыта тематика проблем многофакторности и сложности предметных областей, а также применение алгоритмов и моделей data mining.

This article examines questions of the structure organization of subject area, preparation and presentation of formal data in forecast, expert-oriented, symbolic-statistical models. Analysis of peculiarities of modern technologies of intellectual data analysis. The themes of problems of many factors and complex nature of subject areas are revealed, and also the application of algorithms and models of data mining.

Ключевые слова: специальное программное обеспечение, информационная поддержка, предметная область, анализ данных, качество системы интеллектуального анализа данных.

Key words: special software, Information Support, subject area, data analysis, quality of intellectual analysis data.

СЛОЖНОСТЬ и многофакторность предметной области [1, 3], а также традиционность современных пакетов, опыт и профессиональное искусство эксперта не в состоянии обеспечить эффек-

тивности результатов в широких масштабах корпоративного или государственного применения, больших объемах разнородной информации, разном уровне подготовки пользователей информации и специалистов, эксплуати-

рующих эти системы. Необходимо применять специальное программное обеспечение (специальные автоматизированные информационные системы), направленное на эффективность и результативность реализации именно таких задач.

Анализ существующих в настоящее время реализаций аналитического инструментария показывает, что в большинстве своем соответствующие средства и системы строятся, в основном, как пакеты статистической обработки. Такие системы либо вообще не снабжены аппаратом экспертной аналитики, либо не дают необходимой строгости решения задач, особенно в разрезе прогнозирования и классификации данных. Применение таких систем не достигает требуемой эффективности результатов и не обеспечивает полноты совокупности процедур обработки, анализа, прогнозирования и их увязки в сквозной единый высокоавтоматизированный процесс. Применение этих средств и систем ограничивается разработкой отдельных составляющих конкретных классов систем в разрезе выбранной отрасли знаний.

Сформулируем основные концепции информационной поддержки. Конструктивный подход [5] позволяет выдвинуть ряд требований к программному обеспечению, реализация которых позволяет характеризовать инструментарий как функционально полную комплексную экспертно-аналитическую автоматизированную систему, обладающую свойством адаптивности к предметным областям и средствам представления знаний.

С точки зрения пользователя специального программного обеспечения (СПО) важными моментами являются:

- 1) возможность формального представления данных и знаний в рамках единого информационного фонда;
- 2) автоматизация накопления, изменения, хранения и представления информации об анализируемых предметных областях;
- 3) возможность представления иерархических структур;
- 4) возможность проведения локального анализа в рамках указанных уровней представления с произвольной степенью детализации;

5) удобство эксплуатации с позиций эргономики и технической эстетики.

Структура технологии автоматизированного анализа включает:

2) основу технологии автоматизированного анализа должна составлять открытая совокупность взаимосвязанных процедурных модулей, обеспечивающая необходимый уровень сбора и хранения данных, формализации, обоснования, документирования экспертных решений;

3) методика эксплуатации должна быть ориентирована на совместную обработку и системный анализ разноаспектных специфицированных данных и знаний;

4) средства анализа должны включать обеспечение синтаксического и семантического анализа и проверку полноты моделей представления в рамках определений эксперта;

5) оперативное представление необходимой информации о свойствах и элементах анализируемых объектов (систем данных), подготовка отчетной документации;

6) методические материалы, ориентированные на более эффективное использование СПО;

Выделяют следующие базовые функции СПО с учетом современного уровня развития информационных технологий:

1) обеспечение для пользователя возможности постановки задачи для компьютера путем сообщения только ее условий, без задания программы решения;

2) возможность для пользователя самостоятельно формировать операционную среду решения задачи на базе терминов и понятий из предметной области;

3) обеспечение для пользователя естественных форм представления информации, а также выбор приемлемых для него способов организации диалога;

4) возможность изменения по желанию пользователя структуры диалога (диалог с архитектурой меню, обмен произвольными сообщениями);

5) обеспечение работы в условиях ошибок, допускаемых пользователем в сообщениях, с использованием средств, объясняющих пользователю его ошибки и непонятные места в ходе решения задачи.

6) объектно-ориентированная технология представления информации. Манипуляции алгоритмами, видами документации (диаграмм, графиков), математического аппарата и прочими объектами производятся путем их классификации по каталогам (библиотечный принцип), с возможностью дальнейшего расширения или коррекцией отдельных объектов.

СПО интеллектуального анализа данных — интегральная экспертно-аналитическая система, позволяющая проводить всесторонний статистический анализ разнородных данных, полученных и накопленных на основе эмпирических фактов исследуемой предметной области, оснащенная инструментарием классификации, диагностики и распознавания образов, интерпретации закономерностей и прогнозирования.

СПО представляет инструментальную многокомпонентную систему, настраиваемую на функциональные требования заказчика, и комплектуемую отдельными модулями. В состав (архитектуру) современного СПО могут входить следующие подсистемы и модули.

Подсистема формирования знаний. Система поиска скрытых закономерностей и поддержки принятия решений в предметных областях со сложной системной организацией, ориентированной на решение задач корпоративного типа с неоднородными, семантически различными структурами и большими объемами данных в режиме распределенных систем. Расширение и углубление возможностей аналитического инструментария системы основывается на новейших научных методах отечественной и зарубежной науки и базируется на широком классе методов, известных под классификацией Data Mining (автоматического поиска закономерностей в данных). Методы Data Mining относятся к алгоритмическому решению задач классификации, диагностики, интерпретации закономерностей, распознавания образов.

Набор модулей, образующих основу подсистемы статистического анализа. Данная подсистема, как правило, предоставляет в интерактивном режиме набор отдельных модулей профессионально-ориентированных на

выполнение унифицированных алгоритмов и методик статистической обработки:

— унифицированный модуль управления данными (основные функции: настройка на форматы данных, конвертирование, редактирование, модуль снабжен полным инструментарием реляционной алгебры для манипулирования данными «плоских» таблиц с адаптивным аппаратом настройки на различные форматы хранения);

— унифицированный модуль управления данными статистического анализа и представления результатов (работа с графиками, таблицами, документами для интернет-публикаций);

— расширенный модуль генерации отчетов (инструментарий настройки на виды документации, манипулирующий графиками и таблицами);

— расширенный модуль обработки статистических данных и построения графиков распределения (модуль снабжен большим набором алгоритмов из аппарата теории математической статистики.

— расширенный модуль бизнес-публикаций. Модуль предназначен для создания высококачественной математической и технической отчетности. Большая встроенная библиотека 2D- и 3D-диаграмм, графиков, презентаций. Различные форматы представления и механизмы интеграции документов с внешними пакетами.

Основными требованиями, предъявляемыми к качеству системы интеллектуального анализа данных [2, 4], являются следующие.

— *Интеграция* (использование единых функциональных спецификаций и интерфейсов, общих информационных и системно-технических ресурсов). Использование в производстве и технологиях международных стандартов: CMM, разработанный (Software Engineering Institute), и стандарты серии ISO, принятые Международным комитетом по стандартизации (International Organization for Standardization).

— *Открытость* (архитектурная открытость программного обеспечения для функционального расширения и модернизации). Модульный принцип организации изначально предполагает повышенную открытость,

взаимозаменяемость программ, механизм подключения новых компонент. Архитектурная открытость системы подтверждается, как правило, условиями поставки отдельных модулей, написанных и разработанных разными фирмами с применением различного инструментария разработки.

— *Масштабируемость* (увеличение количества рабочих мест и количества задач без изменения программного обеспечения). СПО может поставляться как для работы на отдельном рабочем месте, так и в локальной сети и в сети Интернет. Архитектура поставляемого СПО тождественна количеству рабочих мест и количеству задач.

— *Адаптивность* (настраиваемость). СПО обладает свойствами совместимости и адаптивности:

— совместимости — работает должным образом не только автономно, в базовой комплектации, расширенной комплектации, но и как составная часть компьютерного обеспечения (различные платформы, операционные среды, утилиты и пр.);

— адаптивности — допускает быструю модификацию с целью приспособления к изменяющимся условиям функционирования. Адаптивность в значительной степени отражается в конструкции СПО, а также в квалификации составления и полноты снабжения документацией. Должен быть предусмотрен способ передачи пользователям системы доработанных или измененных программных модулей без общего обновления всех модулей.

— *Надежность и безопасность* (обеспечение бесперебойной работы, получение достоверных результатов и защита от несанкционированных действий). Надежность, универсальность и защищенность — неотъемлемые свойства поставляемого СПО. Особая политика тестирования новых версий, долгосрочная эксплуатация модулей на различных предприятиях, контроль и техническая поддержка поставщика, развитие новых версий с учетом эволюции информационных технологий позволяют утверждать, что:

— в процессе эксплуатации СПО при всех равных условиях обеспечивается полная повторяемость результатов и правильность работы при любых допустимых вариантах ис-

ходных данных. Выдаваемые числовые данные имеют допустимые отклонения от аналогичных результатов, полученных с помощью «идеальных математических зависимостей»;

— в СПО должны быть предусмотрены специальные средства защиты от ввода неправильных и несанкционированных данных по управлению, обеспечивающие целостность системы. Предусмотрен механизм защиты данных путем ее шифрования при обмене по информационным каналам (клиент-сервер);

— СПО должна сохранять работоспособность при возникновении прогнозируемых сбоев. Это качество особенно важно для программ, предназначенных для решения задач в режиме реального времени и манипулирующего огромными массивами информации, потери которых из-за сбоев могут привести к катастрофическим последствиям;

— должна быть предусмотрена возможность проверки таких свойств программы, как правильность и универсальность. Имеются в виду такие неформальные приемы, как прогоны с остановами в контрольных точках, промежуточные результаты, обсуждения результатов с заинтересованными пользователями и др.

— *Экономичность* (обеспечение важнейших функций при ограничениях на трудоемкость и стоимость создания, эксплуатации, сопровождения и развития). Существование поставляемого продукта на рынке программных услуг на протяжении многих лет, большое количество инсталляций, отработанные технологии разработки и поставки СПО позволят пользователю снизить затраты на внедрение, эксплуатацию, создание новых версий и проводить политику экономичности.

— *Эргономика и техническая эстетика*. Требования к эргономике и технической эстетике основаны на принципах монотонного, постоянного использования системы на рабочем месте пользователя с большой вероятностью психологической усталости. Такие принципы эксплуатации предполагают использование в СПО «не возбуждающих» и «не агрессивных» цветовых схем, а также минимизацию и унификацию одновременно представляемой информации.

Программное обеспечение поддерживает концепцию «дружественного пользовательского интерфейса». В основу реализации такого интерфейса была положена концепция «программной оболочки». Оболочка СПО ориентирована на объектную архитектуру приложений, набор возможностей Drag and Drop (Перетащить и Оставить), длинные имена файлов, настройку панелей управления и др.

СПО — инструментальная объектно-ориентированная программная система. Это означает, что в СПО реализовано рабочее окружение компьютера, моделирующее действительное физическое рабочее окружение в виде виртуального рабочего стола с виртуальными инструментами. Инструментарий предлагает много различных способов выполнения одной и той же задачи. Это многообразие позволяет пользователю самому накладывать некоторые ограничения в зависимости от того, как он хочет работать.

Windows-ориентированный интерфейс основывается на нескольких разновидностях окон: окна программы, вторичные окна, вспомогательные окна, окна диалога и пр. С помощью мыши и кнопок управления можно изменять размер и расположение окон.

СПО является многозадачной системой, и под ее управлением могут одновременно выполняться несколько программ. На экране могут одновременно находиться окна

нескольких программ, поэтому интерфейс системы — многооконный, но в каждый момент времени только одно окно является активным, а остальные — неактивны. Активное окно всегда располагается поверх других.

Для запроса значений тех или иных параметров, а также для вывода информационных сообщений и предупреждений, программы СПО выводят на экран соответствующие окна диалогов. Вид диалога может быть как простым — с некоторым сообщением и одной или двумя кнопками для ответа, так и сложным — с десятками полей различных типов. Тем не менее основные правила по использованию диалоговых запросов такие же, как у всех Windows-совместимых программ.

Особенности организации интерфейса:

- интеграцию вновь включаемых компонент целевых подсистем с существующими модулями и интерфейсами взаимодействия;
- предоставление пользователю оперативной информации о процессе обработки запросов к хранилищу данных;
- предоставление пользователю информации о статусе информационных ресурсов, их полноте и достоверности;
- возможность формирования регламентных и нерегламентированных запросов в терминах предметной области и получения информации с необходимым уровнем обобщения или детализации;
- многоязыковый интерфейс СПО.

Список литературы

1. Гиг Дж. ван. Прикладная общая теория систем. — М.: Мир, 1981.
2. Дюк В. А., Эмануэль В. Л. Информационные технологии в медико-биологических исследованиях. — СПб.: Питер, 2003. — 525 с.
3. Киселев М., Соломатин Е. Средства добычи знаний в бизнесе и финансах // Открытые системы. — 1997. — № 4. — С. 41–44.
4. Кречетов Н. Продукты для интеллектуального анализа данных // Рынок программных средств. — 1997. — № 14–15. — С. 32–39.
5. Математические методы для анализа последовательностей ДНК: пер. с англ. / под ред. М. С. Уотермена. — М.: Мир, 1999. — 349 с.