

Министерство образования Республики Беларусь

Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.65

Нестерчук Илья Владимирович

АЛГОРИТМЫ АНАЛИЗА ДАННЫХ В СИСТЕМЕ ПОДДЕРЖКИ ПРИНЯТИЯ
РЕШЕНИЙ

АВТОРЕФЕРАТ

на соискание степени магистра технических наук

по специальности 1-40 80 02 «Системный анализ, управление
и обработка информации»

Научный руководитель
Севернёв Александр Михайлович
доцент, кандидат технических наук

Минск 2020

ВВЕДЕНИЕ

Во всем мире организации накапливают или уже накопили в процессе своей административно-хозяйственной деятельности большие объемы данных, в том числе и в электронном виде. Эти коллекции данных хранят в себе большие потенциальные возможности по извлечению новой аналитической информации, на основе которой можно и необходимо строить стратегию организации, выявлять тенденции развития рынка, находить новые решения, обуславливающие успешное развитие в условиях конкурентной борьбы. Для некоторых организаций такой анализ является неотъемлемой частью их повседневной деятельности, другие начинают активно приступать к такому анализу.

Хранилище данных (*data warehouse*) — накопитель информации из других систем, на основе которых строятся процессы принятия решений и анализа данных. Целью построения хранилищ данных является своевременное предоставление необходимой информации только тем сотрудникам предприятия, которые действительно в ней нуждаются.

Хранилище данных — это набор данных, предназначенный для поддержки принятия управленческих решений и характеризующийся предметной ориентацией, интеграцией, поддержкой хронологии и неизменяемостью.

Системы поддержки принятия решений (СППР) — это системы, обладающие средствами ввода, хранения и анализа данных, относящихся к определённой предметной области, с целью поиска решений.

Основная задача СППР — предоставление аналитикам инструмента для выполнения анализа данных. СППР не гарантирует правильных решений, а только предоставляет аналитику данные в виде таблиц, отчётов, графиков и т.п. для изучения и анализа.

Трудно переоценить актуальность построения системы принятия решений (СППР) для крупных банков, предприятий нефтяной, газовой отрасли, машиностроения, металлургии, торговле и т.д. Конечно, СППР, основанная на технологии хранилищ данных или витрин данных (а большинство СППР базируются именно на них), будет быстрее окупаться на крупных предприятий или организациях с развитой сетью филиалов и разнообразной деятельностью, внедряющих корпоративные информационные системы.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Задачи исследования. Разработка предложений по усовершенствованию сбора, обработки, обобщения и хранения информации, которые позволят оптимизировать анализ данных и приведут к эффективному принятию решений.

Объект исследования. Системы хранения, обработки и извлечения информации из баз данных и хранилищ данных.

Предмет исследования. Модели и алгоритмы систем поддержки принятия решений на основе хранилищ данных, алгоритмы анализа данных.

Цель и задачи исследований. Целью является повышение эффективности обработки, хранения и анализа больших объемов информации за счет использования технологии хранилищ данных, а также исследование и применение методов интеллектуального анализа данных в СППР.

Структура и объем диссертации. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы и приложений. Работа содержит 83 страницы основного текста, 16 формул, 25 рисунков и 2 приложения. Список литературы содержит 103 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В главе 1 «Анализ состояния проблемы и постановка задачи» проанализирована существующая система управления объекта автоматизации, ее недостатки в рамках действующего процесса. Также показано, что тогда, когда целью является выполнение задач, связанных со сбором, хранением и поиском информации, эффективным средством является применение традиционных реляционных баз данных. В случае же анализа и принятия решения с использованием существующих (исторически накопленных) данных, лучше воспользоваться хранилищем данных с технологией *OLAP*. Сделан вывод о необходимости использования многомерных моделей данных для повышения эффективности обработки информации, а также определен список задач, которые должна решать разработанная система поддержки принятия решений.

В главе 2 «Методы анализа данных, применяемые в СППР, основанных на хранилищах данных» исследуются и решаются задачи интеллектуального анализа данных с целью обнаружения знаний. Для этого используются методы интеллектуального анализа, в число которых входят классификация, поиск ассоциативных правил и др. На примерах возможности обслуживания клиентов

в торговых организациях используются методы классификации (классификационные правила, наивный метод Байеса).

Глава 3 «Проектирование структуры системы поддержки принятия решений» основана на формулировании общих задач и структуры системы, которая состоит из следующих подсистем: подсистема управления метаданными, подсистема извлечения данных, подсистема хранения данных, подсистема информационно-аналитического обслуживания и подсистема администрирования и управления. Также раздел включает в себя описание математического, алгоритмического, информационного, технического и системного обеспечения, а также уровни доступа к системе.

В заключительной главе 4 «Программная реализация системы» представлена структура системы и принципы, лежащие в ее основе, перечень применяемых в проекте технологий; описаны сформированные витрины данных, а также рассмотрена витрина «Главная книга». Также в четвертом разделе рассматриваются методы повышения эффективности обработки данных с целью ускорения времени выполнения запросов.

Диссертация выполнена самостоятельно, проверена в системе «Антиплагиат». Процент оригинальности соответствует норме, установленной кафедрой. Цитирования обозначены ссылками на публикации, указанные в «Списке использованных источников».

ЗАКЛЮЧЕНИЕ

При работе над диссертацией были выполнены следующие задачи:

- проанализированы и исследованы существующие способы хранения данных и принятия решений в банковской сфере; для ускорения процесса анализа данных, представимых в многомерном виде, было предложено применение *OLAP*-технологии в СППР, основанных на ХД.

- для интеллектуального анализа данных предложено использование, технологии «*Data mining*» с применением методов классификации, поиска ассоциативных правил и вероятностных методик Байеса;

- решена задача автоматизации закрытия глобального операционного дня двумя способами: с использованием алгоритма покрытия и методом *Naive Bayes*;

- предложено решение задачи об учёте влияний курсов валют на величины процентных ставок вкладов при помощи вероятностных методик Байеса;

– проанализированы исходные базы данных, а именно: комплексный анализ отрасли, выделение основных сущностей и определение гранулярности данных в каждой БД;

– дано формализованное описание многомерной модели данных и математическое представление гиперкуба данных. Разработаны методы повышения производительности ХД с использованием материализованного представления, индексирования данных и разделения таблиц;

– создана абстрактная модель предметной области с учетом проанализированной информации на этапе комплексного анализа отрасли;

– создана концептуальная архитектура управления потоком данных;

– проведено извлечение данных из источников информации путем реализации механизма *ETL*;

– обеспечена бесперебойная работа и поддержание актуальности данных;

– обеспечено ежедневное представление аналитической информации в требуемых разрезах.

Разработка и внедрение продукта позволит:

– исторически анализировать показатели за любой временной период;

– производить динамический анализ, представлять отчетную информацию в виде таблиц и диаграмм;

– производить агрегирование данных по любым разрезам;

– использовать методы прогнозирования и моделирования;

– получать показатели на основе данных, накапливающихся в *OLTP*-системах;

– сравнивать и согласовывать данные, полученные из разных оперативных систем и внешних источников.

В результате бизнес-пользователи (разработчики отчетов, аналитики, потребители отчетов) смогут сократить количество времени и денежных затрат на консолидацию, подготовку и проверку данных, вместо этого сосредоточиться на их анализе, научном исследовании и принятии решений.

СПИСОК ПУБЛИКАЦИЙ АВТОРА

[1–А.] Нестерчук И.В. Алгоритмы анализа данных в системах поддержки принятия решений / И.В. Фёдоров, И.В. Нестерчук // XXVIII Международная научно-практическая конференция «*Advances in science and technology*» – Москва, 2020.