

ЭКОНОМИКА

Ковалев Сергей Викторович

канд. филос. наук, доцент

НОУ ВПО «Московский институт

управления и информатики»

г. Москва

**МОДЕЛИРОВАНИЕ СИСТЕМЫ «БАЗЫ
ЗНАНИЙ–ХРАНИЛИЩА БИЗНЕС-ДАННЫХ»
В СИСТЕМЕ МОНИТОРИНГА ЭФФЕКТИВНОСТИ
КОРПОРАТИВНОГО УПРАВЛЕНИЯ**

***Аннотация:** в данной статье автором рассматриваются вопросы разработки базы знаний и хранилища данных в виде программно-аппаратной системы анализа бизнес-процессов, многомерного иерархического структурирования данных для прогнозирования и предотвращения наступления риск-ситуаций. Представлен математический аппарат моделирования возможных потерь и (или) приобретений в ходе применения предлагаемой системы.*

***Ключевые слова:** разработка базы знаний, хранилища данных, программно-аппаратная система, бизнес-анализ, информационный куб, многомерные иерархические структуры данных.*

Введение.

Для формирования системы корпоративных данных на предприятии всегда существуют одна или несколько учетных систем, в которых фиксируются все полезные данные для бизнеса, например: продажи, поставки, издержки и прочие сведения [1, 2, 3, 4, 5]. Подобные системы, будь то современные и дорогостоящие ERP (Enterprise Resource Planning – Система Управления Ресурсами Предприятия) и CRM (Customer Relationship Management – Система Управления Взаимоотношениями с Клиентами) или более простые и относительно недорогие, служат по большому счету для накопления и изменения полезных данных. Такие системы

называются транзакционными или OLTP-системами (Online Transaction Processing). Десятки, сотни и даже тысячи пользователей могут одновременно вносить изменения в Базу Данных [1, 2, 3, 4, 5]. Подобные системы обладают следующими особенностями:

- структура базы данных состоит из реляционных таблиц и обычно нормализована, а значит содержит относительно большое количество таблиц;
- системы разработаны специально для быстрого ввода и обработки данных, основные операции в системе – ввод, изменение и удаление данных;
- каждая система выполняет свои специфические функции, между собой системы слабо связаны.

Представим себе, что подобные системы работают на предприятии несколько месяцев и даже лет. Объем данных неуклонно растет, фирма очень дорожит многолетними данными, но никакой реальной пользы они не приносят [1, 2, 3, 4, 5]. Рано или поздно появляется потребность в анализе огромного объема накопившейся информации и все без исключения бизнес-аналитики сталкиваются со следующими трудностями:

- анализ больших объемов данных идет долго и сильно нагружает основной вычислительный сервер (группу серверов) с базами данных, тем самым блокируя на некоторое время работу остальных пользователей;
- аналитические отчеты формируются очень долго, это связано с огромными массивами обрабатываемых данных;
- невозможно получить единый и правильный отчет из нескольких учетных систем одновременно.

Для преодоления этих трудностей разрабатывается специальная База Данных, в которую из всех учетных систем предприятия импортируется только нужная для анализа информация. Такие базы данных называются Хранилищами Данных или информационным кубом. Часто информация извлекается не только из учетных систем, но и из «плоских» файлов (Excel, Access,.dbf). Все это называется Источниками Данных. Хранилище Данных (Data Warehouse) – это большая предметно-ори-

ентированная информационная корпоративная база данных с неизменяемой информацией, которая специально разработана и используется для целей анализа и поддержки принятия решений. Хранилище Данных (ХД) может содержать как элементарные, так и агрегированные данные [1, 2, 3, 4, 5].

Особенности Хранилищ Данных по сравнению с OLTP-системами:

- данные в ХД структурируются и хранятся по объектам бизнеса, а не по приложениям, в которых поддерживаются;
- вся информация об объекте собирается из всех источников, консолидируется и хранится в одном месте;
- информация в ХД не изменяется и не удаляется, загрузка данных в хранилище происходит только в установленные моменты времени, строго по регламенту;
- таблицы в хранилище данных сильно денормализованы (содержат большое количество атрибутов), такая структура таблиц повышает быстроту выполнения аналитических запросов.

При этом задача определения всех необходимых систем – источников данных и их синхронизации (согласования данных) решается один раз на этапе проектирования ХД. Далее система в автоматическом режиме поддерживает ведение какого-либо объекта, и аналитику, работающему с этим бизнес-объектом, уже не надо каждый раз заниматься трудоемкими операциями по сбору и сведению воедино требуемой проектной информации для систематического мониторинга [1, 2, 3, 4, 5]. Вместо этого он может сосредоточиться собственно на анализе данных и принятии решений.

По масштабу хранилища данных обычно разделяют на:

- Корпоративное Хранилище Данных, которое охватывает весь бизнес компании. Данные собираются из всех источников, поэтому реализация хранилища требует больших затрат и усилий. Оно может служить источником для витрин данных.

– Витрины Данных (Data Marts), которые охватывают только часть бизнеса компании (продажи, закупки, бухгалтерия), соответственно они служат источниками информации для отдела. Реализация Витрин данных более проста и контролируема, поэтому они могут быть использованы как этап в построении корпоративного ХД.

Обычно такое разделение связано с уменьшением времени отклика на аналитические запросы. Когда хранилище одновременно используют несколько групп пользователей, обрабатываются огромные массивы данных, с которыми СУБД могут не справляться, но иногда это делают для того, чтобы на начальных этапах внедрения хранилища иметь контроль над качеством данных. Невозможно продумать структуру хранилища и консолидировать информацию, взятую сразу из 5–7 источников данных. Для осуществления систематического мониторинга следует постепенно подключать все новые и новые источники данных. Информация может быть представлена в форме базы данных, в форме электронного конструкторского документа или в форме, пригодной для восприятия человеком, – бумажной или экранной [1, 2, 3, 4, 5].

1. Концепция создания базы знаний инновационного развития корпорации.

В последнюю четверть XX века человечество вступило в новую стадию своего развития – стадию построения постиндустриального общества. Экономически развитые страны провозгласили переход к экономике, основанной на знаниях. Как показывает статистика, в промышленно развитых странах мира от 75 до 100% прироста промышленного производства обеспечивается за счёт использования новых знаний и инноваций, а оживление производства также происходит через интенсификацию и активизацию инновационных процессов. Это особенно характерно в современном ускоряющемся мире, динамично переходящем к экономике знаний [1, 2, 3, 4, 5].

В отличие от ведущих экономических держав инновационная активность реального сектора российской экономики крайне низка. По различным оценкам в России используются только от 8 до 10% инновационных идей и проектов, в то время как в Японии – 95%, в США – 62%. Только одно из 500 запатентованных в России

изобретений находит применение в промышленности. На долю России приходится более 10% всех учёных мира, а это огромный потенциал, но сегодня её доля на мировом рынке наукоёмкой продукции не превышает 0,3% от его объёма, хотя по оценке аналитиков она могла бы составлять 10–12% [1, 2, 3, 4, 5].

Перед экономикой Российской Федерации поставлена важная задача – сделать основной составляющей экономики высокотехнологичные отрасли. Для этого необходимо обеспечить инновационную деятельность предприятий, т. е. осуществить нововведения в области техники, технологии, организации труда и управления, основанные на использовании достижений науки и передового опыта. Наша страна обладает огромным научным потенциалом, многочисленными кадрами специалистов, мощным интеллектуально-информационным ресурсом. Таким образом, необходимо подготовить материалы научных исследований для внедрения и оказать поддержку их внедрению [1, 2, 3, 4, 5].

Самой сутью инноватики является работа со знаниями. Поэтому инструменты работы непосредственно со знаниями, с информационными ресурсами становятся определяющими эффективность инновационного процесса. Без решения проблем генерации, обработки и использования знаний и информации инновационный процесс в целом не возможен.

Проблемы подготовки результатов научных исследований – это отсутствие централизованных хранилищ, контроля за качеством и достоверностью информации, систем поиска хранящейся информации, системы выбора и анализа полученной в результате поиска информации. Проблемы внедрения – это отсутствие удобных инструментов для доступа к хранящейся информации предприятиями, систем контроля и учёта инновационной информации, инструментов для управления инновационными проектами на предприятиях, инструментов для взаимодействия между предприятиями в инновационной сфере [1, 2, 3, 4, 5]. Окончательно устоявшихся определений знаний, информации, данных и соответственно баз данных и баз знаний в настоящее время не существует. Однако в настоящем контексте целесообразно дать следующие определения:

– Данные – это полученные эмпирическим путём и зафиксированные факты, дискретно описывающие ситуацию (объект и др.) вне контекста, т. е. характеризующие отдельные свойства.

– Знания – совокупность сведений, понятий, представлений о чём-либо, полученных в результате обучения, опыта и обычно реализуемых в деятельности.

– База данных – один или более файлов данных с произвольными организацией и доступом, которые служат для хранения неизменяющихся данных в вычислительной системе, для целей периодического мониторинга.

– База знаний (БЗ) – семантическая (смысловая) модель, описывающая предметную область и позволяющая отвечать на такие вопросы из этой предметной области, ответы на которые в явном виде не присутствуют в базе – осуществлять систематический и выборный мониторинг.

– Данные являются первичной информацией, которая хранится в базах данных, число которых в настоящее время достаточно велико и продолжает увеличиваться. На основе данных и баз данных строятся базы знаний.

С позиции используемых средств мировые информационные ресурсы разделяются на информацию, хранимую в традиционных информационных системах (библиотеках, архивах), и информацию, содержащуюся в электронных профессиональных базах, деловых ресурсах Интернет и электронных библиотеках [1, 2, 3, 4, 5]. Традиционные средства, содержащие документированную информацию, – библиотеки и архивы – также имеют компьютеры, но эти средства автоматизируют лишь часть функций поиска, сами документы хранятся в библиотеках и архивах в печатном виде.

Основой мировых информационных ресурсов являются профессиональные базы, которых в настоящее время свыше 20 тысяч. Формируют информацию и продают её поставщикам на коммерческой основе около 8 тысяч информационных агентств. К более чем 60% профессиональных баз имеется непосредственный доступ пользователей через телесети в режиме непосредственного доступа. В связи с тем, что источники информации в профессиональных базах тщательно отбираются,

информация характеризуется высокой степенью достоверности и постоянно обновляется. При хранении в профессиональных базах заранее определено, какая информация будет накапливаться в каком хранилище, что облегчает потребителю её поиск.

Основой деловых ресурсов Интернет являются web-страницы. Этот ресурс представляет собой страницы гипертекста, в которых может также содержаться графическая, звуковая и видеоинформация. В основном этот ресурс содержит рекламную информацию, и он может быть использован как дополнение к профессиональным базам.

В последнее время получили своё развитие электронные библиотеки, однако область их применения существенно ограничена законодательством в области авторского права. Одна и та же информация может быть представлена в библиотеках, в профессиональных базах, в деловых ресурсах Интернет и в электронных библиотеках. Объём информации в профессиональных базах – свыше 10 миллиардов записей. В России в 150 тысячах библиотек содержится свыше миллиарда экземпляров изданий, а количество электронных библиотек лишь 400. В настоящее время в мире издаётся свыше 200 тысяч периодических изданий, из них две трети без ограничений по направленности. Чуть больше 10% периодических изданий представлено в Интернет. Из 16 тысяч журналов в области экономики и бизнеса 12% журналов в полном объёме представлено в Интернет. Число web-страниц в Интернете в 2002 году превысило 4 миллиарда [1, 2, 3, 4, 5].

Приведенные характеристики объёмов информации, накопленных в различных типах мировых информационных ресурсов, говорят о гигантском объёме знаний, созданных мировым сообществом и доступных пользователям. По мнению специалистов, мировой объём накопленной информации удваивается каждые три года [1, 2, 3, 4, 5]. Изменения, которые происходят в обществе и экономике под влиянием увеличения объёмов доступной информации и внедрения современных информационно-телекоммуникационных технологий (ИКТ), наиболее чётко сформулированы в Окинавской хартии глобального информационного общества,

подписанной Великобританией, Германией, Италией, Канадой, Россией, США, Францией и Японией [1, 2, 3, 4, 5].

В настоящее время примером документальной системы могут служить деловые ресурсы Интернет. В тоже время по данным исследований полнота поиска в Интернет с помощью поисковых систем составляет менее 1%. Потребитель, как правило, не представляет, что объём не выданной, но удовлетворяющей его информации, т. е. релевантной информации, на два порядка превышает объём выданной.

Основной объём мировых информационных ресурсов (84%) – текстовая информация – отражает реальный мир фрагментарно. В хранимых документах источник описывает реальный мир в виде фрагмента его части, т.е. то, что он видит, или то, что он получил в результате анализа, обобщения. Отсутствие возможности автоматически обрабатывать смысловое содержание каждого документа, представленного на естественном языке, и, следовательно, обобщать полученную информацию ставит перед потребителем сложную и, чаще всего, невыполнимую задачу – в условиях крайне низкой полноты и большого шума проводить анализ и обобщение фрагментов для решения задачи, стоящей перед потребителем. Человек, обладая аналитическими способностями по выявлению содержания смысла текста, сталкивается с непреодолимыми трудностями, если ему нужно сопоставить смысл нескольких текстов.

Таким образом, в качестве важнейшего технологического инструмента инновационного развития представляется БЗ технологий и инноваций. Она должна носить распределённый характер и включать максимальное количество знаний (сведений и данных) из области науки, техники, технологий, производств, об инновационных продуктах, об участниках инновационного процесса (предприятиях, организациях, заказчиках, потребителях, специалистах) и других возможных участников. Подобная БЗ должна адекватно отражать всё состояние инновационной и технологической сферы. Она должна носить глобальный характер и формироваться как проект национального масштаба. При этом на начальных этапах должны формироваться первичные кластеры отраслей (подотраслей), регионов (отдельных

предприятий), уровней знаний (фундаментальная наука, технологии, патентная деятельность, производство, маркетинг, группы экспертов). Единая идеология разработки и глобальная цель позволят «сшивать» различные кластеры между собой для образования единого информационного пространства.

Объективные предпосылки создания БЗ технологий и инноваций:

- автономность и информационная закрытость деятельности отдельных разработчиков;
- отсутствие объективной информации о проводимых исследованиях, доступной широкому кругу специалистов и других участников инновационной деятельности;
- субъективизм при формировании программ инновационной деятельности;
- неоправданное дублирование исследований;
- отсутствие непрерывных технологических цепочек от фундаментальных исследований до создания конечных потребительских продуктов, что приводит к «замораживанию» отдельных результатов без выхода на потребительский рынок.

Цель создания БЗ – повышение эффективности функционирования инновационной инфраструктуры путём обеспечения потребителей (участников инновационной деятельности) полной и объективной информацией о состоянии технологического развития отраслей, о степени освоения инноваций, о структуре взаимодействия различных составных частей, элементов и обеспечивающих компонентов Хранилища Данных, а также субъектов инновационных процессов [1, 2, 3, 4, 5].

Задачи, решаемые БЗ:

- предоставление потребителям информации о проводимых исследованиях, конъюнктуре рынков, программах инновационной деятельности, о субъектах инновационных процессов, включая разработчиков, внедренцев, экспертов, потребителей;

– представление на основе информации из первичных источников формализованных знаний и построение моделей объектов разработок, технологических цепочек и предметной области в целом.

Требования к БЗ:

- открытость как способность к непрерывному и неограниченному развитию;
- распределённость – территориальная, по предметным областям, по типу пользователей, по уровням знаний;
- надёжность хранения данных путём распределения прав доступа, многократного дублирования информационных ресурсов;
- непротиворечивость;
- открытость для широкого круга пользователей по получению информации;
- возможность ограничения доступа к информации, имеющей гриф секретности или конфиденциальности;
- верификация и экспертиза первичной информации, предоставляемой пользователями для внесения в БЗ;
- полнота информационных ресурсов;
- связность информационных единиц, блоков и др. – отображение целостности предметной области;
- возможность интеграции и взаимодействия с другими информационными ресурсами;
- способность к преобразованию первичной информации и использованию единых форматов данных для представления семантической информации и построения моделей предметной области;
- наличие гибкого настраиваемого интерфейса для различного типа пользователей и типа решаемых задач.

Задачи начального этапа разработки БЗ технологий и инноваций:

- разработка системы интеграции разнородных баз данных, архивов научно-технической информации путём обеспечения доступа к ним в едином поисковом пространстве;

Научные исследования: от теории к практике

- построение интегрированной электронной библиотеки инновационных и технологических ресурсов региона (отрасли), внедрение библиотеки в инновационную систему региона (отрасли);
- разработка сетевой самоорганизующейся системы сбора информации о состоянии технологического развития и инновационной деятельности;
- разработка и внедрение методики сбора информации о новых научно-технических, технологических и инновационных разработках;
- разработка средств формализации описаний научно-технических и технологических объектов для введения данных о них в автоматизированную систему обработки и в базу знаний, а также предложений по стандартизации описаний объектов инновационной деятельности;
- разработка единой системы многоуровневой и многоаспектной классификации инновационных объектов;
- - разработка принципов и методики каталогизации инновационных технологий;
- разработка технологий построения моделей инновационного пространства;
- разработка АСУ для управления инновационными проектами.

Компонентами БЗ являются теоретический базис, содержащий знания об устройстве модели и предметной области – концептуальную основу, и фактическая модель, отражающая реальное состояние предметной области. Эти компоненты не должны противоречить друг другу. Теоретический базис является информационным тезаурусом предметной области [1, 2, 3, 4, 5]. Взаимодействие между двумя компонентами модели осуществляется следующим образом. Входная информация изменяет состояние фактической модели, эти изменения приводят к корректировке теоретического базиса, что в свою очередь позволяет переструктурировать данные фактической модели. Такая структура позволяет строить различные модели-страты, вводить новые правила, сверять модель с реальностью, в том числе уточнять правила – систему аксиом.

Всю информацию в базе знаний можно условно разделить на концептуальную и фактическую. К первой относится информация, отражающая смысловые взаимодействия между элементами. Здесь элементарная комбинация отражается единственным образом, множество таких комбинаций является составной частью теоретического базиса. Комбинация показывает, что связь или отношение данного типа между элементами данного типа возможно для осуществления в данной предметной области. Эта связь может быть реализована в модели предметной области многократно между различными экземплярами одинаковых сущностей в фактической базе знаний. Множество подобных реализаций составляет фактическую информацию [1, 2, 3, 4, 5]. Увеличение экземпляров известных комбинаций в БЗ прибавляет фактическое знание в ней, в то время как появление новых, ранее не отраженных связей (комбинаций) или элементов увеличивает семантическую ёмкость тезауруса, т.е. делает модель «умнее». Предложенная двухуровневая модель предметной области позволяет в фактической БЗ строить модели элементов и систем различных типов: модели на микро-, макро- и метаяровнях, имитационные модели, модели ситуаций, используемые в ситуационном управлении.

В настоящее время обоснованы принципы формирования модели БЗ и разработана методика построения содержательной части предметной области (независимо от её специфики).

Описанный подход к построению моделей позволяет строить БЗ, которые содержат не только фактуальную (текстовую) информацию, а фактически образ среды, что позволяет имитировать поведение последней, в реальном времени исследовать её поведение, исключает внесение в БЗ ложной информации, так как на модели осуществляется верификация данных. Становится возможным строить новые объекты, находить новые связи и эффекты [1, 2, 3, 4, 5].

Всё это приобретает особый интерес при включении интеллектуальных технологий в инновационный процесс. В частности, такая БЗ позволяет строить технологические цепочки, связывать между собой различные результаты, различных разработчиков, находить взаимосвязи между ними и определять разрывы, циклы,

параллелизмы и др. Модель позволяет найти требуемый эффект для замыкания технологического процесса, обнаружить имеющуюся технологию реализации этого эффекта и её разработчика. На модели можно проследить развитие инновационного объекта от идеи до конечного продукта. На модели возможно прогнозировать эффект от реализации программ научно-технического и инновационного развития; в перечнях системных мероприятий, входящих в программы, получать не разрозненный список проектов отдельных исполнителей, а систему взаимоувязанных работ, совместная реализация которых способна дать синергетический эффект от реализации программы и дать обоснованный задел на будущие исследования и мероприятия [1, 2, 3, 4, 5]. Роль экспертов при внесении первичных сведений является ключевой. Подбор экспертов в связи с этим приобретает большую важность, и от этого будет зависеть адекватность БЗ предметной области и её компетентность. БЗ, которая является моделью предметной области, как это описано выше, со временем, т.е. при её наполнении сведениями, сама превращается в эксперта, так как позволяет верифицировать новые сведения и проверять их на непротиворечивость. В результате с помощью интеллектуальных инструментов, средств моделирования и знаний, находящихся в БЗ, становится возможным проверять компетентность экспертов.

Истинность знаний обеспечивается сбалансированным взаимодействием специалистов-экспертов, автоматизированных интеллектуальных средств и способов верификации формализованных знаний непосредственно на внешней среде.

2. Проектирование системы базис-анализа – создание хранилища данных.

Оценка Business Intelligence (BI) – это процесс получения знаний, необходимых для принятия управленческих решений на основе накопленных в компании данных. Общая идея Business Intelligence заключается в объединении данных и превращении их в информацию.

Система Бизнес Анализа (BI-система) – это система, обеспечивающая сбор, очистку и преобразование данных, а также средства конечных пользователей для анализа данных, представления и распространения результатов.

Основными составными компонентами (элементами) систем Бизнес Анализа является хранилища данных, зачастую в BI-системы входят следующие элементы:

- ETL-средства. Механизмы, обеспечивающие автоматическую выгрузку нужных данных из необходимых источников, их преобразование и загрузку в хранилище.

- Средства регламентной отчетности. Часто требуется разрабатывать стандартные отчеты для разных департаментов компании. Отчеты, полученные на основе хранилища, будут консолидированы и непротиворечивы между собой.

- OLAP-анализ. Позволяет разрабатывать OLAP-кубы (многомерные, иерархические структуры данных), на основе которых аналитики получают быстрый доступ к огромным массивам информации.

- Data Mining, или Интеллектуальный Анализ. Мощные технологии обнаружения глубоких (нетривиальных и неочевидных) зависимостей и шаблонов в исходных данных и их дальнейшая интерпретация.

Современные BI-системы используют трехзвенную архитектуру. Это означает наличие клиентского приложения (обычно просто браузера), а также сервера приложений и сервера базы данных. Сервер приложений обеспечивает хорошую масштабируемость, а также существенно снимает вычислительную нагрузку с сервера БД.

В последнее время BI-системы связывают с Системами Поддержки Принятия Решений (DSS, Decision Support System). DSS – это целый класс систем, который нацелен на информирование менеджеров о ключевых показателях в компании [1, 2, 3, 4, 5]. На основе полученной информации можно проверять собственные гипотезы и принимать какие-либо управленческие решения, например, продвижение какого-либо товара или увеличение доли рынка. Рассмотрим более детально перечисленные компоненты. ETL-средства (Extraction, Transformation, Loading; Извлечение, Преобразование, Загрузка). Прежде чем информация попадет в хранилище, необходимо провести следующие процедуры:

- извлечение – следует извлекать только нужные данные для анализа, извлечение происходит из разных источников данных;

Научные исследования: от теории к практике

– преобразование – необходимо привести данные к общему формату и обеспечить соответствие данных между собой, нужно очистить данные от возможных противоречий;

– загрузка – после обработки данных их следует загрузить в хранилище.

Обычно различают 2 вида загрузки данных в хранилище:

– первоначальная загрузка данных производится один раз на этапе построения Хранилища, она подразумевает сбор исторической информации и данных, актуальных на начальном этапе построения хранилища; на этом этапе тестируются выбранная архитектура, возможности и функциональность инструментов;

– регламентная загрузка обеспечивает актуальность данных в хранилище, производится строго по расписанию в назначенное время, интервалы между регламентными загрузками выбираются исходя из задач компании и специфики ее бизнес-процессов, а также частоты обновления данных в источниках.

Современные BI-системы представляют свои встроенные ETL-средства, также существует специальное ПО, встроенное в СУБД, но часто ETL-алгоритмы пишут программисты вручную. Аналитикам в процессе работы требуется оперировать огромными массивами данных, при этом скорость обработки данных должна быть очень высока. Отчеты, полученные на основе обычных запросов к базе данных, не могут удовлетворять таким требованиям. Поэтому появились OLAP-кубы – многомерные иерархические структуры данных.

OLAP-приложение – это система, обеспечивающая представление данных в виде многомерной модели и позволяющая работать с ними интерактивно с гарантированно низким временем отклика. Многомерная модель представления данных очень помогает, когда существует целый набор измерений, по которым необходимо отслеживать важные показатели. Измерение (Dimension) – это последовательность значений одного из анализируемых параметров. Например, для параметра «Регион» – это список городов или населенных центров. Почти всегда измерения можно представить в виде иерархической структуры, например, «Страна–область–город–район», или «Год–квартал–месяц–неделя». На пересечении осей измерений располагаются Меры (Measures) – данные, которые количественно характеризуют

анализируемые факты, например, объемы продаж или поставок, суммы издержек и другие их виды [1, 2, 3, 4, 5].

Для того чтобы значительно увеличить скорость доступа к данным, OLAP-кубы содержат также и агрегированную информацию. В общем случае OLAP-кубы содержат более трех измерений. Для удобного доступа к данным и уточнения информации аналитик может в любой момент времени осуществить определенные действия с OLAP-кубом.

- Срез (Slice) – получение подмножества данных, удовлетворяющих определенному условию, например, получить все поставки или все издержки на хранение определенного товара в регионе.

- Вращение (Rotate) – изменение расположения измерений, представленных в отчете или на отображаемой странице.

- Детализация (Drill Down) и Консолидация (Drill Up) – операции, которые определяют переход на уровень более детальных и агрегированных данных соответственно. При этом происходит переход по иерархиям измерений. Например, можно выполнить детализацию по измерению «Дата» с уровня «Год» на уровень «Месяц» или консолидацию с уровня «Город» на уровень «Область». С помощью этой технологии аналитик может исследовать ситуацию в целом, а затем сконцентрировать свое внимание на каком-либо конкретном бизнес-объекте.

Вместе с базовой технологией существуют три основных типа OLAP-приложений:

- MOLAP (Multidimensional OLAP, Многомерный OLAP) – это классическая форма OLAP, данные хранятся в виде упорядоченных многомерных массивов, MOLAP содержит уже агрегированные показатели, тем самым значительно ускоряет работу с OLAP-кубами;

- ROLAP (Relational OLAP, Реляционный OLAP) работает напрямую с реляционными базами данных, для хранения агрегатов создаются дополнительные реляционные таблицы, благодаря такой схеме хранения данных ROLAP достаточно просто реализовать на основе обычного хранилища без каких-либо дополнительных действий;

Научные исследования: от теории к практике

– HOLAP (Hybrid OLAP, Гибридный OLAP) использует реляционные таблицы для хранения атомарных данных и многомерные массивы для агрегатов.

Совокупность большого количества записей в базе данных, накопленных за несколько лет, может стать источником дополнительной, гораздо более ценной информации, которую нельзя получить на основе одной конкретной записи [1, 2, 3, 4, 5]. Примерами подобной информации являются сведения о том, как зависят продажи определенного товара от дня недели или времени года, какие категории покупателей чаще всего приобретают определенный товар, какая категория клиентов чаще всего вовремя не отдает предоставленный кредит и т.д.

Подобного рода информация обычно используется при прогнозировании, стратегическом планировании, анализе рисков, и ценность ее для предприятия очень высока. Термин Data Mining обозначает не столько конкретную технологию, сколько сам процесс поиска корреляций, тенденций, взаимосвязей и закономерностей посредством различных математических и статистических алгоритмов. Data Mining (дословно «Добыча» данных) – это процесс обнаружения в «сырых» данных ранее неизвестных, нетривиальных, практически полезных и доступных к интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности [1, 2, 3, 4, 5]. Выделяют 5 стандартных типов закономерностей, выявляемых методами Data Mining:

- ассоциация – высокая вероятность связи событий друг с другом (например, один товар часто приобретается вместе с другим);
- последовательность – высокая вероятность цепочки связанных во времени событий (например, в течение определенного срока после приобретения одного товара будет с высокой степенью вероятности приобретен другой);
- классификация – имеются признаки, характеризующие группу, к которой принадлежит то или иное событие или объект (обычно при этом на основании анализа уже классифицированных событий формулируются некие правила);
- кластеризация – закономерность, сходная с классификацией и отличающаяся от нее тем, что сами группы при этом не заданы, они выявляются автоматически в процессе обработки данных;

– временные закономерности – наличие шаблонов в динамике поведения тех или иных данных (типичный пример – сезонные колебания спроса на те или иные товары либо услуги), используемых для прогнозирования.

Синергические решения появляются при разработке новых безопасных технологий в поиске и устранении или защите наиболее уязвимых мест, конструировании оригинальных устройств и др. Соединение структурно-функционального и процессного управления, а также интеграция «жестких» – информационных и инженеринговых – технологий и «мягких» методов снижает зависимость компании от конкретных исполнителей и одновременно мотивирует персонал, повышает управляемость компании, обеспечивает предсказуемость и устойчивость деятельности. В настоящее время существуют 5 крупнейших корпораций, предлагающие BI-решения, известные своей функциональностью и надежностью, например, Oracle после поглощения компаний, таких как Siebel, Hyperion, Sunopsis и других, а также Microsoft, SAP (Business Objects), IBM (Cognos), SAS Institute. В настоящее время создано большое количество форматов для хранения информационных ресурсов, поэтому часто стоит задача автоматизированного преобразования этих ресурсов в другие форматы хранения. Реализованная универсальная система конвертирования позволяет решать эту задачу, тем самым повышая скорость получения необходимых пользователю информационных ресурсов.

Дальнейшее развитие универсальной системы конвертирования является постепенным добавлением новых модулей и автоматизации процесса конвертирования, добавляя, например, элементы искусственного интеллекта или элементы текстового интеллектуального данных (Text Mining). Важно понимать, что грамотное внедрение перечисленных технологий позволит компании выйти на качественно новый уровень BI-системы и создать единое информационное пространство [1, 2, 3, 4, 5]. Используя анализ данных, накопившихся в компании за продолжительный период времени, управленческие решения будут приниматься значительно быстрее, чем у конкурентов. В настоящее время BI-технологии активно используют банки и страховые компании, телекоммуникационные компании, сети розничной торговли, а также многие другие отрасли, например:

Научные исследования: от теории к практике

- для оценки финансовых последствий управленческих решений в процессах перспективного и оперативного планирования в координатах «риск-доходность»;
- формирование единой сбалансированной и непротиворечивой системы ключевых показателей результативности для интегральной оценки последствий управленческих решений;
- осуществление балансировки ключевых показателей результативности на базе единой финансовой модели «предприятие-банк»;
- использование единых форматов плановых, отчетных и прогнозных документов, создаваемых для целей принятия управленческих решений (ПУР);
- организация сбора информации для ПУР в рамках единого информационного пространства (ЕИП), создаваемого на базе информационного хранилища (ИХ).

3. Автоматизация действующих предприятий – сокращение затрат полезного времени.

С учетом применения хранилища данных и баз знаний базовый уровень коэффициента полезного времени (КПВ) автоматизированного оборудования на конкретном технологическом переходе может быть получен путем обобщения опыта работы действующих предприятий как в России, так и за рубежом. Кроме того, выявленные резервы роста частных коэффициентов полезного времени указывают на направления перспективных (первоочередных) теоретических исследований в области совершенствования техники и технологии в промышленности [1, 2, 3, 4, 5]. Решать проблему роста КПВ следует путем повышения не всех K_i , а только их части и в первую очередь тех, которые ниже уровня, определенного по выражению (1). С учетом такой постановки часть K_i следует принять условно-постоянными (p), а оставшуюся часть K_i – переменными ($m-p$) величинами, и, следовательно, соответствующие длительности простоев $t_{i\text{var}}$ принимаются в качестве аргументов в уравнении (1), тогда:

$$KПП = \left[\sum_{i=1}^p K_{i_{const}}^{-1} - (m-1) - \sum_{i=1}^{m-p} K_{i_{var}}^{-1} \right]^{-1} \quad (1)$$

с учетом формулы (1) получим:

$$KПП = \left[\sum_{i=1}^p K_{i_{const}}^{-1} + (1-m) + \sum_{i=1}^{m-p} t_{i_{var}}^{-1} / t_p \right]^{-1} \quad (2)$$

$$\text{Обозначив: } 1 + \sum_{i=1}^p K_{i_{const}}^{-1} - m = \alpha \quad | \quad (3)$$

$$\text{и учитывая, что } t_p / 1 = A_T \quad (4)$$

где: A_T – теоретическая производительность единицы оборудования, получим удобное для анализа $KПВ$ выражение:

$$KВП = (\alpha + \eta A_T)^{-1} \quad (5)$$

Величина α получается в результате алгебраических преобразований при выводе формулы (5). В случае если все частные K_i соответствуют оптимальным значениям, можно говорить о соответствии техники, технологии и организации производства современному уровню развития научно-технического развития в данной области [1, 2, 3, 4, 5].

Блок прогнозирования эффективности внедрения технико-организационных инноваций, направленных на повышение результативности использования оборудования, предполагает проведение анализа и синтеза технических, технологических, организационных инноваций и расчет целесообразности их внедрения. На основе результатов проведенной диагностики выявляются резервы повышения производительности труда, проводится анализ технических, технологических или организационных инноваций, которые, по мнению эксперта должны способствовать повышению производительности труда.

Получено выражение, позволяющее оценить реальные резервы роста $KПВ$ автоматизированного оборудования в результате внедрения технико-экономических мероприятий при сокращении потерь m -го вида в ε раз:

$$T_p = \frac{KПВ}{KПВ_{исх}} = \left[1 - KВП_{исх} t_{n_i} (1 - \varepsilon/1) \right]^{-1} \quad (6)$$

где: $KПВ_{исх}$, $KПВ$ – соответственно коэффициент полезного времени автоматизированного оборудования до и после проведения мероприятий, направленных на его повышение, ε – коэффициент сокращения потерь конкретного вида.

Блок сравнительного анализа результативности использования автоматизированного оборудования различных типов предоставляет возможность выбора наиболее прогрессивного оборудования. Прогрессивность новой техники определяется ее соответствием главному целевому назначению – повышению производительности и облегчению условий труда. Важно отметить, что прогрессивность новой техники оценивается независимо от возможных организационных неувязок (затянувшиеся сроки ввода в эксплуатацию, низкий уровень обслуживания и другие факторы) при полном использовании заложенных в ней возможностей [1, 2, 3, 4, 5]. Сравнительный анализ эффективности использования автоматизированного оборудования различных типов осуществляется путем сопоставления частных коэффициентов полезного времени K_i по видам простоев и группам: $ЧКПВ_{цп}$, $ЧКПВ_m$ и $ЧКПВ_{от}$.

Цикловой частный коэффициент полезного времени $ЧКПВ_{цп}$ характеризует конструктивное совершенство автоматизированного оборудования (наличие/отсутствие холостых ходов, связанных с питанием автоматизированного оборудования, съемом наработанной продукции и других операций):

$$ЧКПВ_{цп} = \left[\sum_{i=1}^S K_{i_{цп}}^{-1} - (S - 1) \right]^{-1} \quad (7)$$

где: S – число цикловых простоев, $K_{i_{цп}}$ – частный коэффициент полезного времени, соответствующий определенному виду цикловых потерь. Технологический частный коэффициент полезного времени $ЧКПВ_t$ характеризует технологическую надежность оборудования (совпадает с коэффициентом технологической надежности, принятым в приборо- и машиностроении):

$$ЧКПВ_t = \left[\sum_{i=1}^y K_{it}^{-1} - (y - 1) \right]^{-1} \quad (8)$$

где: y – число технологических простоев, K_{it} – частный коэффициент полезного времени, соответствующий определенному виду технологических потерь. Далее введем организационно-технический частный коэффициент полезного времени $ЧКПВ_{om}$, который характеризует уровень организации труда и условия эксплуатации:

$$ЧКПВ_{om} = \left[\sum_{i=1}^{\Phi} K_{iom}^{-1} - (\Phi - 1) \right]^{-1} \quad (9)$$

где: Φ – число видов организационно-технических простоев, K_{iom} – частный коэффициент полезного времени, соответствующий определенному виду организационно-технических потерь.

Произведение $ЧКПВ_m$ и $ЧКПВ_{om}$ численно совпадает с так называемым коэффициентом использования автоматизированного оборудования, принятым в приборостроении. В теории надежности существует понятие коэффициента готовности, который можно условно принять равным K_i , если считать его связанным только с текущим ремонтом. С помощью этого коэффициента можно оценить надежность конструкции оборудования. Сравнивая величины показателей, можно судить соответственно о конструктивном и технологическом совершенстве оборудования, а также об условиях эксплуатации, уровне организации труда и надежности автоматизированного оборудования.

Блок контроля соответствия нормы производительности технически обоснованным нормам создан для рационализации организации труда операторов автоматизированного оборудования и позволяет установить соответствие производительности автоматизированного оборудования технически обоснованным нормам. Данный блок применяется совместно с блоком диагностики, выявленные в результате диагностики низкие уровни частных K_i , относящихся к цикловой либо органи-

зационно-технической группе простоев, подвергаются анализу. Проводятся фактические наблюдения и измерения, прежде всего, фотографирование работы оборудования и хронометраж его простоев по причинам, которые обусловили низкий уровень K_i . Первичная обработка этой информации сводится к получению сведений о сомнительных величинах простоев, вошедших в расчет нормы производительности автоматизированного оборудования. В случае необходимости производится перерасчет норм производительности и частных коэффициентов полезного времени K_i [1, 2, 3, 4, 5].

Резервы повышения производительности автоматизированного оборудования в случае сокращения нескольких видов потерь определяются по формуле:

$$\lambda_k = \left[\sum_{j=1}^{\varphi} K_j^{-1} - (\varphi - 1) \right]^{-1} / \left[\sum_{j=1}^{\varphi} K_j^{t-1} - (\varphi - 1) \right] \quad (10)$$

где: K_i – коэффициент полезного времени, характеризующий потери данного вида, которые можно сократить; K_j^t – то же после сокращения потерь времени; φ – число изменяющихся K_i .

Результативность использования системотехнической методологии проектирования технологии и организации производства оценивается прежде всего показателями темпов роста производительности труда, а также снижением затрат на производство продукции. Получено выражение, позволяющее оценить изменение затрат на производство продукции за счет роста производительности труда [1, 2, 3, 4, 5]. В случае однопереходного производства нижеприведенное выражение позволяет непосредственно оценить эффективность мероприятий, направленных на повышение производительности труда, и получить величину снижения себестоимости единицы продукции:

$$\Delta C_{\Pi\Gamma} = \left[1 - I_{3\Pi} / \left[1 - KB\Pi_{\text{исх}} t_{n_i} (1 - \varepsilon/1) \right]^{-1} \right] Y_{3\Pi} 100 \% \quad (11)$$

В случае многопереходного производства (инновационного, наукоемкого, приборостроительного, электротехнического или других видов), выражение (11)

дает возможность получить величину изменения издержек на конкретном переходе.

Экономический эффект от внедряемых инновационных информационных технологий автоматизации оборудования в общем виде достигается за счет дополнительного выпуска продукции, снижения затрат на производство, повышения сортности выпускаемой продукции. Информационная интеграция процессов, обеспечивающая совместное и многократное использование одних и тех же данных, достигается минимизацией числа и сложности вспомогательных процессов и операций, связанных с поиском, преобразованием и передачей информации. Поскольку доля вспомогательных процессов и операций в общем цикле достаточно велика, сокращение связанных с ними затрат времени и средств является существенным фактором экономии. Одним из инструментов интеграции является стандартизация способов и технологий представления данных с тем, чтобы результаты предшествующего процесса могли быть использованы для последующих процессов с минимальными преобразованиями [1, 2, 3, 4, 5].

Основой научно-методического обеспечения качества этой продукции является современная методология менеджмента качества, базирующаяся на теории управления процессами и информационными потоками на всех стадиях жизненного цикла продукции. Для предприятий приборостроения должна быть разработана специализированная методология создания и обеспечения эффективного функционирования современных систем менеджмента качества, соответствующих требованиям стандартов ИСО серии 9000. По опыту передовых зарубежных фирм кардинальным средством повышения качества и конкурентоспособности наукоемкой продукции является применение информационных технологий автоматизации оборудования и технологических процессов [1, 2, 3, 4, 5].

Результатом успешного внедрения CALS-технологий на ведущих западных фирмах является введение требований по применению указанных технологий в процессах разработки, производства, эксплуатации и утилизации продукта производства в качестве обязательного условия получения заказа на разработку новейших видов высокотехнологической продукции.

Научные исследования: от теории к практике

С целью реального применения изложенных в настоящей статье алгоритмов и научно-практических подходов еще в 2008 году в России был разработан программно-аппаратный комплекс «ВТест», позволяющий постоянно совершенствовать автоматизированную систему обработки информации и формирования базы знаний инновационной организации (фирмы) любого профиля деятельности [1, 2, 3, 4, 5].

Спроектированные с учетом требований стратегического анализа автоматизированные системы управления, комплексно охватывающие проблемы и процессы управления организацией, предоставляют возможность руководству оперативно выявлять вопросы, требующие немедленного вмешательства, получать по ним детализированную информацию и в дальнейшем принимать эффективные решения.

Список литературы

1. Ковалев С.В. Методическая база оценки затрат на внедрение информационных технологий / С.В. Ковалев // Информационные технологии моделирования и управления. – 2010. – №1(60). – С. 87–95.
2. Ковалев С.В. История развития и моделирования корпоративных информационных систем / С.В. Ковалев // Информационные технологии моделирования и управления. – 2009. – №7(59). – С. 880–888.
3. Ковалев С.В. Экономическая математика. – М.: ИД «Кнорус», 2010. – 248 с.
4. Новиков Д.А. Управление проектами: организационные механизмы. – М.: Пмсофт, 2007. – 140 с.
5. Новиков Д.А., Иващенко А.А. Модели и методы организационного управления инновационным развитием фирмы. – М.: Ленанд, 2006. – 336 с.