

1. Понятие знаний. Управление знаниями

Знания – результат познания действительности, проверенный практикой или практическая информация, которая активно управляет процессами выполнения задач, решения проблем и принятия решений. При этом управлять знаниями – значит систематически, точно и продуманно формировать, обновлять и применять их с целью максимизации эффективности предприятия и прибыли от активов, основанных на знаниях.

К знаниям можно отнести любое слово, факт, пример, событие, правило, гипотеза или модель, которые усиливают понимание или исполнение в определенной области деятельности или дисциплины.

Знания обладают следующими особенностями:

- имеют отношения к данным и информации;
- дефицитны, их непросто добывать;
- всегда связаны с каким-то контекстом, существуют в его рамках;
- генерируются только людьми;
- динамичны;
- любые знания обладают своей скоростью передачи и восприятия;
- для успешного восприятия требуют четких границ их понимания;
- могут быть очень дороги, цена при этом не фиксирована;
- обладают сроком и целью использования.

Несмотря на многообразие и расплывчатость в определении понятия «знание» многие исследователи всю совокупность знаний разделяют на следующие классы: *эмпирические*; *теоретические*; *личностные*; *организационные*; *неявные*; *явные* знания.

К эмпирическим знаниям относятся наблюдения, наблюдаемые явления.

К теоретическим знаниям относятся законы, теории, абстракции, обобщения.

К личностным профессиональным знаниям относятся: знания познавательного плана («знаю, что»); прикладное мастерство («знаю, как»); системное представление («знание, почему»); личная мотивация творчества («хочу знать, почему»).

К организационным знаниям относятся стратегические доктрины и программы, теоретические знания конкретной сферы деятельности, технологические и производственные стандарты, правила и инструкции, коммерческие знания. Организационные знания являются совокупностью знаний каждого из сотрудников, но не простой суммой, а специально отобранной и тщательно проверенной в практике деятельности групп, подразделений, отделений и компании в целом.

Для анализа деятельности экономического объекта необходимы явные знания, так как экономика – наука, построенная на анализе цифр и использовании различного рода методов расчетов, иногда достаточно простых, а иногда достаточно сложных.

Явные знания – формализованные знания могут быть выражены словами, цифрами и символами, легко изложены и распространены в виде чисел, формул, алгоритмов или всеобщих принципов. Они могут систематизироваться, передаваться, накапливаться. Человек может самостоятельно получить их, пользуясь их представлением в виде документов, инструкций, регламентов, положений, учебных пособий и т.д.

Исследователи полагают, что явные знания, выражаемые словами и цифрами это всего лишь «верхушка айсберга», а знание является в основном неформализованным, то есть не является легко объясняемыми и видимыми.

Неявные (неформализованные) **знания** существует на уровне индивидуума и плохо поддается формализации. Это затрудняет их передачу и использование другими людьми кроме владельца. Эти знания тесно связаны с опытом и действиями конкретного человека.

К неявным знаниям относят опыт, мастерство, культуру мышления, интуицию, информацию, хранящуюся в нейронных структурах головного мозга

как результат генетической наследственности, образование и приобретенный жизненный опыт.

Выявление и использование неявных знаний позволяет решить сложные задачи и увидеть предприятие как живой организм – зачем оно существует, в каком направлении развивается, в каком мире хочет жить и как этот мир создать. Знание подразумевает не только образы и символы, но и духовные ценности, интуицию и эмоции.

Знания могут быть представлены в разных формах:

- *научная* – объективное, системно организованное и обоснованное знание;
- *ненаучная* – разрозненное, несистематическое знание, которое не формализуется и не описывается законами;
- *донаучное* – прототип, предпосылки научного знания;
- *паранаучное* – несовместимое с имеющимся научным знанием;
- *лженаучное* – сознательно использующее домыслы и предрассудки; *антинаучное* – утопичное и сознательно искажающее представление о действительности.

Часто знания оказываются чем-то большим, чем просто информацией и данными о событиях, продуктах или процедурах.

Управление знаниями (Knowledge Management) – это управление процессами поиска, формирования, передачи, использования, хранения, объединения и генерации новых знаний, которые необходимы для успеха организации.

Под *управлением явными знаниями* понимают применение стандартных методов. Неявные знания не поддаются управлению такими методами и требуют выработки новых методов воздействия на людей. К ним относятся формирование стратегического партнерства с потребителями и контрагентами, осуществление сбалансированной инновационной политики, формирование этических основ деятельности организации, формирование продуктивной корпоративной культуры. Одним из подходов к управлению неявными знаниями стало формирование обучающейся организации.

Очевидно, что понятие управление знаниями объединяет в себя технологию менеджмента – стратегия управления предприятием для обеспечения функций по созданию и повышению количества и качества интеллектуальных знаний, и информационную технологию – программное и техническое обеспечение для реализации задач хранения, поиска, обработки, анализа и использования документов, информации и знаний.

Цель управления знаниями – помочь людям лучше работать вместе, используя все возрастающие объемы информации и управляя ими. Основными компонентами управления знаниями являются люди, процессы и технологии.

Управление знаниями включает следующие направления деятельности: обмен знаниями, управление внешними потоками информации, обучение, структуризация знаний на предприятии, совместная работа в социальных сетях и др.

Для сокращения дефицита знаний необходимо решать такие задачи как: *приобретение знаний* (использование уже имеющихся знаний в мире и получение новых знаний путем ведения научно-исследовательских и опытно-конструкторских работ); *усвоение знаний* (обеспечение начального обучения и создание возможностей для обучения на протяжении всей жизни); *передача знаний* используя новые каналы связи.

Этапы управления знаниями:

- *определение* – выявление знаний, имеющих решающее значение для успеха;
- *сбор* – приобретение существующих знаний, опыта, методов и квалификации;
- *выбор* – оценка полезности собранных, упорядоченных знаний;
- *хранение* – отобранные знания классифицируются и вносятся в организационную память (в человеческую, на бумаге, в электронном виде);
- *распределение* – знания извлекаются из корпоративной памяти, становятся доступными для использования;

- *применение* – использование полученных знаний при выполнении заданий, решении проблем, принятии решений, поиске идей и обучении;
- *создание* – выявляются новые знания путем наблюдения за клиентами, обратной связи, причинного анализа, эталонного тестирования, опыта, исследований, экспериментирования, креативного мышления, разработки данных;
- *продажа* – на основе интеллектуального капитала – новые продукты и услуги, которые могут быть реализованы вне предприятия.

В настоящее время выделяют такие *подходы к управлению знаниями*:

- *персонифицирующий* или *интуитивистский подход* – использование в качестве носителей знаний конкретных людей (экспертов), на которых возложена и функция передачи (распространения) знаний другому персоналу;
- *информационный* или *технологический подход* – поиск знаний в огромных базах и массивах данных (о клиентах, поставщиках, операциях и т.п.), накопленных современными предприятиями, и их обработка осуществляются интеллектуальными информационными технологиями¹, позволяющими выявлять скрытые зависимости в данных;
- *логистический подход*, в основе которого лежит семь «R» – обеспечение пригодности правильного знания, в правильном количестве и правильном состоянии, в правильном месте, в правильное время, для правильного клиента и по правильной стоимости.

В условиях информатизации и глобализации конкурентные преимущества, базирующиеся на явных знаниях, перестают быть устойчивыми, они превращаются в краткосрочные преимущества. В то же время преимущества, базирующиеся на неявных знаниях, носят более устойчивый характер, так как неявные знания сложно скопировать, осуществить их трансфер и применить на другом предприятии. Для того чтобы добиться устойчивых преиму-

¹ Под *интеллектуальными информационными технологиями* понимают информационные технологии, в которых предусмотрены возможности: наличие баз знаний, наличие моделей мышления на основе баз знаний, способность формировать вполне четкие решения на основе нечетких, нестрогих, неполных, не доопределенных данных, способность объяснять выводы и решения, а также способность к обучению, переобучению.

ществ, которые сохранялись бы в течение продолжительного времени, необходимо на предприятии создать систему управления знаниями с учетом непрерывного поиска и исследования информации.

Система управления знаниями (СУЗ) – это совокупность организационных процедур, организационных подразделений и компьютерных технологий, обеспечивающих интеграцию разнородных источников знаний, и их коллективное использование в бизнес процессах. Отличительной особенностью СУЗ является интеграция знаний как из внутренних, так и из внешних источников. Источники знаний могут иметь недокументированную форму (неявные знания), документированную текстовую, табличную, графическую и структурируемую форму в виде баз знаний.

Функции системы управления знаниями:

- систематизация и накопление актуальной информации и знаний из различных источников;
- обеспечение доступа сотрудникам к явным и неявным знаниям;
- использование языка профессионального общения на основе метаданных и онтологического описания предметной области;
- обеспечение семантического поиска и фильтрации знаний в корпоративных информационных ресурсах;
- обмен знаниями между специалистами, экспертами, проектными группами и сообществами по интересам;
- распространение знаний в соответствии с установленным регламентом (кому, что, когда) и по запросам;
- хранение и обновление профилей компетентности специалистов предприятия, поддержку процессов экспертизы и консультирования;
- генерация новых идей, их апробация и использование в бизнес-процессах;
- поддержка процесса принятия решений;
- создания сетевой среды, позволяющей повышать квалификацию специалистов в режиме дистанционного обучения;

- повышения корпоративной культуры в реализации управленческих процессов.

В научной литературе существует различные подходы к классификации технологий управления знаниями.

Согласно отчетов Gartner Group² выделяют три основных группы технологий управления знаниями:

- *технологии доступа к информации* – поиск по ключевым словам и с использованием классификации документов по группам, сложные запросы;
- *совместная работа групп и социально-ориентированное ПО* – поддержка коммуникаций и создание единой виртуальной рабочей среды,
- *системы управления контентом (Enterprise Content Management)* – различный инструментарий для гибкой и надежной работы с документами, для структурирования содержимого сайтов и для организации взаимодействия с посетителями сайтов.

Крупнейшими разработчиками средств, поддерживающих эти группы технологий управления знаниями, в настоящее время являются компании: HP Autonomy, Fast Search&Transfer, Endeca, IBM, BEA, Spike Source, Socialtext, Microsoft, Айкумена и другие.

Лидерами в группе «*технологии доступа к информации*» являются продукты компаний Autonomy (IDOL™ Server), Fast Search&Transfer (FAST ESP, Convera RetrievalWare), Endeca, IBM и др.

В настоящее время широко используются:

- *IDOL (Intelligent Data Operating Layer)* – позволяет передавать задачи обработки системе HP StoreAll, осуществлять быстрый анализ на основе свежих данных с использованием меньшего количества компьютерного оборудования, обеспечивает надежное хранение цифровых активов в течение длительного срока и преобразовать машинные коды в понятную бизнес-логику;
- *Adaptive Information Warehouse* – корпоративная система поиска информа-

² *Gartner Group* – исследовательская и консалтинговая компания, специализирующаяся на рынках информационных технологий и известна введением в употребление понятия ERP.

ции, которую можно применять взамен хранилища данных и способную формировать корпоративные данные независимо от места их нахождения, производить оперативную аналитическую обработку данных, генерацию отчетов и т. д.;

- *Галактика ZOOM* предназначена для аналитической обработки текстовых неструктурированных документов, находящихся в подключаемых базах данных и включает классические поисковые системы, системы сбора текстовых данных и системы аналитической обработки информации;
- *Яндекс.Сервер* – осуществляет корпоративный поиск по сайту, обеспечивает полнотекстовый поиск информации с учетом морфологии русского языка на web-сервере или в корпоративной сети;
- *Айкумена–Аналитика* – поисково-аналитическая система для оперативной обработки больших объемов информации, полученной из различных источников;
- *IBM OmniFind* – предоставляет полнотекстовый поиск по ключевым словам и фразам, позволяет задавать правила для автоматической классификации документов и одновременно проводится поиск по всем ресурсам – в сети Интранет, по внешним корпоративным сайтам предприятий, в базах данных, и т.д.;
- *IBM Classification Module* – автоматически классифицирует неструктурированный контент по заданным категориям, создает и настраивает таксономию, выступает дополнением к IBM Omnifind, что повышает эффективность и точность поиска с помощью навигации;
- *WebSphere Portal* – программная среда, реализующая единую точку индивидуального взаимодействия людей и программ.

Лидерами в группе «совместная работа групп и социально-ориентированное ПО» являются продукты компаний IBM, Socialtext (Socialtext), Microsoft, SpikeSource (SuiteTwo), BEA (BEA Aqualogic), и др.:

- *IBM Lotus Notes* обладает функциями планирования (календарь) и работы с электронной почтой, похож на почтовую систему MS Outlook;
-

- *IBM Lotus Sametime* предназначена для общения и совместной работы с возможностью проведения web-конференций, позволяет передавать голосовые и видео-сообщения, определять местонахождение участников, и поддерживает связь с мобильными устройствами;
- *Lotus Quickr* обеспечивает совместную работу нескольких пользователей в интерактивном режиме как внутри предприятия, так и за ее пределами;
- *IBM Lotus Connections* представляет собой решение для социализации бизнес-процессов, расширяющее инновационные возможности и ускоряющее работу с помощью объединения сотрудников, партнеров и заказчиков в сети, содержит такие компоненты: Profiles, Blogs, Communities, Activities и средство создания закладок Bookmarks (Dogear);
- *Socialtext* – социальное программное обеспечение, включающее wiki, социальную сеть, систему микроблоггинга. Организация документов производится только с помощью тэгов, содержит средства работы с загруженными файлами, имеет профайлы пользователей, стартовая страница, полностью совместимая с iGoogle;
- *SuiteTwo* – продукт, ориентированный на бизнес-пользователей, в его состав входят программное обеспечение для корпоративного блоггинга, RSS-агрегатор, поисковые утилиты и сервис Wiki. Все они интегрированы в одну оболочку.

Лидерами в группе «системы управления контентом» являются продукты компаний IBM (Content Manager, FileNet P8), EMC (EMC Documentum), Opentext (Livelink ECM 10):

- *IBM Content Manager* предназначена для хранения, контроля, распределения и интеграции информации. Интегрирует данные различных форматов (документ, изображение, медиа и другие) и совместима с различными профессиональными системами сканирования;
 - *IBM WebSphere Information Integrator* обеспечивает доступ к данным, хранящимся в разнородных системах, интегрируя содержимое порталов, бизнес-приложений и других программ;
-

- *EMC Documentum Content Services* для порталов SAP позволяет архивировать и управлять контентом для всех приложений SAP;
- *Documentum Content Services* представляет собой менеджера хранилищ данных, совместимый с SAP, и набор готовых к работе приложений на базе технологии iView для порталов SAP, который позволяет пользователям создавать контент, обеспечивать доступ к нему и управление в контексте приложений на базе SAP, а для порталов SAP интегрируется с EMC Archive Services для SAP, обеспечивая пользователям доступ к контенту, архивированному из действующего SAP-приложения в рамках портала SAP;
- *Open Text Livelink ECM* — решение для совместной работы и управления корпоративным контентом (системы электронного документооборота, 3-rd party продукт и используется для хранения и обмена файлами на web – платформах, предлагает возможности применения метаданных к документам, аудит всех событий документа и управление «историей версий»).

С развитие сети Интернет у пользователей появилась возможность самим создавать и размещать в ней контент и манипулировать им³.

Для работы со знаниями web 2.0 предусматривает:

- *mail.ru, yandex.ru, gmail.com* – электронная почта;
- *ICQ, google-чат* – системы обмена сообщениями;
- *Google Календарь* – планирование встреч, событий и дел, общение по электронной почте;
- *vkontakte.ru, moikrug* и др.– позволяют узнать о знаниях, опыте, интересах других людей и представить свои знания;
- *Google Docs* – коллективное редактирование документов в on-line режиме;
- другие.

Для структурирования знаний web 2.0 предлагает:

- *WebODE, OntoSaurus, WebOnto OilEd, OntoEdit*, и др. – редакторы онтоло-

³ В 2005 году Тим О’Рейли описал концепцию «эволюционировавшей всемирной паутины», которую он обозначил термином Web 2.0, подчеркнув тем самым ее значимость как следующего поколения интернет-систем.

гий;

- *Argument mappers* позволяет получать автоматически сгенерированное текстовое описание аргумента (простые шаблоны, содержащие аргумент с несколькими присоединенными словами, и относят к инструментам электронного обучения);
- *Mind mappers* позволяет структурировать информацию, выделить опорные понятия, тезисы и моделировать взаимосвязи между ними, что способствует лучшему усвоению и запоминанию.

Для управления знаниями в специализированных системах web 2.0 используют такие системы:

- *ARIS* – система, которая позволяет анализировать и моделировать деятельность предприятия и разрабатывать автоматизированные информационные системы;
- *ОРГ-МАСТЕР* – система бизнес-моделирования, обладающая возможностями по описанию и оптимизации деятельности предприятия;
- *Аналитический CRM* – система, позволяющая формировать отчетность и анализировать результаты маркетинговых мероприятий, эффективность продаж в разрезе продуктов, сегментов клиентов, регионов и т.д.;
- *SAP SRM* – система, предназначенная для планирования закупочных процедур и поисков источников поставки.

2. Понятие искусственного интеллекта

Термин «*интеллект*» (intelligence) происходит от латинского слова intellectus, что означает ум, рассудок, разум; мыслительные способности человека.

Интеллект – это способность мозга человека решать интеллектуальные задачи путем приобретения, запоминания и целенаправленного преобразования знаний в процессе обучения на опыте и адаптации к разнообразным обстоятельствам.

Четкого и однозначного определения понятия «*искусственный интеллект*» нет.

Под *искусственным интеллектом* будем понимать способность автоматизированных систем брать на себя отдельные функции интеллекта человека (например, выбирать и принимать решения на основе ранее полученного опыта и рационального анализа внешних воздействий).

Искусственный интеллект (ИИ), как научное направление, позволяет разрабатывать методы, решать интеллектуальные задачи и направлен на повышение эффективности различных форм умственного труда человека.

Основными направлениями использования ИИ являются:

- *разработка естественно-языковых интерфейсов и машинный перевод* – исследование методов и разработка систем, которые обеспечивают общение человека с компьютером на естественном языке;
- *генерация и распознавание речи* – разработка систем речевого общения позволит повысить скорость ввода информации, разгрузить зрение и руки;
- *обработка визуальной информации* – предназначены для решения задач обработки, анализа и синтеза изображений;
- *обучение и самообучение* – накопление и формирование знаний с использованием процедур анализа и обобщения данных;
- *распознавание образов* – распознавание объектов осуществляется на основании применения специального математического аппарата;
- *интеллектуальные роботы* – робот способный автономно интерпретировать знания, обрабатывать визуальную информацию, совершать операции по достижению поставленных целей;
- *программное обеспечение для разработки систем ИИ* – инструментальные средства для разработки интеллектуальных систем, ориентированные на обработку символьной информации (LISP, SMALLTALK, РЕФАЛ), языки логического программирования (PROLOG), языки представления знаний (OPS 5, KRL, FRL), интегрированные программные среды, оболочки экспертных систем, позволяющие создавать прикладные системы;

- *проектирование компьютеров новой архитектуры* – создание компьютеров, ориентированных на обработку символьной информации;
- *интеллектуальные информационные системы, основанные на знаниях*, предназначены для решения неструктурированных и слабоструктурированных задач⁴, разработки моделей представления знаний, извлечения, структурированных знаний и изучение проблемы создания баз знаний (БЗ).

Задачи, для которых алгоритм решения связан со сложными рассуждениями, логическими обобщениями и выводами, требующие изобретательности и высокой квалификации от человека, который решает ее, получили название *интеллектуальных задач*. Основой решения таких задач является интеллектуальный анализ данных и знаний.

3. Интеллектуальный анализ данных и знаний

Принятие разумных управленческих решений связано с аналитической обработкой огромных массивов данных и знаний. Для минимизации усилий лиц, принимающих решения (ЛПР), в процессе анализа данных последнее время широко стали использовать интеллектуальные *системы анализа данных* (ИСАД).

Интеллектуальный анализ данных (Data Mining) и знаний (Knowledge Discovery in Databases) – процесс обнаружения пригодных к использованию данных и знаний в сверхбольших объемах информации с применением методов математической статистики, теории баз данных, теории искусственного интеллекта и др.

Data Mining (DM) – исследование и обнаружение алгоритмами, средствами искусственного интеллекта в сырых данных скрытых знаний, которые

• ⁴ Условно все задачи делят на: *неструктурированные* - задачи имеют только качественное описание, основанное на суждениях лиц, принимающих решения, количественные зависимости между основными характеристиками задачи не известны; *структурированные* - задачи характеризуются зависимостями, которые могут быть выражены количественно; *слабоструктурированные* – задачи, которые сочетают количественные и качественные зависимости, при этом доминируют малоизвестные и неопределенные стороны задачи.

ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации.

Knowledge Discovery in Databases(KDD) – поиск полезных знаний в "сырых" данных, включающий подготовку данных, выбор информативных признаков, очистку данных, применение методов DM, постобработку данных и интерпретацию полученных результатов.

Типы задач интеллектуального анализа данных:

- *задачи прогнозирования* – исследование перспектив развития каких-либо событий, например, перспективы продаж, нагрузки сервера или времени простоя сервера и др.;
- *задачи маркетингового анализа* – обозначение проблем, поиск способов их решения, формирование базовой информации для принятия решений, например, определение зависимости спроса от стоимости самого товара, затрат на его продвижение и рекламу и т.д.;
- *задачи анализа работы с персоналом* – повышение мотивации к производительному труду, например, установление зависимости производительности труда работника от уровня подготовки, оплаты труда, опыта работы, взаимоотношений с руководством и т.д.;
- *задачи профилирования клиентов* – классификация перспективных клиентов, сотрудничество с которыми наиболее выгодно или стало неэффективным, выработка стратегии поиска подходящих клиентов в будущем;
- *задачи оценки потенциальных клиентов* – выявление характерных особенностей товара, способствующих реальным продажам, определение товаров, которые с высокой долей вероятности могут быть проданы вместе, выработка рекомендаций;
- *задачи сравнительного анализ* конкурирующих предприятий.
- *задачи поиска последовательностей* – определение часто встречающихся событий и прогнозирование следующего возможного события, например, анализ выбора заказчиков во время совершения покупок, прогнозирование следующего возможного заказчика;

- *задачи группирования* – разделение событий на кластеры связанных элементов, анализ и прогнозирование общих черт, например, образование социально-экономических типов явлений; изучение структурных изменений, происходящих в кластерах; выявление связи между изучаемыми признаками.

Процесс интеллектуального анализа данных (ИАД) включает следующие стадии:

- *выявления закономерностей* – поиск скрытых закономерностей без предварительного определения гипотез относительно вида этих закономерностей;
- *использования выявленных закономерностей* для прогнозирования, классификации объекта и для предположений о вероятном развитии процесса в будущем;
- *анализ исключений* – выявление аномалий в раскрытых закономерностях и их исключение.

Интеллектуальный анализ данных представляет собой сложный процесс обнаружения закономерностей в больших объемах данных, которые нельзя обнаружить традиционными способами из-за сложности связей и большого объема.

Современные методы искусственного интеллекта позволяют выделить следующие *типы закономерностей*:

Ассоциация – поиск закономерностей между несколькими событиями, которые происходят одновременно, осуществляется не на основе свойств анализируемого объекта, а на основе связанных событий. Данная закономерность используется для выработки рекомендаций – определение продуктов, которые с высокой степенью вероятности могут быть проданы вместе. Например, исследование, проведенное в крупнейших магазинах, показывает, что 70 % купивших зубную пасту берут также и зубную щетку, а при наличии скидки на зубные щетки приобретают в 90 % случаев. Располагая сведениями о подобной ассоциации, менеджерам легко оценить, насколько действенна предоставляемая скидка.

Последовательность – установление закономерностей между объектами (событиями), связанными во времени. Это необходимо для построения сценариев, например, такие закономерности могут быть использованы для анализа выбора заказчиков во время совершения покупок, прогнозирования следующего возможного события, например, после покупки верхней одежды в 45% случаев в течение недели приобретается новая обувь, а в пределах более недели 60 % покупателей приобретает головной убор. Фактически, ассоциация является частным случаем последовательности с временным шагом, равным нулю.

Классификация – разбиение множества событий (объектов) на априорно заданные группы (классы), элементами которых являются объекты, обладающие примерно одинаковыми свойствами. Классификация позволяет обобщить свойства объектов распознавания их принадлежности определенным совокупностям, например, для выбора наиболее подходящих заказчиков для целевой рассылки покупательских предложений.

Кластеризация – это распределение множества объектов по группам (кластерам) с одновременным определением признаков этих групп. Кластеризация, например, позволяет выявить характерные признаки покупателей, которых заинтересует предложения только брендовых товаров или только акционных товаров.

Прогнозирование – научное предвидение тенденций и перспектив дальнейшего развития объектов (событий) и их будущего состояния на основе знания закономерностей их развития в прошлом и в настоящее время. Прогнозирование позволяет, например, оценить объемы продаж или спрос, предположить нагрузки сервера или время его простоя на предстоящие периоды.

Все рассмотренные закономерности выявляются методами ИАД, которые можно сгруппировать по моделям ИАД.

Модель интеллектуального анализа данных – это набор алгоритмов, которые можно применить к данным для формирования на их основе новых знаний.

Основными моделями ИАД являются: *логическая, эволюционная, имитационная, структурная и статистическая.*

Логическая модель, в основе которой лежит булева алгебра и логический оператор IF (если), представляет собой цепочку доказательств теорем, для которых характерна большая трудоемкость, так как во время поиска доказательства необходим полный перебор вариантов. Хорошие результаты по данной модели получаются при сравнительно небольшом размере базы знаний. Исходные данные хранятся в базе данных в виде аксиом, а правила логического вывода – как отношения между ними. Основными методами для данной модели являются *деревья решений и нечеткая логика.*

Эволюционная модель предназначена для построения прогнозов состояний системы в условиях задания предыстории ее исходной организации. Для этого необходимо задание исходной организации системы, случайные мутации с ее элементами и отбор тех, которые оказывают положительное влияние на развитие системы с учетом некоторого критерия. Поиск такой структуры организации системы происходит в большей степени случайно и нецеленаправленно. Основными методами для данной модели являются *генетические алгоритмы.*

Имитационная модель, в основе которой понятие «черный ящик», т.е. внешнему наблюдателю доступны лишь входные и выходные величины, а структура и внутренние процессы неизвестны, и предназначена для получения подробной статистики о различных аспектах функционирования системы в зависимости от входных данных. Основными методами для данной модели являются: методы теории *массового обслуживания и оперативного планирования*, математические методы решения *дифференциальных уравнений, стохастические* и др.

Структурная модель строится по аналогии структуры человеческого мозга. Основными методами для данной модели являются *нейронные сети.*

Статистическая модель, в основе которой лежат базовые методы, позволяющие осуществлять предварительный анализ природы статистиче-

ских данных – проверка гипотез стационарности, нормальности, независимости, однородности, оценка вида функции распределения, ее параметров и т.п.– и выявлять связи и закономерности. Основными методами для данной модели являются: *корреляционный, регрессионный, кластерный, компонентный, факторный и дисперсионный анализы, динамические модели* и др.

Большинство из перечисленных методов были разработаны в рамках теории искусственного интеллекта и для них характерна точность, масштабируемость, интерпретируемость, проверяемость, трудоемкость, гибкость, быстрота и популярность. Но ни один метод не может обеспечить решение всего спектра задач, поэтому, используют одновременно несколько методов.

В настоящее время, как показывает опыт, знания людям помогают рационально организовывать свою деятельность и быстро решать различные проблемы, возникающие в её процессе, т.е. знания являются основным источником конкурентных преимуществ.

4. 4. Классификация систем искусственного интеллекта

Система искусственного интеллекта (СИИ) – автоматизированная информационная система для решения трудно формализуемых (слабоструктурированных) задач, для которых заранее не известен алгоритм решения.

СИИ представляет собой комплекс программных, лингвистических и логико-математических средств, предназначенный для поддержки деятельности человека в решении интеллектуальных задач.

В состав СИИ входят следующие компоненты:

База знаний (БЗ) – это совокупность фактов и правил вывода, допускающих логический вывод и осмысленную обработку информации.

Извлечение знаний – извлекает знания из БЗ, из неструктурированного текста, из графической информации, накапливать, передавать и преобразовывать для конструирования или расширения БЗ. Потенциальными источниками знаний являются эксперты, учебники, справочники, мультимедийные

документы, базы данных, специальные исследовательские отчеты и информация, доступная из сети Интернет.

Компонент диалогового общения использует БЗ, содержащую правила анализа и синтеза естественно-языковой или графической информации из проблемной области, как средство преобразования неформализованного задания в формализованное.

Компонент формирования цели обрабатывает формализованное задание и определяет возможность или невозможность его выполнения при существующих в данный момент ресурсах системы и состояний ее компонентов. Если система не может выполнить задание, то она формирует сообщение с объяснениями отказа и предложением скорректировать задание.

Компонент вывода на знаниях (машина логического вывода) выполняет поиск решения, используя БЗ, содержащую правила интерпретации знаний. Процесс поиска решения является итерационным, и на каждом шаге осуществляется коррекция модели среды с целью проверки правильности решений.

Компонент обработки внешней и внутренней информации выполняют анализ текущих изменений информации, для получения которой может быть использован искусственный интеллект (ИИ).

Компонент обучения и самообучения позволяет формировать дополнительные знания для базы знаний, которые получены в процессе обучения и самообучения и отсутствовали в ней по неизвестной причине.

Компонент контроля и диагностики выполняет проверку о наличии знаний в БЗ, диагностируют их и подсказывают правильные решения.

Системы искусственного интеллекта проникают во все сферы нашей жизни, поэтому трудно провести строгую их классификацию.

Рассмотрим *виды СИИ*, получивших наибольшее практическое применение.

СИИ, основанные на знаниях:

- *экспертные системы (ЭС)* – системы, основу которых составляет базы знаний с правилами вывода и которые на основании представленных фактов пользователем позволяют решить задачу в какой-либо предметной области, например, диагностика автомобилей, поиск медицинских препаратов и др;
- *системы распознавания образов* (нейрокомпьютерные системы) – системы, в основе которой лежит нейросеть в виде множества *процессоров*, соединенных между собой в соответствии с имеющейся моделью взаимодействия нейронных клеток человека, например, идентификация наземных объектов по результатам аэрофотосъемки из космоса;
- *робототехнические интеллектуальные системы* – системы, имитирующие человеческое восприятие окружающего мира и предназначены для автоматизации человеческого труда во вредных и опасных условиях его деятельности;
- *системы поддержки принятия решений (СППР)* – системы, позволяющие принимать решения по управлению объектами и процессами в сложных условиях с применением моделей человеческого мышления и поведения (например, выработка и принятие решений по устранению внештатной ситуации на борту космического корабля);
- *мультиагентные системы* – системы, позволяющие оперативно получать информацию из различных информационных систем и сетей и решать интеллектуальные задачи специального назначения. Примером таких систем могут быть *интеллектуальные агенты*, способные поддерживать взаимодействие с окружающей программной средой и получать от нее информацию, проявлять собственную инициативу, посылать и получать сообщения от других агентов, вступать с ними во взаимодействие, действовать без вмешательства извне, в том числе – и без вмешательства человека.
- *информационно-поисковые интеллектуальные системы (ИПС)* – системы, в основе которых лежат модели мышления и правила вывода на знаниях, и которые способны формировать соответствующие ответы на запросы пользователя, не умеющего четко сформулировать задание;
- *обучающие системы (тьюторы)* – системы, имеющие в своей базе знаний

все необходимые знания для организации процесса обучения;

- *генетические системы* базируются на концепциях эволюции (естественного отбора) и используют механизм генетического наследования в природе. В основу функционирования генетических систем положены эволюционные алгоритмы генетического поиска;
- *нечеткие интеллектуальные системы* – системы, предназначенные для решения задач с недостоверной и неполной информацией об исследуемых объектах, процессах и явлениях окружающего мира с помощью метода нечеткой логики и позволяющих формализовать нечеткие знания и обрабатывать их.
- *системы управления знаниями* предназначены для поиска, анализа, распространения, использования и обработки знаний об отдельно взятом предприятии. Основой таких систем является корпоративная память, которая получает информацию из различных источников и позволяет избежать повторов и дублирований;
- *онтологические системы* представляют собой иерархические структуры баз знаний специального назначения и предназначены для поиска, извлечения и накопления концептуальных знаний, получаемых из различных систем и сетей, главным образом – из сети Internet. Они представляют пользователям более детальную, обширную и релевантную (запросу) информацию о различных областях человеческой деятельности.

Самоорганизующиеся СИИ, обладающие свойством изменяться в целях самосовершенствования:

- *робототехнические интеллектуальные системы*;
- *нейросистемы* – системы, в основе которых лежат: простейшая нейронная сеть (совокупность нейронов, организованных слоями), архитектура сети и сила связи, принципы управления динамикой нейронов не известны и предназначены для нахождения оптимальных решений при нарастании объемов информационных потоков данных;

Системы эвристического поиска:

- *робототехнические интеллектуальные системы;*
- *системы распознавания образов;*
- *системы общения* – системы, позволяющие устанавливать и развивать контакты, порожденные потребностями совместной деятельности, между людьми и группами. К ним относят системы *обработки текстов, речевого общения, машинного перевода и генерации музыки;*
- *игровые системы* – системы на основе заранее заложенных правил определяют поведение игровых объектов и с учетом разнообразия действий конечный результат может быть неявной поведенческой системой, например, *Mafia, Казино онлайн и др.*

Где в Вашей классификации находятся описанные ниже системы: **VI-системы, интеллектуальные агенты, Big Data системы, системы аналитической обработки данных?**

Для экономических исследований наибольшее распространение получили такие СИИ как экспертные системы, системы поддержки принятия решений, VI-системы, интеллектуальные агенты, Big Data системы, системы аналитической обработки данных, которые рассмотрим далее подробно.

5. Экспертные системы

Экспертная система (ЭС) – это система искусственного интеллекта для решения слабоструктурированных задач на основе знаний экспертов в предметной области.

Основой экспертной системы является база знаний, которая составляется из формализованных знаний экспертов (экспертных знаний), сочетающих теоретическое понимание проблемы и практические навыки ее решения.

ЭС позволяет решать задачи, для которых характерны следующие особенности:

- параметры задачи не могут быть заданы полностью в числовой форме;
- целевую функцию нельзя выразить в точных терминах;
- не существует четкого алгоритмического решения задачи;

- исходные данные отличаются неоднозначностью, ошибочностью, избыточностью и противоречивостью;
- ресурсы (время, память), используемые для решения задачи, ограничены.

Особенности работы экспертных систем:

- ЭС моделируют не физические процесс конкретной предметной области, а механизм мышления человека применительно к решению задач в этой области;
- ЭС помимо выполнения вычислительных операций, формируют определенные соображения и выводы, основываясь на тех знаниях, которыми располагают;
- для решения задач ЭС используют эвристические и приближенные методы, которые, в отличие от алгоритмических, не всегда гарантируют успех, т.к. не требуют исчерпывающей исходной информации, но обеспечивают определенную степень уверенности в верности полученного решения;
- у ЭС нет предубеждений;
- ЭС не делают поспешных выводов;
- ЭС работают систематизировано, рассматривая все детали, часто выбирая наилучшую альтернативу из всех возможных;
- способность к символьным рассуждениям, т.е. представление знаний и отношений между ними в символьном виде;
- компетентность в конкретной предметной области. ЭС должна иметь тот же уровень профессионализма, что и эксперты, применять знания эффективно и быстро;
- способность проверять правильность своих рассуждений и учиться на своих ошибках;
- работать в предметной области, где слабоструктурированные задачи.

Основными областями применения экспертных систем являются: медицина, военное дело, геология, инженерное дело, информатика, компьютерные системы, космос, математика, метеорология, промышленность, сельское хо-

зайство, управление процессами, физика, химия, электроника, юриспруденция и т.д.

С помощью экспертных систем решаются задачи:

- *диагностики*, например, диагностика и терапия сужения коронарных сосудов, диагностика ошибок в аппаратуре и математическом обеспечении ПК и др.;
- *интерпретации данных* (определение смысла данных, результаты которого должны быть согласованными и корректными), например, обнаружение и идентификация различных типов океанских судов, определение основных свойств личности по результатам психодиагностического тестирования и др.;
- *мониторинга* (непрерывная интерпретация данных в реальном масштабе времени и сигнализация о выходе тех или иных параметров за допустимые пределы), например, контроль за работой электростанций, помощь диспетчерам атомного реактора, контроль аварийных датчиков на химическом заводе и др.;
- *проектирования* (подготовка спецификаций на создание объектов с заранее определенными свойствами), например, проектирование конфигураций ПК, проектирование банковских информационных систем, синтез электрических цепей и др.
- *прогнозирования* (вывод вероятных следствий из заданных ситуаций), например, предсказание погоды, оценка будущего урожая, прогнозы в экономике и др.
- *планирования* (нахождение планов действий для объектов, чтобы просчитать последствия планируемой деятельности), например, планирование поведения робота, планирование промышленных заказов и экспериментов и др.
- *обучения* (диагностируют ошибки при изучении какой-либо дисциплины с помощью ПК и подсказывают правильные решения), например, обучение языкам программирования и др.

Обобщенная структура экспертной системы представлена на рис. 1. Реальные экспертные системы имеют более сложную структуру, чем обобщенная структура ЭС.

Механизм логического вывода или решатель предназначен для получения новых фактов на основе сопоставления исходных данных из рабочей памяти и знаний из базы знаний.

Рабочая память или база данных (БД) предназначена для хранения исходных и промежуточных фактов решаемой задачи.

База знаний (БЗ) предназначена для хранения долгосрочных фактов, описывающих рассматриваемую область, правил, описывающих отношения между этими фактами и других типов декларативных знаний о предметной области, множество функций и процедур, реализующих расчетные и другие алгоритмы. Экспертные системы основываются на знаниях специалиста-эксперта о предметной области.

Подсистема приобретения и пополнения знаний автоматизирует процесс наполнения экспертной системы знаниями, осуществляемый пользователем-экспертом, и адаптации базы знаний системы к условиям ее функционирования. Адаптация экспертной системы к изменениям в предметной области реализуется путем замены правил или фактов в базе знаний.

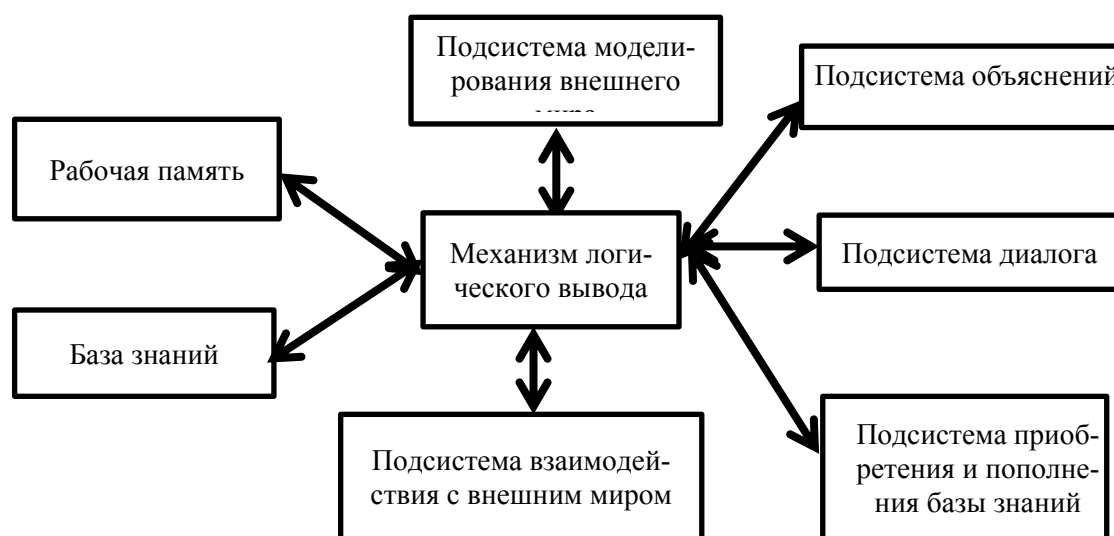


Рис. 1. Обобщенная структура экспертной динамической системы

Подсистема объяснения объясняет, как система получила решение задачи (или почему она не получила решения) и какие знания она при этом использовала, что облегчает эксперту тестирование системы и повышает доверие пользователя к полученному результату, повышается доверие пользователей к полученным результатам;

Подсистема диалога предназначена для организации дружественного интерфейса со всеми категориями пользователей как в ходе решения задач, так и в ходе приобретения знаний и объяснения результатов работы.

ЭС может работать в двух режимах: *приобретения знаний* и *консультации*.

В режиме приобретения знаний эксперт через подсистему приобретение и пополнение наполняет БЗ формализованными знаниями в виде совокупности правил и данных. Данные определяют объекты, их характеристики и значения, существующие в области экспертизы, правила – способы манипулирования данными, характерные для рассматриваемой области.

Пользователь (лицо, принимающее решение – ЛПР) *в режиме консультации* самостоятельно (без эксперта) решает задачи из проблемной области. При этом исходные данные задачи через подсистему диалога поступают в рабочую память. Механизм логического вывода на основе исходных данных и общих сведений о проблемной области из БЗ формирует решение задачи.

Классификация экспертных систем

ЭС можно классифицировать по различным признакам. Рассмотрим наиболее распространенные из них:

по решаемым задачам:

- *диагностические ЭС* – выполняют обнаружение неисправностей в системе. Примеры: MYCIN диагностика и наблюдение за состоянием больного при менингите и бактериальных инфекциях, ANGY – диагностика и терапия сужения коронарных сосудов; CRIB – диагностика ошибок в аппаратуре и математическом обеспечении ЭВМ и др.;

- *проектирующие ЭС* – системы предназначены для достижения конкретных целей при решении задач с большим числом переменных. Среди них наиболее распространенными: VAX – изменяет конфигурацию компьютерных систем, XCON – изменяет конфигурацию компьютерных систем в соответствии с требованиями покупателя, CADHELP – проектирование БИС, SYN – синтез электрических цепей и др.
- *планирующие ЭС* – системы предназначены для планирования деятельности. Среди них наиболее распространенными: STRIPS – планирование поведения робота, ISIS – планирование промышленных заказов, MOLGEN – планирование эксперимента и др.
- *прогнозирующие ЭС* – предсказывают возможные результаты или события на основе данных о текущем состоянии объекта. Например, PI.ANT – оценки будущего урожая, WILLARD – предсказание погоды, ECON – прогнозы в экономике, «Завоевание Уолл-стрита» – анализ конъюнктуру рынка и с помощью статистических методов разработка плана капиталовложений на перспективу и др.
- *интерпретирующие* – системы позволяют получать определенные заключения на основе результатов наблюдения. Например, PROSPECTOR объединяет знания многих (9 экспертов) экспертов; HASP/SIAP определяет местоположение и идентифицирует типы судов в океане по данным акустических систем слежения; АВТАНТЕСТ и МИКРОЛЮШЕР определяет основные свойства личности по результатам психодиагностического тестирования;
- *контролирующие* – непрерывная интерпретация данных в реальном масштабе времени и сигнализация о выходе тех или иных параметров за допустимые пределы. Например, СПРИНТ – контроль за работой электростанций, REACTOR – помощь диспетчерам атомного реактора, FALCON – контроль аварийных датчиков на химическом заводе и другие.
- *обучающие системы ЭС* – системы диагностируют ошибки при изучении какой-либо дисциплины с помощью ПК и подсказывают правильные решения. Например, "Учитель Лиспа" – обучение языку программирования,

PROUST - обучение языку Паскаль и др. Обучающие системы также работают на атомных электростанциях, управляют воздушным движением и осуществляют медицинский контроль, могут быть полезны при регулировании финансовой деятельности предприятия и оказывать помощь при выработке решений.

по связям с реальным миром ЭС:

- *статические* используются в приложениях, где можно не учитывать изменения реального мира, происходящие за время решения задачи, например, программы диагностики неисправностей в автомобиле и др.;
- *квазидинамические* интерпретируют ситуацию, которая меняется с некоторым фиксированным интервалом времени, например, микробиологические ЭС, в которых снимаются лабораторные измерения с технологического процесса один раз в 4-5 часов;
- *динамические* работают в сопряжении с датчиками объектов в режиме реального времени с непрерывной интерпретацией поступающих в систему данных. Например, системы управления гибкими производственными комплексами и мониторинга реанимационных палат и т.д.

по степени интеграции с другими программами:

- *автономные ЭС* работают в режиме консультаций с пользователем и для решения их не требуется привлекать традиционные методы обработки данных. Например, ACES выполняет картографические работы по нанесению обстановки на карты, ASTA помогает аналитику определить тип радара, пославшего перехваченный сигнал анализирует этот сигнал в свете имеющихся у нее общих знаний о физике радаров и специальных знаний о конкретных типах радарных систем, DART помогает обрабатывать разведанные данные о центрах командования, управления и связи противника и дает советы аналитикам по идентификации критических узлов сети командования, управления и связи и помогает обрабатывать сообщения о боевой обстановке, ExpSystem PC помогает пользователю подобрать приемлемую конфигурацию будущего компьютера и др.;

- *гибридные ЭС* представляют программный комплекс, агрегирующий стандартные пакеты прикладных программ (например, математическую статистику, линейное программирование или системы управления базами данных) и средства манипулирования знаниями. Это может быть интеллектуальная надстройка над ППП или интегрированная среда для решения сложной задачи с элементами экспертных знаний.

Большинство ЭС можно отнести одновременно к нескольким типам. Так, *обучающая* система может выполнять *диагностику* и *планирование*: определяет способности обучаемого по основным темам курса, а затем составляет учебный план с учетом полученных данных. *Управляющая* система может применяться для целей *контроля*, *диагностики*, *прогнозирования* и *планирования*.

6. Системы поддержки принятия решений

С ростом масштабов и сложности деятельности предприятий, с внедрением информационных технологий повышаются требования к качеству управленческой деятельности. Основу любой управленческой деятельности составляют решения, принимаемые органами управления на единоличной или коллегиальной основе и направленные на решение задач, стоящих перед предприятием.

На сегодняшний день не существует единого определения понятия «*система поддержки принятия решений (СППР)*». Авторы по-разному трактуют это понятие:

- *СППР* – это интерактивная автоматизированная система, помогающая лицу, принимающему решения, использовать данные и модели для решения слабо структурированных задач.
- *СППР* – это система, которая обеспечивает пользователям доступ к данным и/или моделям, так что они могут принимать лучшие решения.

В рамках данной работы под *системой поддержки принятия решений* будем понимать интерактивную интеллектуальную ИС, предназначенную

для поддержки многокритериальных решений в сложной информационной среде.

Информационная сложность среды, в которой принимается решение, обусловлена необходимостью учета большого объема данных, обработка которых без современных информационных технологий практически невозможна. В этих условиях число возможных решений велико, и выбор приемлемого решения без всестороннего анализа может привести к неправильным решениям.

Системы поддержки решений СППР решают две **основные задачи**:

- *оптимизация* – выбор наилучшего решения из множества возможных, например: выбрать типовой план подготовки мероприятия из существующих шаблонов; сформировать список потенциальных участников обучения из базы данных клиентов; подготовить и разослать текст индивидуального письма; накапливать и анализировать статистику поведения клиентов и др.
- *ранжирование* – упорядочение возможных решений по предпочтительности. Например, СППР, основанная на методе анализа иерархий позволяет структурировать проблему, построить набор альтернатив, выделить характеризующие их факторы, задать значимость этих факторов, оценить альтернативы по каждому из факторов, найти неточности и противоречия в суждениях ЛПР/эксперта, проранжировать альтернативы, провести анализ решения и обосновать полученные результаты. Такая система используется при: оценке качества организационных, проектных и конструкторских решений; определении политики инвестиций в различных областях; распределении ресурсов; проведении анализа проблемы по методу «стоимость-эффективность»; стратегическом планировании; проектировании и выборе оборудования, товаров и др.

Особенности работы в СППР:

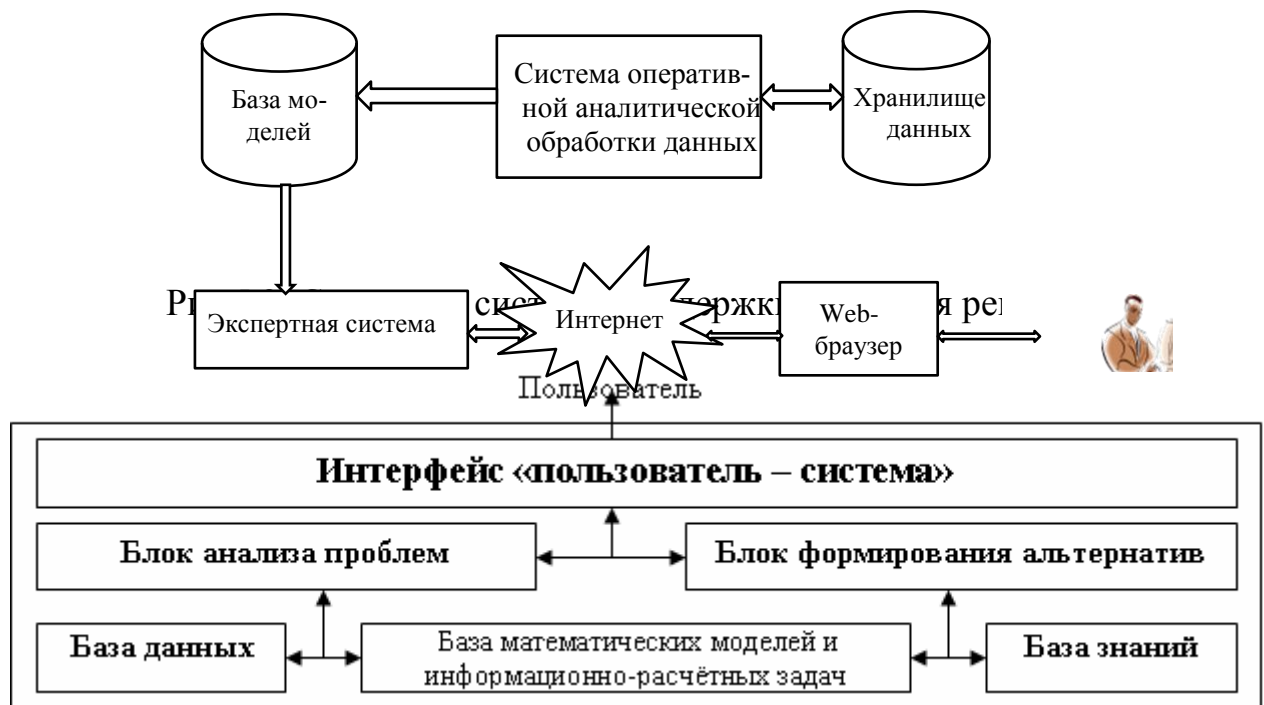
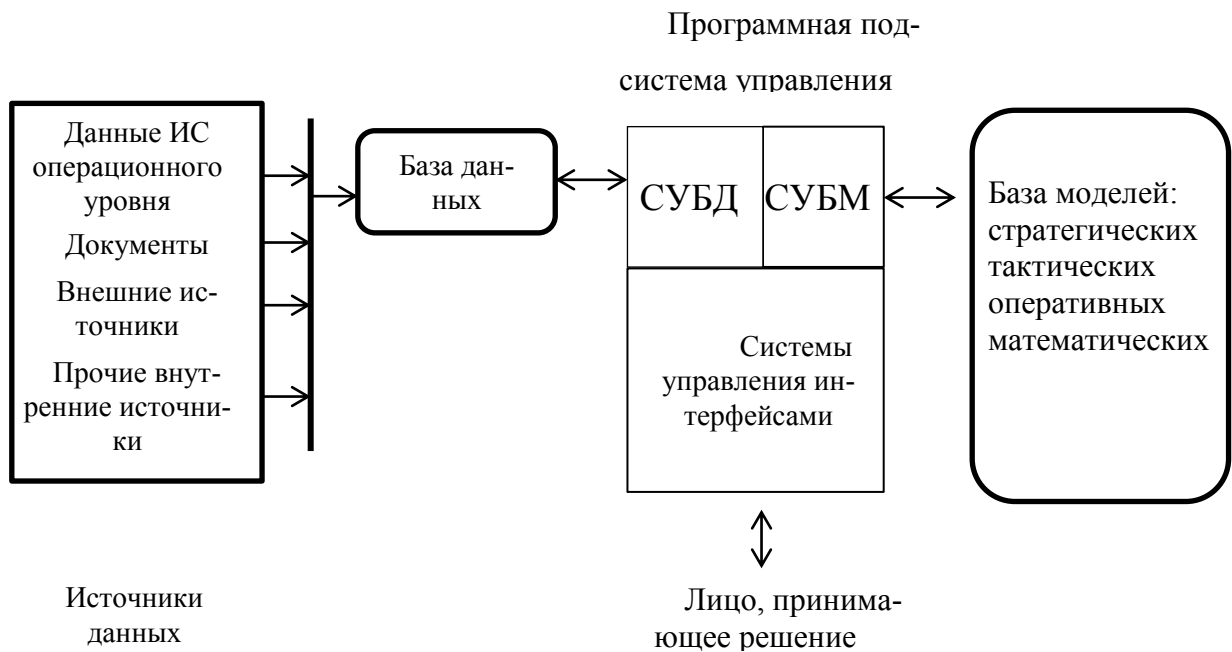
- умение работать с неструктурированными и слабоструктурированными задачами;

- необходимость выбора совокупности критериев, на основе которых оцениваются и сопоставляются возможные решения (альтернативы);
- нетребовательность к квалификации пользователя;
- допускают, чтобы пользователи управляли входом и выходом;
- ускоряют решение проблем в бизнесе;
- способствуют налаживанию контактов внутри организации;
- могут быть использованы менеджерами различного уровня управления (от высшего руководства до руководителей оперативного уровня управления);
- только поддерживают выбранную из возможных альтернатив, а не заменяют окончательный выбор альтернативы менеджером;
- интерактивность и др.

Области применения СППР:

- *торговля* – планирование закупок и хранение, анализ совместных покупок, поиск шаблонов поведения во времени;
- *страховой бизнес* – выявление потенциальных случаев мошенничества, анализ риска, классификация клиентов;
- *банковское дело* – оценка риска кредитования, прогнозирование изменений клиентуры, классификация клиентов, выделение групп клиентов со сходными потребностями, выявление случаев мошенничества и др.);
- *телекоммуникации* – подготовка и принятие комплекса решений, направленных на сохранение своих клиентов и минимизацию их оттока и др.;
- *медицина* – диагностика заболеваний, оценка диагностических тестов, выявление побочных эффектов;
- *бизнес* – построение рекомендательных систем для персонализации пользователей web-сайтов, с целью повышения лояльности покупателей, и, как следствие, повышение продаж, выявление случаев мошенничества и т.д.;
- *промышленное производство* – прогнозирование качества производимого изделия, планирование ремонтов и т.д.) и др.

Обобщенная структура СППР представлена на рис. 2.



Таким образом, перечень основных функций СППР выглядит следующим образом:

1. Извлечение знаний;
2. Верификация знаний;
3. Вывод рекомендаций;
4. Объяснение рекомендаций.

Всё это позволяет нарисовать самую обобщённую функциональную архитектуру СППР в следующем виде:



Где описание всех компонентов?

Для анализа и выработок предложений в СППР используются разные методы – информационный поиск, интеллектуальный анализ данных, поиск знаний в базах данных, рассуждение на основе прецедентов, имитационное моделирование, эволюционные вычисления и генетические алгоритмы, нейронные сети, ситуационный анализ, когнитивное моделирование и др. Некоторые из этих методов были разработаны в рамках искусственного интеллекта и те системы, в основе которых лежат методы искусственного интеллекта, получили название интеллектуальной СППР (ИСППР).

Классификация СППР

По взаимодействию СППР с пользователем:

- *пассивные СППР* – помогают в процессе принятия решений, но не могут выдвинуть конкретного предложения, например: *Executive Information System* предоставляет наборы отчетов, построенных на основании данных из транзакционной информационной системы предприятия и отражающих основные аспекты производственной и финансовой деятельности в режиме реального времени
- *активные СППР* – помогают структурировать проблему, построить набор

альтернатив, выделить характеризующие их факторы, задать значимость этих факторов, оценить альтернативы по каждому из факторов, про ранжировать альтернативы, провести анализ решения и обосновать полученные результаты. Примеры: *Выбор* – аналитическая система, помогающая кроме перечисленных возможностей СППР, найти неточности и противоречия в суждениях ЛПР; *Assistant Choice* – система многокритериального выбора оптимальных решений, позволяющая построить математическую модель для поставленной задачи, которая включает множество альтернатив, множество критериев выбора, функцию предпочтения пользователя и информацию о важности критериев;

- *кооперативные СППР* – предполагают взаимодействие с пользователем – полученное решение пользователь может доработать, усовершенствовать, а затем отправить обратно в систему для проверки и так до получения согласованного решения. Пример: *Combi-PC* позволяет анализировать различные виды продукции, варианты планов, мероприятий, исполнителей при этом пользователь может конструировать метод решения своей задачи как цепочку этапов или подбирать готовый метод из библиотеки.

По сфере использования:

- *автономные СППР* – небольшие системы, которые подходят для управления с персонального компьютера одного пользователя. Например, Microsoft Excel позволяет создать модели для прогнозирования различных элементов организации или финансового состояния. В качестве данных используются финансовые отчеты организации. Начальная модель включает различные предположения относительно будущих трендов в категориях расхода и дохода. После рассмотрения результатов базовой модели менеджер проводит ряд исследований типа «Что, если...?», изменяя одно или большее количество предположений, чтобы определить их влияние на исходное состояние.
- *сетевые СППР* – работают с большими хранилищами данных и используются многими пользователями. Например: Expert Choice decision portal «(ECDP), разработанная только для использование через сеть Интернет, и

позволяет сделать выбор между несколькими альтернативами, основанными на множестве критериев принятия решений и различных атрибутов, EXSYS используется для разработки экспертных систем для оказания консультативных услуг лицам, принимающим решения, и др.

заменила на

СППР по управляемым ими объектам:

- *управляемые сообщениями* (Communication-Driven DSS) – поддерживают группу пользователей, работающих над выполнением общей задачи. Пример: PolyAnalyst, NeuroShell, GeneHunter, BrainMaker, OWL, 4Thought;
- *управляемые данными* (Data-Driven DSS) – ориентированы на доступ к данным, работу и манипуляцию с ними, например, Express Server имеет встроенный язык манипулирования и обработки многомерных данных Express Language (4GL), обеспечивающих построение простых как статистических так и имитационных моделей; Express Spreadsheet Add-In позволяет использовать электронные таблицы MS Excel версии 7.0 и выше в качестве среды разработки клиентских OLAP-приложений.
- *управляемые документами* (Document-Driven DSS); – осуществляют поиск и манипулируют неструктурированной информацией в различных форматах. Пример: *Document management* – управляют, осуществляют поиск и манипулируют неструктурированной информацией, заданной в различных форматах;
- *управляемые знаниями* (Knowledge-Driven DSS); – обеспечивают решение задачи в виде фактов, правил, процедур. Пример: (дополнила) NORBIT SAP BI4Retail – система стратегического анализа данных для принятия управленческих решений, предоставляющая всесторонний анализ информации в режиме реального времени из различных источников данных, возможность формирования аналитических отчетов, за любой период времени и настраиваемой детализацией; SAP BI – обеспечивает анализ бизнес-процессов для принятия качественных управленческих решений как для краткосрочной, так и для долгосрочной перспектив;

- *управляемые моделям (Model-Driven DSS)* – обеспечивают доступ и манипуляцию математическими моделями (статистическими, финансовыми и др.). Например, *Express Objects* - профессиональная инструментальная среда для визуальной объектно-ориентированной разработки OLAP-приложений для архитектуры клиент-сервер под MS Windows. Express Objects поддерживает необходимые объектные механизмы: инкапсуляцию, наследование и полиморфизм
- *гибридные* – обеспечивают моделирование, поиск и обработку данных. Пример: *OLAP-системы* позволяют осуществлять сложный анализ данных, включая поиск, обработку и моделирование данных.

По уровню управления предприятием:

- *СПИП оперативного уровня* или их называют «Информационные Системы Руководства» (Executive Information Systems - EIS) – представляют набор конечных отчетов, созданных на основе данных из транзакционной информационной системы предприятия, и отражающих главные аспекты финансовой и производственной деятельности предприятия в режиме реального времени. Пример: *Project Expert* позволяет моделировать деятельность предприятия, пригодна как для инвестиционного, так и оперативного бизнес-планирования, может быть использована для оценки финансового состояния предприятий, принятия решений при их обращении о предоставлении налоговых льгот, об участии в финансировании строительства нового производства и пр.; *Альт-Инвест* осуществляет подготовку финансовых разделов технико-экономического обоснования и бизнес-планов, моделирование и оптимизацию схемы осуществления проекта, проведение экспертизы инвестиционных проектов; *Marketing Expert* обеспечивает поддержку принятия решения на всех этапах разработки тактического плана маркетинга и контроля за их реализацией, используя для этого GAP-анализ, SWOT –анализ, Portfolio-анализ.
- *СПИП стратегического уровня* предполагают глубокую проработку информации, на основании которой осуществляется планирование и управле-

ние всеми ресурсами предприятия. Например, *ERP, GIS, DocFlow, Business Modeller, SCADA/ DCE, Project Management* и др.

В настоящее время разработано достаточно много СППР. Широкое применение получили такие СППР:

IBM Watson Analytics – позволяет автоматизировать анализ и визуализации данных, обеспечивает быстрое и удобное получение новых знаний, предлагается как настольное приложение, так и мобильная версия для iPad;

• *IBM Cognos Insight Personal Edition* – позволяет исследовать, анализировать и наглядно представлять данные на ПК;

• *IBM Cognos Express* – позволяет формировать отчетность, проводить анализ, планирование, бюджетирование;

• *IBM Cognos Business Intelligence* – позволяет быстро и легко собрать воедино всю значимую информацию, чтобы понять положение дел на предприятии;

• *Парус Бюджет 8* – позволяет поэтапное построение комплексной системы управления предприятием, организует обмен данными с системой автоматизации документооборота на базе Lotus Notes, включает модули: управление финансами, материально—техническое обеспечение, сведение отчетности, штаты, тарификация и расчет оплаты труда и специализированные модули;

• *SAP Business objects Enterprise* – позволяет бизнес-пользователям самостоятельно решать задачи, предоставляет возможность использования бизнес-аналитики всем пользователям - как внутри, так и за пределами компании, проводить интеграцию с существующей ИТ-инфраструктурой.

Несмотря на то, что СППР и ЭС как конкретные программные продукты в ряде случаев выглядят внешне одинаково, они имеют существенное различие.

Отличие СППР от ЭС:

• СППР учитывает личные предпочтения ЛПР, его субъективное видение при решении стоящей перед ним задачи, а ЭС опирается при решении задачи на знания экспертов, т.е. специалистов в конкретной предметной области;

- СППР используют для принятия решений не только знания и аналитические данные, но и оперативные данные, сопровождающие деятельность предприятия и хранящиеся в БД, тогда как ЭС для принятия решений используют только знания.

7. Системы **Business Intelligence**

Рост конкуренции и высокая экономическая нестабильность диктуют новые требования к качеству принимаемых управленческих решений на предприятии. Такие решения должны учитывать все аспекты ведения бизнеса, основываться на достоверной информации и должны приниматься в сжатые сроки.

Для выработки управленческих решений предприятию необходим механизм, позволяющий объединить данные из всех его информационных систем и анализировать их.

Бизнес-анализ как деятельность – это процесс, который включает интеллектуальный анализ данных, аналитическую обработку их в режиме on-line, получение информации из баз данных и составление отчетов.

Business intelligence (BI) – это методы и инструменты для перевода необработанной информации в осмысленную и удобную форму.

BI-система – это совокупность технологий и инструментов необходимых для получения нужной информации в нужное время и которая позволяет собрать «разбросанные» по различным источникам данные, построить единое хранилище информации и предоставить удобные средства построения отчетов и проведения анализа. В результате, у пользователей BI-системы появляется целостная представление о деятельности предприятия, позволяющая им в кратчайшие сроки принимать нужные управленческие решения. BI-системы адаптируются к бизнесу, в отличие от учетных систем, которые его структурируют.

Структура BI-систем включает такие основные компоненты (см. рис. 3):

среда интеграции и хранения информации: интеграция данных и создание семантического слоя – выгрузка данных из учетных систем, проверка, очистка, обработка данных, загрузка данных в хранилищ; хранение данных – хранилища и витрины данных, многомерные (OLAP) базы данных;

среда мониторинга и анализа информации: панели индикаторов – объединение средств отчетности и анализа; система отчетности – создание, управление и доставка параметризованных статических и интерактивных отчетов конечному пользователю; система оперативной аналитической обработки (OLAP); средства интеллектуального анализа данных (Data Mining).

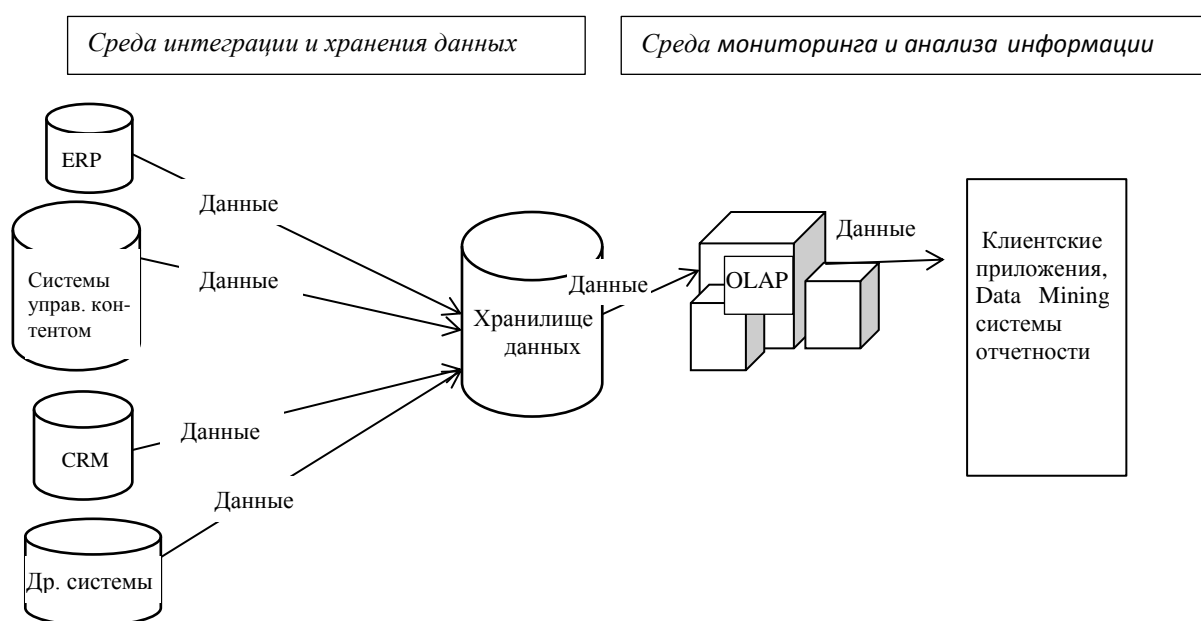


Рис. 3. Обобщенная структура BI-системы

BI-система включает BI-инструменты и BI-приложения.

BI-инструменты – это генераторы запросов и отчетов, а развитые BI-инструменты – это прежде всего инструменты оперативной аналитической обработки, корпоративные BI-наборы и BI-платформы. Средства генерации запросов и отчетов в основном на современном этапе развития информационных технологий поглощаются и замещаются корпоративными BI-наборами. Многомерные OLAP-механизмы или серверы, а также реляционные OLAP-механизмы являются BI-инструментами и инфраструктурой для BI-платформ. Большинство BI-инструментов применяются конечными пользователями для доступа, анализа и генерации отчетов по данным. Разработ-

чики приложений используют BI-платформы для создания и внедрения BI-приложений. Примером BI-приложения является информационная система руководителя (EIS).

BI-платформы предлагают наборы инструментов для создания, внедрения, поддержки и сопровождения BI-приложений. BI-приложения обычно ориентированы на конкретную функцию организации или задачу – анализ и прогноз продаж, финансовое бюджетирование, прогнозирование, анализ рисков, анализ тенденций и т.п. В них встроены BI-инструменты (OLAP, генераторы запросов и отчетов, средства моделирования, статистического анализа, визуализации и data mining). Многие BI-приложения извлекают данные из ERP-приложений.

BI-систем используются для решения таких бизнес-задач как:

- формулировка иерархии показателей эффективности;
- сопоставительный анализ на основе эталонных показателей (бенчмаркинг), что позволяет информировать руководителей предприятий о достижениях бизнес-целей;
- наглядность и удобство работы с информацией для бизнес-пользователей;
- поддержка развития бизнес-процессов и структурных изменений предприятия;
- возможность моделирования бизнес-ситуаций в единой информационной среде;
- проведение оперативного анализа по нестандартным запросам;
- снижение рутинной нагрузки на персонал;
- устойчивая работа при увеличении объема обрабатываемой информации;
- оценка эффективности различных направлений бизнеса;
- оценка достижимости поставленных целей;
- оценка эффективности использования ресурсов на предприятии;
- оценка эффективности операционной, инвестиционной и финансовой деятельности;

- бизнес-моделирование и оценка инвестиционных проектов;
- управление затратами, налоговое планирование, планирование капитальных вложений;
- возможность использования нескольких аналитических решений для различных направлений деятельности в масштабах всего предприятия;
- способность извлекать, анализировать и консолидировать данные практически из любых источников.

Наиболее известными BI-системами являются:

- *SAP BI* (SAP Business Objects) – предоставляет доступ к информации, содержащейся в различных системах или базах данных предприятия, что позволяет решать задачи финансирования, бюджетирования, казначейства и бизнес-анализа. На базе этого программного продукта реализованы такие решения – бизнес-аналитика для оптовой торговли и дистрибуции, NORBIT SAP BI4Retail - бизнес-аналитика для розничной торговли (ритейл, FMCG).
- *SAP BPC* (SAP Business Planning and Consolidation) – позволяет автоматизировать планирование, бюджетирование и прогнозирование на предприятии, анализу и консолидацию данных, формирование отчетов.
- *QlikView* – выполняет всесторонний анализ процесса продаж, клиентской базы, и оценку эффективности взаимодействия с каждым клиентом, анализа, области потенциальных сделок, принятых заказов, их оплаты и выполнения, структурный анализ данных произвольного периода, динамики и сезонности изменения показателей с течением времени, а также ассортиментный, рейтинговый ABC-XYZ и Парето-анализ во всевозможных аналитических разрезах.
- *Microsoft Power BI* (входит в составе Office 365) – выполняет сбор, упорядочение и визуализацию данных, выделяет основные задачи.
- *Roistat* – собирает данные из CRM-системы, рекламных площадок и сайта компании и на их основе формирует отчеты по основным бизнес-показателям.

- *Ваш финансовый аналитик* (on-line сервис) – выполняет финансовый анализ, формирует аналитические отчеты по данным налоговых отчетов, настраивать шаблоны и поддерживает протокол SSL.
- *Seneco* (облачный сервис) – служит для управления финансами: управленческий учет, финансовая аналитика и планирование денежных средств.
- *Watson Analytics* – позволяет собирать данные, обрабатывать их в облаке, структурировать и выдавать в удобочитаемом виде.
- *Microsoft Assistant* выполняет бизнес анализ и управление запасами в малых и средних компаниях, занимающихся продажей товаров, интегрируется с системами учета, помогает спланировать оптимальный график закупок и продаж, выявить выгодных клиентов и менеджеров.
- *IBM Cognos Business Intelligence* формирует обобщенную систему процесса принятия решений для всего предприятия, раскрывает взаимосвязь всей предыдущей деятельности с текущей и позволяет найти наилучшее решение, преобразующее стратегию в действие.
- *HandyKPI* (on-line сервис) – выполняет контроль ключевых показателей бизнеса.

8. Интеллектуальные агенты

Слово агент происходит от латинского слова «аґеґе» – двигатель. В настоящее время однозначного определения данного понятия не существует, но многие ученые согласны с определением М.Дж. Вуприджа и М.Дж. Дженнингса: «агент – это компьютерная система, помещенная в некоторую среду и обладающая способностями к автономным действиям в этой среде, для того чтобы обеспечить достижения намеченной в ней целей» В искусственном интеллекте под термином *У* понимаются сущности, получающие информацию через систему сенсоров о состоянии управляемых ими процессов и осуществляющие влияние на них.

Интеллектуальный агент – программа, самостоятельно выполняющая задание, указанное пользователем, в течение длительных промежутков времени.

Интеллектуальный агент обладает следующими свойствами:

- *автономность* – способность функционировать без вмешательства со стороны и осуществлять контроль внутреннего состояния и своих действий;
- *адаптивность* – способность обучаться;
- *способность к рассуждениям* – владение механизмами вывода;
- *коммуникативность* – умение общаться с другими агентами;
- *мобильность* – способность перемещаться в электронной среде;
- *реактивность* – способность реагировать на изменение исходных данных;
- *активность* – способность генерировать цели и действовать рациональным образом для их достижения;
- *владение базовыми знаниями* – наличие знаний о себе, об окружающей среде, в том числе о других агентах;
- *наличие убеждений* – способность менять базовые знания во времени;
- *наличие цели* – совокупность состояний, на достижение которых направлено текущее поведение агента;

- *наличие желаний* – состояния и/или ситуации, достижение которых для агента важно;
- *наличие обязательств* – задачи, которые берет на себя агент по просьбе и/или поручению других агентов;
- *правдивость* – агент не будет сознательно сообщать ложную информацию;
- *благожелательность* – агенты не имеют антагонистических целей, и поэтому каждый агент будет всегда стараться выполнить то, что запросили от него другие агенты;
- *рациональность* – агент действует для достижения целей согласно своим убеждениям.

Классификация интеллектуальных агентов

По типу алгоритма поведения:

- *агенты с простым поведением* – действуют только на основе текущих знаний, по схеме: IF (условие) THEN (действие), например, *компьютерные вирусы* (*Aids-8064*, *Ionkin-2372*, «Черная дыра», «Тринадцатая пятница», «Вирус замедления», «Израильский» и др.); *боты* (сокращение от слова робот), которые предназначены для выполнения работы, однообразной и повторяемой, с максимально возможной скоростью (игровые боты, боты для интернет-аукционов, чат-боты для оповещения прогноза погоды, результатов спортивных соревнований, курсов валют, биржевых котировок и т.п.), *поисковые роботы* (Google, Yndex, поисковая система Mail.ru, AltaVista, АПОРТ, CNN) и др.);
- *агенты с поведением, основанным на модели* – оперируют со средой, частично поддающейся наблюдению, например, *Cron* в UNIX и «*Планировщик задач*» в Windows занимаются тем, что запускают указанные пользователем задания в определённые моменты времени;
- *целенаправленные агенты* – оперируют со средой, которая частично поддается наблюдению, а внутри хранят информацию только о желаемых ситуациях, например, *Менеджеры SNMP* обрабатывают данные о конфигурации и

функционировании управляемых систем и преобразуют их во внутренний формат, удобный для поддержания протокола SNMP;

- *практичные агенты* – различают только состояния, когда цель достигнута, и когда не достигнута, например, *Web-сервис BookSearch* позволяет искать книги и получать простую информацию о них, *SmartX*™ – многофункциональный торговый терминал с интуитивно понятным интерфейсом, имеет встроенную систему риск-менеджмента, расширенный функционал и доступ к торговому счету через мобильные приложения и web-кабин. *SmartWeb* на платформе HTML5 – торговая система ценными бумагами, которая позволяет управлять заявками, выставленными с других терминалов и имеет возможность добавления на график индикаторов и других инструментов технического анализа;

- *обучающиеся агенты* – могут приспосабливаться к изменяющимся обстоятельствам, т.е. могут: обучаться и развиваться в процессе взаимодействия с окружающей средой; приспосабливаться в режиме реального времени; быстро обучаться на основе большого объема данных; пошагово приспосабливать новые способы решения задач; обладать базой примеров с возможностью ее пополнения; анализировать себя в терминах поведения, ошибки и успеха, например, *Forward Bank* – дистанционное обучение для агентов, которые предлагают клиентам продукты и услуги банка, *Mirapolis Learning Management System* – современная система управления дистанционным обучением.

Чтобы активно выполнять свои функции, интеллектуальные агенты обычно имеют иерархическую структуру, включающую много субагентов (выполняющих низкоуровневые функции):

- *временные агенты* – позволяют принимать оперативные решения в режиме реального времени, например, Windows Server 2003. Библия пользователя. – позволяет в процессе работы получать подсказки для правильного решения, так в разделе «Порты системных служб» содержится описание каждой системной службы с указанием логического имени, а также используемых портов и протоколов. Чтобы просмотреть описание необходимо щелкнуть на имя системной службы в приведенном списке и получите ответ;
- *сенсорные агенты* – обрабатывают сенсорные сигналы на основе нейросетей, например, *WM_TOUCH* (Сенсорный интерфейс рабочего стола Windows 7 и Windows 8), который может использоваться для указания того, что с экраном могут иметь контакт один или более указателей, таких как палец или стилус (небольшая металлическая или пластиковая палочка со специальным силиконовым наконечником, которым нужно касаться сенсорой поверхности); *WM_POINTER* (сенсорный интерфейс рабочего стола Windows 8) являются частью прикладных интерфейсов Direct Manipulation и относятся к рабочему столу Windows 8. Они могут использоваться для записи отдельных сенсорных указателей, а также жестов и манипуляций.
- *обрабатывающие агенты* – работают с метаданными, которые характеризуют свойства и содержание ресурсов сети Интернет, решают проблемы распознавания, например, *Semantic Web* – это общедоступная глобальная семантическая сеть, формирующая на базе сети Интернет путём стандартизации представления информации в виде, пригодном для автоматической обработки;
- *мировые агенты* – объединяют в себе признаки нескольких классов агентов, например, *Federated search* осуществляет поиска разнородных данных, используя при этом простой набор команд, *Fuzzy agents* собирает информацию и передает оператору.

Агенты, которые имеют ограничения по обучаемости и принятию решений называются *полуинтеллектуальными агентами*. К ним относят:

- *роботы по закупкам* – просматривают сетевые ресурсы, собирают информацию о товарах и услугах, например, Amazon.com, Bot-2: Первый облачный бот для торговли на BTC-E, BITFINEX, POLONIEX и BTCC и др.;
- *пользовательские или персональные агенты* – действуют в интересах клиентов – проверяют почту, сортируют письма по важности, используя заданные критерии, и оповещают, когда поступает важное письмо (службы Apple ID, FaceTime и др.); играют в компьютерной игре в качестве оппонента; собирают новости; ищут информацию по выбранному предмету; самостоятельно заполняют web-формы, сохраняя информацию для последующего использования; просматривают web-страницы, ища и подсвечивая ключевую информацию; дискутируют на различные темы;
- *наблюдающие и управляющие агенты* – следят за состоянием инвентаря, планированием, составлением расписания, ведут наблюдение за компьютерными сетями и конфигурацией каждого компьютера, формируют отчеты. Примеры: *Monitoring and Surveillance Agents, NASA's Jet Propulsion Laboratory*;
- *добывающие информацию агенты* – действуют в хранилище данных, собирая информацию для последующего использования. Например: *Data mining agent*.

Область использования интеллектуальных агентов:

- *бизнес*: управление распределенными или сетевыми предприятиями; сложная и многофункциональная логистика; виртуальные организации и Интернет-порталы по продаже продуктов и услуг; управление учебным процессом в системах дистанционного обучения; компании с развитыми дистрибьюторскими и транспортными сетями (например, в Procter&Gamble); управление каналами распределения; моделирование предпочтений пользователей.
- *управление*: сокращение количества уровней управления и высвобождение работников среднего звена управления, упразднение ряда функций; снижение административных расходов; рационализация решения управленческих задач за счет внедрения математических методов обработки данных и систем

искусственного интеллекта; экономия времени на планирование деятельности и принятие решений; повышение квалификации и информационной грамотности управленцев и др.

- *маркетинг*: создание новых возможностей по получению и распределению информации (наглядность, скорость передачи сообщений, достоверность); информационная поддержка изучения рынка и поиска новых ниш; возможность идентификации и изучения потребителей конкретного продукта или услуги; повышение способности гибко реагировать на спрос и оперативно удовлетворять новые желания потребителей и др.

Наиболее известными системами, в основе которых **интеллектуальные агенты**, являются:

- *Intelligent Hedger* позволяет быстро разработать систему страхования от рисков;
- *Nereid* – оптимизирует работу с валютными опционами, оказывает дилерскую поддержку;
- *PMIDSS* – позволяет осуществлять долгосрочное планирование инвестиций и выбор портфеля ценных бумаг;
- «*IC-Битрикс: Управление сайтом*» представляет собой программное ядро для всестороннего управления веб-проектами любой сложности;
- «*IC-Битрикс: Корпоративный портал*» - позволяет создать корпоративный портал компании, с возможностью доработки штатного функционала под потребности компании и др.

9. Системы Big data

В связи с внедрением информационных технологий во все сферы деятельности человека объем записываемых в хранилища данных постоянно растет. А это означает, что такими же темпами должны изменяться средства хранения информации и появляться новые возможности для наращивания ее объема. Сегодня основной поток информации генерируют не люди, а роботы,

находящиеся в постоянном взаимодействии друг с другом. Они задают темп роста объема данных, что приводит к появлению потребности наращивать количество рабочих серверов, расширять и внедрять новые центры хранения данных (data-центры).

Понятие «большие данные» (Big Data) носит относительный характер. Условно можно сказать, что *Big Data* – это набор информации, по объему превосходящей жесткий диск одного персонального компьютера, не поддающейся обработке классическими инструментами, применяемыми для меньших объемов.

Большие данные (Big data) – технологии обработки структурированных и неструктурированных данных огромных объемов и значительного многообразия для получения воспринимаемых человеком результатов.

В бизнесе Big Data используется для поддержки принятия решений руководителем на основании анализа финансовых показателей из учетной системы предприятия или маркетологом.

Big Data характеризуется следующими параметрами:

- *объем* – большой объем информации не позволяет обрабатывать и хранить ее традиционными способами, а требует для этого новых подходов и усовершенствованных инструментов;
- *скорость* – высокая скорость обработки большого объема информации в режиме on-line;
- *многообразие* – одновременной обработки структурированной и неструктурированной информации разнообразных форматов;
- *ценность* – полезность информации для совершенствования бизнеса предприятия

Методы анализа больших объемов данных:

- *Data Mining* – комплекс подходов для обнаружения скрытых полезных знаний, которые нельзя получить стандартными способами;

- *Crowdsourcing* – комплекс подходов для мобилизации ресурсов посредством информационных технологий с целью решения задач, стоящих перед бизнесом и обществом в целом;
- *Data Fusion & Integration* – набор методов, позволяющие объединять данные различного происхождения для последующего совместного анализа;
- *Machine Learning* – набор методов, позволяющих строить алгоритмы обучения;
- *Методы распознавания образов* – наиболее используемыми методами: *сравнение с образцом* – по ближайшему среднему, классификация по расстоянию до ближайшего соседа и другие структурные методы распознавания; *статистические методы* – A/B-тестирование, анализ временных рядов, метод на основе байесовского решающего правила и другие; *нейронные сети*;
- *Пространственный анализ* – комплекс методов, использующих топологическую, геометрическую и географическую информацию в данных;
- *Визуализация данных* – представление информации в виде рисунков, диаграмм, с использованием интерактивных возможностей и анимации как для получения результатов, так и для использования в качестве исходных данных для дальнейшего анализа.

Big data-система (BD-система) – информационная система, предназначенная для анализа больших объемов данных и управления ими. Она представляет собой комплексную платформу, включающую программное и техническое обеспечение.

Для реализации технологии Big Data служат комплексные платформы, включающие следующие ресурсы:

- программное обеспечение;
- оборудование;
- сервисные услуги.

Наиболее распространенное программное обеспечение ***Big Data***:

- *SQL* – язык структурированных запросов, позволяющий работать с базами данных, создавать и модифицировать их, а управлением массива данных занимается соответствующая СУБД;
- *NoSQL* включает в себя ряд подходов, направленных на реализацию базы данных, имеющих отличия от моделей, используемых в традиционных, реляционных СУБД;
- *Hadoop* используется для реализации поисковых и контекстных механизмов высоконагруженных сайтов – Facebook, eBay, Amazon и др.
- *SAP HANA* – система, обеспечивающая высокую скорость обработки запросов, что уменьшает затраты на поддержку аналитических систем;
- *Apache Chukwa* – система агрегации и анализа больших данных, построенная на основе Hadoop Distributed File System (HDFS) и платформы MapReduce, включает мощный гибкий инструментарий для отображения результатов анализа, обладает масштабируемостью и надежностью;
- *Apache Hadoop* – распределенная система обработки данных, работающая на кластерах;
- *Apache Hive* – система обобщения информации, нерегламентированных запросов и анализа больших объемов данных, размещенных в совместимых с Apache Hadoop файловых системах. Реализованный в Hive механизм позволяет осуществлять запросы на SQL-подобном языке HiveQL;
- *High Performance Computing Cluster* – решает задачи, связанных с оценкой риска;
- *MapReduce* – программное обеспечение для параллельных вычислений над очень большими наборами данных в компьютерных кластерах;
- *Revolution Analytics* – пакеты для математической статистики, могут использовать в облачной системе Azure и других приложениях;
- *PROMT Analyser* – позволяет работать с большими данными в информационно-аналитических системах, анализировать внутренние ресурсы предприятия, и данные, получаемые из закрытых источников, для оценки критичности ситуаций, осуществляет поиск, извлечение, обобщение и структуризацию

информации из практически любого текстового контента на разных языках как в корпоративных системах, так и во внешних источниках, а также оптимизировать поисковые системы и службы поддержки.

Оборудование Big Data включает:

- серверы для размещения хранилищ данных;
- инфраструктурное оборудование – источники бесперебойного питания, комплекты серверных консолей и др.

Сервисные услуги Big Data включают в себя услуги по построению архитектуры системы базы данных, обустройству и оптимизации инфраструктуры и обеспечению безопасности хранения данных.

Технология обработки Big Data позволяет решать задачи:

- хранение поступающей информации;
- структурирование разрозненного контента: текстов, фотографий, видео, аудио и всех иных видов данных;
- анализ информации, создание аналитических отчетов.

Область использования систем Big Data:

- *розничная торговля* – для управления поставками товара, его хранения, продаж, прогнозирования спроса, подготовки отчетности и др.);
- *банковская сфера* – для оценки кредитоспособности заемщика, для кредитного скоринга и андеррайтинга;
- *телекоммуникации* – для проведения сегментации клиентов, анализа их трафика и др.;
- *горнодобывающая и нефтяная промышленность* – для оценки эффективности разработки месторождения, отслеживания графика капитального ремонта и состояния оборудования, прогнозирования спроса на продукцию и цены и др.

Big Data активно внедряются как зарубежными компаниями: Nasdaq, Facebook, Google, IBM, VISA, Master Card, Bank of America, HSBC, AT&T, Coca Cola, Starbucks, Netflix, тСургутнефтегаз, Газпром нефть, так и отече-

ственными: РУП Белтелеком, СП ЗАО «Международный деловой альянс» (ИВА) и др.

10. Системы оперативной аналитической обработки данных

Основой для принятия управленческих решений на предприятии являются базы данных оперативной информации. Современный уровень развития аппаратных и программных средств сделал возможным повсеместное ведение их на всех уровнях управления. 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 для предприятия.

Современные предприятия для принятия управленческих решений используют как старые централизованные системы, так и новые распределенные системы, основой для которых является хранилище оперативных данных, которое объединяет данные из различных источников в простые и удобные для доступа пользователя.

Хранилище данных (Data Warehouse) – предметно-ориентированная информационная база данных, специально разработанная и предназначенная для подготовки отчётов и бизнес-анализа. Содержимое хранилища данных формируется из данных транзакционной системы (OLTP)⁵ – данные из этой системы копируются в хранилище данных, а при построении отчётов и аналитическом анализе не используются ресурсы транзакционной системы, что не нарушает её стабильность.

Хранилище данных имеет свою структуру хранения данных. Основными являются трехрядная и двухрядная структуры.

В *трехрядная структура* процесса хранения данных включает такие стадии: *сбор* – извлечение данных из внутренних и внешних источников; *обработка* – очищение собранных данных, фильтрация и суммирование по-

⁵ OLTP-системы предназначены для ввода, структурированного хранения и обработки информации в режиме реального времени.

средством специального ПО; *хранение* – помещение очищенных, профильтрованных и суммированных данных в дополнительную специальную многомерную базу данных (БД), которая удобна для пользователя информационной системы.

Такая структура хранения данных на основе многомерной БД приемлема для предприятий, где данные хранятся в различных системах, причем одни и те же данные представлены по-разному в различных системах; база данных большая и разнообразная; данные хранятся в трудных для расшифровки форматах.

Двухрядная структура хранения данных, в отличие от трехрядной, не имеет многомерной БД, поэтому в ней отсутствуют стадии обработки и хранения, а извлеченные данные из внутренних и внешних источников сразу анализируются. Такая структура хранения данных подходит для малых организаций.

Обновление хранилища данных может быть:

- полным – загрузка новых данных происходит после полного удаления старых данных. Процесс происходит с определённой периодичностью, при этом актуальность данных может несколько отставать от OLTP-системы;
- частичным – обновляются только те данные, которые изменились в OLTP-системе.

Для того чтобы существующие хранилища данных способствовали ускорению процесса принятия управленческих решений аналитик должен иметь развитые инструменты доступа к данным хранилища и их обработки.

Оперативная аналитическая обработка (OLAP, online analytical processing) – технология обработки данных, заключающаяся в подготовке агрегированной информации на основе больших массивов данных, структурированных по многомерному принципу. Реализации технологии OLAP являются компонентами программных решений класса Business Intelligence. Для хранения базовых данных и информации об измерениях, которые могут потребоваться для ответов на любые запросы используют OLAP-куб.

Особенности OLAP систем:

- работа в режиме реального времени;
- многомерное представление данных OLAP-кубом;
- поддержка технологии Big data;
- поддержка технологии Business Intelligence;
- генерация данных из различных источников.

Классификация OLAP-систем:

по типу исходной БД OLAP делятся на:

- ***MOLAP*** – системы оперативной аналитической обработки, например, SAS System Essbase, Oracle Express Server, работающие только со своими собственными многомерными базами данных. Они обеспечивают полный цикл OLAP-обработки. Они либо включают в себя, помимо серверного компонента, собственный интегрированный клиентский интерфейс, либо используют для связи с пользователем внешние программы работы с электронными таблицами;
- ***ROLAP*** – системы оперативной аналитической обработки реляционных данных позволяют представлять данные, хранимые в реляционной базе, в многомерной форме, обеспечивает преобразование информации в многомерную модель через промежуточный слой метаданных, например, DSS Suite, DSS/Server и DSS/Agent, MetaCube, DecisionSuite, ИнфоВизор и др.;
- ***HOLAP*** – гибридные системы совмещают достоинств и минимизируют недостатки, присущих MOLAP и ROLAP, например, Media/MR объединяет аналитическую гибкость и скорость ответа MOLAP с постоянным доступом к реальным данным, свойственным ROLAP;
- ***Инструменты генерации запросов и отчетов для настольных компьютеров***, которые интегрированы с внешними средствами, выполняющими такие функции. Они осуществляют выборку данных из исходных источников, преобразуют их и помещают в динамическую многомерную БД, функционирующую на клиентской станции конечного пользователя, например, BusinessObjects, BrioQuery и PowerPlay.

Интеллектуальный анализ данных (Data Mining) - это процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей (шаблонов информации). При этом накопленные сведения автоматически обобщаются до информации, которая может быть охарактеризована как знания.

В общем случае процесс ИАД состоит из таких стадий:

- выявление закономерностей (свободный поиск);
- использование выявленных закономерностей для предсказания неизвестных значений (прогностическое моделирование);
- анализ исключений, предназначенный для выявления и толкования аномалий в найденных закономерностях.

по режиму анализа данных:

- *статические* – на этапе проектирования они ориентированы на разработку отчетных шаблонов, которые заполняются данными по требованию пользователя, например, информационная система руководителя или Executive Information Systems (EIS);
- *динамические* – ориентированы на разработку динамической модели построения отчетов с использованием технологий анализа OLAP.

Статическая система оперативной аналитической обработки данных на практике очень сложна, так как на этапе проектирования шаблонов невозможно учесть все требования, которые предъявляют пользователи и которые могут возникнуть в процессе анализа. Результатом работы статической системы оперативной аналитической обработки данных являются многостраничные отчеты, которые не в полной мере удовлетворяют аналитика и появляются новые вопросы, требующие продолжения изучения и анализа имеющихся исходных данных

Динамическая система оперативной аналитической обработки данных ограничена множеством измерений, по которым можно производить анализ.

Работа аналитиков с динамическими системами заключается в интерактивной последовательности формирования запросов и изучения их результа-

тов, каждый из которых может породить потребность новой серии запросов. Так как задачей анализа является минимизация недоступных для отчетности наборов данных при максимуме удовлетворенных потребностей пользователей, решением проблемы будет являться комбинация этих систем.

Общая структура информационно-аналитической системы, построенной на основе хранилища данных, представлена на рис. 4..

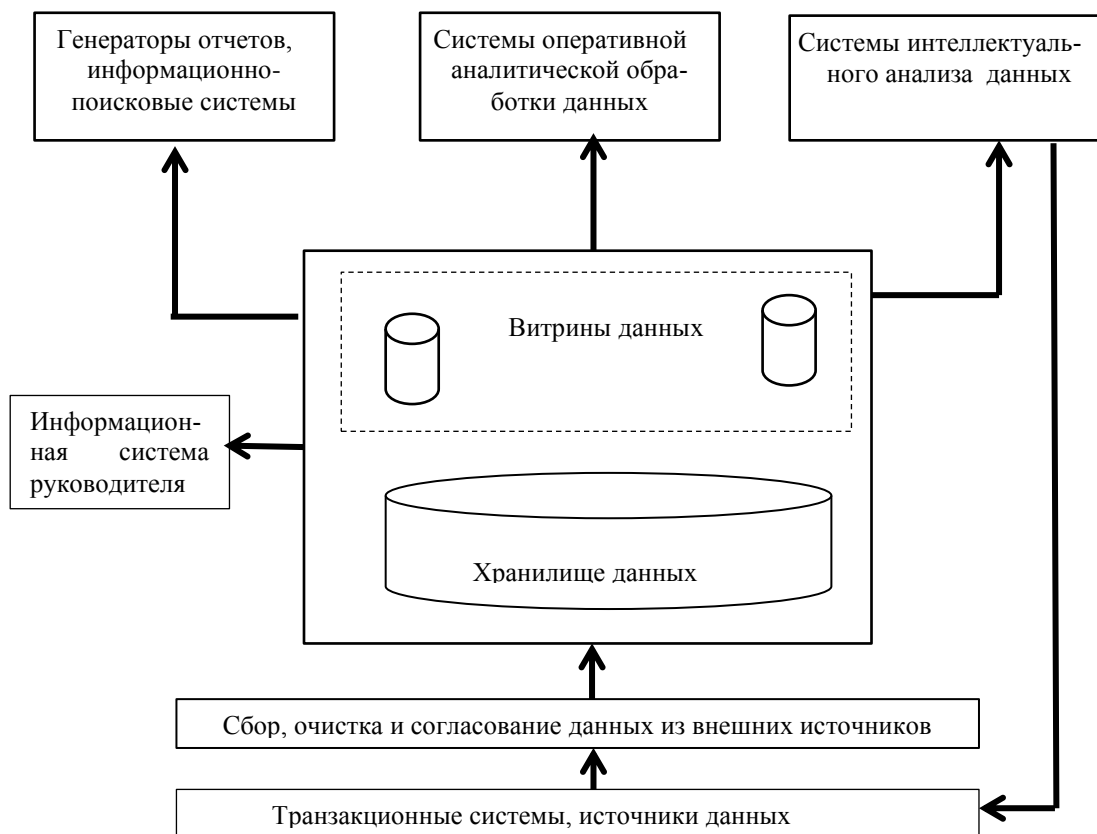


Рис. 4. Общая структура информационно-аналитической системы

OLAP-системы могут быть как самостоятельными программными продуктами, так и являться компонентами автономных/настольных СУБД и/или предоставлять им OLAP-функции из внешних систем

В настоящее время широкое распространение получили такие системы оперативной аналитической обработки данных: Microsoft SQL Server Analysis Services, Hyperion Essbase, Cognos PowerPlay, BusinessObjects, MicroStrategy, SAP BW, Cartesis Magnitude, Oracle Express, OLAP Option, IBM Cognos TM1. Существует несколько open-source решений, включая Mondrian и Palo.