

# Информационные технологии на службе разведки.

## Обзор современных программных средств обеспечения принятия управленческих решений

**И. А. Печенкин,**  
директор департамента технологий  
управления знаниями  
ООО «Информбюро»



### Конкурентная разведка – объективная необходимость для современного бизнеса

Реалии современного делового мира таковы, что многие предприниматели и руководители предприятий приходят к пониманию: без глубокого анализа информации, которая сегодня буквально наводняет рынки, невозможно успешное ведение бизнеса. Потoki информации, генерируемой участниками деловой активности, при ее квалифицированной обработке, анализе и синтезе выводов, способны вооружить предприятие конкурентным преимуществом по отношению к другим игрокам на рынке, которые не обладают нужной информацией в нужное время.

«...В любом бизнесе при любых условиях, прежде чем вкладывать деньги, развивать или изменять направление бизнеса, выбирать партнеров по бизнесу, необходимо активно собирать информацию для принятия решения. Задачи конкурентной разведки мы вынуждены решать каждый день и, на мой взгляд, будем решать теперь всегда (проректор московской академии безопасности предпринимательства В. С. Абрамов)» [2].

«Времена прибыли в 500 % от вложенного безвозвратно прошли, рыночные ниши и люфты сегодня активно заполняются, и к поиску новых заповедных мест надо уже подходить, опираясь на стройную систему экономической разведки, позволяющей получить данные о рынках сбыта, конкурентах, партнерах, кон-

трагентах, новых технологиях, законодательных актах и т. д. и т. п.» [1].

По мнению директора службы безопасности Ассоциации российских банков «Амулет» А. С. Крылова, рынок конкурентной разведки России только формируется: «До этого не было объективных экономических обстоятельств. Жесткая конкуренция несколько лет тому назад заменялась разборками. Сейчас рынок все больше приобретает структурированный, цивилизованный характер. И теперь даже для малого бизнеса, чтобы относительно спокойно принять решение: идти в тот или иной регион, развиваться, вкладывать деньги, требуется предварительная информационная проработка» [2].

Конкурентная разведка, которая на западе уже представляет собой отдельную область экономики, приобретает все большее значение и в нашей стране. Что же есть разведка по своей сути? Разведка – это технологии сбора, обработки, анализа информации и синтеза знаний, необходимых для реализации конкретных целей. Научно-техническая революция, свершившаяся в области информационных технологий в конце XX столетия, открыла широчайшие возможности для генерации и распространения информации в глобальных масштабах. Соответственно, технологии сбора, обработки, анализа информации и синтеза выводов по ней совершенствуются в ногу с научно-техническим прогрессом. Порождение спроса на информационные технологии, облегчающие извлечение и обработку

информации, вызвало естественное предложение компаний-разработчиков программного обеспечения различных решений в этой области.

Какие программные продукты представлены на рынке программных средств для выполнения информационных исследований? Какие из них и какими способами решают поставленные задачи? К кому из поставщиков таких решений можно обратиться по тому или иному вопросу? Для того чтобы ответить на эти вопросы, необходимо вначале рассмотреть основные принципы работы эксперта аналитика и принципы построения информационных систем для автоматизации разведывательных исследований.

## Основные принципы работы программных средств для конкурентной разведки

### Принципы организации информационно-аналитической деятельности аналитика

Какие основные принципы построения программных средств для выполнения разведывательных исследований? Эти принципы базируются на особенностях организации разведывательных исследований на предприятии. Для того чтобы говорить о принципах организации разведывательных исследований (в дальнейшем – разведка) сначала надо определить цели создания подразделений конкурентной разведки (информационно-аналитических служб). А. И. Доронин в своей книге «Бизнес разведка» дает следующее определение этих целей:

«Цели создания подразделений конкурентной разведки:

а) обеспечение руководства достоверной, объективной и полной информацией о:

- намерениях партнеров, смежников, клиентов, контрагентов;
- сильных и слабых сторонах конкурентов;

б) сбор данных, позволяющих оказывать влияние на позицию оппонентов в ходе деловых переговоров;

в) оповещение о возможном возникновении кризисных ситуаций;

г) мониторинг и контроль хода реализации заключенных договоров и достигнутых ранее договоренностей» [1].

Весь процесс сбора, обработки, анализа информации и синтеза знаний представляет собой ряд последовательных мероприятий, повторяющихся от одной цели разведки к другой. Этот процесс получил название разведывательный цикл. В его состав входят следующие основные этапы:

- целеуказание и планирование;
- сбор данных;
- обработка данных (превращение в информацию);
- анализ и синтез информации (превращение в знания);
- распространение полученных знаний.

Собранная информация на первом этапе представляет собой разведданные, после ее обработки (сортировки, калибровки) она превращается в информацию, и только после анализа информации и синтеза на ее основе выводов экспертом, она становится знаниями. Знаниями, дающими реальную почву руководству для принятия решений.

Таким образом, мы подошли к пониманию того, какие основные функции реализуются сотрудником информационно-аналитического подразделения в ходе выполнения разведки. К этим функциям относятся:

- сбор данных;
- накопление данных;
- их структуризация и категоризация;
- поиск и извлечение необходимых данных;
- анализ данных – работа с отчетами по извлеченным данным: получение отчетов по различным критериям запросов, сопоставление результатов;
- создание отчетов с формулировкой выводов для руководства по результатам анализа полученной информации.

Важнейшим условием успешной работы эксперта-аналитика – является наличие информационного поля хозяйствующего субъекта, которое должно представлять собой ряд структурированных и неструктурированных информацион-

ных массивов, необходимых для извлечения из них необходимых данных. Как правило, эти массивы формализованы в виде различных банков и баз данных, которые позволяют строить информационные модели, давать практические рекомендации по управлению текущей ситуацией и прогнозировать ее дальнейшее развитие.

Основная технология аналитика – это установление причинно-следственных связей между различного рода данными. Построение причинно-следственных цепочек позволяет обработанные данные обратить в информацию и, сделав выводы в предметной области, синтезировать соответствующие рекомендации для руководства предприятия.

Наиболее точная картина будет формироваться, если банк данных будет включать в себя данные, получаемые из внешних источников и данные, получаемые из внутренних источников. К данным, получаемым из внутренних источников, могут относиться отчеты сотрудников по всем аспектам взаимоотношений с внешними контрагентами.

Определив основные функции аналитика, при выполнении им разведки, мы подошли к пониманию функций программных средств, которые должны отвечать задачам разведки.

### Функциональность программных средств, отвечающая принципам организации информационно-аналитического процесса

Если проецировать полученную функциональность процесса аналитической деятельности на существующие информационные технологии, то легко обнаружить, что из всех существующих программных средств автоматизации бизнес-процессов предприятий наиболее полно отвечает задачам бизнес-разведки (конкурентной разведки) информационная технология, получившая название – технология управления знаниями (Knowledge Management, КМ-технология). Данная технология реализует идею сбора и накопления всей доступной предприятию информации, как из внутрен-

них, так и из внешних источников. Такое накопление знаний дает предприятию значительное конкурентное преимущество в виде банка знаний предприятия.

Банк знаний предприятия в совокупности с прилагающимися системами сопряжения с другими информационными системами предприятия, с мощной системой поиска информации и приложениями для построения отчетов в различных срезах, позволяет руководству оперативно получать нужную информацию и, соответственно, оперативно реагировать на вызовы внешней и внутренней среды. Такая оперативность, помноженная на выигрывш во времени принятия решения относительно конкурентов, плюс преемственность знаний (более быструю передачу накопленных знаний) между поколениями сотрудников предприятия, дают организации мощнейшее оружие конкурентной борьбы.

Какую функциональность несет в себе технология КМ? Большинство программных средств, обеспечивающих КМ-технологии, как правило, реализуют следующие типичные функции:

- сбор данных из источников различных форматов (БД, неструктурированные источники и т. д.);
- накопление и хранение данных;
- рубрикацию архивов;
- поиск данных, в том числе и нечеткий поиск;
- построение отчетов в различных срезах выборки, в том числе и многомерный анализ данных;
- построение причинно-следственных цепочек данных, позволяющих определять тенденции и направления развития ситуации.
- моделирование данных, ситуационный и сценарный анализ («что будет, если»).

Функция сбора данных реализуется через создание необходимых конверторов, настраиваемых на сбор данных из различных информационных систем и использования так называемых поисковых роботов. Основное предназначение поисковых роботов – сбор данных на заранее predetermined тематику из открытых источников web-пространства.

Накопление и хранение данных происходит через создание специализированных хранилищ электронных данных. При этом разработчики могут использовать как общеупотребительные хранилища данных SQL, Oracle, файловые серверы, так и предложить собственные решения в этой области (Hummingbird, Documentum, Lotus Notes и т. д.). Как правило, специализированные хранилища обладают собственными системами категоризации и рубрикации хранимых данных.

Рубрикация может выполняться ручным вводом необходимых категорий (рубрик) или автоматически. Наиболее передовые технологии уже позволяют выполнять автоматическую рубрикацию документов без элементов «обучения» (настройки) системы. Функции поиска реализуются за счет алгоритмизации математических моделей ассоциативно-статистического и синтактико-семантического подходов анализа данных. Построение отчетов выполняемо программными инструментами построения запросов, многомерного анализа (построения выборок по нескольким критериям отбора одновременно) и средствами построения отчетов на заданные темы. Построение причинно-следственных цепочек реализуется с использованием математических моделей нейронных сетей. Программные средства моделирования данных, ситуационного или сценарного анализа позволяют моделировать финальные состояния объектов исследования в зависимости от изменения тех или иных данных в цепи преобразований.

Таким образом, решения в области программного обеспечения для автоматизации процессов бизнес-разведки условно можно классифицировать по ряду признаков.

#### По функциональности:

- полнофункциональные;
- частичные (реализующие одну или несколько функций КМ-технологии).

#### По форме представления обрабатываемых данных:

- для работы со структурированными данными (цифровые данные);

- для работы с неструктурированными данными (текстовые, видео, графика);
- комбинированные.

#### По степени автоматизации логических операций:

- комплексы поиска и сбора данных (поисковые системы);
- аналитические комплексы (содержащие автоматизированные процедуры или методы анализа данных);
- поисково-аналитические комплексы.

### Решения на базе информационных технологий в помощь эксперту-аналитику

В настоящее время на рынке информационных технологий в области КМ и ВІ (Business Intelligence) представлены в основном компании-разработчики программного обеспечения (ПО) зарубежного происхождения или российские компании, выполнявшие разработки в интересах зарубежных заказчиков. Лишь незначительная часть игроков рынка может похвастаться собственными разработками в этой области. Это объясняется несколькими причинами:

- математический аппарат программного обеспечения в данной предметной области относится к НИОКР в области моделирования искусственного интеллекта, что требует привлечения персонала высочайшей квалификации;
- низкая капиталоемкость российского рынка ИТ не позволяет выделить российским компаниям необходимые инвестиции на НИОКР в данной области.

Соответственно, о наличии полнофункциональных систем КМ и ВІ возможно говорить только у крупных зарубежных разработчиков – это Documentum, Hummingbird, IBM, FileNet и т. д. Предприятия, способные установить у себя такие системы, получают в свои руки мощный информационно-аналитический инструмент, способный обеспечить руководство информацией для принятия управленческих решений, базирующийся на всем объеме информации

предприятия, получаемой как из внешних источников, так и из внутренних систем и документов.

### Обзор полнофункциональных систем

#### Решение от Hummingbird

Канадская компания Hummingbird, Inc представляет полнофункциональный программный комплекс Enterprise Information Management Systems, (EIMS). EIMS



включает в себя линейку программных продуктов, реализующих те или иные функции КМ-технологии. В состав комплекса входит ряд программных модулей.

Hummingbird KM – Управление знаниями. Предоставляет пользователям и проектным группам возможность интегрировать информационные репозитории для поиска информации и анализа документации. Использование решения позволяет осуществлять единый сквозной поиск по всем источникам неструктурированных информационных ресурсов, включая Lotus Notes, Microsoft Exchange Server, web-сайты, файловые репозитории, системы управления документами, мультимедийным библиотекам и т. п.

Hummingbird BI – Бизнес-аналитика.

BI Query – Запрос. Приложение для создания специальных запросов и построения отчетов, позволяющее пользователю формулировать вопросы о данных и объединять результаты в отчетах.

BI Web – Веб-решение для построения запросов и отчетов с дополнительными OLAP-возможностями.

BI Analyze – Аналитическое приложение оперативной аналитической обработки (OLAP; On-Line Analytical Processing) для многомерного анализа данных, позволяющее пользователям получать аналитическую информацию по данным через срезы данных по различным категориям.

BI Server – сервер корпоративных приложений, который обеспе-

чивает безопасность, планирование, распространение, уведомление и централизованное администрирование программными приложениями.

Hummingbird DM – Решение по документообороту. Оптимизирует процессы, связанные с получением, обменом, коллективным использованием и обеспечением защиты корпоративных информационных ресурсов, выраженных в виде электронных документов.

Hummingbird RM – Управление записями. Создает среду для управления полным жизненным циклом всех физических и электронных информационных интеллектуальных активов корпорации от их создания до элементарного перемещения. Пользователи получают возможность обеспечить быструю категоризацию документов.

Hummingbird Collaboration – Среда коллективной работы для системы управления корпоративной информацией. Используя такую среду, организации могут создать высоконадежную, интуитивно понятную среду для проектных групп, необходимую для публикации, хранения, управления, распространения и обмена документами. При интеграции с решением по документообороту и с Порталом (см. ниже), пользователь получает возможность задействовать все функциональные и инструментальные средства для получения целостной информации по проекту и результативной работы проектной команды.

Hummingbird ETL – Универсальное решение интеграции данных. Конвертор данных из любого источника в любую целевую систему. Может выполнять операции, связанные с обменом данными независимо от формата данных, синтаксиса или целевой системы, от поддержки XML до универсальной возможности подключения к любым базам данных, включая многомерную оперативную аналитическую обработку. Обеспечивает тесную интеграцию любого проекта с любыми корпоративными системами, не требует дополнительного программирования и перемещения данных.

Hummingbird Portal – Портал. Платформа, позволяющая интегрировать существующую информацию и приложения в едином интерфейсе, обладающая инструментами, которые обеспечивают лучший доступ, управление, обмен и понимание этой информации. Портал создан для управления документами и корпоративной информацией, интеллектуальными активами, доступом к главному хранилищу и работой в сетевой среде, управления взаимодействием и т. п.

#### Решение от Documentum

Американская корпорация Documentum (Nasdaq: DCTM) предлагает платформы для управления неструктурированной информацией, составляющей основной объем информационных ресурсов предприятия и хранящейся в виде фай-



лов различного формата, таких как офисные документы, отсканированные бумажные документы, web-страницы, графические изображения, чертежи, видеофайлы и т. д.

Платформа Documentum является основой для построения различных бизнес-приложений для решения широкого спектра задач, связанных с созданием, хранением и использованием неструктурированной информации. Такие бизнес-приложения создаются в четырех основных областях: управление документами и создание электронных архивов, автоматизация основных и вспомогательных бизнес-процессов предприятия, проектирование и поддержка жизненного цикла изделия, автоматизация профессиональной деятельности по созданию содержания в рекламной и издательской деятельности, электронных информационных ресурсах и т. п.

Платформа Documentum основана на трехуровневой архитектуре, которая является важнейшим условием при построении современных информационных систем. Можно выделить три основных элемента

архитектуры Documentum – это хранилище содержания, службы управления содержанием и приложения для работы с содержанием.

Content Server – Хранилище содержания. Позволяет хранить и управлять различными типами неструктурированного содержания: документами, web-содержанием, XML, мультимедийными данными, отсканированными изображениями, записями и др. Хранилище обеспечивает полную безопасность хранимого содержания, масштабируемость, открытость, гибкость, позволяя предприятиям объединить все свои информационные системы и приложения в единую корпоративную информационную среду.

Службы управления содержанием – это набор автоматизированных функций, которые производят различные действия над объектами содержания в хранилище и исполняются сервером Documentum (Content Server).

Управление версиями документов. Любой объект Documentum может быть представлен неограниченным числом версий. Каждая версия – это отдельный объект хранилища со своими атрибутами и файлом содержания, что позво-

решение, которое позволяет просматривать и аннотировать документы различных форматов без использования специализированных приложений, в которых эти документы создавались.

Управление жизненным циклом документа – встроенный механизм, который управляет жизненным циклом тех типов содержания, которые были определены на предприятии (например, договор, чертеж, счет, заявка и т. д.).

Службы Workflow позволяют автоматизировать процесс передачи объектов содержания между исполнителями, определяя, что, кто и когда должен выполнить на своем участке работы. При этом исполнителем может быть не только пользователь, но и любая информационная система или приложение, которое получит задачу и произведет ее автоматическую обработку. Для проектирования маршрутов Workflow Documentum предлагает простой в использовании графический инструмент, при помощи которого можно визуально создавать маршруты бизнес-процессов любой сложности.

Content Aggregation Services (CAS) – Служба сбора содержания – содержание можно извле-

цесс рассылки содержания из хранилища Documentum через Интернет различным получателям.

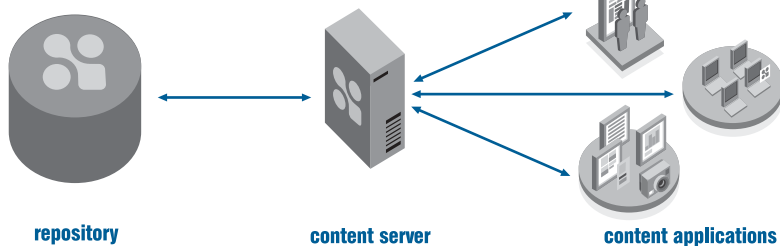
Documentum Media Services Documentum – Специальные службы для управления массивами мультимедийных данных. В основе служб лежит серверный компонент Media Server, которая отвечает за обработку мультимедийных данных. В базовой конфигурации служб поддерживаются следующие форматы: EPS, GIF, JPEG, MAC, OS/2 BMP, PCD, PCX, PDF, PICT, PNG, PSD, RLE, SCT, SGI, SVG, TGA, TIFF, Windows BMP, AIFC, AIFF, AU, IRCAM, WAV, Real Windows Media Audio, MPEG-1, MPEG-2, AVI, Real Windows Media Video.

Documentum Media Server – это расширяемый набор библиотек, который позволяет производить различные операции над мультимедийными данными в хранилище Documentum. Используя заранее определенные шаблоны, бизнес-пользователи, не владеющие специальным инструментарием для обработки изображений, таким как PhotoShop, могут производить над хранимыми данными различные операции, работая в привычных рабочих приложениях.

Documentum Records Management – Технологии управления записями. Позволяют организовать архивное хранение официальной документации в электронном виде. К любому типу содержания в хранилище может быть привязана учетная запись, которая будет гарантировать законную подлинность требуемых объектов в хранилище Documentum. Законная подлинность юридически обеспечивается необходимыми стандартами и сертификатами безопасности, которыми обладает Records Manager.

Adobe Forms – Управление формами документов. Это приложение для обработки электронных форм, которое позволяет создавать формы любой сложности на основе PDF и HTML. Adobe Forms имеет собственный визуальный графически редактор форм, где можно проектировать их внутреннюю логику, структуру и внешний вид.

## ECM Platform



ляет независимо отслеживать и управлять последовательными изменениями его состояния.

Виртуальные документы. Механизм виртуальных документов позволяет визуально или автоматически строить иерархические структуры взаимосвязей непосредственно между объектами содержания, вне зависимости от того, в какой части хранилища (в каких папках) они располагаются.

Модуль Brava!. Просмотр и аннотирование документов. Представляет собой технологическое

решение, которое позволяет просматривать и аннотировать документы различных форматов без использования специализированных приложений, в которых эти документы создавались. Службы позволяют проводить сканирование внешних хранилищ информации, извлечение необходимых данных по заданным параметрам и их перенос в хранилище Documentum для дальнейшей классификации и обработки.

Content Distribution Services (CDS) – Служба распределения информации. Представляет собой дополнительный сервер, который позволяет автоматизировать про-

Captiva InputAccel и Adobe FineReader – специализированные продукты для сканирования и распознавания входящих бумажных форм. После завершения процесса распознавания документа информация из него извлекается и сохраняется в виде атрибутов документа в хранилище Documentum. Затем отсканированная форма и распознанные данные могут быть привязаны к какому-либо бизнес-процессу, переданы в CRM или ERP системе либо зарегистрированы в постоянном неизменяемом архиве.

Technology Services Group (TSG) использует другой подход к обработке электронных форм, который основан на настраиваемом Мастере. Мастер разбивает процесс заполнения формы на шаги, что позволяет пользователю перемещаться по шагам вперед и назад, заполняя определенные категории полей и отвечая на вопросы Мастера.

Клиентские приложения Documentum:

- Documentum Desktop и Webtop – стандартные клиентские приложения открывают доступ к хранилищу Documentum и всем основным службам по управлению содержанием из среды Windows Explorer и web-браузеров;
- Documentum WebPublisher – специализированное приложение для пользователей, которые управляют структурой и содержанием web-сайтов и порталов;
- Digital Asset Manager – приложение для пользователей, работающих с массивами мультимедийных данных;
- Documentum Administrator предназначен для администраторов и открывает доступ ко всем функциям администрирования платформы Documentum.
- набор модулей интеграции с бизнес-приложениями для пользователей, которым необходимо работать в специализированных приложениях или системах, например, в MS Office или SAP, осуществляя при этом доступ к хранилищу неструктурированного содержания Documentum;
- модули интеграции с основными порталами, которые представлены сегодня на рынке.

### Решение от «Город ИНФО»

Компания «Город-Инфо» была основана в 1996 году. Стратегическим направлением деятельности «Город-Инфо» является Web-интеграция – новый подход к построению корпоративных информационных систем на базе web-технологий, которые призваны существенно повысить отдачу от использования корпоративных информационных



систем, способствовать повышению эффективности деятельности российских предприятий, снизить затраты на поддержку и развитие информационной системы.

В архитектуру программных продуктов «Город-Инфо» изначально заложен модульный принцип построения информационной системы, то есть возможность постепенного расширения функциональности внедренного решения путем добавления новых модулей, функционирующих на единой технологической платформе. Все модули тесно интегрированы друг с другом, а также с имеющимися приложениями и базами данных. Доступ к системе осуществляется через Корпоративный портал Axiom.Portal. Такой подход позволяет автоматизировать не только некоторые отдельные «конечные» функции (например, складской учет), но бизнес-процессы в целом. Например, система CRM позволяет не только поддерживать базу данных клиентов, но и интегрировать ее с бухгалтерскими и складскими программами, что позволит эффективно управлять финансовыми потоками и процессом обслуживания клиента.

Компания «Город-Инфо» работала и предлагает своим заказчикам следующие программные продукты, реализованные в рамках технологии web-интеграции:

Axiom.Portal – корпоративный информационный портал;

Axiom.CRM – систему управления взаимодействием с клиентами;

Axiom.Logistics – систему автоматизации деятельности экспедиторской компании;

Axiom.Docstore – единое хранилище документов;

Axiom.Docflow – систему управления документооборотом;

Axiom.Clipping – мониторинг внешних информационных источников;

Axiom.CM – систему управления содержанием (content management);

Ситуационно-кризисные центры (СКЦ) – средства создания систем оперативного мониторинга и визуализации информации;

Интразнание – систему дистанционного обучения.

### Системы поддержки принятия решений

Подклассом программных продуктов технологии управления знаниями можно считать системы поддержки принятия решений (DSS). Как правило, это программные комплексы, реализующие функции, аналогичные функциям KM-технологии, однако они больше тяготеют к работе со структурированными данными и, зачастую, не имеют в себе систем документооборота предприятий. Документами для них являются записи в БД, управляемые надстроенной над ними СУБД.

### Решение от Oracle

Компания Oracle в отдельном представлении не нуждается. Концепция построения систем поддержки принятия решений, предлагаемая Oracle, объединяет все компоненты, необходимые для создания и управления Хранилищем

# ORACLE®

Данных, а также для использования накопленной в нем информации.

В качестве механизма хранения в реляционных Хранилищах и Витринах Данных используется сервер Oracle8i, в многомерных Витринах – Express Server. Центральным инструментальным средством

вом является Oracle Warehouse Builder, построенный на базе современной архитектуры Common Warehouse Metadata, который позволяет спроектировать структуру целевого Хранилища, создать процедуры извлечения, согласования и загрузки данных из различных источников и сгенерировать метаданные для средств доступа, таких как Discoverer и Express.

Для проектирования Хранилища можно также использовать знакомый многим инструмент Oracle Designer, а затем автоматически перенести описание проекта в репозиторий метаданных Warehouse Builder.

Средства доступа к данным покрывают весь спектр аналитических задач: для стандартной отчетности используется Reports, для генерации нерегламентированных отчетов и запросов – Discoverer, для сложного многомерного анализа – продукты семейства Express, а для задач «извлечения знаний» – Darwin Data Mining Suite. Кроме того, существуют готовые аналитические приложения для решения специализированных задач: Sales Analyzer, Financial Analyzer, Activa и Balanced Scorecard. В настоящее время готовятся к выпуску новые аналитические приложения, такие как Demand Planning, Value Based Management и др. Конечные пользователи могут получать информацию как с помощью традиционных клиент-серверных технологий, так и в интранет-архитектуре с использованием масштабируемого сервера приложений Application Server.

Направление Хранилищ Данных и систем бизнес-анализа является сегодня для Oracle одним из наиболее приоритетных. Будучи поставщиком полного технологического решения в данной области, Oracle выпускает новые продукты и постоянно совершенствует существующие. Кроме программного обеспечения, корпорация Oracle разработала специальную методологию построения Хранилищ Данных – Data Warehouse Method, дающую четкую последовательность этапов внедрения и позволяющую получить гарантированный результат.

### Решение от Cognitive Technologies Ltd.

Компания Cognitive Technologies Ltd. была создана в 1993 году на базе Лаборатории искусственного интеллекта Института системного анализа РАН (бывший ВНИИСИ АН СССР), руководимой профессором, д. т. н., В. Л. Арлазаровым. Под его непосредственным руководством создавались программа «КАИССА» (первый чемпион мира по шахматам среди компьютеров) и СУБД ИНЕС (более 2500 инсталляций в стране и в мире). В компании работает 327 сотрудников, в том числе пять докторов наук и девятнадцать кандидатов наук.

Программный комплекс Cognitive «Эксперт» предназначен для построения информационно-аналитических систем. Использование комплекса позволяет компаниям и организациям развивать и автоматизировать следующие виды деятельности:

- сбор и обработку неструктурированной информации, получаемой из Интернета, печатных материалов, СМИ и т. д.;
- аналитико-прогностические исследования, направленные на оценку перспектив развития в различных сферах деятельности;



- конкурентную разведку;
- анализ рынков сбыта/предложения товара;
- анализ рынка фирм поставщиков/покупателей.

Перечисленные модули решают следующие основные задачи.

Ввод бумажных документов:

- потоковое и ручное сканирование документов;
- конвертация бумажных документов в электронную форму;
- распознавание фрагментов документов «на лету» при заполнении их реквизитов

Ввод электронных документов:

- потоковый ввод документов с интернет-сайтов;
- ввод документов с электронных носителей информации;

- ввод документов из баз данных.
- Средства обработки и поиска документов:
- автоматическое рубрицирование документов;
  - интеллектуальный поиск документов.

Средства генерации отчетов и визуализации результатов исследований:

- интегрированный генератор отчетов;
- мощные средства визуализации информации.

Средства высокоуровневого анализа:

- инструментарий поддержки ведения системного анализа;
- средства, задающие структуру ведения анализа рынков сбыта, рынка фирм поставщиков/покупателей, других видов конкурентной разведки.

Средства статистического анализа:

- анализ временных рядов;
- корреляционный анализ;
- автоматическое ранжирование результатов частотного анализа по реквизитам с определением различных частотных характеристик для последующего анализа.

Администрирование программного комплекса:

- общее управление программным комплексом;
- управление базами данных;
- управление уровнями доступа и правами различных групп пользователей.

### Решение от Business Objects

Программные продукты в области систем поддержки принятия решений DSS/OLAP фирмы Business Objects (Франция), представленной в России компанией «ТЕРН».

Business Objects может быть использован для получения информации как из баз данных SQL, в том числе организованных как Data Warehouse, так и быть средством front-end для складов данных, использующих специализированные многомерные СУБД (Arbor Essbase, Oracle Express).

BusinessObjects существует для платформ Windows (3.1x, 95, NT)

и UNIX (Solaris 2.5, IRIX 6.2 и, в ближайшем будущем, HP UX и AIX), семантический уровень BusinessObjects позволяет работать и с базами данных, организованными как



RDBMS DataWarehouse (Informix MetaCube, Sybase IQ и др.).

Инструментом описания семантики в BusinessObjects является программа Designer. С его помощью программист или администратор, знающий как информация «лежит» в базе, создает каталог терминов конечного пользователя (Universe) и определяет для каждого термина метод получения данных (фрагменты запроса SQL). Здесь же задаются исходные иерархии измерений и исходные форматы отображения объектов в отчетах.

Universe могут быть построены по любой базе, доступной через ODBC32 либо напрямую (Oracle, Sybase, Informix, Microsoft SQL Server, DB2, CA-Ingres, Teradata, Red Brick, Rdb). При прямом доступе используются особенности диалекта SQL конкретной RDBMS. Созданные в Designer Universe используются в клиентских приложениях Reporter, BusinessQuery и WebIntelligence для построения запросов в терминах бизнеса.

Reporter – Построение отчетов. Уровень подготовки специалиста, создающего отчеты, может быть примерно как у среднего пользователя Excel.

Следующим уровнем наращивания системы BusinessObjects является объединение рабочих мест в систему с централизованным администрированием прав пользователей, общими репозиториями Universe и документов. Эта функциональность обеспечивается продуктом Supervisor. База будет содержать несколько доменов. Домен Universe позволяет производить дистрибуцию Universe от разработчиков к пользователям.

Домен документов является местом централизованного хранения и дистрибуции разработанных от-

четов между пользователями BusinessObjects. Через этот домен пользователь может отправить документ на обработку заданий (обновление отчета, печать, экспорт отчета на web, рассылка по списку и выполнение скриптов) по расписанию специальным продуктом Document Agent Server (DAS).

Все операции администрирования могут производиться с одного рабочего места администратора Supervisor. Одновременно могут администрироваться более 10 тыс. пользователей.

BusinessObjects WebIntelligence (проект Darwin) – клиентское место пользователя, которое позволяет обращаться к корпоративной информации, пользуясь стандартным web-браузером с любой платформы (Windows, Mac, UNIX), имеющим встроенную виртуальную машину Java. Нетехнический пользователь может строить и просматривать отчеты, задавая произвольные запросы к базам данных опять же в терминах своего бизнеса. WebIntelligence использует те же Universe, что и другие средства BusinessObjects.

В WebIntelligence реализована схема тонкого клиента, работающего с удаленным сервером приложений. Средством формулирования запросов и работы с отчетом является апплет Java, код которого автоматически передается в браузер в момент первого обращения к серверу приложений. Обновления кода происходят в дальнейшем только при появлении его новых версий. Формирование выборок и построение документа производится на сервере приложений WebIntelligence, который передает изображение документа клиенту в виде страниц HTML. Сервер WebIntelligence использует стандартные средства доступа к корпоративным базам, такие как SQL\*Net for Oracle, Open Client for SQL Server или ODBC.

#### Поисково-аналитические комплексы

Другой ветвью технологии управления знаниями являются поисково-аналитические комплексы, в большей мере сконцентрированные на задачах поиска и извлече-

ния информации из источников различных форматов.

#### Решение от Convera

Компания Convera Technologies International Ltd. (панее – Excalibur) по данным отчетов Garthner Group уже много лет является мировым лидером в области разработки технологий поиска информации и извлечения знаний из текстов и мультимедийных данных любой природы (графические и видеоизображения, звук в файлах практически всех известных форматов).

Для управления знаниями, представленными в текстовом виде, компания Convera предлагает программный продукт RetrievalWare (RW) с набором дополнительных сервисных программ. Семейство продуктов RW обеспечивает поиск, анализ и выделение информации посредством задания пользователем поисковых запросов на естествен-

# CONVERA™

ном языке к информации, хранящейся как в неструктурированном виде, так и формализованных базах данных, расположенной как в локальной сети организации, так и в сети Internet.

В RW реализована уникальная технология «нечеткого» поиска, позволяющая находить информацию не на основе точного совпадения запроса с данными, а на степени похожести запроса, с содержанием в источниках информации, что позволяет находить информацию, введенную, например, с опечатками. Кроме того, при проведении ретроконверсии бумажных документов отпадает необходимость их правки после распознавания текста, что в свою очередь позволяет перевести объемные архивы бумажных документов в электронный вид за обозримый срок (до нескольких тысяч машинописных листов в день с одного рабочего места) и за разумные инвестиции.

В RW реализована возможность ассоциативного поиска на основе семантической сети. При этом можно использовать несколько се-

мантических сетей одновременно, например, сеть на основе общей лексики русского языка и сеть понятий, используемых в приборостроении. Каждый пользователь может с легкостью создавать свои собственные семантические сети в дополнение к общим.

В семантической сети RW можно определять вид семантической связи для понятий, например, слова синонимы или антонимы, слова, связанные отношением «род-вид», и т. п. Данное свойство обеспечивает высокую точность извлечения информации и позволяет автоматически находить документы (или записи в БД) не только по терминам заданным в запросе, но и по другим терминам, связанным по смыслу с заданным.

RW имеет возможность включать в единое поисковое пространство как информацию, хранимую в файловой системе, так и СУБД (Oracle, MS SQL, Sybase, Informix, Teradata, ODBC DBS), почтовых и корпоративных систем (MS Exchange, Lotus) и систем документооборота (StaffWare, Documentum, FileNet Panagon). С помощью RW можно организовывать доступ и индексировать удаленные хранилища данных. Это свойство RW позволяет создавать единое корпоративное информационное пространство.

Развитая система безопасности, наследующая свойства безопасности источников информации, в совокупности с web-технологией позволяет использовать RW как средство для создания территориально распределенных автоматизированных систем.

RW обеспечивает пользователю просмотр документов более чем в 250 форматах, среди которых как широко известные doc, rtf, txt, pdf, html, так и специфические, например, форматы САПР (dxf, dwg). В последней версии RW реализована возможность поиска информации в архивах (ZIP и пр.).

В RW реализована возможность динамической рубрикации всей поступающей информации на основе запросов, созданных пользователями. Таким образом, реализуется функция «профайлинга» и значи-

тельно сокращается время ознакомления с вновь поступившей информацией, так как она представляется пользователю в структурированном виде, то есть предварительно разложенной по рубрикам.

В RW реализована функция кросс-языкового поиска. Пользователю достаточно задавать вопросы на родном языке, система на основе установленного соответствия семантических сетей для разных языков возвращает документы на других языках. RW может автоматически извлекать атрибуты из текстовых документов определенной структуры и помещать их в СУБД (создавать формуляры для документов).

RW имеет дополнительные сервисные программы.

Сервис RW FileRoom полностью интегрирован с RW и предназначен для обеспечения работы с бумажными архивами. В нем совместно хранятся отсканированные образы документов и текстовые файлы, содержащие результаты оптического распознавания. Документам приписываются учетные карточки. Структура электронного архива может повторять структуру бумажного за счет использования виртуальных шкафов, ящиков и папок, что позволяет по электронному образу документа легко находить место хранения физического аналога.

Состав комплекса:

- RW Internet Spider – специальное приложение для поиска в обозначенных областях Интернета и Интранета;
- RW WebExpress – специальное приложение для обслуживания провайдеров, обеспечения поиска по содержимому web-сайта и электронной торговли через Интернет;
- RW CDExpress – приложение для создания портативных баз данных на компакт-дисках, содержащих поисковой механизм RW;
- ScreeningRoom (SR) – средство управления видео архивом, позволяющее наряду с визуальным поиском выделять из видеоизображений текст, соответствующий субтитрам или телетексту и преобразовывать в текст со-

провожающий аудиодорожку, существует возможность создавать аннотацию к видеосюжетам storyboard;

- средства для системных интеграторов и разработчиков программных систем, использующих решения компании Convera – RetrievalWare SDK и Visual RetrievalWare SDK, позволяют разрабатывать дополнительные функциональные возможности к RW для обеспечения решения задач конкретной организации.

### Решение от NeurOK

NeurOK – российско-американская группа высокотехнологичных компаний. Более 50 человек работают в офисах компании в McLean (США), Москве и Снежинске. «НейрОК Интелсофт» поставляет готовые решения и выполняет заказные проекты по созданию информационно-поисковых систем управления знаниями на основе платформы NeurOK Semantic Suite.

NeurOK Semantic Suite – это эффективное решение задач отслеживания, организации и управления информационными потоками и массивами. Оригинальная технология распознавания тематического контекста позволяет автоматизировать многие рутинные процессы обработки информации.

Автоматическая рубрикация информации. Документы в различных форматах в реальном времени



автоматически раскладываются по тематическим рубрикам в зависимости от их содержания и атрибутов. С каждой рубрикой ассоциирован запрос, обычно содержащий несколько десятков или сотен относящихся к данной проблеме терминов. Кроме того, рубрика-запрос может фильтровать документы по наличию в них обязательных терминов и по другим формальным атрибутам, например, времени. Это позволяет, например, формировать специализированные рубрики типа «Все о ...», содержащие все публикации, в которых упоминалось то

или иное имя. Другой пример: рубрика «Последние известия», объединяющая новости по временному, а не тематическому признаку.

Semantic Teacher – специализированный редактор каталогов. Дерево каталога можно создавать и редактировать вручную с помощью этого продукта.

Semantic Miner – компонента для автоматического создания тематического каталога. Обозначается лишь число уровней и желаемое количество рубрик на каждом из иерархических уровней.

Semantic Explorer – система визуализации содержания текстовых коллекций и навигации в них. Она, в частности, позволяет ограничивать поисковые запросы рамками той или иной тематики тематического дерева Иерархия тематических категорий облегчает работу с текстовыми массивами, являясь своеобразным оглавлением баз данных.

Semantic Annotator – модуль выявляет и отмечает в документе наиболее значимые термины и фразы, в максимальной степени отражающие содержание документа. Модуль автоматического аннотирования встраивается в качестве опции в информационно-поисковую систему Semantic Explorer.

Semantic Scanner – программный робот с широким выбором настроек автоматически отслеживает обновления в источниках информации, будь то интернет-страницы или новостные ленты информационных агентств. Автоматический обход и мониторинг интернет-сайтов, в том числе новостных лент, производится данной компонентой. Администратор системы определяет периодичность обхода сайтов, глубину просмотра дерева ссылок, а также правила фильтрации и преобработки контента.

Собранная информация индексируется поисковым сервером NeuroOK Semantic Engine и рубрицируется системой Semantic Explorer. Последняя также предоставляет развитые средства навигации и поиска в массивах собранной информации.

Диалоговая поисковая система. Поиск по ключевым словам уточняется указанием контекста запроса,

причем система сама предлагает варианты уточняющих терминов, делая поиск все более направленным.

Поиск документов по подобию. Одной из функций серверного компонента NeuroOK Semantic Engine является ассоциативный поиск – выдача по любому отрывку текста набора тематически связанных с ним документов. Этот поиск опирается на технологии распознавания образов, лежащих в основе семантической платформы NeuroOK.

Интеграция с базами данных. Полнотекстовый поиск по содержанию документов совмещается с традиционными запросами к базам данных по атрибутам с помощью оригинальной технологии NeuroOK CARE.

Сервер приложений семантической платформы NeuroOK Semantic Engine использует разработанное в NeuroOK ассоциативное расширение реляционных СУБД, нацеленное на решение задач распознавания образов. Соответствующий компонент NeuroOK Content Addressed Retrieval Engine (NeuroOK CARE) создает и поддерживает специальный ассоциативный индекс, обеспечивающий быстрый доступ к информации по всей совокупности ее атрибутов.

#### Решение от «Информбюро»

Компания «Информбюро» представляет программный комплекс Intellectum.BIS™ – информационную систему для работы с массивами информации в электронном виде.

Intellectum.BIS™ предназначен для обеспечения автоматизированного управления информационными ресурсами предприятия. В основном, он ориентирован на аналитические подразделения предприятий и организаций различных отраслей и видов деятельности, но может использоваться и на любых участках, деятельность которых связана с хранением, поиском, анализом и предоставлением данных из информационных хранилищ.

Возможности продукта:

- сбор данных из источников различных форматов (БД, неструктурированных источников, web-ресурсов и т. д.);
- накопление и хранение данных, их каталогизация;

- поиск данных, в том числе и нечеткий поиск;



- рубрикация архивов;
- построение отчетов в различных срезах выборки;
- построение причинно-следственных цепочек данных, позволяющих определять тенденции и направления развития ситуации.

Возможности подключения источников информации в Intellectum.BIS:

- любые электронные документы предприятия;
- базы электронных документов СМИ;
- web-ресурсы;
- специальные базы (телефонные, регистрационные);
- базы по грузоперевозкам, внешнеэкономической деятельности.

Состав комплекса:

- сервер комплекса – обеспечивает обработку процедур в рамках группы пользователей, при первичной установке по умолчанию рассчитан на работу с 5 клиентскими местами;
- клиентское место – интерфейс работы пользователя с комплексом, реализованный на технологии «web-клиент»;
- модуль администрирования – реализует функции администрирования комплексом по разграничению прав и возможностей клиентских мест, приему и обработке поступающих данных, доставке информации к удаленным клиентским местам;
- модуль приема информации – реализует функции приема информации от различных источников, распознавания формата документа и источника поступления информации, разбития информации на статьи (если это текстовые данные), ввод информации в хранилище данных;
- почтовый робот – реализует функцию взаимодействия комплекса

с почтовой системой по приему информации от различных источников информации;

- интернет-робот WEBHunter™ – реализует функцию сбора информации с заданных web-ресурсов с заданной периодичностью.

Архитектура:

- сервер баз данных;
- сервер приложений;
- пользовательские места (Клиент, Администратор).

#### Решение от «Галактики»

В 1999 году в корпорации «Галактика» создан новый продукт – автоматизированная система поиска и аналитической обработки информации Galaktika-Zoom.

Назначение – обеспечение руководителей высшего звена важной и актуальной для принятия решения аналитической информацией, выделенной из массива текстовых данных. Инструментарий «Ситуационной комнаты», позволяющий получить выверенные управленческие решения с предсказуемыми последствиями.

Возможности:

- поиск и формирование информационных массивов по конкретным аспектам исследуемой проблематики, осуществляемый



по ключевым словам с учетом их морфологии;

- анализ объективных смысловых связей отобранных данных,
- получение «образа» проблемы – моментальной многомерной фотографии в информационном потоке в форме ранжированного списка значимых слов, употребляемых совместно с темой проблемы;
- сравнение нескольких состояний проблемы, выявление закономерностей и тенденций (или частностей и случайностей) динамики развития изучаемой проблемы.

Характеристики:

- высокое быстродействие;
- полный морфологический анализ;

- полнотекстовый поиск, включая возможность работы со «стоп-словами»;
- масштабируемость аппаратных средств: поддержка многопроцессорной обработки и кластерных систем;
- разнообразие архитектур: «мобильные» варианты (CD, DVD), «клиент-сервер», «интернет-сервер».

#### Решение

##### от НИЦИ при МИД России

Информационно-аналитический программный комплекс (ИАПК) «Дипломат» разработан в НИЦИ при МИД России под руководством проф. В. С. Кретова.

ИАПК «Дипломат» позволяет вести оперативную обработку поступающей информации и формировать фонд взаимосвязанных баз данных (слабо и сильноформализованной информации) по политическим, экономическим, социальным проблемам и конфликтам, событиям (визиты, заявления, ноты и др.), объектам политической системы (государства, международные организации, общественно-политические организации, промышленные предприятия, политические деятели и т. д.) и документам. В составе ИАПК имеется информационно-поисковая система, позволяющая автоматически индексировать входную информацию.

В комплексе реализованы следующие функции:

- оперативная обработка поступающих на вход документов любого вида (текстовых документов, рисунков, аудио-, видеоматериалов и т. д.);
- формирование и работа со специализированными базами данных: «Международные объекты», «Объекты интереса», «Персоны», «Абстрактные объекты», «Административно-территориальные, географические объекты», «События», «Документы», «Проблемы»;
- расчет и графическая репрезентация количественных показателей: взаимодействия элементов исследуемой области, развития во времени конфликтов, других процессов;

- оперативная подготовка справочных и аналитических документов на основе систематизированной (с помощью этой системы) входной информации.

Теоретической базой создания комплекса являются результаты исследований, полученные, прежде всего, в таких разделах науки, как теория политического конфликта, системный анализ, психология и искусственный интеллект.

Комплекс представляет собой систему гибридного (естественного и искусственного) интеллекта. Его ядром является модель области исследования, построенная в виде объектно-ориентированной сетевой структуры.

Для повышения достоверности информации используется процедура обработки фактов. При этом для отслеживания и анализа проблем и ситуаций применяется трехуровневая схема обработки входной информации «документ – факты – значения» формализованных переменных. В качестве документа могут быть использованы сообщения СМИ, записи бесед, фото-, видеоматериалы и т. д. При обработке документов выделяются факты, характеризующие в то числе и значения формализованных переменных. Факты укладываются в соответствии со структурной схемой, соответствующей тематическим направлениям изучения той или иной целостности. При этом для каждой целостности существует два типа структурных схем, отражающих ее объективные и субъективные модели. Структурные схемы, отражающие объективные и субъективные модели объектов, могут иметь одинаковые элементы, но в целом они различны.

#### Поисковые комплексы

Системы этой категории автоматизируют функцию поиска информации, оставляя задачу реализации функции аналитики на эксперте.

#### Решение от «Интегрум-Техно»

Информационное агентство «Интегрум-Техно» – оператор крупнейшей в России полнотекстовой онлайн-службы баз данных, содержащей политическую, коммер-

ческую, юридическую, научную, адресно-справочную информацию.

В основе работы агентства лежит технология с использованием информационной поисковой системы (ИПС) «Артефакт», предназначенной для накопления и хранения больших объемов текстовой и графической информации с целью выполнения эффективного поиска по всей коллекции накопленных баз данных.

Основными единицами информации в ИПС «Артефакт» являются произвольные файлы, в том числе индексируемые файлы в формате html или text, которые могут содержать произвольную информацию на естественном языке: русском, ан-



глийском или других европейских языках. Необязательное разделение документа на текстовые поля требует конвертирования входящих структурированных данных и позволяет структурировать документ.

Образ документа помещается в базу данных на этапе загрузки. ИПС «Артефакт» является полнотекстовой системой в том смысле, что все слова, составляющие документ базы данных, учитываются при поиске.

«Артефакт» позволяет учесть в запросе любые факты, связанные с естественной структурой текста (разделение на слова, предложения, абзацы и поля, взаимное расположение слов и расстояние между ними), и предоставляет пользователю все необходимые логические операторы для формирования запроса (И, ИЛИ, НЕ, скобки, контекстные операторы). ИПС Артефакт поддерживает также операцию усечения слов (поиск по фрагменту слова): окончания, начала и двустороннее усечение.

ИПС проводит распознавание дат, заданных в различных форматах, и реализует как поиск точной даты, так и поиск по диапазону дат, заданному в запросе.

Лингвистический аппарат основан на морфологическом словаре А. А. Зализняка.

В «Артефакте» реализован также механизм поиска с учетом опечаток (наиболее типичных, встречающихся в 9 случаях из 10). Существует возможность поиска не только слов с единой основой, но и однокоренных слов.

Размер отдельной базы данных ограничивается лишь операционной средой пользователя и аппаратным обеспечением. Реализовано использование многобазового поиска (неограниченного числа баз данных одновременно).

ИПС «Артефакт» является системой типа «клиент-сервер» и может работать в интернет/интранет режимах. Во всех случаях клиентом является браузер (например, Microsoft Internet Explorer или Netscape Navigator). Используемая операционная система – Microsoft Windows NT.

Представление результатов поиска в ИПС «Артефакт» многоступенчатое: сначала показываются заголовки документов и релевантные фрагменты текста, затем текст документов, в которых релевантные фрагменты и слова также выделены. Просматривать найденные документы можно, как только найден хотя бы один документ, не дожидаясь окончания поиска.

При увеличении числа баз данных существует возможность их объединения в группу. Для каждой базы данных можно создать краткое описание, что облегчает работу с ней конечного пользователя.

#### Решение от «МедиаЛингва»

ЗАО «МедиаЛингва» – ведущий разработчик программного обеспечения, реализующего лингвистические, поисковые и мультимедиа технологии. Основанная в августе 1995 года, компания обладает опытом создания готовых программных продуктов для дома и офиса, технологий для разработчиков, а также сложных корпоративных систем.

Продуктами компании, представляющими интерес для настоящего обзора, являются поисковая система «Следопыт 3.0», позволяющая быстро и качественно находить документы на персональном компьютере, и «Серверный следопыт» для полнотекстового поиска

документов на корпоративном web-узле или внутреннем сервере интрасети.



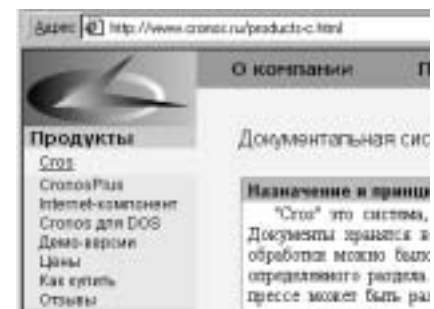
## МедиаЛингва

«Следопыт» – поисковая система, предоставляющая возможность поиска нужной информации на отдельном web-сайте или сервере корпоративной интрасети. Поиск осуществляется по содержанию документов и их атрибутам, а также по размеру, имени, дате создания, по отправителю или получателю почтового сообщения. Программа может обрабатывать файлы практически всех форматов: doc, rtf, html, xls, pdf, zip, pst, а также папки (как сами сообщения, так и вложения) Microsoft Outlook. В системе реализован морфологический поиск, то есть для каждого слова учитывается вся парадигма. Фильтр для формата pdf при работе с русским языком является в «Следопыте» одним из лучших.

#### Решение от «Кронос-Информ»

Система полнотекстового поиска CROS 4.01 предназначена для накопления и обработки текстовых документов различных форматов. Хранение документов в базах данных системы обеспечивает двух-, трехразовое уменьшение необходимого объема дисковой памяти. Предусмотрено автоматическое определение форматов документов Microsoft Word версий 6.0, 7.0, 97, 2000, а также rtf и html. Помимо этого, определяется тип кодировки (DOS, Win, KOI8, Unicode).

CROS обеспечивает навигацию по найденным документам, под-



держивает защиту информации от несанкционированного доступа и способен работать в локальной сети. При этом отсутствуют ограничения на количество иерархических областей поиска, осуществляется сортировка найденных документов по дате, имени, типу и атрибутам, которые задаются самим пользователем.

### Системы класса Business Intelligence

Business Intelligence или BI-системы в основном специализируются на реализации технологии Data Mining (извлечение данных). Другими словами, наиболее успешным будет их применение для обработки больших баз данных. Эти системы, как правило, обладают мощной автоматизацией процессов аналитики.

### Решение от Megaputer Intelligence

Решение PolyAnalyst – это система, предназначенная для автоматического и полуавтоматического анализа числовых и текстовых баз данных с целью обнаружения в них закономерностей, которые необходимы для принятия оптимальных ре-



шений в бизнесе и в других областях человеческой деятельности.

PolyAnalyst является клиент/серверным приложением. Пользователь работает с клиентской программой PolyAnalyst Workplace. Математические модули выделены в серверную часть – PolyAnalyst Knowledge Server. Такая архитектура предоставляет естественную возможность для масштабирования системы от однопользовательского варианта до корпоративного решения с несколькими серверами.

PolyAnalyst работает с разными типами данных: числами, логическими переменными, категориальными переменными, текстовыми

строками, датами, а также со свободным текстом. PolyAnalyst может обрабатывать исходные данные из различных источников, таких как файлы Microsoft Excel 97/2000, любая ODBC-совместимая СУБД, SAS data files, Oracle Express, IBM Visual Warehouse.

Модули PolyAnalyst используют различные алгоритмы Data и Text Mining, в том числе, модуль Text Categorizer – каталогизатор текстов, который позволяет автоматически создать иерархический древовидный каталог имеющихся текстов и пометить каждый узел этой древовидной структуры как наиболее индикативный для текстов, относящихся к данному узлу.

Модуль Link Terms обеспечивает связь понятий. Он позволяет выявлять связи между понятиями, встречающимися в текстовых полях изучаемой базы данных, и представлять их в виде графа, который может быть использован для выделения записей, реализующих выбранную связь.

Модуль Link Analysis выявляет корреляционные и антикорреляционные связи между значениями категориальных и булевых полей.

Благодаря уникальной технологии «эволюционного программирования» и другим интеллектуальным алгоритмам PolyAnalyst с успехом применяется в различных типах бизнес-задач, в социологических исследованиях, в прикладных научных и инженерных задачах, в банковском деле, в страховании и медицине.

### Решение от Cognos Business Intelligence

В классе информационно-аналитических систем компания АйТи предлагает комплекс решений Cognos Business Intelligence производства корпорации Cognos, признанного лидера в области создания ин-

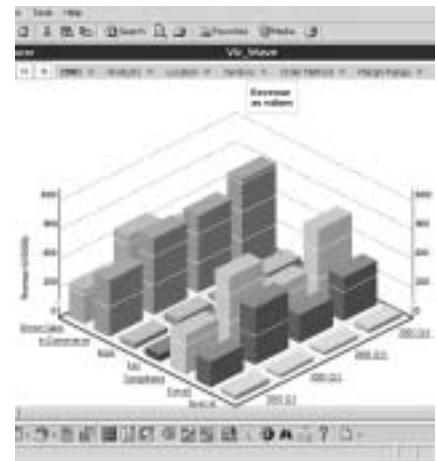


струментальных средств аналитической обработки данных.

Решения Cognos Business Intelligence предоставляют мощные и эффективные инструменты оператив-

ного анализа, реализованные на платформе OLAP-технологии.

Способность к интеграции с компонентами других корпоративных информационных систем позволяет не только использовать инструментальные средства Cognos BI как самостоятельный инструмент сбора и анализа данных, но и существенным образом расширить аналитиче-



ские возможности других систем, будь то системы финансово-экономического планирования, CRM- или ERP-системы. Простые и удобные в использовании, аналитические инструменты семейства Cognos BI дают менеджерам практически безграничные возможности в областях:

- сбора и консолидации данных из всего многообразия внутренних и внешних источников;
- многогранного и оперативного интеллектуального анализа информации под любым углом зрения, обеспечивающего реальное многоуровневое видение текущей ситуации;
- формирования запросов и отчетов различной сложности, комбинирующих данные, необходимые для принятия решений и поступающие из разных источников;
- визуального представления информации с возможностью выделения только необходимой информации из всего объема имеющихся данных и расстановки нужных акцентов;
- графического оформления отчетов: от таблиц и многомерных диаграмм вплоть до географических карт для отображения региональных зависимостей.

## Все системы хороши – выбирай на вкус

Анализ спектра программных средств, представленных на рынке информационных технологий, демонстрирует нам широкий выбор возможностей использования средств для автоматизации аналитической деятельности. Основными факторами, определяющими выбор того или иного программного продукта, являются:

- цели и задачи аналитического подразделения;
- объемы и состав обрабатываемой информации;
- размеры предприятия;
- стоимость программных решений, удовлетворяющих вышеперечисленным требованиям и имеющийся бюджет.

Как уже отмечалось выше, наибольшими возможностями по сбору информации и извлечению из нее знаний будут обладать полнофункциональные программные комплексы, так как они обеспечат аналитику целостную информационную картину, пополняемую как внешними источниками, так и внутренней информацией предприятия.

Однако не следует думать, что на выполнение бизнес-разведки способны только крупные корпорации и концерны. Во-первых, сила действия всегда равна силе противодействия, то есть величина компании и ее возможности по влиянию на рынок, на котором она работает, будет прямо пропорциональны количеству информации, необходимой для аналитической обработки. И если крупному предприятию необходимо учитывать макроэкономические факторы, политическую ситуацию, состояние мировой экономики, состояние отраслей поставщиков сырья и комплектующих, то малое предприятие может ограничиваться мониторингом рынка, конкурентов, правовой и налоговой базы для принятия правильных управленческих решений. Во-вторых, существует достаточное количество программных средств, имеющих модульную структуру, которая позволит на начальном этапе использовать минимально необходимую недорогую

функциональность, дополняемую организационными мерами, с возможностью наращивания мощности комплекса по мере роста бизнеса предприятия. В этом случае, важно правильно распределить приоритетность наращивания информационных задач.

В качестве первого шага в направлении создания регулярной бизнес-разведки можно предложить анализ открытой прессы, поставленный на регулярную основу и имеющий вполне определенные цели: мониторинг действий конкурентов, мониторинг состояния отрасли. Трудно найти делового человека, который не следит за новостями. Но мало кто понимает, что данное занятие, поставленное на систематическую основу и имеющее определенные цели получения информации, является составной частью разведывательного цикла, или говоря более простым языком – информационно-аналитической работы.

Простое систематическое «чтение» прессы (если хотите – «мониторинг прессы», «пресс-клиппинг») позволяет получить массу информации, имеющей практическую ценность для ведения бизнеса. Какие мониторинги могут выполняться обычным представителем малого бизнеса и что они ему дают? В качестве элементарных – это мониторинг упоминаний предприятия в СМИ и мониторинг упоминаний о конкурентах в СМИ.

Какие программные средства помогут предпринимателю в этом? На первом этапе достаточно установить у себя поисковые комплексы и наладить канал поступления информации в них. Такое простейшее решение при определенных навыках анализа информации и работы с поисковым комплексом даст значительный результат. Причем, он будет тем качественнее, чем больший массив информации будет привлекаться для работы. При выборе комплекса необходимо принимать в расчет возможность дальнейшего наращивания его мощности и функциональности, качество его технического сопровождения и перспективы в будущем (устойчивость компании-раз-

работчика). Правильный учет указанных факторов даст гарантии предприятию, что инвестиции на автоматизацию аналитической деятельности не будут выброшены на ветер, а выльются в конкурентное преимущество компании, обеспечат рост бизнеса, снизят риски хозяйствования. ■

### ЛИТЕРАТУРА

1. А. И. Доронин, *Бизнес разведка*. – «Ось-89», М. – 2003.
2. В. Шарлот, *Конкурентная разведка в России: понимать, осуществлять, учить. Современные проблемы становления*, [www.sec.ru](http://www.sec.ru), 2002.
3. *Информация и знания: невидимая грань* И. Ашманов, [www.newasp.omskreg.ru](http://www.newasp.omskreg.ru)
4. *Программные продукты Hummingbird*, <http://www.uni.ru/dist/software/hummingbird/>
5. *Программные продукты Documentum*, [http://www.documentum.ru/products/client\\_applications.html](http://www.documentum.ru/products/client_applications.html)
6. *Информационно-аналитические системы Cognitive Эксперт*, <http://www.cognitive.ru/decisions/ias.htm>
7. *Информационно-аналитические Системы поддержки принятия решений (DSS) и средства их реализации*. ЗАО «ТЕРН» (Москва), <http://www.ellips.ru/seminar/1999/tern1.htm>
8. *Добыча знаний Д. Ландэ «Телеком» 1-2/2004*, <http://www.visti.net/~dwl/art/cz/index1.html>
9. *Семейство продуктов Convera. ВестьМетаТехнология*, <http://www.vest-meta.ru/product/Convera/products.asp>
10. *Управление знаниями – решения для бизнеса. НейроОк Интелсофт*, <http://soft.neurok.ru/products/index.shtml>
11. *Программы анализа. Информбюро*. <http://www.informburo.net/index.php?option=displaypage&Itemid=96&op=page&SubMenu=>
12. *Автоматизированная система поиска и анализа информации. Галактика-зум*. <http://zoom.galaktika.ru/content.htm>
13. *Российская дипломатия и информационные технологии*. А. И. Смирнов, <http://www.polpred.ru/free/smi/8.htm>
14. *Услуги и продукты. Интегрум-Техно*, <http://www.integrum.ru/products/>
15. *Поисковая система «Следопыт». Медиалингва*, <http://www.sledopyt.ru/>
16. *Документальная система поиска информации CROS. Кронос-Информ*, <http://www.cronos.ru/products-c.html>
17. *Polyanalyst & Data Mining. Меганьютер Интеллидженс*, [http://www.megaputer.ru/doc.php?production/pa/polyanalyst\\_info.html](http://www.megaputer.ru/doc.php?production/pa/polyanalyst_info.html)
18. *Аналитические системы. АйТи*, <http://www.it.ru/service/strategy/analytic.html>