

Школа веб-разведки

Конкурентная разведка — важнейшая функция современного менеджмента и главное условие динамичного развития бизнеса. В открытых источниках и специализированных БД, доступных в Интернете, содержится большая часть информации, необходимой для проведения конкурентной разведки, однако остается открытым вопрос ее нахождения и эффективного использования.

Некоторое время назад в широком обиходе бизнес-структур на Западе появились термины «бизнес-разведка» (Business Intelligence, или BI) и «конкурентная разведка» (Competitive Intelligence, или CI). Этими терминами стали обозначать сбор и анализ информации (о рынках товаров и услуг, клиентах, контрагентах, конкурентах и технологиях), необходимой для успешного развития бизнеса. Зыбкая грань, которую пытаются провести между этими терминами и промышленным шпионажем, лежит в легитимности или противозаконности методов и средств, используемых в процессе сбора такой информации. Следует отметить также весьма тонкую разницу (на которую указывают некоторые западные авторы систем или публикаций) между Business

Intelligence и Competitive Intelligence. Из публикаций и описаний систем, где упоминаются эти термины, можно сделать вывод, что BI (см. «7 шагов внедрения Business Intelligence», ТЕЛЕКОМ 1–2/07) направлена больше на изучение внутренней маркетинговой, финансовой, экономической информации и данных о клиентах, в то время как термином CI чаще обозначаются процессы, связанные с добыванием внешней информации и знаний непосредственно о конкурентах. На постсоветском пространстве, где еще окончательно не сформировалась

понятийная база,
как пра-
в и -

ло, не делают разницы между терминами «деловая», «конкурентная» или «экономическая» разведка и понимают весь комплекс мероприятий, связанный с информационно-аналитическим обеспечением управления предпринимательскими рисками, выявления угроз, возможностей и других факторов, влияющих на получение конкурентных преимуществ в бизнесе.

Немного теории...

Методики и технологии ведения конкурентной разведки (КР) в бизнесе весьма близки к используемым в традиционной разведывательной деятельности. В соответствии с общепринятой теорией разведывательной информации, основным принципом, который положен в основу разведдеятельности как разновидности информационной работы, есть четкое разделение таких понятий, как данные, информация и знания. Под данными понимают «сырые» необработанные сведения, основанные на фактах. Это могут быть статистические данные, факты из биографии или сведения из финансовой отчетности предприятия. Информация представляет собой уже определенным образом отобранные, обработанные и проанализированные данные. Конечным же информационным продуктом любой разведки должны быть знания —



синтезированные выводы, рекомендации и решения.

Информация характеризуется качественными, количественными и ценностными показателями. К качественным обычно относят: достоверность (чистоту), объективность и однозначность информации. А к количественным — ее полноту (отсутствие невыясненных пробелов) и релевантность (степень соответствия существу поставленных вопросов и задач). Ценностными характеристиками информации являются ее стоимость и актуальность.

Процесс превращения сырых данных в знания и доведение их до конечных потребителей принято называть разведывательным циклом. В своем классическом понимании разведчик принято делить на пять основных этапов:

- 1) целеуказание, планирование, определение источников информации;
- 2) сбор, добывание данных;
- 3) обработка разведанных — превращение их в разведывательную информацию;
- 4) анализ и синтез разведанной информации — превращение ее в знания: выводы, рекомендации, решения;
- 5) доведение информации до конечных потребителей.

Не вдаваясь в подробности, следует также отметить некоторые клю-

чевые особенности указанных выше этапов превращения данных в знания, на которые хотелось бы обратить внимание. А именно:

- ✓ целеуказание и планирование целесообразно делить на три уровня — стратегический, тактический и оперативный;
- ✓ на этапе сбора информации крайне важно задействовать как можно большее количество независимых и первичных источников;
- ✓ процесс обработки данных предполагает учет, классификацию, отбор, верификацию и оценку добытых сведений;
- ✓ разведчикл, не требующий в некоторых случаях глубокой прора-

ботки, или в условиях ограниченного времени может быть не полным и заканчиваться на выдаче потребителям не знаний (в виде окончательных выводов, рекомендаций или проектов решений), а просто обработанной информации в виде информационных справок;

✓ в разведывательном документе не должно быть ссылок на конфиденциальные источники информации, поскольку это может привести к их расшифровке;

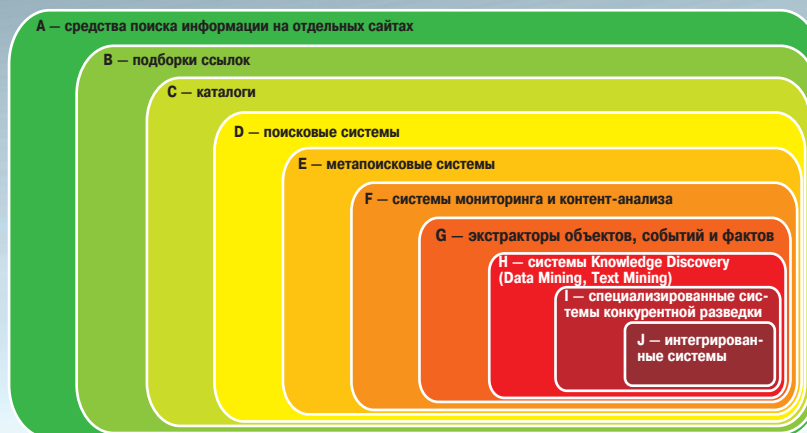
✓ выводы и рекомендации должны быть четкими, краткими и однозначными, а прогнозы носить вероятностный характер;

✓ доведение информации до конечных потребителей должно быть в виде, адаптированном к восприятию заказчика, и форме, легко доступной их пониманию (в связи с этим любопытно заметить, что ЦРУ, например, предоставляло Рональду Рейгану ежедневную информацию в виде видеофильма, который снимали каждый день, поскольку бывший киноактер воспринимал такую подачу информации более адекватно).

Без Интернета как без рук

У конкурентной разведки в последние десятилетия появился и развился до невиданных ранее масштабов новый информационный источник — веб-пространство. Сегодня, как утверждают эксперты, Интернет по количеству информации находится на первом месте, опережая СМИ, отраслевые издания и получаемые от коллег новости, специальные обзоры, закрытые базы данных.

Иерархия поисковых систем



Средства поиска информации в Интернете делятся на несколько подгрупп — по принципам работы и обеспечиваемому функционалу

Задачи конкурентной разведки

Основными целями и задачами, которые ставятся перед подразделениями экономической разведки «среднестатистической» бизнес-корпорации на постсоветском пространстве являются:

- 1) сбор и своевременное обеспечение надежной и всесторонней информацией руководства и бизнес-подразделений о внешней и внутренней среде предприятия;
- 2) выявление факторов риска, угроз, которые могут затронуть экономические интересы бизнеса или помешать его нормальному функционированию;
- 3) выявление новых возможностей и других факторов, влияющих на получение конкурентных преимуществ;
- 4) выработка прогнозов и рекомендаций влияния конкурентной среды на деятельность бизнес-структуры;
- 5) усиление благоприятных и локализация неблагоприятных факторов конкурентной среды на деятельность бизнес-структуры.

Примечание. Здесь сформулированы цели и задачи конкурентной разведки, рассчитанные на легитимную деятельность соответствующих подразделений. Хотя справедливости ради следует отметить, что вся наша недавняя история, к сожалению, пестрит многочисленными примерами использования служб безопасности и экономической разведки коммерческих структур в противозаконных целях и с применением таких же методов: от незаконной прослушки и видеосъёмки до ликвидации конкурентов. Авторы статьи своей целью видят пропаганду ведения конкурентной разведки исключительно в рамках закона.

Последние исследования интернет-пространства показали, что доступные через традиционные информационно-поисковые системы 20 млрд веб-страниц — это лишь видимая часть айсберга. Количество веб-сайтов в Сети увеличивается со скоростью более миллиона узлов в месяц. Около 40 % всей информации в Интернете доступно бесплатно. Навигацию по этому информационному пространству обеспечивают более миллиона поисковых систем и каталогов, но и они охватывают лишь малую часть информационных ресурсов. А скрытых и невидимых ресурсов Сети значительно больше.

Google, сможет проиндексировать всю информацию примерно лишь через 300 лет.

Традиционные системы поиска в Интернете отлично справляются с простыми однократными запросами, но, как правило, в большинстве случаев практически не применимы для нужд конкурентной разведки. По некоторым оценкам, более 97 % критичной для КР онлайн-информации невозможно найти с помощью традиционных информационно-поисковых систем.

Второй, после описанной выше, проблемой конкурентной веб-разведки является то, что информация

Около 40 % всех охваченных Интернетом данных доступно бесплатно. При этом число веб-сайтов растет со скоростью более 1 млн узлов в месяц

Это прежде всего информация из многочисленных баз данных, динамически генерируемые страницы, файлы разнообразных форматов, которые могут представлять большой интерес для любой разведки, не говоря уже о государственных и негосударственных виртуальных частных сетях (VPN), построенных с использованием зашифрованных каналов связи и других средств защиты поперек TCP/IP.

Проблемы конкурентной разведки в Сети

Первой и наиболее существенной проблемой является то, что колоссальные объемы информации в Интернете затрудняют поиск и выбор действительно нужных сведений. Ведь сами по себе необработанные, необобщенные и непроверенные данные, как уже говорилось выше, не могут служить подспорьем для принятия решений.

По оценкам специалистов, ресурсы всемирной паутины (проиндексированные поисковиками и скрытые) составляют более 550 млрд документов и постоянно растут. По словам Эрика Шмидта, главы Google, в настоящее время из пяти миллионов терабайт существующей информации проиндексировано лишь 170 Тбайт, и даже такая мощная поисковая система, как

в Интернете имеет явно выраженный динамичный характер: она размещается на сайтах, модифицируется и удаляется во времени. Частичное решение этих проблем возможно при применении систем контент-мониторинга информационных потоков в Интернете. Использование при сборе информации в Сети системы контент-мониторинга в среднем увеличивает число найденных релевантных документов в 20 раз.

Третья задача, которую необходимо решить в целях КР, — автоматическое извлечение понятий из формализованных массивов информации (таблиц, БД) и неструктурированных текстов. Перспективным направлением решения данной проблемы в системах КР является применение технологий Knowledge Discovery (поиск знаний), использующих концепции Data Mining и Text Mining.

Четвертой проблемой является выявление неочевидных закономерностей и связей. В настоящее время известно несколько путей решения извлечения понятий из текстов и выявления их взаимосвязей (как практических, так и теоретических). Одним из них есть построение матриц и графов взаимосвязей и гистограмм распределения понятий. Матрицы можно рассматри-

По ту сторону веба

В качестве примера «теневого» веб-пространства можно привести американские закрытые сети INTERLINK, POLICYNET, SIPRNET (Secret Internet Protocol Router Network), информация о которых просочилась в открытые источники. К скрытому вебу можно отнести также такие сети, как Ares, BitTorrent, DirectConnect, EMule, Gnutella, Kazaa/FastTrack, Napster, Piolet и др.

Что касается закрытых частных сетей Virtual Private Network, то этот сегмент сегодня бурно развивается. По мере роста ценности информации и обострения рыночной конкуренции продажи шлюзов SSL VPN постоянно увеличиваются, сообщают аналитики. Лидирует в данной области небезызвестный игрок на рынке защиты сетей и контента — компания Juniper Networks.

вать как неориентированные графы и применять к ним соответствующие математические методы. Как правило, узлы этих графов — коэффициенты, которые пропорциональны количеству документов входного информационного потока, одновременно соответствующие обоим понятиям, или количеству других понятий, употребляемых совместно с данными. Другим перспективным решением этой задачи можно считать построение визуального графа связей — информационного портрета найденных релевантных документов с использованием AJAX-технологии.

Пятая проблема — это поиск информации в «скрытом» веб-пространстве, где, как было замечено выше, содержится несравнимо большее количество данных, потенциально интересных для конкурентной разведки, чем в открытой части Всемирной паутины. Американский журнал Business 2.0 опубликовал список из семи технологий, которые должны «изменить мир» в ближайшем будущем. Среди них — Wi-Fi-мобильники, WiMAX, AJAX и поиск в невидимой сети (в скрытом веб-пространстве). В связи с этим интересно заметить, что две последние технологии крайне полезны для нужд конкурентной разведки.

Таким образом, как мы видим, не вся потенциально открытая «несекретная» информация является хорошо доступной, скорее — наоборот. Извлечение необходимой в каждом конкретном случае информации является сложной задачей. По мнению экспертов, только порядка 10–15 % нужной информации имеется в Интернете в готовом виде, остальные 85–90 % можно получить в результате сравнения, интеграции и анализа многочисленных разрозненных данных.

Вопрос нахождения в вебе и эффективного использования информации, необходимой для проведения КР, пока не решен — из-за присущих Интернету недостатков: непропорционального роста уровня информационного шума, засилья паразитной информации, слабой структурированности и связанности между собой информации, ее динамичности и отсутствия гарантии ее целостности, многократного дублирования, отсутствия возможности смыслового поиска, ограниченности доступа к скрытому вебу. В целом же возможности Интернета оцениваются экспертами в области конкурентной разведки достаточно высоко.

За внешним хаосом информации, доступной сегодня в Интернете, скрываются глубинные связи, достоверная информация и дезинформация. Разобраться с этими данными можно, только применяя современные средства анализа содержания текстов (контент-анализа), контент-мониторинга, глубинного анализа данных и текстов, а мониторинг и качественная обработка информационных потоков способны принести ощутимое конкурентное преимущество.

Средства поиска информации в Интернете

Итак, чтобы получить крупницы необходимой информации в Сети, необходимо «перелопатить» огромные массивы сырых данных. Естественно, что для облегчения этой задачи нужны некие поисковые инструменты.

Все имеющиеся средства поиска информации в вебе могут быть условно разделены на несколько подгрупп, а именно:

A — средства поиска информации на отдельных сайтах;

B — подборки ссылок;

C — каталоги;

D — поисковые системы;

E — метапоисковые системы;

F — системы мониторинга и контент-анализа;

G — экстракторы объектов, событий и фактов;

H — системы Knowledge Discovery (Data Mining, Text Mining);

I — специализированные системы конкурентной разведки;

J — интегрированные системы.

Если средства поиска на отдельных сайтах и подборки ссылок (зоны A и B) не требуют специальных пояснений, то для отражения сути вопроса по остальным средствам поиска хотелось бы сказать несколько слов.

По своей сути каталоги, поисковые системы и метапоисковые системы являются сайтами со специализированными базами данных, в которых хранится информация о других интернет-сайтах и документах на них. В подобных БД хранятся не сами документы, а информация (аннотация) и ссылки на них. В результате запроса к таким системам выдается список адресов (а иногда и краткое описание документов), где может присутство-

XV Міжнародна спеціалізована виставка

ТЕХНОЛОГІІ ІІІ ТИСЯЧОЛІТТЯ

ІНФОРМАТИКА І ЗВ'ЯЗОК 2007

www.informatika.net.ua
www.pe.com.ua

10-12 ЖОВТНЯ
МВЦ > Броварський пр-т, 15 > Київ > Україна

Генеральні інформаційні спонсори:



Організатори:



Прем'єр Експо
04050, Київ, вул. Пимоненка, 13-Б
Тел. +380 44 451 4160
Факс: +380 44 451 4161
E-mail: Rsologub@pe.com.ua
www.informatika.net.ua
www.pe.com.ua

вать запрашиваемая информация. Как правило, поиск может производиться только по ключевым словам и фразам. Нажимая на ссылку, найденную в результате запроса, вы попадаете на оригинал документа, находящегося на исходном сайте. Естественно, если документ со временем изменился или сайт прекратил свое существование, то и первоначально найденный оригинал документа через некоторое время может быть не найден.

Основное отличие поисковых систем (зона D) от каталогов (зона

C) — автоматический «робот» или «паук», который постоянно сканирует Интернет и накапливает новую информацию в индексных файлах базы данных. В каталоги же информация заносится инициативно — либо авторами новых сайтов, либо обслуживающим персоналом самих каталогов. Так как пользование такими системами, как правило, бесплатное, то на сегодняшний день это самые популярные поисковые инструменты в Сети.

Метапоисковые системы (зона E) являются «поисковыми система-

ми по поисковым системам». Так как разные поисковые системы по-разному индексируют хотя и обширные и часто пересекающиеся, но все-таки разные сегменты Сети, то, естественно, и результат поиска с помощью метапоисковой системы будет априори более полным, чем с помощью одной отдельно взятой поисковой системы. Считать это плюсом или минусом поисковой работы зависит от поставленных вами целей и количества найденных документов. Если в результате поиска найдены тысячи или миллионы документов, то «вручную» вы все равно вряд ли сможете просмотреть более нескольких сотен.

Вторым поисковым преимуществом таких систем является то, что одним запросом вы обеспечиваете поиск во многих поисковых системах, не утруждая себя многочисленными повторениями одного и того же запроса в разных поисковиках.

Системы мониторинга и контент-анализа (зона F) обеспечивают регулярный поиск и загрузку информации по заданным темам и с заданных сайтов, а также анализ содержания загруженных документов. Такие системы, как правило, обладают развитым языком запросов, что позволяет существенно детализировать и конкретизировать запросы по сравнению с обычными поисковыми системами. Кроме того, такие системы хранят в своих базах данных полные тексты исходных документов, что обеспечивает сохранность этих документов во времени и возможность их обработки и контент-анализа как в текущем времени, так и в ретроспективе. Существенным преимуществом таких систем является также то, что сложные запросы, состоящие из десятков или сотен поисковых слов и выражений, однажды составленные аналитиком-знатоком предметной области, могут быть сохранены в виде каталогизированного запроса или рубрики и в дальнейшем вызываться автоматически или вручную из сохраненного списка для проведения поиска или контент-анализа.

С помощью контент-анализа такие системы позволяют устанавливать пересекающиеся связи между темами, понятиями и объектами, поставленными на мониторинг, выявлять эмоциональную окраску доку-

С миру по нитке: голые факты

В 1990 году, в последний момент, непосредственно перед подписанием контракта расстроилась 200-миллионная сделка между японской компанией NEC и Индонезией на поставку телекоммуникационного оборудования. Контракт был передан американской AT&T.

В 1994 году Франция неожиданно потеряла заказ на поставку в Саудовскую Аравию военной техники на сумму \$5 млрд. Этот заказ получили американцы из «Макдонелл Дуглас».

В 1999 году сорвалась сделка по поставке в одну из стран Персидского залива европейских Airbus A-310. Договор был заключен с американской компанией «Боинг».

В конце 90-х годов XX в. американские спецслужбы активно собирали информацию по слиянию концернов «Даймлер» и «Крайслер», а также финансовых структур «Дойче банка» и американского «Банкерс Траст».

Это лишь немногие, ставшие известными, факты промышленного шпионажа с использованием системы «Эшелон» Агентства национальной безопасности США (АНБ) и спецслужб других государств, озвученные в ходе слушаний по результатам расследования, проводившегося депутатами Европарламента в 2000 году. По данным этого расследования, ущерб от промышленного шпионажа с использованием системы «Эшелон» нанес странам Европы убытки в размере как минимум \$20 млрд.

Главный этап развития этой системы пришелся на постсоветскую эпоху, когда наряду с военно-политическими целями первостепенными задачами разведывательных служб стали борьба с терроризмом, организованной преступностью и... промышленный шпионаж.

АНБ США признало, что передавало американским корпорациям значимую информацию о европейских и других конкурентах. При этом, как выяснилось, в разработке словарей-ловушек для системы «Эшелон» принимали участие американские фирмы, заинтересованные в получении конкурентной информации. Бюджет АНБ с конца 90-х годов стал превышать бюджет ЦРУ и ФБР, вместе взятых, и исчисляется миллиардами долларов.

По результатам опроса в 1999 году 500 крупнейших компаний США, 90 % из них подтвердили, что занимаются конкурентной разведкой, для чего создали у себя соответствующие подразделения и ввели в руководство посты директоров по бизнес-разведке. При этом затраты этих корпораций на разведку составили в среднем 1–1,5 % от оборота и оказались вполне рентабельными.

В качестве классического примера рентабельности экономической разведки можно привести исторический факт 1971 года, когда французская внешняя разведка (DGSE) получила сведения о готовящейся девальвации доллара и незамедлительно довела эти данные до Президента Ж. Помпиду, который дал соответствующие указания Центробанку. В результате французская казна заработала сотни миллионов на продаже долларов и скупке франков на мировых валютных рынках.

В то же время к явным неудачам разведывательных структур — как государственных, так и корпоративных — можно отнести отсутствие соответствующих прогнозов и предупреждений о надвигающемся кризисе в России в августе 1998 года, который серьезно подорвал экономику как отдельных предприятий и граждан, так и страны в целом.

ТЕЛЕКОМ-ИНФО

ментов, оценивать динамику появления во времени тех или иных документов, проводить сравнительный анализ информационной активности по различным тематикам и многое другое.

Если мониторинговые системы могут выделять из информационного потока поставленные на мониторинг известные объекты, то экстракторы объектов, событий и фактов (зона G) умеют выделять из потока информации неизвестные ранее объекты, события или факты, которые соответствуют определенному заранее заданному типу. Примерами типов объектов могут служить географические понятия, персоны, структуры и организации, события (дорожно-транспортные происшествия, катастрофы, международные встречи, обычные или необычные факты). Обычным фактом в данном случае можно, например, считать выезд автомобилей за черту города, а необычным — выезд за ту же городскую черту автомобиля без номерных знаков.

Системы Knowledge Discovery (зона H), использующие технологии Data Mining и Text Mining, умеют выявлять новые знания и закономерности. Такая система, например, способна самостоятельно, без участия человека, сделать вывод о факте знакомства между людьми, основываясь на имеющихся в системе данных об окончании ими одной и той же школы, одного и того же класса в одном и том же населенном пункте. Правда, пока сами правила, по которым такая система делает выводы, все-таки создаются и задаются людьми.

Специализированные системы конкурентной разведки (зона I) могут включать в себя одно или несколько из перечисленных выше поисковых средств, специально «заточенных» под эти специфические задачи. Кроме того, потребности конкурентной разведки предполагают в качестве источников информации помимо полнотекстовых документов из Интернета еще и доступные в нем базы данных, собственные, принадлежащие структуре, документы, таблицы и БД, а также формализованные и неформализованные документы и БД, добытые из других источников.

О роли открытых источников

По устоявшемуся, но все же ошибочному мнению, сформированному многочисленными шпионскими сериалами и детективами, вся полезная разведывательная информация добывается из строго секретных источников агентурным или оперативным путем и обязательно сопровождается стрельбой, погонями, взломами кодов, замков и шифров или в крайнем случае соблазнением прекрасной дивноокой и длинноногой носительницы секретов. На самом деле это не так. Известное признание адмирала Захариаса — заместителя начальника разведки ВМС США в годы второй мировой войны — опровергает это. Так, по его оценке, 95 % информации разведка военно-морских сил черпала из открытых источников, 4 % — из официальных и только 1 % — из конфиденциальных. Справедливости ради надо сказать, что часто именно этот один процент является тем золотым недостающим звеном, который позволяет сложить целостную картину разрозненной мозаики всех разведанных. И если такое соотношение справедливо для военной разведки, то тем более оно будет правильным для конкурентной разведки бизнес-структур.

Разведывательная информация может быть получена из официальных источников, неофициальных открытых источников, СМИ, объявлений, рекламы, внутрифирменных, банковских, правительственных отчетов, баз данных, от экспертов, путем анализа или специальной обработки данных, текстов по прямым или косвенным признакам. Правда, при этом количество разнородных сведений, которые необходимо переработать, чтобы получить крупницы знаний, огромно. А потому в настоящее время конкурентная разведка немыслима без использования специализированных информационных технологий.

На Западе самый обычный, ничем не выдающийся из общей массы человек зарегистрирован в более чем 300 базах данных: прописка, страховка, водительские права, кредитные бюро, информационные, рейтинговые, рекрутинговые агентства, бюро по трудоустройству, медицинские и полицейские учеты, супермаркеты, клубы, банки, системы управления взаимоотношений с клиентами (CRM) и т. д. В интересах конкурентной разведки и маркетинга анализируются не только рынки товаров и услуг, но также вкусы и предпочтения отдельных клиентов. Информация, хранящаяся в различных базах данных о юридических лицах, еще более обширна.

В целях конкурентной разведки необходимо анализировать данные из всех доступных источников информации, но, так как эти вопросы выходят за рамки данной статьи, мы не будем касаться других источников информации, кроме Интернета.

Под интегрированными средствами конкурентной разведки (зона J) понимается интегрированная система, включающая в себя не только все доступные поисковые средства, но также интегрированный банк выявленных (добытых) и логично связанных между собой дан-

ных, информации и знаний. С точки зрения создания информационно-аналитических систем такая система концептуально должна предполагать реализацию следующих трех принципов:

- ✓единое информационное пространство взаимосвязанных объектов и фактов независимо от типа их источников или контента;
- ✓сохранение связи объектов и фактов с релевантными данными и источниками информации (аргументированность);
- ✓исторически пространственная модель банка данных системы, что предполагает наличие у всех объектов учета атрибутов времени и места, а также невозможность их безвозвратного удаления из системы.

Справедливости ради следует отметить, что, согласно отчету Fuld's Intellgence Software Report 2006, известных коммерческих версий полноценных интегрированных систем, позволяющих решать весь комплекс задач конкурентной разведки, пока не существует, по крайней мере на Западе. ●

Продолжение статьи читайте в ближайшем номере

Дмитрий Ландэ,
Виктор Прищепа