

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ  
ИНСТИТУТ ПРОБЛЕМ РЕГИСТРАЦИИ ИНФОРМАЦИИ

**А. Г. Додонов, Д. В. Ландэ,  
В. В. Прищепа, В. Г. Путятин**

**КОНКУРЕНТНАЯ РАЗВЕДКА  
В КОМПЬЮТЕРНЫХ СЕТЯХ**

**Киев – 2013**

УДК 004.5  
ББК 22.18, 32.81, 60.54  
С95

А.Г. Додонов, Д.В. Ландэ, В.В. Прищеп, В.Г. Путятин  
Конкурентная разведка в компьютерных сетях. – К.: ИПРИ НАН  
Украины, 2013. – 250 с.

Книга посвящена рассмотрению вопросов интернет-разведки – сегменту конкурентной разведки, охватывающему процедуры сбора и обработки информации, проводимые с целью поддержки принятия управленческих решений, повышения конкурентоспособности исключительно из открытых источников в компьютерных сетях – веб-пространства, блогосферы, форумов, социальных сетей. Рассматриваются различные вопросы информативно-аналитической деятельности в сетевой среде, ориентированной на задачи конкурентной разведки. В качестве теоретических основ конкурентной разведки в сетевой среде рассматриваются элементы теорий сложных сетей, анализа социальных сетей, информационного и математического моделирования, социологии.

Для широкого круга специалистов в области информационных технологий и безопасности.

*Рекомендовано к изданию ученым советом Института проблем  
регистрации информации НАН Украины  
(протокол № 9 от 3 сентября 2013 года)*

*Рецензенты:*

д.т.н., профессор В.В.Мохор  
д.т.н., профессор А.Я. Матов  
д.ю.н., профессор К.И. Беляков

**ISBN 978-966-00-1087-1**

© А.Г. Додонов, Д.В. Ландэ,  
В.В. Прищеп, В.Г. Путятин, 2013

## Оглавление

---

---

Введение .....	5
1. Цели, задачи и проблемы интернет-разведки .....	11
2. Системы и технологии конкурентной разведки .....	16
2.1. Средства поиска информации в Интернет .....	20
2.2. Концепция глубинного анализа текстов .....	26
2.3. Мониторинг информационного пространства .....	30
2.4. Определение взаимосвязей .....	32
2.5. Технологии конкурентной разведки .....	39
2.6. Примеры применения технологий конкурентной разведки .....	64
3. Источники информации .....	72
3.1. Веб-пространство .....	75
3.2. Глубинный веб .....	79
3.3. Специальные базы данных .....	89
3.4. Социальные медиа .....	95
4. Социальные сети .....	98
4.1. Анализ социальных сетей .....	99
4.2. Основные социальные сети .....	108
4.3. Мониторинг социальных сетей .....	112
5. Управление репутацией в сетях .....	117
5.1. Проблема управления репутацией компаний .....	117
5.2. Моделирование репутации в сетях .....	122
5.3. Живучесть информации в Интернете .....	129
6. Правовые вопросы конкурентной разведки .....	137
6.1. Конкурентная разведка в правовом поле .....	137

6.2. Конкурентная разведка и защита коммерческой тайны .....	139
6.3. Конкурентная разведка и защита персональных данных .....	142
7. Информационные операции .....	149
7.1. Информационное влияние.....	153
7.2. Этапность информационных операций.....	155
7.3. Моделирование информационных операций .....	159
7.4. Выявление информационных операций .....	171
7.5. Противодействие информационным операциям.....	184
7.6. Конкурентная разведка в антимонопольной деятельности.....	185
Краткий глоссарий .....	192
Литература.....	209
Веб-сайты по тематике конкурентной разведки .....	215
Адреса упоминаемых веб-ресурсов .....	216

## Введение

---

---

Конкурентная разведка (Competitive Intelligence) в компьютерных сетях охватывает процедуры сбора и обработки информации, проводимые с целью поддержки принятия управленческих решений, повышения конкурентоспособности коммерческих организаций исключительно из открытых источников из компьютерных сетей, большинство из которых являются надстроенными над сетью Интернет, так называемыми, оверлейными. Поэтому часто в качестве синонима конкурентной разведки в дальнейшем будет использоваться термин интернет-разведка. Таким образом, данная книга фактически посвящена проблематике конкурентной разведки, но с одним существенным ограничением – все источники информации, необходимые для проведения разведывательной деятельности, являются открытыми и доступными в компьютерных сетях. Более того, большая часть инструментария, программ обработки информации, также доступна через современные компьютерные сети. В англоязычной литературе такой вид конкурентной разведки принято называть разведкой по открытым источникам (Open Sources INTelligence, OSINT) [Берд, 2007].

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 немислима без использования специализированных информационных технологий.

По мнению бывшего директора Центрального разведывательного управления США (ЦРУ) Р. Хилленкерта «80 % разведывательной информации получается из таких источников как книги, журналы, научно-технические обзоры, фотографии, коммерческих аналитических отчетов, газет, теле- и радиопередач...».

По другим оценкам, в любой разведке от 35 до 95 % всей информации добывается из открытых источников. При этом доля затрат на работу с открытыми источниками, например, в разведывательном бюджете США, составляет лишь около 1 %.

Значимость такой разведки по открытым источникам отметил еще президент США Линдон Джонсон (Lyndon Baines Johnson) 30

июня 1966 г., когда произнес речь на церемонии принятия присяги директором ЦРУ Ричардом М. Хелмсом (Richard McGarragh Helms): «Высшие достижения не являются результатом потихоньку пересказанной тайной информации, а происходят из терпеливого, ежедневного изучения печатных источников».

Общеизвестно, что основное отличие конкурентной разведки от промышленного шпионажа – это легитимность и соблюдение этических норм [Дудихин, 2004]. В интернет-разведке данное положение доведено до абсолюта – исключительно все источники информации в этом случае доступны и легальны.

Интернет-разведка, как, впрочем, и вся конкурентная разведка, представляет собой особый вид информационно-аналитической работы, позволяющей собирать разностороннюю бизнес-информацию без применения тех специфических методов оперативно-розыскной деятельности, которые являются исключительной прерогативой правоохранительных органов.

Вместе с тем, методы ведения интернет-разведки, методики и технологии ее проведения весьма близки к используемым в традиционной разведывательной деятельности спецслужбами.

Применение интернет-разведки в коммерческой компании оправдывается не только соображениями информационной безопасности, но важно и для решения задач менеджмента и маркетинга тем, что обеспечивает:

- наблюдение за репутацией компании (с точки зрения клиентов, конкурентов, госорганов);
- активное участие в формировании имиджа компании, информационного поля вокруг компании;
- отслеживание появления нового конкурента, технологии или канала сбыта;
- выявление возможных слияний и поглощений;
- оценка потенциальных рисков при инвестициях;
- опережение шагов конкурентов в рамках маркетинговых кампаний;
- опережение конкурентов в тендерах;
- выявление каналов утечки информации.

Зыбкая грань между понятиями конкурентная разведка и промышленный шпионаж, состоит в легитимности или законности методов и средств, используемых в процессе сбора целевой информации [Ландэ, Прищепа, 2007]. Следует отметить также весьма тонкую разницу между бизнес-разведкой (business intelligence, BI) и конкурентной разведкой. Из публикаций и описаний систем, где упоминаются эти термины, можно сделать вывод, что бизнес-

разведка направлена больше на изучение «внутренней» маркетинговой, финансовой, экономической информации и информации о клиентах, в то время как конкурентная разведка чаще охватывает процессы, связанные с добыванием «внешней» информации и знаний непосредственно о конкурентах. Однако на практике в постсоветском пространстве, где еще окончательно не сформировалась понятийная база, не делают различия между терминами «деловая», «конкурентная» или «экономическая» разведка, и под конкурентной разведкой 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.

Родоначальником современной конкурентной разведки считается компания Ксерокс (Xerox), столкнувшаяся с конкуренцией со стороны японских производителей [Прескотт, 2003]. В начале 70-х годов XX века, после выхода японцев на американский рынок, менеджеры Ксерокс заметили, что компания стала утрачивать позиции на рынке. Ситуацию исправили изменения, основанные на сборе актуальной информации о рынке и конкурентах. Ксерокс, благодаря своему японскому филиалу, создал систему работы (бенчмаркинг), а затем адаптировал и применил к бизнесу разведывательные технологии. При этом одним из основных условий организации этого процесса было неотступное соблюдение закона, так как репутация компании могла рухнуть гораздо раньше, чем можно было бы воспользоваться экономическими преимуществами промышленного шпионажа. Вскоре эти методы работы начали применяться и другими американскими компаниями. Затем конкурентная разведка стала применяться в Европе, а в дальнейшем и во всем мире.

Игнорирование возможностей конкурентной разведки на начальном этапе дорого обходилось даже для крупнейших компаний [Джиллад, 2010]. Так после создания фотоаппарата, который выдавал готовый снимок, компания Polaroid стала почивать на лаврах. Когда аналитический отдел компании представил отчет, в котором указал на перспективы развития фотоиндустрии и зарождении цифровой эры, руководство компании назвали эту информацию «футуристической чепухой». Прошло некоторое время и в октябре 2001 года компания Polaroid начала первую процедуру банкротства.

Аналогично в 70-х годах XX века «Большая тройка» американских производителей автомобилей не прореагировала на появление на рынке японских производителей автомобилей. Однако, сами

американцы выбрали небольшие, экономичные и надежные японские автомобили, и американские корпорации понесли значительные убытки.

Конкурентные разведчики из корпорации Samsung узнали из открытой прессы, что последний американский завод по производству гитар может закрыться из-за более дешевых корейских инструментов, и американское правительство готовится защитить своих производителей с помощью таможенных пошлин. Вовремя узнав это, представители Samsung успел ввезти в США большое количество гитар, а в результате введения ввозных пошлин, еще и поднять цены на этот музыкальный инструмент.

По устоявшемуся ошибочному мнению, вся полезная разведывательная информация добывается из секретных источников агентурным или оперативным путем – на самом деле это не так. Известное признание адмирала Захариаса – заместителя начальника разведки Военно-морских сил 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 источников. 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.

В то же время, анализ рассекреченного отчета ЦРУ за 1987 год «Enterprise-Level Computing in Soviet Economy» (SOV C87-10043) дает представление о том, какой колоссальный объем данных необходимо было обрабатывать аналитикам. Для составления отчета постоянно на протяжении года сканировалось 347 открытых источников; для создания сводки объемом в одну страницу ежедневно обрабатывался информационный массив объемом примерно 7 млн. слов.

Сегодняшнее развитие информационных технологий сделало компьютерную разведку доступной даже для относительно небольших компаний, сегодня она распространена на всех уровнях экономики.

В арсенале тех, кто сегодня полноценно занимается конкурентной разведкой, нет специальной аппаратуры, шпионской техники. Их основной инструмент – компьютер, подключенный к сети Интернет. Деятельность подразделений конкурентной разведки компаний все больше основывается на последних достижениях в



области искусственного интеллекта в сочетании с наработками в областях психологии, социологии, экономики.

Ощутимые преимущества, получаемые за счет использования конкурентной разведки, подтверждают результаты опроса, проведенного еще в 1999 г. среди 500 крупнейших компаний США. Почти 90 % компаний подтвердили, что создали у себя подразделения конкурентной разведки. При этом затраты корпораций на разведку составляют в среднем 1–1,5 % от оборота и вполне рентабельны [Ландэ, Прищепа, 2007].

В настоящее время создаются многочисленные профессиональные объединения (сообщества) специалистов в области конкурентной разведки. Наиболее известные из таких сообществ, занимающихся организацией конференций, тренингов, – это Strategic and Competitive Intelligence Professionals, SCIP (www.scip.org, рис. 1) в США и Competia (www.competia.com) в Канаде.



Рис. 1 – Фрагмент веб-сайта организации SCIP (www.scip.org)

В России известны Российское общество профессионалов конкурентной разведки, РОПКР (www.rscip.ru) и Сообщество Практиков Конкурентной разведки, СПКР (razvedka-open.ru), в Украине – Харьковская областная общественная организация «Общество аналитиков и профессионалов конкурентной разведки». В Украине ведется подготовка специалистов в области конкурентной разведки

в Харьковском национальном университете радиоэлектроники, где готовят магистров по специальности «Консолидированная информация».

В настоящее время конкурентная разведка не ограничивается изучением конкурентов, а проводит анализ всей среды, окружающей организацию или предприятие. Изучается политическая обстановка, особенности законодательства, кадровые перемещения, новые технологии, собственные клиенты и поставщики компании и т.п., подбираются эксперты по специальным вопросам.

## **1. Цели, задачи и проблемы интернет-разведки**

Основными целями интернет-разведки, как сегмента конкурентной разведки [Кочергов, 2009], являются:

1. Информационное обеспечение процесса выработки управленческих решений на стратегическом и тактическом уровнях;
2. Выявление, прогноз и предотвращение потенциальных угроз бизнесу;
3. Выявление (совместно со службой безопасности) попыток конкурентов получить доступ к закрытой информации компании.
4. Определение благоприятных возможностей для бизнеса;
5. Управление рисками, обеспечение эффективного реагирования компании на быстрые изменения окружающей среды, интернет-пространства;
6. Промышленная контрразведка, предупреждение разведывательной деятельности конкурентов в сетевой среде, аналитическая поддержка службы безопасности компании.

Для достижения перечисленных целей в рамках интернет-разведки должны быть решены следующие задачи:

- сбор и своевременное обеспечение руководства и бизнес-подразделений компаний надежной и всесторонней информацией из сетевых источников о «внешней» и «внутренней» среде предприятия;
- выявление факторов риска, угроз, которые могут затронуть экономические интересы бизнеса или помешать его нормальному функционированию;
- выявление новых возможностей и других факторов, влияющих на получение конкурентных преимуществ;
- усиление благоприятных и локализация неблагоприятных факторов конкурентной среды на деятельность бизнес-структуры;
- выработка прогнозов и рекомендаций по влиянию конкурентной среды на деятельность бизнес-структуры.

Конкурентная разведка становится современным направлением исследования поведения конкурентов на рынке, позволяющим создавать модели рынка, его участников, определения характеристик и оптимизации тактики и стратегии развития субъектов хозяйствования на определенных рынках. Для достижения таких целей требуется использование эффективных приемов работы с информацией и ее элементами. Информация в этом смысле становится как объектом исследования рынка, так и основой для создания его модели.

Выше сформулированы цели и задачи интернет-разведки, рассчитанные на легитимную деятельность соответствующих структур. Вся система конкурентной разведки должна позволять руководству, аналитическому, маркетинговому отделам компании не только оперативно реагировать на изменения ситуации на рынках, но и оценивать дальнейшие возможности своего развития. Конкурентная разведка обеспечивает переход от традиционного интуитивного принятия решений на основе недостаточной информации к управлению, основанному на знаниях. При этом она также обеспечивает снижение рисков, безопасность бизнеса, а также приобретение конкурентных преимуществ. Современная система конкурентной разведки позволяет не только осуществлять мониторинг информации, но и моделировать стратегию конкурентов, выявлять их партнеров, поставщиков, понимать условия сотрудничества.

Основные задачи систем конкурентной разведки заключаются в нахождении и обобщении информации о конкурентах, рынках, товарах, бизнес-тенденциях и операциях по таким основным объектам:

- партнеры, акционеры, смежники, союзники, контрагенты, клиенты, конкуренты (личности и компании);
- объединения компаний, слияния, поглощения, кризисные ситуации и т.п.;
- кадровый состав, как своей компании, так и партнеров, конкурентов и т.д., а также кадровые изменения, их динамика;
- торговый оборот, бюджет и его распределения по пунктам;
- заключенные договора, соглашения или договоренности.

Интерес при проведении конкурентной разведки вызывает не только непосредственная сфера деятельности компаний, но и сферы их влияния и интересов. Эти знания могут применяться, например, для оказания влияния на позиции партнеров и оппонентов в ходе деловых переговоров. Большое значение имеет информация, относящаяся к политике конкурентов, их намерениям, сильным и слабым сторонам, продукции и услугам, ценам, рекламным кампаниям, другим параметрам рынка.

Современные открытые сетевые ресурсы, веб-сайты, социальные сети превращаются в настоящее время в основной источник и эффективный инструмент для конкурентной разведки. Они позволяют в режиме реального времени не только отслеживать действия компаний-конкурентов, но и выявлять последние тенденции по интересующей тематике. Назовем лишь некоторые способы использования интернет-ресурсов для решения задач конкурентной разведки:

*1. Получение новостей по целевой тематике.*

Современные сетевые сервисы, такие как Google News, Яндекс.Новости, социальные сети типа Twitter, FaceBook позволяют получать новости, отобранные в соответствии с информационными потребностями пользователей. Например, при использовании социальной сети Twitter, можно воспользоваться поисковым режимом, и ввести запрос, например «банкротство». После этого пользователь получит список сообщений, в некоторых случаях снабженных аккаунтами пользователей, чьи сообщения релевантны введенному запросу. Таким образом, можно определить экспертов, которых можно сгруппировать в соответствии со своими информационными потребностями. Затем, следуя за группой экспертов, можно получить достаточно широкий охват проблемы, несколько точек зрения, новые информационные ресурсы.

*2. Выявление тенденций.*

По выбранным с помощью поисковых возможностей информационным ресурсам (веб-сайтам, блогам, форумам) можно вручную или с использованием специальных аналитических инструментов выявлять тенденции в выбранной сфере.

*3. Получение рассылки целевых документов по электронной почте.*

Многие из новостных агрегаторов и социальных сетей (в частности, Twitter) предоставляют возможность качественных персонализированных периодических рассылок, охватывающих сообщения, комментарии, блоги экспертов.

*4. Построение сетей информационных связей*

Для задач конкурентной разведки важно не только получение целевой информации (сообщений), но и понимание связей, которые обнаруживаются при анализе информации. Важен не только объект анализа, но и связанные с ним информационные ресурсы, профили в социальных сетях, «друзья», группы обсуждений и т.п. В некоторых случаях можно посмотреть, кто является подписчиком данных профилей, кто интересуется той же тематикой и, следовательно, может стать новым источником для получения целевой информации.

*5. Получение ответов на вопросы.*

Социальные сети можно использовать как способ получения ответов на конкретные вопросы, в том числе и по вопросам методологии конкурентной разведки. Если вопрос поставлен корректно, то с большой вероятностью можно получить ответ на него от других пользователей.

*6. Фильтрация мусора.*

Для конкурентной разведки не всегда интересны общеизвестные, зачастую ложные данные и информация, интересные большинству, а ведь именно на такие данные ориентированы социальные сети. При использовании сетевых ресурсов в качестве мощнейшей базы для конкурентной информации особое внимание следует уделять отработке запросов, выбору источников, экспертов, установлению связей.

Отметим некоторые проблемы, связанные с конкурентной разведкой в веб-пространстве.

Первой и наиболее существенной проблемой является то, что колоссальные объемы информации в Интернет затрудняют поиск и выбор действительно необходимых сведений. Сами по себе необработанные, необобщенные и непроверенные данные не могут обеспечить качественную поддержку при принятии решений.

По оценкам специалистов, охваченные в настоящее время поисковыми системами веб-ресурсы составляют свыше триллиона документов и постоянно растут. По словам Эрика Шмидта (Eric Emerson Schmidt) – председателя совета директоров Google, даже такая мощная поисковая система как Google сможет проиндексировать всю имеющуюся сегодня информацию лишь примерно через 300 лет.

Традиционные поисковые системы в Интернет отлично справляются с простыми однократными запросами, однако, как правило, слабо применимы для нужд конкурентной разведки. По некоторым оценкам [Ландэ, Прищепа, 2007], более 97 % критичной для конкурентной разведки онлайн-информации невозможно найти с помощью традиционных информационно-поисковых систем.

Второй проблемой конкурентной разведки в Интернет является то, что информация в Интернет имеет явно выраженный динамичный характер: информация размещается, модифицируется и удаляется. Частичное решение этих проблем возможно при применении систем контент-мониторинга информационных потоков в Интернет.

Третья проблема, которую необходимо решить в целях конкурентной разведки, – автоматическое извлечение понятий из формализованных массивов информации (таблиц, баз данных) и неструктурированных текстов. Перспективным направлением решения этой проблемы в системах конкурентной разведки является использование технологий Knowledge Discovery, Data Mining и Text Mining [Ландэ, 2005], [Печенкин, 2004].

Четвертой проблемой является выявление неочевидных закономерностей и связей. В настоящее время известно несколько пу-

тей решения проблем извлечения понятий из текстов и выявления их взаимосвязей, как практических, так и теоретических. Одним из этих путей является построение матриц и графов взаимосвязей, к которым можно применять соответствующие математические методы. Как правило, узлы этих графов – коэффициенты, которые пропорциональны количеству документов из входного информационного потока, которые одновременно соответствуют исследуемым понятиям.

Пятой проблемой является поиск информации в «скрытом» веб-пространстве, где содержится несравнимо большее количество данных, потенциально интересных для конкурентной разведки, чем в открытой части сети Интернет. 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% необходимой информации имеется в Интернете в готовом виде, остальные 85-90% можно получить в результате сравнения, интеграции и анализа многочисленных разрозненных данных.

Итак, в Интернет содержится большая часть информации, необходимой для проведения конкурентной разведки, однако остается открытым вопрос ее нахождения и эффективного использования. Причина – присущие сети Интернет недостатки [Ландэ, 2005]:

- непропорциональный рост уровня информационного шума;
- засилье паразитной информации;
- слабая структурированность и связность информации;
- динамичность информации;
- отсутствие целостности информации;
- многократное дублирование информации;
- отсутствие возможности смыслового поиска;
- ограниченность доступа к «скрытому» веб.

Несмотря на это возможности Интернета оцениваются экспертами в области конкурентной разведки достаточно высоко.

## **2. Системы и технологии конкурентной разведки**

---

---

Конкурентная разведка использует в своем арсенале различные средства, наиболее развитыми из которых являются специализированные информационно-аналитические системы (ИАС). Типовая ИАС конкурентной разведки (рис. 2) включает следующие компоненты:

- комплексы контент-мониторинга информации из открытых сетей (веб-пространства, социальных, пиринговых сетей и т.п.);
- средства экстрагирования понятий (компаний, персон, событий и т. п.) из полнотекстовых документов;
- средства выявления и визуализации информационных связей, выявления аномалий, неочевидных закономерностей;
- средства формирования аналитических документов, которые предоставляются лицам, принимающим решения (ЛПР).

Содержательная часть, информационная база информационно-аналитической системы конкурентной разведки формируется комплексом контент-мониторинга. Особенности современных комплексов контент-мониторинга заключаются в том, что они должны охватывать огромные объемы информации из динамически возрастающих информационных потоков в сетях при наличии шумовой информации, большой части слабодоступных ресурсов, так называемого «скрытого Интернета» [Ландэ, 2010]. При этом с помощью комплексов контент-мониторинга в рамках конкурентной разведки, как правило, решаются следующие задачи:

- мониторинг деятельности партнеров, конкурентов, регулирующих органов;
- контроль медиаприсутствия и медиаактивности участников рынков;
- нахождение информации об участниках рынков;
- выявление новых продуктов на рынках;
- выявление новых игроков на рынках;
- организация ретроспективного информационного фонда документов для их последующего использования в аналитической деятельности.

Процесс превращения сырых данных в знания и доведение их до конечных потребителей принято называть разведывательным циклом. В своем классическом понимании разведывательный цикл (разведцикл) принято разделять на пять основных этапов:



- целеуказание, планирование, определение источников информации;
- сбор, добывание данных;
- обработка разведывательных данных (разведданных) – превращение их в разведывательную информацию;
- анализ и синтез разведывательной информации – превращение ее в знания – выводы, рекомендации, решения;
- доведение информации до конечных потребителей.

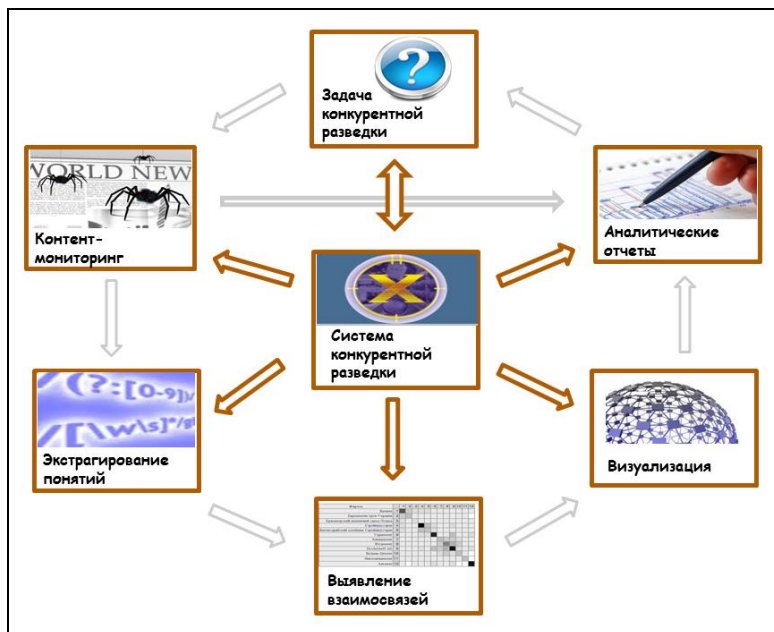


Рис. 2 – Типовые компоненты системы конкурентной разведки

Не вдаваясь в подробности, следует также отметить некоторые ключевые особенности указанных выше этапов, 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;

– разведчик, в некоторых случаях может не требовать глубокой проработки, например, в условиях ограниченного времени, может быть не полным и заканчиваться выдачей 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 (*любопытно заметить, что ЦРУ, например, предоставляло Р. Рейгану ежедневную информацию в виде видеofilmа, который снимали каждый день, поскольку бывший киноактер воспринимал такую подачу информации более адекватно*).

Итак, открытые источники являются наиболее используемым каналом информации, при их использовании возрастает объективность добываемой информации, однако, резко возрастают трудозатраты на извлечение нужной информации. Следовательно, для их использования в конкурентной разведке нужны специализированные методики и системы. Такие специализированные методики и системы создавались учеными в интересах спецслужб на протяжении многих лет, как на Западе, так и в бывшем Советском Союзе. Перевод в последние 10–20 лет значительного объема мировой информации из бумажного вида в электронный, широкое использование и рост сети Интернет, новые информационные технологии сделали интернет-разведку одним из самых перспективных направлений разведывательной деятельности. Тот факт, что так поступают практически все спецслужбы мира, лишь подтверждает перспективность этого направления для конкурентной разведки.

Для поиска и сбора информации в компьютерных сетях в интересах разведки по всему миру используются специальные мониторинговые системы сбора данных, т.н. «процессоры сбора данных». На компьютерном сленге их еще называют «роботами» или «пауками». Такая программа-робот сама обходит по заданному графику указанные URL-ссылки в сети Интернет, скачивает с них данные, а затем извлекает из них нужную информацию, используя целый арсенал средств лингвистического, семантического и статистического анализа. Такие системы автоматически перехватывают

любую поставленную на мониторинг информацию, как только она появится в доступном сегменте Сети.

При организации конкурентной разведки в сети Интернет широкое распространение получило использование направления науки, возникшего на стыке искусственного интеллекта, статистики и теории баз данных, как Knowledge Discovery (обнаружение знаний), использующего концепции Data Mining (добыча знаний в формализованных базах данных) и Text Mining (добыча знаний в неструктурированных текстах). Уникальными особенностями концепции и технологий Data и Text Mining, является то, что с их помощью можно добывать из «сырых» данных ранее неизвестные, неочевидные, полезные на практике и доступные для интерпретации знания, необходимые для принятия решений в различных сферах деятельности. Одним из первых рассекреченных подобных комплексов стала французская система «TAIGA» (Traitement Automatique de l'Information Geopolitique d'Actualite – автоматическая система обработки актуальной геополитической информации) [Доронин, 2003]. Этот программный комплекс на протяжении 11 лет использовался в интересах французской разведки, после чего был заменен более современным, рассекречен и разрешен к коммерческому использованию. Новый более совершенный комплекс Noemic, поставленный на вооружение французской разведки, способен обрабатывать информацию со скоростью более 1 миллиарда знаков в секунду. Американский аналог этих программных комплексов Topic, что в переводе значит «Тема», также уже рассекречен и передан для коммерческого использования.

Аналогичные системы создавались в бывшем Советском Союзе. Достаточно вспомнить такие системы как «Барометр», «Эльбрус» и другие. Продолжается создание и использование таких систем в России и других странах постсоветского пространства.

На первый взгляд может показаться, что все перечисленные примеры – это системы, которые либо используются государственными структурами, либо слишком дороги, чтобы их могли использовать «среднестатистические» компании. На самом деле все не совсем так. На современном рынке представлен целый ряд, как западных коммерческих продуктов, так и продуктов производства стран бывшего СССР, способных в том или ином объеме выполнять подобные задачи в интересах конкурентной разведки коммерческих структур.

## **2.1. Средства поиска информации в Интернете**

Для того, чтобы получить крупницы необходимой пользователю информации в Сети необходимо обработать огромные массивы сырых данных. Естественно, что для облегчения этой задачи нужны некие поисковые инструменты.

Поиск информации в сети Интернет без использования поисковых средств, только путем просмотра отдельных веб-сайтов, во-первых, носит выборочный и/или случайный характер (к тому же информация на отдельных сайтах может носить весьма субъективный или даже заказной характер), во-вторых, крайне не продуктивен – вряд ли можно обойти и просмотреть более сотни веб-сайтов за день непрерывной работы, даже если знать их адреса.

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

- средства поиска информации на отдельных сайтах;
- подборки ссылок;
- каталоги;
- поисковые системы;
- метапоисковые системы;
- системы мониторинга и контент-анализа;
- экстракторы объектов, событий и фактов;
- системы Knowledge Discovery, Data Mining, Text Mining;
- специализированные системы конкурентной разведки;
- интегрированные системы.

По своей сути каталоги, поисковые системы и метапоисковые системы являются веб-сайтами со специализированными базами данных, в которых хранится информация о других веб-ресурсах и документах, хранящихся на них. В результате запроса к таким системам выдается список адресов, а иногда и краткое описание документов (сниппеты), где может присутствовать запрашиваемая информация. Как правило, поиск может производиться только по ключевым словам и фразам. Активизируя на ссылку, найденную в результате запроса, пользователь попадает на оригинал документа, размещенного на некотором веб-сайте. Естественно, что если документ со временем изменился или веб-сайт прекратил свое существование, то и первоначально заиндексированный поисковой системой документ через некоторое время может быть не найден.

Основное отличие поисковых систем от каталогов – наличие автоматического «робота», постоянно сканирующего веб-пространство и накапливающего новую информацию в индексных файлах базы данных. В каталоги же информация заносится вручную –

либо владельцами сайтов, либо обслуживающим персоналом самих каталогов. Пользование такими системами, как правило, бесплатное, это самые популярные поисковые инструменты в сети Интернет.

Метапоисковые системы являются поисковыми системами по поисковым системам. Так как отдельные поисковые системы различным образом 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 целей и количества найденных документов. Если в результате поиска найдены тысячи или миллионы документов, то «вручную» все равно вряд ли можно просмотреть свыше нескольких сотен.

Вторым поисковым преимуществом таких систем является то, что одним запросом обеспечивается поиск во многих поисковых системах, не требуя многочисленных повторений одного и того же запроса в разных поисковых системах.

Системы мониторинга и контент-анализа обеспечивают регулярный поиск и «скачивание» информации по заданным темам и с заданных сайтов, а также анализ содержания «скачанных» документов. Такие системы, как правило, обладают развитым языком запросов, что позволяет существенно детализировать и конкретизировать запросы по сравнению с обычными поисковыми системами. Во-вторых, 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 документов, проводить анализ динамики появления во времени тех или иных документов, проводить сравнительный анализ информационной активности по различным тематикам и многое другое.

Если мониторинговые системы как системы фильтрации могут выделять из информационного потока известные объекты, то экстракторы объектов, событий и фактов умеют выделять из потока информации объекты, неизвестные заранее, события или факты, которые лишь соответствуют определенному заранее типу, например, географические понятия, персоны, структуры и организации, события (дорожно-транспортные происшествия, катастрофы, международные встречи). При этом факты могут классифицироваться как обычные или необычные. Примером обычного факта в данном случае можно считать выезд автомобилей за черту города, а примером необычного факта – выезд за ту же городскую черту автомобиля без номерных знаков.

Системы типа Knowledge Discovery, технологии Data Mining и Text Mining, умеют выявлять новые знания и закономерности. Такая система, например, может 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 и одного итого же класса в одном и том же населенном пункте. Правда, сами правила, по которым такая система делает выводы, все-таки создаются и задаются пока что людьми.

Специализированные системы для конкурентной разведки могут включать в себя одно или несколько из перечисленных выше поисковых средств, специально «заточенных» под эти специфические задачи. Кроме того, потребности конкурентной разведки предполагают использование в качестве источников информации, кроме полнотекстовых документов, еще и доступных в сети Интернет баз данных (БД), собственных, принадлежащих компании, документов, таблиц и баз данных, а также формализованных и неформализованных документов и БД, добытых из других источников.

В странах Европейского союза (ЕС) обычный, ничем не выдающийся из общей массы человек зарегистрирован в более 300 базах данных, таких как прописка (место жительства), страховка, водительские права, банки, кредитные бюро, информационные, рейтинговые, рекрутинговые агентства, бюро по трудоустройству, медицинские и полицейские учеты, супермаркеты, клубы, системы управления взаимоотношений с клиентами коммерческих фирм (так называемые CRM-системы) и т.п. В интересах конкурентной разведки и маркетинга анализируются не только рынки товаров и услуг, но и вкусы и предпочтения отдельных клиентов. Хранящаяся в различных базах данных информация 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 работы, мы не будем касаться других источников информации кроме Интернета.

Под интегрированными средствами конкурентной разведки мы здесь понимаем интегрированную систему, включающую в себя не только все доступные поисковые средства, но также интегрированный банк выявленных (добытых) и логично правильно связанных между собой данных, 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.

Перефразируя девиз одной из самых популярных российских поисковых систем Яндекс «Найдется все!», можно сказать, что для конкурентной разведки «все» равнозначно ничего, если не хуже. Поэтому правильным девизом для систем конкурентной разведки могли бы стать слова «Найдется только то, что нужно и ничего более!».

По словам российского эксперта в области конкурентной разведки А. Масаловича «... из 23 видов поисковых задач, интересующих аналитика спецслужб «Яндекс» удовлетворительно решает одну».

В этом легко убедиться, задав реальный, требуемый в рамках информационно-аналитической деятельности запрос любой из традиционных сетевых поисковых систем, будь-то Google, Яндекс, Bing. Поисковые системы отлично справляются с простыми однократными запросами. Когда же предметная область сложна или слишком широка (например – «политика», «экономика»), или, наоборот, предельно узка и отдалена во времени (например, усло-

вия бартерной сделки некоторых компаний пятилетней давности), а требуется обобщить все информационные темы и поводы по данной тематике, оценить их во временной динамике, найти взаимосвязи с другими объектами, составить целостную картину об интересующем объекте, выделить нестандартное событие из общего массива, то можно убедиться, что:

- выдача поисковых систем либо перегружена тысячами бесполезных ссылок, либо наоборот недостаточна;

- информация в сети Интернет не хранится долго, необходимую информацию, присутствующую на целевом сайте месяц назад, сегодня можно там не обнаружить;

- поисковая система не сохраняет просмотренные аналитиком ссылки и ему каждый раз приходится начинать рутинную работу с нуля после вынужденного перерыва;

- поисковая система не всегда отличает действительно важную информацию от информационного шума;

- поисковая система не всегда способна обобщать или сравнивать информацию по смыслу или другим содержательным критериям;

- поисковые системы не охватывают некоторые веб-ресурсы или отдельные виды информации (например, информацию из баз данных), а некоторые веб-ресурсы, наоборот, всегда показываются на первых страницах выдачи, хотя их содержание не интересно авторам запросов;

- поисковые системы могут выполнять поиск информации только по непосредственно введенному запросу и не всегда могут повторять их автоматически в заданное время без участия пользователя;

Не всегда у пользователя есть возможность объединять результаты поиска разных сетевых поисковых систем – в этом случае страдает полнота полученной информации.

По оценкам экспертов [Кузнецов, 2006] большую часть критичной для бизнеса информации из сети Интернет невозможно найти с помощью традиционных информационно-поисковых систем. Точнее традиционные сетевые информационно-поисковые системы не в полной мере справляются с задачами конкурентной разведки. Поэтому разрабатываются специализированные информационно-поисковые системы, ориентированные на задачи сетевой аналитики, конкурентной разведки. Список таких общедоступных систем приведен по адресу ([http://hrazvedka.ru/category/poisk\\_soft](http://hrazvedka.ru/category/poisk_soft)). Приведем описание некоторых из них:



**Website-Finder** ([www.softpedia.com/get/Internet/Search-engine-tools-submitting/Website-Finder.shtml](http://www.softpedia.com/get/Internet/Search-engine-tools-submitting/Website-Finder.shtml)) – программа, которая дает возможность поиска веб-сайтов, плохо индексируемых поисковой системой Google. Для каждого ключевого слова ищется 30 веб-сайтов. Программа проста в использовании, есть бесплатная версия.

**Global Supplier Directory by Solusource** ([www.worldindustrialreporter.com/solusource](http://www.worldindustrialreporter.com/solusource)) – веб-интерфейс для конкурентной разведки от компании Thomas. Позволяет найти информацию, имеющуюся в ретроспективных базах данных Thomas (охват – более 100 лет) по компаниям, продуктам и отраслям.

**dtSearch** ([www.dtsearch.com](http://www.dtsearch.com)) – поисковая программа, позволяющая обрабатывать терабайты текста, как на локальном диске, так и в сетевом окружении. Поддерживает статические и динамические данные. Позволяет искать во всех форматах MS Office.

**Strategator** ([www.strategator.com](http://www.strategator.com)) – программа, осуществляющая поиск, фильтрацию и агрегацию информации о компаниях из десятка тысяч веб-источников. Ищет по данным из США, Великобритании, стран Европейского союза.

**InfoNgen** ([www.infongen.com](http://www.infongen.com)) – агрегатор, охватывающий в режиме просмотра свыше 35 тысяч онлайн-источников, легко настраиваемый на уникальные темы. Объединяет мониторинг, фильтрацию и агрегацию информации по запросам конкретного пользователя. Предоставляет информацию на восьми языках, обеспечивает перевод на английский язык.

**Visual.ly** ([visual.ly](http://visual.ly)) – система поиска инфографики в веб-пространстве.

**Quixey** ([www.quixey.com](http://www.quixey.com)) – англоязычная функциональная поисковая система, обеспечивающая вывод ссылок на программы, сервисы и расширения для браузеров, соответствующих запрашиваемой задаче.

**Поиск mail.ru** ([go.mail.ru/realtime](http://go.mail.ru/realtime)) – поиск в mail.ru по обсуждениям тем, событий, объектов, субъектов в режиме реального времени.

**Weblib** ([weblib.in.ua](http://weblib.in.ua)) – экспериментальная, эффективно работающая метапоисковая система по документам, представленным в формате PDF.

**Zanran** ([www.zanran.com/q](http://www.zanran.com/q)) – поисковая система для данных, извлекаемых из PDF-файлов, таблиц Excel, веб-страниц в формате HTML.

**CIRadar** ([www.ciradar.com/Competitive-Analysis.aspx](http://www.ciradar.com/Competitive-Analysis.aspx)) – коммерческая англоязычная система поиска информации для конкурентной разведки в «глубинном» веб. Реализована как веб-сервис.

**Мультитендер** (multitender.ru/tenders) – специализированная поисковая система по госзакупкам, тендерам и аукционам в Российской Федерации (РФ).

**Public.Ru** (public.ru) – интернет-библиотека СМИ, профессиональный анализ информации, архив СМИ с 1990 года (70 млн. статей из 4600 источников).

**РосГосЗаплаты** (www.rosspending.ru) – поисковая система, обеспечивающая мониторинг российских государственных контрактов на федеральном и региональном уровнях, имеет возможность поиска по ключевым словам, организациям, продуктам в базах данных госзаказчиков, поставщиков и контрактов.

**Cluuz** (cluuz.com) – поисковая система с расширенными возможностями для применения в конкурентной разведке, в том числе и на русском языке. Позволяет не только находить, но и визуализировать, устанавливать связи между людьми, компаниями, доменами, e-mail, адресами и т.п. (рис. 3).

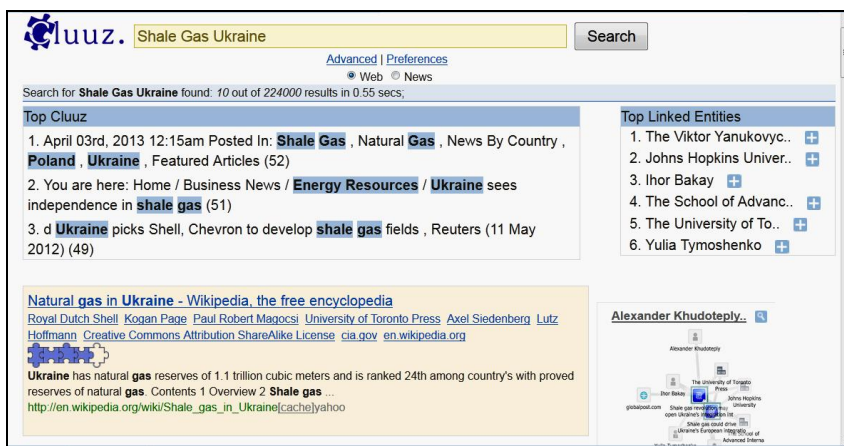


Рис. 3 – Фрагмент страницы поиска системы Cluuz (cluuz.com)

## 2.2. Концепция глубинного анализа текстов

Задача, которую необходимо постоянно решать при проведении конкурентной разведки – автоматическое извлечение понятий и фактов из формализованных массивов информации (таблиц, БД) и неструктурированных текстов, представленных в веб-пространстве, выявление глубинных связей между отдельными понятиями. Для этого предполагается использование в системах конкурентной

разведки технологий Knowledge discovery, концепции глубинного анализа данных и текстов (Data Mining, Text Mining).

Важная задача технологии Text Mining связана с извлечением из текста его характерных элементов или свойств, которые могут использоваться в качестве метаданных документа, ключевых слов, аннотаций. Другая важная задача заключается в отнесении документа к некоторым категориям из заданной заранее схемы классификации. Text Mining также обеспечивает новый уровень семантического поиска документов.

Согласно сложившейся в настоящее время методологии, к основным элементам Text Mining относятся [Ландэ и др., 2009]: классификация (Classification), кластеризация (Clustering), построение семантических сетей, извлечение фактов, понятий (Feature Extraction), реферирование (Summarization), ответы на запросы (Question Answering), тематическое индексирование (Thematic Indexing) и поиск по ключевым словам (Keyword Searching). Также в некоторых случаях этот набор дополняется средствами поддержки и создания таксономии (Taxonomies), тезаурусов (Thesauri) и онтологий (Ontology).

При классификации текстов используются статистические корреляции для создания правил размещения документов в определенные категории. Задача классификации – это классическая задача распознавания, где по некоторой контрольной выборке система относит новый объект к той или иной категории. Особенность же концепции Text Mining заключается в том, что количество объектов и их атрибутов могут быть очень большими – предусматривается применение интеллектуальных механизмов оптимизации процесса классификации.

Кластеризация базируется на признаках документов, применении лингвистических и математических методов без использования заданных заранее категорий. Результатом кластеризации может быть таксономия или визуальная карта, которая обеспечивает эффективный охват больших объемов данных. Кластеризация в Text Mining рассматривается как процесс выделения компактных подгрупп объектов с близкими свойствами. Средства кластеризации позволяют находить признаки и разделять объекты по подгруппам на базе этих признаков. Кластеризация, как правило, предшествует классификации, поскольку позволяет определить группы объектов.

При построении семантических сетей предполагается анализ связей между понятиями, экстрагируемыми из документов. Понятиям соответствует появление определенных дескрипторов (ключе-

вых фраз) в документах. Связи между понятиями могут устанавливаться в простейшем случае путем учета статистики их совместного упоминания в различных документах.

Извлечение или экстрагирование фактов (понятий) предназначено для получения некоторых фактов из текста с целью улучшения классификации, поиска, кластеризации и построения семантических сетей.

Автоматическое реферирование (Automatic Text Summarization) [Хан, 2000] – это составление кратких изложений материалов, аннотаций или дайджестов, т.е. извлечение наиболее важных сведений из одного или нескольких документов и генерация на их основе лаконичных, понятных и информационно наполненных отчетов.

На основе методов автоматического реферирования возможно формирование поисковых образов документов. По автоматически построенным аннотациям больших текстов – поисковым образам документов – может проводиться поиск, характеризующийся высокой точностью (естественно, за счет полноты). В некоторых случаях вместо поиска в полных текстах массива больших по размеру документов оказывается целесообразным поиск в массиве специально созданных аннотаций. Хотя поисковые образы документов часто оказываются образованиями, лишь отдаленно напоминающими исходный текст, не всегда воспринимаемый человеком, но за счет вхождения наиболее весомых ключевых слов и фраз, они помогают приводить к вполне адекватным результатам при проведении полнотекстового поиска.

Уникальными особенностями концепции и технологий Text Mining, является то, что с их помощью можно извлекать из «сырых» данных неочевидные, полезные на практике и доступные для интерпретации знания, необходимые для принятия решений в различных сферах деятельности, в том числе в области экономической конкуренции.

На современном рынке представлен целый ряд, как западных продуктов, так и систем производства постсоветских стран, способных в той или иной объеме осуществлять глубокий анализ текстов.

В последнее время все основные западные бренды, специализирующиеся на разработке информационных хранилищ и баз данных, корпоративных систем управления расширили свои линейки продуктов системами или модулями Text Mining. О наличии таких модулей заявляют SAP, Oracle, SAS, IBM и другие компании.

Процесс конкурентной разведки можно рассматривать как построение сети из исследуемых объектов и связей между ними. Ре-

зультаты должны представлять собой аналитическую информацию, которая может быть использована для принятия решений. Аналитическая информация может быть представлена в виде наглядных схем – семантических сетей, дайджестов, наборов сюжетных линий, взаимосвязей ключевых понятий, компаний, лиц, технологий и т.п.

Задачи конкурентной разведки породили спрос на специальные информационные технологии, обеспечивающие возможность извлечения и обработки необходимой информации, что в свою очередь вызвало поток предложений систем со стороны разработчиков программного обеспечения.

Сегодня решать задачи конкурентной разведки на основе информации из сети Интернет помогают общедоступные и специальные программы и сервисы, например, в последнее время приобрели популярность так называемые «персонализированные разведпорталы», способные отбирать информацию по самым узким, специфическим вопросам и темам и предоставлять ее заказчикам.

В настоящее время декларированы технологии и системы «компьютерной конкурентной разведки», идея которых заключается в автоматизации и ускорении процессов извлечения необходимой для конкурентной борьбы информации из открытых источников и ее аналитической обработки.

При ведении конкурентной разведки находят все более широкое применение новые направления науки и технологий, получившие названия: «управления знаниями» (Knowledge Management) и «обнаружение знаний в базах данных» (Knowledge Discovery in Databases) или иначе, Data и Text Mining – «глубинный анализ данных или текстов».

Если системы управления знаниями реализуют идею сбора и накопления всей доступной информации, как из внутренних, так и из внешних источников, то Data и Text Mining, как уже было показано, позволяют выявлять неочевидные закономерности в данных или текстах – так называемые латентные (скрытые) знания. В целом эти технологии еще определяют как процесс обнаружения в «сырых» данных ранее неизвестных, но полезных знаний, необходимых для принятия решений. Системы этого класса позволяют осуществлять анализ больших массивов документов и формировать предметные указатели понятий и тем, освещенных в этих документах.

Характерная задача конкурентной разведки, обычно включаемая в системы Text Mining – это нахождение исключений, то есть поиск объектов, которые своими характеристиками сильно выделяются из общей массы.

Еще один класс важных задач, решаемых в рамках технологии Text Mining – это моделирование данных, ситуационный и сценарный анализ, а также прогноз [Ланде, Фурашев, 2012].

Для обработки и интерпретации результатов Text Mining большое значение имеет визуализация. Часто руководитель компании не всегда адекватно воспринимает предлагаемую ему аналитическую информацию, особенно если она не вполне совпадает с его пониманием ситуации. В связи с этим служба конкурентной разведки должна стремиться представлять информацию в виде, адаптированном к индивидуальному восприятию заказчика.

Визуализация обычно используется как средство представления контента всего массива документов, а также для реализации навигации по семантическим сетям при исследовании, как отдельных документов, так и их классов.

### 2.3. Мониторинг информационного пространства

Современные методы контент-мониторинга – это адаптация концепции глубинного анализа текстов (Text Mining) и классических методов контент-анализа к условиям формирования и развития динамических информационных массивов, например, потоков информации в сети Интернет. Типовая задача контент-мониторинга – построение диаграмм динамики появления понятий (отражения событий) во времени. Рассмотрим, как в системе контент-мониторинга InfoStream [Григорьев, 2007] отслеживаются публикации, относящиеся к российско-украинскому газовому кризису 2008–2009 годов. Для этого был составлен запрос **«газов~криз & geo.UA»**, введенный через веб-интерфейс системы (рис. 4).

На этой диаграмме видно, что пик кризиса пришелся на середину января 2009 года и был связан с подписанием соответствующего договора в Кремле и реакцией на это Секретариата Президента Украины (рис. 5).

На примере рынка нефтепродуктов рассмотрим, как из массивов текстовой информации из сети Интернет могут быть выявлены документы, содержащие максимальное количество ценовой информации по данному рынку.

Для получения списка основных сюжетов, относящихся к рынку нефтепродуктов, был введен запрос **«(нефтепродукт|бензин) & цены»**, который уточнялся специальными признаками **«numb.medium|numb.large»**, означающие в системе InfoStream средний или высокий уровень присутствия в документах цифровой информации (рис. 6). После этого достаточно перейти в режим

просмотра и проанализировать документы, ссылки на которые выданы системой (рис. 7).

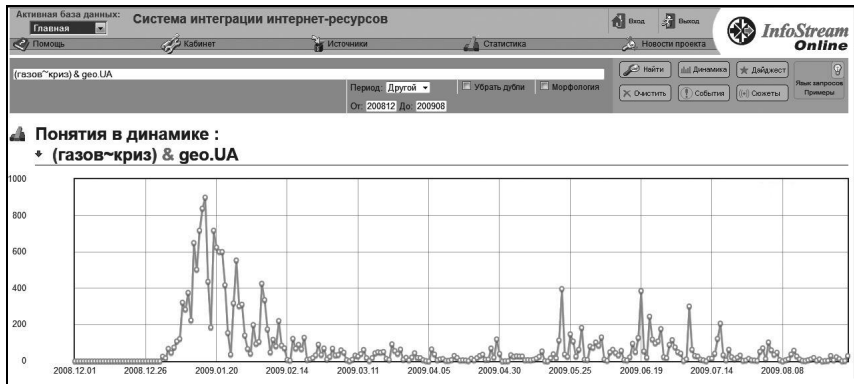


Рис. 4 – Диаграмма динамики понятия во времени

Обзор основных сюжетов  
 ( (газов-кризис) & geo.UA ) & (2009.01.16) ;  
 документов - 903, сюжетов - 112

**Секретариат Президента: газовый кризис начался с приходом Тимошенко в Кабмин**  
 МедиаПорт 2009.01.16 17:35  
 Заместитель главы Секретариата Президента Роман Бессмертный заявляет, что начало сегодняшнего газового кризиса следует искать в 2005 году с приходом Юлии Тимошенко на премьерскую должность. Бессмертный отметил, что на момент прихода Юлии Тимошенко на премьерскую должность в 2005 году между Украиной и Россией была полностью сформирована договорная правовая база и необходимо было лишь ежегодно подписывать Дубли - Похожие документы - Оригинал

**Всего в сюжете сообщений:** 90  
**Первое сообщение:** Хайбэй, 2009.01.16 01:33  
**Ключевые слова:** ГАЗ ПРЕЗИДЕНТ УКРАИНЫ РОССИЙСКИЙ ТИМОШЕНК ЕВРОП МЕДВЕДЕВ ПРЕМЬЕР ГАЗОВ ТРАНЗИТ КРИЗИС РЕШЕН УКРАИНСК СЕКРЕТАРИАТ БЕЗСМЕРТН ЕВРОПЕЙСК МОСКВ МЕЖДУНАРОДН КОНФЛИКТ КИЕВ

2009.01.19 07:30 **Российские национал - патриоты готовятся "стирнуть" у коммунистов последний козырь - результаты Всесоюзного референдума 17 марта 1991 года *Славянская Европа***

2009.01.18 00:23 **Собое мнение: "украинский вопрос" придется решать и без Путина с Медведевым**  
*"Forum.msk.ru"*

2009.01.16 23:53 **Тимошенко взяла на себя ответственность за газовые переговоры *Политика de facto***

2009.01.16 22:20 **Мини-саммит в Киеве *EuroNews***

2009.01.16 21:36 **Тимошенко взяла на себя ответственность за газовые переговоры *"Lenta.Ru"***

2009.01.16 20:35 **Секретариат Ющенко просит прокуратуру проверить Тимошенко *"Космопольская Правда" в Украине***

2009.01.16 20:31 **У Ющенко снова вспомнили о любви между Тимошенко и Путиным *Настоящий Дозор***

2009.01.16 20:30 **Европейский бизнес готов разделить с Россией риск по транзиту топлива через Украину *Первый канал***

2009.01.16 20:23 **Тимошенко берет на себя преодоление газового кризиса. Заявление Премьера *Главное***

2009.01.16 20:10 **Банковская вновь призывает ГПУ "проверить" Тимошенко *Цензор.Нет***

2009.01.16 20:06 **Бессмертный требует судить Тимошенко как врага нации *Обозреватель***

2009.01.16 20:02 **Секретариат просит ГПУ начать шить дело Тимошенко *From-UA.com***

Рис. 5 – Основная сюжетная цепочка по запросу

(v) Обзор основных сюжетов  
 (нефтепродукт | Бензин) & цены) & (Большая цифровая насыщенность);  
 документов - 38, сюжетов - 8

2013.06.24 04:49 Бензин в Приморье стал дешевле Delta.RU  
 14  
 2013.06.27 15:30 Цены на крупнооптовом рынке нефтепродуктов Украины 27 июня понизились: Passarge.com.ua

2013.06.25 10:29 Цены на автомобильное топливо в Крыму на 25 июня 2013 г Соблюдя Крыма  
 8  
 2013.06.27 18:42 Сжиженный газ за неделю подешевел на 1,5% Терминал

2013.06.25 13:25 Цены на бензин и дизтопливо в Киеве 25 июня не изменились РБК-Украина  
 5  
 2013.06.27 13:55 Цены на бензин и дизтопливо в Киеве 27 июня незначительно изменились РБК-Украина

1. **Средние розничные цены на топливо в Украине на утро 25 июня**  
 По итогам 21 июня зафиксированы следующие розничные цены на топливо. Топливо Цена, грн. Мин/Мак Бензин А-76/80 10.23 9.80/10.50 Бензин А-92 10.54 10.20/10.79 Бензин А-95 10.91 10.50/11.29 Бензин А-95+ 11.09 10.85/12.59 Дизельное топливо 9.97 9.29/10.29 По данным консалтинговой группы "А-95".  
 Сюжет полностью (14)

2. **Розничная цена сжиженного газа стала меньше**  
 По результатам мониторинга рынка нефтепродуктов в Украине 25 июня 2013 г. отмечается снижение розничных цен на сжиженный нефтяной газ СПБТ, используемый в качестве моторного топлива, за 25.06.2013 г. на АГЭС в Украине подешевел на 0,45% (2,4 коп./л) до 5,34 гривн.  
 Сюжет полностью (8)

3. **Цены на бензин и дизтопливо в Киеве 26 июня незначительно изменились**  
 РБК-Украина Розничные цены на бензин и дизельное топливо 26 июня 2013 г. по сравнению с предыдущим торговым днем незначительно изменились. Об этом свидетельствуют данные мониторинга ценового департамента "Консалтинговой группы А-95". Компания А-80 А-92 А-95 А-95+ ДТ LPG КЛЮ 10,33 10,60 10,83 11,23 10,08 5,09 (+0,10) WOG 10,49 10,79 11,29 12,38 10,29 5,29 (+0,20) Лукойл 10,30 10,69 11,14 12,09 10,19 5,00 ТНК н/л 10,58 10,99 11,99 10,09 Укрнафта 10,10 10,40 10,70  
 Сюжет полностью (5)

Рис. 6 – Фрагмент цепочки основных сюжетов

**Цены на топливо на 27.06.2013**

По данным консалтинговой группы "А-95", средние на бензин и дизельное топливо на АЗС в Днепропетровске, грн./л на 27 июня 2013 года:

Компания	А-80	А-92	А-95	ДТ (Н-02,62)
Укрнафта	10,10	10,40	10,70	9,80
Веста Сервис	10,10	10,34	10,59	9,59
Формула Ритейл		10,59	10,94	9,99
Лукойл		10,69	11,14	10,19
Нефтек	10,29	10,59	10,99	9,99
Альфа-Нафта		10,35	10,65	9,80
<b>Средняя по области</b>	<b>10,16</b>	<b>10,49</b>	<b>10,84</b>	<b>9,89</b>

Средние цены на топливо по областям Украины:

Область	А-80	А-92	А-95	ДТ (Н-02,62)
АР Крым	10,24	10,58	10,98	9,99
Винницкая	10,25	10,61	11,00	10,02
Волинская	10,23	10,54	10,90	9,95
Днепропетровская	10,24	10,45	10,77	9,85
Донецкая	10,25	10,70	11,16	10,14
Житомирская	10,29	10,49	10,85	9,95
Закарпатская	10,23	10,54	10,91	9,86
Запорожская	10,20	10,47	10,82	9,89
Ивано-Франковская	10,23	10,61	11,01	10,05
Киев	10,23	10,60	10,98	10,04
Кировоградская	10,24	10,53	10,90	9,96
Луганская	10,25	10,62	10,94	10,00
Львовская	10,23	10,60	11,00	10,05

Рис. 7 – Документ с ценовой информацией

## 2.4. Определение взаимосвязей

Важной задачей конкурентной разведки является выявление неочевидных закономерностей и связей из текстов веб-страниц и



выявление их взаимосвязей, построение матриц и графов взаимосвязей.

Существующие доступные фактографические базы данных структурированной информации не всегда могут прийти на помощь исследователю-аналитику. Для оперативного определения фактов и сущностей, моделирования информационных связей между ними наиболее перспективным подходом оказывается учет информации, знаний, которые содержатся в неструктурированных текстовых документах, в частности, в Интернет.

Сегодня, когда практически у всех заинтересованных пользователей уже накоплен большой опыт работы с традиционными информационно-поисковыми системами, оказалось очевидным, что факты или понятия, которые ищутся с помощью таких систем, сами по себе зачастую бессмысленны. Например, если пользователь интересуется информационные связи Сбербанка России с другими банками или частными лицами, то он не знает, какие банки или фамилии ему указать в запросе, а все документы, содержащие словосочетание «Сбербанк России», указать физически невозможно. В таких случаях информационные связи, количество которых выходит за рамки статистического фона, как правило, отражают реальность.

Интерпретируют обычно не сами понятия или факты, а взаимосвязи между ними. Важным оказывается не столько исследование самих понятий, сколько исследование их взаимосвязи. Известно, что именно взаимосвязь способствует пониманию мотивационно-целевых особенностей, то есть пользователя интересует не понятие само по себе, а понятие в окружении, чтобы сразу иметь представление о предметной области, при необходимости направить уточняющий поиск в нужном направлении. Элементы такого подхода можно видеть, например, в «облаках» системы Quintura ([www.quintura.ru](http://www.quintura.ru)), но там отображаются не понятия/сущности, а наиболее часто используемые слова. Подобные решения, реализованные в виде «информационных портретов», содержащих опорные слова, используются в таких системах, как «RCO Zoom» ([www.rco.ru](http://www.rco.ru)), на веб-сайте интегратора новостей Webground ([webground.su](http://webground.su)).

База данных практически любой традиционной информационно-поисковой системы может рассматриваться в виде графа, вершинами которого выступают объекты – термины, понятия, дескрипторы и др., а ребрами – их связи. Вместе с тем, основа поиска в этих случаях – поиск вершин, то есть поиск объектов. Поиск по взаимосвязям, ребрам, кажется на первый взгляд менее эффектив-

ным. Действительно, если предположить, что в графе  $N$  вершин, то число ребер теоретически может составлять  $N(N - 1)/2$ , то есть, если предположить, что вершин всего 100 тыс., то ребер может оказаться около 5 млрд., что соответствует достаточно большой базе данных даже по современным понятиям. Вместе с тем, если в качестве вершин графа использовать такие понятия, как имена людей и названия компаний из новостных документов, то оказывается, что соответствующая матрица инцидентности оказывается очень разреженной. Измерения показали, что при количестве отдельных понятий, извлеченных из 5 млн. новостных документов, равном примерно  $N = 1,5$  млн., количество связей составило всего лишь  $v = 4$  млн.

Кроме того, как показали эксперименты, распределение степеней вершин (степень вершины – количество исходящих из нее ребер) в подобных графах – степенное, что свидетельствует о, так называемой, безмасштабности, то есть о том, что многие характеристики (в частности, соотношение количества вершин и ребер), должно оставаться на одном уровне. Поэтому в качестве основы построения базы данных связей оказывается технически возможным использование ребер рассматриваемого графа – связей между отдельными понятиями.

В качестве массивов документальной информации для такой системы могут использоваться данные, поступающие от систем контент-мониторинга, таких как InfoStream, Webscan или «Яндекс.Новости» а также результаты мониторинга специализированных веб-служб, таких как базы данных биографий людей (например, peoples.ru, file.liga.net/person, openua.net), организаций (например, www.yellowpages.kiev.ua, yrag.ru, baza.kompass.ua), служб трудоустройства и т.п.

Информационные взаимосвязи между понятиями выявляются путем обработки документальных массивов и, могут храниться в специальной базе данных. Набор понятий, используемый при построении базы данных связей, формируется путем экстрагирования данных из доступного пользователю текстового массива, что придает системе целостность.

В корпоративной информационной инфраструктуре база данных связей может использоваться различным образом, например, отдельно, либо ее возможности могут быть дополнены возможностями существующих полнотекстовых и/или фактографических баз данных (рис. 8). При этом основным результатом работы является построение так называемых «карт связей», а в качестве побочного

эффекта, реализующего «режим доказательства», может рассматриваться извлечение самих документов как источников связей.

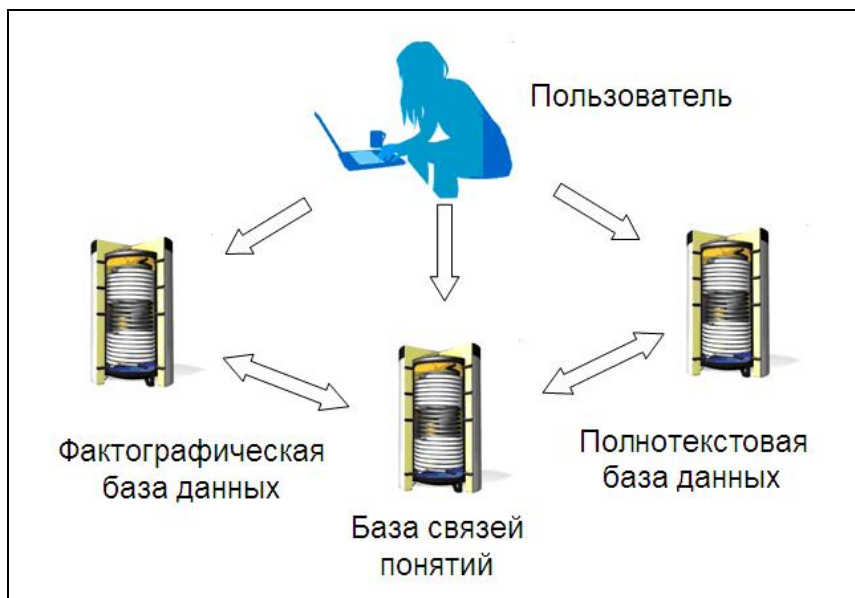


Рис. 8 – Место базы данных связей понятий в информационной инфраструктуре

При проектировании баз данных связей используются перспективные решения в области создания информационно-аналитических систем, в частности, теория и технологии глубинного анализа тестов – Text Mining, в том числе методы экстрагирования информации (Information Extraction), технологии баз данных сверхбольших объемов (Big Data), концепция «сложных сетей» (Complex Networks).

В рамках теории сложных сетей изучаются характеристики, связанные с топологией сетей, но и статистические феномены, распределение весов отдельных вершин (в качестве которых можно рассматривать сущности, понятия, факты) и ребер, эффекты протекания и проводимости в сетях и т.п.

На рис. 9 схематически представлены возможные технологические этапы формирования базы данных связей [Ландэ, Брайчевский, 2010].

С помощью программы-робота осуществляется сканирование выбранных веб-ресурсов, содержащих информацию, относящуюся к объектам исследований.

После этого осуществляется экстрагирование необходимых пользователям понятий, например, наименований брендов, компаний, электронных адресов и т.п.

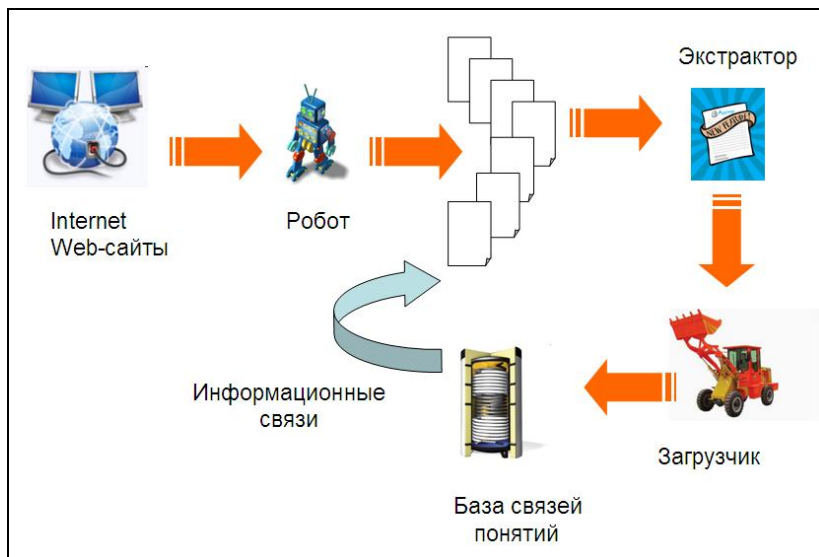


Рис. 9 – Схема формирования базы данных связей

Отобранные понятия и соответствующие отношения между ними загружаются в базу данных связей, которая также содержит ссылки на документы-первоисточники. Средства экстрагирования понятий, как правило, ориентированы на обработку документов, сканируемых из сети Интернет, представленных на различных языках.

Предложенный подход к поиску, естественно, влечет за собой некоторые особенности в реализации архитектуры базы данных связей понятий. Кроме того, архитектура базы данных связей должна быть ориентирована на такие возможные применения, как выявление неявных связей (не выявленных явно комплексом экстрагирования понятий), поиск отдельных объектов, а также взаимосвязь с существующими фактографическими базами данных.

Можно назвать несколько систем, в которых частично реализован данный подход:

– **PolyAnalyst** ([www.megarputer.ru](http://www.megarputer.ru)) – позволяет решать проблемы прогнозирования, классификации, группирования объектов, проводить анализ связей, многомерный анализ и интерактивное создание отчетов. Система PolyAnalyst (и ее компонента – система TextAnalyst) обеспечивает лингвистический и семантический анализ текста, выявление сущности, визуализацию связей, систематизацию документов, резюмирование и обработку запросов на естественном языке;

– **Businessobjects Text Analysis** ([www.businessobjects.com/product/catalog/text\\_analysis/features.asp](http://www.businessobjects.com/product/catalog/text_analysis/features.asp)) – программа, позволяющая извлекать информацию о 35-типах объектов и событий, включая людей, географические названия (топонимы), компании, даты, денежные суммы, email-адреса и выявлять связи между ними;

– **Attensity suite** ([www.attensity.com](http://www.attensity.com)) – технология извлечения информации из неструктурированных текстов. Она позволяет выявлять информацию, содержащуюся в неструктурированном тексте и превращать ее в структурированные данные, имеющие связи, которые могут быть проанализированы.

Вариант такой системы в настоящее время реализован и используется в качестве компоненты системы конкурентной разведки X-SCIF украинской компании «Информационная корпоративная служба», которая позволяет пользователю в онлайн-режиме получать карты связей для выбранных объектов и помогает интерпретировать результаты. Предусматривается, что пользователь вводит в качестве запроса объект. Запрос направляется к базе данных связей, откуда выбираются соответствующие ему фрагменты – карты связей (уровень детализации и временная ретроспектива должны указываться параметрически).

После выявления релевантных объектов и связей выполняются процедуры их автоматической группировки (кластеризации) и визуализации, результаты предъявляются пользователю в виде карт связей, которые представляются в виде динамических (чаще всего, Java-диаграмм) графов связей.

В частности, в системе конкурентной разведки X-SCIF граф связей строится с помощью апплетов Java и представляет собой графический объект, который содержит в своем составе узлы и ребра. Каждый элемент графа связей имеет контекстное меню, которое является дополнительным элементом управления в интерфейсе пользователя (рис. 10).



лили выявить (после обращения к документам-первоисточникам) неочевидные на первый взгляд факты, например, то, что УкрСиббанк и УкрСоцбанк являются банками-партнерами.

Представленный подход может рассматриваться как основа построения так называемых «вертикальных» (предметно-ориентированных) информационно-поисковых систем, в которых изначально решены вопросы оперативности, отсеивания информационного шума. Рассматриваемая реализация имеет свойство масштабирования по трем параметрам: объему баз данных, составу понятий, которые используются, и по инфраструктурному окружению.

Анализируя связи в сети, можно определить многие неочевидные свойства, например, выявить наличие кластеров, определить их состав, различия в связности внутри и между кластерами, идентифицировать ключевые элементы, которые связывают кластеры между собой и т.п. Серьезным препятствием при анализе является неполнота информации о связях между отдельными узлами сети. Вместе с тем сегодня уже существуют алгоритмы, с помощью которых становится возможным с высокой вероятностью восстановить отсутствующие фрагменты связей. Даже не имея полного описания информационной сети, можно получать репрезентативную выборку «реальных» связей и по ней достроить всю сеть. Представленный подход реализует связующее звено между полнотекстовыми и фактографическими базами данных.

## 2.5. Технологии конкурентной разведки

Система конкурентной разведки должна позволять руководству, аналитическому, маркетинговому отделам компании не только оперативно реагировать на изменения ситуации на рынках, но и оценивать риски и возможности, прогнозировать их и принимать решения о дальнейших путях развития, обеспечить переход от традиционного интуитивного принятия решений на основе недостаточной информации к управлению, основанному на достоверных прогнозах и знаниях.

Одним из основных общих требований к системе конкурентной разведки должно быть соответствие цикла обработки информации в такой системе классическому информационному разведывательному циклу. Т.е. система должна самостоятельно или с участием оператора обеспечивать:

- выбор тематики и направлений разведки (целеуказание);
- выбор источников информации (веб-сайты, блоги, форумы и т.д.);

- автоматический поиск и скачивание информации по заданным направлениям мониторинга и указанным источникам по запланированному расписанию (планирование и сбор данных);
- обработку собранных данных и превращение их в информацию;
- контент анализ и синтез информации – превращение ее в знания;
- своевременную доставку информации к конечным потребителям.

Так как в целях конкурентной разведки необходимо анализировать данные из всех доступных источников информации, в которых эта информация может быть представлена в различных видах и форматах, то крайне важным требованием к системе является обеспечение ею единого информационного пространства взаимосвязанных объектов и фактов независимо от типа их источников или контента. Два других требования касаются сохранения связи объектов и фактов с релевантными данными и источниками информации (аргументированность) и обеспечения исторически-пространственной модели банка данных системы, что предполагает наличие у всех объектов атрибутов времени, места и источника данных, а также невозможность их безвозвратного удаления из системы с течением времени.

Основными объектами учета и мониторинга в системах конкурентной разведки, как правило, являются:

- источники информации (официальные сайты, интернет-издания, персональные сайты организаций или лиц, Интернет представительства печатных СМИ, информагентств, теле- и радиоканалов, открытые базы данных, объекты учета и т.д.);
- географические регионы;
- рынки и направления бизнеса;
- структуры (предприятия, организации и т.д.);
- персоны (конкуренты, контрагенты, партнеры, сотрудники, кандидаты и т.д.);
- нормативно-законодательная база и факты ее нарушения;
- политико-экономическая ситуация;
- криминальная обстановка;
- другие специализированные индивидуальные тематики.

Безусловно, система конкурентной разведки, использующая Интернет как один из источников информации, должна настраиваться под специфику деятельности компании. Она должна включать в себя соответствующую классификацию, гибкие механизмы поиска, оперативной доставки данных, а также качественной оцен-



ки информации. Одной из самых важных задач анализа информации является определение ее достоверности, т.е. задача анализа и фильтрации шума и ложной информации. Без таких оценок всегда есть риск принять неверные решения. После анализа достоверности информации должны следовать оценки ее точности и важности. Главным критерием достоверности данных на практике является подтверждение информации другими источниками, заслуживающими доверия.

Даже поверхностный анализ основных требований к системам конкурентной разведки в Сети, показывает, что традиционные поисковики в системе Интернет не могут считаться полноценными инструментами конкурентной разведки в Интернет.

Информационные системы конкурентной разведки можно также условно классифицировать по наличию в них модулей автоматического и экспертного извлечения фактов. Соотношение между автоматически извлекаемыми системой и вручную (с помощью экспертов) фактами, событиями, объектами учета в разных системах разное. Автоматически извлекаемые системой факты называют А-фактами, факты, извлекаемые экспертами, – Э-фактами [Киселев, 2005].

Существующие на рынке системы конкурентной разведки отличаются как по своей полноте и соответствию полному разведциклу, так и своему инструментарию и соответственно своей цене. Кроме того, системы могут быть предназначены для использования в качестве инструментария исключительно собственными силами внутреннего подразделения конкурентной разведки предприятия, либо предполагать вынесение части задач на аутсорсинг специализированными структурами конкурентной разведки. Выбор систем, подходов и методик конкурентной разведки остается за потребителем, и в каждом случае индивидуален. Да это и понятно, нельзя же сравнивать потребности и выполняемые задачи аналитика спецслужбы и сотрудника, к примеру, маркетингового отдела малого предприятия.

В настоящее время в мире существует ряд систем, которые частично реализуют решения приведенных выше задач мониторинга субъектов, извлечения фактов, построения связей, однако некоторые из них не выдерживают критики по функциональности, некоторые имеют слишком высокую цену. Кратко остановимся на возможностях некоторых подобных систем, реализованных в настоящее время.

Система **RCO** ([www.rco.ru](http://www.rco.ru)) – основное назначение – выявление фактографической информации из неструктурированных тек-

стов. Обладает широким спектром алгоритмов и технологий интеллектуальной обработки текстов, представленных на естественном языке. В частности, технологии RCO позволяют решать задачи выявления именованных объектов, связей и фактов из массивов неструктурированных данных. RCO Fact Extractor – это персональное приложение для Windows, которое предназначено для аналитической обработки текста на русском языке и выявления фактов, связанных с заданными объектами – лицами и организациями. Основная сфера применения программы – это задачи из области конкурентной разведки, требующие высокоточного поиска информации.

**RCO Zoom** (рис. 11) – поисково-аналитическая система с функциями контент-анализа в реальном времени и транзакционного хранилища документов. RCO Zoom обладает инструментарием для проведения эффективного оперативного поиска и аналитических исследований информации.

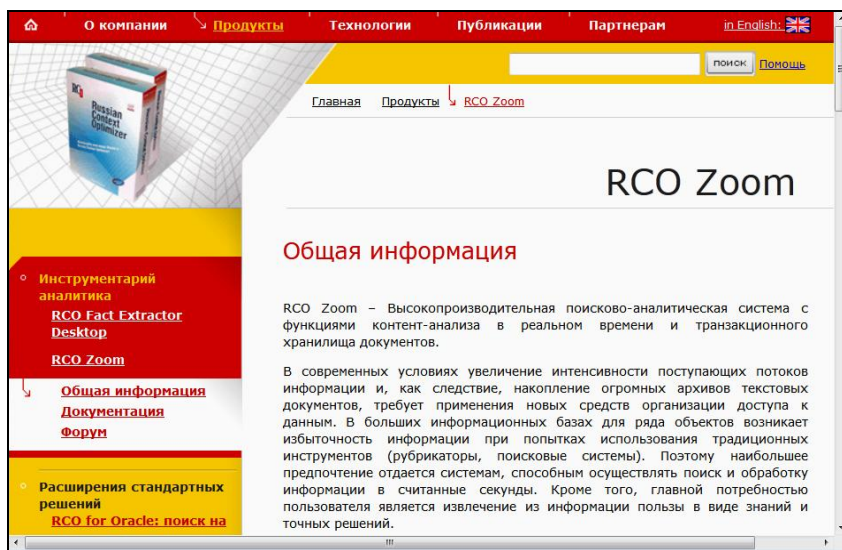


Рис. 11 – Фрагмент веб-сайта с описанием возможностей системы RCO Zoom

Система RCO Zoom позволяет работать с огромными массивами текстовой информации в реальном времени (объем базы – до сотен гигабайт, время поиска и обработки – секунды). Средство отображения – информационный портрет дает возможность полу-

чить ключевые слова, формулировать и проверять гипотезы, разделять объекты, выделять статистические инварианты в первом приближении.

«**Медиалогия**» ([www.mlg.ru](http://www.mlg.ru)) – сервис, обеспечивающий онлайн-доступ к базе СМИ с возможностью производить самостоятельный мониторинг СМИ и экспресс-анализ полученных сообщений. С помощью «Медиалогии» можно осуществлять оперативный мониторинг СМИ компании, ее топ-менеджеров, брендов, конкурентов и др. Возможности фильтров позволяют настроить мониторинг и оценивать тональность прессы, журналов, ТВ и интернет-изданий практически под любые информационные задачи.

**PolyAnalyst** ([www.megarputer.ru](http://www.megarputer.ru)) – основной продукт компании Megarputer Intelligence, который позволяет получить полезные знания из массивов текстовых и структурированных данных, передает эти знания в доступной форме и в виде оперативно применяемых моделей для поддержки принятия решений. Система PolyAnalyst охватывает весь традиционный цикл анализа данных, начиная с импорта и предварительной обработки, и заканчивая отчетами. Функциональность PolyAnalyst, в частности, включает возможности создания многоступенчатых сценариев анализа данных и формы отчетов посредством простого drag-and-drop интерфейса, формировать аналитические прогнозы, предоставлять клиентам результаты, полученные из прогнозов, в доступных для понимания формах.

**RetrievalWare** ([www.verticalsearchworks.com](http://www.verticalsearchworks.com)) – система, обеспечивающая поиск, анализ и выявление информации с помощью задания пользователем поисковых запросов на естественном языке как к массивам информации, хранящихся в неструктурированном виде, так и в формализованных базах данных. В RetrievalWare реализована возможность выявления семантической связи для понятий, например, слова синонимы или антонимы, слова связанные отношениями «род-вид» и т.д. Данное свойство обеспечивает высокую точность выявления понятий и семантическое расширение запросов пользователей. В настоящее время компания Convera, известная как поставщик этого аналитического продукта, объединилась с Firstlight ERA, образовав специализирующуюся на семантическом поиске в веб-пространстве компанию Vertical Search Works.

**Autonomy IDOL Server** ([www.autonomy.com](http://www.autonomy.com)) был создан в компании Autonomy (рис. 12), купленной недавно корпорацией Hewlett-Packard. Идеологическая основа Autonomy IDOL (Intelligent Data Operating Layer) – это Meaning-Based Computing (MBC),

что переводится как «компьютерные системы, основанные на смысле». В Autonomy принята определенная иерархия технологий работы с текстами, где на нижнем уровне стоит простой поиск по ключевым словам, а верхний уровень занимают интеллектуальные технологии МВС. Autonomy IDOL включает в себя свыше 500 различных функций для извлечения смысла из текстов и более 400 коннекторов к различным источникам данных, поддерживает работу со встроенными узлами Hadoop и СУБД Vertica. Для визуализации результатов используются как собственные средства пользовательского интерфейса, так и технологии компаний QlikView или Tableau.



Рис. 12 – Фрагмент веб-ресурса Autonomy IDOL

**Endeca Latitude** ([docs.oracle.com/cd/E29220\\_01](http://docs.oracle.com/cd/E29220_01)) – технология, реализующая концепцию Text Mining, приобретенная корпорацией Oracle. Ядром технологии Endeca Latitude является MDEX Engine – гибридная аналитическо-поисковая система управления базами данных (СУБД), для которой характерна высокая масштабируемость. С помощью СУБД Oracle Essbase входные данные проходят предварительную обработку и агрегируются, а MDEX нацелена на поиск и обнаружение знаний и является гибридом поиско-

вой и аналитической СУБД, рассчитанной на работу с динамическими данными. Принципиальное отличие MDEX от обычных СУБД состоит в приближении хранящихся в ней записей к реальностям окружающего мира. Таким образом, Endeca сфокусирована на раскрытии содержимого данных (Data Discovery) в противоположность традиционному анализу.

**Vivisimo** ([www-01.ibm.com/software/data/information-optimization](http://www-01.ibm.com/software/data/information-optimization)) – технология, включающая три взаимосвязанных продукта, объединенных в платформу IBM InfoSphere Data Explorer (ранее известную как Velocity Platform): Search Engine – многофункциональная поисковая система, роботы которой способны собирать и обрабатывать файлы различных типов (HTML, TXT, RTF, PDF, MS Word, Excel, PowerPoint, WordPerfect и т.д.), а также извлекать информацию из реляционных СУБД; Clustering Mashine – машина кластеризации, группирующая результаты работы поисковых машин в Google, Autonomy, FAST и Ultraseek, а также тексты в различных форматах; Content Integrator – интегратор, обеспечивающий объединение результатов поиска (федерализацию), умеющий работать с метаданными.

**Attensity suite** ([www.attensity.com](http://www.attensity.com)) – технология выявления понятий, фактографической информации из неструктурированных текстов. Технология Attensity suite позволяет выявлять информацию, скрытую в неструктурированном тексте, и переводить ее в структурированные данные, имеющие связи, которые могут быть проанализированы теми же методами, что и другие виды структурированных данных. В настоящее время технология ориентирована на обработку так называемых «больших данных» (Big Data) и информации из социальных сетей.

**Statistica Text Miner** ([www.sas.com/text-analytics/text-miner](http://www.sas.com/text-analytics/text-miner)) – расширение известной программы Statistica Data Miner, предназначенное для перевода неструктурированных текстов в информацию, пригодную для принятия решений. Кроме того, система Statistica Text Miner позволяет устанавливать принадлежность документов к той или иной смысловой категории, идентифицировать группы сходных по смыслу документов, устанавливать связи между полученными характеристиками документов и другими индикаторами и т. д. Программное обеспечение для Text Mining полностью интегрируемо с любым другим программным обеспечением, принадлежащим линии продуктов Statistica.

**Rocket AeroText** ([www.rocketsoftware.com/products/rocket-aerotext](http://www.rocketsoftware.com/products/rocket-aerotext)) – программа, позволяющая выявлять элементы информации, такие как сущность (Entities), взаимные связи (Relation-

ships) и события (Events), в неструктурированных текстах, а также выявлять неявные взаимосвязи и события в текстах.

В последнее время все основные западные бренды, специализирующиеся на разработке хранилищ и баз данных, корпоративных системах управления, расширили свои линейки продуктов модулями Business Intelligence (BI) или, дословном переводе – деловой разведки. О наличии таких модулей заявляют Oracle, SAS, SAP, IBM и другие бренды.

По заказу группы аналитиков Гарвардского университета российские разработчики из «Инфорус» создали информационно-аналитическую систему (ИАС) **Avalanche** ([www.toga-centre.ru/avl3.htm](http://www.toga-centre.ru/avl3.htm)), предназначенную для мониторинга изменений, происходящих в Интернете. Она собирает информацию с веб-страниц по заданному алгоритму и складывает эту информацию в собственную базу данных.

Технология Avalanche базируется на трех компонентах: автономном интеллектуальном поисковом роботе, создании «умных» папок и встроенной базе данных, позволяющей преобразовать их в «персональную энциклопедию». При работе с ИАС Avalanche формируется модель, требуемой пользователю области в виде набора «умных папок», каждая из которых «знает», что должно в нее попадать, и обеспечивает отсутствие дублирования. Наполнением «умных» папок занимается специализированный поисковый робот, который запускается с компьютера в соответствии с установленными, требуемыми пользователю настройками. Робот может запускаться и автоматически в определенное установленное для него время. В Avalanche предусмотрены «тонкие» настройки, которые позволяют производить более детальный мониторинг.

Информационно-аналитическая система «**Семантический архив**» ([www.anbr.ru/products/semarchive](http://www.anbr.ru/products/semarchive)) представляет собой инструмент для создания интегрированного хранилища информации с возможностью хранения досье на объекты мониторинга, происходящие события, а также текстовые документы.

Гибко настраиваемая онтологическая модель данных позволяет работать с разными тематиками и сферами деятельности. ИАС «Семантический архив» позволяет хранить информацию, импортированную из различных реляционных баз данных, вводить информацию из любых других источников: Интернет, СМИ, базы данных, онлайн библиотеки и системы (Спарк, Интегрум и др.), любой документ, собственные сведения экспертов.

Созданное хранилище служит аналитикам для поиска информации, добавления конфиденциальных собственных данных, вы-

явления взаимосвязи между объектами и событиями, получения аналитических отчетов, визуализации: схем, графиков и карт.

«Семантический архив» имеет модульную структуру, что позволяет легко подобрать и настроить нужную конфигурацию системы.

Российская система управления досье **XFiles** предназначена для решения задачи выделения достоверных фактов из различных источников, заполнения ими досье на объекты мониторинга и их последующей аналитической обработки. Она используется для обеспечения процессов принятия решений при наличии большого объема «сырого» контента, что характерно для деятельности органов государственной власти, правоохранительных органов, крупных коммерческих компаний.

Система XFiles предполагает реализацию трех принципов:

1) единое информационное пространство взаимосвязанных фактов или гипотез независимо от типа их контента (содержимого источников информации);

2) связь фактов или гипотез с релевантными источниками информации (аргументированность фактов и гипотез);

3) исторически-пространственная информационная модель базы данных фактов и гипотез. Это означает наличие атрибутов времени и места для каждого факта, а также невозможность их безвозвратного удаления из системы.

Система XFiles применяется и как аналитический инструмент, и как система поддержки корпоративной безопасности подразделений, занимающихся бизнес-разведкой. Система позволяет извлекать факты из сети Интернет, корпоративных источников документов, из учетных баз данных и других. Система Xfiles имеет веб-интерфейс для работы пользователей. В Xfiles реализована семантическая сеть, отражающая лишь взаимосвязи между объектами. В системе отсутствует направление мониторинга СМИ, интернет-роботов и др. Базируется на средствах портала управления знаниями Hummingbird.

Среди самых развитых порталов управления знаниями, применяемых для решения задач конкурентной разведки, нельзя не назвать систему **Hummingbird Enterprise** канадской компании Hummingbird ([www.hummingbirdenterprise.com](http://www.hummingbirdenterprise.com)). Среди множества компонент системы можно выделить Hum-mingbird Portal – платформу, позволяющую интегрировать информацию из информационного хранилища и приложения в едином веб-интерфейсе. Эта платформа является полнофункциональным порталом знаний.

Основной особенностью системы **i2 Analyst Notebook** корпорации IBM ([www-01.ibm.com/software/industry/i2software/support.html](http://www-01.ibm.com/software/industry/i2software/support.html)) является возможность визуального проектирования структуры данных для хранения данных о различных персонах и организациях.

В базе данных предусматривается возможность хранения определенных событий, происходящих с ними и имеющиеся взаимосвязи. Система i2 Analyst's Notebook позволяет быстро и эффективно проводить анализ системы взаимосвязанных объектов и динамики последовательных событий, отображая результаты исследования в виде удобных для понимания схем и диаграмм.

Информация отображается на диаграмме в виде объектов, к которым при необходимости можно добавить дополнительные атрибуты и карточки данных с комментариями. Объекты на диаграмме могут представляться не только в виде пиктограмм, но и в виде фотографий, файлов, аудиозаписей, видеозаписей и т.д.

Программа позволяет создавать диаграммы с помощью запросов к реляционным базам данных, а также импорта данных из внешних файлов. При помощи имеющихся в Analyst's Notebook функций можно объединять элементы диаграмм, искать существующие между ними связи, использовать систему поиска элементов, проследивать «путь», объединяющий объекты, и т.п.

Analyst's Notebook снабжен редактором, позволяющим в графической форме сформулировать запрос для поиска объектов и выявления их связей, создавать шаблоны интересующих событий.

Analyst's Notebook предоставляет целый ряд удобных форматов визуализации, каждый из которых по-своему проясняет смысл информации и демонстрирует связи между объектами.

Систему Analyst's Notebook можно интегрировать в уже работающие у пользователя приложения.

Система Analyst's Notebook обеспечивает:

- поиск общих элементов и взаимосвязей, скрытых в данных;
- простоту интерпретации сложной информации;
- графическое отображение результатов;
- создание динамических диаграмм;
- распространение диаграмм в печатном и электронном виде.

Говоря о западных продуктах, лидирующих в области Business Intelligence, следует отметить, что под этим термином, как правило, понимается набор инструментальных средств анализа статистических цифровых данных и других корпоративных отчетов и их визуализации, в отличие от Competitive Intelligence (конкурентной



разведки), которая является гораздо более широким направлением информационной деятельности.

На российском рынке, кроме упомянутых выше, в сегменте инструментальных средств конкурентной разведки в Интернет, представлено довольно много подобных систем. Перечислим только наиболее известные из них. К российским системам, заслуживающим внимания, можно отнести такие информационно-аналитические системы, как «*Интегрум*», «*Аналитический курьер*», «*Астарта*», «информационно-программные комплексы *Intellectum.BIS*, «*Арион*», *X-files*, *Cronos*, *CronosPlus*, *T2 BANC* и др.

На украинском рынке в этом сегменте представлены такие системы, как «*Сфера*», *X-SCIF*, «*Энциклопедия деловой информации*», «*ICEBERG*», «*Пиранья*», «*Страбис*» и др.

Хотелось бы отметить, что далеко не все из названных систем имеют полный функционал и соответствующие модули, обеспечивающие выполнение всего спектра задач конкурентной разведки.

В качестве одной из наиболее полнофункциональных отечественных систем, обработка информации в которой соответствует классическому информационному разведывательному циклу, можно назвать систему *X-SCIF*.

Рассмотрим, как реализуются этапы разведки с помощью данной системы, для чего остановимся на описании возможностей системы *X-SCIF* чуть подробнее.

Онлайновая инструментальная корпоративная система мониторинга, агрегации и анализа информации *X-SCIF* (далее – ИКС *X-SCIF*) представляет собой программно-технический комплекс, предназначенный для решения задач автоматизированного сбора, обработки, создания интегрированного банка данных и анализа разнообразной информации.

Система обеспечивает:

- мониторинг информации с заданных пользователем веб-сайтов (веб-страниц) в сети Интернет (Интранет) по заданным темам;
- поиск новых источников информации в сети Интернет по заданным пользователем тематикам и их последующую постановку на мониторинг;
- создание и сохранение сложных запросов по заданным темам, в виде каталогизированного списка или рубрики, для последующего проведения автоматического мониторинга, поиска или контент-анализа;
- приведение отобранной информации к единому формату и ее загрузку в хранилище;

- фильтрацию, классификацию, кластеризацию, рубрицирование и анонсирование загруженной полнотекстовой информации;
- автоматическое экстрагирование (извлечение) из полученной информации сущностей (объектов и фактов);

- создание, на базе загруженной в систему неформализованной полнотекстовой и формализованной фактографической информации, интегрированного банка данных (хранилища) объектов, фактов, событий и документов, связанных между собой различными видами и мотивами связей, с учетом атрибутов достоверности, актуальности, а также, весовых коэффициентов таких связей;

- сквозной поиск информации по запросам или темам пользователя, охватывающий, как поиск по внутреннему интегрированному банку данных, ранее уже извлеченной и накопленной информации, так и онлайн-поиск в сети Интернет (поисковые системы, веб-сайты, блоги, социальные сети) и другим подключенным внешним источникам данных (официальные банки данных госорганов, БКИ и т.д.);

- аналитическую обработку информации (позволяет анализировать совместное упоминание и выявлять неявные связи между объектами, отождествлять объекты и группировать информацию по сюжетам, строить цепочки и графы связей, анализировать информационную активность, эмоциональную окраску документов, пересечение заданных рубрик или тем, автоматически создавать информационный портрет отобранных по запросу документов, выделяя упоминаемые в них объекты, источники, регионы и т.д., вычислять индекс информационного благоприятствования и многое другое);

- генерацию выходных форм по заданным пользователем параметрам (позволяет автоматически создавать электронное досье, схемы связей, дайджесты, обзоры, сравнительные диаграммы, информационные справки и агрегированные отчеты);

- оперативную доставку результатов запросов по различным каналам (в состав системы входит виртуальный офис, с собственным удаленным криптозащищенным хранилищем документов и почтовой системой, что позволяет обеспечивать, как «онлайн-безопасный доступ по зашифрованному каналу к документам, хранящимся в облаке, так и «оффлайновое» получение результирующих документов по электронной почте).

Структурно ИКС X-SCIF состоит из нескольких подсистем, ориентированных на соответствующие потребности корпоративных заказчиков, а именно:

– X-Stream – подсистема мониторинга web-сайтов, созданная на основе технологии InfoStream, а также полнотекстового банка данных (архива) неформализованной информации (статей, сообщений и т.д.), который автоматически пополняется с 1996 года и является наиболее полным среди существующих электронных архивов в Украине;

– X-Files – интегрированный банк данных для накопления разнообразной формализованной справочно-фактографической информации, экстрагированной и агрегированной из всех доступных системе источников информации, стоящих на мониторинге, а также системы сквозного поиска по внутренним и внешним источникам (веб-сайтам, блогам, онлайн-базам данных, социальным сетям и т.д.).

– X-Office – система виртуального офиса, обеспечивающая безопасный доступ к корпоративным ресурсам из любой точки мира без установки дополнительного программного обеспечения. Система включает в себя «облачное» файловое хранилище документов и защищенную корпоративную веб-почту. Дополнительно в виртуальный офис может быть интегрирован сервер VoIP-телефонии для ведения конфиденциальных переговоров.

– X-Scoring – предскаринговая система, которая позволяет в автоматическом режиме проводить верификацию данных и предварительную проверку благонадежности контрагентов (физических и юридических лиц).

Остановимся на рассмотрении каждой из подсистем более детально.

Подсистема X-Stream, построенная на основе технологии InfoStream компании ElVisti, предназначена для мониторинга информации в сети Интернет по заданным пользователем параметрам, поиска информации по запросам или темам пользователя, оперативной доставки результатов поиска, и, таким образом, минимизации усилий и экономии времени, потраченного на поиск и обработку необходимой информации. Подсистема X-Stream предоставляет пользователю доступ к информации по интересующей его тематике одновременно с большого числа веб-сайтов, включая и те избранные, которые он привык просматривать ежедневно. В настоящее время осуществляется автоматический мониторинг более 7000 источников, поток информации превышает 80 000 документов в сутки. Территориальный охват – русско-англо- и украиноязычные издания Украины, России и других стран ближнего и дальнего зарубежья. При необходимости может быть охвачен любой веб-сайт, доступный в сети Интернет. Информация из системы

никогда не удаляется, а переносится в архив. Архив публикаций ведется непрерывно с 1996 года и составляет в настоящее время свыше 85 миллионов документов.

Отличие данной подсистемы от конкурирующих продуктов – ее объемы и возможность индивидуальной настройки. Она ориентирована не только на быструю доставку общих лент новостей, которых много в веб-пространстве, но и на осуществление мониторинга по индивидуально заданным пользователем параметрам или архивного поиска.

Просмотр информации осуществляется через единый унифицированный интерфейс. Пользователь может в режиме реального времени не только получать результаты поиска, но и формировать дайджесты, информационные досье, строить сюжетные цепочки, анализировать взаимосвязь рубрик, информационную активность, информационные связи и совместное упоминание объектов и т.д.

Ниже приведены примеры вывода результатов поиска (рис. 13), просмотра отдельного материала (рис. 14) и результата аналитической обработки результатов поиска – дайджеста (рис. 15).

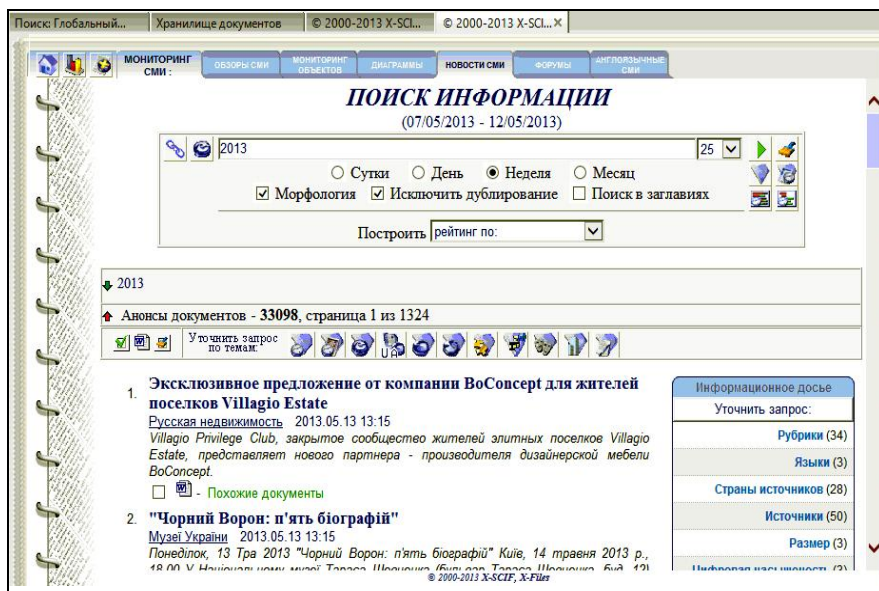


Рис. 13 – Вывод результатов поиска



Рис. 14 – Просмотр статьи

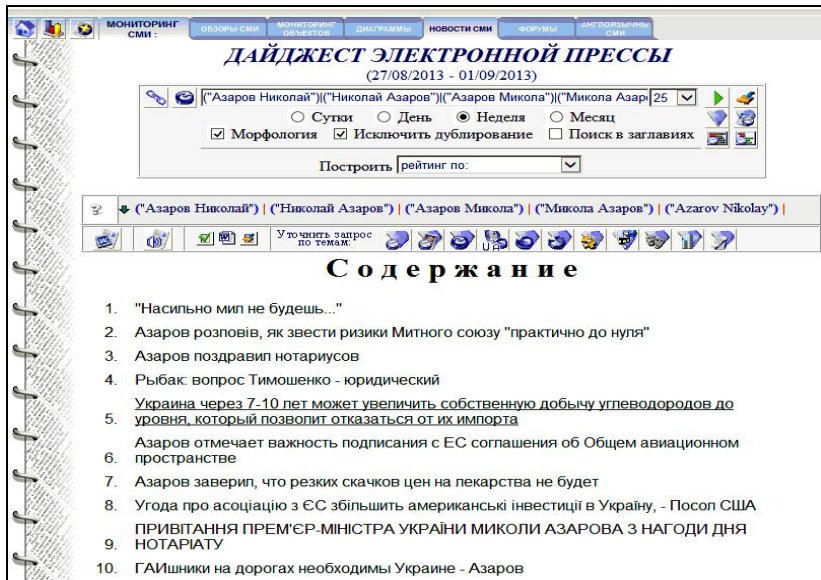


Рис. 15 – Дайджест

Использование подсистемы X-Stream позволяет:

- оперативно получать необходимую информацию по мере ее появления в Интернет, анализировать события, своевременно на них реагировать;

- формировать собственные информационные каналы, которые обусловлены запросами на информационно-поисковом языке, формировать архивы для последующей обработки и ретроспективного анализа;

- анализировать поток информации, поступающей в режиме реального времени;

- своевременно выявлять тенденции развития и состояние рынков товаров или услуг;

- отслеживать информацию о деятельности отдельных организаций, партий, движений, их PR-активность;

- оценивать возможные сферы влияния конфликтных или кризисных ситуаций, осуществлять информационный контроль вероятных источников рисков;

- находить и проверять потенциальных партнеров и клиентов.

Следующим структурным элементом ИКС X-SCIF является подсистема X-Files (не следует путать с известной российской системой). Данная подсистема предназначена для накопления и хранения формализованной информации, полученной из всех доступных источников, осуществления сквозного поиска и дальнейшей аналитической обработки найденной информации.

Информация, полученная из различных источников, обрабатывается, формализуется, приводится к единому виду и записывается в интегрированный банк данных, который структурно охватывает объекты и связи между ними. Его структура, разработанная с учетом практических потребностей аналитиков, включает в себя более 40 типов объектов учета и более 1000 мотивов связей между ними.

Поиск необходимой информации осуществляется средствами глобального поиска, который выполняется по всем доступным банкам данных, а также предусматривает автоматическое получение информации от онлайн-поставщиков информации.

На рис. 16 приведен интерфейс ввода запросов для глобального поиска. Подсистема позволяет представлять результаты поиска в различной форме, наиболее удобной для решения текущей задачи.

Одной из наиболее распространенных форм представления отобранных данных является информационное досье (рис. 17). Данная форма позволяет отображать информацию об объекте учета интегрированного банка данных в виде, в котором представлены все реквизиты данного объекта учета, а также все связанные с ним

записи в других банках данных. Формат вывода информационного досье приведен на рис. 18.

**Глобальный поиск**

<b>Поисковые значения</b>	
Код предприятия/учредителя	<input type="text"/>
Наименование предприятия, органи	<input type="text"/>
Телефон/факс	<input type="text"/>
МФО	<input type="text"/>
Счет	<input type="text"/>
Адрес	<input type="text"/>
Фамилия	<input type="text"/>
Имя	<input type="text"/>
Отчество	<input type="text"/>
Дата рождения	>= <input type="text"/> <= <input type="text"/>
Свидетельство налогоплательщика	<input type="text"/>
Идентификационный номер	<input type="text"/>
№ паспорта	<input type="text"/>
Другое удостоверение	<input type="text"/>
Гос. номер авто	<input type="text"/>
Номер кузова	<input type="text"/>
Электронный адрес	<input type="text"/>
Контекстный поиск	<input type="text"/>
Поиск в Интернет и архивах	<input type="text"/>

Рис. 16 – Интерфейс ввода запросов глобального поиска

**1. Информационная справка - Сообщения СМИ**

<b>Ключевые реквизиты</b>	
Вид сообщения	СООБЩЕНИЕ О БАНКРОТСТВЕ
Название	"Голос України" N 87 (3337) від 14 травня 2004
Дата	14.05.2004
Анонс	18.03.2004 р. господарським судом Харківської області (61022, м. Харків, Держпром, ( п.д.) порушено провадження по справі N Б-19/22-04 про визнання банкрутом Сільськогосподарського
<b>Информация</b>	
Текст сообщения	18.03.2004 р. господарським судом Харківської області (61022, м. Харків, Держпром, 8 п.д.) порушено провадження по справі N Б-19/22-04 про визнання банкрутом Сільськогосподарського товариства з обмеженою відповідальністю "Агрокомбінат Богодухівський" (62103, м. Богодухів, Харківської обл., вул. Залізнична, 14, код ЄДРПОУ 22660116, п/р 26008301747 у Першому ХФ АКЕ "Базис", МФО 351599). Розпорядником майна призначено арбітражного керуючого Панасюка І. В. (ліцензія АА N 047594 від 03.07.01, адреса: м. Харків, вул. Полтавський шлях, буд. 154, кв. 84). Заяви кредиторів приймаються протягом місяця з дня публікування.
Источник	<a href="#">Голос України</a>
<b>Дополнительная информация</b>	
Особенности информации	ВОЛЬШАЯ ЦИФРОВАЯ НАСЫЩЕННОСТЬ
Дата ввода	17.07.2012
Дата редактирования	17.07.2012
	<b>Связанная информация</b>
<b>СМИ о ЮИ</b>	
Мотив связи	По сообщениям СМИ
<b>Дополнительные реквизиты</b>	
Дата возникновения связи	14.05.2004
Достоверность	ДОСТОВЕРНАЯ ИНФОРМАЦИЯ
<b>Ключевые реквизиты</b>	
Код ОКПО	22660116
Наименование объекта	Сільськогосподарське Товариство з обмеженою Відповідальністю "Агрокомбінат "Богодухівський"
Правовая форма	ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ
Адрес (текст)	ХАРКІВСЬКА ОБЛ., БОГОДУХІВСЬКИЙ Р-Н, М.БОГОДУХІВ ВУЛ. ЗАЛІЗНИЧНА БУД. 14
Страна	УКРАЇНА

Рис. 17 – Информационное досье

(заключение на кандидата/сотрудника)

Дата: 13.05.2013 № \_\_\_\_\_ на вх. № \_\_\_\_\_

Место работы, должность

Организация: ██████████ ТОВАРИСТВО З ОБМЕЖЕНОЮ ВІДПОВІДАЛЬНІСТЮ  
██████████


Должность: Охранник

фотография, краткая характеристика,  
результаты проверки

Род занятий, специализация: СПЕЦІАЛІСТЫ, ОХРАННИК,  
ТЕЛОХРАНИТЕЛЬ

Краткая характеристика: Здесь текст характеристики

Семейное положение: ЖЕНАТ  
Хобби: Рыбалка



Установочные данные:

(дата и место рождения, гражданство, адрес регистрации, адрес проживания и др.)

Дата рождения: ██████████

Место рождения: ТУРКМЕНИСТАН, м.Красноводськ

Гражданство: УКРАИНА

Адреса регистрации, проживания:  
УКРАЇНА, м. КИЇВ, БРАТИСЛАВСЬКА, ██████████

Телефон:  
38044 ██████████  
██████████

Адрес регистрации:

Рис. 18 – Формат вывода информационного досье

Еще одним примером формы представления данных является графическое досье. Отобранные объекты, вместе со своими связями, отображаются в виде графа, в котором вершинами выступают объекты учета, а ребрами – связи между соответствующими объектами (рис. 19). Такая форма представления позволяет осуществлять аналитические исследования как явных, так и неявных связей объектов учета, представлять их на экране в виде графов в различных масштабах, печатать схемы этих графов т.д. Также в данном режиме пользователю доступны все инструменты по редактированию, вводу и удалению информации, обеспечивающие интуитивное и быстрое редактирование формализованных данных.

Для аналитической обработки больших объемов однотипной информации в системе предусмотрен механизм агрегированных форм. Он позволяет на основании исходной информации, которая плохо поддается непосредственному анализу, строить агрегированные формы, графики, проводить расчет интегральных характеристик.



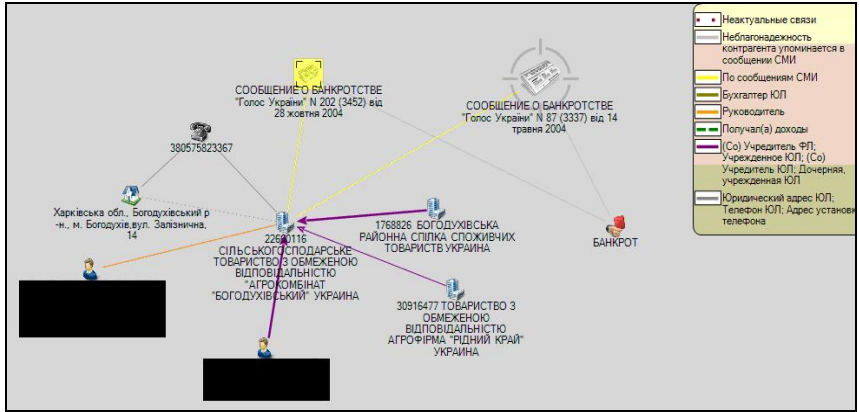


Рис. 19 – Визуализация графов связей объекта учета

Одной из ключевых возможностей подсистемы X-Files является модуль автоматизированного ввода и распознавания полнотекстовых документов. Он позволяет без участия оператора создавать объекты учета (лица, компании, телефоны, адреса, электронные адреса и другие) и устанавливать связи между ними на основании неформализованных документов (тексты, анкеты, карточки и т.д.).

Для обеспечения удаленного взаимодействия пользователей, и обеспечения эффективной совместной работы предназначена подсистема X-Office. В ее состав, в свою очередь, входят такие подсистемы:

- «Корпоративная веб-почта». Обеспечивает работу с корпоративной почтой из любого места по шифрованному каналу связи без необходимости настройки и «следов» на компьютере;

- «Файловое хранилище документов» (рис. 20). Представляет собой удаленное защищенное файловое хранилище, доступное из любой точки только членам закрытой группы. Обеспечивает доступ к личным и корпоративным документам с возможностью совместной работы нескольких пользователей (рис. 21). Хранилище документов предоставляет возможность сквозного поиска по содержанию документов. Разграничение доступа к тексту документа производится согласно профилю доступа или с разрешения автора документа;

- «Переговорная» (рис. 22). Обеспечивает пользователям системы возможность общения внутри закрытой группы в текстовом, голосовом и видео режиме по защищенному протоколу связи. Так-

же доступна возможность совершать звонки на стационарные и мобильные телефоны вне группы с невозможностью определения исходящего абонента.

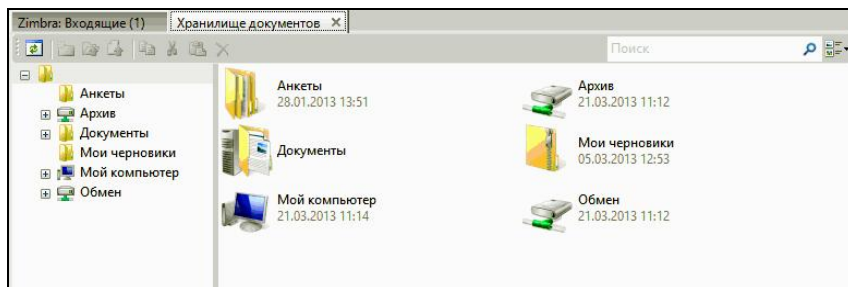


Рис. 20 – Интерфейс работы с хранилищем корпоративных документов

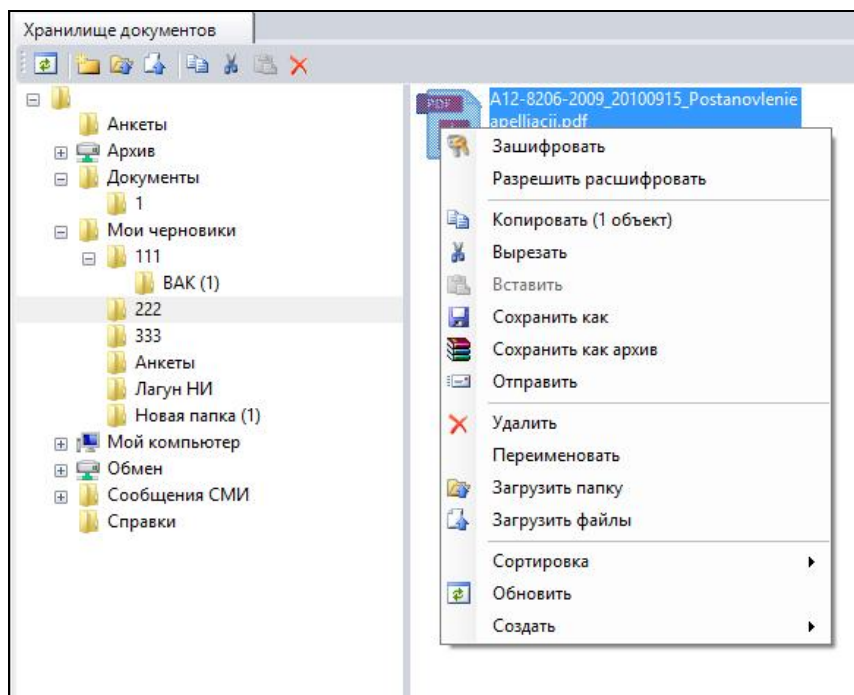


Рис. 21 – Операции, доступные в файловом хранилище

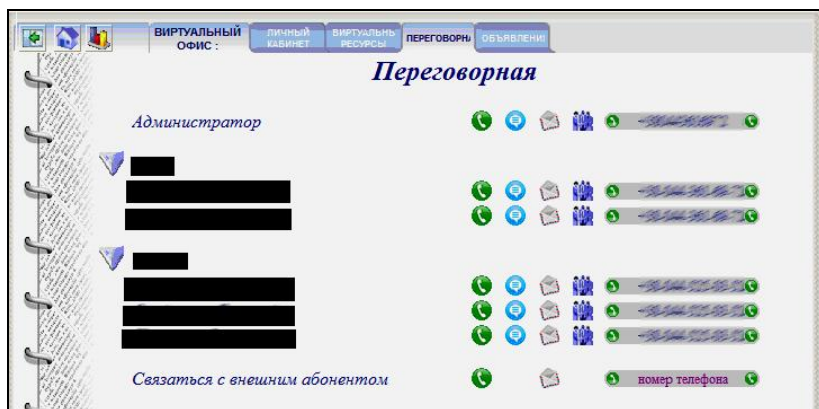


Рис. 22 – Интерфейс подсистемы «Переговорная»

Подсистема X-Office обладает рядом особенностей, которые дают ее пользователям существенные преимущества по сравнению с аналогичными программно-аппаратными комплексами. Авторизация в подсистеме происходит по комбинации отпечатка пальца и пароля, а для коммуникаций используются шифрованные каналы связи и только доверенные сертификаты, что предотвращает возможность анализа трафика на уровне интернет-провайдера или в любой другой точке перехвата. По завершении работы с подсистемой на компьютере пользователя не остается никаких следов ее функционирования.

Файловое хранилище корпоративных документов, которое входит в состав подсистемы X-Office, предоставляет пользователю возможность работы с удаленным облачным хранилищем так же просто, как с папкой на локальном диске его компьютера. Подсистема делает работу с различными источниками (локальный диск, корпоративные документы, файловые хранилища) прозрачной для пользователя. Редактирование офисных документов возможно без установки и настройки дополнительного программного обеспечения. Разграничение доступа к различным файлам осуществляется на основании шаблонов доступа, заданных администратором. В случае наличия особо секретной информации пользователь имеет возможность дополнительно ограничить к ней доступ средствами шифрования прямо из интерфейса программы.

Для проверки благонадежности контрагентов в автоматическом режиме доступна подсистема X-Scoring. Система реализована на основании технологии XML Web service, что обеспечивает про-

зранию интеграцию с большинством существующих систем на стороне заказчика. Несмотря на большой объем постоянно пополняющейся информации, на основании которой принимается решение о благонадежности контрагента, система дает ответ менее чем за 3 сек.

Алгоритм принятия решения может гибко настраиваться под потребности конкретного заказчика. Типовой алгоритм проведения проверки и принятия решения о благонадежности контрагента состоит из следующих этапов:

- оценка экономической платежеспособности клиента по предоставленным им сведениям;

- автоматическая проверка по банку данных, нацеленная на проверку соответствия предоставляемых заемщиком сведений и выявление возможных попыток мошенничества со стороны недобросовестных заемщиков;

- детальная проверка контрагента, заявка которого прошла все предыдущие этапы.

Следует отметить, что не всегда полнофункциональные системы конкурентной разведки являются доступными и даже необходимыми, ввиду их стоимостных характеристик или других причин. Вместе с тем, отдельные задачи конкурентной разведки могут быть частично решены вполне доступными средствами. Использование новых подходов, а также открытых, доступных и относительно недорогих информационных источников, позволяет уже сегодня эффективно поддерживать принятие управленческих решений по очень многим, в том числе и стратегическим, направлениям бизнеса.

Технологии конкурентной разведки завтрашнего дня сегодня во многих случаях уже реализованы в виде военных информационных технологий. Так, например, для осуществления разведки на государственном уровне в США еще в 2005 году в рамках Национальной разведки (National Intelligence) была создана специальная структура – Центр открытых источников (Open Source Center). В настоящее время все разведывательные центры США работы с открытыми источниками объединены в эту единую информационную систему. В 2006 г. информационные ресурсы данной системы получили название Intelink-U [Кондратьев, 2010].

Несмотря на то, что материалы в данной системе добываются из общедоступных открытых источников, она предназначена далеко не для всех. Информация из системы распространяется по сетям ограниченного доступа.

В состав системы Intelink-U входят многочисленные базы данных, среди которых:

- база данных CIRC, содержащая свыше 10 млн. статей научно-технической тематики, включая информацию о патентах, стандартах, военном вооружении и военной технике;
- база данных DTED, содержащая большое количество разнообразных карт, полученных от Национального управления геопространственной разведки;
- материалы центров и пунктов информационной службы зарубежного вещания FBIS;
- база данных периодических изданий IC ROSE;
- информационные порталы научно-исследовательских и учебных заведений;
- онлайн-справочники информационной службы Jane's Information Group;
- ресурсы негосударственного информационно-аналитического агентства STRATFOR's, предоставляющего, среди прочего и регулярные обновления по районам развертывания авианосных и экспедиционных ударных групп ВМС США.

Внимания заслуживают и способы наполнения подобных информационных ресурсов. Например, для сбора информации с сайтов всемирной сети Интернет, ее систематизации, перевода и архивирования во Всемирной информационной библиотеке (World Basic Information Library – WBIL), обеспечение функционирования которой возложено на Управление изучения Вооруженных сил (ВС) иностранных государств FMSO Командования учебного и научных исследований по строительству сухопутных войск (CB) (Training and Doctrine Command – TRADOC), привлекается личный состав резерва СВ и других видов Вооруженных сил.

Для обеспечения потребностей министерства обороны и разведывательного сообщества США создана база данных HARMONY, содержащая библиографические справки по всем имеющимся источникам информации (метаинформацию) о зарубежных странах. База данных HARMONY характеризуется простотой использования, возможностью быстрого поиска необходимых документов, быстрого обмена данными внутри правительственных структур США.

Всемирная информационная библиотека (World Basic Information Library, WBIL) представляет собой специальную программу разведывательного сообщества США, управление которой возложено на отдел изучения вооруженных сил иностранных государств (Foreign Military Studies Office, FMSO) командования учебного и научных исследований по строительству сухопутных войск (Training and Doctrine Command, TRADOC). Персонал с правом доступа к

базе данных может осуществлять сбор информации из сети Интернет, ее систематизирование и архивирование в библиотеке WBIL, применяя аналитический инструментарий Pathfinder. Система Pathfinder позволяет за несколько минут проводить анализ 500 тыс. документов из различных баз данных.

Совокупность названных технологий позволяет сотрудникам военной разведки США получать доступ к огромным массивам данных, удовлетворять потребности в разведывательной информации.

Рассмотрим некоторые проекты Агентства национальной безопасности (АБН) США, ориентированные на сферы добычи данных, аналитики и прогнозирования [Черных, 2013].

Проект тотальной интеграции информационных потоков предполагал создание программных средств, обеспечивающих решение весьма сложной задачи, не решенной по настоящее время в гражданском секторе – интеграцию всех информационных потоков в едином хранилище с одной стороны, и их разделение по специальным критериям, с другой стороны. К настоящему времени, согласно мнению экспертов, задача полностью решена. В военной сфере для извлечения информации из потоков данных, поступающей по каналам АБН, таким как «Эшелон», «Титан», «Буря», «Эйнштейн», «Интернет Игл» и др. применяется система управления базами данных Prosecutor's Management Information System (PROMIS), разработанная в Inslaw Inc. под руководством Билла Гамильтона (В. Hamilton). Более 30 лет программа, имеющая в начале 570000 строк программного кода, непрерывно совершенствуется собственными разработчиками АНБ.

Программа PROMIS способна одновременно интегрировать неограниченной объем информации, получаемой от неограниченного количества программ и содержащегося в любом количестве баз данных, независимо от их типов, языков, на которых написаны оригинальные программы, архитектуры операционных систем и платформ, откуда извлекается информация. По-видимому, аналогов PROMIS в мире не существует.

Другим важным направлением является добыча знаний в реальном режиме времени. Сегодня, согласно имеющейся информации, машины могут извлекать знания примерно о 40 млн. сущностей (объекты, субъекты, события и др.) и более чем 2,5 трлн. параметров.

По заказу DARPA компания Raytheon создала самоорганизующуюся базу знаний, которая позволяет автоматически составлять досье на граждан и организации, собирая информацию из откры-

тых источников. С конца 2011 г. головными разработчиками таких баз знаний стали IBM и Recorded Future. Уже сегодня по заказу отдельных родов войск США им удалось создать эффективные системы раннего предупреждения кризисов.

Очень большие надежды связываются с проектом по автоматизированному выявлению аномальных процессов, протекающих в различных масштабах. Источниками информации для программы являются как обычно веб 1 и веб 2, а также анализ потокового видео, финансовых транзакций и т.п.

Большое внимание АНБ и американское разведывательное сообщество уделяли и уделяют проекту по анализу информации и прогнозированию в реальном масштабе времени. Наиболее известной реализованной автоматизированной системой (с участием человека-эксперта) в рамках этого проекта является Palantir, разработанная компанией Palantir Technologies, которая предназначена для анализа и визуализации данных. Система обеспечивает сбор потоков данных со всех доступных информационных каналов обо всех регистрируемых событиях, касающихся людей: банковские транзакции, транзакции по кредитным карточкам, телефонные звонки, электронная почта, информация с камер видеонаблюдения, информация о транзакциях во всех федеральных и муниципальных базах данных и т.п. В каждом потоке данных средствами интеллектуального анализа данных выявляются необычные события, вероятность которых мала, и события из наперед заданных «тревожных списков». Затем происходит объединение и увязка информации о необычных событиях из разных баз данных. И в результате, если расчетная вероятность всего комплекса связанных необычных событий окажется ниже некоего заданного порога вероятности, выдается сигнал тревоги, указывающий на конкретное лицо, с которым связан весь комплекс событий.

Программа Palantir постепенно находит применение и в гражданском секторе. Так, представитель банка JPMorgan Chase Гай Чарелло (Guy Chiarello) говорит, что программы Palantir помогают банку выявлять мошенников еще до того, как случилось преступление. Контракт с Palantir, по его словам, – это «лучшая сделка за последнее время».

При этом основные инвесторы проекта Алекс Карп (Alex Karp) и Питер Тиль (Peter Thiel) говорят, что величайшая проблема, которую им удалось решить с помощью своих программ, – это возможность борьбы с терроризмом и насилием при сохранении гражданских свобод. Задействованная в их работе информация засекречена, и пользователи имеют доступ лишь к отдельным ее фрагментам.

## 2.6. Примеры применения технологий конкурентной разведки

Для качественного проведения конкурентной разведки методами анализа текстов из сети Интернет необходимо сформулировать цели, построить базы данных для наблюдений и проведения исследований, сформулировать запросы. Заметим, что не следует ограничиваться одной информационно-поисковой системой даже для анализа такой информации, как интернет-ресурсы. Рекомендуем использовать лучшие глобальные и специальные информационно-поисковые системы, такие как Google ([www.google.com](http://www.google.com)), Yahoo! ([www.yahoo.com](http://www.yahoo.com)) или Яндекс ([www.yandex.ru](http://www.yandex.ru)). Для специальных потребностей рекомендуется также использовать законодательные, адресно-номенклатурные, ценовые базы данных, доступные как из сети Интернет, так и в локальных версиях.

Покажем, как формируются запросы, относящиеся к конкурентной проблематике, на примере поисковых предписаний к системе контент-мониторинга InfoStream ([www.infostream.ua](http://www.infostream.ua), рис. 23).

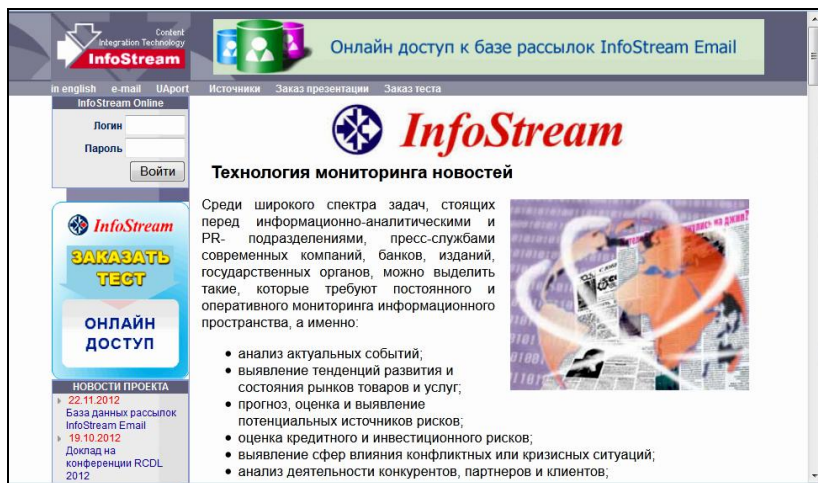


Рис. 23 – Веб-сайт системы контент-мониторинга InfoStream

Обычно поиск информации о компании или персоне всегда начинается с указания различных способов написания названия компании или Ф.И.О. персоны. Порой поиска в оперативных и ретроспективных данных по таким «примитивным» запросам вполне



достаточно, однако задача усложняется, если необходимо исследовать состояние отдельной отрасли, отдельного региона, или даже целой страны. В таких случаях в соответствии с проблематикой строятся запросы, которые затем итеративно уточняются.

В качестве примера приведем ряд понятий, а затем поставим им в соответствие фрагменты запросов и рассмотрим фрагменты текстов, публикуемые различными источниками, которые затем можно использовать при построении разного рода аналитических справок.

После нахождения документов, содержащих упоминания анализируемых фирм или брендов, можно путем уточнения запросов выявить некоторые важные характеристики, относящиеся к деятельности этих компаний. В качестве примера ниже приведены уточняющие запросы, относящиеся к финансовому положению компаний, упоминаемых в веб-пространстве:

Уставн~капитал~/2/грн  
Уставн~капитал~/2/руб  
Уставн~фонд~/2/руб  
Уставн~фонд~/2/долл  
Принадлежит~/2/акций

В первых двух запросах подразумевается нахождение документов, в которые входят фрагменты, содержащие словосочетания «уставной капитал» или «уставный фонд», с указанием значения в долларах или гривнях («~/2/»; на языке запросов это означает состояние в 2 или менее слов между выражениями).

В результате поиска получены текстовые документы, содержащие такие фрагменты:

По данным «ВГЗ», со 2 октября 2008 года ООО «Ремтекс» заключило 24 договора на общую сумму 60,73 млн. грн. Старт значительного роста доходов предприятия зафиксирован в начале 2012 года.

ООО «Ремтекс» основано депутатом Криворожского райсовета Юрием Винклером, Марком Поповым и Инной Рожковой. **Уставной капитал предприятия 12.000 грн.** Совладельцем «Криворожспецремонта» является Алексей Рожков, зарегистрированный с вышеуказанной совладельцем «Ремтекс» Инной Рожковой по одному адресу. Кроме того в состав учредителей фирмы входит Евгений Донченко – депутат Криворожского райсовета.

*Еженедельник «Вгору» 2013.06.19*

ЗАО «Михайлов и Партнеры. Управление стратегическими коммуникациями» зарегистрирована в Москве в 2002 году. Компания, получающая многомиллионные контракты, имеет **уставной капитал 25 тысяч рублей**. 75% акций принадлежат медиа-менеджеру Сергею Михайлову, остальные – пяти партнерам. Компания «Михайлов и партнеры» позиционирует своих сотрудников как специалистов в спектре внешних и внутренних коммуникаций, помогающих развивать своим клиентам отношения со СМИ и общественностью.

*PrimaMedia 2013.06.20*

Совместное предприятие «Илим Палп Энтерпрайз» (СП ТОО «Pim Pulp Enterprise», IPE) было зарегистрировано Комитетом по внешним связям мэрии Санкт-Петербурга 30 апреля 1992 г. (реестровый No АОЛ-1546) с **уставным фондом 1 млн. руб.** Учредителями являлись: ТОО «Техноферм» – 50% уставного фонда, фирма INTERTSEZ S.A. (Швейцария) – 40%, целлюлозный завод ТПО «Усть- Илимский ЛПК» – 10%. Позднее в число учредителей вошло АОЗТ «Финцелл», имея долю в 40%. Поскольку акционерный пакет Медведева в «Финцелле» составлял 50%, то, соответственно, по состоянию на 1994 г. ему лично принадлежало не менее 20% «Илим Палп Энтерпрайз».

*Вопросик.net 2013.04.26*

ИМС в 2004 году создали СКМ Ахметова и «Интерпайп» Пинчука. Летом того же года компания выиграла конкурс по приватизации «Криворожстали». За 93,02% акций комбината консорциум заплатил около 800 млн долларов. В 2005 году хозяйственные суды всех инстанций по представлению Генпрокуратуры отменили итоги конкурса. Осенью 2005-го «Криворожсталь» на повторном приватизационном аукционе досталась Mittal Steel индийского миллиардера Лакшми Миттала за 4,8 млрд долл. (позже комбинат был переименован в «АрселорМиттал Кривой Рог»). На 31 декабря 2012 года «Метинвесту» принадлежали 49% ИМС, говорится в отчетности группы Ахметова. Активы компании – 49 млн долл., **уставный фонд – 15 млн долл.**

*Главред 2013.06.06*

ОАО «Газпром нефть» (ранее «Сибнефть») – дочернее нефтедобывающее предприятие ОАО «Газпром», которому **принадлежит свыше 95 % акций** компании. Уставный капитал «Газпром нефти» равен 7,59 млн. руб. и разделен на 4,74 млрд. обик-

новенных акций номиналом 0,0016 руб.

«Газпрому нефти» в России принадлежат Омский НПЗ, доля в Московском НПЗ (38,8 %), 50 % акций «Славнефти» (контролируется на паритетных условиях с ТНК-ВР), которой, в свою очередь, принадлежат 42,5 % акций Мозырского НПЗ (Беларусь), а также 56,15 % акций сербской вертикально интегрированной нефтяной компании Nafta Industrija Srbije (NIS), которая владеет двумя НПЗ, заводом по производству сжиженного газа, 480 АЗС и нефтебазами.

*OilNews 2013.06.26*

Информация о слияниях и приобретениях в той или иной сфере бизнеса, позволяющая следить за экспансией конкурентов в новые рыночные ниши, может быть получена в результате отработки таких уточняющих запросов:

Приобр~/2/акций  
приобр~/2/пакет~акций  
прода~/2/пакет~акций  
(слияние~компаний) & (акций,активов)

Выполнение этих уточняющих запросов позволяет получить документы, содержащие, например:

В 2009-м бизнес-консультант Алексей Тимофеев набрался смелости и позвонил Новинскому с предложением создать в его холдинге нефтегазовое подразделение. Решимость Тимофеева принесла плоды. Новинский взял его на работу и за последние два года инвестировал в нефтегазовый сектор около \$300 млн. Smart Energy **приобрела 54 % акций** британской Regal Petroleum, ведущей бизнес в Украине, и компанию «Пром-Энерго Продукт», добывающую газ в Харьковской области. «Смарт» закрывает сделку с бизнесменом Игорем Вороновым по приобретению предприятия «Укргазвидобуток» – одного из крупнейших частных добытчиков газа в Украине.

*Журнал «Forbes Украина» 2013.06.29*

С октября 2006 года, напомним, председателем совета директоров ОАО АВТОВАЗ был глава госкорпорации «Ростехнологии» (ныне – «Ростех») Сергей Чемезов.

Напомним, компания Renault **приобрела блокирующий пакет акций** ОАО АВТОВАЗ у ГК «Ростехнологии» в конце 2007 года. В настоящее время держателем контрольного пакета акций

российского автогиганта является альянс Renault- Nissan. «Ростеху» принадлежит блокирующий пакет (25 % + 1 акция).

*Финам.info 2013.06.28*

Британский разработчик программного обеспечения Autonomy, принадлежащий Hewlett-Packard, заявил о **продаже оставшегося пакета акций** поисковой системы Blinkx, специализирующейся на поиске видеофайлов, за 54,84 млн фунтов стерлингов (\$84,48 млн).

Сообщается, что компания разместила 45,7 млн. акций Blinkx по цене 120 пенсов за каждую, что на 5% ниже последней цены закрытия бумаг.

*Финам.ru 2013.06.26*

Американский оператор сети медицинских клиник Tenet Healthcare Corp. покупает Vanguard Health Systems Inc. за \$1,8 млрд. денежными средствами.

Tenet заплатит \$21 за **акцию** Vanguard, что на 70 % выше стоимости бумаг компании на закрытие рынка в пятницу. Помимо этого, оператор сети клиник возьмет на себя долг Vanguard в размере \$2,5 млрд., с учетом которого сумма сделки достигнет \$4,3 млрд., сообщает «Финмаркет».

Ожидается, что приобретение позитивно отразится на прибыли Tenet уже в первый год после завершения сделки. Синергетический эффект от **слияния компаний** оценивается в \$100-200 млн. в год.

*Росбалт 2013.06.24*

Для выявления публикаций об изменении финансового состояния и банкротства можно использовать такие уточняющие запросы:

выпуск~/2/акц

(увелич~уставн)&(фонд,капитал)

повыс~/1/дол~/2/акц

объяв~/2/банкротств

Обработка подобных запросов позволила найти такие документы:

Глава ОАО «Интер РАО ЕЭС» Борис Ковальчук сообщил журналистам, что энергохолдинг не исключает продажи принадлежащих ему пакетов акций ТГК-6 и ТГК-7, входящих в «КЭС-Холдинг» в том случае, если компания не получит разрешения на

**выпуск дополнительного выпуска акций** в пользу «Роснефтегаза». Вырученные в результате продажи средства будут направлены на покрытие дефицита инвестиционной программы «Интер РАО ЕЭС».

По оценкам «Интер РАО ЕЭС», вышеупомянутые пакеты стоят на сегодняшний день порядка 900 млн. долларов. Их покупателем может выступить как сам «КЭС-Холдинг», так и стратегический инвестор. По словам Ковальчука, претендентов на роль стратегического инвестора имеется несколько, однако кто это именно, глава «Интер РАО ЕЭС» уточнять не стал.

*MAonline.ru 2013.06.28*

Как сообщалось ранее, в начале мая текущего года акционеры «Россетей» в ходе внеочередного общего собрания приняли решение об увеличении уставного капитала компании путем размещения дополнительного выпуска обыкновенных акций объемом 161 млрд 066 млн 440 тыс 775 штук номинальной стоимостью 1 рубль каждая.

Совет директоров «Россетей» утвердил цену размещения дополнительной эмиссии акций компании в размере 2,79 рубля за акцию допвыпуска.

*Нефть России 2013.06.28*

Выручка в сегменте ИТ-услуг составила 483,8 млн. долл., что на 1,5 % выше показателей предыдущего года. После корректировки выручки в сегменте путем исключения из расчетов показателей ООО «Группа Борлас», увеличивших выручку за соответствующий период на 31,2 млн. долл. (в феврале 2012 года группа обменяла 100 % акций «Группы Борлас» на 14,09 % акций дочерней компании группы – IBS IT Services, **повысив свою долю фактического владения акциями** компании до 97,02 %), рост выручки сегмента без учета показателей проданных активов за финансовый год, составил 8,6 % в долларовом выражении и 11,1 % в рублевом выражении.

Выручка в сегменте разработки программного обеспечения составила 314,6 млн. долл., что на 16,0 % выше показателей предыдущего года.

*Открытые системы 2013.06.03*

Всего около немецких 40 изданий в настоящее время частично или полностью используют платный доступ к журналистским материалам на своих сайтах. Эксперты отрасли считают, что ре-

шение Bild приведет к тому, что за ней последуют многие другие. Толчком к этому может стать угроза закрытия. Напомним, что недавно была закрыта газета Financial Times Deutschland, а Frankfurter Rundschau и агентство DAPD **объявили о своем банкротстве.**

МедиаБизнес 2013.06.12

Методы контент-мониторинга – это адаптация классических методов контент-анализа к условиям динамических информационных массивов, например потоков информации из сети Интернет.

Типичная задача контент-мониторинга – построение диаграмм динамики появления понятий по времени. Рассмотрим, как в системе InfoStream отслеживались кризисные явления на рынке нефтепродуктов Украины в июне 2011 года. Для этого был составлен запрос «**кризис & гречка & Украина**», который был введен через веб-интерфейс системы. В специальном режиме «Динамика» была получена соответствующая диаграмма появления понятия (рис. 24).



Рис. 24 – Динамика появления понятия

На приведенной диаграмме видно, что массовое появление сообщений о кризисных явлениях произошло 12-го января 2011 г. (в то время, как сами цены на гречневую крупу резко возросли лишь в середине марта). Безусловно, оперативное получение такого типа данных должно было помочь аналитикам при построении краткосрочных прогнозов.

Аналогично можно проводить мониторинг финансового рынка. К примеру, простой запрос «падение~ курса~евро», относящийся к фрагменту информационного потока за период с апреля по

август 2010 года выдал диаграмму, свидетельствующую о динамике падения курса европейской валюты (рис. 25). Как видим, пик публикаций приходится на 17 мая 2010 г., когда, в частности, украинские банки ограничили операции с евро из-за падения курса.

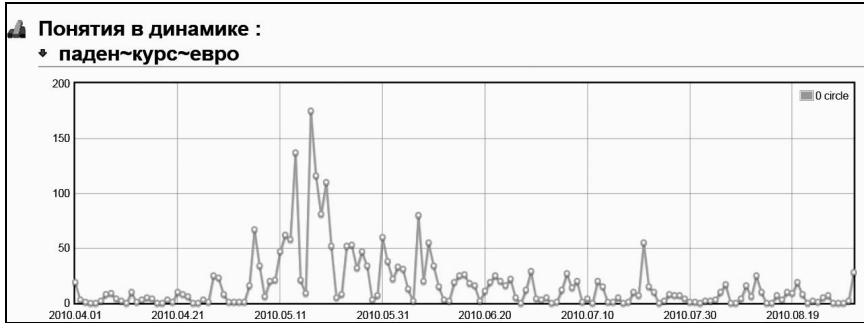


Рис. 25 – «Падение курса евро» в динамике

### 3. Источники информации

---

---

В информационно-аналитической работе важное значение имеет возможность доступа к источникам данных, информации и знаний. При этом главной проблемой является нахождение содержательных и надежных источников из всех общедоступных. Когда такие источники найдены, включаются механизмы превращения данных в знания, для чего применяются соответствующие технологии. Под данными обычно понимают «сырые», необработанные сведения, основанные на фактах. Это могут быть статистические данные, факты из биографий ключевых персон или, например, сведения об отчетности отдельных компаний. Информация представляет собой уже определенным образом обработанные и проанализированные данные. Конечным же информационным продуктом любой аналитической работы являются знания – синтезированные выводы, рекомендации для принятия решений.

Информация, как было указано выше, может быть получена из официальных, открытых источников, СМИ, объявлений, рекламы, фирменных, банковских, правительственных отчетов, баз данных, от экспертов путем анализа или специальной обработки данных, текстов.

Ниже приведен подробный перечень видов информационных источников, которые чаще всего используются при конкурентной разведке [Нежданов, 2009].

1. Пресс-релизы компаний, официальные заявления от имени компаний о новых технологиях, новых направлениях, сделках, перспективах. Такие пресс-релизы создаются компаниями для собственной популяризации, привлечения внимания потенциальных клиентов, инвесторов, ищущих выгодные варианты вложения своих средств. Часто в таких заявлениях присутствует информация о намерениях, планируемых событиях. Пресс-релизы доступны на веб-сайтах компаний, в PR-службах, на общих и профильных специализированных площадках для размещения пресс-релизов.

2. Интервью сотрудников компаний, соответствующие материалы в СМИ. В интервью интерес представляют планы компаний. При этом со стороны службы конкурентной разведки допускается инициирование интервью кого-то из сотрудников объекта интереса.

3. Высказывания сотрудников компаний на форумах, в блогах, в частных беседах. При этом могут выявляться планы компаний, кадровая политика, атмосфера в коллективе и т. п. Источники информации: 1) интернет-ресурсы (специализированные форумы,



блоги сотрудников), блоги экспертов, группы в социальных сетях; 2) выставки, конференции, курсы повышения квалификации, профессиональные мероприятия.

4. Тендеры, закупки. Предметы закупок, оборудование, исполнители. Источники информации: 1) интернет-ресурсы (веб-сайты компаний, торговые площадки, профильные форумы); 2) партнеры исследуемой компании, те, кто участвовал в их тендерах, у клиентов и поставщиков.

5. Патенты, авторские свидетельства компании и ее сотрудников. Для задач конкурентной разведки интересно их содержание, направленность, списки соавторов. Информация размещается на соответствующих сайтах. Для России: [www.ru-patent.info](http://www.ru-patent.info), для Украины: [www.base.ukrpatent.org/searchINV](http://www.base.ukrpatent.org/searchINV), для СНГ [www.eapo.org/ru](http://www.eapo.org/ru). Патентование возможно в любой стране, предпочтительные варианты – страна регистрации организации, страна ведения бизнеса, кроме того США, Евросоюз, Россия, Япония и Китай.

6. Разработки компании: ведущиеся, финансируемые, разработки, которыми компания интересуется. Наблюдению подлежат попытки компании проводить исследования: закупка специфического оборудования, прием на работу специалистов, переговоры, посещения соответствующих организаций и т.д.

7. Активность компании на рынке слияний и поглощений (M&A). Информация о том, какие организации поглощаются, планируют поглотить или ведут переговоры о поглощении. Информацию можно получить в Федеральной антимонопольной службе (ФАС) России или Антимонопольном комитете (АМК) Украины, по новостным сообщениям на веб-ресурсах посвященных M&A.

8. Вакансии компании (открывающиеся, закрывающиеся), сообщения об активном поиске сотрудников, требования к вакансиям, условия. Источник информации: веб-сайт компании, сайты по поиску работы и на сайты агентств, с которыми компания сотрудничает.

9. Курсы повышения квалификации, обучение персонала – указание на приоритеты в развитии компании. Интерес представляет то, чему обучают, каких специалистов приглашают для обучения, какие требования выдвигают при привлечении обучающихся, какие сроки обучения, какое количество персонала обучается.

10. Благодарности и награды компании и ее сотрудников.

11. Участие в мероприятиях (выставки, конференции, круглые столы, презентации). Выяснение, в каких мероприятиях участвуют компании, их направленность, круг участников.

12. Участие в организациях (союзы, ассоциации, конфедерации и т.п.) – информация о том, в каких объединениях участвует компания, как активно участвует, что получает от участия, на что рассчитывает, как использует.

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

Деятельность конкурентной разведки основана на использовании только легитимных источников информации, которых вполне достаточно для принятия управленческих решений в сфере бизнеса, необходимо лишь провести некоторую информационно-аналитическую обработку имеющихся открытых данных. Среди таких источников информации можно назвать: данные статистики, материалы с веб-сайтов, социальных сетей, СМИ, отраслевых отчетов и т.д.

Многие службы конкурентной разведки не всегда могут отделить нелегитимную часть информации от легальной, а заказчик, как правило, интересуется конечными результатами, источники для него выступают лишь в качестве подтверждений, промежуточных данных. Вместе с тем, солидные заказчики сами заинтересованы в том, чтобы информация добывалась законными средствами, чтобы аналитический отчет был легален.

У конкурентной разведки в последние десятилетия появился и развился до невиданных ранее масштабов новый информационный источник – веб-пространство сети Интернет. Сегодня по оценкам экспертов Интернет по количеству информации находится на первом месте, опережая СМИ, отраслевые издания и получаемые от коллег новости, специальные обзоры, закрытые базы данных. При этом в открытых источниках и специализированных базах данных, доступных в Интернет, содержится большая часть информации, необходимой для проведения конкурентной разведки, однако остается открытым вопрос ее нахождения и эффективного использования. Последние исследования информационного веб-пространства показали, что доступный через традиционные информационно-поисковые системы триллион веб-страниц – это лишь «поверхностная видимая часть айсберга». Около 40 % всей информации в Интернете доступно бесплатно. Навигацию по данному информа-

ционному пространству обеспечивают более миллиона поисковых систем и каталогов, но и они охватывают лишь малую часть информационных ресурсов. Скрытых и невидимых (deep, invisible) ресурсов сети Интернет значительно больше – это, прежде всего динамически-генерируемые страницы, файлы разнообразных форматов, информация из многочисленных баз данных. К «скрытому» веб можно отнести и такие сети, как BitTorrent, DirectConnect, EMule, Napster и др.

Сегодня для конкурентной разведки основными источниками информации служат Интернет, пресса, а также открытые базы данных. Очень популярны среди специалистов по конкурентной разведке базы данных государственных и статистических органов, торгово-промышленных палат, органов приватизации и т.д. Большую пользу приносят и отдельные доступные базы данных других органов власти. В последнее время все более популярны базы данных на основе архивов СМИ, в том числе и сетевых. В России, например, большой популярностью пользуется крупнейшая архивная база данных СМИ службы «Интегрум» (integrum.ru), содержащая несколько сотен миллионов документов. С помощью другой российской базы данных «Лабиринт» (labyrinth.ru), составленной на основе публикаций ведущих бизнес-изданий, можно получить обширную информацию о конкретных персонах, организациях и компаниях.

Традиционно конкурентная разведка опирается на следующие источники информации, как опубликованные документы открытого доступа, которые содержат обзоры товарного рынка, информацию о новых технологиях, создании партнерств, слияниях и приобретениях, объявлениях о рабочих вакансиях, о выставках и конференциях, и т.п. Широко используются сведения, находящиеся в документах, уже имеющихся в компаниях, ведущих конкурентную разведку, результаты маркетинговых исследований, информация, полученная на конференциях, при общении с клиентами и коллегами. Большая часть этих данных попадает в сетевую прессу, пресс-релизы или публикуются на корпоративных веб-сайтах.

Поэтому в последнее время большую популярность получают базы данных на основе архивов масс-медиа, в том числе (и преимущественно) сетевых.

### **3.1. Веб-пространство**

Веб-пространство, основанное на физической инфраструктуре сети Интернет и протоколе передачи данных HTTP, объединяет

сотни миллионов веб-серверов, подключенных к сети Интернет (рис. 26). В начале существования веб-пространства на небольшом количестве веб-сайтов публиковалась информация отдельных авторов для относительно большого количества посетителей. Сегодня ситуация резко изменилась, произошел переход к веб второго поколения. Сами посетители веб-сайтов активно участвуют в создании контента, что привело к резкому росту объемов информации и динамики веб.

Сегодня в веб уже существует свободно доступная для пользователей информационная база такого объема, который ранее трудно было представить. Более того, объемы этой базы превышают на порядки все то, что было доступно десятилетие назад. В августе 2005 года компания Yahoo! объявила о том, что проиндексировала около 20 млрд. документов. Достижение компании Google в 2004 году составляло менее 10 млрд. документов. Сегодня Google заиндексировала свыше триллиона веб-документов. По данным службы Netcraft Web Server Survey ([news.netcraft.com](http://news.netcraft.com)), в настоящее время количество веб-серверов превышает 670 млн.

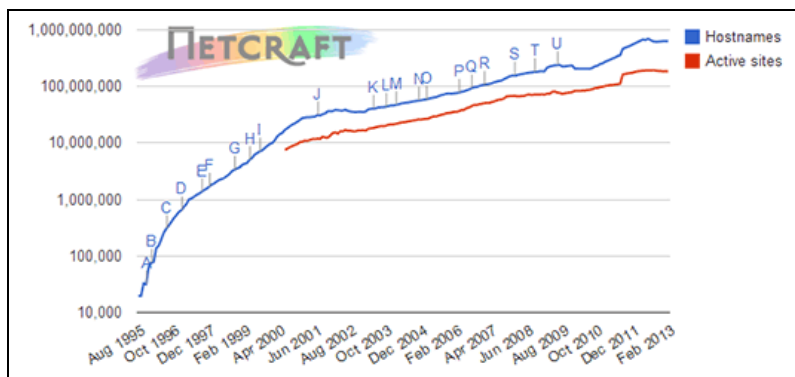


Рис. 26 – Динамика роста количества веб-серверов по логарифмической шкале (Netcraft, январь 2013 года)

В открытых источниках и специализированных базах данных, доступных в веб-пространстве, содержится большая часть информации, необходимой для проведения аналитических исследований, однако остаются открытыми вопросы ее нахождения и эффективного использования. При использовании веб-пространства как мощнейшего источника информации, как уже было отмечено ранее, самыми существенными являются проблемы объема, навига-

ции, наличия информационного шума и динамического характера информации в Интернет.

Возможности доступа к интернет-ресурсам, привлекающим своей открытостью, объемами и содержательной многогранностью, на первый взгляд кажутся безграничными. Однако важные события в различных областях свидетельствуют об обратном. Именно в кризисных ситуациях Интернет довольно часто подводит. Существует множество проблем – от перегруженности сетевой инфраструктуры – до вирусных атак, уязвимостей и отказов в обслуживании отдельных веб-серверов. Целый ряд проблем порожден также объемами, разнообразием представления и динамикой контентного сегмента информационного пространства.

Несмотря на такие качества, как открытость и доступность, существующую инфраструктуру веб-пространства нельзя признать надежной и достоверной. Назовем еще несколько проблем, присущих веб-пространству:

- не решена задача доступа пользователей к разнородным веб-ресурсам из «одного окна» для получения обобщенного представления потоков информации по необходимой тематике;
- не обеспечена возможность своевременного «напоминания» и «проталкивания» профильной для пользователя информации, публикуемой на большом количестве веб-сайтов;
- достаточно большая вероятность отказа в обслуживании критически важных веб-ресурсов в самое необходимое время.

Известно, что сегодня существуют технологии интеграции контента, которые позволяют частично решать названные проблемы, обеспечивая эффективный поиск и навигацию в веб-пространстве, мониторинг и агрегацию открытых веб-ресурсов.

Для профессионального поиска в веб-пространстве и мониторинга информации используются специализированное программное обеспечение, информационно-поисковые системы и сервисы. Приведем некоторые примеры программных продуктов:

**Copernic Agent** ([www.copernic.com/en/products/agent](http://www.copernic.com/en/products/agent)) – программа, позволяющая проводить метапоиск, используя, как заявлено на веб-сайте компании, 1000 поисковых систем, объединять результаты, устранять дубликаты, блокировать нерабочие ссылки, показывать наиболее релевантные результаты (рис. 27).

**Avalanche** ([www.tora-centre.ru](http://www.tora-centre.ru)) – семейство программных средств для веб-мониторинга. Технология Avalanche базируется на трех основных решениях: концепции «умных папок» (Smart Folders), автономном интеллектуальном поисковом роботе и встроенной базе данных («персональной энциклопедии»).



Рис. 27 – Фрагмент веб-сайта программы Copernic Agent

**Newprosoft Web Content Extractor** ([www.newprosoft.com](http://www.newprosoft.com)) – программа сканирования и извлечения данных из веб-сайтов.

**Portable Offline Browser** от MetaProducts Corporation ([www.portableofflinebrowser.com](http://www.portableofflinebrowser.com)) – программа, позволяющая скачивать необходимые веб-сайты и мультимедиа-информацию, в том числе Flash-анимацию, скрипты и активное содержимое страниц.

**Neiron Search Tools** ([neiron.ru/toolbar](http://neiron.ru/toolbar)) – программная надстройка, объединяющая результаты информационно-поисковых систем Google и Yandex, которая позволяет осуществлять конкурентный анализ, базирующийся на оценке эффективности сайтов и контекстной рекламы.

**WebSite-Watcher** ([www.aignes.com](http://www.aignes.com)) – программа, позволяющая проводить мониторинг веб-сайтов, форумов, локальных файлов, обеспечивающая фильтрацию информации, а также удобную визуализацию результатов мониторинга.

В качестве сервисных решений можно назвать:

**WatchThatPage** ([watchthatpage.com](http://watchthatpage.com)) – бесплатный сервис, позволяющий автоматически собирать новую информацию с веб-ресурсов, поставленных на мониторинг.

**Diphur Monitor Everything** ([www.diphur.com](http://www.diphur.com)) – бесплатный сервис мониторинга любых веб-сайтов, уведомляющий об их обновлении и доставляющий пользователям обновления.

**Newspaper Map** ([newspapermap.com](http://newspapermap.com)) – сервис, объединяющий геолокацию и информационно-поисковую систему по медиа-

ресурсам. При решении задач конкурентной разведки пользователь может выбрать интересующий его регион, язык, список онлайн версий газет и журналов, непосредственно выходить на документы. Сервис поддерживает русский язык, имеет удобный интерфейс.

**InfoStream** ([www.infostream.ua](http://www.infostream.ua)) – сервис контент-мониторинга веб-ресурсов России и Украины, предоставляющий доступ в поисковом режиме к информации из 6000 источников, классификацию информации, экстрагирование понятий (персон, компаний, топонимов), формирование сюжетных цепочек, оценку тональности сообщений, анализ динамики публикаций по определенным объектам.

**Agregator.pro** ([agregator.pro](http://agregator.pro)) – агрегатор информации с новостных и медийных порталов. Может использоваться в конкурентной разведке для отслеживания интересующих объектов, получения частоты и контекста упоминания отслеживаемого объекта в СМИ, анализа динамики обращений по времени.

**WebGround** ([webground.su](http://webground.su)) – агрегатор новостной информации из русскоязычного сегмента веб-пространства. Может использоваться в конкурентной разведке для отслеживания интересующих тематик, получения тематических сюжетов, ретроспективного анализа развития тематики во времени (рис. 28).

WebGround  
Проект мониторинга Рунета

Рубрики Сюжеты Источники Ваша тема

Мониторинг Интернета: \* Только значимая информация в реальном времени \* Самый большой объем потока новостей \* Обновлено 18.08.2013 18:00 Мск

**Новые сюжеты**

РПЦ заявила о массовых расстрелах в подтопленных регионах  
Алексей Навальный отказался от дебатов  
В Украине ограблен автобус с китаянками и вьетнамками  
Полница Таиланда изъела почти миллион таблеток метамфетамина стоимостью более \$8 млн  
Опоздали: четверо из шести пилотов выжили в Дагестане  
Британка вернула своих дипломатов в Иран  
Илиана обещившего угнать самолет девочку в окрестностях Корана, освободили  
Пав Ги Муа выразил обеспокоенность по поводу пилотской ситуации в Египте  
Саргсян: Армяне выиграли победу Алиева на выборах в Азербайджане

**Затонувший на Иртыше теплоход поднят со дна реки**  
Затонувший в субботу при столкновении с баржей на реке Иртыш теплоход поднят на поверхность, сообщает региональное министерство строительства, транспорта и жилищно-коммунального комплекса Омской области.  
По информации регионального Минстроя, катер уже поднят на баржу, сейчас на судне работают следователи. Угрозы загрязнения водного бассейна топливом нет, - говорится в сообщении, размещенном на сайте министерства.  
После подъема к осмотру судна должны приступить следователи.  
Характер повреждения корабля могут помочь в прояснении обстоятельств и причин трагедии.  
По предварительным данным, в результате столкновения прогнучного катера "Полесье-8" с грузовой баржей погибли шесть пассажиров.  
Всего на борту катера находились 57 человек, в том числе четыре члена экипажа.  
ВВС Русская служба 16:30 [Сюжет \(264\)](#)

**Як-52 потерпел крушение в Якутии в ходе показательных выступлений**  
На месте падения самолета проходил фестиваль "Берег радости". При крушении получили травмы 2 человека: девушка 1995 года рождения и мужчина 1959 года рождения, находившиеся на берегу. Они госпитализированы. Об этом сообщается из сайта регионального ГУ МЧС.

**Темы дня**  
ТЕПЛОХОД БАЖКА  
ОМСКАЯ ОБЛАСТЬ  
ИРТЫШ  
ОМСКИЙ  
ХАБАРОВСК  
АМУРСКАЯ ОБЛАСТЬ  
ИСТОРИЧЕСКИЙ МАКСИМУМ  
ХАБАРОВСКИЙ КРАЙ  
ПАВОДОК  
МЕРТВАЯ ПЕТЛЯ  
ПОКАЗАТЕЛЬНОЕ ВЫСТУПЛЕНИЕ  
ТРЕНИРОВОЧНЫЙ САМОЛЕТ  
ТОМОИТ  
ВЫШШИЙ ПИЛОТАЖ  
ТАМОЖЕННОЕ АДМИНИСТРИРОВАНИЕ  
ТОРГОВЫЕ ВОЙНЫ  
ТАМОЖЕННАЯ СЛУЖБА  
УКРАИНСКИЕ ТОВАРЫ  
РОССИЙСКАЯ ТАМОЖНЯ  
ПАТРИАРШИЙ СОВОР  
УПЦ  
КАТОЛИЧЕСКАЯ ЦЕРКОВЬ  
УКРАИНСКИЕ ГРЕКО  
ДИНА  
ПРИНЦЕССА  
ДОДИ  
НОВАЯ ИНФОРМАЦИЯ

Знайти:  технологии для менеджеров  Наступить  Попретиде  Посетить все  Зарукаваннам регистру  Досягнути кнця сторинки, продовжено з початку

Рис. 28 – Фрагмент агрегатора новостей WebGround

## 3.2. Глубинный веб

Последние исследования веб-пространства показали, что доступные через традиционные информационно-поисковые системы более триллиона веб-страниц – это лишь «поверхностная видимая часть айсберга».

Важной проблемой является поиск информации в «скрытом» или «глубинном» веб-пространстве, где, как было замечено выше, содержится несравнимо большее количество данных, потенциально интересных для конкурентной разведки, чем в открытой части Интернета.

Это, прежде всего, динамические веб-страницы, информация из многочисленных баз данных, которые могут представлять большой интерес для аналитической работы. К разряду «скрытого» веб относятся и полнотекстовые информационные системы типа LexisNexis или Factiva.

К «скрытым» ресурсам сети Интернет можно отнести также пиринговые сети, такие как BitTorrent, EDonkey, EMule, Gnutella, Kazaa.

Как уже было отмечено ранее, необходимой (в том числе и для конкурентной разведки) информации в сети Интернет значительно больше, чем ее охватывают универсальные поисковые машины.

Предполагается, что в отличие от «познаваемой» части сети Интернет, «скрытая» часть оказалась в сотни раз более объемной.

Бизнес-аналитик часто сталкивается с ситуацией, когда ему известно о существовании в веб-пространстве какого-то документа, но не может найти его с помощью традиционных поисковиков, какими сегодня можно считать такие системы, как Google, Yahoo!, Bing, Яндекс, Рамблер или Мета. Однако, вспомнив или найдя в закладках адрес (URL) этого документа, он без труда выходит на него. То есть в веб-пространстве этот документ есть, а найти его привычным способом нельзя. Пользователь столкнулся с невидимым (*invisible*) для поисковых систем ресурсом.

### 3.2.1. Что такое глубинный веб?

Совокупность источников в веб-пространстве, недоступных пользователям традиционных поисковых систем, образует так называемый «глубинный веб» – понятие, введенное Джиллом Иллсвортом (Jill Ellsworth) в 1994 г. Т.е. под глубинным веб (*invisible web*, *deep web*, *hidden web*) принято понимать ту часть веб-пространства, которая не индексируется роботами (*web crawlers*) поисковых систем. Используя аналогию, информация, будучи недоступной для поиска, находится «в глубине» (англ. – *deep*). При этом не стоит пу-



тать глубинный веб с ресурсами, вовсе недоступными из сети Интернет – это темный веб (dark web), и речь о нем здесь идти не будет. Некоторые ресурсы, доступ к которым открыт лишь для зарегистрированных пользователей, также относятся к глубинному веб.

В 2000 году американская компания BrightPlanet ([www.brightplanet.com](http://www.brightplanet.com)) опубликовала сенсационный доклад, в котором утверждается, что в веб-пространстве в сотни раз больше страниц, чем их удалось проиндексировать самыми популярными на то время поисковыми системами. Компания разработала программу LexiBot, которая позволяет сканировать некоторые динамические веб-страницы, формируемые из баз данных, и, запустив ее, получила неожиданные данные. Выяснилось, что в глубинном веб находится в 500 раз больше документов, чем доступно через поисковые системы. Конечно, эти цифры неточны. Кроме того, стало известно, что средняя страница глубинного веб на 27 % компактней средней страницы из видимой части веб-пространства.

Сегодня ситуация изменилась, например, ведущие поисковые системы могут индексировать документы, представленные в форматах, содержащих текст. Конечно, это, прежде всего, pdf, rtf и doc. В 2006 году Google запатентовала способ поиска в глубинном веб: «Searching through content which is accessible through web-based forms» (рис. 29). По мнению разных авторов к видимому веб относится лишь 20–30 % веб-пространства.

The screenshot shows a WIPO IP Services page for patent WO/2006/108069. The page title is "SEARCHING THROUGH CONTENT WHICH IS ACCESSIBLE THROUGH WEB-BASED FORMS". The patent is owned by GOOGLE, INC. and was published on 12.10.2006. The inventors listed are HALEVY, ALON Y., MADHAVAN, JAYANT, and KO, DAVID H. The agent is PARK, A. RICHARD. The patent is in the field of searching through content accessible through web-based forms.

Latest bibliographic data on file with the International Bureau		
<b>Pub. No.:</b>	<b>WO/2006/108069</b>	<b>International Application No.:</b> PCT/US2006/012734
<b>Publication Date:</b>	<b>12.10.2006</b>	<b>International Filing Date:</b> 04.04.2006
<b>IPC:</b>	<b>G06F 17/30</b> (2006.01)	
<b>Applicants:</b>	<b>GOOGLE, INC.</b> , [US/US]; 1600 Amphitheatre Parkway, Bldg. 47, Mountain View, CA 94043 (US) ( <i>All Except US</i> ). <b>HALEVY, Alon Y.</b> , [US/US], (US) ( <i>US Only</i> ). <b>MADHAVAN, Jayant</b> [IN/US], (US) ( <i>US Only</i> ). <b>KO, David H.</b> , [CN/US], (US) ( <i>US Only</i> ).	
<b>Inventors:</b>	<b>HALEVY, Alon Y.</b> , (US). <b>MADHAVAN, Jayant</b> , (US). <b>KO, David H.</b> , (US).	
<b>Agent:</b>	<b>PARK, A. Richard</b> , 2820 Fifth Street, Davis, CA 95616 (US).	
<b>Priority Data:</b>	60/669,292 06.04.2005 US	
<b>Title:</b>	SEARCHING THROUGH CONTENT WHICH IS ACCESSIBLE THROUGH WEB-BASED FORMS	

Рис. 29 – Фрагмент веб-ресурса WIPO с описанием патента Google на поиск в глубинном веб

### 3.2.2. Причины возникновения

В глубинном веб находятся веб-ресурсы, не связанные с остальными ресурсами гиперссылками – например, страницы, динамически создаваемые по запросам к базам данных, документы из баз данных, доступные пользователям через поисковые веб-формы (но не по гиперссылкам). Такие документы остаются недоступными для робота, неспособного в режиме реального времени правильно заполнить поля формы значениями (формировать запросы к базам данных).

Вот что говорится о глубинном веб в книге [Price, 2001]: «Большинство страниц невидимого Интернета могут быть проиндексированы технически, но не индексируются, потому что поисковые системы решили их не индексировать... Большинство «невидимых» сайтов имеют высококачественный контент. Просто эти ресурсы не могут быть найдены с помощью поисковых машин общего назначения...

... Некоторые сайты используют технологию баз данных, что действительно сложно для поисковой машины. Другие сайты, однако, используют сочетание файлов, которые содержат текст и мультимедиа, а поэтому часть из них может быть проиндексирована, а часть – нет.

... Некоторые сайты могут быть проиндексированы поисковыми машинами, но это не делается потому, что поисковые машины считают это непрактичным – например, по причине стоимости или потому, что данные настолько короткоживущие, что индексировать их просто бессмысленно – например, прогноз погоды, точное время прибытия конкретного самолета, совершившего посадку в аэропорту и т.п.»

Основные ограничения, связанные с роботами поисковых машин можно объяснить следующими основными причинами: для публичных поисковых служб важнее обеспечить точность поиска, чем полноту, важнее обеспечить получение ответа на запрос в приемлемое время, чем точность. Отсюда – ограничения на глубину сканирования веб-ресурсов, попытки «фильтрации» контента по содержанию, отсеивание страниц, содержащих излишние выходные гиперссылки и т.п. При этом часто с водой выплескивается и ребенок. Общеизвестно, что ценность ресурсов глубинного веб зачастую выше ценности ресурсов видимой части веб-пространства.

Можно упомянуть еще один источник пополнения глубинного веб – владельцы сознательно не хотят, чтобы их веб-ресурсы находили с помощью поисковых систем. Чаще всего такие веб-ресурсы

представляют нечто не совсем законное, хакерские форумы, архивы неавторизованного контента и т.п. Понятно, что многие из таких ресурсов очень интересны для изучения бизнес-аналитиками.

Многие компании сначала подключаются к общей Сети, и лишь потом тратят большие средства на защиту. Владельцы сайтов могут попытаться запретить индексацию тех или иных страниц своих ресурсов, прописав запрещающую команду в файле robots.txt, но поисковые системы могут ее проигнорировать. Поэтому такие ресурсы либо удаляют, либо удаляют гиперссылки, переводя ресурсы в глубинный веб. Например, недавно бизнес-каталоги когда Auto.ru и Drom.ru отказались отдавать свои объявления «Яндексу», т.е. защищая свои информационные активы компании перевели свои ресурсы в глубинный веб.

### **3.2.3. Виды ресурсов глубинного веб**

Существует несколько типов ресурсов глубинного веб, например, как было отмечено выше, это могут быть быстро устаревающие веб-страницы. Кроме того, к глубинному веб относятся веб-ресурсы, представляющие собой мультимедийную информацию. Как известно, в данное время еще не существует удовлетворительных алгоритмов поиска не текстовой информации. Динамически генерируемые по запросу страницы также часто попадают в глубинный веб. Зачастую без запроса таких страниц не существует, они генерируются при запросе к базам данных. Получается, что информация вроде бы и присутствует в веб-пространстве, но возникает она лишь в момент обработки запроса, а универсального алгоритма заполнения их роботами поисковых форм не существует. И, наконец, если на веб-ресурс не ведут никакие ссылки, то роботы поисковых систем никаким образом не могут узнать об его существовании.

Основатель компании BrightPlanet Майкл Бергман (Michael K. Bergman) смог выделить 12 разновидностей глубинных веб-ресурсов, относящихся к классу онлайн-баз данных. В списке оказались как традиционные базы данных (патенты, медицина и финансы), так и публичные ресурсы – объявления о поиске работы, чаты, библиотеки, справочники. Бергман причислил к глубинным ресурсам и специализированные поисковые системы, которые обслуживают определенные отрасли или рынки, базы данных которых не включаются в глобальные каталоги традиционных поисковых служб.

К глубинному веб также относятся многочисленные системы интерактивного взаимодействия с пользователями – помощи, консультирования, обучения, требующие участия людей для формирования динамических ответов от серверов. К ним также можно отнести и закрытую (полностью или частично) информацию, доступную, пользователям Сети только с определенных адресов, групп адресов, иногда городов или стран. К «скрытой» части Сети многие причисляют и веб-страницы, зарегистрированные на бесплатных серверах, которые индексируются, в лучшем случае, лишь частично – поисковые системы во избежание рекламного спама не стремятся обходить их в полном объеме.

К глубинному веб также относится категория так называемых «серых» сайтов, функционирующих на основе динамических систем управления контентом (Dynamic Content Management Systems). В поисковых системах обычно ограничивается глубина индексирования таких сайтов во избежание возможного циклического просмотра одних и тех же страниц.

### **3.2.4. Примеры ресурсов глубинного веб**

Как же найти веб-ресурсы, размещенные в глубинном веб? Если ресурсы требуют заполнения специальных форм, дополненных, например, капчами, то необходимо выйти на базу данных, предположительно содержащую необходимые документы. Найти базы данных – источники скрытого веб можно с помощью обычных поисковых систем, обобщив запрос и введя уточняющие слова, такие как «база данных», «банк данных», «database» и т.п.

Приведем общеизвестный пример: пользователю требуется статистика по катастрофам самолетов в Аргентине. Естественный запрос к традиционной поисковой системе выдает огромный список газетных заголовков. На запрос «aviation database», можно сразу выйти на базу данных NTSB Aviation Accident Database ([www.nts.gov/ntsb/query.asp](http://www.nts.gov/ntsb/query.asp)).

Для поиска в глубинном веб, а именно в том его сегменте, который составляют базы данных, сегодня уже существуют некоторые специализированные ресурсы. Лидером среди навигаторов в глубинном веб является сайт CompletePlanet ([www.completeplanet.com](http://www.completeplanet.com)) компании BrightPlanet. Этот сайт является крупнейшим каталогом, насчитывающим свыше 100 тысяч ссылок. Компания BrightPlanet также создала персональную утилиту для поиска в онлайн-базах данных LexiBot, которая может обеспечивать поиск в нескольких тысячах поисковых систем «глубинного» веб. Метапоис-

ковый пакет DeerQueryManager (DQM) этой же компании обеспечивает поиск более чем по 70 тысячам «скрытым» веб-ресурсам.

Исследование, проведенное еще в 2006 г. [He, 2007] показало, что глубинный веб охватывает более 300 тыс. сайтов, связанных с более 450 тыс. базами данных, не охватываемых традиционными поисковыми системами. К наиболее интересным для бизнес-аналитиков ресурсам глубинного веб относятся: базы данных юридических и физических лиц; отраслевые базы данных; репутационные базы данных (черные и белые списки); криминологические базы данных; базы данных товаров и услуг; каталоги продукции и т.п. К всемирно известным бизнес-ресурсам, размещенным в глубинном веб, относятся: amazon.com, ebay.com, realtor.com, cars.com, imdb.com.

Приведем еще несколько примеров баз данных и каталогов глубинного веб:

**FindLaw** (www.findlaw.com) – один из наиболее популярных в мире юридических веб-сайтов – огромный каталог правовых ресурсов, содержащий аннотированный список свободно доступных баз данных нормативно-правовых документов, для которых данный ресурс является «точкой входа». Фрагмент веб-сайта сервиса FindLaw приведен на рис. 30.



Рис. 30 – Фрагмент веб-сайта сервиса FindLaw

**About.com** (www.about.com) – портал, охватывающий тысячи, снабженных комментариями, ссылок на веб-ресурсы, в том числе и

на ресурсы глубинного веб (имеется ссылка «Invisible Web»). На портале предоставляется возможность поиска в каталоге. Ресурс также включает несколько статей по проблематике глубинного веб: «What is the Invisible Web?», «Finding the Invisible Web», «Top Places to Search the Invisible Web» и др.

**Politicalinformation.com** ([www.politicalinformation.com](http://www.politicalinformation.com)) – сервис, обеспечивающий оперативный поиск в 5000 отобранных веб-сайтах политической направленности, предоставление новостей из нескольких десятков авторитетных источников.

**Infomine** ([infomine.ucr.edu](http://infomine.ucr.edu)) – сервис обеспечивает добычу информации из баз данных, электронных журналов (блогов), электронных досок объявлений, электронных книг, списков рассылок, электронных каталогов и т.п., преимущественно познавательно-образовательного характера. Обеспечивает как общий поиск, так и поиск по тематическим категориям.

Особенность большинства «скрытых» ресурсов заключается в их узкой специализации. Для поиска в них используются те же механизмы, что и для «поверхностного» веб, однако, в большинстве случаев, роботы поисковых систем для глубинного веб включают уникальные для каждого такого ресурса модули доступа к данным.

Традиционная поисковая система чаще всего может выдать адрес базы данных, но не скажет, какие документы конкретно содержатся в ней. Типичный пример – информационно-поисковые системы по украинскому ([zakon.rada.gov.ua](http://zakon.rada.gov.ua)) или российскому законодательству ([www.kodeks.ru](http://www.kodeks.ru)). Тысячи документов из баз данных становятся доступны только после входа в систему, а роботы стандартных поисковых систем не в состоянии заиндексировать контент баз данных.

Парадоксально, но в качестве одного из ресурсов глубинного веб можно рассматривать и архив ресурсов открытого веб-пространства. Такой архив – Internet Archive с 1996 года создает компания Alexa ([www.archive.org](http://www.archive.org)). Сегодня объем базы данных Alexa превышает 350 млрд. веб-страниц (рис. 31).

Технология хранилища Alexa включает ряд современных средств управления гигантским документальным хранилищем. Например, с помощью технологии Alexa выполняется кластеризация веб-ресурсов, т.е. формирование коллекций документов, близких по тематикам. Особый интерес у пользователей сервиса Alexa вызывает «Машина времени» (Wayback Machine), открывающая доступ к временным срезам веб-пространства. Одно из наиболее интересных практических применений этой технологии – восстановление документов, некогда опубликованных в веб-пространстве,

но впоследствии удаленных. При этом рост глубинного веб грозит серьезными проблемами полноты в хранилище системы, связанными с увеличивающимся количеством сайтов, эксплуатирующих различные технологии управления контентом, динамической публикацией документов из баз данных и т.п.

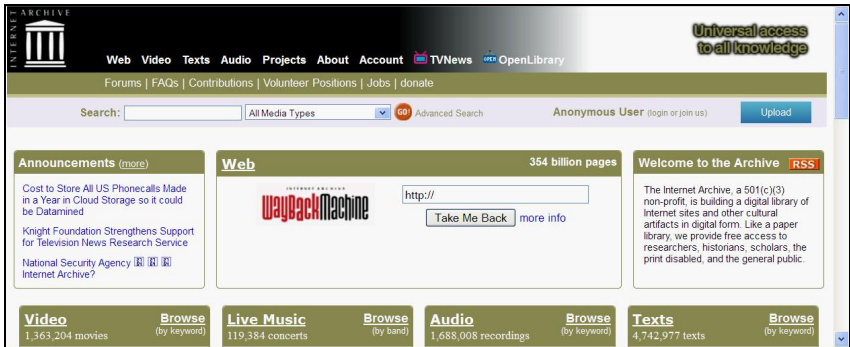


Рис. 31 – Заглавная страница веб-сайта www.archive.org

### 3.2.5. Сервисы работы с глубинным веб

Традиционные поисковые системы стремятся сузить пространство глубинного веб, постепенно захватывая такие ниши, как блоги, научные сайты, информационные агентства. Так, в качестве вспомогательных сервисов для поиска по глубинному веб от Google можно рекомендовать: Google Book Search (books.google.com) – поиск книг, Google Scholar (scholar.google.com) – поиск научных публикаций, Google Code Search (code.google.com) – поиск программного кода.

Система Goldfire Research от компании Invention Machine Corp. (inventionmachine.com) позволяет обрабатывать контент глубинного веб, размещенный на более чем 2000 сайтов правительственных, академических, исследовательских и коммерческих организаций США. Система Goldfire Research обладает информацией о механизмах доступа к базам данных глубинного веб и автоматически генерирует запросы к ним.

Исследовательская поисковая система Infovell из Калифорнийского университета Беркли (www.infovell.com) позволяет искать в глубинном веб по «ключевым фразам», от параграфов до целых документов, или даже наборам документов общим объемом до 25 тысяч слов. Система Infovell не зависит от языка, пользователи мо-

гут искать страницы на английском, арабском, китайском языках или же вводить в строке поиска математические уравнения, химические формулы.

Российская компания «Р-Техно» создала систему «it2b.интернетошпионопаук 3000+», предназначенную для загрузки данных из невидимого сегмента сети Интернет. На основе этой системы построен поисковый сервис Web Insight ([www.r-techno.com/rtechno/online-services/webinsight](http://www.r-techno.com/rtechno/online-services/webinsight)), обеспечивающий поиск по официальным сайтам и базам данных России и ближнего зарубежья, а именно, по документам Федеральной налоговой службы (ФНС), Федеральной службы судебных приставов (ФССП), Пенсионного фонда, Федеральной антимонопольной службы (ФАС), Трудовой инспекции, Федеральной регистрационной службы (ФРС), Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий (МЧС), Арбитражного суда, Министерства внутренних дел (МВД), Федеральной службы безопасности (ФСБ). Известны также такие базы данных службы «Р-Техно», как «Розыск Интерпола»; «Компании США, уличенные в мошенничестве»; «Недобросовестные поставщики ФАС», «Должники металлургической отрасли» и т.п.

Существующие средства анализа и продвижения веб-ресурсов позволяют по-новому подойти к оценке соотношения объемов видимого и глубинного веб. Так на веб-сайте [www.cy-pr.com](http://www.cy-pr.com) приводится информация о реальном количестве документов на исследуемом веб-сайте, представленном в RUNet, и о количестве документов, заиндексированных различными поисковыми системами, в том числе, Google и Яндекс. Получив репрезентательную выборку по сайтам, например, по рейтингу Рамблера top100 ([top100.gambler.ru](http://top100.gambler.ru)), можно получить оценку соотношения видимой и глубинной части в RUNet-сегменте веб-пространства.

Как показывают расчеты, объем информации, оказавшейся в глубинной части веб-пространства, превышает объем информации из видимой части примерно в 3-5 раз. Оказывается, за редким исключением, что чем крупнее ресурс, тем большая его часть относится к глубинному веб. В этом смысле небольшие веб-ресурсы выигрывают в доступности. Так как большая доля новостных документов оказывается в глубинном веб, то для задач бизнес-аналитики требуются специальные сервисы доступа к такой информации. Именно такой сервис предоставляют службы интеграции новостного контента – архивы сетевых СМИ. Российские и украинские бизнес-аналитики активно используют крупнейшие архивы



информации из открытых источников «Интегрум» (integrum.ru) и InfoStream (www.infostream.ua). Именно использование открытых источников позволяет конкурентной разведке действовать в рамках правового поля, но, при этом, иметь высокую эффективность.

Можно констатировать, что чем быстрее растет веб-пространство, тем хуже оно охватывается традиционными каталогами и поисковыми машинами. Из-за роста количества веб-сайтов и порталов, использующих базы данных, динамических систем управления контентом, появления новых версий форматов представления информации глубинный веб растет очень интенсивно. С одной стороны, Интернет как огромное хранилище увеличивает объем информации, доступной «в принципе», но с другой стороны – растет информационный хаос, увеличивается энтропия сетевого информационного пространства. Все меньшая часть информационных ресурсов становится доступной пользователям реально.

Ведущие поисковые системы по-прежнему пытаются найти технические возможности для индексации содержимого баз данных и доступа к закрытым веб-сайтам, однако, их задачи объективно расходятся с задачами бизнес-аналитиков – ориентация традиционных поисковых служб на массовый сервис в данном случае оправдана. Таким образом, ниша для систем поиска в глубинном веб становится все шире.

### 3.3. Специальные базы данных

Как правило, для успешного ведения конкурентной разведки должен быть создан и поддерживаться банк данных, включающий следующие основные базы данных [Ландэ, 2005]:

1. Конкуренты (действующие и потенциальные);
2. Информация о рынке (тенденции, номенклатурная, ценовая, адресная информация);
3. Технологии (продукты, выставки, конференции, ГОСТы, качество);
4. Ресурсы (сырье, человеческие и информационные ресурсы);
5. Законодательство (международные, центральные, региональные и ведомственные нормативно-правовые акты);
6. Общие тенденции (политика, экономика, региональные особенности, социология, демография).

Сегодня для конкурентной разведки основными источниками информации служат Интернет, пресса, а также открытые базы данных. Но если доступ к обычным интернет-ресурсам можно считать условно бесплатным, то, в большинстве случаев, доступ к базам

данным требует не только регистрации, но и оплаты таких услуг. Кроме того практически все они могут быть отнесены к так называемому «скрытому» веб-пространству.

Очень популярны среди специалистов по конкурентной разведке базы данных таможенных, налоговых и статистических органов, органов юстиции и судов, торгово-промышленных палат, органов приватизации и фондовых рынков, информационных, рейтинговых, аналитических и других агентств и т.д. Большую пользу приносят и отдельные доступные базы данных других контролирующих органов и организаций.

Традиционно конкурентная разведка опирается на такие источники информации, как опубликованные документы открытого доступа, которые содержат обзоры товарного рынка, информацию о новых технологиях, создании партнерств, слияниях и приобретениях, объявлениях о рабочих вакансиях, о выставках и конференциях, и т.п. Поэтому в последнее время все более популярны базы данных на основе архивов СМИ, в том числе и сетевых.

В «Большую тройку» мировых служб, занимающихся предоставлением пользователям доступа к деловой и аналитической информации, входят **LexisNexis**, **Factiva** и **Internet Securities**.

Крупнейшая в мире полнотекстовая онлайн-информационная система **LexisNexis** ([www.lexisnexis.com](http://www.lexisnexis.com)), которая содержит свыше 2 миллиардов документов из 45 тыс. источников с архивом глубиной более 30 лет по бизнес-информации и более 200 лет по правовой информации, относится к разряду «скрытого» веб (рис. 32). Каждую неделю в архивы добавляется еще 14 млн. документов. В отличие от неструктурированных массивов «поверхностного» веб, пользователи LexisNexis могут использовать мощные инструменты поиска для получения достоверной и классифицированной информации.

Служба **Factiva** ([global.factiva.com](http://global.factiva.com)), подразделение компании Dow Jones, в настоящее время принадлежит компании News Corporation, занимается предоставлением доступа к деловой и аналитической информации. В основе службы Factiva имеется более 35 тыс. первичных источников из 159 стран мира. В базе данных Factiva содержатся материалы более чем по 36,5 млн. компаний, а также полная подборка информации Investext.

Компания **Internet Securities** ([www.internetsec.com](http://www.internetsec.com)), бренд ISI Emerging Markets, охватывает 80 тематических информационных разделов, формируемых из 16 тыс. источников информации – тексты статей, финансовые и аналитические отчеты, корпоративная информация, макроэкономическая статистика, данные по рын-

кам (рис. 33). Основные продукты ISI Emerging Markets: CEIC Data, Emerging Market Information Service (EMIS), Islamic Finance Information Service (IFIS), IntelliNews, ISI Compliance Edition, ISI Deal-Watch.

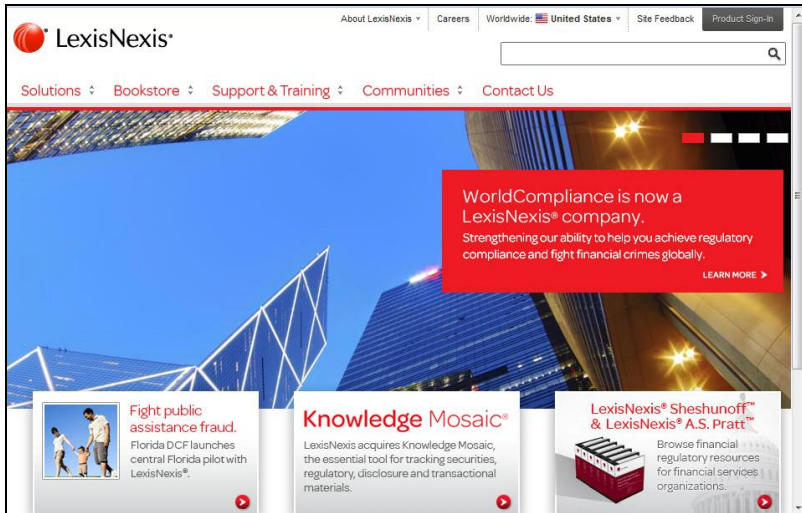


Рис. 32 – Фрагмент веб-сайта службы LexisNexis



Рис. 33 – Фрагмент веб-сайта службы Internet Securities

В России большой популярностью пользуются такие крупнейшие службы, как «**Интегрум**» (более 10 тыс. источников, сервисы «Анализ СМИ», «Архив СМИ», «Лента СМИ», БД «Компании», «Связи»), «**Медиология**» (13 тыс. источников: СМИ, телевидение, радио, газеты, журналы, информагентства, интернет, блоги, база данных из 30 тыс. объектов: компаний, персон и брендов), «**Яндекс.Новости**» (служба автоматической обработки и систематизации новостей, свыше 4000 источников, не допускается участие в работе службы материалов, содержащих сообщения неновостного характера), Public.Ru – крупная онлайн библиотека русскоязычных СМИ. С 2000 года **Public.Ru** создает свою базу данных, которая хранит архивные материалы российских изданий с 1990 года. В архивах базы данных доступно более 70 млн. статей русскоязычных СМИ собранных из 4600 источников. Основные виды источников содержащихся в базе данных: федеральные издания; региональные издания; информационные агентства; телеканалы; радиостанции; интернет-издания.

В Украине эту нишу занимает система контент-мониторинга интернет-СМИ **InfoStream** (свыше 6 тыс. источников информации, более 100 миллионов документов в архиве).

Украинская корпорация «**Media-proctip**» (550 источников информации, 25 региональных информационных бюро), которая осуществляет анализ информационного пространства Украины, обеспечивает предоставление медиа-обзоров. Систематизация информационных сообщений происходит по четырем объектам привязки: предметные сферы; политические субъекты; личности; территории.

Информационно-мониторинговая система **Web-Observer** в «базовой комплектации» охватывает 500 источников. Система внедрена в информационном агентстве УНИАН, на ней базируется сервис «УНИАН-монитор».

Украинская система интернет-мониторинга **MonitorIX** обеспечивает мониторинг интернет-источников (как сетевая информационно-поисковая система), СМИ (130 изданий), ТВ, блогов и форумов. Предоставляет клиентам результаты оперативного и архивного мониторинга.

Приведем еще один пример зарубежной базы данных из «скрытого» веб. Корпорация LexisNexis предоставляет сервис **Auto TrackXP**, вошедший в список двадцати крупнейших «скрытых» веб-сайтов мира (по рейтингу BrightPlanet). Auto TrackXP представляет собой базу данных объемом 30 Терабайт (ТБ), охватывающую практически все аспекты гражданской жизни США. База данных

Auto TrackXP содержит информацию практически о каждом гражданине США. TestProfiles.com – часть ChoicePoint Online – содержит личные характеристики и сведения о компетентности граждан США. Например, чтобы определить, не завладел ли человек чужими документами, на основе системы организован платный сервис ProCheck, позволяющий сопоставить информацию из различных источников и государственных каталогов. Для частных любителей составления «досье» ChoicePoint предлагает более скромный, но не менее любопытный набор сервисов ([www.choicetrust.com](http://www.choicetrust.com)). Подозрительные пациенты с помощью Doctor Check имеют возможность самостоятельно выбрать или проверить квалификацию врачей 40 различных специализаций. Отчет, получаемый с помощью системы, может, например, служить для страховой компании поводом в отказе выдачи полиса.

Система широко используется как легальный ресурс для задач конкурентной разведки. Вместе с тем, сегодня американцы повсеместно выражают возмущение, обнаруживая существование подобных сервисов, видя в этом нарушение своих гражданских прав.

Сервис **Insight Profiles** ([www.insightprofiles.com](http://www.insightprofiles.com)) – содержит личные характеристики и сведения о способностях и компетентности граждан США. Например, чтобы определить, не завладел ли человек чужими документами, организован платный сервис **ProCheck** ([procheck.com](http://procheck.com)), позволяющий сопоставить информацию из различных источников и государственных каталогов.

В России и Украине популярны такие базы данных, как:

«**Лабиринт**» ([www.labyrinth.ru](http://www.labyrinth.ru)) – база данных, составленная на основе публикаций ведущих бизнес-изданий, предназначена для помощи при выполнении аналитических и исследовательских работ, написании статей, комментариев, докладных записок, пр. В БД представлены биографии российских деятелей, справки об организациях и компаниях, информация о субъектах Российской Федерации и другой справочный материал;

«**Компасс**» ([www.kompass.com/ru](http://www.kompass.com/ru)) – база данных, позиционирующая себя как международную информационную B2B (типа «бизнес для бизнеса») поисковую систему, обеспечивает поиск по компаниям, товарам и услугам, руководителям с целью формирования баз данных целевого маркетинга и сбыта, потенциальных клиентов;

«**КАРЕ**» ([kare.pulscen.com.ua](http://kare.pulscen.com.ua)) – БД предприятий Украины – 384000 компаний, БД агропромышленного комплекса Украины – 218000 компаний;

База данных **Dun & Bradstreet (D&B)**. Национальные представительства **D&B** в Украине – это компания «Бизнес-мониторинг», которая входит в состав группы компаний «Авеста-Украина» и представляют направление деловой информации. Компания обеспечивает подключение к Системе Профессионального Анализа Рынков и Компаний (СПАРК-Россия) и к базе данных Dun & Bradstreet (D&B). В России подключение к этой базе данных обеспечивает информационное агентство Интерфакс ([www.dnb.ru](http://www.dnb.ru));

Базы данных международной корпорации **Creditreform**, представленной в России компанией «Кредитреформ РУС» ([www.creditreform-rus.ru](http://www.creditreform-rus.ru)), а в Украине информационным агентством «Кредитреформ Украина» ([www.creditreform.ua](http://www.creditreform.ua)). Компании обеспечивают доступ к международной сети содействия информационному бизнесу BIGNet (Business Information Group Network). Эта сеть объединяет независимые организации, которые предоставляют бизнес-справки по всему миру в режиме онлайн (свыше 8 млн. в год), а также доступ к базе данных BIGNet, а также к собственной базе данных ([www.crefoport.ru](http://www.crefoport.ru)), содержащей данные о 30 млн. компаний.

**Europages** ([www.europages.eu](http://www.europages.eu)) – Европейская бизнес-директория – информационно-поисковая B2B-система, охватывающая свыше 2 млн. поставщиков, производителей и дистрибьюторов в Европе и во всем мире.

Задача полного перечисления всех источников информации практически невыполнима, так как этот рынок очень динамичен, постоянно появляются новые базы данных, происходит слияние существующих источников, поглощение слабых сильными. Вместе с тем, одно из правил конкурентной разведки формулируется таким образом: «чем большим количеством независимых источников подтверждается информация – тем более она достоверна».

Наряду с базами данных, одним из самых эффективных источников информации могут служить отчеты и справки аутсорсинговых компаний, профессионально занимающихся конкурентной разведкой и сбором сведений о коммерческих структурах и рынках. Их продукция, на самом деле, и является результатом конкурентной разведки.

В мире существует множество таких специальных компаний. Одной из таких крупнейших компаний, которой принадлежит около 80 % западного рынка, является американская компания, **Dun & Bradstreet (D&B)**, чья база данных упоминалась нами выше. Справка по любой компании в этой службе оценивается из расчета в среднем 100 долларов и выше. Более серьезный анализ рынка или конкурента может обойтись и в 10 тыс. долларов. Сроки испол-

нения – от нескольких часов (информация присутствует в базе данных) – до нескольких суток для справок и до нескольких месяцев для аналитической работы.

На европейском рынке не менее известны названная выше ирландская компания **Creditreform**, немецкая **Schufa Holding AG** (479 млн. документов в БД, в том числе, 66 млн. о физических лицах), австрийская **Intercredit Information Holding**, латвийская **Coface IGK** (известна IGK System – база данных должников, включающая сведения о текущих долгах, судебных исках, а также процессах неплатежеспособности) и многие другие. Некоторые из этих компаний совмещают функции конкурентной разведки с другими видами деятельности, например, обязанностями кредитных бюро.

Общей проблемой при обращении за информационными справками в западные агентства, имеющие представительства в России и Украине, является то, что, как правило, информация, предоставляемая в отношении западных нерезидентов, намного обширнее и качественнее, чем та, что предоставляется в отношении отечественных фирм. В связи с чем, в таких случаях целесообразно обращаться к местным информационным компаниям, результаты оказываются дешевле и качественнее.

На российском рынке в сфере конкурентной разведки пользуются популярностью информационные отчеты компаний «**P-Tехно**», «**Медиология**», «**Специальная Информационная Служба**», «**Интегрум**», «**Кронос-Информ**» и многих других.

В Украине также существует целый ряд подобных компаний, среди которых можно назвать «**Авеста-Украина**», «**СИДКОН**», «**Межбанковская служба безопасности «СКИФ»**» и другие.

Все отечественные и зарубежные информационные компании имеют свои представительства и принимают заказы в Интернете, в связи с чем их можно отнести к специфическим интернет-источникам.

Следует также отметить, что, несмотря на то, что в случае заказа услуг аутсорсинговой компании, она делает большую часть информационной работы за клиента, окончательные выводы и решения, рекомендации для принятия управленческих решений все-таки остаются за ним. Только клиент может обладать всей необходимой полнотой внешней и инсайдерской информации.

### 3.4. Социальные медиа

Социальные медиа представляют собой совокупность онлайн-сервисов и интернет-приложений, которые позволяют пользовате-

лям общаться друг с другом в том числе, и в режиме реального времени. При этом пользователи могут обмениваться между собой мнениями, новостями, информацией, в том числе и мультимедийной.

Социальные медиа базируются на идеологической и технологической базе веб 2.0, позволяющих создание и обмен контентом, созданным самими пользователями (User-Generated Content), в отличие от предшествующей концепции веба, предполагающей, как и в случае традиционных СМИ, централизованное создание контента, поставляемого пользователям-читателям.

Очевидно, социальные медиа являются самым ценным источником информации для конкурентной разведки, предоставляя абсолютно на легальных условиях разностороннюю информацию о людях, событиях, компаниях, брендах, продуктах. Получившие в последнее время широкое распространение такие явления, как информационные операции, активное информационное противодействие в рамках конкурентной борьбы, сетевая мобилизация, во многих случаях базируется на манипулировании данными именно в социальных медиа.

Выделяют семь разновидностей социальных медиа, это социальные сети; блоги; форумы; сайты отзывов; серверы фото- и видеохостинга; виртуальные службы знакомств и геосоциальные сети. Следует отметить, что четкие границы между этими разновидностями размыты.

Под социальной сетью в сети Интернет (social networking service) понимается онлайн-сервис, предназначенный для построения, отображения и организации социальных взаимоотношений, обеспечивающий предоставление широкого спектра возможностей для обмена информацией, возможность пользователя предоставить информацию о самом себе (создать свой профиль), построить связи, найти друзей по интересам, подключить родственников, коллег, одноклассников и т. п.

Под блогом (blog, от web log) понимают веб-сайт, основное содержание которого – это периодически добавляемые пользователями записи (текст, изображения или мультимедиа). Для блогов характерны недлинные записи (особенно, в случаях так называемых «микроблогов») временной значимости, блоги обычно публичны и предполагают сторонних читателей, которые могут вступить в публичную полемику с автором (в комментариях к блогозаписи или своих блогах). Совокупность всех блогов в сети Интернет называют блогосферой.



Веб-форумы представляют собой веб-приложения, предназначенные для организации общения посетителей некоторых интернет-ресурсов (веб-сайтов или порталов). На ресурсах веб-форума пользователи задают интересующие их темы, которые затем обсуждаются и другими пользователями путем размещения сообщений (постинга) внутри этих тем.

Веб-сайты отзывов создаются с целью повышения эффективности и качества предоставляемых (не обязательно в интернет-среде) услуг и товаров. Пользователи, посещая веб-сайты отзывов, оставляют там свои сообщения, участвуют в анкетированиях, формируют мнения о той или иной услуге или товаре.

Фотохостинг (photo hosting) – это веб-сайт, позволяющий публиковать любые изображения (чаще всего, цифровые фотографии) в сети Интернет. Основное преимущество фотохостинга – удобство демонстрации размещенных фотографий. Соответственно, видеохостинг – это веб-сайт, позволяющий загружать и просматривать видеoinформацию в веб-браузере. Видеохостинг набирает популярность в связи с развитием широкополосного доступа в Интернет.

Виртуальная служба знакомств представляет собой интернет-сервис, оказывающий услуги по виртуальному знакомству пользователей с целями общения, создания семьи, серьезных отношений и др. При использовании виртуальной службы знакомств пользователь создает анкету, в которой указывает свой псевдоним (никнейм) и другие параметры, запрашиваемые службой (пол, возраст, цель знакомства, интересы, фотографии). После регистрации пользователь может общаться с другими пользователями, получать сообщения и отвечать на них.

Геосоциальные сети (GeoSocial Network) – это разновидность социальных сетей, в которых пользователи оставляют данные о своем местонахождении, что позволяет объединять и координировать их действия на основании информации о том, какие люди присутствуют в тех или иных местах, какие события происходят в этих местах.

## 4. Социальные сети

---

---

Термин «социальная сеть» обозначает сосредоточение социальных объектов, которые можно рассматривать как сеть (или граф), узлы которой – объекты, а связи – социальные отношения. Этот термин был введен в 1954 году социологом из «Манчестерской школы» Дж. Барнсом (J. Barnes) в работе «Классы и сборы в норвежском островном приходе». Во второй половине XX столетия понятие «социальная сеть» стало популярным у западных исследователей, при этом как узлы социальных сетей стали рассматривать не только представителей социума, но и другие объекты, которым присущи социальные связи. Сегодня термин «социальная сеть» обозначает понятие, оказавшееся шире своего социального аспекта, оно включает, например, многие информационные сети, в том числе и WWW. Рассматривают не только статистические, но и динамические сети, для понимания структуры которых необходим учет принципов их эволюции.

Сегодня под термином «социальные сети» (*Social Networks*) понимают, прежде всего, онлайн-сервисы в сети Интернет, предназначенные для формирования, отображения и упорядочения социальных взаимоотношений. Особенности социальных сетей:

- 1) предоставление пользователям широкого спектра возможностей для обмена информацией;
- 2) создание профилей пользователей, в которых требуется указать некоторое количество персональной информации;
- 3) друзьями в социальной сети становятся преимущественно не виртуальные, а реальные друзья.

Веб-ресурс социальной сети предоставляет возможности:

- 1) активного общения;
- 2) создания публичного или закрытого профиля (*Profile*) пользователя, содержащего персональные данные;
- 3) организации и ведения пользователем списка других пользователей, с которыми у него имеются некоторые социальные отношения;
- 4) просмотра связей между пользователями внутри социальной сети;
- 5) образования групп пользователей по интересам;
- 6) управления содержимым в рамках своего профиля;
- 7) синдикации контента;
- 8) подключения различных приложений.

### 4.1. Анализ социальных сетей

В последнее время выделилось отдельное научное направление – анализ социальных сетей (SNA, Social Networks Analysis), которое базируется, с одной стороны, на социологии, а с другой – на теории сложных сетей (Complex Networks) [Newman, 2003]. В рамках теории сложных сетей изучаются сетевые характеристики не только с точки зрения топологии сетей, но и статистические феномены, распределение весов отдельных узлов и ребер, эффекты протекания и проводимости. Несмотря на то, что в рассмотрение теории сложных сетей попадают различные сети (электрические, транспортные, информационные), наибольший вклад в развитие этой теории внесли исследования социальных сетей [Ландэ и др., 2009]

В теории сложных сетей выделяют три основных направления:

- исследование статистических свойств, которые характеризуют поведение сетей;
- создание моделей сетей;
- предсказание поведения сетей при изменении структурных свойств.

В прикладных исследованиях чаще всего применяются такие типичные для сетевого анализа характеристики, как размер сети, сетевая плотность, степень центральности и т. п.

При анализе сложных сетей, как и в теории графов, исследуются:

- параметры отдельных узлов;
- параметры сети в целом;
- сетевые подструктуры.

Для отдельных узлов выделяют следующие параметры:

- входная степень узла – количество ребер графа, которые входят в узел;
- выходная степень узла – количество ребер графа, которые выходят из узла;
- расстояние от данного узла до каждого из других;
- среднее расстояние от данного узла до других;
- эксцентричность (eccentricity) – наибольшее из геодезических расстояний (минимальных расстояний между узлами) от данного узла к другим;
- посредничество (betweenness), показывающее, сколько кратчайших путей проходит через данный узел;
- центральность – общее количество связей данного узла по отношению к другим.

Для анализа сети в целом используют такие параметры, как:

- число узлов;
- число ребер;
- геодезическое расстояние между узлами;
- среднее расстояние от одного узла к другим;
- плотность – отношение количества ребер в сети к возможному максимальному количеству ребер при данном количестве узлов;
- количество симметричных, транзитивных и циклических триад;
- диаметр сети – наибольшее геодезическое расстояние в сети и т.д.

Существует несколько актуальных задач математического исследования социальных сетей, среди которых можно выделить следующие основные:

Важной характеристикой сети является функция распределения степеней узлов  $P(k)$ , которая определяется как вероятность того, что узел  $i$  имеет степень  $k_i = k$ . Сети, характеризующиеся разными  $P(k)$ , демонстрируют различное поведение.  $P(k)$  в некоторых случаях может быть распределением Пуассона ( $P(k) = e^{-m} m^k / k!$ , где  $m$  – математическое ожидание), экспоненциальным ( $P(k) = e^{-k/m}$ ) или степенным ( $P(k) \sim 1/k^\gamma$ ,  $k \neq 0$ ,  $\gamma > 0$ ).

Сети со степенным распределением степеней узлов называются безмасштабными (scale-free). Именно безмасштабные распределения часто наблюдаются в реально социальных сетях. При степенном распределении возможно существование узлов с очень высокой степенью, что практически не наблюдается в сетях с пуассоновским распределением.

Расстояние между узлами определяется как количество шагов, которые необходимо сделать, чтобы по существующим ребрам добраться от одного узла до другого. Естественно, узлы могут быть соединены прямо или опосредованно. Кратчайшим путем  $d_{ij}$  между узлами  $i$  и  $j$  называется наименьшее расстояние между ними. Для всей сети можно ввести понятие среднего пути, как среднее по всем парам узлов кратчайшего расстояния между ними:

$$l = \frac{2}{n(n+1)} \sum_{i \geq j} d_{ij},$$

где  $n$  – количество узлов,  $d_{ij}$  – кратчайшее расстояние между узлами  $i$  и  $j$ .

Венгерскими математиками П. Эрдёшем (P. Erdős) и А. Реньи (A. Rényi) было показано, что среднее расстояние между двумя вершинами в случайном графе растет как логарифм от числа его узлов [Erdős, 1960].

Некоторые сети могут оказаться несвязными, т.е. найдутся узлы, расстояние между которыми окажется бесконечным. Соответственно, средний путь может оказаться также равным бесконечности. Для учета таких случаев вводится понятие среднего инверсного пути (его еще называют «эффективностью сети») между узлами, рассчитываемое по формуле:

$$il = \frac{2}{n(n-1)} \sum_{i>j} \frac{1}{d_{ij}}.$$

Сети также характеризуются таким параметром как диаметр или максимальный кратчайший путь, равный максимальному значению из всех  $d_{ij}$ .

Д. Уаттс (D. Watts) и С. Стрoгaтц (S. Strogatz) в 1998 году определили такой параметр сетей, как коэффициент кластерности [Watts, 1998], который характеризует уровень связности узлов в сети, тенденцию к образованию групп взаимосвязанных узлов, так называемых клик (clique). Кроме того, для конкретного узла коэффициент кластеризации показывает, сколько ближайших соседей данного узла являются также ближайшими соседями друг для друга.

Коэффициент кластерности может определяться как для каждого узла, так и для всей сети. Для сети коэффициент кластерности определяется как нормированная по количеству узлов сумма соответствующих коэффициентов для отдельных узлов.

Коэффициент кластерности для отдельного узла сети определяется следующим образом. Пусть из узла выходит  $k$  ребер, которые соединяют его с  $k$  другими узлами, ближайшими соседями. Если предположить, что все ближайшие соседи соединены непосредственно друг с другом, то количество ребер между ними со-

ставляло бы  $\frac{1}{2}k(k-1)$ . Т.е. это число, которое соответствует максимально возможному количеству ребер, которыми могли бы соединяться ближайшие соседи выбранного узла. Отношение реального

количества ребер, которые соединяют ближайших соседей данного узла к максимально возможному (такому, при котором все ближайшие соседи данного узла были бы соединены непосредственно друг с другом) называется коэффициентом кластерности узла  $i - C(i)$ . Естественно, эта величина не превышает единицы.

Посредничество (betweenness) – это параметр, показывающий, сколько кратчайших путей проходит через узел. Эта характеристика отражает роль данного узла в установлении связей в сети. Узлы с наибольшим посредничеством играют главную роль в установлении связей между другими узлами в сети. Посредничество  $b_m$  узла  $m$  определяется по формуле:

$$b_m = \sum_{i \neq j} \frac{B(i, m, j)}{B(i, j)},$$

где  $B(i, j)$  – общее количество кратчайших путей между узлами  $i$  и  $j$ ,  $B(i, m, j)$  – количество кратчайших путей между узлами  $i$  и  $j$ , проходящих через узел  $m$ .

О «структуре сообщества» можно говорить тогда, когда существуют группы узлов, которые имеют высокую плотность ребер между собой, притом, что плотность ребер между отдельными группами – низкая. Традиционный метод для выявления структуры сообществ – кластерный анализ. Существуют десятки приемлемых для этого методов, которые базируются на разных мерах расстояний между узлами, взвешенных путевых индексах между узлами и т.п. В частности, для больших социальных сетей наличие структуры сообществ оказалось неотъемлемым свойством.

К свойствам реальных социальных сетей относятся и так называемые «слабые» связи. Аналогом слабых социальных связей являются, например, отношения с далекими знакомыми и коллегами. В некоторых случаях эти связи оказываются более эффективными, чем связи «сильные». Так, группой исследователей из Великобритании, США и Венгрии, был получен концептуальный вывод в области мобильной связи, заключающийся в том, что «слабые» социальные связи между индивидуумами оказываются наиболее важными для существования социальной сети [Vjorneborn, 2004].

Для исследования были проанализированы звонки 4.6 млн. абонентов мобильной связи, что составляет около 20 % населения одной европейской страны. Это был первый случай в мировой практике, когда удалось получить и проанализировать такую боль-

шую выборку данных, относящихся к межличностной коммуникации.

В социальной сети с 4.6 млн. узлов было выявлено 7 млн. социальных связей, т.е. взаимных звонков от одного абонента другому и обратно, если обратные звонки были сделаны на протяжении 18 недель. Частота и продолжительность разговоров использовались для того, чтобы определить силу каждой социальной связи.

Было выявлено, что именно слабые социальные связи (один-два обратных звонка на протяжении 18 недель) связывают воедино большую социальную сеть. Если эти связи проигнорировать, то сеть распадется на отдельные фрагменты. Если же не учитывать сильных связей, то связность сети нарушится (рис. 34). Оказалось, что именно слабые связи являются тем феноменом, который связывает сеть в единое целое.

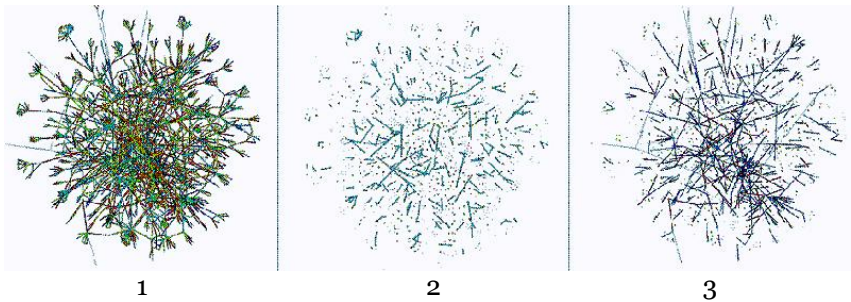


Рис. 34 – Структура сети:

- 1) полная карта сети социальных коммуникаций;
- 2) социальная сеть, из которой изъяты слабые связи;
- 3) сеть, из которой изъяты сильные связи: структура сохраняет связность

Несмотря на огромные размеры некоторых социальных сетей во многих из них существует сравнительно короткий путь между двумя любыми узлами – геодезическое расстояние. В 1967 г. психолог С. Милгран в результате проделанных масштабных экспериментов вычислил, что существует цепочка знакомств, в среднем длиной шесть звеньев, практически между двумя любыми гражданами США [Milgram, 1967].

Д. Уаттс и С. Стрататц обнаружили феномен, характерный для многих реальных сетей, названный эффектом малых миров (Small Worlds) [Watts, 1998]. При исследовании этого феномена ими была предложена процедура построения наглядной модели сети, которой присущ этот феномен. Три состояния этой сети пред-

ставлены на рис. 35: регулярная сеть – каждый узел которой соединен с четырьмя соседними, та же сеть, у которой некоторые «ближние» связи случайным образом заменены «далекими» (именно в этом случае возникает феномен «малых миров») и случайная сеть, в которой количество подобных замен превысило некоторый порог.

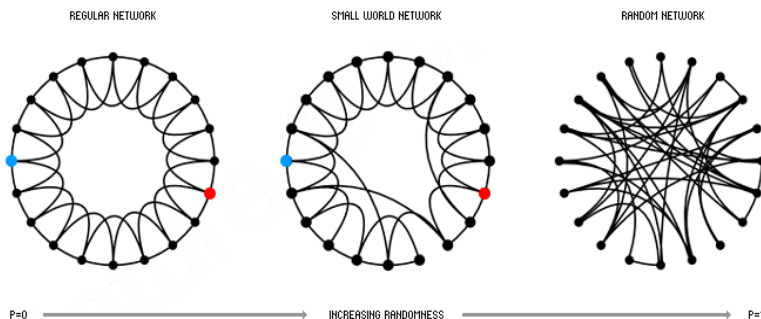


Рис. 35 – Модель Уаттса-Строгатца

В реальности оказалось, что именно те сети, узлы которых имеют одновременно некоторое количество локальных и случайных «далеких» связей, демонстрируют одновременно эффект малого мира и высокий уровень кластеризации.

С концепцией «малых миров» связан также практический подход, называемый «сетевой мобилизацией», которая реализуется над структурой «малых миров». В частности, скорость распространения информации благодаря эффекту «малых миров» в реальных сетях возрастает на порядки по сравнению со случайными сетями, ведь большинство пар узлов реальных сетей соединены короткими путями.

Экспертами по проблемам безопасности эффект «малых миров» в последнее время все чаще связывается с сетями террористических организаций, так называемыми оверлейными сетями, т.е. сетями, надстроенными поверх сети Интернет.

Анализируя связи в сети, можно узнать о ее важных свойствах, например, выявить наличие кластеров, определить их состав, различия в связности внутри и между кластерами, идентифицировать ключевые элементы, связывающие кластеры между собой и т.д. Вместе с тем серьезным препятствием при анализе является неполная информация о связях между отдельными узлами сети.



Недавно группа исследователей из Института Санта Фе (Santa Fe Institute) представила алгоритм, с помощью которого становится возможным автоматическое получение информации об иерархической структуре подобных сетей [Clauset, 2008]. Новый метод восстановления связей может поступить на вооружение как спецслужб, так и подразделений конкурентной разведки компаний. Так, например, зная только о половине связей между террористами, можно будет с высокой вероятностью восстановить недостающие звенья всей цепочки.

Даже не имея полного описания системы можно получать репрезентативную выборку связей и по ней пытаться достроить всю сеть. Анализ получающегося графа позволяет выявить потенциально важные связи, которые не удалось обнаружить в реальной сети. Например, имея информацию только о половине контактов участников сети между собой, можно с вероятностью 0,8 прогнозировать те связи, о которых сначала ничего не было известно. Очевидно, что данный метод может быть очень полезным при выявлении скрытых сетевых группировок, и таким образом поставить дело обеспечения государственной и коммерческой безопасности на качественно новый уровень.

Для анализа сложных сетей понятий, упоминаемых в отдельных документах из информационных потоков, могут применяться методы глубинного анализа текстов, а точнее контент-мониторинга и экстрагирования таких понятий, как персоны, компании, топонимы (географические названия) и т.п.

Одним из направлений анализа социальных сетей является визуализация, которая имеет важное значение, поскольку зачастую позволяет делать важные выводы относительно характера взаимодействия субъектов-узлов, не прибегая к точным методам анализа. При отображении модели сети может оказаться целесообразным:

- размещение узлов сети в двух измерениях;
- пространственное упорядочение объектов в одном измерении в соответствии с некоторыми количественными свойствами;
- использование общих для всех сетевых диаграмм методов для отображения количественных и качественных свойств объектов и отношений.

В качестве примеров визуализации сетей можно рассмотреть некоторые разработки компании TouchGraph. Так, например, система TouchGraph Amazon отражает сеть, порожденную книгами и связями между ними (по тематикам, авторам, издательствам). Компания TouchGraph также реализовала интерфейс для построения социальной сети на основе Livejournal – TouchGraph LiveJour-

nal Browser. В случае визуализации WWW средствами TouchGraph SEO Browser (www.touchgraph.com/seo) ребрами выступают не гиперссылки, а отношения подобия.

Сервис TouchGraph SEO Browser представляет собой Java-апплет, позволяющий визуализировать связи подобия между веб-сайтами, которые рассчитываются в поисковой системе Google. В этом интерфейсе можно увидеть все веб-сайты, связанные отношением подобия с заданным понятием, при этом пользователь может указывать глубину связей и отображать взаимосвязи различных понятий. TouchGraph SEO Browser – весьма полезный инструмент также при поиске веб-сайтов, связанных с исходным общей тематикой (рис. 36).

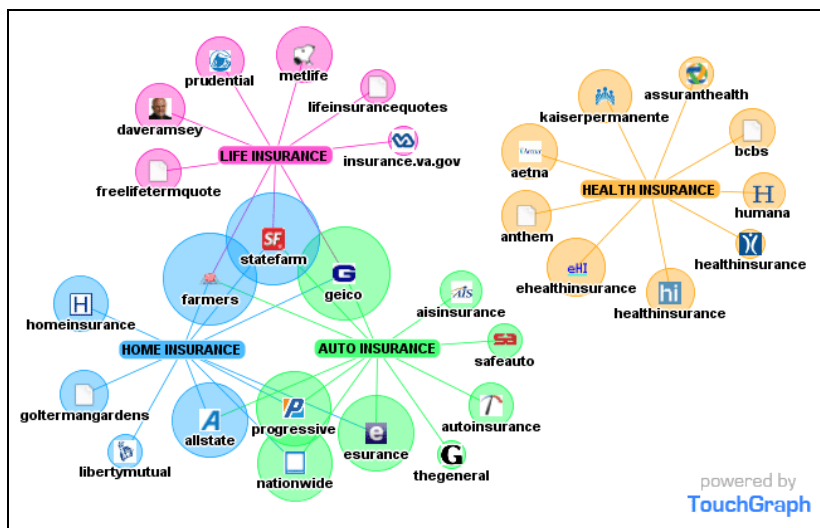


Рис. 36 – Фрагмент отображения связей веб-серверов (от TouchGraph)

В качестве примера применения возможностей анализа социальных сетей приведем фрагмент исследования сети связей понятий (фамилий персон), экстрагируемых из корпусов неструктурированных текстов – массивов документов, сканируемых из сети Интернет системой контент-мониторинга InfoStream [Григорьев, 2007].

При построении сети понятий использовались алгоритмы автоматического извлечения понятий из неструктурированных текстов. Следует отметить, что подходы к извлечению различных типов понятий из текстов существенно отличаются как по контексту

их представления, так и по структурным признакам. Так, для выявления принадлежности документа к рубрике тематического классификатора могут использоваться специальным образом составленные запросы, включающие логические и контекстные операторы, скобки и т.п.

Выявление географических названий предполагает использование таблиц, в которых помимо шаблонов написания этих названий используются коды стран, названия регионов и населенных пунктов.

Еще один вид понятий, такой как «персоны», экстрагируется из текстов на основании правил, учитывающих таблицы допустимых имен и фамилий, шаблоны инициалов, возможные варианты совместного написания инициалов/имен и фамилий.

Следует отметить, что система InfoStream включает средства извлечения понятий и, среди прочего, предоставляет пользователям результаты в виде «информационных портретов», включающих такие понятия, как ключевые слова, географические названия, фамилии персон, названия фирм и т.д. В рамках этой системы анализируются свойства сетей, образованных понятиями, связанных друг с другом упоминаниями в тех же документах.

Сеть, образованная понятиями, извлекаемыми из потоков текстов, не статичен, а зависит от объемов документов, из которых извлекаются соответствующие понятия. Следовательно, для понимания структуры такой сети необходимо учитывать ее эволюцию.

Ребра исходной сети приписываются весовые значения, равные количеству документов, в которых встречаются упоминания персон, отвечающих узлам. Для предотвращения «шума», ребра с весом, меньше чем 2 не учитывались. При развитии сети с фиксированным количеством персон, при увеличении количества рассмотренных документов среднее расстояние между узлами, соответственно уменьшается, достигая своего логического насыщения.

Интересен тот факт, что узлы рассматриваемой сети персон с максимальным количеством исходящих ребер преимущественно имеют наибольший уровень посредничества и не могут рассматриваться в качестве основы для построения кластеров при автоматической группировке, а скорее как элементы, соединяющие отдельные группы узлов.

Информационная сеть персон (наиболее упоминаемых) и их связей, полученных путем анализа выходного потока по экономической проблематике за определенный период времени, представлена в виде графа (рис. 37).

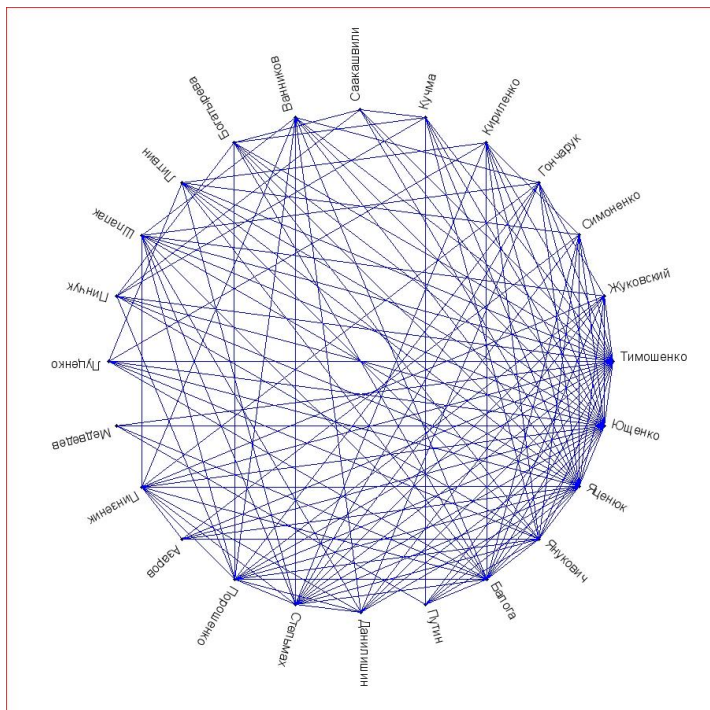


Рис. 37 – Отображение информационной сети персон

Полученные эмпирические результаты могут быть полезными, например, при моделировании экономико-социальных процессов, выявлении и визуализации неявных связей отдельных объектов или субъектов. Феномен стабилизации сети связей на практике позволяет путем анализа небольшого массива документов выявлять устойчивые связи, снижать влияние шумовых факторов. Вместе с тем пока остается открытым вопрос оценки корреляции полученных информационных взаимосвязей персон, рассчитанных путем подсчета частоты документов, в которых персоны упоминаются совместно, и взаимосвязей реальных.

## 4.2. Основные социальные сети

В список крупнейших социальных сетей, которые могут быть интересными для конкурентной разведки, можно включить:

- Facebook;

- ВКонтакте;
- Google+;
- Мой Круг;
- LinkedIn;
- Badoo;
- Livejournal;
- Twitter.

**Facebook** ([www.facebook.com](http://www.facebook.com)) – крупнейшая социальная сеть, основанная в 2004 году М. Цукербергом и его компаньонами. Начиная с сентября 2006 года социальная сеть доступна для пользователей сети Интернет. 4 октября 2012 года аудитория *Facebook* составила 1 миллиард пользователей. Суточная активная аудитория превышает 525 миллионов человек. Около 500 млн. человек в месяц используют мобильные приложения *Facebook*. Каждый день в социальной сети пользователи оставляют 3,2 миллиарда «лайков» и комментариев и публикуют 300 миллионов фотографий. На сайте зафиксировано 125 млрд. «дружеских связей». Ежемесячное количество просмотров страниц *Facebook* превышает 1 триллион.

«**ВКонтакте**» ([vk.com](http://vk.com)) – крупнейшая в Рунете социальная сеть, созданная П. Дуровым в 2006 г., позиционирующая себя как «современный, быстрый и эстетичный способ общения в сети». По данным на начало 2013 года ежедневная аудитория «ВКонтакте» составляет более 43 миллионов пользователей. Пользователям «ВКонтакте» доступен характерный для многих социальных сетей набор возможностей: создавать профиль с информацией о себе, производить и распространять контент, управлять настройками доступа, взаимодействовать с другими пользователями приватно и публично, отслеживать через ленту новостей активность друзей и сообществ. Социальная сеть «ВКонтакте» предлагает сторонним ресурсам использовать специально разработанные инструменты – виджеты – для глубокой интеграции с социальной сетью. Эти решения позволяют встраивать в веб-сайты систему комментариев для пользователей, сообщества, системы опросов, а также возможность легко поделиться ссылкой на материал с другими пользователями и авторизоваться на сайте. У сайта имеется мобильная версия, расположенная по адресу [m.vk.com](http://m.vk.com). 24 мая 2013 г. Роскомнадзор внёс домен [vk.com](http://vk.com) и его IP-адрес в «Единый реестр запрещённых сайтов», однако уже через несколько часов удалил его оттуда, обосновав ошибку человеческим фактором.

**Google+** ([plus.google.com](http://plus.google.com)) – социальная сеть от компании Google, официально начавшая свою работу в 2011 г. На начало 2012 г.

количество зарегистрированных в Google+ пользователей превышало девяносто миллионов человек. Сервис предоставляет возможность общения через Интернет с помощью специальных компонентов: *Круги*, *Темы*, *ВидеоВстречи*, *Мобильная версия*. Основными принципами действия сервиса являются пользователи, приватность и живое общение. Информация, которой делятся участники сети, влияет на персонализированные результаты поиска Google. В основе работы Google+ лежит концепция кругов (*Circles*), благодаря которым пользователь регулирует свое общение. На основе кругов пользователь делится контентом, определяя, какой круг будет иметь доступ к информации, а какой нет. Обмен пользовательскими материалами идёт в специальной ленте (*Stream*), в которой можно следить за обновлениями участников кругов. Google была представлена также и мобильная версия социальной сети, в которой есть две уникальные функции: мгновенная загрузка фото и Чат (*Huddle*). Социальная сеть Google+ позволяет получать хорошие позиции в поиске Google.

«**Мой Круг**» ([www.moikrug.ru](http://www.moikrug.ru)) – русскоязычная социальная сеть, направленная на установление деловых контактов между людьми. Архитектура сети представляет собой круги пользователей, где первый круг – это близкие друзья пользователя, которым он доверяет свою контактную информацию, второй круг – это друзья друзей пользователя, а третий соответственно друзья друзей его друзей. Сеть «Мой Круг» была создана в 2005 г. В 2007 г. проект был куплен компанией «Яндекс». Теперь он является одним из сервисов Яндекса. В 2009 году к сервису был подключен экспорт вакансий из крупнейших тематических сайтов: [hh.ru](http://hh.ru) и [rabota.mail.ru](http://rabota.mail.ru). В 2011 году социальная сеть «Мой круг» завершил процесс интеграции с социальными платформами – появилась возможность поиска друзей через аккаунты Facebook, Twitter, LiveJournal и LinkedIn.

**LinkedIn** ([www.linkedin.com](http://www.linkedin.com)) – социальная сеть для поиска и установления деловых контактов. В LinkedIn зарегистрировано свыше 200 миллионов пользователей из 200 стран, представляющих 150 отраслей бизнеса. Социальная сеть LinkedIn, основанная Р. Хоффманом, была запущена в эксплуатацию в 2003 г. Эта социальная сеть предоставляет возможность зарегистрированным пользователям создавать и поддерживать список деловых контактов. Контакты могут быть приглашены как из сайта, так и извне, однако LinkedIn требует предварительного знакомства с контактами. В случае, когда пользователь не имеет прямой связи с контактом, он может быть представлен через другой контакт. Список контактов

LinkedIn может использоваться для: расширения связей, поиска компаний, людей и групп по интересам, публикации резюме и поиска работы рекомендовать пользователей (быть рекомендованными), публиковать вакансии, создавать группы по интересам. Социальная сеть LinkedIn также позволяет публиковать информацию о деловых поездках и конференциях.

**Badoo** (badoo.com) – социальная сеть знакомств, основанная в 2006 г. российским бизнесменом А. Андреевым. По состоянию на 2013 г. в Badoo зарегистрировано более 180 миллионов пользователей. Зарегистрировавшись, пользователь может общаться в чате, загружать на сайт свои фотографии, связываться с друзьями в своем регионе или за его пределами. Существуют также премиум-услуги, которые являются платными. Они предоставляются тем, кто хочет иметь большую популярность, расширить круг знакомств. За время своего существования компания Badoo выпустила несколько продуктов под свободной лицензией, включая различные улучшения языка программирования PHP, сервер Pinba, собирающий статистику в реальном времени, бесплатный быстрый шаблонизатор Blitz для PHP.

**«Живой Журнал», ЖЖ, LiveJournal, LJ** (www.livejournal.com) – платформа для ведения онлайн-дневников (блогов), созданная в 1999 г. американским программистом Б. Фицпатриком. ЖЖ предоставляет пользователям возможность публиковать свои и комментировать чужие записи, вести коллективные блоги («сообщества»), добавлять в друзья («френдить») других пользователей и следить за их записями в «ленте друзей» («френдленте»). Серверы ЖЖ находятся в США и система принадлежит американской компании LiveJournal, Inc., хотя ей и владеет российская компания SUP Media. По данным статистики LiveJournal.com на конец 2012 г. в «Живом Журнале» зарегистрировано более 40 млн. пользователей. Среди настроек, функций и опций «Живого Журнала» следует выделить: разные типы записей и возможность их комментирования; указание расширенных сведений о пользователе; друзья и лента друзей; картинки пользователей; функции безопасности аккаунта.

**Твиттер** (англ. *Twitter* – «щебетать») (twitter.com) – сервис, позволяющий пользователям отправлять короткие текстовые заметки (до 140 символов), используя веб-интерфейс, SMS, средства мгновенного обмена сообщениями или сторонние программы-клиенты. Созданный Дж. Дорси в 2006 г., по состоянию на начало 2011 года сервис насчитывает свыше 200 млн. пользователей, из них 50 млн. пользуются Твиттером ежедневно. 55 % пользуются

Твиттером на мобильных гаджетах. Особенностью Твиттера является публичная доступность размещённых сообщений; это называется микроблоггингом.

### 4.3. Мониторинг социальных медиа

Мониторинг социальных медиа – важнейший этап для успешного развития бизнеса, продвижения в Интернет, конкурентной разведки. С помощью социальных медиа можно узнать наиболее полную информацию об аудитории товара или услуги, ее мнении о работе компании.

Рассмотрим несколько сервисов для эффективного мониторинга социальных медиа, сосредоточив внимание на наиболее доступных:

**Seesmic** (seesmic.com) – бесплатный сервис мониторинга социальных медиа. Поддерживает мониторинг таких ресурсов, как: Twitter, Facebook, LinkedIn, Chatter, Google Buzz, Ping.fm. Есть приложения как для веб, так и для персонального компьютера, iPhone, Android, Windows Mobile.

**Socialmention** (www.socialmention.com) – платформа бесплатного поиска и анализа информации в социальных сетях. Система ищет упоминания в выбранных сетях или во всех сетях сразу. Предоставляет анализ тональности упоминаний, связанные ключевые слова, популярные источники и многое другое. Охват системы – более 100 социальных медиа, включая социальные сети, социальные закладки, блоги, форумы и многое другое.

**Hootsuite** (hootsuite.com) – многофункциональный сервис для работы с социальными медиа. Система Hootsuite позволяет работать с аккаунтами Twitter, Facebook, LinkedIn, MySpace и Foursquare, с блогами на WordPress. Сервис HootSuite является сертифицированным партнёром Twitter. Обеспечивает постинг (posting) по расписанию, возможность отслеживать сообщения по ключевым словам и упоминаниям. Система HootSuite также предоставляет полноценную интеграцию с Facebook. Система HootSuite является условно-платной, есть бесплатная версия (аналитика, 5 социальных профилей, 2 RSS/Atom ленты. Доступна на мобильных платформах: iPhone, Android, Blackberry. Все мобильные программы бесплатны.

**YouScan** (www.youscan.ru) – система мониторинга русскоязычных социальных медиа. Система YouScan отслеживает упоминания в блогах, форумах, социальных сетях (Facebook, ВКонтакте), Twitter, YouTube, и предоставляет результаты мониторинга в ана-



литическом интерфейсе с функциями одновременной работы нескольких сотрудников. Представляет отчеты по количеству сообщений с упоминаниями ключевых слов, авторов, источников, тональности. Система YouScan предлагает 5 тарифных пакетов, в том числе один бесплатный (Freemium – количество тем – 1; пользователей – 2).

**BuzzLook** (buzzlook.ru) – сервис мониторинга социальных медиа: «ВКонтакте», Facebook, LiveJournal, Flickr, YouTube и Twitter, позволяющий: следить за репутацией бренда; изучать деятельность конкурентов в сети; отвечать на вопросы клиентов в социальных сетях; собирать предложения от клиентов; поддерживать онлайн-сообщества.

**IQBuzz** (www.iqbuzz.ru) – сервис для мониторинга социальных медиа – большого количества источников и площадок, таких как LiveInternet, LiveJournal, Twitter, Яндекс.Блоги, сервисы видеохостинга RuTube и YouTube, различные новостные, развлекательные, специализированные, тематические и региональные порталы. Система обеспечивает круглосуточный мониторинг, позволяет получать информацию практически в режиме реального времени. Система IQBuzz позволяет определять тональность пользовательских сообщений, анализировать социально-демографические характеристики их авторов на основании информации из профайлов социальных сетей. Имеется возможность подключения по запросам пользователей новых источников для мониторинга.

**Socialbakers** (www.socialbakers.com) – сервис сбора статистики о работе социальных сетей, называющий себя «сердцем статистики Facebook». Система Socialbakers известна своими рейтингами брендов на Facebook, в разных категориях. Кроме Facebook сервис Socialbakers предоставляет возможность бесплатного мониторинга информации в таких сетях, как в Twitter, Google+, LinkedIn.

**SocialSeek** (socialseek.com) – простой в использовании бесплатный сервис мониторинга нескольких социальных медиа в режиме реального времени. Обеспечивает поиск в новостях, блогах, Twitter, Facebook, Youtube.

**Socialpointer** (www.socialpointer.com) – простой сервис мониторинга в социальных сетях, новостях, блогах. Имеется базовая аналитика.

**PeerIndex** (www.peerindex.net) – бесплатный сервис анализа социальных медиа, прежде всего Twitter, Facebook, LinkedIn. Определяет размеры «социального капитала» или влияния компании, профессионала, публикации и др.

**PostRank** ([www.postrank.com](http://www.postrank.com)) – сервис компании Google, позволяющий в режиме реального времени анализировать данные по темам, тенденциям, событиям, имеющим отношение к личности или бизнесу.

**Topsy** ([topsy.com](http://topsy.com)) – бесплатный сервис поиска в режиме реального времени по социальным медиа.

**HowsSciable** ([www.howsociable.com](http://www.howsociable.com)) – бесплатный инструмент мониторинга брендов и ключевых слов в 32 социальных сетях.

**Twitalyzer** ([www.twitalyzer.com](http://www.twitalyzer.com)) – аналитическая программа-клиент для Твиттера, позволяющая отслеживать количество переходов, анализировать позитивные и негативные комментарии, сегментировать аудиторию. Интегрирована с системой Google Analytics, выводит интерактивные диаграммы и графические инструменты.

**WildFire** ([monitor.wildfireapp.com](http://monitor.wildfireapp.com)) – многофункциональный онлайн-сервис для коммерческого медиа-маркетинга в социальных сетях, включающий инструмент Wildfire Messages, предназначенный для создания, мониторинга и управления сообщениями. Позволяет настроить отложенный постинг сообщений в социальные сети по расписанию. Предоставляет полноценный функционал для продвижения брендов в различных социальных сетях.

**Kurrently** ([www.kurrently.com](http://www.kurrently.com)) – бесплатная поисковая система по социальным сетям Twitter и Facebook, позволяющая отслеживать, распространять целевую информацию по социальным сетям.

**Trackur** ([www.trackur.com](http://www.trackur.com)) – коммерческий онлайн-инструмент мониторинга и анализа социальных медиа. Позволяет отслеживать репутацию брендов по новостным веб-сайтам, блогам, форумам, социальным сетям Twitter, Google+ и Facebook.

**Babkee** ([www.babkee.ru](http://www.babkee.ru)) – система мониторинга упоминаний в социальных медиа. Позволяет решать такие задачи, как оценка репутации бренда; анализ эффективности рекламных кампаний в сети Интернет; проведение маркетинговых исследований рынка, конкурентов и целевой аудитории; реагирование на обращение пользователей и их поддержка. Система позиционируется как уникальная услуга оценки значимости сообщений. Есть возможность бесплатного использования.

**Buzzware** ([www.buzzware.ru](http://www.buzzware.ru)) – сервис мониторинга социальных медиа, позволяющий исследовать мнения пользователей о брендах, которые они выражают в блогах и социальных сетях. Сервис можно использовать для репутационного анализа, изучения конкурентов, получения представлений о пользовательском опыте

и ожиданиях и, конечно же, для оценки успешности проведенных кампаний в сети Интернет.

**SemanticForce** ([www.semanticforce.net](http://www.semanticforce.net)) – сервис, обеспечивающий мониторинг неструктурированных источников – комментариев в сетевых СМИ и интернет-магазинах (рис. 38). Выдает более 20 видов аналитических отчетов. Сервис SemanticForce интегрирован с внешними системами: Klout, Copiny, GoogleAnalytics.

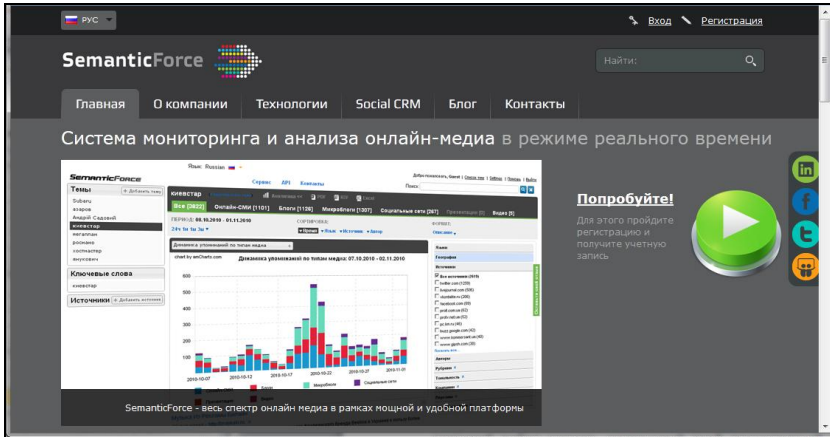


Рис. 38 – Фрагмент веб-сайта службы SemanticForce

**Brandspotter** ([www.brandspotter.ru](http://www.brandspotter.ru)) – система мониторинга социальных медиа, предлагающая стандартный набор услуг, в том числе, определение эмоциональной окраски высказываний, статистика мониторинга сообщений по темам, платформам, авторам.

**Крибрум** ([www.kribrum.ru](http://www.kribrum.ru)) – система мониторинга социальных сетей, позволяющая отслеживать и анализировать упоминания бренда, продуктов или услуг, ключевых персон, событий, географических названий. Содержит инструментарий автоматического оценивания эмоциональной окраски сообщений и построения интерактивных отчетов. Система сообщения, в которых бренд упоминается лишь вскользь. Данные начинают отображаться в системе через 2–4 часа после их публикации.

**Wobot** ([wobot.ru](http://wobot.ru)) – сервис, позволяющий проследить ретроспективу мнений в социальных сетях. Доступен широкий набор метрик, социальный граф пользователей. Обладает самообучающимся механизмом, позволяющий определять тональность сообщений.

***TweetDeck*** ([tweetdeck.com](http://tweetdeck.com)) – бесплатное кроссплатформенное приложение – инструмент для управления и отслеживания сообщений в социальных сетях Twitter, Facebook, MySpace, LinkedIn. Поддерживает многоканальный колоночный интерфейс, всевозможные фильтры, в том числе, по ключевым словам. Аналитика отсутствует.

## 5. Управление репутацией в сетях

---

### 5.1. Проблема управления репутацией компаний

Репутация представляет собой социальную оценку группы субъектов о человеке, группе людей или компании, сформировавшуюся на основе некоторых критериев.

Репутация компании – это комплекс оценочных представлений целевых аудиторий о компании, сформированный на основе факторов репутации, имеющих значение для этой аудитории.

С другой стороны, успех, финансовая отдача компании напрямую связаны с ее репутацией. Так исследование, проведенное австралийскими учеными Г. Даулингом и П. Робертсом [Roberts, 2002], выявило два преимущества благоприятного имиджа компании. Сравнив данные рейтинга 500 лучших и наиболее уважаемых компаний США, ежегодно составляемого американским журналом Fortune, за 1984–1995 годы с финансовыми показателями компаний за этот же период, ученые выявили взаимосвязь между репутацией фирмы и ее финансовым уровнем. Выяснилось, что чем выше репутация у компании, тем, во-первых, дольше период, на протяжении которого она получает максимальный доход от своей деятельности, и, во-вторых, тем меньше времени компании нужно для достижения средних по отрасли финансовых показателей при внедрении инноваций.

Репутационный капитал (Reputational Capital) – понятие не только маркетинговое, не меньшее отношение оно имеет и к финансам. В соответствии с Международными стандартами финансовой отчетности (МСФО) деловая репутация, или гудвилл (goodwill), представляет собой разницу между ценой, заплаченной за предприятие покупателями, и «справедливой стоимостью» (данная величина зачастую значительно отличается от простой стоимости всех активов фирмы). Например, в российских правилах бухгалтерского учета под репутацией понимается «разница между покупной ценой организации и стоимостью по балансу всех ее активов и обязательств».

Чтобы иметь возможность выяснять нематериальную цену компании, разрабатываются экспертные оценки репутации. Стоимость репутации может определяться экспертами, например, таким образом. Сначала рассчитывается доход, полученный компанией за счет бренда (разница между реальной прибылью и доходами, кото-

рые можно получить, продавая небрендовый товар), а затем полученная сумма умножается на специально рассчитанный коэффициент (зависящий от положения компании в отрасли, стабильности финансовых показателей и т.д.) Результат и есть цена брэнда, являющегося важной частью репутации.

Существуют и косвенные оценки уровня репутации компаний, например, основанные на результатах опроса руководителей фирм и аналитиков, оценивающих компании по таким параметрам, как качество менеджмента и продукта, способность привлечь и удерживать квалифицированные кадры, финансовая стабильность, эффективное использование активов, инвестиционная привлекательность, применение новых технологий и т.п.

Понятие «Управление репутацией в Интернете» (Online Reputation Management, ORM) по сути представляет собой комплекс мероприятий по обнаружению в сети негативного контента и сведения его к минимуму в социальных медиа и в результате поисковой выдачи. Это, своего рода, PR-кампания в киберпространстве. Ветвью ORM является SERM (Search Engine Reputation Management) – поисковое управление репутацией. На Западе такие услуги практикуются очень активно, и рост ORM в год составляет порядка 35–40 %.

Сегодня по статистике компании Google 70 % пользователей ищут отзывы о товарах и услугах, прежде чем купить их. Исторически первой компанией, которая стала практиковать двустороннюю связь с клиентами в социальных сетях, стала компания eBay. На основе обратной связи был составлен рейтинг продавцов, на который могли опираться покупатели при принятии решения о покупке. В России ярким примером отображения репутации компании, базирующейся на отзывах пользователей, можно назвать систему Яндекс.Маркет. Больше половины пользователей Интернета при выборе того или иного продукта, компании, заказчика, исполнителя и т.д. опираются на информацию, предоставленную другими пользователями.

Работы по управлению репутацией проводят как специализированные PR-агентства, работающие на просторах веб-пространства, так и подразделения SEO-агентств, которые запускают PR-кампании, направленные на поиск и устранение негативного контента. Кроме того, такие услуги предоставляют и частные лица – фрилансеры, специалисты в области интернет-маркетинга и SEO. В крупных компаниях существуют свои собственные отделы, работа которых направлена на управление репутацией фирмы, брэнда, товара, услуги.

Понятие «Управление репутацией в Интернете» (ORM) уже стало устоявшимся термином и на эти цели на Западе ежегодно выделяется часть бюджета большинства крупных компаний. Вместе с ростом влияния социальных медиа на взгляды и предпочтения людей растет и необходимость крупных компаний следить за своим имиджем в сети. На этом фоне не кажется странным рост рынка ORM на 40 % ежегодно.

Основная задача управления репутацией – формирование положительного имиджа о компании и ее продукте. Так как сложно охватить абсолютно все пользовательские отзывы и убрать весь негатив, обычно усилия концентрируются в трех областях: поисковой выдаче, отзывах в электронных СМИ и упоминаниях в социальных медиа. Приходится работать как с контентом, создаваемым редакторами различных изданий, так и простыми пользователями. Для создания целостного положительного образа информация из этих трех источников должна быть положительной или нейтральной.

Управление репутацией в поисковых системах – Search Engine Reputation Management (SERM) – комплекс мероприятий, направленных на исключение негативных отзывов о компании, товаре или услуге в результатах выдачи поисковой системы.

Услуга управления репутацией в поисковых системах необходима:

- компаниям, желающим исключить или минимизировать негативные отзывы о своей деятельности (продукции);
- компаниям, желающим сформировать положительные отзывы или увеличить их количество и видимость для целевой аудитории.

Негативная информация, наносящая вред репутации в сети, может быть различного происхождения. Условно выделяют три основные группы происхождения негативного контента [Ковальчук, 2012]:

- неумышленный негатив – это могут быть как отзывы недовольных клиентов, которые не имеют помыслов нанести вред репутации компании, а просто не удовлетворены итогами сотрудничества, так и неосторожно размещенные в Интернете фотографии с корпоративных праздников, высказывания сотрудников в адрес клиентов и т. п. Обычно такой негатив не представляет большой угрозы, но игнорировать его ни в коем случае нельзя;
- умышленный негатив с целью ударить по репутации – в этом случае классический пример – отрицательные отзывы уволенных

или уволившихся сотрудников, недовольных концепцией компании.

– черная PR-кампания – самый опасный вид негативного контента, наносящий серьезный удар по репутации. Такие PR-кампании проводят специалисты, которые тщательно изучают бизнес конкурента и точно знают, где скрыта ахиллесова пята. Организуются крупные рейдерские захваты, способные привести к полному краху не только репутацию, но и весь бизнес в целом. Данную услугу у PR-специалистов заказывают крупные серьезные конкуренты.

Самыми уязвимыми тематиками в плане притяжения негативных отзывов можно назвать:

- банки, финансовые институты;
- деятели политики и шоу-бизнеса;
- туризм, путешествия (отзывы об отелях, курортах, туроператорах, авиаперевозчиках);
- мобильная техника и связь (операторы, телефоны, электронные планшеты);
- бытовая техника;
- заведения общепита (рестораны, кафе, бары).

Соответственно, размещается и распространяется негативный контент на различных площадках:

- блоги и форумы;
- социальные сети;
- тематические веб-сайты и порталы;
- специализированные сервисы отзывов (otzyv.ru, irecommend.ru, otzovik.com и др.).

Бороться с негативным контентом призвана поисковое управление репутацией – SERM. Задача SERM состоит в вытеснении из результатов поиска веб-страниц с нежелательной информацией, в результате чего целевая аудитория перестанет видеть такие ресурсы, так как пользователи не будут выходить на них с помощью поисковых систем. Для достижения этой цели создаются материалы с позитивным контентом, предполагая, что они вытеснят негативные нежелательные сообщения. Для размещения позитивного контента (с целью вытеснения негативного) используются самые авторитетные веб-ресурсы:

- крупные новостные ресурсы;
- тематические порталы;
- отраслевые форумы;
- персональные блоги и личные сайты потребителей.

Продвигаются страницы с позитивным контентом при помощи стандартных инструментов поисковой оптимизации (Search Engine



Optimization, SEO), таких как ссылочные биржи, покупка, обмен ссылками на статьи с тематическими ресурсами, размещение анонсов, новостей и др. При этом позитивный контент следует размещать регулярно, так как негативный контент способен проявляться вновь и портить репутацию.

Управление репутацией в сети обычно начинают с мониторинга поисковой выдачи и социальных медиа с целью выявления информации по заданному объекту. Существует несколько методов мониторинга:

- ручной мониторинг поисковых систем путем ввода целевых поисковых запросов;
- использование систем оповещения, интегрированных с поисковыми системами, например, Яндекс.Блоги ([blogs.yandex.ru](http://blogs.yandex.ru)) и Google Оповещения ([google.ru/alerts](http://google.ru/alerts)). В этих случаях релевантная информация поступает на электронную почту подписчика;
- использование специальных средств мониторинга репутации компаний в социальных сетях.

В качестве пространства мониторинга для управления репутацией выбирают сетевые ресурсы, где размещаются отзывы потребителей:

- социальные сети;
- блоги и форумы;
- тематические веб-сайты и порталы;
- специальные сервисы отзывов.

Одним из критериев качества услуги мониторинга репутации является полнота охвата – доля информации об объекте, исследуемая во время работы от общего объема информации в сети об объекте. По-прежнему основным инструментом поиска информации являются традиционные поисковые системы, они охватывают значительную часть интернет-контента, а также некоторую часть социальных медиа.

Сегодня во всем мире существуют сотни систем мониторинга репутации, среди которых можно назвать системы Babkee, Brand-spotter, BuzzLook, Buzzware, IQBuzz, Крибрум, SemanticForce, Wobot, Youscan. В исследованиях Кена Барбери (Ken Burbary) и Адама Коэна (Adam Cohen) [Burbary, 2009-2013] приведен список из 230 систем мониторинга репутации, для многих из которых предлагаются бесплатные тестовые периоды для оценки качества их работы.

## 5.2. Моделирование репутации в сетях

В последнее время в рамках теории анализа социальных сетей большое внимание уделяется оценке репутации отдельных субъектов (агентов, узлов социальных сетей) и уровня доверия между ними [Расторгуев, 2006], [Губанов, 2009].

Формально социальная сеть представляет собой граф, в котором множество вершин – это совокупность агентов, субъектов – индивидуальных или коллективных, а множество ребер представляет собой совокупность отношений, совокупности социальных связей между агентами.

При моделировании социальных сетей возникает необходимость учета динамики социальных связей – взаимного влияния агентов.

Влияние в данном случае рассматривается как процесс и результат изменения индивидом (субъектом влияния) поведения другого субъекта – объекта влияния, его установок и оценок в ходе взаимодействия [Расторгуев, 2006]. Таким образом, влияние – это способность воздействовать на чьи-либо представления или действия [Новиков, 2002]. Различают направленное и ненаправленное влияние [Новиков, 2007]. Направленное влияние использует в качестве механизмов воздействия на другого человека убеждение и внушение. При этом индивид – субъект влияния – ставит перед собой задачу добиться определенных результатов от объекта влияния. Ненаправленное (нецеленаправленное, косвенное) влияние – это влияние, при котором индивид не ставит перед собой задачу добиться определенных результатов от объекта влияния.

Целенаправленное влияние участников социальной сети (или субъектов, не входящих в сеть, но использующих ее в качестве инструмента информационного воздействия) является частным случаем информационного управления, заключающегося в формировании у управляемых субъектов такой информированности, чтобы принимаемые ими на ее основе решения были наиболее выгодны для управляющего субъекта.

Возможности влияния одних участников социальной сети на других ее участников существенно зависят от репутации первых. Репутация – «создавшееся общее мнение о достоинствах или недостатках кого-либо, чего-либо, общественная оценка» [Ермаков, 2005]. Репутацию можно рассматривать как «весомость» мнения сообщества об отдельном агенте или группе агентов, определяемую его взглядами и деятельностью (активностью). При этом репутация может быть как индивидуальной, так и коллективной.

Репутация возрастает, если выбор агента (ответы на некоторые ключевые вопросы) совпадает с тем, чего от него ожидает сообщество, и понижается в противном случае.

Приведем формальное определение репутационной модели [Губанов, 2009].

Пусть  $\{a_1, a_2, \dots, a_n\}$  – множество агентов – узлов социальной сети, которые влияют друг на друга. Матрицу влияния обозначим как  $A = \|a_{ij}\|_{i=1, n}^{j=1, n}$  ( $a_{ij} \geq 0$  обозначает степень доверия  $i$ -го агента  $j$ -му). При этом очевидно, что если  $i$ -й агент влияет на  $j$ -го, а  $j$ -й влияет на  $k$ -го, то это означает следующее:  $i$ -й агент косвенно влияет на  $k$ -го (транзитивность), что позволяет строить цепочки косвенных влияний.

Предположим, что у каждого агента в начальный момент времени имеется мнение по некоторому ключевому вопросу. Пусть мнение сообщества агентов сети отражает вектор начальных мнений  $b^0$  размерности  $n$ . Мнение каждого агента меняется под влиянием мнений других агентов социальной сети.

Будем считать, что мнение  $i$ -го агента в момент времени  $t$  равно

$$b_i^t = \sum_{j=1}^n a_{ji} b_j^{t-1}$$

В [Ермаков, 2005] показано, что при многократном обмене мнениями, мнения агентов сходятся к результирующему вектору мнений  $B = \lim_{t \rightarrow \infty} b^t$ . Таким образом, справедливо соотношение  $B = Ab$ .

Обозначим  $r_i$  – параметр, описывающий репутацию  $i$ -го агента в социальной сети (сообществу), которую можно определить как нормированную сумму его влияний на всех остальных агентов социальной сети (предполагается,  $a_{ij} \geq 0$ ,  $i, j = 1, \dots, n$ ), т.е.

$$r_i = \frac{\sum_{j=1, j \neq i}^n a_{ij}}{R}, \quad j = 1, \dots, n.$$

Здесь  $R = \sum_k \sum_{j \neq k} a_{kj}$ ,  $k, j = 1, \dots, n$  – суммарное взаимное влияние друг на друга всех членов социальной сети.

В соответствии с приведенным выражением агент  $i$  имеет тем более высокую репутацию, чем выше его влияние на всех остальных членов социальной сети.

Моделирование с использованием гиперкомплексных числовых систем (ГЧС) позволяет применять развитый инструментарий из этой области математики.

В рамках модели, основанной на использовании ГЧС, каждый субъект (узел социальной сети) характеризуется своим отношением к ряду важных (ключевых) вопросов (пусть их количество равно  $N$ ). Тогда, по аналогии с [Lande, 2012], субъекту  $A$  можно поставить в соответствие гиперкомплексное число с базисом размерности  $2N$  :

$$A = e_1 w_1^+ + e_2 w_1^- + \dots + e_{2N-1} w_N^+ + e_{2N} w_N^- .$$

При этом каждому вопросу приписываются весовые значения  $w_i^+$  и  $w_i^-$ , которые соответствуют уровню положительного отношения субъекта к данному вопросу ( $w_i^+$ ) или отрицательного ( $w_i^-$ ), что является естественным расширением приведенного выше подхода. Оба значения могут быть в интервале  $[0, 1]$ , в отдельных случаях можно предположить, что  $w_i^+ + w_i^- = 1$ .

Предлагается использовать ГЧС размерности  $2N$  с базисом  $\{e_1, e_2, \dots, e_{2N}\}$  и законом умножения, который можно представить в виде таблицы:

	$e_1$	$e_2$	$e_3$	$e_4$	...	$e_{2N-1}$	$e_{2N}$
$e_1$	$e_1$	$e_2$	0	0	...	0	0
$e_2$	$e_2$	$e_1$	0	0	...	0	0
$e_3$	0	0	$e_3$	$e_4$	...	0	0
$e_4$	0	0	$e_4$	$e_3$	...	0	0
...	...	...	...	...	...	...	...
$e_{2N-1}$	0	0	0	0	0	$e_{2N-1}$	$e_{2N}$
$e_{2N}$	0	0	0	0	0	$e_{2N}$	$e_{2N-1}$

Модель субъекта социальной сети в данном случае рассматривается как гиперкомплексное число вида:

$$D = e_1 w_1^+ + e_2 w_1^- + e_3 w_2^+ + e_4 w_2^- + \dots + e_{2N-1} w_N^+ + e_{2N} w_N^-.$$

Можно рассмотреть оценку близости мнений двух субъектов  $Est(A, B)$  (соответствующими гиперкомплексными числами)

$$A = e_1 a_1^+ + e_2 a_1^- + \dots + e_{2N-1} a_N^+ + e_{2N} a_N^- \text{ и } B = e_1 b_1^+ + e_2 b_1^- + \dots + e_{2N-1} b_N^+ + e_{2N} b_N^-:$$

$$Est(A, B) = Norm\left(\frac{1}{N} \left(\sum_{i=1}^N (e_{2i-1} a_i^+ + e_{2i} a_i^-)(e_{2i-1} b_i^+ + e_{2i} b_i^-)\right)\right),$$

где  $Norm(\bullet)$  – функция нормы гиперкомплексного числа:  $Norm(e_{2i-1}) = e_1$ ,  $Norm(e_{2i}) = -e_1$ .

Отношение большей части участников социальной сети (общества) к выбранным вопросам также представляется в виде гиперкомплексного числа  $Q = q_1 e_1 + q_2 e_2 + q_3 e_3 + \dots + q_{2N} e_{2N}$ , как и отдельный субъект  $D$ . Чем больше значение  $Est(Q, D)$ , тем субъект более лояльный, «релевантный» обществу.

Можно ввести и другую оценку близости между гиперкомплексными числами, по аналогии с нормой различий между обычными векторами в векторном пространстве:

$$Est_1(A, B) = Norm\left(\frac{1}{N} \left(\sum_{i=1}^N (e_{2i-1} a_i^+ - e_{2i-1} b_i^+)^2 (e_{2i} a_i^- - e_{2i} b_i^-)^2\right)\right).$$

В этом случае субъект будет более лояльным по отношению к обществу, если оценка  $Est_1(Q, D)$  будет меньшей.

Вместе с этим, вторая оценка по содержанию менее пригодна для задач выявления взаимного влияния субъектов. Например, при сравнении отношения субъекта к обществу с запросом, даже при полностью нулевых значениях весовых коэффициентов, относящихся к значению всей социальной сети (обществу), сумма в приведенном для выражения  $Est_1(Q, D)$  не будет нулевой, т.е. полностью зависит от субъекта.

Поэтому ограничимся применением первой оценки.

Рассмотрим для примера некоторые упрощенные частные случаи, при которых анализируются отношения общества и субъекта к одному вопросу.

1. Пусть значение, соответствующее обществу, имеет вид:

$$Q = \frac{1}{2}e_1 + \frac{1}{2}e_2, \text{ т.е. отношение к выбранному вопросу в обществе}$$

может быть как позитивным, так и негативным, с равной вероятностью. Пусть отношение субъекта к этому же вопросу однозначно позитивное, а именно:  $D = e_1$ . В этом случае  $Est(Q, D) = 0$ , что соответствует полной неопределенности.

2. Пусть значение, соответствующее обществу, как и в преды-

дущем случае, имеет вид:  $Q = \frac{1}{2}e_1 + \frac{1}{2}e_2$ . Пусть отношение субъекта

к этому же вопросу также имеет вид:  $D = \frac{1}{2}e_1 + \frac{1}{2}e_2$ . В этом случае

$$Est(Q, D) = \frac{1}{4} + \frac{1}{4} - \frac{1}{2} = 0, \text{ что, как и в предыдущем случае, соответ-}$$

ствует полной неопределенности.

3. Пусть значение, соответствующее обществу, имеет вид:

$$Q = e_1, \text{ а отношение субъекта к этому же вопросу: } D = \frac{4}{5}e_1 + \frac{1}{5}e_2,$$

$$\text{тогда } Est(Q, D) = \frac{4}{5} - \frac{1}{5} = \frac{3}{5}.$$

Следует отметить, что не все нулевые элементы приведенной выше «идеальной» таблицы в реальности могут быть нулевыми, однако предполагается, что данная таблица будет разреженной. Редкие ненулевые элементы в ней могут характеризовать взаимосвязь различных вопросов.

Значения коэффициентов при базисных элементах образов субъектов социальной сети могут соответствовать вероятностям позитивного (или негативного) отношения субъектов к соответствующим вопросам. В этом случае путем перенумерации базисов таблицу умножения ГЧС можно представить в следующем виде:

	$e_1$	$e_2$	$e_3$	$e_4$	...	$e_{4N-3}$	$e_{4N-2}$	$e_{4N-1}$	$e_{4N}$
$e_1$	$B_1$								
$e_2$									
$e_3$									
$e_4$									
...									
...									
$e_{4N-3}$						$B_N$			
$e_{4N-2}$									
$e_{4N-1}$									
$e_{4N}$									

где блок  $B_1$  (а в общем случае и любой ненулевой блок) будет иметь вид, дополненный коэффициентами  $w_i^j$ , вычисляемыми в процессе обучения модели:

	$e_1$	$e_2$	$e_3$	$e_4$
$e_1$	$w_1^1 e_1$	$w_1^2 e_2$	$w_1^3 e_1$	$w_1^4 e_1$
$e_2$	$w_2^1 e_2$	$w_2^2 e_1$	$w_2^3 e_1$	$w_2^4 e_1$
$e_3$	$w_3^1 e_1$	$w_3^2 e_1$	$w_3^3 e_3$	$w_3^4 e_4$
$e_4$	$w_4^1 e_1$	$w_4^2 e_1$	$w_4^3 e_4$	$w_4^4 e_3$

При этом  $w_1^1$  – вес положительного отношения к вопросу  $t_1$ ,  $w_1^2$  – вес негативного отношения к вопросу  $t_1$ ,  $w_2^1, w_2^2$  – веса взаимосвязей наличия противоречивых одновременных положительных и негативных отношений к вопросу  $t_1$ .

Тогда для документа  $A = a_1 e_1 + a_2 e_2 + a_3 e_3 + a_4 e_4$  и документа  $B = b_1 e_1 + b_2 e_2 + b_3 e_3 + b_4 e_4$  оценка близости будет иметь следующий вид:

$$\begin{aligned} Sim(A \cdot B) = Norm & \left( \frac{1}{2} (e_1 (w_1^1 (a_1 b_1 + a_2 b_2) + w_2^1 (a_1 b_3 + a_2 b_4 + a_3 b_1 + a_4 b_2) + \right. \\ & + w_2^2 (a_1 b_4 + a_2 b_3 + a_3 b_2 + a_4 b_1)) + e_2 (w_1^2 (a_1 b_2 + a_2 b_1)) + \\ & \left. + e_3 (w_3^1 (a_3 b_3 + a_4 b_3)) + e_4 (w_3^2 (a_3 b_4 + a_4 b_4)) \right). \end{aligned}$$

При этом функция *Norm* соответствует той ГЧС, таблица которой используется для поиска. Следует учитывать, что предложенная таблица может быть составной частью таблицы большей размерности. Переход от заполненной таблицы умножения к слабозаполненной (разреженной) можно осуществить изоморфным переходом [Калиновский, 2012], что значительно сократит количество операций при вычислении функции близости между субъектами.

В общем случае уровень доверия (близость) между субъектами  $A = e_1 a_1^+ + e_2 a_1^- + \dots + e_{2N-1} a_N^+ + e_{2N} a_N^-$  и  $B = e_1 b_1^+ + e_2 b_1^- + \dots + e_{2N-1} b_N^+ + e_{2N} b_N^-$ , которую можно трактовать как степень доверия, также характеризуется функцией, приведенной выше с учетом возможного наличия ненулевых недиагональных элементов таблицы:

$$Sim(A, B) = Norm \left( \frac{1}{N} (e_1 a_1^+ + e_2 a_1^- + \dots + e_{2N-1} a_N^+ + e_{2N} a_N^-) \right) + (e_1 b_1^+ + e_2 b_1^- + \dots + e_{2N-1} b_N^+ + e_{2N} b_N^-).$$

Репутация субъекта  $A_i = e_1 a_{i1}^+ + e_2 a_{i1}^- + \dots + e_{2N-1} a_{iN}^+ + e_{2N} a_{iN}^-$  в рамках всей социальной сети (т.е. по отношению к обществу) при этом определяется, как нормированная сумма уровней доверия со всеми остальными субъектами:

$$Trust(A_i) = \frac{\sum_{j \neq i} Sim(A_i, A_j)}{\sum_{k \neq j} Sim(A_k, A_j)}.$$

Для оценки уровня взаимного влияния субъектов социальной сети могут также использоваться другие методы, среди которых можно выделить: расчет меры взаимной информации (mutual information), расчет модифицированного коэффициента Dice (modified Dice coefficient), входжение правдоподобия (log likelihood



ratio), оценку  $\chi^2$  (Chi-square test). Вместе с тем, без специальных модификаций никакой из этих методов не позволяет учитывать одновременно уровень позитивного и негативного отношения одного субъекта к ключевым вопросам, учитывать взаимную зависимость ключевых вопросов, вплоть до смысловой синонимии.

Применение модели определения репутации в социальных сетях на базе использования ГЧС может обеспечивать: возможность обучения системы, учет некоторой смысловой синонимии, омонимии в ключевых вопросах на уровне расширения таблиц умножения соответствующих ГЧС; возможность применения имеющихся наработок в области гиперкомплексных числовых систем, в том числе изоморфных ГЧС, но более пригодных для вычислений.

### **5.3. Живучесть информации в Интернете**

С проблемой управления репутацией в сети Интернет тесно связано понятие живучести информации. В свою очередь, для управления живучестью информационных объектов необходимо моделирование их жизненного цикла: формирования и развития, реакции на деструктивные воздействия, восстановления, разрушения.

Под живучестью понимают способность информационной системы (или ее части, фрагмента) адаптироваться к новым непредусмотренным условиям, противостояния нежелательным влияниям при одновременной реализации основной функции – целевого информирования. Кроме того, с живучестью информационных объектов сегодня связывают такая социально важная проблема, как обеспечение информационной безопасности.

Существует несколько механизмов, обеспечивающих живучесть информационных объектов в Интернете.

Ниже рассматриваются некоторые наиболее распространенные механизмы обеспечения живучести, которые в реальности применяются не в чистом виде, а как правило, в комбинированном.

Для изучения проблем, связанных с живучестью необходимо четко определить как само это понятие, так и привести формальную модель, на основании которой можно рассчитывать уровень живучести для таких трудно формализуемых сущностей, как информационные объекты.

### 5.3.1. Механизмы обеспечения живучести информационных объектов

Понятие живучести информационной составляющей сети Интернет подразумевает способность информационных объектов (новостных сообщений, статей, документов, видеороликов и т.д.) своевременно выполнять свои функции (информирования) в условиях действия дестабилизирующих факторов. Такими факторами могут быть устранение отдельных объектов из информационного пространства, потеря ими свойств актуальности, доступности [Додонов, 2011], [Knight, 2003]. Рассмотрим некоторые из них.

1. Копирование данных при размещении их на целевой ресурс. То есть автор размещает информацию, которая копируется хостинг-провайдером на некоторое количество зеркальных серверов. Пример – скандально известная служба WikiLeaks (несколько сотен серверов, на которых хранятся фрагменты копий).

2. Перепечатка информации (републикации, «копипаст») на другие сайты с целью их информационного наполнения. В качестве примера приводится соотношение оригинальной информации и общего объема информации, сканируемой системой InfoStream [Григорьев, 2007] за первые четыре месяца 2012 г. по дням (рис. 39). При этом следует отметить, что наиболее важная и интересная информация перепечатывается сотни раз, в то время как неактуальная, неинтересная информация практически не дублируется.

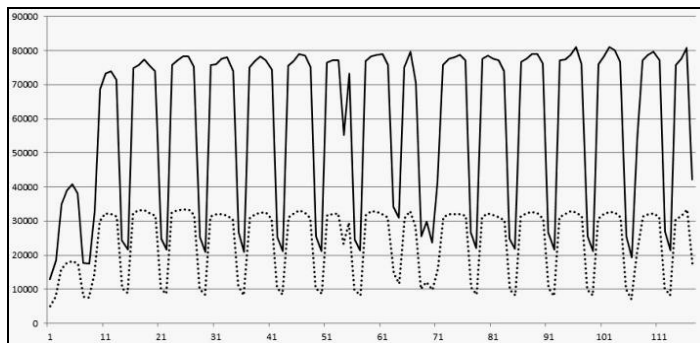


Рис. 39 – Соотношение оригинальной информации (пунктирная линия) и общего объема информации (сплошная линия), сканируемой системой InfoStream

3. Размещенная однажды информация навсегда попадает в архивные службы Интернета типа Архив Интернета (archive.org), ко-

торый накапливает сетевую информацию. Библиотека Конгресса США ([www.loc.gov](http://www.loc.gov)) купила права на хранение всех публичных сообщений социальной сети Twitter с 2006 года и всех твиттов, которые будут опубликованы впредь. Библиотека Конгресса также реализует и национальный проект сохранения и распространения цифрового контента Digital Preservation ([www.digitalpreservation.gov](http://www.digitalpreservation.gov) – 1400 коллекций данных).

4. Информация часто остается в кешах поисковых систем, даже если она удалена с веб-страницы или страницы социальной сети. Информация индексируется глобальными информационно-поисковыми системами и остается у них в кеш-памяти, откуда она доступна пользователям. Лишь относительно недавно у администраторов веб-ресурсов появилась возможность самостоятельного удаления своего контента из кешей Google и Яндекс. Часто многое, например, о человеке можно узнать в его блоге, онлайн-репутация – сегодня модный бренд. Что касается социальной сети Twitter, то [twitFlink](http://www.twitfink.com) ([www.twitfink.com](http://www.twitfink.com)), к примеру, который быстро соберет и выдаст твитты пациента. Сервис Google Replay позволяет находить и просматривать тематические сообщения в микроблогах за указанный период времени.

5. И наконец, информация с веб-сайта может сохраняться на локальных компьютерах конечных пользователей, которые получили к ней доступ либо непосредственно, либо через интеграторов информации.

### 5.3.2. Формальные модели живучести информационных объектов

Известно, что живучесть информационного объекта оценивается как вероятность того, что объект будет неповрежденным в течение определенного периода времени  $t$  при определенных условиях [Li, 2012].

Если информационный объект сохраняется на  $n$  серверах (носителях информации), то вероятность разрушения этого объекта оценивается как:

$$F_{lost}(t) = \prod_{i=1}^n F_i(t).$$

В этом произведении  $F_i(t)$  – вероятность потери информационного объекта на  $i$ -м сервере за время  $t$ .

Соответственно живучесть оценивается как:

$$S_n(t) = 1 - F_{\text{lost}}(t) = 1 - \prod_{i=1}^n F_i(t).$$

Допуская, что вероятность уничтожения информационных объектов пропорциональна времени их существования, и то, что время их разрушения имеет степенное распределение (в соответствии с законом Парето), можно считать целесообразным и обоснованным исследование модели со степенным распределением потерь информационных объектов, что принципиально отличается от подходов, в которых используется пуассоновский поток ошибок (теория систем массового обслуживания) и распределение ошибок по Вейбулу. В этом случае, живучесть можно оценивать как:

$$S_n(t) = 1 - \prod_{i=1}^n F_i(t) = 1 - \prod_{i=1}^n C t^{-\beta} = 1 - C^n t^{-n\beta},$$

где  $C$ ,  $\beta$  – некоторые константы.

Закономерности статистического распределения времени жизни информационных объектов позволяют делать выводы, связанные с их живучестью, а именно учитывать явления самого подобия, нерегулярности потери информации, наличие «тяжелого хвоста» в распределении, которое характеризует чрезвычайно большое количество фактически устаревших информационных объектов и т.п.

При анализе жизненного цикла информационного объекта можно использовать еще два больших класса моделей: булевы и марковские.

В булевой модели можно предположить, что копии информационного объекта содержатся на  $n$  серверах, при этом  $i$ -му серверу соответствует булева переменная  $x_i$ , которая может принимать значения  $\{0,1\}$ , то есть  $x_i = 1$ , если информационный объект на сервере  $i$  активен, и 0 – в противном случае. Состояние информационного объекта определяется структурной функцией его доступности (булевой функцией)  $S(x_1, x_2, \dots, x_n)$ , которая принимает значения 1, если информационный объект доступен, и 0 в противном случае.

Если доступность информационного объекта рассматривать как функцию времени, то состояние информационного объекта на  $i$ -м сервере можно рассматривать как случайный процесс  $x_i(t)$ , принимающий в произвольные моменты времени  $t \geq 0$  значения 0 и 1. Для системы определяется вероятность ее безотказной работы по приведенным выше формулам.

Среди недостатков булевых моделей можно назвать предположение только о двух состояниях информационного объекта – активности (доступности) и неактивности. Кроме того, в общем случае характер отказов отдельных копий информационного объекта зависит от состояния других копий.

Информационный объект можно описать также марковской моделью. Пусть система (множество копий информационного объекта) имеет  $m$  возможных состояний. Обозначим множество состояний через  $M = \{z_1, z_2, \dots, z_m\}$ . Для любого фиксированного момента времени  $t \geq 0$  состояние системы  $z(t)$  интерпретируется как случайная величина. Заданы множество всех состояний  $M$ , вектор распределения начальных вероятностей  $p(0)$  и функция переходных вероятностей. Определяется вероятность активности, «жизни» системы в заданный момент времени  $t$  (готовность системы). Применимость марковских моделей также имеет свои границы. Интенсивности переходов между отдельными состояниями системы могут быть нестационарными, принимаемые при расчете допущения относительно распределения интенсивности отказов могут значительно снизить точность полученных результатов; число состояний системы может быть так велико, что расчет становится практически невозможным.

Оценка живучести информационных объектов может проводиться на всех этапах их жизненного цикла. Существует несколько подходов, к проведению оценки живучести, имеющих наиболее общий характер. Живучесть можно оценить относительно некоторого стандартного внешнего воздействия либо относительно множества внешних воздействий. В этом случае решается задача нахождения множества характеристических векторов состояний информационного объекта (в простейшем случае – распределение по серверам), в которых реализуется конфигурация, обеспечивающая выполнение цели функционирования. Мощностью этого множества может служить мерой живучести всего информационного объекта.

При анализе живучести информационных объектов рассматривается проблема информирования по их различным аспектам, независимо от наличия или отсутствия неблагоприятных факторов. В связи с этим, в качестве количественного критерия оценки живучести целесообразно использовать отношение количества функций, выполняемых объектом при наличии определенных неблагоприятных воздействий либо множества таких воздействий, к общему количеству функций информационного объекта, с учетом критичности выполняемых и не выполняемых функций. Критичность

каждой конкретной функции определяется индивидуально для каждого конкретного информационного объекта исходя из его специфики. Количественный показатель живучести конкретного информационного объекта в заданных условиях можно вычислять по формуле:

$$S = \sum_{i \in \Delta} \alpha_i / \sum_{j \in \Theta} \alpha_j ,$$

где  $\Theta$  – множество всех функций информирования,  $\Delta$  – множество функций информационного объекта, выполняемых в заданных условиях ( $\Delta \subseteq \Theta$ ),  $\alpha_n$  – критичность  $n$ -й функции. Таким образом, количественная оценка живучести информационного объекта **будет изменяться** в интервале  $[0, 1]$ , живучесть тем выше, чем больше ее количественная оценка.

### 5.3.3. Цифровые следы и тени

Удаление информационного объекта с веб-ресурса не может гарантировать его исчезновения из Интернета. Остаются не только «цифровые следы» и «цифровые тени».

Выражение «цифровые следы» (Digital Footprint) относится к той информации, которая оставляется самим пользователем при работе в Сети и по которой можно не только его идентифицировать, но и «привязать» к определенным действиям, событиям, восстановить какие-то фрагменты биографии.

Часто пользователи по доброй воле указывает свои Ф.И.О., «привязывая» дальнейшую информацию к собственной личности, дату рождения, семейное положение, образование, профессию, места предыдущей работы и много чего еще, включая и контактные телефоны, и адреса электронной почты. Кроме «цифровых следов», которые пользователи оставляют сами, информация о пользователях постоянно тиражируется и без всякого его участия.

Информация о пользователе, создаваемая без его участия, получила название «цифровой тени» (Digital Shadow), которые возникают и накапливаются всякий раз, когда кто-то ищет пользователя через поисковые системы, когда происходит электронная почтовая рассылка по спискам, в которых он фигурирует и во многих других случаях. Индексация роботами поисковых машин страниц с информацией пользователя и их последующее кеширование – это тоже создание «цифровой тени», доступной каждому. Кроме «цифровых теней открытого доступа», создаются и копятся «цифровые

тени ограниченного доступа» – записи камер наблюдения, банковские транзакции, биллинги интернет-магазинов, сервисов продажи билетов, телефонных звонков и др.

По оценке аналитической компании International Data Corporation (IDC), специализирующейся на исследованиях рынка информационных технологий, объем «цифровой тени», т.е. информации о пользователе Интернет, которая создается без его участия, уже в 2007 г. превысил объем информации, которую создает сам пользователь.

С проблемой репутации в Интернете ежедневно сталкивается все больше пользователей. Об этом свидетельствует и появление особых сайтов (например, [www.suicidemachine.org](http://www.suicidemachine.org)), позволяющих одновременно удалить регистрацию и все сделанные записи на различных форумах и в социальных сетях. Такая операция называется «покончить с собой в Интернете». Однако эта система пока несовершенна. С недавнего времени эту заботу берут на себя специальные компании, так называемые «интернет-чистильщики», которые налаживают контакты с администрацией ведущих поисковых систем и социальных сетей, отдельных веб-сайтов, используют программные интерфейсы взаимодействия с кешами поисковых систем.

В качестве иллюстрации можно привести данные администрации социальной сети (сервиса микроблогов) Twitter о количестве запросов об удалении контента. По данным аналитиков, за первое полугодие 2013 года правительства разных стран направили в Twitter 1157 запросов о предоставлении информации. За аналогичный период 2012 г. эта цифра составляла 849. При этом в 10 раз выросло количество запросов об удалении контента. По числу запросов об удалении информации лидирует Россия. Кроме того, отмечается резкий рост правительственных запросов. 78 % всех запросов об информации (902) приходится на долю США. На втором месте и третьем месте находится Япония (87) и Великобритания (26).

Понятие живучести информационного объекта подразумевает его способность своевременно выполнять свои функции (в данном случае – информирования) в условиях действия дестабилизирующих факторов. Такими факторами могут быть устранение отдельных информационных объектов из информационного пространства, потеря их актуальности, доступности. Необходимо отметить, что привлечение внимания аудитории к другой теме, порождение другого информационного объекта также может снизить актуальность текущего информационного объекта.

При этом следует учитывать, что самая важная информация, попав в Интернет, остается там практически навсегда, и как показывает практика, рассчитывать на ее легкое удаление или изменение не приходится. Лучшим методом оказывается вытеснение нежелательной информации новыми сюжетами, проведение специальных мероприятий по содержательному исправлению ошибок [Додонов, 2010].

Учитывая эффект сверхживучести информации в сети Интернет, стоит учитывать несколько важных моментов, при борьбе с негативным контентом при управлении репутацией в сети:

- нельзя просто проигнорировать негативный контент; как известно, информационные сообщения, особенно негативной направленности, многократно дублируются в сети. Поэтому необходимы опровержения, позитивный контент;

- интернет-чистильщики – службы устранения негатива из сети Интернет могут «механически» лишь частично решить проблему. Негативная информация **всеравно** где-то останется и когда-то всплывет. Поэтому следует вытеснить негативный контент позитивным;

- позитивный контент должен быть правдивым, объективным. Интернет – отличный детектор лжи;

- необходимо размещать «выталкивающую негатив» позитивную информацию в сети на различных целевых ресурсах, заботясь о гиперссылках на эту информацию.

Живучесть информационных объектов и систем трудно заметить в нормальных условиях функционирования. Это свойство рельефно проявляется только в случаях потери информации, возникновения нарушений в структуре информационной системы, отказе ее составляющих, отдельных функций, целенаправленных деструктивных влияний. В зависимости от класса систем, их сложности, степени организованности, а также от выбранного уровня анализа свойство живучести может оцениваться как устойчивость, надежность, адаптивность, отказоустойчивость.

Наблюдаемый в настоящее время процесс в области интеллектуализации автоматизированных систем, перехода от простой обработки данных к процессам поддержки принятия решений требует новых подходов. Именно поэтому особое место занимают задачи, связанные с обеспечением живучести, как информационных систем, так и информационных объектов в сетевой среде.



## **6. Правовые вопросы конкурентной разведки**

### **6.1. Конкурентная разведка в правовом поле**

Безусловно, конкурентная разведка как сфера деятельности должна осуществляться в рамках правового поля государства. Основой для этого являются конституционные права на поиск, получение, передачу и использование информации во всех цивилизованных государствах. При этом следует отметить, что в ряде стран законодательство, ограничивающее деятельность по сбору и обработке информации, практически ставит конкурентную разведку под запрет.

Вместе с тем, в Российской Федерации «каждый имеет право свободно искать, получать, передавать, производить и распространять информацию любым законным способом. Перечень сведений, составляющих государственную тайну, определяется федеральным законом» (Конституция Российской Федерации, глава 2, ст. 29.4). В Украине также «каждый имеет право свободно собирать, хранить, использовать и распространять информацию устно, письменно или любым другим способом – по своему выбору» (Конституция Украины, раздел 2, ст. 34).

Таким образом, как в РФ, так и в Украине правовое регулирование в информационной сфере, к которой, безусловно, относится и конкурентная разведка, основывается на следующих принципах:

- 1) свобода поиска, получения, передачи, производства и распространения информации любым законным способом;
- 2) установление ограничений к доступу информации только законами государства;
- 3) открытость информации о деятельности государственных органов и органов местного самоуправления и свободный доступ к такой информации, кроме случаев, установленных законами государства;
- 4) по категории доступа информация подразделяется на открытую (общедоступную) и с ограниченным доступом. В свою очередь, информация с ограниченным доступом по своей правовой природе также подразделяется на два вида: сведения, составляющие государственную тайну; конфиденциальная информация.

Несмотря на то, что конкурентная разведка сегодня является признанной сферой деятельности, узаконенного понятия «конкурентная разведка» в России и Украине сегодня не существует, хотя деятельность по сбору, хранению, обработке и распространению

информации регулируется целым рядом законодательных и нормативных актов:

Закон РФ «О средствах массовой информации» от 27.12.1991. № 2124-4 (ред. от 11.07.2011), ст. 4, 40, 41, 47, 49 – 52;

Закон Украины «О печатных средствах массовой информации» от 16.11.1992 г. № 2782-XII, ст. 6, 25.

Закон РФ «О частной детективной и охранной деятельности в Российской Федерации», от 11.03.1992 г. № 2487-1 (ред. от 03.12.2011), разделы 2-3;

Закон Украины «Об охранной деятельности» № 4616-VI от 22 марта 2012 г., ст. 9, 13, 19;

Закон РФ «О коммерческой тайне» от 29.07.2004 г. № 98-ФЗ (ред. от 11.07.11);

Гражданский кодекс Украины (ст. 505), Уголовный кодекс Украины (ст. 231, 232), Кодекс Украины об административных правонарушениях (ст.163, ст.163);

Закон РФ «Об информации, информатизации и защите информации» от 27.07.2006. № 149 (ред. от 06.04.2011), ст. 10, 11, 12, 20;

Закон Украины «Об информации» от 02.10.1992 г. № 2657-XII (с изменениями от 13.01.2011 г.), ст. 5-7.

Нельзя забывать, что осуществление мероприятий по обеспечению безопасности бизнеса даже в рамках конкурентной разведки иногда может быть воспринято как проведение оперативно-розыскной деятельности, проводить которую, согласно Законам Украины «Об оперативно-розыскной деятельности» от 18.02.1992 г. № 2135-XII и Федеральному закону РФ от 12.08.1995 № 144-ФЗ могут лишь субъекты, указанные в отдельных статьях данных Законов. При этом перечень субъектов является исчерпывающим, а проводить оперативно-розыскную деятельность другими юридическими и физическими лицами запрещается.

В то же время действующими Уголовными кодексами России и Украины предусмотрено уголовную ответственность за незаконный сбор с целью использования или использование сведений, составляющих коммерческую тайну, а также за разглашение коммерческой тайны.

При достаточно широком толковании норм законодательства любые процедуры сбора, обработки и хранения информации о конкурентах становятся, с одной стороны, легитимными, практически безнаказанными, а, с другой стороны, затруднительными. В России и в Украине фактически закрыт доступ к большому пласту свободно доступной в большинстве стран бизнес-информации, например, о

недвижимости (имеющейся и заложенной), земельных участках, наличии банковских счетов и т.п. В этих странах большую часть сведений можно получить только путем консультаций с соответствующими экспертами.

При этом сегодня к деятельности компаний, занимающихся конкурентной разведкой, наблюдается повышенное внимание со стороны государственных контролирующих органов.

Это связано с несколькими группами правовых проблем, которые можно сгруппировать, выделив проблемы, связанные с:

- 1) защитой коммерческой тайны;
- 2) защитой персональных данных;
- 3) соблюдением авторских прав;
- 4) возможностью конкуренции на рынке самой конкурентной разведки.

Также можно выделить три класса основных проблем авторского права, имеющих отношение к конкурентной разведке, это проблемы, связанные с:

- правомерностью использования входной информации (источников информации), на основании которой формируются отчеты – результаты конкурентной разведки; проблемы,
- авторскими правами на результаты конкурентной разведки;
- правами на применение (использование) специализированного программного обеспечения, необходимого для проведения конкурентной разведки.

Кроме того, одна из проблем, стоящая перед службами конкурентной разведки в России и Украине – практически полное отсутствие антидемпингового законодательства. Несмотря на то, что приход на этот рынок крупных международных игроков затруднен ввиду отсутствия необходимых связей, баз данных, архивов и даже лингвистической и правовой подготовки, с их стороны возможно проявление демпинга на услуги конкурентной разведки.

Ситуация может измениться, если будет создана четкая правовая база для деятельности, связанной со сбором и аналитической обработкой информации и, в частности, для конкурентной разведки.

## **6.2. Конкурентная разведка и защита коммерческой тайны**

Важное значение для становления конкурентной разведки имел ряд статей Закона РСФСР «О конкуренции и ограничении монополистической деятельности на товарных рынках» № 948-1 от 22.03.1991, которые запрещают незаконное использование инфор-

мации, составляющей коммерческую тайну, определяют функции и права государственных антимонопольных органов, содержат санкции, применяемые к нарушителям правил добросовестной конкуренции, и т.д. В Законе Украины «О защите от недобросовестной конкуренции» № 236/96-ВР от 07.06.1996 определяется состав информации, которая вводит в заблуждение (ст. 15-1), запрещает «Неправомерный сбор коммерческой информации», «Разглашение коммерческой информации» «Неправомерное использование коммерческой информации» (гл 4, ст. 16, 17, 19, соответственно).

Федеральный закон РФ от 20 февраля 1995 г. «Об информации, информатизации и защите информации» хотя и не содержит специальных правил, посвященных охране коммерческой тайны, но включает целый ряд норм, имеющих к ней непосредственное отношение. В частности, в нем раскрываются понятия информационных ресурсов, конфиденциальной информации, защиты информации и т.д. Трудовой кодекс РФ от 30.12.2001 N 197-ФЗ допускает включение в трудовой договор условий неразглашения работником сведений, составляющих служебную или коммерческую тайну, ставших известными ему в связи с исполнением своих должностных обязанностей. В постановлении Кабинета Министров Украины от 9 августа 1993 года № 611 «О перечне сведений, которые не составляют коммерческой тайны» определен целый класс документов, касающихся деятельности бизнес-структур, которые являются фактически открытыми, в частности, учредительные документы, формы отчетности, информация об участии учредителей и должностных особ в других компаниях и т.п. Также следует выделить аналогичное постановление Правительства РСФСР от 5 декабря 1991 г. № 35 «О перечне сведений, которые не могут составлять коммерческую тайну», указывающее на те виды информации, которые не могут засекречиваться участниками информационного обмена.

Зачастую усилия конкурентной разведки направлены на получение коммерческой тайны конкурентов. И хотя в различных законодательных актах даются различные формулировки, можно согласиться с тем [Иващенко, 2006], что коммерческая тайна характеризуется такой совокупностью признаков: информация является секретной, является неизвестной и не является легкодоступной для лиц, которые обычно имеют дело с видом информации, к которой она относится; в связи с тем, что является секретной, она имеет коммерческую ценность. Таким образом, коммерческая тайна – это информация, которая является полезной и не является общеизвестной обществу. Она имеет действительную или коммерческую ценность, с которой можно иметь прибыль и для защиты которой

владелец принимает меры во всех сферах жизни и деятельности». Таким образом, можно сказать, что деятельность конкурентной разведки иногда направлена на добычу информации, которая не является общедоступной и охраняется законом. Эти деяния нарушают огромное количество статей Уголовного Кодекса Российской Федерации, в частности, статью 183 «Незаконное получение и разглашение сведений, составляющих коммерческую, налоговую или банковскую тайну» и Уголовного Кодекса Украины, в частности, статью 231 «Незаконный сбор с целью использования или использование сведений, составляющих коммерческую или банковскую тайну».

Таким образом, коммерческая разведка может легитимно использовать лишь те методы и способы сбора и обработки информации, которые не противоречат законодательству, т.е. основные функции конкурентной разведки — качественный сбор, систематизация и, главное, анализ информации, а не слежка, подкупы и незаконные хакерские взломы.

Впервые право на сохранение коммерческой тайны было провозглашено Законом СССР от 4 июня 1990 г. «О предприятиях в СССР». В ст. 33 указанного Закона раскрывалось понятие коммерческой тайны как не являющихся государственными секретами сведений, связанных с производством, технологической информацией, управлением, финансами и другой деятельностью предприятий, разглашение (передача, утечка) которых может нанести ущерб их интересам.

Статья 2 Закона РСФСР от 24 декабря 1990 г. «О собственности в РСФСР» упоминала о секретах производства (ноу-хау, торговые секреты) как одном из признаваемых законом объектов интеллектуальной собственности.

Чуть позже Законом РСФСР от 22 марта 1991 г. «О конкуренции и ограничении монополистической деятельности на товарных рынках» получение, использование, разглашение научно-технической, производственной или торговой информации, в том числе коммерческой тайны, без согласия ее владельца было квалифицировано в качестве одной из форм недобросовестной конкуренции (ст. 10).

В настоящее время российское и украинское законодательство об охране служебной и коммерческой тайны представляет собой совокупность статей, которые содержатся в различных правовых актах, посвященных в целом регулированию иных общественных отношений.

Закон РФ «О частной детективной и охранной деятельности в РФ» от 21 мая 1993 г. и Закон Украины «Об охранной деятельности» № 4616-VI от 22 марта 2012 г. содержат специальные нормы, возлагающие на работников соответствующих структур обязанность по неразглашению сведений, ставших им известными в связи с выполнением служебных обязанностей и составляющих коммерческую тайну предприятий и организаций.

Гражданский кодекс РФ определяет коммерческую тайну как информацию, имеющую действительную или потенциальную коммерческую ценность в силу ее неизвестности третьим лицам, к которой нет свободного доступа на законном основании и по отношению к которой обладатель информации принимает меры к охране ее конфиденциальности (п. 1 ст. 139).

Гражданский кодекс Украины, в свою очередь, определяет коммерческую тайну (с. 505 п. 1) как информацию, «которая является секретной в том смысле, что она в общем или в определенной форме и совокупности э неизвестной и не является легкодоступной для лиц, которые обычно имеют дело с видом информации, к которому она относится, в связи с этим имеет коммерческую ценность и была предметом адекватных существующим обстоятельствам мер, относящихся к сохранению ее секретности, предпринятых особой, которая законно контролирует эту информацию».

В соответствии с этими определениями, как только информация в результате каких-либо действий попадает, например, на страницы любого веб-сайта, она перестает считаться коммерческой тайной, так как становится легкодоступной.

Хотя во многих статьях Уголовного кодекса Украины (ст. 231, 232, 232-1, 361, 363) установлена уголовная ответственность как за разглашение коммерческой тайны, так и за незаконный сбор и использование сведений, к ней относящихся, однако, существующая нормативно-правовая база четко не регламентирует, какие именно сведения о финансово-хозяйственной деятельности предприятия являются коммерческой тайной (за исключением разве что банковской тайны, определение которой дано в ст. 60 Закона Украины «О банках и банковской деятельности»).

### **6.3. Конкурентная разведка и защита персональных данных**

Государственные учреждения, банки, крупные корпорации не всегда могут обеспечить защиту хранящихся у них баз персональных данных, в результате чего, огромный поток конфиденциальной

информации поступает на рынок. Обеспечение безопасности персональных данных – объективная потребность. Сегодня персональные данные, информация о людях превращается в самый дорогой товар. Такая информация в руках злоумышленника – мощное оружие. То есть персональные данные необходимо защищать.

Персональные данные – важная составляющая более широкого понятия – приватность. Поэтому защита персональных данных, это составляющая часть обеспечения приватности. Приватность, наряду со свободой слова и другими правами, является одной из основных ценностей человечества.

На сегодня основными европейскими документами в области защиты персональных данных являются Конвенция Совета Европы «О защите личности в связи с автоматической обработкой персональных данных» и Директива Европарламента «О защите физических лиц при автоматизированной обработке персональных данных», ETS № 108, 1981 г., которая является обязательной для всех государств-членов Европейского союза и которая является предметом для подражания в области законодательства, в том числе странами Восточной Европы. Страны Евросоюза последовательно приводят свое законодательство в соответствие с Директивой. В Великобритании еще в 1998 году был принят «Закон о защите персональных данных» – «Data Protection Act 1998». Его техническая реализация – проект стандарта «Specification for the management of personal information in compliance with the Data Protection Act 1998» (BS 10012, 2009). Параллельно с англичанами свою версию стандарта по безопасности персональных данных выпустили в США. Проект документа по защите персональных данных для американских государственных структур – «Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)» (SP 800122) регламентирует выполнение Законов «The Privacy Act of 1974» и «Privacy Protection Act of 1980». Канада выпустила «Privacy Code» – набор документов для реализации законодательства по защите сведений о частных лицах (The Privacy Act и PIPEDA).

В государствах-членах Евросоюза определения персональных данных, как правило, максимально широкие, в результате чего гражданами на практике зачастую не выполняется соответствующее законодательство из-за излишней «нагрузки». Соответствующие органы государственной власти, как правило, не предпринимают никаких действий, кроме особых случаев. Важными остаются вопросы возникновения коллизий между требованиями приватности и интересами свободы слова. Современными европейскими законами, как правило, запрещается сбор, хранение, использова-

ние и распространение без согласия субъекта данных именно критичных персональных данных.

Право на приватность гарантируется Конституциями РФ и Украины. Статья 23 Конституции Российской Федерации гласит «1. Каждый имеет право на неприкосновенность частной жизни, личную и семейную тайну, защиту своей чести и доброго имени. 2. Каждый имеет право на тайну переписки, телефонных переговоров, почтовых, телеграфных и иных сообщений. Ограничение этого права допускается только на основании судебного решения»; в статье 32 Конституции Украины говорится: «Никто не может подвергаться вмешательству в его личную и семейную жизнь, кроме случаев, предусмотренных Конституцией Украины». Кроме того в Конституции Украины предусмотрена защита еще некоторых аспектов приватности. Так, статья 30 защищает неприкосновенность жилища (территориальная приватность), статья 31 – тайну переписки, телефонных разговоров, телеграфной и другой корреспонденции (коммуникационная приватность), статья 32 предусматривает запрет сбора, хранения, использования и распространения конфиденциальной информации о лице без его согласия (информационная приватность), а статья 28 предусматривает запрет подвергать лицо без его свободного согласия медицинским, научным или другим исследованиям (защищая некоторые элементы физической приватности). Аналогично, статья 24 Конституции РФ определяет, что «сбор, хранение, использование и распространение информации о частной жизни лица без его согласия не допускаются».

Конституционные нормы определяют исчерпывающий перечень оснований для вмешательства в приватность и условий для такого вмешательства. Однако в постсоветских государствах существует много отраслевых норм права, противоречащих требованиям их Конституций. Именно такие нормы не соответствуют международным стандартам, практике Европейского законодательства.

В соответствии с российским и украинским законодательством персональными данными в России и в Украине является Ф.И.О. в сопровождении любых других идентификационных данных, например, адреса, телефона или образовательного статуса.

Для выяснения, какое же отношение имеет физическое лицо или компания к защите персональных данных, большое значение имеет определение субъектов отношений, связанных с персональными данными (статья 4 Закона Украины № 2297-VI): «Субъектами отношений, связанными с персональными данными, являются:

- субъект персональных данных;
- владелец базы персональных данных;



- распорядитель базы персональных данных;
- третье лицо;
- уполномоченный государственный орган по вопросам защиты персональных данных;
- другие органы государственной власти и органы местного самоуправления, к полномочиям которых относится осуществление защиты персональных данных».

В России введен новый специальный субъект – «оператор». По-видимому, он соответствует одновременно «владельцам» и «распорядителям» из украинского Закона.

Как в российском, так и в украинском законодательствах предусмотрен уведомительный характер обработки персональных данных. Владелец или распорядитель (оператор) до начала обработки персональных данных обязан уведомить уполномоченный орган по защите прав субъектов персональных данных о своем намерении осуществлять обработку персональных данных. Затем данные о владельцах или распорядителях (операторах) вносятся в специальный реестр операторов. Информация, содержащаяся в реестре операторов, становится общедоступной.

Законы о персональных данных касаются большинства населения как участников процесса «обработки» данных. А так как субъектом персональных данных является каждый человек, то Закон носит всеобщий характер и касается каждого.

Оба законодательных акта (украинский и российский) имеют прямое отношение к сфере информационных технологий и телекоммуникаций, оба содержат спорные, противоречащие сложившейся практике, казалось бы неосуществимые положения. Требования закона распространяются на все юридические и физические лица, и интернет-сфера не является исключением. Закон о защите персональных данных может изменить принципы работы украинских интернет-ресурсов: сервисов электронной почты, знакомств, онлайн-магазинов и социальных сетей, хотя сами участники рынка надеются, что сайты не подпадут под действие закона. Владельцам интернет-ресурсов для соблюдения всех положений закона о персональных данных необходимо тщательно образом продумывать организацию своей деятельности. В настоящее время существует немало веб-служб, в рамках которых происходит сбор, хранение, использование персональных данных. Соблюдение требований закона является непростой задачей для владельцев этих ресурсов, в частности, чиновники имеют возможность обязать интернет-компанию брать письменное согласие на использование анкетных данных у каждого пользователя. Не секрет, что на многих сайтах

размещается информация, содержащая персональные данные людей (например, ищущих работу, знакомства), в том числе и относящиеся к специальным категориям, например, национальность или вероисповедание. Задача тех, кто обеспечивает подобные сервисы, легитимно обрабатывать подобную информацию и одновременно защищать ее согласно требованиям законодательства.

В частности, персональные данные широко используются в социальных сетях и сервисах электронной почты. Например, владельцам веб-ресурсов весьма сложно соблюсти требование закона о получении согласия каждого пользователя на обработку его персональных данных. При этом закон возлагает именно на оператора обязанность доказывания факта получения им такого согласия.

Современная интернет-компания собирает и обрабатывает разные категории персональных данных – своих сотрудников, своих контрагентов по договорам и некоторые данные пользователей своих сервисов. Люди, размещающие информацию о себе в социальных сетях или службах знакомств, сознательно делают ее открытой для всех пользователей ресурса, и по закону ее можно трактовать как «общедоступную», а значит, соблюдения особого режима конфиденциальности в отношении ее не требуется, но в социальных сетях есть и информация, которую пользователь скрывает, делая ее доступной только для отдельной группы пользователей («друзей»). В этом случае интернет-ресурс должен предусматривать для нее специальные средства защиты.

В практике конкурентной разведки приходится сталкиваться с многочисленными противоречиями и казусами в существующем законодательстве, например, в украинском Законе «О защите персональных данных» (часть 9 ст. 6) говорится: «использование персональных данных в исторических, статистических или научных целях может осуществляться только в обезличенном виде». То есть записи в отчетах конкурентной разведки должны выглядеть примерно так: «Персона А провела переговоры с персоной Б». В научных отчетах нельзя делать ссылок на других коллег, даже при наличии их письменного согласия. Вызывает определенные сложности и необходимость оповещать орган власти «о каждом изменении сведений, необходимых для регистрации соответствующей базы», которая среди прочего включает информацию обо всех распорядителях (пользователях) такой базы данных.

Кроме того, многие службы конкурентной разведки, совершенно на законных условиях создающие базу данных персональных данных для решения обозначенной ими задачи, обязаны уничтожить плоды своей работы, достигнув цели. А ведь, если основная

цель, например, при оказании услуг клиентам – это выполнение этих самых заявок, но сопутствующая цель любой уважающей себя организации – это и наработка базы клиентов. И эта база часто имеет собственное коммерческое значение. Известны многочисленные случаи легальной перепродажи баз данных клиентов, например, при прекращении деятельности фирмы-владельца. Однако, например, в российском Федеральном законе определено (ст. 21): «в случае достижения цели обработки персональных данных оператор обязан незамедлительно прекратить обработку персональных данных и уничтожить соответствующие персональные данные в срок, не превышающий трех рабочих дней с даты достижения цели обработки персональных данных». В украинском законодательстве такой строгой статьи нет, однако также предусмотрены условия уничтожения персональных данных, среди которых (ст. 15), «прекращение правоотношений между субъектом персональных данных и владельцем или распорядителем базы...». А это означает, например, что оператор – исполнитель услуги должен уничтожить всю наработанную за время выполнения услуги базу данных.

Поэтому владельцы и распорядители подобных баз данных переформулируют свои цели специальным образом, например, как «оказание услуги с возможностью хранения персональных данных в течение гарантийного срока...». Таким образом, соблюдаются нормы законодательства и обеспечиваются интересы исполнителя – владельца или распорядителя (оператора) базы персональных данных.

Подразделения конкурентной разведки занимаются обработкой персональных данных, которые находятся в открытых источниках в сети Интернет, т.е. являются общедоступными. Для их обработки согласия субъекта персональных данных не требуется. Однако при этом обязанность доказательства, что обрабатываемые персональные данные являются общедоступными, возлагается на владельца или распорядителя. А это значит, что необходимо либо накапливать доказательства того, что данные взяты из общедоступных источников, либо получать согласие от субъекта персональных данных и затем хранить этот документ. Кроме того, нужно иметь документ, подтверждающий общедоступность источника персональных данных. При этом остается без ответа вопрос доказательства того, что владелец информационного ресурса (веб-сайта) обладает письменным согласием на обработку.

Как никогда острой стала проблема криминализации отдельных служб конкурентной разведки. Многие службы безопасности

сегодня пользуются базами данных с информацией о персонах. Такие базы используются с вполне благими целями, например, для проверки данных о сотрудниках, партнерах и конкурентах. Очевидно, такими базами данных они будут пользоваться и в дальнейшем, однако будут вынуждены нарушать закон и «уходить в подполье». Технически возможности использования и ведения подобных баз данных предоставляют многочисленные системы типа Cronos (оболочки, распространяемые вполне легально). С помощью подобных инструментальных средств любому заинтересованному пользователю Интернет становятся доступны многочисленные базы данных, которые работают под этими оболочками.

При этом владельцам и распорядителям персональных баз данных сегодня экономически выгоднее выполнить требования законодательства и не подвергать свой бизнес возможным рискам со стороны контролирующих органов и недобросовестной конкуренции. Очевидно, реализовать требования законодательства отдельным организациям будет достаточно сложно. В этом случае на помощь им могут прийти специализированные организации, компании-интеграторы, работающие на рынке информационной безопасности.

## 7. Информационные операции

---

---

В последние годы благодаря многочисленным документам и публикациям Министерства обороны США стал популярен термин «информационные операции», прежде всего потому, что информационные технологии играют постоянно увеличивающуюся роль в военных операциях. При этом информационные операции определяются как «акции, направленные на воздействие на информацию и информационные системы противника, и защиту собственной информации и информационных систем» [DoD, 2003]. Информационные операции рассматриваются как объединение основных возможностей радиоэлектронной войны, компьютерных сетевых операций, психологических операций, военных действий и операций по обеспечению безопасности с целью воздействовать, разрушать, исказить информацию, необходимую для принятия противником решений, а также защищать собственную информацию.

Информационные операции охватывают целый комплекс процессов, проводимых в самых разных областях. При этом необходимо отметить, что информационные операции – существенная и традиционная составляющая боевых операций. Несмотря на то, что формальное определение в документах Департамента обороны США ориентировано на военные аспекты информационных операций, оно вполне применимо практически для любой области жизни.

Ниже будут рассматриваться такие информационные операции, которые реализуются с помощью информационных систем (ИС). Живучесть этих ИС во многом определяет живучесть информационных операций, которые реализуются в виде информационных воздействий на сознание людей.

Информация является отражением вложенного в нее смысла, поэтому сегодня информация превратилась из абстрактного термина в объект, цель и средство информационных операций, стала критическим понятием в проблематике безопасности. Бывший министр обороны США Уильям Коэн 18 марта 1999 г. заявил, что «способность армии использовать информацию, чтобы доминировать в будущих сражениях, даст США новый ключ к победам в течение многих лет, если не в течение нескольких поколений» [Hill, 2000].

При моделировании и проведении информационных операций необходимо учитывать значение ценности информации для лиц, принимающих решение. Ценность информации включает ее своевременность, точность и «аналитичность». С практической точ-

ки зрения ценность информации также может быть определена как ее значимость или применимость, пригодность к использованию. Под применимостью информации понимается обеспечение доступа ЛПР к готовой к использованию информации. Стандарт ISO 9241 (ISO – Международная Организация по Стандартизации) определяет применимость в терминах эффективности и удовлетворения потребностей указанного набора пользователей для решения указанного набора задач в специфическом окружении. На практике большая часть полезной информации поступает к ЛПР от информационно-аналитических систем, обеспечивающих ориентацию в ситуации и поддержку при принятии решений. Согласно полемому уставу военного ведомства США «Информационные операции» (FM 100-6), «ориентация в ситуации означает комбинацию ясного представления о диспозиции своих и вражеских сил с оценкой ситуации и намерений со стороны командования».

Информационные операции осуществляются в некоторой социальной среде, соответственно, для успешного их проведения необходимо адаптироваться к этой среде, преодолеть определенный барьер не очень сильного внимания к информационному воздействию. Этот барьер возникает благодаря так называемой иммунной системе среды, которая может не пропустить информационные воздействия, если она достаточно мощная и/или уже научилась защищаться от подобных воздействий. К подготовительным действиям для проведения информационных операций может относиться создание «иммунодефицита» социальной среды путем воздействия через информационное пространство, например, с помощью материалов в СМИ. Очень часто информационные воздействия используют механизмы «вирусного маркетинга», например, в виде слухов, когда сенсационно поданная дезинформация распространяется с огромной скоростью. Именно иммунная система оказывает противодействие подобным информационным операциям. Очень часто с иммунной системой общества отождествляют государство, призванное обеспечивать безопасность этого общества, т.е. при наличии сильного государственного аппарата вероятность успеха антиобщественных информационных операций существенно снижается. Читатель прекрасно знает, как происходило противодействие подобным информационным процессам в тоталитарных государствах. В демократическом обществе, естественно, тоталитарные методы не применимы. В этом случае иммунитет достигается за счет «обучения», т.е. демократическое общество должно пройти через многие информационные атаки, воздействия, влияния стереотипов, чтобы выработать необходимый иммунитет.

Уровень готовности к проведению информационных операций сегодня считается ключевым фактором успеха проведения любой социальной процедуры, кампании.

Особой целью при проведении информационных операций являются информационно-аналитические системы субъекта воздействия. Оказывая влияния на такие системы, можно добиться того, что принимающие решение лица из лагеря противника примут неадекватные выводы, и требуемый социальный процесс изменит траекторию в необходимом оказывающей влияние стороне направлении [Горбулін, 2009] (рис. 40).

В данном случае к непосредственным информационным воздействиям может быть отнесено размещение в информационном пространстве документов, компрометирующих противоположную сторону, реклама (в том числе скрытая) своих преимуществ, искаженные данные о внешней среде, искаженная информация о намерениях и т.д.

Социальные процедуры и процессы, как правило, сложно оценивать и моделировать, так как их результаты относятся к психологическим и социологическим, а не физическим. Именно этот факт также определяет проблематичность прогнозирования результатов моделирования информационных операций. Кроме того, экспериментирование с информационными воздействиями в рамках информационных операций более сложны и опасны, чем при моделировании физических процессов.

Действия для достижения эффективности влияния на процессы принятия решения противником иногда необходимо предпринимать в течение длительного времени, прежде чем они вступят в силу.

Одна из основных компонент информационных операций – социальное влияние, охватывающее все многообразие процессов влияния. Существенные изменения в убеждениях или отношении

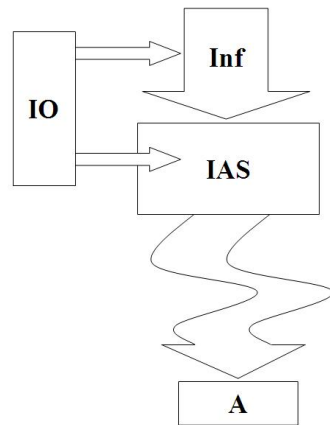


Рис. 40 – Воздействие на информационно-аналитическую систему противника: Inf – информационное пространство; IAS – информационно-аналитическая система; А – абонент системы – ЛПР; IO – информационные воздействия

людей к некоторой проблеме или явлению, как ожидается, будут вести к изменению в поведении, связанном с этой проблемой.

В 1948 году Харольд Д. Лассвелл [Lasswell, 1948] разработал модель трансмиссии коммуникаций, состоящую из пяти компонент:

- источник – персона, которая влияет или убеждает другие персоны;
- сообщение – с помощью чего источник пробует убедить цель;
- цель – человек, на которого источник пробует влиять;
- канал – метод доставки сообщений;
- воздействие – реакция цели на сообщение.

Хотя Лассвелл прежде всего интересовался массовой коммуникацией, его модель передачи информации может применяться в межличностной коммуникации типа циркулярных моделей Шеннона–Вивера (Shannon–Weaver) и Осгуда–Шрамма (Osgood–Schramm), которые включают петли обратной связи в процессе коммуникаций, утверждая, что коммуникация является циркулярным, а не линейным процессом [Schramm, 1974], [Osgood, 1954].

Моделирование объективных факторов социального влияния требует междисциплинарных подходов, имеющих отношение к информатике, маркетингу, политологии, социальной психологии. Самые известные модели формирования общественного мнения и социального влияния базируются на теории Латэйна динамического социального воздействия [Latane, 1981], [Latane, 1997], развитой многими другими авторами, прежде всего, в работах [Nowak, 1990], [Lewenstein, 1993], [Kasperski, 2000], [Sobkowicz, 2003].

Пытаясь обосновать механизм социального влияния сообщений Латэйн [Latane, 1981] подчеркнул важность трех признаков отношений источника и цели:

- сила – социальная сила, вероятность или уровень влияния на индивидуумов;
- непосредственность – физическое или психологическое расстояние между индивидуумами;
- число источников – количество источников, стремящихся к цели.

Современное состояние моделирования информационных операций характеризуется рядом открытых проблем, основные из которых относятся к пониманию понятий информационного влияния и воздействия.



## 7.1. Информационное влияние

Универсальными характеристиками объектов являются его состояние и возможность воздействия на другие объекты. Реализация возможности воздействия требует определенных условий, которые принято называть его влиянием. При этом объект, который может осуществлять свою волю, называют субъектом, а управлением принято называть воздействие по отношению к объекту воздействия, применяемое с определенной целью.

Когда индивидуум является целью влияния одного или более источников, динамическая социальная теория воздействия утверждает, что уровень социального влияния на индивидуума может быть представлен уравнением, являющимся основой так называемой индивидуум-ориентированной модели:

$$I_i = -S_i\beta - \sum_{j=1, j \neq i}^N \frac{S_j O_j O_i}{d_{i,j}^\alpha},$$

где  $I_i$  — величина (количество) социального давления, оказываемого на индивидуума  $i$ , ( $-\infty < I_i < \infty$ );  $O_i$  и  $O_j$  представляет мнение индивидуума ( $i$  и  $j$ , соответственно) по актуальному вопросу — +1 или -1 — поддержку или возражение относительно данного вопроса, соответственно.  $S_i$  ( $S_j$ ) представляет силу индивида  $i$  ( $j$ ) или влияние ( $S_i > 0$ ,  $S_j > 0$ );  $\beta$  — сопротивление индивидуума к изменениям ( $\beta > 0$ );  $d_{i,j}^\alpha$  — расстояние между индивидуумами  $i$  и  $j$  ( $d_{i,j}^\alpha \geq 1$ );  $\alpha$  — показатель сокращения расстояния ( $\alpha \geq 2$ );  $N$  — общее количество агентов (индивидуумов, составляющих сообщество). Значение  $\beta$ , тенденция сохранять собственное мнение или сопротивляться изменению определяет то, что индивидуумы в рамках модели могут требовать больших или меньших объемов социального давления для изменения их мнения. Большие уровни значения  $\alpha$  соответствуют эффекту возрастания расстояния между источником и целью, что влияет на объем социального давления на цель.

На основе введенных терминов формулируется понятие «информационного поля объекта» [Кононов, 2003], описываются его характеристики. Это дает возможность определить информационное воздействие как воздействие на информационное поле объекта.

Исследуя информационные поля объектов и субъектов социальных систем, можно определить информационные влияния и управления. При этом информация может рассматриваться и как объект, и как средство воздействия. Использование информации как средства воздействия требует в процессе управления осуществить подготовку данных, производство соответствующей информации, а лишь затем реализовывать созданную информацию в виде воздействия (влияния).

Одним из основных методов ведения информационных операций является информационное влияние, оказываемое с целью информационного управления. Под информационным управлением в данном случае понимается механизм управления, когда управляющее воздействие носит неявный, косвенный информационный характер и объекту управления дается определенная информационная картина, под влиянием которой он формирует линию своего поведения. Таким образом, информационное управление — это способ воздействия, побуждающий людей к упорядоченному поведению, выполнению требуемых действий.

В соответствии с [Кононов, 2003], [Кульба, 2004] процесс информационного влияния одного объекта на другие целесообразно декомпозировать на следующие этапы:

- генерация источником влияния данных, информационных элементов и информационных совокупностей;
- передача информации источником влияния;
- прием информации реципиентом;
- генерация совокупности данных, информационных элементов и новых совокупностей объекта влияния;
- соответствующие активные действия объекта влияния.

Информационные воздействия на элементы систем можно классифицировать по таким признакам, как источники возникновения, длительность воздействия, природа возникновения и т.п.

Для выбора конкретных способов реализации информационного управления необходимо конкретизировать задачи, решаемые с помощью информационного воздействия, провести анализ процесса формирования информационных операций и выработать критерии их оценки. Информационное управление рассматривают как процесс, охватывающий такие три взаимосвязанных направления:

- управление обменом данными между реальным миром и виртуальным миром субъекта влияния;
- управление виртуальным миром субъектов влияния, механизмами принятия решений;

– управление процессом преобразования решений в действия субъектом влияния в реальном мире.

Информационное воздействие может быть двух основных видов:

1) изменение в требуемую сторону данных, которые использует информационно-аналитическая система объекта воздействия при принятии решений;

2) непосредственное влияние на процесс принятия решения объекта воздействия, например, на процедуры принятия решения или отдельные лица, принимающие решения.

Важнейшее значение для проведения информационных операций имеет окружающая среда, состояние объектов информационного воздействия, их взаимное влияние. В частности, если в качестве объектов информационных операций выбирается некоторое электоральное поле, то важно учитывать все электоральные группы, входящие в это поле, которые представляют сторонников (или противников) тех или иных политических сил. Несмотря на то, что в дальнейшем будут рассматриваться и некоторые модели, в которых в явном виде постулируется однородность среды, в общем случае по отношению к информационным операциям окружающая среда может состоять из областей:

- доминирующего восприятия;
- повышенной чувствительности;
- индифферентности к соответствующим информационным воздействиям.

## 7.2. Этапность информационных операций

Остановимся отдельно на этапности информационных операций. Очевидно, не существует единственного «стандартного» плана проведения как наступательных, так и оборонительных информационных операций. Можно лишь рассмотреть примерную, полученную путем обобщения некоторых уже реализованных информационных операций последовательность действий при их осуществлении.

На практике информационная операция как процесс информационного воздействия на массовое сознание, как правило, реализуется следующим образом: в результате предварительной разведки вырабатывается план следующего этапа — оперативного управления и намечаются соответствующие мероприятия оперативной разведки, которые являются приближенной моделью решения, после чего реализуется оперативное управление противником. На

этапе оперативной разведки определяется уровень отклонения первоначальной модели от реальности, и если оно незначительно, то реализуется первоначальный план. В противном случае строится новый план оперативного управления и управления противником. Далее цикл повторяется до тех пор, пока оперативная разведка не подтвердит используемую модель. При этом окончательное решение принимается с определенным оперативным риском.

Таким образом, процесс информационного воздействия охватывает такие основные этапы [Чхартишвили, 2004] (рис. 41):

- предварительная разведка (preliminary intelligence, PI);
- выявление текущей обстановки, состояния противника (Op);
- управление противником (management of enemy, M) (информационное воздействие на противника с целью передачи ему сведений соответствующих замыслу управляющего);
- оперативная разведка (operational intelligence, OI) (проверка результатов рефлексивного управления);
- оперативное управление (operational management, OM) – действия управляющего для достижения требуемой цели.

При планировании или моделировании социальных процессов, в частности информационных операций, всегда необходимо учитывать, что общее поведение социальных систем невозможно определить, оперируя исключительно рафинированными математическими моделями. Это главным образом обусловлено тем, что такие процессы в большой степени зависят от социально-психологических факторов.

Различают два основных типа информационных операций – наступательные и оборонительные. Однако, на практике, большая часть информационных операций является смешанной. Кроме того, большинство процедур информационных операций относятся одновременно к наступательным и оборонительным. Каждый из типов информационных операций, включая приведенные выше основные этапы, подразумевает некоторые особенности и уточнения.

Особенностью наступательных информационных операций (информа-

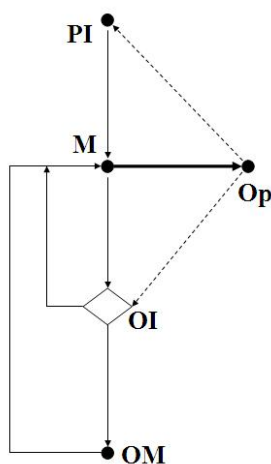


Рис. 41 – Основные этапы информационных операций

ционных атак) является то, что объекты воздействия таких операций определены и планирование основывается на достаточно точной информации об этих объектах. Информационная атака чаще всего требует нахождения или создания информационного повода (для оборонительных информационных операций поводом может являться сама информационная атака противника), раскрытка этого повода, т.е. пропаганда (в отличие от мер контрпропаганды при оборонительных информационных операциях), а также необходимость принятия мер по препятствию информационному противодействию.

Таким образом, план типовой информационной операции включает совпадающие на верхнем уровне для информационных операций обеих типов такие этапы, как оценка, планирование, исполнение и завершающая фаза. Приведем более детальный перечень компонент информационных операций.

В наступательных информационных операциях можно выделить такие основные фазы:

1. Оценка необходимости проведения операции:
  - 1) определение цели, прогноз достижимости, степени влияния;
  - 2) сбор информации.
2. Планирование.
3. Исполнение информационного воздействия:
  - 1) нахождение или создание информационного повода;
  - 2) раскрытка информационного повода (пропаганда);
  - 3) оперативная разведка;
  - 4) оценка воздействия;
  - 5) препятствие информационному противодействию;
  - 6) корректировка информационного воздействия.
4. Завершающая фаза:
  - 1) анализ эффективности;
  - 2) использование позитивных результатов информационного воздействия;
  - 3) противодействие отрицательным результатам.

Типовая оборонительная информационная информация охватывает такие основные этапы:

1. Оценка:
  - 1) анализ возможных уязвимостей (целей);
  - 2) сбор информации о возможных операциях;
  - 3) определение возможных «заказчиков» информационных воздействий;

- определение сфер общих интересов объекта и потенциальных «заказчиков»;
- ранжирование потенциальных заказчиков по их интересам.

2. Планирование:

1) стратегическое планирование оборонительной операции (явное или неявное):

- определение критериев информационных воздействий;
- моделирование информационных воздействий с учетом: связей объекта; динамики воздействия; «особых» (критичных) точек воздействия;
- прогнозирование следующих шагов;
- расчет последствий.

2) тактическое планирование контропераций.

3. Исполнение — отражение информационного воздействия:

- 1) выявление и «сглаживание» информационного повода;
- 2) контрпропаганда;
- 3) оперативная разведка;
- 4) оценка информационной среды;
- 5) корректировка информационного противодействия.

4. Завершающая фаза:

- 1) анализ эффективности;
- 2) использование позитивных результатов информационного воздействия;
- 3) противодействие отрицательным результатам.

Оперативное управление информационными операциями с использованием информационно-аналитических систем можно проиллюстрировать с помощью диаграммы, представленной на рис. 42.

В соответствии с приведенной диаграммой информация из реального мира (R) поступает в информационное пространство, в частности, в средства массовой информации (I) либо непосредственно экспертам (E), также через средства

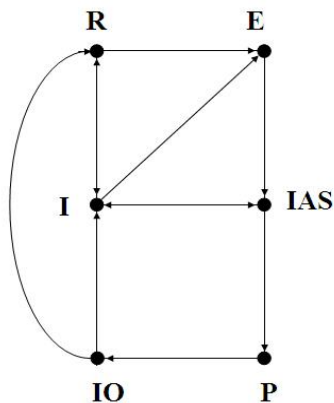


Рис. 42 – Диаграмма оперативного управления с использованием информационно-аналитических систем

массовой информации. От экспертов или непосредственно из информационного пространства (например, с помощью средств контент-мониторинга) информация поступает в информационно-аналитическую систему (IAS). Информационно-аналитическая система передает лицам, принимающим решения (Р), данные, которые определяют меры информационного воздействия на информационное пространство и непосредственно на объекты реального мира (людей, окружающую среду, компьютерные системы и т. д.).

### 7.3. Моделирование информационных операций

Моделирование можно рассматривать как один из способов решения проблем, возникающих в реальном мире, в частности, при планировании и проведении информационных операций. Чаще всего моделирование применяется в случаях, если эксперименты с реальными объектами невозможны, либо слишком затратные. Моделирование охватывает отображение реальной проблемы в мир абстракции, изучение, анализ и оптимизацию модели, и отображение оптимального решения обратно в реальный мир.

При моделировании существуют два альтернативных подхода — аналитическое и имитационное моделирование. Идеальные аналитические модели допускают строгое аналитическое решение или, по меньшей мере, постановку, например в виде систем дифференциальных уравнений. Однако, аналитические решения не всегда достижимы. Поэтому, особенно в последнее время, и особенно при решении задач из области социальной динамики все чаще применяются методы имитационного моделирования (*Simulation Modeling*). Имитационное моделирование представляет собой более мощное и практически незаменимое средство анализа социальных процедур. Имитационную модель можно рассматривать как множество правил, определяющих будущее состояние системы на основании текущего. При этом процесс моделирования заключается в наблюдении эволюции системы во времени по данным правилам, и, соответственно, оценки адекватности модели, когда это возможно.

Наиболее перспективным направлением моделирования информационных операций является математическое описание самоорганизации среды восприятия и распространения информации с учетом сложившихся в текущий момент условий. Самоорганизующиеся среды, для которых отсутствует центральный механизм управления, а развитие идет за счет множества локальных взаимодействий, изучаются теорией сложных систем. Эта теория охваты-

вает такие отрасли знаний, как нелинейная физика, термодинамика неравновесных процессов, теория динамических систем. Взаимодействия между отдельными элементами сложных систем определяют возникновение сложного поведения при отсутствии централизованного управления. Для исследований подобного поведения применяются самые современные методы, которые охватываются междисциплинарной основой современной методологии — концепцией сложности. В настоящее время к теоретическим и технологическим основам этой концепции относятся теории детерминированного хаоса, фракталов и сложных сетей, синергетика, волновой (вейвлет) анализ, многоагентное моделирование, теория самоорганизованной критичности (изучающей динамическое развитие до критического состояния, характеризуемого сильными пространственно-временными флуктуациями, без внешнего управления [Bak, 1996]), теория перколяции (Percolation – протекание) и т.п.

Моделирование социальных процедур (информационные операции, безусловно, относятся к таковым) предполагает проведение вычислительных экспериментов, так как чаще всего возникают существенные ограничения, затрудняющие проведение «полевых» натуральных экспериментов.

При моделировании информационных операций вычислительный эксперимент позволяет сократить операции по уточнению ограничений, подбору исходных данных, выбору правил функционирования компонент модели и т.д. В этом случае появляется возможность учета случаев, трудно реализуемых на практике, используя реальные данные лишь для идентификации параметров математической модели. Вместе с тем математическое моделирование имеет свои ограничения, реальный мир оказывается сложным для моделирования с достаточным уровнем детализации и точности, т.е. более или менее достоверные математические модели настолько сложны и многопараметричны, что не поддаются анализу и оценкам точными методами.

Отработать математические модели при планировании информационных операций можно лишь в процессе моделирования конкретных процедур, постоянно сопоставляя их с реальностью.

Выраженная цель методологии оценки информационных операций состоит в том, чтобы обеспечить своевременный и точный анализ возможных несоответствий между запланированной операцией и фактическим воздействием. Когда обнаруживаются существенные различия, которые влияют на вероятности успеха операции, аналитическая система должна сообщать об этом лицам, принимающим решения, для того, чтобы откорректировать текущие



планы и решения. Вместе с тем, при планировании информационных операций нельзя действовать методом проб и ошибок, поэтому необходимо развивать методы, позволяющие обобщать ретроспективные данные, и на их основе проверять адекватность моделей.

В основу успешных моделей информационных операций закладываются синергетические подходы. Действительно, общество является сложной системой, каждая компонента которой характеризуется множеством признаков, имеет множество степеней свободы. При этом важным свойством этой системы является самоорганизация, которая является результатом взаимодействия таких компонент, как случайность, многократность, положительная и отрицательная обратная связь.

Особенностью математического моделирования информационных операций следует считать сравнительную простоту интерпретации получаемых результатов. Такие понятия, как «численность электората», «политический вес» и т.д., воспринимаются на интуитивном уровне даже без знакомства с точными (насколько они тут возможны) определениями. А это позволяет делать подобный анализ актуальных ситуаций предметом широкого обсуждения.

В силу того, что некоторые решения являются неустойчивыми по отношению к своим параметрам, значения таких параметров необходимо определять с высокой точностью. Для этого требуется комплекс методик, основанных не только на обработке больших объемов статистических данных, но и на разносторонних социологических исследованиях.

В настоящее время реалистичной выглядит постановка задачи, состоящая в использовании математических моделей для прогнозирования возможных сценариев динамики социальных процессов на качественном уровне. В такой формулировке моделирование динамики занимает как бы промежуточный уровень между тем, что изложено здесь, и точным прогнозированием. И все же потребуются выбор значений параметров, которые бы в некотором разумном приближении соответствовали изучаемой ситуации, причем в большинстве случаев продуктивным оказывается использование относительных величин. Так, конечно, не получить достоверных данных о будущем развитии событий, но, скорее всего, можно составить более или менее адекватную картину того, что и как может произойти. А это уже не мало.

Для достижения успеха при этом отдельные информационные воздействия необходимо рассматривать как части единой информационной операции, точно так же, как артобстрел или авиацион-

ные атаки можно рассматривать как согласованные части военной операции.

При этом информационным операциям присущи такие основные особенности:

- информационные операции – это междисциплинарный набор методов и технологий в таких областях, как информатика, социология, психология, международные отношения, коммуникации, военная наука;

- до сих пор не существует стандартов проведения информационных операций;

- в развитии технологий информационных операций заинтересованы не только оборонные ведомства, но и многие правительственные и коммерческие организации;

- задача формирования научного подхода к информационным операциям является насущной и актуальной.

При проведении информационных операций существенно выявление содержания (знаний), вкладываемого в информацию, с учетом самых разнообразных аспектов – социальных, политических, религиозных, исторических, экономических, психологических, ментальных, культурных, присущих различным слоям общества. Поэтому в настоящее время имеет смысл рассматривать информационные операции шире, как операции, базирующиеся на знаниях (Knowledge Operations) [Burke, 2001].

Обычная сетевая информационная атака в веб-среде сегодня производится следующим образом: как правило, создается и некоторое время функционирует веб-сайт (назовем его «первоисточником»), при этом он публикует вполне корректную информацию. В час X на его странице появляется документ, обычно компромат на объект атаки, достоверный либо сфальсифицированный. Затем происходит так называемая «отмывка информации». Документ перепечатывают интернет-издания двух типов – заинтересованные в атаке и те, кому попросту не хватает информации для заполнения своего информационного поля. В случае претензий все перепечатающие издания ссылаются на «первоисточник» и, в крайнем случае, по просьбе/требованию объекта атаки удаляют со своих веб-сайтов информацию. Первоисточник при необходимости также снимает информацию либо вовсе ликвидируется (после чего оказывается, что он зарегистрирован в Интернет на несуществующее лицо). Вместе с тем информация уже разошлась, задача первоисточника выполнена, атака стартовала.

Современное информационное пространство представляет собой уникальную возможность получения любой информации по

выбранному вопросу при условии наличия соответствующего инструментария, применение которого позволяет анализировать взаимосвязь возможных событий или событий, которые уже происходят, с информационной активностью определенного круга источников информации. С другой стороны, при ретроспективном анализе любого процесса или явления интерес представляют определенные характеристики его развития, а именно:

- количественная динамика, присущая процессу или явлению, например, количество событий в единицу времени, или количество сообщений, имеющих отношение к нему;

- определение критических, пороговых точек, которые соответствуют количественной динамике явления;

- определение проявлений в критических точках, например, выявления основных сюжетов публикаций в СМИ относительно выбранного процесса или явления;

- после выявления основных проявлений явления в критических точках, эти проявления ранжируются, и исследуется динамика развития отдельных определенных проявлений до и после определенных критических точек;

- осуществляется статистический, корреляционный и фрактальный анализ общей динамики и динамики отдельных проявлений, на основе которых осуществляются попытки прогнозирования развития явления и отдельных его проявлений.

Для исследования взаимосвязи реальных событий и публикаций о них в сети Интернет авторами использовалась система InfoStream, обеспечивающая интеграцию и мониторинг сетевых информационных ресурсов.

Количество веб-публикаций в день по какой-либо теме, а особенно изменения (динамика) этой величины порой позволяют даже небольшим специалистам в предметной области делать более-менее точные выводы.

Получить данные подобной динамики можно, например, ежедневно заходя на сайты интеграторов новостей ([news.yandex.ru](http://news.yandex.ru), [webground.su](http://webground.su), [uaport.net](http://uaport.net)). Конечно, в лучшем положении пользователи профессиональных систем мониторинга типа Интегрум или InfoStream. Именно на основе последней системы получена удивительная статистика по количеству веб-публикаций по тематике эпидемий гриппа в разные периоды.

В качестве примера рассмотрим информационную кампанию, направленную против «Проминвестбанка», которая началась в конце сентября 2008 г.

С помощью системы контент-мониторинга InfoStream ([www.infostream.ua](http://www.infostream.ua)) [Григорьев, 2007], сканирующей все основные информационные веб-сайты Украины в режиме реального времени, была определена динамика публикаций на веб-сайтах сообщений, в которых упоминался «Проминвестбанк» за три месяца – сентябрь, октябрь и ноябрь (рис. 43). Эта динамика свидетельствует о небольшом количестве публикаций за первую половину сентября, однако затем пошел ряд публикаций, компрометирующих председателя правления В. Матвиенко, что вызвало относительно небольшой резонанс.

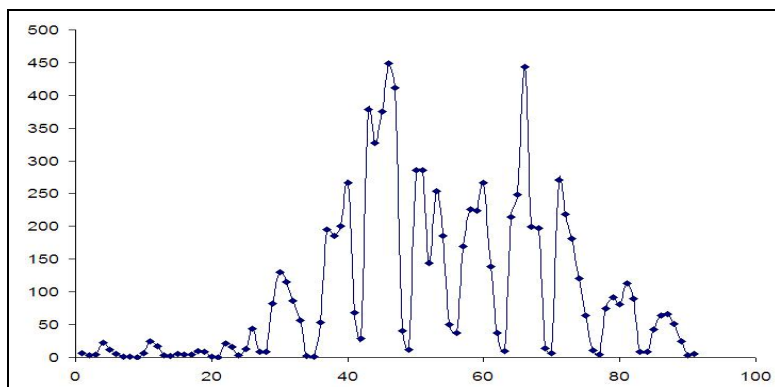


Рис. 43 – Динамика публикаций по теме «Проминвестбанк» за три месяца 2008 г.

Как оказалось впоследствии, эти публикации были лишь «арт-подготовкой». 26 сентября появились первые сообщения о возможном банкротстве банка (рис. 44), количество которых вполне соответствовало лавинообразному процессу, ограниченному лишь числом веб-сайтов, способных публиковать подобную информацию. Впрочем, этот процесс вышел на стабильно-средний уровень к декабрю 2008 г.

Нельзя утверждать, что лишь информационная атака через сеть Интернет привела банк к печальному состоянию, однако именно первые тревожные сообщения подорвали доверие многих вкладчиков, заставили их массово забирать свои сбережения из банка.

30 сентября появилось сообщение, что для спасения Проминвестбанка Национальный Банк Украины (НБУ) решил выделить

ему 5 млрд. гривен рефинансирования, а 5 декабря появилось сообщение, что у «Проминвестбанка» появился новый владелец (рис. 45). После этого объемы публикаций о «Проминвестбанке» существенно сократились, что свидетельствует не столько об его оздоровлении, сколько о системном кризисе банковской системы Украины, «уронившему» многие другие кредитные и банковские учреждения.

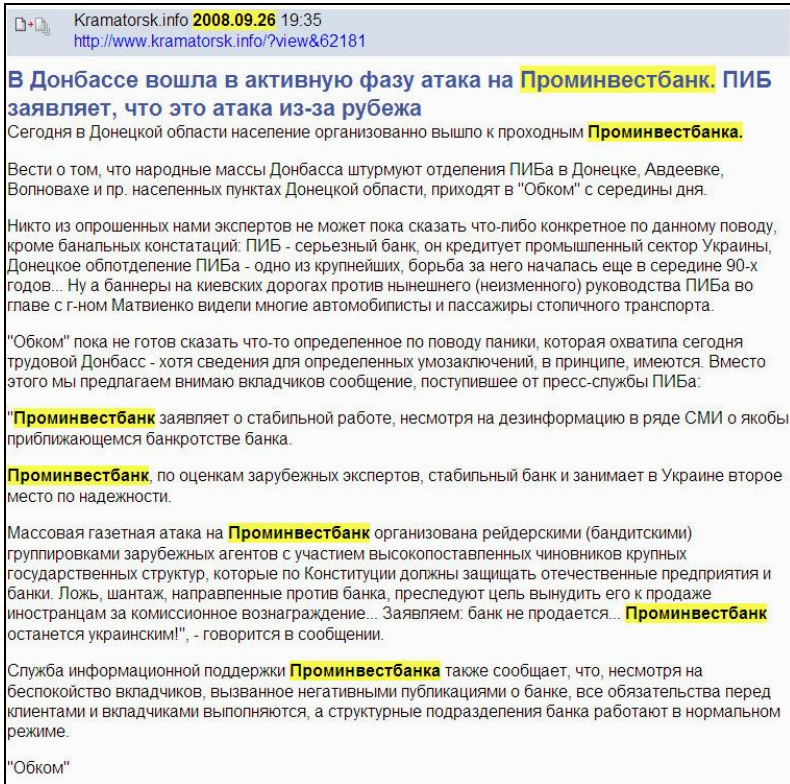


Рис. 44 – Одно из первых тревожных сообщений

Буквально через неделю после описанных выше событий в Украине произошла еще одна публичная знаковая информационная атака, в этот раз на рынке страхования. Это была настоящая информационная операция против Национальной акционерной страховой компании (НАСК) «Оранта». В этом случае первоисточником компромата оказался не веб-сайт, а информационное сообщение,

разосланное электронной почтой тысячам пользователей Интернет. В результате применения специальных технических приемов, оно разошлось с обозначением адреса пресс-службы объекта атаки. И так, 10 декабря 2008 года в районе 11:30 в виде спама было разослано информационное сообщение, в котором говорилось о том, что страховая компания «Оранта» заявляет о банкротстве. По предварительным данным, информация разлетелась по 1000 адресам, естественно, данные попали к конкурентам и в СМИ. В сообщении говорилось, что компания с 31 декабря 2008 года прекращает выполнять взятые перед клиентами обязательства.

The screenshot shows the InfoStream Online search engine interface. The search query is "(проминвестбанк)&(2008.11.05)". The search results are displayed in a list format, with the first seven items visible. Each item includes a title, a date, and a brief description of the news article. On the right side, there is a sidebar with various filters and a classification system.

**Активная база данных:** Система интеграции интернет-ресурсов

**Главная** | Кабинет | Источники | Статистика | Новости проекта

Вход | Выход

InfoStream Online

Поиск: (проминвестбанк)&(2008.11.05)

Найти | Динамика | Действие

Очистить | События | Сохранить

Язык запросов: Примеры

От: 200809 | До: 200811

Найдено документов - 443, страница 1 из 30

Статистика слов

**ПРОМИНВЕСТБАНК** - 26463, 2008.11.05 - 58209

Добавить канал

**Информационный портрет**

Уточнить запрос

Языки (1)

Страны источников (1)

Источники (1)

Размер (1)

Цифровая насыщенность (1)

География (1)

Компании (1)

Слова (12)

Классификатор-навигатор

ОБЕКОМ  
ВОЛНОВАХА  
АТАКА  
ФАЗА  
РУБЕЖ  
ВКЛАДЧИК  
ДОНБАСС  
ГОТ  
ВОЛНОВАХА  
АТАКА  
ФАЗА  
РУБЕЖ  
ВКЛАДЧИК  
ДОНБАСС  
УМОЗАКЛЮЧЕНИЕ  
ВОЛНОВАХА  
АТАКА  
ФАЗА  
РУБЕЖ  
ВКЛАДЧИК  
ДОНБАСС  
ПРОХОДНАЯ  
ВОЛНОВАХА  
АТАКА  
ФАЗА  
РУБЕЖ

- Матвиенко поделился "Проминвестбанком" с братьями Клюевыми**  
 Политбюро 2008.11.05 22:08  
 40% акций "Проминвестбанка" досталось братьям Клюевым. Эту информацию газете "Сегодня" подтвердил источник в руководстве Партии Регионов. По словам информатора, денег за это братья Клюевы не заплатили.  
 Похожие документы - Оригинал
- ВЧЕРА ОФИЦИАЛЬНЫЙ И РЫНОЧНЫЙ КУРСЫ ДОЛЛАРА ПОЧТИ СРАВНИЛИСЬ, И В ОБМЕННИКАХ АМЕРИКАНСКУЮ ВАЛЮТУ ПРОДАВАЛИ ПО 5,85 ГРИВНИ**  
 Газета "Факты и комментарии" 2008.11.05 21:30  
 А "Проминвестбанк" обрел новых владельца Роберт ВАСИЛЬ "ФАКТЫ" Национальный банк в среду продолжил свою деятельность по укреплению курса гривны на наличном и межбанковском рынке и параллельно по ослаблению официального курса американской валюты. Курс доллара, установленный Нацбанком, вчера вырос приблизительно на 3,5 копейки и достиг значения 5,8261 гривны за доллар.  
 Похожие документы - Оригинал
- "Проминвестбанк" сменил собственника**  
 Газета "День" 2008.11.05 21:22  
 Правительство утверждает, что не будет расходовать деньги налогоплательщиков на рекапитализацию банка Наталья БИЛОУСОВА, "День" Вчера об этом официально сообщил на своем сайте Национальный банк Украины (НБУ).  
 Похожие документы - Оригинал
- Кабмин утвердил правила для капитализации банков**  
 УРА-Информ 2008.11.05 21:13  
 Кабинет министров Украины утвердил порядок участия государства в капитализации банков. Соответствующее постановление от 4 ноября 2008 г. N 960 размещено на сайте правительства.  
 Похожие документы - Оригинал
- Дмитрий Фирташ покупает почти 90% банка "Надра"**  
 Вовремя.info 2008.11.05 20:17  
 Вчера, 4 октября, украинский предприниматель Дмитрий Фирташ, совладелец швейцарского газового трейдера RosUkrEnergy, подписал предварительное соглашение о покупке 86,7% акций банка "Надра".  
 Похожие документы - Оригинал
- У Проминвестбанка поменялся владелец (5.11.2008 18:00)**  
 INTV (рус.) 2008.11.05 19:45  
 new У Вас есть видео, которое Вы хотите показать всему миру? Вам сюда В Проминвестбанка изменился владелец. Факт продажи акций банка подтвердили в НБУ. СТВ  
 Похожие документы - Оригинал
- Население будет покупать доллары по официальному курсу**  
 АМИ Новости-Украина 2008.11.05 19:45  
 Национальный банк Украины своим постановлением N353 от 5 ноября обязал коммерческие банки продавать населению наличные доллары по курсу не выше официального, сообщает "Українські новини".  
 Похожие документы - Оригинал

Рис. 45 – Сообщения, завершившие экстремальную динамику интенсивности публикаций по теме «Проминвестбанк»

В связи со случившимся НАСК «Оранта» обратилась в правоохранительные органы с просьбой расследовать данный инцидент и наказать виновных. Произошедшее с НАСК «Оранта» очень напоминало ситуацию с «Проминвестбанком», с этим согласились многочисленные эксперты. Ведь как банковский бизнес, так и страховой основываются на доверии клиентов, которое легче всего поддается именно информационными атаками. По словам Олега Спилки, председателя наблюдательного совета НАСК «Оранта», «Это мероприятие готовилось целенаправленно для того, чтобы дискредитировать страховую компанию и подорвать ее репутацию». Не вдаваясь в детали возможных целей атаки (смена владельцев, борьба за блокирующий пакет акций, уничтожение компании и т.п.), с помощью ретроспективного анализа проследим за динамикой публикаций в сети Интернет, в которых упоминалась НАСК «Оранта».

На рис. 46 приведена посуточная динамика количества соответствующих публикаций. На этой диаграмме, кроме всего прочего, отчетливо виден спад интенсивности публикаций по данной теме в начале декабря 2008 г., что вполне можно воспринимать как некоторое «затишье перед бурей».

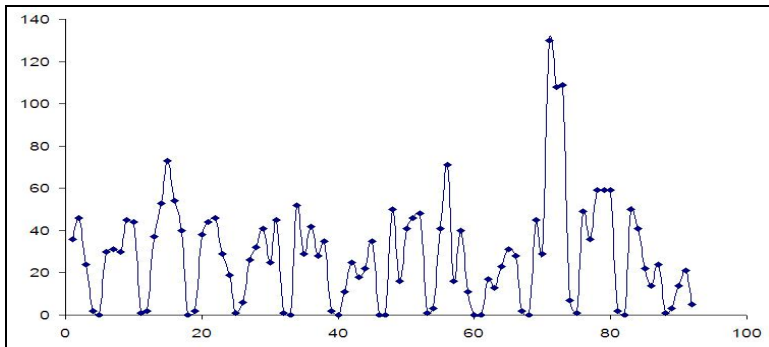


Рис. 46 – Интенсивность публикаций в Интернет по теме «Оранта»

Для анализа временных рядов в рамках исследования авторов применялся  $\Delta L$ -метод. На рис. 47 представлена скейлограмма динамики рассматриваемого процесса с помощью метода ( $\Delta L$ -метода) за второе полугодие 2008 года. Несмотря на отдельные пики в 16 и 55 день квартала, все же наибольший интерес представляет экстремум, приходящийся именно на 10–12 декабря.

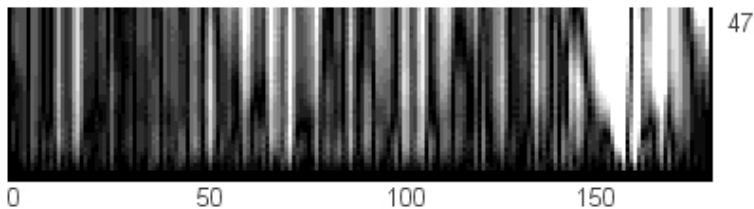


Рис. 47 –  $\Delta L$  -диаграмма ряда публикаций по теме «Оранта»

Более детальная статистика публикаций по теме «Оранта» за декабрь 2008 года получена через интерфейс пользователя системы контент-мониторинга InfoStream (рис. 48).

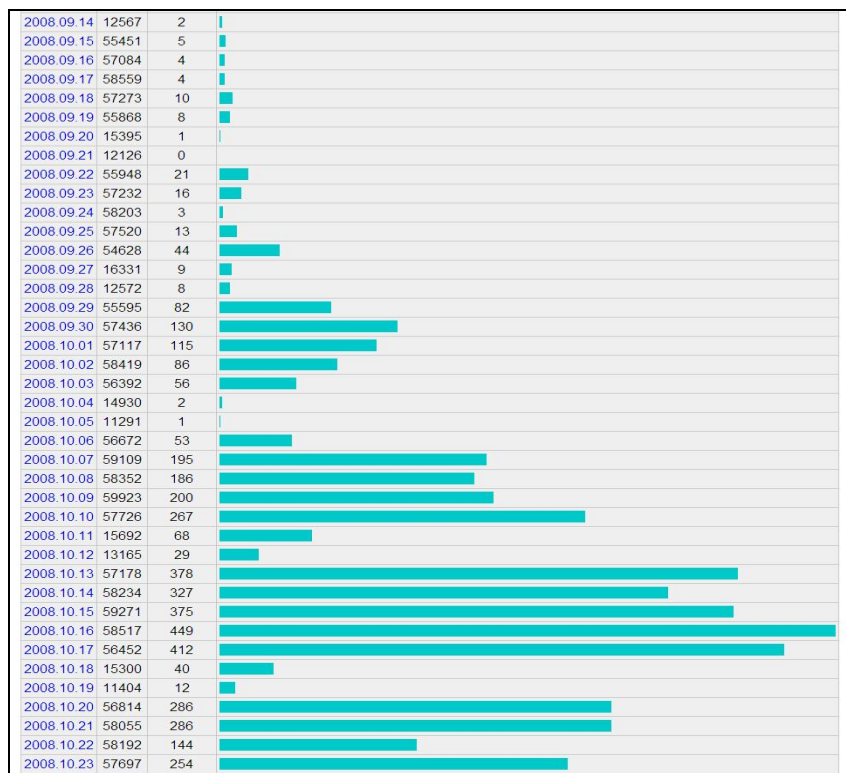


Рис. 48 – Детальная диаграмма интенсивности публикаций по теме «Оранта»



Проследим за ходом информационной операции, рассматривая сообщения, публикуемые в разные промежутки времени.

На рис. 49 приведен список публикаций по теме «Оранта» в течение первых часов атаки. По словам Олега Спилки, в течение двух часов с начала атаки все почтовые серверы НАСК «Оранта» были выведены из строя, поэтому опровержение в сети задержалось.

Информационный портрет	
Уточнить запрос	
Рубрики (1)	
Языки (1)	
Страны источников (2)	
1. <b>Крупнейшая страховая компания Украины заявила о своем банкротстве</b> PRO-test 2008.12.10 16:56 Крупнейшая страховая компания классического страхового рынка Украины НАСК "Оранта" заявила о своем банкротстве. Об этом говорится в письме компании, поступившем в адрес редакции. В связи с действием обстоятельств непреодолимой силы, национальная страховая компания Оранта уведомляет всех своих клиентов о невозможности выполнения взятых на себя обязательств после 31 декабря 2008 года и ограниченным выполнением обязательств по случаям, наступившим и (или) наступающим в будущем. Похожие документы - Оригинал	<input type="checkbox"/> AND <input type="checkbox"/> OR <input type="checkbox"/> NOT <input type="checkbox"/> ИтерМедиа Консалтинг <input type="checkbox"/> УРА-Информ Донбасс <input type="checkbox"/> УРА-Информ <input type="checkbox"/> Экономические новости <input type="checkbox"/> Украина криминальная <input type="checkbox"/> Донецкий коммуникационный ресурс <input type="checkbox"/> Новости Луганска <input type="checkbox"/> ОБКОМ <input type="checkbox"/> Попыткаконтроль <input type="checkbox"/> 24 UA <input type="checkbox"/> УРА-Информ Харьков <input type="checkbox"/> ОЛИГАРХ.NET <input type="checkbox"/> Zaxid.net <input type="checkbox"/> "Портал эксклюзивных новостей" <input type="checkbox"/> Страхование в России <input type="checkbox"/> "Дедап" <input type="checkbox"/> Громадський Спротив України <input type="checkbox"/> Перший Діловий <input type="checkbox"/> PRO-test <input type="checkbox"/> Ракурс плюс <input type="checkbox"/> Sxid.info <input type="checkbox"/> Новости N
2. <b>Страховая компания "Оранта" объявила о банкротстве</b> ИтерМедиа Консалтинг 2008.12.10 15:29 В связи с обстоятельствами, которые нельзя преодолеть, национальная страховая компания "Оранта" сообщает всем своим клиентам о невозможности выполнения взятых на себя обязательств после 31 декабря 2008 года и ограниченном выполнении обязательств по случаям, которые наступили и (или) наступают с 1 сентября по 31 декабря 2008 года. Похожие документы - Оригинал	
3. <b>"Оранта" говорит, что не объявляла о банкротстве</b> ИтерМедиа Консалтинг 2008.12.10 15:29 Сегодня от имени руководителя пресс-службы наблюдательного совета НАСК "Оранта" Елены Кулаковой на множество Интернет-адресов было направлено SPAM-сообщение о банкротстве крупнейшей в стране страховой компании "Оранта". Похожие документы - Оригинал	
4. <b>Страховая катастрофа: украинская "Оранта" обанкротилась</b> "Дедап" 2008.12.10 15:16 Крупнейший классический страховщик Украины объявил о своей несостоятельности. Об этом говорится в пресс-релизе, обнародованном украинской компанией. Приводим текст заявления полностью: "Уважаемые клиенты!" Похожие документы - Оригинал	
5. <b>Страховая катастрофа: украинская Оранта обанкротилась</b> Страхование в России 2008.12.10 15:06 Крупнейший классический страховщик Украины объявил о своей несостоятельности. Об этом говорится в пресс-релизе, обнародованном украинской компанией. Приводим текст заявления полностью: Уважаемые клиенты! Похожие документы - Оригинал	
6. <b>"Оранта" обанкротилась</b> "Портал эксклюзивных новостей" 2008.12.10 14:39 Страховая компания "Оранта" обанкротилась. Издание приводит текст пресс-релиза полностью: "Уважаемые клиенты! В связи с действием обстоятельств непреодолимой силы, национальная страховая компания Оранта уведомляет всех своих клиентов о невозможности выполнения взятых на себя обязательств после 31 декабря 2008 года и ограниченном выполнении обязательств по случаям, наступившим и (или) наступающим с 1 сентября по 31 декабря 2008 года" Похожие документы - Оригинал	
Рубрикатор (2)	
Цифровая насыщенность (1)	

Рис. 49 – Первые часы атаки. Самые «оперативные» источники

В 12:31 на сайте «Экономические новости» появляется странное «обновленное» сообщение с парадоксальным последним предложением (рис. 50).

Далее руководство НАСК «Оранта» опубликовало в Интернете первые опровержения, не спеша обвинять конкурентов в происшедшем, а затем все же признав атаку целенаправленной и выгодной третьим лицам.

На рис. 51 приведен список публикаций, посвященных опровержению сообщения о банкротстве за следующий день (11 декабря), а также наиболее активных источников, опубликовавших эти сообщения. Безусловный интерес аналитиков вызывает сравнение источников, приведенных на рис. 49 и 51.

Экономические новости 2008.12.10 12:31  
<http://economic-ua.com/articles/46840>

### Страховая компания "Оранта" стала банкротом (обновлено)

В Интернете появились сообщения о том, что страховая компания "Оранта" стала банкротом.

"Уважаемые клиенты!  
 В связи с действием обстоятельств непреодолимой силы, национальная страховая компания **Оранта** уведомляет всех своих клиентов о невозможности выполнения взятых на себя обязательств после 31 декабря 2008 года и ограниченным выполнением обязательств по случаям, наступившим и (или) наступающим с 1 сентября по 31 декабря 2008 года.  
 В связи с начатой процедурой **банкротства** действие всех страховых полисов ограничивается сроком до 31 декабря 2008 года, вне зависимости от даты, указанной в договоре.  
 Страховые возмещения по случаям, наступившим с 1 сентября по 1 декабря 2008 года, будут выплачены в период от одного до трех лет, от даты судебного решения о **банкротстве**. С 1 января 2008 года ответственность по полисам НАСК "Оранта" будет переполжена на ряд партнерских страховых компаний. Список партнеров будет опубликован на нашем сайте.  
 Клиентам, у которых срок действия договоров заканчивается позже 31-го декабря, необходимо прибыть в ближайшее отделение компании и перезаключить договор страхования с нашими партнерами. До 31-го декабря на перезаключение договоров по абсолютно всем видам страхования нашими партнерами предоставляются скидки". На официальном сайте компании, данная информация не подтвердился.  
 Как сообщили "ЭН" в самом НАСК "Оранта", это не правдивая информация.

Рис. 50 – Опровержение?

Дальнейший спад публикаций по теме НАСК «Оранта» и возвращение его на нормальный «средний» уровень свидетельствует о том, что компания своими осторожными и точными действиями смогла с успехом противостоять информационной операции.

<ol style="list-style-type: none"> <li>1 "Оранта" - не банкрот                  Бердские ведомости 2008.12.11 19:49                  Смирнова Екатерина В среду, 10 декабря, в интернет-изданиях появилась информация о том, что крупнейшая страховая компания классического страхового рынка Украины НАСК "Оранта" заявила о своем банкротстве и уведомила своих клиентов о невозможности выполнения взятых на себя обязательств после 31 декабря 2008 года.                  Похожие документы - Оригинал</li> <li>2 Официальный пресс-релиз НАСК Оранта по поводу мнимого банкротства                  News.kompass.ua 2008.12.11 17:35                  10 декабря, в 11:30 от имени руководителя пресс-службы наблюдательного совета НАСК Оранта Алены Кулаковой на множество Интернет-адресов было направлено СПЛМ-сообщение о банкротстве крупнейшей в стране страховой компании Оранта.                  Похожие документы - Оригинал</li> <li>3 Страховая компания "Оранта" подверглась массовой информационной атаке                  Простобанк.ua 2008.12.11 17:07                  Многие интернет-пользователи получили по электронной почте сообщение, в котором якобы от имени руководителя пресс-службы "Оранты" говорилось о том, что страховщик начал процедуру банкротства и с 31 декабря 2008 года прекращает выполнять перед своими клиентами все взятые на себя обязательства.                  Похожие документы - Оригинал</li> <li>4 "Оранта" исключает причастность конкурентов к информации о якобы банкротстве компании                  УНИАН 2008.12.11 16:46                  Руководство национальной акционерной страховой компании (НАСК) "Оранта" исключает причастность конкурентов к распространению через спам-рассылку информации о якобы банкротстве компании.                  Похожие документы - Оригинал</li> <li>5 "Оранта" намерена привлечь к ответственности распространителей лже-информации                  УНИАН 2008.12.11 16:46                  Руководство национальной акционерной страховой компании (НАСК) "Оранта" обратилось в Генеральную прокуратуру, Службу безопасности Украины, а также к премьер-министру Украины с просьбой о расследовании инцидента и привлечении к ответственности распространителей информации о якобы банкротстве компании.                  Похожие документы - Оригинал</li> <li>6 "Оранта" просит Генпрокуратуру найти автора спама о банкротстве                  "Мир домов" 2008.12.11 16:57                  Утром 10 декабря от имени руководителя пресс-службы СК "Оранта" было Елены Кулаковой было отправлено на множество интернет-адресов спам-сообщение о банкротстве страховой компании.                  Похожие документы - Оригинал</li> </ol>	<table border="1"> <thead> <tr> <th colspan="2">Информационный портрет</th> </tr> <tr> <th colspan="2">Уточнить запрос</th> </tr> </thead> <tbody> <tr> <td>Рублики (6)</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Языки (2)</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Страны источников (3)</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Источники (19)</td> <td><input type="checkbox"/></td> </tr> <tr> <td>AND</td> <td><input type="checkbox"/></td> </tr> <tr> <td>META - Украина</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Пресс-релизы</td> <td><input type="checkbox"/></td> </tr> <tr> <td>СТРАХНАДЗОР</td> <td><input type="checkbox"/></td> </tr> <tr> <td>PR - это жизнь</td> <td><input type="checkbox"/></td> </tr> <tr> <td>TRISTAR.com ua</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Экономична правда</td> <td><input type="checkbox"/></td> </tr> <tr> <td>UkrBiz.net</td> <td><input type="checkbox"/></td> </tr> <tr> <td>УНИАН</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Одеські вісті</td> <td><input type="checkbox"/></td> </tr> <tr> <td>fofinSURER.com</td> <td><input type="checkbox"/></td> </tr> <tr> <td>"Страхование сегодня"</td> <td><input type="checkbox"/></td> </tr> <tr> <td>B2Blogger.com</td> <td><input type="checkbox"/></td> </tr> <tr> <td>"Комерсант-Украина"</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Prbloom</td> <td><input type="checkbox"/></td> </tr> <tr> <td>"Вікна" Калушчини</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Fin.org.ua</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Міністерство економіки України</td> <td><input type="checkbox"/></td> </tr> <tr> <td>НАСК "Оранта"</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Страхование Украины</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Украина деловая</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Размер (2)</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Цифровая насыщенность (3)</td> <td><input type="checkbox"/></td> </tr> <tr> <td>Тональность (2)</td> <td><input type="checkbox"/></td> </tr> <tr> <td>География (25)</td> <td><input type="checkbox"/></td> </tr> </tbody> </table>	Информационный портрет		Уточнить запрос		Рублики (6)	<input type="checkbox"/>	Языки (2)	<input type="checkbox"/>	Страны источников (3)	<input type="checkbox"/>	Источники (19)	<input type="checkbox"/>	AND	<input type="checkbox"/>	META - Украина	<input type="checkbox"/>	Пресс-релизы	<input type="checkbox"/>	СТРАХНАДЗОР	<input type="checkbox"/>	PR - это жизнь	<input type="checkbox"/>	TRISTAR.com ua	<input type="checkbox"/>	Экономична правда	<input type="checkbox"/>	UkrBiz.net	<input type="checkbox"/>	УНИАН	<input type="checkbox"/>	Одеські вісті	<input type="checkbox"/>	fofinSURER.com	<input type="checkbox"/>	"Страхование сегодня"	<input type="checkbox"/>	B2Blogger.com	<input type="checkbox"/>	"Комерсант-Украина"	<input type="checkbox"/>	Prbloom	<input type="checkbox"/>	"Вікна" Калушчини	<input type="checkbox"/>	Fin.org.ua	<input type="checkbox"/>	Міністерство економіки України	<input type="checkbox"/>	НАСК "Оранта"	<input type="checkbox"/>	Страхование Украины	<input type="checkbox"/>	Украина деловая	<input type="checkbox"/>	Размер (2)	<input type="checkbox"/>	Цифровая насыщенность (3)	<input type="checkbox"/>	Тональность (2)	<input type="checkbox"/>	География (25)	<input type="checkbox"/>
Информационный портрет																																																															
Уточнить запрос																																																															
Рублики (6)	<input type="checkbox"/>																																																														
Языки (2)	<input type="checkbox"/>																																																														
Страны источников (3)	<input type="checkbox"/>																																																														
Источники (19)	<input type="checkbox"/>																																																														
AND	<input type="checkbox"/>																																																														
META - Украина	<input type="checkbox"/>																																																														
Пресс-релизы	<input type="checkbox"/>																																																														
СТРАХНАДЗОР	<input type="checkbox"/>																																																														
PR - это жизнь	<input type="checkbox"/>																																																														
TRISTAR.com ua	<input type="checkbox"/>																																																														
Экономична правда	<input type="checkbox"/>																																																														
UkrBiz.net	<input type="checkbox"/>																																																														
УНИАН	<input type="checkbox"/>																																																														
Одеські вісті	<input type="checkbox"/>																																																														
fofinSURER.com	<input type="checkbox"/>																																																														
"Страхование сегодня"	<input type="checkbox"/>																																																														
B2Blogger.com	<input type="checkbox"/>																																																														
"Комерсант-Украина"	<input type="checkbox"/>																																																														
Prbloom	<input type="checkbox"/>																																																														
"Вікна" Калушчини	<input type="checkbox"/>																																																														
Fin.org.ua	<input type="checkbox"/>																																																														
Міністерство економіки України	<input type="checkbox"/>																																																														
НАСК "Оранта"	<input type="checkbox"/>																																																														
Страхование Украины	<input type="checkbox"/>																																																														
Украина деловая	<input type="checkbox"/>																																																														
Размер (2)	<input type="checkbox"/>																																																														
Цифровая насыщенность (3)	<input type="checkbox"/>																																																														
Тональность (2)	<input type="checkbox"/>																																																														
География (25)	<input type="checkbox"/>																																																														

Рис. 51. Сообщения с опровержением

#### 7.4. Выявление информационных операций

Для оперативного анализа информационной обстановки с целью выявления информационных операций применяются специализированные системы мониторинга информационного пространства (контент-мониторинга). Такие системы обеспечивают, во-первых, оперативность, которую не могут обеспечить традиционные поисковые системы (время индексации сетевого контента даже лучшими из них составляет от нескольких суток до нескольких недель). Во-вторых, полноту (как в плане источников, так и представления материалов источников), которую не всегда обеспечивают обычные агрегаторы новостей. И, в-третьих, необходимые аналитические средства, которые позволяют пользователю создавать аналитические отчеты, базирующиеся на публикациях по заданной тематике в необходимый период времени.

В плане профилактики информационных операций следует внимательно следить за динамикой публикаций о целевой компании, если есть возможность, с учетом тональности этих публикаций, пользоваться доступными аналитическими средствами, например, вейвлет-анализом. При этом следует ориентироваться на возможные модели информационных атак, например, если эта модель охватывает фазы: «фоновые публикации» — «затишье» — «артподготовка» — «затишье» — «атака» (рис. 52), то уже по первым трем компонентам можно с большой вероятностью предсказать грядущие события.

Приведенный выше план, очевидно, является идеальным, ориентированным исключительно на данные контент-мониторинга веб-ресурсов.

Конечно, в лучшем положении находятся пользователи профессиональных систем контент-мониторинга. Многие современные информационно-аналитические системы содержат в своем составе средства отображения статистики вхождения в базы данных понятий, соответствующих пользовательским запросам. В частности, авторами использовалась подсистема статистики в рамках системы контент-мониторинга веб-пространства InfoStream, реализующая данную функциональность.

При изучении трендов информационных операций в качестве временных рядов рассматриваются именно ряды по количеству тематических публикаций за определенный промежуток времени (чаще всего – за сутки), соответствующие этим информационным операциям. Поэтому для выявления трендов исследуются инфор-

мационные потоки, соответствующие тематикам информационных операций – тематические информационные потоки.

Приведенные в [Горбулін, 2009] тренды сообщений, соответствующие этапам информационной операции, приведены на рис. 53. При этом аналитикам следует ориентироваться на такие модели, например, если мониторинг позволяет определить фазы: «фон» – «затишье» – «артподготовка» – «затишье» – «атака», то уже по первым трем компонентам можно с большой вероятностью предсказать будущие события.

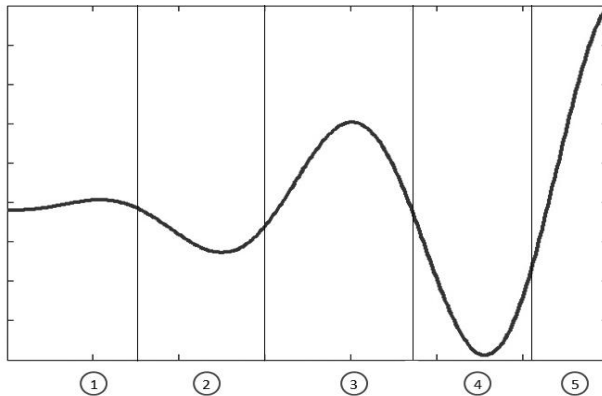


Рис. 52 – Динамика количества тематических сообщений во время проведения информационной операции: 1 – фон; 2 – затишье; 3 – «артподготовка»; 4 – затишье; 5 – атака/триггер роста

Следует отметить, что подобная динамика количества тематических сообщений при проведении информационных операций хорошо описывается известным уравнением распространения электромагнитных волн:

$$y = A + Bx \sin(x),$$

где  $x$  – время,  $A$  и  $B$  – константы, определяемые эмпирически.

Как известно, в настоящее время инновационная деятельность также косвенно измеряется количеством публикаций, относящимся к инновациям, существует несколько моделей инновационных процессов, среди которых можно выделить модель диффузии инноваций [Bhargava, 1993]. Вместе с тем, внедрение инноваций также можно считать информационными операциями. Поэтому обратимся к результатам соответствующих исследований. На рис. 53

приведена обоснованная в [Хорошевский, 2012] диаграмма количества публикаций, соответствующая тренду инновационной деятельности.

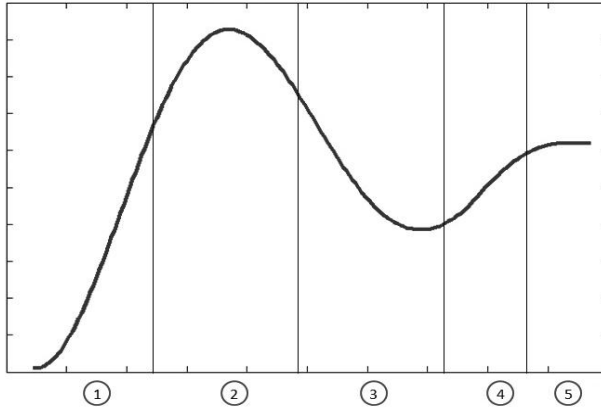


Рис. 53 – Диаграмма количества публикаций, соответствующих тренду инновационной деятельности: 1 – атака/триггер роста; 2 – пик завышенных ожиданий; 3 – утрата иллюзий; 4 – общественное осознание; 5 – продуктивность/фон

Объединяя графики, соответствующие началу информационной операции (рис. 52) и тренду инновационной деятельности (рис. 53), можно получить полный график, соответствующий отображению информационных операций в информационном пространстве (рис. 54).

Предложенные модели полностью соответствуют реальным данным, которые экстрагируются системами контент-мониторинга [Додонов, 2009], [Ландэ, 2007]. Поэтому приведенные зависимости могут быть использованы как шаблоны для выявления информационных операций – как путем анализа ретроспективного фонда сетевых публикаций, так и для оперативного мониторинга появления некоторых их признаков в реальном времени. Как известно, для выявления информационных операций следует внимательно следить за динамикой публикаций по целевой теме и, если есть возможность, пользоваться доступными аналитическими средствами, средствами цифровой обработки данных и распознавания образов, например, вейвлет-анализом или полиномами Кунченко [Чертов, 2009].

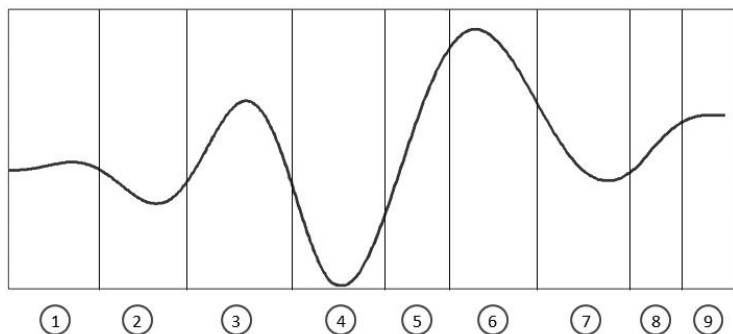


Рис. 54 – Обобщенная диаграмма, соответствующая всем этапам жизненного цикла информационных операций: 1 – фон; 2 – затишье; 3 – «артподготовка»; 4 – затишье; 5 – атака/триггер роста; 6 – пик завышенных ожиданий; 7 – утрата иллюзий; 8 – общественное осознание; 9 – продуктивность/фон

В качестве примера, на рис. 55 показана динамика публикаций в RUNet – тематических информационных потоков по запросам «Банки, Кипр», «Офшор», «Вирджинские острова» за март-апрель 2013 года, в период известных кризисных событий, полученная с помощью системы InfoStream. Как видно из рис. 55, пик публикаций, связанных с банковским кризисом на Кипре приходится на 17-18 марта 2013 года, в то время, как большинство публикаций по Вирджинским островам пришелся на 4-5 апреля, когда там, со значительно меньшими масштабами, стали проявляться события, подобные кипрским. При этом следует отметить слабую коррелированность динамики информационных потоков, связанных с Кипром и Вирджинскими островами. В этом случае коэффициент взаимной корреляции соответствующих числовых рядов составил всего 0,3. При этом отмечается высокий уровень взаимной корреляции рядов соответствующих тематикам «Офшор» и «Банки Кипра» (0,73), а также «Офшор» и «Вирджинские острова» (0,77).

По-видимому, проявления информационных операций в области офшорных банков в данном случае лучше всего увидеть при анализе более общей тематики – «Офшоры». На графике соответствующего числового ряда четко видны две области локальных экстремумов, соответствующих кризисным ситуациям на Кипре и на Вирджинских островах, а также фазы, соответствующие «затишьям» и «артподготовкам».

Можно высказать предположение, что если динамика частного информационного потока в какой-то момент начинает существенно отличаться от динамики потока, соответствующего более общей тематике (как в рассматриваемом случае, «Банки Кипра» и «Офшор»), то возможно проявление признаков начала информационной операции, относящейся к узкой тематике.

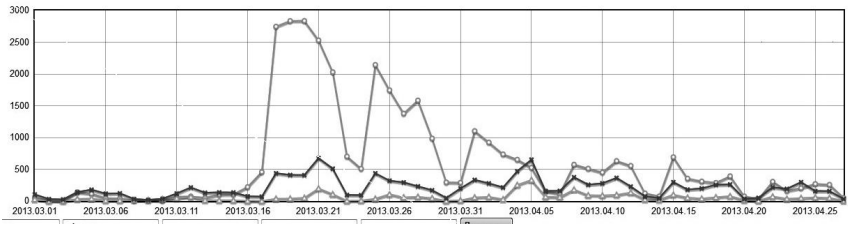


Рис. 55 – Диаграмма динамики тематических информационных потоков по запросам: о – «Банки Кипра»; Δ – «Вирджинские острова»; х – «Офшор»

При проведении вейвлет-анализа [Астафьева, 1996], [Buckheit, 1995] (рис. 56) было принято решение использования вейвлета «Мексиканская шляпа», как близкого по форме к диаграмме, приведенной на рис. 54.

Рассматриваемые процессы четко просматриваются как на вейвлет-спектрограммах, так и на соответствующих им скелетонах (графиках линий экстремумов).

Приведенные модели и методы пригодны для описания общих тенденций динамики информационных процессов, однако, проблема прогнозирования остается открытой. По-видимому, более реалистичные модели могут быть получены с учетом дополнительного набора факторов, большинство которых не воспроизводятся во времени. Вместе с тем, структура правил, лежащих в основе функционирования большинства из доступных моделей, позволяет вносить соответствующие коррективы, например, искусственно моделировать случайные отклонения.

Отметим, что воспроизведение результатов во времени является серьезной проблемой при моделировании информационных процессов и составляет основу научной методологии. В настоящее время только ретроспективный анализ уже реализованных информационных операций остается относительно надежным способом их верификации.

Естественно, на практике ориентация лишь на единственный тип источников может привести к дефициту информации, необхо-

димой для принятия решений, неточностям, а порой – к дезинформированности. Лишь применение комплексных систем, базирующихся на использовании многочисленных источников и баз данных, наряду с приведенными выше возможностями системы контент-мониторинга, может гарантировать эффективную информационную поддержку при противодействии информационным операциям.

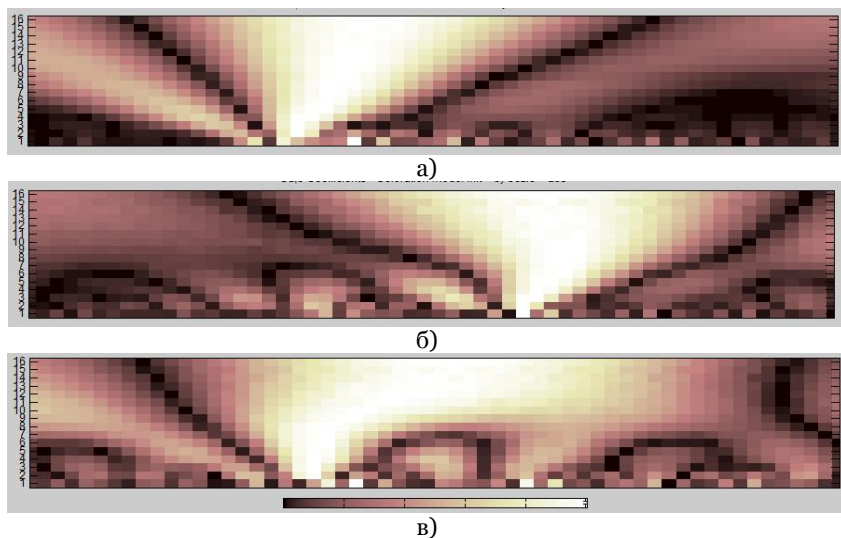


Рис. 56 – Вейвлет-спектрограммы, соответствующие динамике тематических информационных потоков по запросам: а – «Банки Кипра»; б – «Вирджинские острова»; в – «Офшор»

Выделенные образцы поведения рядов интенсивностей тематических публикаций могут рассматриваться как шаблоны (образцы) функциональной зависимости. Эти шаблоны можно взять в качестве единого базисного элемента некоторого линейного пространства, т.е. в качестве порождающего элемента  $e$  для моделирования с помощью полиномов Кунченко [Чертов, 2009].

Тогда как линейную комбинацию линейно-независимых преобразований  $f_1(e), f_2(e), \dots, f_n(e)$  соответствующего порождающего элемента можно построить полином  $P_n$  приближения  $n$ -го порядка к части выходного сигнала  $f_s(e)$ :



$$P_n = \sum_{\substack{k=0, \\ k \neq s}}^n c_k f_k(e),$$

где коэффициенты  $c_k$  определяются из условия обеспечения минимума расстояния между строящимся полиномом и сигналом. Элемент  $c_0$  определяется выражением:

$$c_0 = \frac{\langle f_s(e), f_0(e) \rangle - \sum_{k=1, k \neq s}^n c_k \langle f_k(e), f_0(e) \rangle}{\langle f_0(e), f_0(e) \rangle},$$

а другие коэффициенты  $c_k$  – как решение системы линейных уравнений:

$$\sum_{k=1, k \neq s}^n c_k F_{i,k} = F_{i,s}, \quad i = \overline{1, n}, \quad i \neq s,$$

где центрированные корреляты  $F_{i,k}$  также рассчитываются с помощью соответствующих преобразований:

$$F_{i,k} = \langle f_i(e), f_k(e) \rangle - \frac{\langle f_i(e), f_0(e) \rangle \cdot \langle f_k(e), f_0(e) \rangle}{\langle f_0(e), f_0(e) \rangle}.$$

Числовой характеристикой, которую можно использовать в критериях качества сопоставления сигнала с выделенным шаблоном, т. е. как меру приближения полинома Кунченко  $P_n$  к сигналу  $f_s(e)$ , можно считать коэффициент эффективности  $d_n$ :

$$d_n = \frac{\sum_{k=1, k \neq s}^n c_k \langle f_k(e), f_s(e) \rangle}{\langle f_s(e), f_s(e) \rangle}.$$

Рассмотренный метод распознавания определенных образцов с помощью построения пространства с порождающим элементом и поиска коэффициентов соответствующего полинома Кунченко может быть использован в любой проблемной области, в которой

можно априори во временном ряду выделить определенные характерные шаблоны.

Таким образом, построив типовые модели поведения рядов интенсивности тематических публикаций во время проведения информационных операций и сопоставив шаблоны, полученные на их основе, можно использовать метод на основе полиномов Кунченко для определения (и предупреждения) возможной информационной атаки.

Динамика тематических информационных потоков определяется комплексом как внутренних, так и внешних нелинейных механизмов, которые должны быть отражены при моделировании (возможно, в неявном виде). Зачастую удовлетворительным оказывается упрощенное понимание тематического информационного потока как некоторой зависимой от времени величины, поведение которой описывается в аналитическом виде нелинейными уравнениями. Сегодня при моделировании информационных потоков используются преимущественно аналитические нелинейные модели, применяются методы нелинейной динамики, теории клеточных автоматов, перколяции, самоорганизованной критичности [Ландэ, 2009], [Додонов, 2011].

Для анализа динамики реальных тематических информационных потоков (ТИП), и, соответственно, оценки их моделей необходимо каким-то образом получить соответствующую статистику, представленную в виде временных рядов.

Динамику реальных тематических информационных потоков (ТИП), например, отображает мультиагентная модель, в рамках которой отдельные документы ТИП ассоциируются с агентами, жизненный цикл которых – с жизненным циклом документов в информационном пространстве. Соответственно, все пространство мультиагентной модели ассоциируется с тематическим информационным потоком.

Предполагается, что в течение дискретных моментов времени происходит эволюция популяции агентов. При этом отдельные агенты могут:

- 1) «самозародиться» (рождаться по причинам, возникающим вне рассматриваемого мультиагентного пространства);
- 2) «порождать» новых агентов;
- 3) «умирать» – исчезать из пространства агентов (соответствует утере актуальности документов);
- 4) получать ссылки от других агентов.

Каждый агент обладает «потенциалом», зависящим от его возраста (времени жизни на текущий момент –  $t$ ), от авторитетности

(ссылок, проставленных на него –  $ns$ ) и плодовитости (количества порожденных непосредственно им агентов –  $k$ ). Потенциал агента  $Pot$  определяется формулой:

$$Pot = \frac{1 + ns + k}{t}.$$

На рис. 57 приведен пример возможной динамики мультиагентной системы: процессы рождения новых агентов от существующих обозначены сплошными стрелками, процессы проставления ссылок на агентов представлены пунктирными стрелками, живые агенты – черными кругами, «мертвые» агенты к моменту  $t = 5$  – незаполненными окружностями.

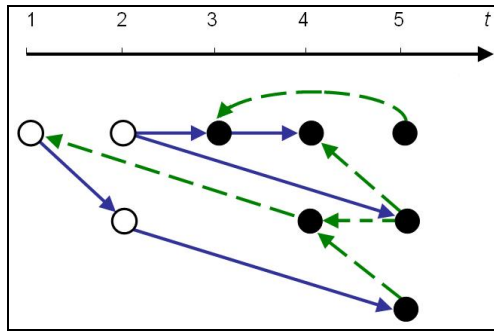


Рис. 57 – Фрагмент мультиагентного пространства

Итак, управляющие параметры модели следующие:

- вероятность «самозарождения»  $P_1$ ;
- вероятность «рождения» от существующего:  
 $P_2 \cdot Pot$ ;
- вероятность «смерти» агента:  $P_3 / Pot$ ;
- вероятность ссылки на агента:  $P_4 \cdot Pot$ .

Варьирование этими четырьмя параметрами  $P_1$ ,  $P_2$ ,  $P_3$  и  $P_4$  позволили смоделировать типовые профили поведения ТИП.

На рис. 58 представлены результаты численного моделирования количества агентов (ось ординат на графике) в рассматриваемой мультиагентной системе в зависимости от количества тактов модели (ось абсцисс).

Рассматриваемая модель эволюции пространства агентов при различных значениях управляющих параметров согласуется с динамикой реальных тематических информационных потоков, определенных с помощью системы InfoStream.

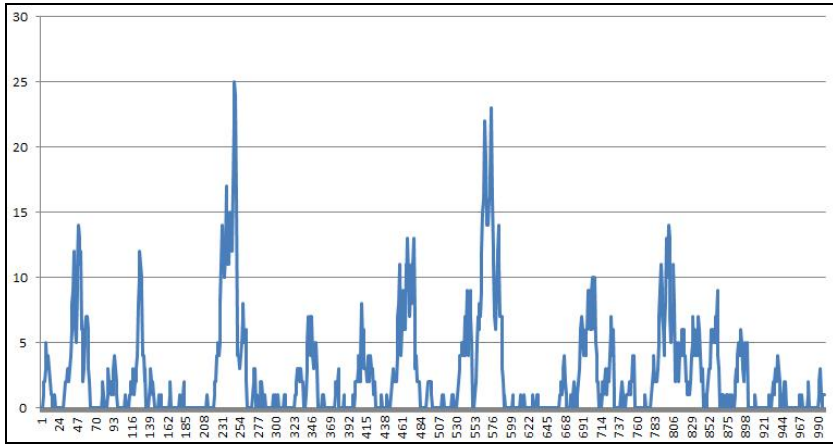


Рис. 58 – Динамика изменения количества агентов в модели

Наряду с исследованием огибающих динамики ТИП большой практической интерес представляет неравномерность, изрезанность соответствующих графиков, которая может свидетельствовать об отклонениях от естественной природы, информационных операциях, манипулировании [Горбулін, 2009]. В частности, для отображения неравномерностей во временном ряду использовался метод SCA (Smoothing, Cellular Automata) [Ландэ, 2013], основанный на учете аномальных значений и концепции одномерных клеточных автоматов. С помощью этого метода не детектируются абсолютные амплитудные всплески, однако он хорошо показал себя на «изрезанных» структурах данных, близких к фрактальным.

К таким данным относятся, в частности, временные ряды, связанные с объемами публикаций в веб-пространстве по определенным тематикам, которые рассматриваются ниже как иллюстрация метода.

В предлагаемой модели каждому значению исходного ряда измерений  $x_0(t)$  (обозначим исходный ряд, как  $X_0 = \{x_0(t)\}$ ) соответствует одна клетка клеточного автомата. По ряду измерений строится сглаженный по приведенному ниже правилу ряд

$X_1 = \{x_1(t)\}$ . Затем ряду  $X_1$  ставится в соответствие ряд  $X_2$  (получаемый из  $X_1$  по тому же алгоритму сглаживания) и т.д. Правило сглаживания пиков заключается в том, что значения, которые принимают элементы рядов измерений  $x_k(t) \in X_k$  ( $k$  – шаг сглаживания,  $t$  – номер элемента ряда измерений) составляют:

$$x_k(t) = \begin{cases} x_{k-1}(t), & \text{if } x_{k-1}(t) \leq \frac{x_{k-1}(t-1) + x_{k-1}(t+1)}{2}, \\ \frac{x_{k-1}(t-1) + x_{k-1}(t+1)}{2}, & \text{if } x_{k-1}(t) > \frac{x_{k-1}(t-1) + x_{k-1}(t+1)}{2}. \end{cases}$$

Цвет клетки с номером  $t$  одномерной клеточной структуры, соответствующей  $X_k$ , белый, если  $x_k(t)$  совпадает с  $x_{k-1}(t)$ , в противном случае – черный. Таким образом, каждой клетке соответствует значение  $x_k(t)$  и значение ее цвета. (Необходимо отметить, что такую систему нельзя считать каноническим клеточным автоматом, так как в общем случае клеткам может соответствовать бесконечное множество значений  $x_k(t)$  и два значения цвета).

Рассмотрим результаты выполнения данного алгоритма для простейших структур, которые, как показывает практика, охватывает все возможные варианты визуализации.

Очевидно, если значения ряда измерений в рассматриваемой зоне представляют собой вогнутое (выпуклое вниз множество), то сразу же, на первой итерации получим  $\forall t: x_1(t) = x_0(t)$  и выполнение алгоритма прерывается.

Если область значений представляют собой выпуклое вверх множество, то визуальное представление клеточных автоматов принимает вид сплошной черной полосы (рис. 59: вертикальная ось – номер шага итерации, а горизонтальная ось – номер элемента ряда измерений).

Единичные всплески значений в исходном ряде измерений (рис. 60,а) и области изрезанности (рис. 60,б) могут вызывать появление структур типа «шахматной доски».

Кроме того, диаграммы, формируемые в результате визуализации в соответствии с предложенным алгоритмом, позволяют выявлять периодические составляющие.

Предложенный метод SCA является относительно простым в программной реализации и линейным по сложности, так как базируется на алгоритме сглаживания пиков и концепции клеточных

автоматов. Он позволяет визуально выявлять единичные и нерегулярные «всплески», резкие колебания, скачки значений, зоны неустойчивости количественных показателей в разные периоды времени. Метод SCA испытывался при анализе временных рядов, связанных с объемами публикаций в веб-пространстве по определенным темам.

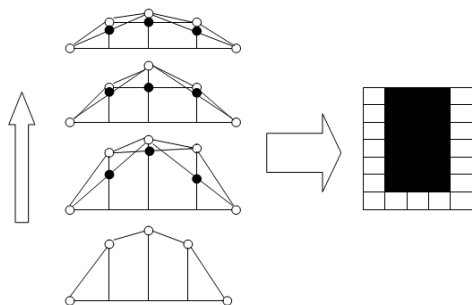


Рис. 59 – Выпуклое вверх множество точек

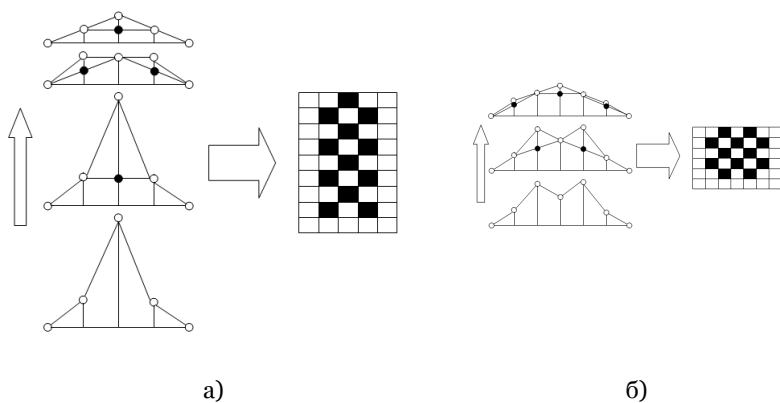


Рис. 60 – Появление структур типа «шахматной доски»

На диаграммах, формируемых в соответствии с SCA, выпуклое вверх множество принимает вид сплошной черной полосы, выпуклое вниз множество – белой полосы, а области изрезанности, неустойчивости могут вызывать появление «клетчатых» структур.

Отображение реального временного ряда измерений, соответствующего посуточным объемам публикаций в веб-пространстве по некоторой заданной теме (точки ряда – объемы публикаций за сутки) с помощью метода SCA представлено на рис. 61. Здесь четко отслеживаются недельные периодичности ТИП (минимумы – праздники, субботы и воскресенья), а также области неравномерности, резких колебаний объемов публикаций, свойственных ТИП в пред- и посткризисные периоды.

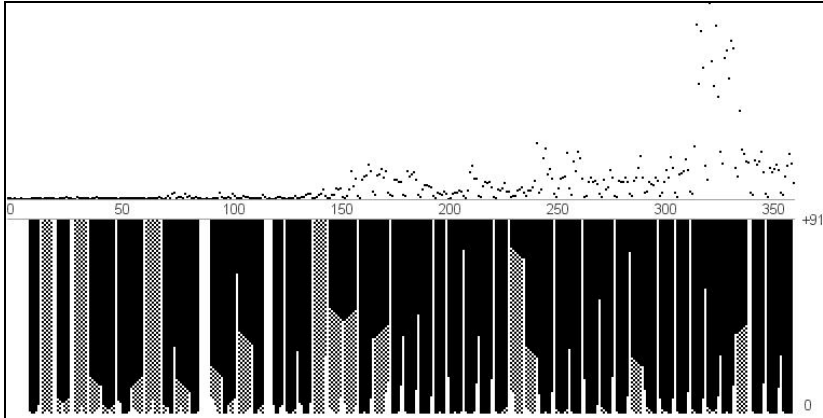


Рис. 61 – Отображение реальной динамики публикаций в течение года с помощью метода SCA

На рис. 62 представлена SCA-визуализация динамики количества агентов (документов) – результатов мультиагентного моделирования.

Следует отметить, что предлагаемая модель:

1) не учитывает конкуренции агентов внутри пространства агентов (предполагается только сотрудничество путем проставления ссылок и порождения новых агентов);

2) конкуренция разных тематических информационных потоков учитывается лишь неявно, как причина, обуславливающая параметры функционирования рассматриваемой мультиагентной системы.

В предложенной модели учитывается общеизвестная практика проведения информационных кампаний в социальных сетях, заключающаяся в регистрации большого числа аккаунтов-роботов (роя), от имени которых проставляются ссылки (лайки) на мате-

риалы, публикуемые от имени аккаунтов из того же роя и на целевые документы.

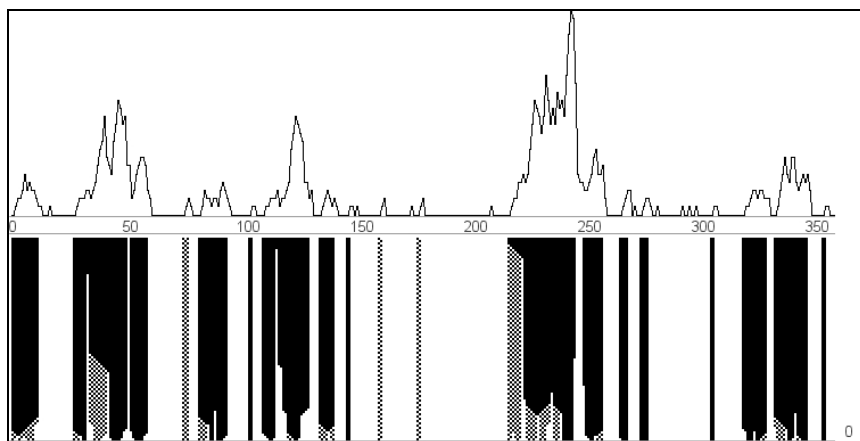


Рис. 62 – Отображение результатов мультиагентного моделирования с помощью метода SCA

Естественно, на практике ориентация лишь на единственный тип источников и математических моделей может привести к дефициту информации, необходимой для принятия решений, неточностям, а порой — к дезинформированности. Лишь применение комплексных систем, базирующихся на использовании многочисленных источников, баз данных, математических моделей, наряду с приведенными выше возможностями систем контент-мониторинга может гарантировать эффективную информационную поддержку при противодействии информационным операциям.

## 7.5. Противодействие информационным операциям

Рассмотренные практические примеры позволили выработать некоторую общую методику проведения оборонительной информационной операции с использованием системы контент-мониторинга веб-ресурсов. Допустим, объектом агрессивной информационной операции является компания «АБВ». Предлагается такие 12 шагов противодействия:

- 1) сбор информации с публикациями в «чужих» (не имеющих отношения к «АБВ», неаффилированных) СМИ о компании;



2) построение графика – динамики появления сообщений о компании «АБВ» в сетевых СМИ;

3) анализ динамики с ретроспективой в 6–12 месяцев с помощью методов анализа временных рядов. После этого анализируется контент публикаций в пороговых точках, определяются моменты, длительность, периодичность воздействия, привязка моментов воздействия к другим событиям из области интереса объекта;

4) определение источников, публикующих наибольшее количество негатива (публикаций с отрицательной тональностью) о компании «АБВ»;

5) определение «первоисточников» публикаций в СМИ – тех источников, которые первыми опубликовали негативную информацию;

6) определение вероятных «заказчиков» – владельцев или лиц, влияющих на издательскую политику отдельных СМИ;

7) определение сфер общих интересов компании «АБВ» и потенциальных «заказчиков» (путем выявления общих информационных характеристик – пересечений «информационных портретов» системы InfoStream, строящихся для объекта и «заказчика»), ранжирование потенциальных «заказчиков» по их интересам;

8) определение критериев информационных воздействий на основе самых рейтинговых интересов;

9) моделирование информационных воздействий, для чего находятся связи «заказчика» – наиболее связанные с ним персоны и организации, анализируется динамика воздействия со стороны заказчика и строится прогноз этой динамики, анализируется контент публикаций в пороговых точках кривой динамики – определяются критичные точки воздействия.;

10) прогнозируются дальнейшие шаги воздействия путем анализа аналогичной динамики публикаций для других компаний в ретроспективной базе данных системы InfoStream;

11) с учетом реалей и публикаций из ретроспективной базы данных оцениваются вероятные последствия;

12) организуется информационное (и не только) противодействие. Примеры публикаций в контексте противодействия находятся в ретроспективной базе данных.

## **7.6. Конкурентная разведка в антимонопольной деятельности**

Антимонопольная деятельность, создание в государстве конкурентной среды, предполагают борьбу с проявлениями монополиз-

ма на рынках товаров и услуг, в том числе, отражение соответствующих информационных операций, проводимых монополистами, проведение наступательных информационных операций.

Для осуществления антимонопольной деятельности со стороны государства, создания конкурентной среды необходимо использовать все доступные и легальные информационные и программные средства. Однако сегодня наблюдается реальный дефицит оперативной рыночной информации, определяемый как слабыми коммуникациями между отдельными органами власти, так и неполнотой, неточностью соответствующих официальных баз данных. С другой стороны, существует огромный информационный ресурс – веб-пространство.

Очевидно, что несмотря на такие преимущества, как оперативность и широкий охват информации, этот ресурс не может быть доказательным источником, однако, его нельзя отвергать в некоторых важных приложениях. Оперативность, свойственная веб-среде, в частности, имеет решающее значение при реализации концепции управления OODA, также известного как цикл Бойда. В переводе аббревиатура OODA означает «Наблюдение – Ориентация – Решение – Действие» [Ивлев, 2008]. Концепция OODA находит во всем мире широкое применение в управлении информационным противоборством, предотвращении информационных операциям. Очевидно, и в антимонопольной деятельности эта концепция может и должна найти применение путем реализации центров быстрого реагирования на монопольные проявления.

Общеизвестно, что антимонопольная деятельность – это комплекс мероприятий, направленных на ограничение деятельности монополий в рамках всего государства, а также создание соответствующего законодательства, в то время, как конкурентная разведка направлена на повышение конкурентоспособности лишь отдельных субъектов хозяйствования. Согласно этому частные задачи конкурентной разведки могут быть обобщены до уровня антимонопольной деятельности на уровне государства следующим образом:

- 1) сбор информации и своевременное информационное обеспечение соответствующих государственных органов;
- 2) выявление факторов риска, угроз конкурентной среды государства;
- 3) выявление факторов, влияющих на получение отдельными компаниями монопольных преимуществ;
- 4) выработка прогнозов и рекомендаций, влияющих на развитие конкурентной среды;

5) усиление благоприятных и локализация неблагоприятных факторов для развития конкурентной среды.

С помощью методов конкурентной разведки, которая становится современным направлением исследования поведения конкурентов на рынке, создаются альтернативные модели рынка для определения характеристик его участников и оптимизации тактики и стратегии развития субъектов хозяйствования на определенных рынках. Достижение таких целей требует использования эффективных приемов работы с информацией и ее элементами. Информация в этом смысле становится объектом в процессе исследования рынка и создания его модели.

Все приведенные задачи реализуются в рамках замкнутой схемы взаимодействия рыночной среды и виртуального информационного пространства.

Как известно, рыночная реальность находит свое отражение в виртуальном информационном пространстве, именно с ним работают эксперты-аналитики, которые готовят информацию, прогнозы для ЛПР, которые, в свою очередь, обеспечивают целенаправленное воздействие на рыночную среду.

По-видимому, все указанные функциональные компоненты конкурентной разведки могут использоваться и для общих задач, стоящих перед антимонопольными органами государства.

Возможности использования средств конкурентной разведки, в частности, средств контент-мониторинга, в антимонопольной деятельности проиллюстрируем на примере гречневого ценового коллапса в начале 2010 г. в Украине. Антимонопольный комитет Украины только в октябре 2011 г. (через полтора года!) обнаружил и наказал участников сговора на рынке гречки (рис. 63), тогда как сотни пользователей системы контент-мониторинга InfoStream могли видеть фигурантов дела уже в феврале 2010 г. в «информационном портрете» этой системы (рис. 64).

Безусловно, система поддержки антимонопольной деятельности, как и системы конкурентной разведки, использующие Интернет как один из информационных ресурсов, должна настраиваться под специфику конкретных рынков. Она должна включать соответствующую классификацию, гибкие механизмы поиска, оперативной доставки данных, а также качественной оценки информации.

Одной из важнейших задач анализа информации при этом является определение ее достоверности, т.е. решение задачи анализа и фильтрации шума и ложной информации. После анализа достоверности информации должны следовать оценки ее точности и важности. Главным критерием достоверности данных на практике

является подтверждение информации другими источниками, заслуживающими доверия.

The screenshot shows a document search interface with the following content:

- Document title:** AMKU нашел вредителей
- Source:** Укрудром 2011.10.25 09:56, [http://www.ukrudrom.ua/digest/AMKU\\_nashel\\_vreditely.html](http://www.ukrudrom.ua/digest/AMKU_nashel_vreditely.html)
- Text:**

Гречневый ценовой коллапс стал новостью номер один среди покупателей. Теперь же нашлись и виновники событий полугодичной давности.

Антимонопольный комитет Украины (АМКУ) оштрафовал восемь предприятий в целом на 590 тыс. грн. за антиконкурентные согласованные действия на рынке гречневой крупы. Как сообщают в Антимонопольном комитете, в частности, к ответственности привлечены такие субъекты хозяйствования, как "Родной продукт" (ТМ "Хуторок"), "Сельхозсервис" (ТМ "Фабрика круп"), ДП "Крупозавод Озерьянка" (ТМ "Озеряночка"), "Сквириский комбинат хлебопродуктов" (ТМ "Сквирянка"), ЗАО "Нива" (ТМ "Добродия"), КП "Белоцерковхлебопродукт", фермерское хозяйство "Дар земли", ДП "Новоукраинский комбинат хлебопродуктов" ГАК "Хлеб Украины". Как установил АМКУ с января по февраль 2010 г. указанные субъекты хозяйствования бесхозяйственно одновременно повысили оптово-отпускные цены на гречневую крупу.

Такие действия предприятий, утверждая в комитете, привели к общему экономически необоснованному повышению цен на гречку и привели к ущемлению интересов потребителей. "В результате ответчики прекратили нарушения", - сообщает в письме комитета.

Напомним, что прошлой зимой в Украине резко выросли цены на гречневую крупу - до 20-25 грн. Кабмин решил закупить гречку в Китае, после чего цены на крупу снизились до 16-18 грн. Однако китайская гречка оказалась не дешевой. Предпродажная себестоимость продукта достигала почти \$2100/т.
- Metadata:**
  - О документе: Рубрики (2): Агротроном, Экономика Украины
  - Языки (1): русский
  - Размер (1): средний
  - Цифровая насыщенность (1): малая
  - География (2): Украина, Китай
  - Персоны (5): Ярославский Шпаген, Присяжнюк Колесник, Арасланов

Рис. 63 - Виновники кризиса найдены (октябрь 2011 г.)

The screenshot shows a search results page with the following content:

- Search criteria:** Ст: 201001 До: 201002
- Search results:**
  - Найдено документов - 76, страница 1 из 6
  - Статистика слов: ЦЕН - 1686954, ГРЕЧКА - 4086, COUNTRY.UA - 1227403
- Search results list:**
  - В Запорожье гречка подорожала из-за поста?**  
 Репортаж Запорожская 2010.02.26 14:11  
 Подорожание продукта питания, особенно гречки, цена которой возросла почти на 20%, объясняется повышением закупочной цены, которую запорожские покупатели не устаканивают и не контролируют.  
 Похожие документы - Оригинал
  - Александр Син: Да пусть они задавятся своей гречкой!**  
 Полтсовет 2010.02.26 12:56  
 Мария Жартовская Заместитель губернатора Запорожской области Александр Син возмущен ростом цен на продукты питания в Запорожской области. В частности, за неделю стоимость продуктов выросла во всех районах области. Особенно Александра Сина возмущит рост цен на гречку. Стоимость гречки за последнюю неделю выросла на 20%, она уже в три раза дороже, чем крупа.  
 Похожие документы - Оригинал
  - Александр Син: Да пусть они задавятся своей гречкой!**  
 Полтсовет 2010.02.26 10:57  
 Мария Жартовская Заместитель губернатора Запорожской области Александр Син возмущен ростом цен на продукты питания в Запорожской области. В частности, за неделю стоимость продуктов выросла во всех районах области. Особенно Александра Сина возмущит рост цен на гречку. Стоимость гречки за последнюю неделю выросла на 20%, она уже в три раза дороже, чем крупа.  
 Похожие документы - Оригинал
  - Обмелели молочные реки**  
 Рабочая газета 2010.02.26 09:13  
 Несмотря на заверания премьер-министра Тимошенко о том, что после выборов цены на продукты питания стремглав полетят вниз, этого не происходит. До 7, а то и 10 гривен за kilo подорожала в последние дни картошка.  
 Похожие документы - Оригинал
  - За несколько дней "гречка" в Донецкой области подорожала почти вдвое**  
 КИД 2010.02.25 16:53
- Informational profile:**
  - Уточнить запрос
  - Рубрики (7)
  - Языки (2)
  - Страны источников (2)
  - Источники (50)
  - Размер (2)
  - Цифровая насыщенность (1)
  - География (17)
  - Персоны (16)
  - Компании (21)
- AND NOT:**
  - Нива \*\*\*\*
  - Сквириский комбинат хлебопродуктов \*\*\*\*
  - ИПС \*\*\*
  - Родной продукт \*\*\*
  - Сельхозсервис \*\*
  - Сельхозсервис \*
  - Озерьянка
  - Сельгоспсервис
  - Ридный продукт

Рис. 64- Отражение «гречневого кризиса» 2010 г.

Условия исследования состояния рынка с помощью электронных средств, в частности, при проведении антимонопольной дея-

тельности должны соответствовать современным условиям конкурентной разведки:

Во-первых, должны применяться методы и программные средства исследования информации полученной из открытых источников с соблюдением требований законодательства и этических норм.

Во-вторых, успех или неудача в решении практической задачи моделирования состояния рынка зависят от упрощения интегрированной информации, которую необходимо обработать.

В-третьих, достижение успеха при исследовании рынков связано с проблемой преодоления сложности доступа к информационным ресурсам из открытых источников, в том числе из сети Интернет.

Методология выявления антиконкурентных действий участников рынка по результатам анализа модели состояния рынка должна соответствовать возможностям имеющихся компьютерных средств и методов конкурентной разведки.

Например, данные, информацию и знания, получаемые в результате антимонопольных исследований, должны представляться в виде, соответствующем по структуре и форме разведывательной информации.

Современные средства, применяемые в конкурентной разведке в сетевой среде, обеспечивают:

- доступность необходимой части информации;
- огромный охват информации;
- оперативность, учет динамики информационных потоков.

В то же время, эти средства не могут заменить все инструменты, необходимые для антимонопольной деятельности. Для принятия решений в этой области требуется использование комплексных систем, которые позволяют добывать и обобщать информацию об объектах исследований из разных источников.

Таким образом, можно сделать вывод о том, что конкурентная разведка дополняет технологию поиска данных и информации в интернет-пространстве и целевое экстрагирование полезных понятий о состоянии и развитии товарного рынка с методами сбора, хранения, обработки и анализа данных, создает пространство интегрированной информации для анализа и формирования конкурентной политики.

Цели и средства антимонопольной деятельности обуславливают практические требования к созданию новых механизмов и технологий и требуют объединения различных по природе инструментов конкурентной разведки в соответствии с различными алгоритмами исследования.

## Заключение

---

---

Актуальность конкурентной разведки в последнее время значительно возросла. Это связано с такими процессами, как глобализация экономики, а, следовательно, и конкуренции, виртуализация экономики, развитие информационных технологий.

Широкому внедрению систем компьютерной конкурентной разведки способствуют и законодательные акты многих стран мира. Так, например, в США ещё в 1996 году был принят Закон о свободе информации, который обязал федеральные ведомства обеспечить гражданам свободный доступ ко всей своей информации. Ограничения касаются лишь материалов, имеющих отношение к национальной обороне, личных и финансовых документов, а также документов правоохранительных органов. Отказ в доступе к информации можно обжаловать в суде. Информация должна быть представлена в десятидневный срок, а споры разрешаться в течение 20 дней.

Во всем мире уже свыше 20 лет считают, что конкурентная разведка – это важнейшая функция современного менеджмента и главное условие динамичного и устойчивого развития бизнеса. Вместе с тем, как утверждает, гендиректор компании «Р-Техно» Роман Ромачев, «Если 10 лет назад конкурентные разведчики в первую очередь проверяли наличие у бизнес-партнеров криминальных связей, то сейчас они, как и на Западе, в большей степени добывают коммерческую информацию». Это подтверждают и данные исследования, проведенного Международным Центром конференций (МЦК) OnConferepse: большинство компаний используют конкурентную разведку для изучения состояния рынка (74 % респондентов) и конкурентов (64 %). Поиск, сбор и анализ информации помогает сформировать целостную картину конкурентной среды, установить причинно-следственные связи.

В настоящее время конкурентная разведка в сети Интернет обеспечивает доступность, огромный охват информации и высокую оперативность. Но она не может заменить другие виды и инструментальные средства бизнес-разведки. Для принятия серьезных решений необходимо использование комплексных систем, которые разрешают компоновать и обобщать информацию об объекте исследований, полученную из разных источников с применением разных технологий.

Об актуальности конкурентной разведки на основе интернет-ресурсов говорят многочисленные публикации, тренинги, конфе-

ренции. Сегодня задачи конкурентной разведки стимулируют развитие систем управления знаниями, глубинного анализа данных и текстов, с другой стороны наиболее развитые из этих систем в явном виде содержат аналитические блоки, специально ориентированные на задачи конкурентной разведки. Поэтому у пользователей имеется широкий выбор средств автоматизации аналитической деятельности. Причем уровни функциональности таких систем, может быть очень разнообразным – от простых информационно-поисковых программ, необходимых на этапе становления систем конкурентной разведки, до дорогих и ресурсоемких систем управления знаниями и глубинного анализа данных и текстов.

В настоящее время уже очевидно, что реальный прорыв в области интенсификации информационно-аналитической работы, как и в науке, возможен лишь в результате агрегирования различных направлений.

## Краткий глоссарий

---

---

**Автоматическое реферирование** [Automatic text summarization] – автоматическое формирование краткого изложения исходного текстового материала либо путем выделения фрагментов информационного наполнения и последующего их соединения, либо методом генерации текста на основании выявления знаний из оригинала.

**Авторское право** [Copyright] – совокупность правовых норм, регулирующих отношения, возникающие в связи с созданием, использованием (изданием, исполнением, показом и т. д.) произведений науки, литературы или искусства – результатов творческой деятельности людей. Программы для компьютеров и базы данных также охраняются авторским правом.

**Анализ социальных сетей** [Social Networks Analysis, SNA] – методология анализа социальных сетей. Предметом анализа в SNA, в отличие от большинства традиционных социологических исследований, являются не атрибуты отдельных личностей, а структура их взаимосвязей в рамках того или иного сообщества (рабочей группы). В рамках анализа социальных сетей рассматриваются социальные отношения с точки зрения теории сетей, состоящих из узлов – личностей, участников сети и связей – отношений между ними.

**Антимонопольная деятельность** [Antitrust Activities] – комплекс мер, направленных на ограничение деятельности монополий, а также создание соответствующего законодательства.

**База данных** [Database] – совокупность данных, организованных по определенным правилам, предусматривающим общие принципы описания, хранения и манипулирования, независимая от прикладных программ. Является информационной моделью предметной области.

**База данных аналитическая** [Analytical Database] – база данных, которая содержит информацию, получаемую из других баз данных в форме итоговой информации, представляет наибольший интерес для пользователя или группы пользователей.

**База данных полнотекстовая** [Full-Text Database] – база данных, в которой хранятся записи полнотекстовых документов или их частей.

**База данных фактографическая** [Factographic Database] – база данных, содержащая фактографические данные – информацию, относящуюся непосредственно к предметной области.



**Бенчмаркинг** [Benchmarking] - инструмент анализа конкурента. Процесс определения, понимания и адаптации имеющихся примеров эффективного функционирования компании с целью получения информации, которая помогла бы предпринять шаги, направленные на улучшение деятельности компании. В равной степени включает в себя два процесса: оценивание и сопоставление. Одно из направлений стратегически ориентированных маркетинговых исследований.

**Бизнес-процесс** [Business Process] – система последовательных, целенаправленных и регламентированных видов деятельности, в которой посредством управляющего воздействия и при поддержке определенных ресурсов входы процесса преобразуются в выходы, представляющие ценность для потребителей.

**Бизнес-разведка** [Business Intelligence, BI] – 1) сбор и обработка данных из разных источников для выработки управленческих решений в целях повышения конкурентоспособности коммерческой организации; 2) структурное подразделение предприятия, выполняющее эти функции.

**Блог** [Blog, Web Log] – сетевой дневник одного или нескольких авторов, состоящий из записей в обратном хронологическом порядке. С помощью сервиса блогов можно создать свой онлайн-дневник, читать и комментировать дневники других пользователей, принимать участие в сообществах по определенным темам, создавать свои сообщества.

**Блогосфера** [Blogsphere] – совокупность (коллекция) всех блогов в сети Интернет; общее название для совокупности блогов.

**Большие данные** [Big Data] – серия подходов, инструментов и методов обработки данных больших объемов и значительного многообразия для получения человеко-читаемых результатов, эффективных в условиях их непрерывного прироста, распределения по многочисленным узлам вычислительной сети. В качестве определяющих характеристик для больших данных отмечают «три V»: объем, в смысле величины физического объема, скорость в смысле скорости прироста, так и необходимости высокоскоростной обработки и получения результатов, многообразие, в смысле возможности одновременной обработки различных типов структурированных и полуструктурированных данных.

**Веб-аналитика** [Web Analytic] – измерение, сбор, анализ, представление и интерпретация информации о посетителях веб-сайтов в целях их улучшения и оптимизации. Основная задача веб-аналитики – мониторинг работы веб-сайтов, на основании которого определяется веб-аудитория и изучается поведение веб-

посетителей для принятия решений по развитию и расширению функциональных возможностей веб-ресурса.

**Веб-пространство** [Web Space] - совокупность сайтов в Интернете (гипертекстовое пространство Интернета, www-пространство).

**Веб-сайт** [Website] – набор веб-страниц, составляющих единое целое (посвященных одной тематике либо принадлежащих одному и тому же автору), как правило, размещенных на одном и том же сервере, имеющих одно и то же доменное имя и связанных между собой перекрестными ссылками. Для прямого доступа клиентов к веб-сайтам на серверах разработан протокол HTTP.

**Веб-форум** [Web-Forum] – класс веб-приложений для организации общения посетителей, веб-сайт, предназначенный для проведения онлайн-дискуссий.

**Виртуальная служба знакомств** [Online Dating Service] – интернет-сервис, предоставляющий пользователям Интернета услуги по виртуальному общению с другими пользователями, аналог реальных служб знакомств.

**Визуализация** [Visualization] – комплекс методов представления результатов анализа данных в наиболее удобной для восприятия и интерпретации форме. Может использоваться для мониторинга процесса построения и работы различных аналитических моделей, проверки гипотез и других целей, связанных с проведением анализа.

**Входная степень узла** [In-Degree] – количество ребер графа, которые входят в узел.

**Выходная степень узла** [Out-Degree] – количество ребер графа, которые выходят из узла.

**Геосоциальная сеть** [Geosocial Networking] – вид социальных сетей, в которых используются геокодирование. Пользователи оставляют данные о своем местонахождении, что позволяет объединять и координировать их действия на основании того, какие люди присутствуют в тех или иных местах, или какие события происходят в этих местах.

**Глубинный анализ данных** [Data Mining] – технология анализа данных в базах или хранилищах данных, основанная на статистических методах и служащая для выявления заранее неизвестных закономерностей, а также для поддержки принятия стратегически важных решений.

**Глубинный анализ текстов** [Text Mining] – технология извлечения информации из текстовых данных на основе обнаружения в них закономерностей. Как правило, включает этапы струк-

турирования исходного текста (обычно путем синтаксического анализа, добавления одних лингвистических структур и удаления других с последующей вставкой результатов в базу данных), поиска закономерностей в данных, оценивания и интерпретации результатов.

**Глубинный веб** [Invisible Web, Deep Web, Hidden Web] – часть веб-пространства, не индексируемую роботами поисковых систем. Информация, будучи недоступной для поиска, находится «в глубине» (англ. – Deep). Состоит из веб-страниц, динамически генерируемых по запросам к онлайн базам данных.

**Граф связей** [Communication Graph] – в социальных сетях – граф, предназначенный для идентификации связей между их участниками. С помощью графа можно визуализировать эти связи. Граф связей строится благодаря обмену контентом между людьми.

**Дайджест** [Digest] – информационный продукт (издание, статья, подборка), содержащий краткие аннотации и основные положения статей или в котором сжато передается содержание самых интересных публикаций за определенный период.

**Дескриптор** [(от лат. Descriptio – описание); Descriptor] – лексическая единица (слово, словосочетание, код) информационно-поискового языка, служащая для выражения основного смыслового содержания документов (текста). Используется для координатного индексирования документов и информационных запросов с целью последующего поиска.

**Диаметр графа** [Graph Diameter] – максимальное из расстояний между парами его вершин. Расстояние между вершинами определяется как наименьшее число ребер, которые необходимо пройти, чтобы добраться из одной вершины в другую.

**Живучесть системы** [System Survivability] – способность системы выполнять установленный минимальный объем своих функций при внешних воздействиях, не предусмотренных условиями нормальной эксплуатации, осуществлять выбор оптимального режима функционирования за счет собственных внутренних ресурсов, перестройки структуры, изменения функций отдельных подсистем и их поведения.

**Извлечение знаний** [Knowledge Extraction] – процесс получения из данных знаний в виде зависимостей, правил, моделей. Этапы: консолидация, очистка, трансформация, моделирование и интерпретация полученных результатов.

**Извлечение (экстрагирование) информации** [Information Extraction] – разновидность информационного поиска, при

которой из электронных документов выделяется некая структурированная информация, т.е. категоризированные, семантически значимые данные по какой-либо проблеме или вопросу.

**Извлечение фактов, понятий** (Feature Extraction) – технология, обеспечивающая получение информации в структурированном виде. Включает три основных метода: Entity Extraction – извлечение слов или словосочетаний, важных для описания содержания текста; Feature Association Extraction – выявление связей между извлеченными понятиями; Event and Fact Extraction – извлечение сущностей, распознавание фактов и событий.

**Имидж компании** [Corporate Image] – устойчивые представление, которое компания создает о себе с помощью рекламы, формируя благоприятное представление у целевой аудитории. Это устойчивое представление потребителей, клиентов, партнеров и общественности о престиже компании, качестве её товаров и услуг, репутации руководителей.

**Имитационное моделирование** (Simulation Modeling) – метод исследования, при котором исследуемая система замещается моделью, с достаточной точностью описывающей реальную систему. Эту модель используют для экспериментов с целью получить информацию о реальной системе. Имитационное моделирование – частный случай математического моделирования. Существует класс объектов, для которых по разным причинам не разработаны аналитические модели или методы решения относительно полученной модели. В этом случае математическая модель замещается имитатором или имитационной моделью – логико-математическим описанием объекта.

**Интернет** [Internet] – глобальная информационная сеть, части которой логически связаны единым адресным пространством, основанным на стеке протоколов TCP/IP, их последующих расширений или других IP-совместимых протоколов. Обеспечивает, исползует или делает доступным, публично или частным образом, коммуникационный сервис высокого уровня. Состоит из множества взаимосвязанных компьютерных сетей.

**Интернет-разведка** [Internet Intelligence] – сегмент конкурентной разведки, охватывающий процедуры сбора и обработки информации, проводимые с целью поддержки принятия управленческих решений, повышения конкурентоспособности коммерческих организаций исключительно из открытых источников из компьютерных сетей, большинство из которых являются надстроеными над сетью Интернет.

**Интернет-чистильщики** [Internet-Cleaners] – специалисты или службы, которые могут удалить из информационных ресурсов сети Интернет данные (как правило, негативную информацию о заказчике).

**Информационная безопасность** [Information Security] – состояние информации, информационных ресурсов и информационных систем, при котором с требуемой вероятностью обеспечивается защита информации (данных) от утечки, хищения, утраты, несанкционированного уничтожения, искажения, модификации (подделки), копирования, блокирования и т.п. Имеет три основные составляющие: конфиденциальность, целостность и доступность.

**Информационно-аналитическая деятельность** [Informational-Analytical Activity] – отрасль человеческой деятельности, призванная обеспечить информационные потребности общества с помощью аналитических и информационных технологий за счет обработки входной информации и получения качественно нового знания.

**Информационно-аналитическая система, ИАС** [Information-Analytical System, IAS] – класс информационных систем, предназначенных для аналитической обработки данных, а не для автоматизации повседневной деятельности организации. Объединяет, анализирует и хранит информацию, извлекаемую как из баз данных организации, так и из внешних источников. Входящие в состав ИАС хранилища данных обеспечивают преобразование больших объемов детализированных данных в обобщенную информацию, пригодную для принятия решений.

**Информационно-поисковая система, ИПС** [information retrieval system, IRS] – система, предназначенная для обеспечения поиска и отображения документов, представленных в базах данных. Ядро ИПС составляет поисковый механизм – программный модуль, который осуществляет поиск по запросу. ИПС, интегрированные с веб-технологиями, являются основой построения информационно-поисковых веб-серверов.

**Информационное воздействие** [Informational Influence] – возбуждение (торможение) в управляемой системе таких процессов, которые стимулируют желательный для управляющей стороны выбор. Этот способ воздействия на субъекта не предполагает, например, прямого выведения из строя части элементов его системы, но представляет собой передачу ему такой информации, которая натолкнет его на выбор определенного решения, при котором эти элементы потеряют свою эффективность.

**Информационные операции** [Information Operations] – информационное воздействие на массовое сознание (как на враждебное, так и на дружеское), воздействие на информацию, доступную конкуренту и необходимую ему для принятия решений, а также на информационно-аналитические системы (ИАС) конкурента, в том числе действия, направленные на физическое поражение ИАС, вывод из строя средств компьютерно-телекоммуникационной инфраструктуры.

**Информационные ресурсы** [Information Resources] – отдельные документы и отдельные массивы документов, документы и массивы документов в информационных системах, зафиксированные на соответствующих носителях информации, а также языковые средства, применяемые для описания конкретной предметной области и для доступа к данным и знаниям.

**Информационные объекты, ИО** [Informational Objects] – объекты, содержащие (несущие) информацию. Могут описываться непосредственно или в виде алгоритма их порождения.

**Информационный портрет** [Informational Portrait] – документ, который характеризует в компактной форме основное содержание текста - описанные в нем предметы, лица, ситуации и т.п.

**Капча** [Captcha] – нечеткое графическое изображение букв и цифр, которые требуется ввести с клавиатуры в определенное поле.

**Классификация** [Classification] – система распределения объектов по классам в соответствии с определенным признаком (основание классификации). Объекты необходимо классифицировать для выявления общих свойств информационного объекта, который определяется информационными параметрами (реквизиты). При классификации нужно соблюдать требования: полнота охвата; однозначность реквизитов; возможность включения новых объектов.

**Кластерный анализ** [Cluster Analysis] – многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов и затем упорядочивающая объекты в сравнительно однородные группы (кластеры).

**Клики** [Cliques] – подгруппы или кластеры, в которых узлы связаны между собой сильнее, чем с членами других клик.

**Коммерческая тайна** [Trade Secret] – сведения конфиденциального характера из любой сферы деятельности государственного или частного предприятия, разглашение которых может нанести материальный или моральный ущерб ее владельцам или пользователям (юридическим лицам).

**Конкурентная разведка** [Competitive Intelligence] – спланированные действия по систематическому сбору и анализу информации, проводимые с целью поддержки принятия управленческих решений, повышения конкурентоспособности коммерческих организаций.

**Конкурентная среда** [Competitive Environment] – результат и условия взаимодействия большого количества субъектов рынка. Образуется не только и не столько собственно субъектами рынка, взаимодействие которых вызывает соперничество, но в первую очередь – отношениями между ними.

**Консалтинг** [Consulting] – деятельность специализированных маркетинговых компаний, консультирующих производителей, продавцов, покупателей по вопросам в сфере экономики, управления, сбыта, ценообразования, продвижения продукции и др.

**Консалтинговая компания (фирма)** [Consulting Company (Firm)] – компания (фирма), выполняющая консультационные услуги по исследованию и прогнозированию рынков, разработке маркетинговых программ, поиску путей выхода из кризисных ситуаций; и др.

**Консолидированная информация** [Consolidated Information] – полученные из нескольких источников и интегрированные разнотипные информационные ресурсы (знания), которые в совокупности обладают признаками полноты, целостности, непротиворечивости и составляют адекватную информационную модель проблемной области с целью ее анализа обработки и использования в процессах поддержки принятия решений.

**Контент** [Content] – содержательное наполнение информационных ресурсов (напр., веб-сайтов) – тексты, графика, мультимедиа. Параметрами контента является его объем, актуальность и релевантность.

**Контент-анализ** [Content Analysis] – анализ содержания документов, который нацелен на измерение ряда качественных и количественных характеристик текста и на анализ зависимостей между ними.

**Контент-мониторинг** [Content Monitoring] – систематическое, непрерывное во времени сканирование и контент-анализ информационных ресурсов.

**Конфиденциальная информация** [Sensitive Information] – информация, которая представляет собой коммерческую или личную тайны и охраняется ее владельцем.

**Конфиденциальность** [Confidentiality] – свойство защищенности информации от несанкционированного доступа и попыток ее

раскрытия пользователями, не имеющими соответствующих полномочий.

**Коэффициент кластеризации** [Clustering Coefficient] – величина, соответствующая уровню связности узлов в сети. Показывает, сколько ближайших соседей данного узла являются ближайшими соседями друг для друга, и равна отношению реального количества ребер, которые соединяют ближайших соседей данного узла, к максимально возможному.

**Коэффициент посредничества** [Betweenness] – параметр, показывающий, сколько кратчайших путей проходит через узел. Указывает на роль данного узла в установлении связей в сети.

**Коэффициент центральности** [Centrality] – параметр, который показывает «важность» или «влияние» определенного узла (кластера узлов) внутри графа (сети). Стандартные методы измерения «центральности» охватывают расчет центральности по посредничеству, центральность по близости, центральность собственного вектора, центральность по степени и др.

**Лицо, принимающее решение, ЛПР** [Decision Maker] – субъект (менеджер), наделенный определенными полномочиями и несущий ответственность за последствия принятого и реализованного управленческого решения. ЛПР – один или несколько человек (коллектив), на которых лежит ответственность за принятое решение.

**Малый мир** [Small Worlds] – один из видов графов, в котором большинство узлов не являются попарно соседними, но могут связываться друг с другом благодаря небольшому количеству переходов по ребрам графа. Граф (сеть) считается малым миром, если расстояние между двумя любыми случайно выбранными узлами в большинстве случаев не превышает двоичного логарифма от общего количества узлов.

**Математическое моделирование** [Mathematical Modeling] – процесс построения и изучения математических моделей – математических представлений о реальности.

**Медиаактивность** [Media Activity] – деятельность индивида по поиску, получению, потреблению, передаче, производству, распространению информации.

**Метапоисковая система** [Metasearch System] – поисковая система, не имеющая своего индекса, способная передавать запросы пользователя одновременно нескольким поисковым серверам, отбирать самые релевантные результаты, объединить их и представлять пользователю в виде документа со ссылками.



**Мультиагентная система, МАС** [Multi-Agent System, MAS] – система, образованная несколькими взаимодействующими интеллектуальными агентами. МАС могут быть использованы для решения таких проблем, которые сложно или невозможно решить с помощью монолитной системы.

**Недобросовестная конкуренция** [Unfair Competition] – нарушение общепринятых правил и норм конкуренции. Силовые и незаконные методы конкуренции (лишение конкурентов сырья, рынков сбыта, сбивание цен, промышленный шпионаж и т.п.).

**Неструктурированная текстовая информация** [Unstructured Text Information] – нестандартизированный и неформализованный текст, состоящий из предложений на естественном языке. Содержание текста – полнотекстовое изложение идей, смыслов и сюжетов (свободный текст).

**Обнаружение знаний** [Knowledge Discovery] – методика извлечения знаний (сведений) из источников информации.

**Обнаружение знаний в тексте** [Knowledge Discovery in Text] – процесс обнаружения новых, потенциально полезных и понятных шаблонов в неструктурированных текстовых данных.

**Обнаружение знаний в базах данных** [Knowledge Discovery in Databases, KDD] – процесс обнаружения полезных знаний в базах данных. Эти знания могут быть представлены в виде закономерностей, правил, прогнозов, связей между элементами данных и др. Главным инструментом KDD являются технологии Data Mining.

**Общество аналитиков и профессионалов конкурентной разведки** [Analysts Society and Competitive Intelligence Professionals] – областная общественная организация специалистов в области конкурентной разведки, созданная на базе кафедры Социальной информатики Харьковского национального университета радиоэлектроники в 2002 г.

**Онтология** [Ontology] – формализация некоторой области знаний с помощью концептуальной схемы, состоящей из структуры данных, их связи и правила, принятые в этой области. Сферы применения – моделирование бизнес-процессов, Семантический веб, искусственный интеллект.

**Открытые источники** [Open Sources] – информационные источники, легально распространяющие сведения, доступ к которым возможен на законных основаниях. Легальность и законность рассматривается только в контексте юрисдикции территории, на которой ведутся хозяйственные или иные операции.

**Персональные данные** [Personal Data] – любая информация, относящаяся к определенному или определяемому на основании такой информации физическому лицу, в том числе его фамилия, имя, отчество, год, месяц, дата и место рождения, адрес, семейное, социальное, имущественное положение, образование, профессия, доходы, другая информация. Персональные данные относятся к категории конфиденциальной информации; сведения или совокупность сведений о физическом лице, которое идентифицировано или может быть конкретно идентифицировано.

**Пиринговая сеть** [Peer-to-Peer, P2P] – компьютерная сеть, основанная на равноправии участников. В такой сети отсутствуют выделенные серверы, а каждый узел (peer) является как клиентом, так и сервером. В отличие от архитектуры клиент-сервера, позволяет сохранять работоспособность сети при любом количестве и любом сочетании доступных узлов.

**Поисковая оптимизация** [Search Engine Optimization, SEO] – комплекс мер для улучшения позиций веб-сайта в результатах выдачи сетевых поисковых систем по определенным запросам пользователей.

**Построение семантической сети** [Construction of Semantic Network] – одна из основных задач, решаемая в Text Mining – поиск ключевых понятий текста и установление взаимоотношений между ними, формирование структуры для представления знаний в виде ориентированного графа, в котором вершины – это понятия, а дуги – отношения. По такой сети можно осуществлять контекстную навигацию.

**Промышленная контрразведка** [Industrial Counter-Espionage] – деятельность по предупреждению промышленного шпионажа.

**Промышленный шпионаж** [Industrial Espionage] – форма недобросовестной конкуренции, при которой осуществляется незаконное получение, использование, разглашение информации, составляющей коммерческую, служебную или иную охраняемую законом тайну с целью получения преимуществ при осуществлении предпринимательской деятельности, а равно получения материальной выгоды.

**Разведка** [Exploration, Intelligence] – практика и теория сбора информации о противнике или конкуренте для обеспечения безопасности и получения преимуществ в области вооруженных сил, политики или экономики. Разведка может использовать как легальные методы сбора информации, так и нелегальные операции, попадающие под понятие «шпионаж».

**Разведывательная деятельность** [Intelligence Activities] – деятельность, включающая сбор информации, оценку ее достоверности и объединение отдельных фактов в общую картину.

**Разведывательная информация** [Intelligence Information] – осмысленные сведения, основанные на собранных, оцененных и истолкованных фактах, полученные в результате отбора, сопоставления, логической увязки и обобщения разведывательных данных и сведений в соответствии с заданием потребителя.

**Разведывательный цикл** [Intelligence Cycle] – в рамках конкурентной разведки – процессы, описывающие: целеуказание, сбор, обработку и анализ разведывательной информации, доведение целевой информации и выводов до заказчика.

**Релевантность** [Relevance] – мера соответствия получаемого результата желаемому. В информационном поиске – мера соответствия результатов поиска задаче поставленной в поисковом запросе.

**Репутация** [Reputation] – социальная оценка группы субъектов о человеке, группе людей или компании, сформировавшуюся на основе некоторых критериев.

**Репутация компании** [Company Reputation] – это комплекс оценочных представлений целевых аудитории о компании, сформированный на основе факторов репутации, имеющих значение для этой аудитории.

**Ретроспективная информация** [Retrospective Information] – сведения, содержащиеся в массивах данных, накопленных за значительный период времени, или полученные в результате поиска в этих массивах.

**Ретроспективный анализ** [Retrospective Analysis] – анализ, заключающийся в изучении тенденций, сложившихся за определенный период времени в прошлом.

**Риск** [Risk] – ситуативная характеристика деятельности, состоящая в неопределенности ее исхода и возможных неблагоприятных последствий в случае неуспеха.

**Российское общество профессионалов конкурентной разведки, РОПКР** [Russian Society of Competitive Intelligence Professionals, RSCIP] – зарегистрированное в г. Моск-ве в 2002 г. юридическое лицо в виде некоммерческой организации в организационно-правовой форме некоммерческого партнерства. В состав основных целей РОПКР входит содействие продвижению, становлению, признанию, легитимности и развитию конкурентной разведки в России и странах СНГ, создание условий для объединения специалистов профессионально занимающихся конкурентной развед-

кой и специалистов других сфер деятельности, интересующихся теорией и практикой конкурентной разведки, создание условий для признания новой массовой профессии – специалист по конкурентной разведке.

**Сбор данных** [Data Collection] – процесс идентификации и получения данных от различных источников, группирования полученных данных и представление их в форме, необходимой для ввода в компьютер.

**Семантическая сеть** [Semantic Network] – способ представления знаний в виде ориентированного графа, в котором вершины соответствуют семантическим единицам языка (понятиям, объектам, действиям, ситуациям и т.п.), а ребра – свойствам или отношениям между ними.

**Сетевая аналитика** [Network Analytics] – совокупность средств и методов сбора из сетевой среды (в частности, из сети Интернет), преобразования, хранения, анализа, моделирования, доставки и трассировки данных, информации и знаний при работе над задачами, связанными с принятием решений.

**Сетевая мобилизация** [Network Mobilization] – процесс объединения усилий участников социальных сетей для решения некоторых проблем, например, организации массовых выступлений, отражения агрессии, помощи пострадавшим и т.п. Возможности сетевой мобилизации зависят от структуры сети, ее топологии, параметров, динамики информации, циркулирующей в ней, возможности и вероятности восприятия информации узлами сети, возможности преобразования информации в узлах сети, возможности восстановления связей в сети после деструктивного воздействия на них.

**Система конкурентной разведки** [Competitive Intelligence Systems, CIS] – инфраструктура для осуществления конкурентной разведки; комплексная информационно-аналитическая система (ИАС) поддержки принятия решений в части анализа изменений условий ведения бизнеса и политической деятельности, на основе которого вырабатывается стратегия и тактика превентивных мероприятий, направленных на достижение конкурентных преимуществ и предотвращение влияния негативных факторов экономической и политической среды.

**Слабые социальные связи** [weakly social connections] – свойство социальных сетей, заключающееся в наличие связей (ребер с малыми весами) между удаленными в каком-то смысле узлами (например, отношения с далекими знакомыми и коллегами). Если эти связи проигнорировать, то сеть распадется на отде-

льные фрагменты. Слабые связи являются тем феноменом, который связывает социальную сеть в единое целое.

**Слияния и поглощения** [mergers and acquisitions, M&A] – класс экономических процессов укрупнения бизнеса и капитала, происходящих на макро- и микроэкономическом уровнях, в результате которых на рынке появляются более крупные компании взамен нескольких менее значительных.

**Сообщество практиков конкурентной разведки, СПКР** [Community Practice Competitive Investigation; CPCI] – сообщество практиков конкурентной разведки в России, существующее с 2004 г. Изначально СПКР было де-факто создано на Интернет-Форуме Бизнес-разведчиков, затем расширилось, принимая в свои ряды новых специалистов из России, Украины, Беларуси. Одним из приоритетов в деятельности СПКР является активная пропаганда и продвижение конкурентной разведки в России и странах СНГ.

**Социальная сеть** [Social Network] – социальная структура, состоящая из узлов (которыми являются социальные объекты) и связей между ними. Объектами сетей могут быть предприятия, люди, Интернет-ресурсы и т.д. Существует множество социальных сетей, имеющих свою специфику, свойственные только им особенности, однако современные методы анализа данных применимы для любой из них вне зависимости от специфики.

**Социальные медиа** [Social Media] – совокупность онлайн-сервисов и интернет-приложений, которые позволяют пользователям общаться друг с другом в том числе, и в режиме реального времени. При этом пользователи могут обмениваться между собой мнениями, новостями, информацией, в том числе и мультимедийной. Социальные медиа базируются идеологической и технологической базе веб 2.0, позволяющих создание и обмен контентом, созданным самими пользователями (User-Generated Content).

**Стратегическая деловая разведка** [Strategic BI] – разведка, оказывающая помощь управлению в разработке их целостных планов и в проверке эффективности процесса видения. Охватывает сканирование окружения, анализ структуры отрасли, конкурентный анализ, анализ сценариев (планов действий), управление вопросами, технологическое прогнозирование, разработку типов конкурентных личностей и др.

**Сценарий** [Script, Scenario] – план выполнения процесса; определяет последовательность команд, которая указывает программе, как и в каком порядке, выполнять ту либо иную процедуру.

**Сценарное планирование** [Scenario Planning] – планирование вариантов развития событий (сценариев).

**Тактическая деловая разведка** [Tactical BI] – разведка, помогающая компании в ее повседневной работе, используемая сотрудниками непосредственно на своих участках на уровне ежедневного контроля. Охватывает анализ нужд покупателя, цены конкурента, анализ продукции и услуг конкурента, производства конкурента и др.

**Теория сложных сетей** [complex networks] – междисциплинарная область знаний, возникшая на базе эмпирических исследований реальных сетей, прежде всего, компьютерных и социальных. В рамках этой теории сложная сеть представляет собой граф (сеть) с нетривиальными топологическими особенностями, которые не встречаются в простых сетях, таких как решетки или случайные графы, но часто встречаются в реальности. Теория сложных сетей изучает характеристики сложных сетей, учитывая не только топологию сетей, но и статистические феномены, распределение весов отдельных вершин и ребер, эффекты протекания и проводимости в сетях и т.п.

**Управление знаниями** [Knowledge Management] – процессы, благодаря которым создаются, сохраняются, распределяются и применяются основные элементы интеллектуального капитала, необходимые для успеха организации. Существует пять основных технологий, которые поддерживает управление знаниями: бизнес-разведка (Business Intelligence), сотрудничество (Collaboration), трансфер знаний (Knowledge Transfer), обнаружение знаний (Knowledge Discovery) и определение экспертов (Expertise Location).

**Управление репутацией** [Reputation Management] – методы мониторинга репутации персоны или компании, выявления фактов которые вредят ей, и использование каналов обратной связи с потребителем для реакции или раннего выявления возможных негативных последствий для репутации.

**Управление репутацией в Интернете** [Online Reputation Management, ORM] – один из современных способов манипулирования интернет-контентом (популяризация информативных площадок, написание пресс-релизов, статей и отзывов) с целью создания положительного или отрицательного образа компании или персоны в интернете. Для осуществления ORM требуются специалисты: копирайтеры, редакторы, SEO-специалисты, дизайнеры и программисты. Цель специалистов ORM – получить привлекательный образ компании или персоны, повысить прибыльность и эффективность предприятия.

**Управление рисками** [Risk Management] – процесс принятия и выполнения управленческих решений, направленных на сниже-

ние вероятности возникновения неблагоприятного результата и минимизацию возможных потерь, вызванных его реализацией. Цель риск-менеджмента в сфере экономики – повышение конкурентоспособности хозяйствующих субъектов с помощью защиты от реализации рисков.

**Управленческое решение** [Management Decision] – директивный акт целенаправленного воздействия на объект управления, основанный на анализе данных, характеризующих конкретную управленческую ситуацию, определение цели действий, и содержащий программу достижения цели.

**Фактографическая база данных** [Factographic Database] – база данных, содержащая фактографические данные – информацию, относящуюся непосредственно к предметной области.

**Фактографическая информация** [Factographic Information] – описание фактов, сгруппированных по определенным системообразующим признакам.

**Фотохостинг** [Photo Hosting] – веб-сайт, позволяющий публиковать изображения (например, цифровые фотографии) в сети Интернет. Фотохостинг может использоваться для размещения, хранения и показа изображений другим пользователям сети. Основное преимущество, которое предоставляет фотохостинг пользователям – удобство демонстрации фотографий. Автор может легко поделиться гиперссылкой, ведущей на фотографию, с любым человеком, имеющим доступ к сети Интернет.

**Цифровая тень** [Digital Shadow] – информация о пользователе, создаваемая без его участия, которая возникает и накапливается, когда кто-то ищет пользователя через поисковые системы, происходит электронная почтовая рассылка по спискам, в которых он фигурирует и во многих других случаях. Кроме «цифровых теней открытого доступа», создаются и копируются «цифровые тени ограниченного доступа» – записи камер наблюдения, банковские транзакции, биллинги интернет-магазинов, сервисов продажи билетов, телефонных звонков и др.

**Цифровой след** [Digital Footprint] – информация, которая остается самим пользователем при работе в Сети и по которой можно не только его идентифицировать, но и «привязать» к определенным действиям, событиям, восстановить какие-то фрагменты биографии.

**Экспертная оценка** [Expert Estimates] – основанная на суждениях специалистов количественная или качественная оценка процессов или явлений, не поддающихся непосредственному измерению.

**Экстрактор** [Extractor] - программа, собирающая данные из исходных систем (выделение в тексте сложных элементов и специальных конструкций, отличающихся особым видом написания, – наименований юридических лиц, товаров, адресов, номеров и т.п.).

**Эксцентричность узла** [Eccentricity] – наибольшее из геодезических расстояний (минимальных расстояний между узлами) от данного узла сети к другим.



## Литература

---

---

[Астафьева, 1996] Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения // Успехи физических наук, 1996. – 166. – № 11. – Р. 1145-1170.

[Берд, 2007] Берд К. Модель OSINT // Компьютерра, 2007. – № 22.

[Горбулін, 2009] Горбулін В.П., Додонов О.Г., Ланде Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія. – К.: Інтертехнологія, 2009. – 164 с.

[Григорьев, 2007] Григорьев А.Н., Ландэ Д.В., Бороденков С.А., Мазуркевич Р.В., Пацьора В.Н. InfoStream. Мониторинг новостей из Интернет: технология, система, сервис: научно-методическое пособие. – Киев: Старт-98, 2007. – 40 с.

[Губанов, 2009] Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Модели репутации и информационного управления в социальных сетях // Математическая теория игр и ее приложения, 2009. – № 2. – С. 14-37.

[Джилад, 2010] Джилад Б. Конкурентная разведка. Как распознавать внешние риски и управлять ситуацией – СПб.: Питер, 2010. – 320 с.

[Додонов, 2009] Додонов О.Г., Ланде Д.В., Путятин В.Г. Інформаційні потоки в глобальних комп'ютерних мережах. – К: Наук. думка, 2009. – 295 с.

[Додонов, 2010] Додонов А.Г., Ландэ Д.В. Живучесть информационных сюжетов // Материалы XI Международной научно-практической конференции «Информационная безопасность». – Ч. 2. – Таганрог: Изд-во ТТИ ЮФИ, 2010. – С. 179-183.

[Додонов, 2011] Додонов А.Г., Ландэ Д.В. Живучесть информационных систем. – К.: Наук. думка, 2011. – 256 с.

[Додонов, 2013] Додонов А.Г., Ландэ Д.В., Коженевский С.Р., Путятин В.Г. Компьютерные информационно-аналитические системы и хранилища данных. Толковый словарь. – К.: Феникс; ИПРИ НАН Украины, 2013. – 554 с.

[Доронин, 2011] Доронин А. Бизнес-разведка. – М.: Ось-89, 2003. – 704 с.

[Дудихин, 2004] Дудихин В.В., Дудихина О.В. Конкурентная разведка в Интернет. – М.: АСТ, НТ Пресс, 2004. – 240 с.

[Ермаков, 2005] Ермаков Н.С., Иващенко А.А., Новиков Д.А. Модели репутации и норм деятельности. М.: ИПУ РАН, 2005. – 67 с.

[Иващенко, 2006] Основы методики розслідування незаконного збирання та розголошення комерційної таємниці // Юридичний журнал, 2006. – № 8. – С. 48-66.

[Ивлев, 2008] Ивлев А.А. Основы теории Бойда. Направления развития, применения и реализации (монография). – М., 2008. – 64 с.

[Калиновский, 2012] Калиновский Я.А., Бояринова Ю.Е. Высокоразмерные изоморфные гиперкомплексные числовые системы и их использование для повышения эффективности вычислений. – К.: Инфодрук, 2012. – 183 с.

[Киселев, 2005] Киселев С. Модель информационной системы бизнес-разведки // Открытые системы, 2005. – № 5-6. – С. 60-66.

[Ковальчук, 2012] Ковальчук А. Практика и секреты заработка в Интернете. Управление репутацией // Выпуск 30, 2012 (on-line: <http://www.trustlink.ru/subscribe/show/35>)

[Кондратьев, 2010] Кондратьев А. Разведка с использованием открытых источников информации в США // Зарубежное военное обозрение, 2010. – №9. – С. 28-32.

[Кононов, 2003] Кононов Д.А., Кульба В.В., Шубин А.Н. Базисные понятия моделирования информационного управления в социальных системах // Труды международной научно-практической конференции «Теория активных систем». – М.: Институт проблем управления им. В.А. Трапезникова РАН, 2003. –Т 2. – С. 125-129.

[Кочергов, 2009] Кочергов Д. Один шаг, который может стать последним // Экономика бизнеса, 2009. – № 13 (9279).

[Кузнецов, 2006] Кузнецов С.В. Как вести бизнес-разведку в «невидимом» интернете? // «СNews», 07.09.06.

[Кульба, 1999]: Кульба В.В., Малюгин В.Д., Шубин А.Н., Вус М.А. Введение в информационное управление. Учебно-методическое издание. – СПб.: Изд-во С.-Петербургского ун-та. 1999. – 116 с.

[Кульба, 2004] Кульба В.В., Кононов Д.А., Косяченко С.А., Шубин А.Н. Методы формирования сценариев развития социально-экономических систем. – М.: СИНТЕГ, 2004. – 296 с.

[Ландэ, 2005] Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа. – М.: Диалектика, 2005. – 272 с.

[Ландэ, 2007] Ландэ Д.В., Снарский А.А., Брайчевский С.М., Дармохвал А.Т. Моделирование динамики новостных текстовых потоков // Интернет-математика 2007: Сборник работ участников конкурса. – Екатеринбург: Изд-во Урал. ун-та, 2007. – С. 98-107.

[Ландэ, 2009] Ландэ Д.В., Снарский А.А., Безсуднов И.В. Интернетика: Навигация в сложных сетях: модели и алгоритмы. – М.: Либроком (Editorial URSS), 2009. – 264 с.

[Ландэ, 2010] Ландэ Д.В. Глубинный web – информационная среда для бизнес-аналитика // Информационные технологии для менеджмента, 2010. – № 9. – С. 28-32.

[Ландэ, 2013] Ландэ Д.В. Метод визуализации зон нестабильности в рядах измерений // Информационные технологии и безопасность. Оценка состояния: Материалы международной научной конференции ИТБ-2013. – К.: ИПРИ НАН Украины, 2013. – С. 105-113.

[Ландэ, Брайчевский, 2010] Ландэ Д.В., Брайчевский С.М., Дармохвал А.Т., Жигало В.В. Архитектура системы охвата информационных связей объектов мониторинга // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». – Вып. 9 (16). – М.: Изд-во РГГУ, 2010. – С. 272-278.

[Ландэ, Прищепа, 2007] Ландэ Д., Прищепа В. Школа веб-разведки. Инструменты и источники // Телеком, 2007. – № 7-8. – С. 46-49.

[Ланде, Фурашев, 2012] Ланде Д.В., Фурашев В.М. Основи інформаційного і соціально-правового моделювання: монографія. – К.: ПанТот, 2012. – 144 с.

[Нежданов, 2010] Нежданов И. Технологии разведки для бизнеса. – М.: Ось-89, 2009. – 400 с.

[Новиков, 2002] Новиков Д.А., Чхартишвили А.Г. Теория управления организационными системами – М.: Синтег, 2002. – 227 с.

[Новиков, 2007] Новиков Д.А. Теория управления организационными системами. 2-е изд. – М.: Физмалит, 2007. – 584 с.

[Печенкин, 2004] Печенкин И.А. Информационные технологии на службе разведки // Конфидент, 2004. – № 4. – С. 28-41.

[Прескотт, 2003] Прескотт Джон Е., Миллер Стивен Х. Конкурентная разведка: Уроки из окопов. – М.: Альпина Паблишер, 2003. – 336 с.

[Расторгуев, 2006] Расторгуев С.П. Информационная война. Проблемы и модели. Экзистенциальная математика. – М.: Гелиос АРВ, 2006. – 240 с.

[Хан, 2000] Хан У., Мани И. Системы автоматического реферирования // Открытые системы, 2000. – № 12.

[Хорошевский, 2013] Хорошевский В.Ф. Семантические технологии: ожидания и тренды // Открытые Семантические технологии проектирования интеллектуальных систем – Open Semantic Technologies for Intelligent Systems (OSTIS-2012): материалы II Между-

нар. научн.-техн. конф. (Минск, 16-18 февраля 2012 г.). – Минск: БГУИР, 2012. – С. 143-158.

[Черных, 2013] Черных Е. Нам от АНБ не спрятаться, не скрыться // Комсомольская правда, 24 июня 2013.

[Чертов, 2009] Четов О.Р. Поліноми Кунченка для розпізнавання образів // Вісник НТУУ «КПІ» Інформатика, управління та обчислювальна техніка, 2009. – № 50. – С. 105-110.

[Чхартишвили, 2004] Чхартишвили А.Г. Теоретико-игровые модели информационного управления. М.: ЗАО «ПМСОФТ», 2004. – 227 с.

[Bak, 1996] Bak P. How nature works: The science of self-organized criticality. – New York: Springer-Verlag Inc., 1996. – 212 p.

[Bhargava, 1993] Bhargava S.C., Kumar A., Mukherjee A. A stochastic cellular automata model of innovation diffusion // Technological forecasting and social change, 1993. – 44. – № 1. – P. 87-97.

[Bjorneborn, 2004] Bjorneborn L., Ingwersen P. Toward a basic framework for webometrics. Journal of the American Society for Information Science and Technology, 2004. – 55(14): 1216-1227.

[Buckheit, 1995] Buckheit J., Donoho D. Wavelab and reproducible research // Stanford University Technical Report 474: Wavelets and Statistics Lecture Notes, 1995. – 27 p.

[Burbary, 2009] Burbary K., Cohen A. A Wiki of Social Media Monitoring Solutions // (on-line: <http://wiki.kenburbary.com/>)

[Burke, 2001] Burke M.M. Knowledge Operations: above and beyond Information Operations // 6th International Command and Control Research and Technology, June 19 – 21, 2001. – 16 p.

[Clauset, 2008] Clauset, A., Moore, C., Newman, M.E.J. Hierarchical structure and the prediction of missing links in networks // Nature, 2008. – 453, 98-101.

[DoD, 2003] Information operations roadmap – DoD US, 30 october 2003. – 78 p.

[Erdős, 1960] Erdős P., Rényi A. On the evolution of random graphs, Publ. Math. Inst. Hungar. Acad. Sci. 5, 1960. – P. 17-61.

[He, 2007] He B., Patel M., Zhang Z., Chang K. C.-C. Accessing the Deep Web: A Survey // Communications of the ACM (CACM), 50(5):94-101, 2007.

[Hill, 2000] Hill J.M.D., Surdu J.R., Ragsdale D.J., Schafer, J.H. Anticipatory planning in information operations // Systems, Man, and Cybernetics, 2000 IEEE International Conference, 2000. – 4. – P. 2350-2355.

[Kacperski, 2000] Kacperski K., Holyst J.A. Physica A. Phase transitions as a persistent feature of groups with leaders in models of opin-

- ion formation // *Statistical Mechanics and its Applications*, 2000. – 287, Issues 3-4. – P 631-643.
- [Knight, 2003] Knight J.C., Strunk E.A., Sullivan K.J. *Towards a Rigorous Definition of Information System Survivability* // *Proceedings of the DARPA Information Survivability Conference and Exposition (DISCEX'03)*, 2003.
- [Lande, 2012] Lande D.V., Kalinovskiy Ya.A., Boyarinova Yu. E. The model of information retrieval based on the theory of hypercomplex numerical systems // *Preprint Arxiv 1205.3031*. (on-line: <http://arxiv.org/abs/1205.3031>)
- [Lasswell, 1948] Lasswell H.D. *The structure and function of communication in society* // *The Communication of Ideas*. / Ed.: L. Bryson. – New York: Harper and Brothers, 1948.
- [Latane, 1981] Latane B. *The psychology of social impact* // *American Psychologist*, 1981. – 33. – P. 343-356.
- [Latane, 1997] Latane B., Nowak A. *Causes of polarization and clustering in social groups* // *Progress in communication sciences*, 1997. – 13. – P. 43-75.
- [Lewenstein, 1993] Lewenstein M., Nowak A., Latane B. *Statistical mechanics of social impact* // *Physical Review*, 1993. – A, 45. – P. 763-776.
- [Li, 2012] Li Y., Miller E.L., Long D.D.E. *Understanding Data Survivability in Archival Storage Systems* // *Proceedings of the 5th Annual International Systems and Storage Conference (SYSTOR 2012)*, June 4–6, 2012, Haifa, Israel.
- [Milgram, 1967] Milgram S. *The small world problem*, *Psychology Today*, 1967. – 2. – P. 60-67.
- [Newman, 2003] Newman M.E.J. *The structure and function of complex networks* // *SIAM Review*, 2003. – 45. – P. 167-256.
- [Nowak, 1990] Nowak A., Szamrej J., Latane B. *From private attitude to public opinion: A dynamic theory of social impact* // *Psychological Review*, 1990. – 97. – P. 367-376.
- [Osgood, 1954] Osgood Ch. E. *Psycholinguistics. A Survey of Theory and Research Problems* // *Supplement to the International Journal of American Linguistics*. Vol. 20. No 4. Oct. 1954, mem. 10. Baltimore: Waverly Press, 1954.
- [Price, 2001] Price G., Sherman C., Sullivan D. *The Invisible Web: Uncovering Information Sources Search Engines Can't See*. – Information Today, Inc., 2001. – 439 p.
- [Roberts, 2002] Roberts P.W., Dowling G. R. *Corporate reputation and sustained superior financial performance* // *Strategic Management Journal*, 2002. – 23. – № 12. – P. 1077–1093.

[Sobkowicz, 2003] Sobkowicz P. Effect of leader's strategy on opinion formation in networked societies // Preprint Arxiv (on-line: <http://arxiv.org/pdf/cond-mat/0311566>)

[Schramm, 1974] Schramm W., D.F.Roberts (eds.) The Process and Effects of Mass Communication. Univ. of Illinois Press, 1974.

[Watts, 1998] Watts D.J., Strogatz S.H. Collective dynamics of «small-world» networks. // Nature, 1998. – 393. – P. 440-442.

## **Веб-сайты по тематике конкурентной разведки**

1. Сообщество Практиков Конкурентной Разведки, СПКР ([razvedka-open.ru](http://razvedka-open.ru))
2. Российское общество профессионалов конкурентной разведки, РОПКР ([www.rscip.ru](http://www.rscip.ru))
3. Международное Общество профессионалов конкурентной разведки SCIP ([www.scip.org](http://www.scip.org))
4. Академия конкурентной разведки Fuld-Gilad-Herring, Кембридж ([www.academyci.com](http://www.academyci.com))
5. Институт конкурентной разведки, Германия ([www.institute-for-competitive-intelligence.com/start.html](http://www.institute-for-competitive-intelligence.com/start.html))
6. Бизнес-школа Skema, Франция (<http://www.skema-bs.fr/faculte-recherche/centre-intelligence-economique-et-influence>)
7. Объединение профессионалов конкурентной разведки со штаб-квартирой в Канаде Competia ([www.competia.com](http://www.competia.com))
8. Харьковская областная общественная организация «Общество аналитиков и профессионалов конкурентной разведки» ([www.scip.org.ua](http://www.scip.org.ua))
9. «Knowledge Camp & Competitive Intelligence Camp» – украинский BarCamp по конкурентной разведке и менеджменту знаний ([barcamp2010.scip.org.ua](http://barcamp2010.scip.org.ua))
10. Частная разведывательная компания «Р-Техно» ([www.r-techno.com](http://www.r-techno.com))
11. Агентство конкурентной разведки «Информант» ([www.informnn.ru](http://www.informnn.ru))
12. Технологии разведки для бизнеса «IT2B» ([www.it2b.ru](http://www.it2b.ru))
13. Альт-маркетинг: библиотека материалов по конкурентной разведке ([alt-marketing.ru/articles/index-competitiveintelligence.shtml](http://alt-marketing.ru/articles/index-competitiveintelligence.shtml))
14. Информационная корпоративная служба ([z-filez.info](http://z-filez.info))
15. Конкурентная разведка, сайт Е.Л. Ющука ([ci-razvedka.ru](http://ci-razvedka.ru))
16. Конкурентная разведка в Интернете. Авторский курс А. Масаловича (<http://www.tora-centre.ru/razvedka.htm>)

## Адреса упоминаемых веб-ресурсов

---

---

***www.about.com*** – About.com – портал, охватывающий ссылки на доступные веб-ресурсы и ресурсы глубинного веб.

***agregator.pro*** – Agregator.pro – агрегатор информации с новостных и медийных порталов.

***www.aignes.com*** – WebSite-Watcher – программа мониторинга веб-сайтов, форумов, локальных файлов.

***www.anbr.ru/products/semarchive*** – информационно-аналитическая система «Семантический архив».

***www.archive.org*** – интернет-архив (Internet Archive).

***www.attensity.com*** – Attensity suite – технология выявления понятий, фактографической информации из неструктурированных текстов.

***www.autonomy.com*** – Autonomy IDOL Server компании Autonomy.

***www.babkee.ru*** – Babkee – система мониторинга упоминаний в социальных медиа.

***badoo.com*** – Badoo – социальная сеть знакомств.

***www.base.ukrpatent.org/searchINV*** – интерактивная БД «Изобретения (полезные модели) в Украине».

***baza.kompass.ua*** – «Компасс Украина» – справочник компаний Украины

***blogs.yandex.ru*** – Яндекс.Блоги.

***books.google.com*** – Google Book Search – поиск книг.

***www.brandspotter.ru*** – Brandspotter – сервис, предлагающий определение эмоциональной окраски высказываний, статистика по темам, платформам, авторам, сравнение с конкурентами.

***www.businessobjects.com/product/catalog/text\_analysis/features.asp*** – Businessobjects Text Analysis – программа, позволяющая извлекать информацию о бизнес-объектах.

***buzzlook.ru*** – BuzzLook – русскоязычный сервис мониторинга социальных медиа.

***www.buzzware.ru*** – Buzzware – инструмент мониторинга социальных медиа, позволяющий исследовать мнения пользователей.

***www.ciradar.com/Competitive-Analysis.aspx*** – CIRadar – систем поиска информации для конкурентной разведки в «глубинном Интернете».

***cluuz.com*** – Cluuz – поисковая система для применения в конкурентной разведке, позволяет находить и визуализировать связи между объектами.



**[code.google.com](http://code.google.com)** – Google Code Search – поиск программного кода.

**[www.copernic.com/en/products/agent](http://www.copernic.com/en/products/agent)** – Copernic Agent – метапоисковая программа.

**[www.creditreform-rus.ru](http://www.creditreform-rus.ru)** – международная корпорация Creditreform, обеспечивающая доступ к международной сети содействия информационному бизнесу BIGNet.

**[www.digitalpreservation.gov](http://www.digitalpreservation.gov)** – американский национальный проект сохранения и распространения цифрового контента Digital Preservation.

**[www.diphur.com](http://www.diphur.com)** – Diphur Monitor EveryThing – сервис мониторинга веб-сайтов, уведомляющий об их обновлениях.

**[docs.oracle.com/cd/E29220\\_01](http://docs.oracle.com/cd/E29220_01)** – Endeca Latitude – технология, реализующая концепцию Text Mining.

**[www.dtsearch.com](http://www.dtsearch.com)** – dtSearch – поисковая программа, позволяющая обрабатывать статические и динамические данные во всех форматах MS Office.

**[www.eapo.org/ru](http://www.eapo.org/ru)** – Евразийская патентная база.

**[www.europages.eu](http://www.europages.eu)** – Europages – Европейская бизнес-директория.

**[www.facebook.com](http://www.facebook.com)** – Facebook – крупнейшая социальная сеть.

**[global.factiva.com](http://global.factiva.com)** – служба Factiva, предоставляющая доступ к деловой и аналитической информации.

**[file.liga.net/person](http://file.liga.net/person)** – ЛИГА.Досье – биографии политиков, звезд и известных людей.

**[www.findlaw.com](http://www.findlaw.com)** – FindLaw – каталог, содержащий список свободно доступных баз данных нормативно-правовых документов.

**[go.mail.ru/realtime](http://go.mail.ru/realtime)** – Поиск mail.ru – поиск в mail.ru.

**[www.google.com](http://www.google.com)** – глобальная информационно-поисковая система Google.

**[google.ru/alerts](http://google.ru/alerts)** – Google Оповещения.

**[hootsuite.com](http://hootsuite.com)** – Hootsuite – многофункциональный сервис для работы с социальными медиа.

**[www.howsociable.com](http://www.howsociable.com)** – HowsSciable – инструмент мониторинга брендов и ключевых слов с помощью 32 социальных сетей.

**[www.hummingbirdenterprise.com](http://www.hummingbirdenterprise.com)** – система Hummingbird Enterprise.

**[www-01.ibm.com/software/data/information-optimization](http://www-01.ibm.com/software/data/information-optimization)** – Vivisimo – технология объединенного поиска информации во многих источниках и организации доступа к информационным ресурсам.

**[www-01.ibm.com/software/industry/i2software/support.html](http://www-01.ibm.com/software/industry/i2software/support.html)** – i2 Analyst Notebook – система визуального проектирования структуры для хранения данных о персонах и организациях, событиях и взаимосвязях.

**[infomine.ucr.edu](http://infomine.ucr.edu)** – Infomine – сервис добычи информации из баз данных, блогов, электронных каталогов и т.п.

**[www.infongen.com](http://www.infongen.com)** – InfoNgen – агрегатор информации, настраиваемый на уникальные темы.

**[www.infostream.ua](http://www.infostream.ua)** – InfoStream – система контент-мониторинга веб-ресурсов.

**[www.infovell.com](http://www.infovell.com)** – поисковая система Infovell, позволяющая искать в глубинном веб.

**[www.insightprofiles.com](http://www.insightprofiles.com)** – сервис Insight Profiles, предоставляющий характеристики и сведения о способностях и компетентности граждан США.

**[www.integrum.ru](http://www.integrum.ru)** – крупнейшая архивная база данных СМИ службы «Интегрум».

**[www.internetsec.com](http://www.internetsec.com)** – служба Internet Securities, поставляющая бизнес-информацию от 16 тыс. источников.

**[inventionmachine.com](http://inventionmachine.com)** – система Goldfire Research – система обработки контента глубинного веб.

**[www.iqbuzz.ru](http://www.iqbuzz.ru)** – IQBuzz – сервис для мониторинга социальных медиа с возможностью подключения по запросам пользователей новых источников.

**[kare.pulscen.com.ua](http://kare.pulscen.com.ua)** – «КАРЕ» – база данных предприятий Украины.

**[www.kodeks.ru](http://www.kodeks.ru)** – информационно-поисковая система по российскому законодательству.

**[www.kribrum.ru](http://www.kribrum.ru)** – Крибрум – технология, позволяющая отслеживать и анализировать упоминания брендов, продуктов, услуг и т.п.

**[www.kurrently.com](http://www.kurrently.com)** – Kurrently – поисковая система в режиме реального времени по Twitter и Facebook.

**[www.labyrinth.ru](http://www.labyrinth.ru)** – российская база данных «Лабиринт», составленная на основе публикаций ведущих бизнес-изданий.

**[www.lexisnexis.com](http://www.lexisnexis.com)** – крупнейшая в мире полнотекстовая онлайн-информационная система LexisNexis.

**[www.linkedin.com](http://www.linkedin.com)** – LinkedIn – социальная сеть для поиска и установления деловых контактов.

**[www.livejournal.com](http://www.livejournal.com)** – «Живой Журнал», ЖЖ, LiveJournal, LJ – платформа для ведения онлайн-дневников (блогов).

**[www.loc.gov](http://www.loc.gov)** – библиотека Конгресса США.

**[www.megaputer.ru](http://www.megaputer.ru)** – PolyAnalyst – семейство продуктов для глубокого анализа данных.

**[www.mlg.ru](http://www.mlg.ru)** – «Медialogия» – сервис, обеспечивающий онлайн-доступ к базе СМИ с возможностью производить самостоятельный мониторинг и экспресс-анализ СМИ.

**[www.moikrug.ru](http://www.moikrug.ru)** – «Мой Круг» – русскоязычная социальная сеть, направленная на установление деловых контактов между людьми.

**[monitor.wildfireapp.com](http://monitor.wildfireapp.com)** – WildFire – онлайн-сервис медиа маркетинга в социальных сетях.

**[multitender.ru/tenders](http://multitender.ru/tenders)** – Мультитендер – поисковая система по госзакупкам, тендерам и аукционам в РФ.

**[neiron.ru/toolbar](http://neiron.ru/toolbar)** – Neiron Search Tools – программная надстройка, объединяющая результаты информационно-поисковых систем Google и Яндекс.

**[www.newprosoft.com](http://www.newprosoft.com)** – Newprosoft Web Content Extractor – программа сканирования и извлечения данных из веб-сайтов.

**[newspapermap.com](http://newspapermap.com)** – Newspaper Map – сервис, объединяющий геолокацию и информационно-поисковую систему по медиа-ресурсам.

**[news.yandex.ua](http://news.yandex.ua)** – Яндекс.Новости – служба автоматической обработки и систематизации новостей компании Яндекс.

**[openua.net](http://openua.net)** – Открытая Украина: биографии выдающихся граждан Украины.

**[www.peerindex.net](http://www.peerindex.net)** – PeerIndex – сервис анализа социальных медиа, определяет размеры влияния компаний.

**[peoples.ru](http://peoples.ru)** – «Люди» – биографии различных людей.

**[www.politicalinformation.com](http://www.politicalinformation.com)** – Politicalinformation.com – сервис поиска в 5000 отобранных веб-сайтах политической направленности.

**[www.portableofflinebrowser.com](http://www.portableofflinebrowser.com)** – Portable Offline Browser – программа, позволяющая скачивать необходимые веб-сайты и мультимедиа-информацию.

**[www.postrank.com](http://www.postrank.com)** – PostRank – система глобального социального медиа-анализа.

**[public.ru](http://public.ru)** – Public.Ru – интернет-библиотека СМИ.

**[www.quintura.ru](http://www.quintura.ru)** – Quintura – визуальная поисковая система.

**[www.quixey.com](http://www.quixey.com)** – Quixey – поисковая система, обеспечивающая вывод ссылок на программы, сервисы и расширения для браузеров.

**[www.r-techno.com/rtechno/online-services/webinsight](http://www.r-techno.com/rtechno/online-services/webinsight)**

– система «it2b.интернетошпионаж 3000+», предназначенная для выгрузки данных из невидимого сегмента сети Интернет.

**[www.rco.ru](http://www.rco.ru)** – RCO – система выявления фактографической информации из неструктурированных текстов.

**[www.rco.ru/product.asp?ob\\_no=7348](http://www.rco.ru/product.asp?ob_no=7348)** – RCO Zoom – поисково-аналитическая система с функциями контент-анализа в реальном времени и транзакционного хранилища документов.

**[www.rocketsoftware.com/products/rocket-aerotext](http://www.rocketsoftware.com/products/rocket-aerotext)** – Rocket AeroText – программа, позволяющая выявлять сущности, их взаимные связи и события, в неструктурированных текстах.

**[www.ru-patent.info](http://www.ru-patent.info)** – база патентов на изобретения РФ.

**[www.rosspending.ru](http://www.rosspending.ru)** – РосГосЗатраты – мониторинг российских государственных контрактов.

**[www.sas.com/text-analytics/text-miner](http://www.sas.com/text-analytics/text-miner)** – Statistica Text Miner – технология перевода неструктурированных текстов в информацию, пригодную для принятия решений.

**[scholar.google.com](http://scholar.google.com)** – Google Scholar – поиск научных публикаций.

**[seismic.com](http://seismic.com)** – Seismic – сервис мониторинга социальных медиа.

**[www.semanticforce.net](http://www.semanticforce.net)** – SemanticForce – сервис мониторинга неструктурированных источников информации.

**[www.socialbakers.com](http://www.socialbakers.com)** – Socialbakers – сервис, мониторинга и сбора статистики о работе социальных сетей.

**[www.socialmention.com](http://www.socialmention.com)** – Socialmention – платформа для поиска и анализа информации в социальных сетях.

**[www.socialpointer.com](http://www.socialpointer.com)** – Socialpointer – сервис мониторинга в социальных сетях, новостях, блогах.

**[socialseek.com](http://socialseek.com)** – SocialSeek – инструмент мониторинга множества социальных ресурсов в режиме реального времени.

**[www.softpedia.com/get/Internet/Search-engine-tools-submitting/Website-Finder.shtml](http://www.softpedia.com/get/Internet/Search-engine-tools-submitting/Website-Finder.shtml)** – Website-Finder, программа, поиска веб-сайтов, которые плохо индексируются системой Google.

**[www.strategator.com](http://www.strategator.com)** – Strategator – программа, осуществляющая поиск, фильтрацию и агрегацию информации о компаниях.

**[topsy.com](http://topsy.com)** – Topsy – сервис поиска в режиме реального времени по социальным медиа.

**[www.tora-centre.ru/avl3.htm](http://www.tora-centre.ru/avl3.htm)** – Avalanche – система интернет-мониторинга и конкурентной разведки.

**www.trackur.com** – Trackur – инструмент для мониторинга и анализа социальных медиа. Позволяет отслеживать, например, репутацию брендов.

**tweetdeck.com** – TweetDeck – инструмент для управления и отслеживания постов в Twitter, Facebook, MySpace, LinkedIn.

**www.twitalyzer.com** – Twitalyzer – аналитическая программа-клиент для Твиттера.

**twitter.com** – Твиттер – крупнейший сервис микроблогов.

**www.verticalsearchworks.com** – Система RetrievalWare – система поиска, анализа и выявления информации с помощью задания пользователем поисковых запросов на естественном языке.

**visual.ly** – Visual.ly – поисковик инфографики в веб-пространстве.

**vk.com** – «ВКонтакте» – крупнейшая в Рунете социальная сеть, позиционирующая себя как «современный, быстрый и эстетичный способ общения в сети».

**watchthatpage.com** – WatchThatPage – сервис, позволяющий автоматически собирать новую информацию с веб-ресурсов.

**weblib.in.ua** – Weblib – мультипоисковая система по документам, представленным в формате PDF.

**webground.su** – Webground – интегратор русскоязычных новостей.

**www.webscan-global.com** – Webscan Technologies – услуги системного мониторинга в Интернете.

**wobot.ru** – Wobot – сервис, позволяющий проследить ретроспективу мнений в социальных сетях.

**www.worldindustrialreporter.com/solusource** – Global Supplier Directory by Solusource – веб-интерфейс для конкурентной разведки от компании Thomas.

**www.yellowpages.kiev.ua** – «Желтые страницы» Киева.

**www.yahoo.com** – глобальная информационно-поисковая система Yahoo!

**www.yandex.ru** – глобальная информационно-поисковая система Яндекс.

**www.youscan.ru** – YouScan – система профессионального мониторинга русскоязычных социальных медиа.

**ypag.ru** – «Желтые страницы» СНГ.

**zakon.rada.gov.ua** – информационно-поисковая система по украинскому законодательству.

**www.zanran.com/q** – Zanran – поисковая система для PDF-файлов, таблиц EXCEL, данных на страницах HTML.

Наукове видання

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ  
ІНСТИТУТ ПРОБЛЕМ РЕЄСТРАЦІЇ ІНФОРМАЦІЇ

ДОДОНОВ Олександр Георгійович  
ЛАНДЕ Дмитро Володимирович  
ПРИЩЕПА Віктор Володимирович  
ПУТЯТІН Володимир Григорович

## **КОНКУРЕНТНА РОЗВІДКА У КОМП'ЮТЕРНИХ МЕРЕЖАХ**

Монографія

(Російською мовою)

Київ, Видавництво «ЛАТ & К», 2013

Комп'ютерна верстка Рассоленко М.Д.

Підп. до друку 00.00.00. Формат ....х..../.....  
Папір офс. № 1. Гарн. 00000. Ум. друк. арк. 00,00.  
Обл.-вид. арк. 000. Замовлення 0000  
Наклад 300 прим. Замовлення №

Видавництво «ЛАТ & К»  
Свідоцтво про внесення до Державного реєстру  
Серія ДК № 181 від 15.09.2000 р.,  
тел.: +38 044 235 75 28, факс: +38 044 235 000 9  
e-mail: lk@ukr.net