Дудихин В. В., Дудихина О. В.

Конкурентная разведка в Internet

Советы аналитика

ДМК Пресс
издательство
Москва

УДК 004.738.52
ББК 32.973.202
Д81

Дудихин В. В., Дудихина О. В.
Д81 Конкурентная разведка в Internet. Советы аналитика. – М.: ДМК Пресс. – 192 с.: ил.

ISBN 5-94074-178-9

В книге рассмотрены методы получения информации из Internet для ведения конкурентной разведки – достаточно новой отрасли информационного бизнеса. В очень доступной форме описаны некоторые приемы аналитической работы, способы нахождения и использования текстовой информации и примеры их применения для решения практических задач.

Издание предназначено для широкого круга читателей.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельца авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но, поскольку вероятность технических ошибок все равно остается, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможный ущерб любого вида, связанный с применением содержащихся здесь сведений.

Все торговые знаки, упомянутые в настоящем издании, зарегистрированы. Случайное неправильное использование или пропуск торгового знака или названия его законного владельца не должно рассматриваться как нарушение прав собственности.

ISBN 5-94074-178-9

© Дудихин В. В., Дудихина О. В.
© ДМК Пресс

Содержание

От авторов 5

Введение 8

Глава 1. Классификация источников информации 14

1.1. Информационные агентства 16
1.2. Средства массовой информации 24
1.3. Специализированные базы данных 39
1.3.1. Информационно-поисковая система «Артефакт» 41
1.3.2. Информационно-аналитическая система «Галактика-Zoom» 46
1.3.3. Информационная система «Россия» 55
1.3.4. Другие информационные системы 61
1.4. Поисковые Internet-системы, Internet-каталоги и метапоисковые системы 64
1.4.1. Классификация поисковых средств 64
1.4.2. Яндекс 73
1.4.3. Апорт 75
1.4.4. Rambler 78
1.5. Ординарные Web-сайты 84

Глава 2. Методы и приемы получения информации 89

2.1. Общие положения 89
2.2. Критерии выбора баз данных и поисковых систем 90
2.3. Типовой алгоритм поиска 91

2.4. Основные понятия языков запросов ... 92
2.4.1. Операнды ... 92
2.4.2. Скобки ... 94
2.4.3. Операторы ... 94
2.4.4. Запрос по полям (зонам ... 97

Глава 3. Аналитическая обработка информации ... 99

3.1. Оценка информации ... 99
3.2. Методы и приемы аналитической обработки информации ... 102
3.3. Основные типовые задачи ... 104
3.3.1. Поиск данных о личности ... 105
3.3.2. Поиск информации о фирме ... 120
3.3.3. Поиск информации о политической или общественной организации ... 123
3.3.4. Аналитические справки о состоянии отраслей промышленности ... 140
3.4. Идеи Шерлока Холмса – в жизнь ... 159
3.5. Контент-анализ текстов ... 161
3.5.1. Простейший контент-анализ ... 162
3.5.2. Выполнение контент-анализа с помощью системы «Артефакт» ... 164
3.5.3. Проведение контент-анализа с помощью системы «Галактика-Zoom» ... 169

Заключение ... 181

Литература ... 182

От авторов

Omne ignotum pro magnifico est.
Tacit. «Agricola»[1]

Наше пособие адресовано самому широкому кругу читателей и посвящено тому, как лучше и с наименьшими затратами получить информацию с помощью Internet. Надеемся, что оно будет интересно сотрудникам самых разных фирм, занимающихся сбором информации, аналитикам, предлагающим различные стратегии построения бизнеса, и обычным гражданам, желающим узнать как можно больше о нашем прекрасном мире.

Руководителю любого уровня для принятия правильных управленческих решений необходимы логика, интуиция и информация. Допустим, логике можно научиться. Интуиция – чутье от Бога, она или есть или ее нет, а вот информация необходима всем и всегда.

После осознания сказанного возникает вопрос: откуда эффективнее всего можно черпать информацию в настоящее время? Далее выясняется, что информация – продукт скоропортящийся, дорогостоящий, и ее получением и обработкой должен постоянно заниматься квалифицированный персонал. Например, базы данных становятся действительно полезными для решения реальных задач только после того, как их объем перерастает некую критическую границу. Чтобы ее достичь, нужно потратить очень много сил и средств. Это могут позволить себе даже не все крупные компании. Где же выход? До недавнего времени его не существовало. Однако менее десяти лет назад

[1] Все неизвестное представляется величественным (Тацит, «Agricola»).

появился Internet – уникальное явление в человеческой жизни. В его необъятных просторах есть, наверное, вся или почти вся информация. Однако найти ее там далеко не просто.

Эта книга посвящена описанию источников информации в Internet, а также методам и приемам ее поиска и обработки. В настоящее время уже написаны сотни, а может быть, тысячи подобных книг о том, как искать сайты в Internet – неком самостоятельном и самодостаточном мире. Ведь не секрет, что сегодняшний Internet многим разработчикам высоких технологий напоминает «башню из слоновой кости». Мир, существующий там, скорее, плод фантазии и результат самовыражения молодых и талантливых людей. Очень большая (если не бо^льшая) часть его ресурсов имеет весьма опосредованное отношение к реальной жизни. Это похоже на сладкий миф, идеальное или, наоборот, критическое, негативное представление окружающей нас действительности. Одним словом – виртуальная реальность в полной ее красе. Однако и в Internet имеется сегмент, содержащий полезную для практической жизни информацию. Методам ее нахождения и эффективного использования посвящена наша книга.

Преимущество такого подхода особенно ощутимо для представителей малого и среднего бизнеса. Они получают доступ к колоссальнейшим информационным хранилищам либо почти бесплатно (так как нет необходимости платить за хранение и обновление информации, требуется лишь весьма умеренная сумма за пользование Internet), либо оплатив удаленный доступ к онлайновым базам данных (что на несколько порядков дешевле содержания своего собственного информационного центра).

Рассказывая о *конкурентной разведке* как о совокупности приемов сбора и обработки информации законными и этичными методами, книга в основном посвящена той ее части, которая может быть проведена с помощью Internet. В книге приводится классификация возможных источников информации в Internet:

- информационных агентств;
- средств массовой информации;
- специализированных онлайновых баз данных;
- поисковых систем и каталогов.

Так как в одном небольшом издании невозможно описать все информационные агентства и средства массовой информации, то рассматривались наиболее типичные источники этого класса. Подобным же образом были продемонстрированы принципы применения

специализированных баз данных на примере трех характерных представителей программных продуктов.

Целая глава книги посвящена методам и приемам получения информации. Сформулированы общие положения, критерии и типовые алгоритмы. Перечислены основные понятия и некоторые специфические особенности языков запросов различных поисковых систем.

Особенности функционирования поисковых Internet-систем и каталогов рассматривались на примере трех наиболее известных в настоящее время русскоязычных порталов – Яндекс, Апорт и Rambler.

И наконец, в заключительной главе очень кратко говорится об оценке информации на предмет достоверности и некоторых методах и приемах ее аналитической обработки, в том числе о типовых задачах, основанных на контент-анализе. Упор сделан на самые простейшие подходы, базирующиеся на здравом смысле и не требующие специальных знаний и квалификации. В этой части книги приведен ряд типичных документов, востребованных в современном бизнесе, политических и PR-технологиях.

Авторы видели свою цель в том, чтобы дать некоторые ориентиры не только руководителям разного уровня при возможной организации информационно-аналитического подразделения в фирме, но и любому деловому человеку, интересующемуся данной сферой деятельности. Насколько это удалось – судить читателям. Естественно, как и любое творение рук человеческих, эта книга неидеальна. Что-то в ней получилось лучше, что-то еще лучше. Надеемся, что читатель простит нам оплошности и воспримет написанное в преломлении своего жизненного опыта, который и является главным компасом в жизни конкретного человека.

В книге собраны и систематизированы сведения из самых различных источников, аккумулировавшиеся длительное время и успешно опробованные на практике. Мы считали важным отразить в книге именно те подходы, методы и приемы, которые не требуют больших специальных знаний, сложной и дорогостоящей аппаратуры и в той или иной степени доступны любому человеку, имеющему выход в Internet.

В настоящем издании, к сожалению, пришлось ограничиться рассмотрением только русскоязычного Internet. Надеемся, что о других доменах Всемирной паутины удастся написать в ближайшее время.

Выражаем искреннюю признательность всем тем, без чьей помощи эта книга, возможно, не увидела бы свет.

Введение

Сегодня не только в бизнесе, но и в частной жизни громадное значение приобретает анализ возможных ситуаций, выяснение их тенденций, вычисление возможных рисков при принятии решений. Недостаточно обладать информацией в виде маркетинговых исследований или бизнес-справок о фирмах потенциальных партнеров. Кто или что является причиной возникновения событий, откуда исходит или может исходить угроза, как ее избежать или хотя бы минимизировать? Важность решения этих задач стала причиной возникновения так называемой конкурентной разведки – мощного инструмента информационно-аналитической поддержки принятия решений.

Считается, что *конкурентная разведка* (competitive intelligence) – это сбор и обработка информации легальными и этичными способами, при выполнении которых явления и тенденции рассматриваются через призму конкуренции и конкурентной борьбы.

В России для конкурентной разведки пока нет устоявшегося термина, поэтому иногда встречаются выражения типа «деловая разведка», «бизнес-разведка» и т.п. Как правило, при ведении конкурентной разведки опираются на следующие источники информации:

- опубликованные документы открытого доступа (книги, журналы, газеты, фирменные и рекламные издания и т.п.) и электронные источники, причем особый интерес представляют онлайновые базы данных;
- сведения из неопубликованных (конфиденциальных) документов, уже имеющихся на фирме, ведущей конкурентную разведку, и данные, которыми располагают ее сотрудники (например,

результаты маркетинговых исследований; информация, полученная на выставках или конференциях в процессе общении с реальными или потенциальными клиентами, и т.д.);

- сведения, поступающие от внешних «агентов», например лиц, работающих в научных центрах, в других фирмах, в органах управления и т.д., а также то, что получено с помощью иных методов и средств. Нужно сказать, что граница между этичным и неэтичным использованием этих источников достаточно эфемерна.

Еще раз повторимся: в этой книге рассматривается та часть конкурентной разведки – сбор информации и аналитическая работа, которую можно выполнить с помощью компьютерных технологий и Internet.

Важно отметить, что конкурентная разведка – это не пресловутый промышленный шпионаж. Действует она исключительно в рамках существующих законов. Использование новых информационных технологий, абсолютно открытых источников и относительно дешевых информационных ресурсов позволяет принимать правильные решения, а тот факт, что так поступают все спецслужбы мира, лишь подтверждает перспективность этого направления деятельности.

Информационно-аналитическая деятельность консалтинговых компаний и подразделений конкурентной разведки весьма наукоемка. Многие крупные корпорации и фирмы давно обзавелись собственными телекоммуникационными сетями и базами данных, используют самые последние достижения научно-технического прогресса, вплоть до систем искусственного интеллекта.

По информационным сетям крупнейших компаний ежедневно передаются десятки тысяч сообщений с разных уголков мира. Информационно-аналитические системы консалтинговых и иных компаний накапливают информацию и анализируют поведение ведущих европейских государственных деятелей и бизнесменов, обеспечивают доступ к данным о конкурентах для многих тысяч своих сотрудников.

При ведении конкурентной разведки особую популярность получило новое направление науки, возникшее на стыке искусственного интеллекта, статистики и теории баз данных, – knowledge discovery in databases (обнаружение знаний в базах данных, или концепция Data Mining либо Text Mining, если речь идет о текстовых базах данных, которая обычно переводится как «добыча» или «раскопка данных»). В основе этой технологии – концепция шаблонов (паттернов), отражающих некие фрагменты взаимоотношений в данных.

Особенность Data Mining в том, что найденные шаблоны должны отражать неочевидные, неожиданные закономерности, так называемые скрытые знания. В целом технологию Data Mining обычно определяют как процесс обнаружения в «сырых» данных ранее неизвестных, нетривиальных, полезных на практике и доступных для интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

На основе этих технологий разработаны программные продукты, пригодные для решения задач конкурентной разведки, например система «Галактика-Zoom». Так как наша цель – показать варианты эффективного использования доступных средств, то мы ограничимся лишь простейшими приемами и возможностями российского Internet. Конечно, иные источники информации и высокие технологии не менее интересны, но столь необъятны, что для пользы дела на первом этапе надо остановиться на чем-то конкретном.

Жизнь ежедневно ставит перед современным человеком массу вопросов. Если он живет в России, то вопросов как минимум в два раза больше. На многие из них нет прямых ответов. Но они могут быть получены путем анализа, сопоставления, обобщения информации, собранной из различных источников.

В большинстве стран конкурентная разведка – развитая отрасль деятельности, приносящая солидные прибыли. Отслеживание конъюнктуры рынка, новых продуктов или услуг, стратегии и планов реальных и потенциальных конкурентов – непременный атрибут работы уважающей себя фирмы. Имеются сведения, что из 500 крупнейших компаний США 90% занимаются ведением конкурентной разведки. Сбор информации и наблюдение за конкурентами позволяет быть в курсе их планов, находить принципиально новые подходы ведения бизнеса, получать существенные преимущества и даже захватить лидерство в отрасли.

До возникновения Internet люди тратили массу времени, чтобы «откопать» требуемые сведения из газет, журналов, справочников. Благодаря новым информационным технологиям ситуация значительно упростилась. Если в прошлом веке появление автоматического стрелкового оружия во многом способствовало краху колониальной системы, поскольку автомат Калашникова в руках папуаса