Структурированный мониторинг открытых персональных данных в сети интернет

*Дорофеев А. В., Марков А. С.**

Аннотация. *Статья посвящена вопросам конкурентной разведки и мониторингу информационной безопасности персонала организации. В целях оптимизации поиска информации о конкретном человеке в сети Интернет предложено использовать структурированный метод сбора и анализа информации, основанный на декомпозиции данных и связей между данными, касающимися конкретного человека. Представлена классификация открытых источников информации в глобальной сети Интернет. Кратко рассмотрены Интернет-ресурсы, косвенно или напрямую касающиеся информации о конкретных людях. Рассмотрена циклическая модель сбора и анализа информации, позволяющая уточнять данные и связи между данными. Разработана концептуальная модель, предоставляющая возможность эксперту выполнять обоснованный последовательный поиск информации. Подробно рассмотрены основные процессы сбора и анализа информации, а также используемые методические приемы и алгоритмы структурированного анализа. Отдельное внимание уделено методу формирования и проверке гипотез на различных этапах сбора и анализа информации. Приведены примеры из реальной жизни, подтверждающие результативность предложенного методического подхода к анализу информации о конкретном человеке в Интернете, в том числе социальных сетях.*

Ключевые слова: *мониторинг безопасности, безопасность персонала, структурированный анализ, разведывательный цикл, социальные сети, Интернет-ресурсы, деловая разведка, бизнес-разведка, досье.*

Введение

Развитие современных Интернет-сервисов привело к тому, что глобальная сеть Интернет стала кладезем информации о людях. Наличие информации в открытых источниках позволяет многое узнать о человеке еще до возникновения каких-либо деловых и личных отношений. Однако разрозненность и динамичность данных в сети Интернет, неопределенность и неустойчивость связей между объектами и событиями усложняют получение достоверного целевого результата о конкретном человеке. Применение общесистемного подхода в некоторых случаях связано с временными издержками по причине необходимости обработки чрезвычайно большого количества информации [1]. Для исключения указанных недостатков в работе предложен структурированный поэтапный анализ информации, ориентированный на целенаправленное получение достоверных знаний о конкретном человеке.

1. Понятие структурированного анализа

Под структурированным анализом (structured analysis) мы будем понимать экспертный целенаправленный пошаговый подход к сбору и обработке данных и связей между данными, позволяющий получить максимально полную и достоверную информацию о конкретном человеке. В некотором смысле, в рамках структурированного анализа выполняется преобразование неясных знаний о данных и связях данных в точные, путем декомпозиции (структуризации) данных и связей между ними.

Традиционно анализ информации о конкретном человеке включает начальный этап (определение источников, сбор и оценка априорных данных) и последующие итерационные шаги по обработке (анализ, формулировка выводов, формирование гипотез) данных в целях получения максимально объективной и полной информации.

* ***Дорофеев Александр Владимирович***, директор по развитию ЗАО «НПО «Эшелон», Российская Федерация, г. Москва.
E-mail: mail@uc-echelon.ru
Марков Алексей Сергеевич, доктор технических наук, старший научный сотрудник, профессор МГТУ им. Н.Э. Баумана, Российская Федерация, г. Москва.
E-mail: a.markov@bmstu.ru

Обеспечение полноты и непротиворечивости анализа обеспечивается соответствующей классификацией открытых источников данных.

2. Классификация источников информации

При классификации открытых источников обычно задаются вопросами:

- кто разместил данные: сам пользователь, аффилированная или третья сторона;
- данные непосредственно связаны с человеком или косвенно;
- какая степень доверия к источнику?

В последнем случае следует выделить несколько наиболее значимых источников для сбора информации, а именно:

- социальные сети (например: vk.com и linkedin.com);
- личные блоги и сайты (в том числе: wordpress.com и blogger.com);
- микро-блог (twitter.com);
- сайты знакомств (loveplanet.ru, mamba.ru и т.д.);
- сайты для размещения контента (youtube.com, slideshare.net, instagram.com);
- форумы по интересам (например, cyberforum.ru);
- программы обмена мгновенными сообщениями или «мессенджеры» (whatsapp, viber).

Что касается государственных служащих, то следует добавить наборы открытых государственных данных (open data [2]), размещаемых в информационных системах общего пользования.

С учетом целевой задачи и ее контекста первичную информацию о человеке можно условно представить кортежем:

$$D = \langle Gn, Rl, Cn, Ph, Jb, Ed, Ml, Zn, Hb, Sp, Ow, Ot \rangle,$$

который включает множества общих идентификационных данных Gn, данных о семейных отношениях Rl, контактной информации Cn, физических данных *Ph*, данных о работе *Jb*, образовании *Ed*, отношении к военной службе *Ml*, приятельском окружении *Zn*, хобби *Hb*, спорте *Sp*, частной собственности *Ow*, специальной информации *Ot*, особенно проблемной. Определение подмножеств и их допустимых значений легко детализировать, например, для случая программной реализации [3].

Приведем сводную таблицу, позволяющую определить какая именно информация о конкретном человеке на каких Интернет-ресурсах может быть размещена (табл.1.)

Далее важно проверить косвенные признаки с целью получения новых данных (табл. 2).

Помимо информации, публикуемой самим человеком, в сети Интернет могут быть размещены данные о нем помимо его воли. Это могут быть данные, раскрываемые в соответствии с требованиями законодательства, либо информация, публикуемая знакомыми, друзьями или недругами человека. Так, во многих социальных сетях можно разместить фотографию и указать на ней других пользователей данной сети.

В табл. 3 приведены примеры источников информации, публикуемые государством.

Также надо понимать, что для того, чтобы сформировать нужную информацию, необходимо уже что-то предполагать (или знать) и понимать взаимосвязи имеющихся данных [4]. Для этого можно построить соответствующую ассоциативную карту (mind maps), т.к. связи на данном этапе анализа еще строго не конкретизированы [5]. Пример ассоциативной карты приведен ниже (рис.5).

Отметим, что в настоящей статье авторы не рассматривают случаи раскрытия персональных данных (личной и семейной тайн, тайны частной жизни и других) в нарушение законодательства [6].

2. Разведывательный цикл

Полученная выше информация, по сути, представляет для нас первичную карту знаний, а структурированные техники анализа создают маршрут движения по ней. Известны насколько структурных моделей сбора и анализа информации о человеке (в частности, разведывательные циклы ЦРУ и ФБР [7, 8]), которые включают 4-6 этапов. На наш взгляд, удобно воспользоваться предлагаемой ниже 7-ми процессной циклической моделью (рис.1).

Предлагаемая модель включает следующие последовательные временные процессы (этапы):

Рис. 1. Циклическая семипроцессная модель поиска информации

Таблица 1.

Сводная информация о человеке

Первичные данные	Интернет-сегменты						
	Соц. Сети	Блоги	Микро-блог	Сайты зна-комств	Контент-сайты	Форумы	«Мес-сендже-ры»
Общие сведения							
ФИО	+			+			
Ник (идентификатор в сети)	+	+	+	+	+	+	
Дата рождения	+						
Место рождения	+						
Место жительства	+		+	+			
Место нахождения			+	+			
Семья/Отношения							
Супруг/а	+			+			
Родители	+						
Дети	+			+	+		
Контактные данные							
Телефон (моб/раб/раб)	+					+	+
E-mail (раб/лич)	+					+	
Физическое состояние							
Фотографии	+	+		+	+		+
Рост	+			+			
Вес	+			+			
Состояние здоровья	+			+	+		
Психотип	+	+		+			
Работа							
Название организации	+	+			+		
Адрес организации	+						
Должность	+						
Отношение к организации		+					
Предыдущие места работы (резюме)	+						
Образование							
Школа	+						
Университет	+						
Награды	+						
Послевузовское образование	+						
Сертификаты	+						

Первичные данные	Интернет-сегменты						
	Соц. Сети	**Блоги**	**Микро-блог**	**Сайты зна-комств**	**Контент-сайты**	**Форумы**	**«Мес-сендже-ры»**
Курсы	+						
Научная деятельность		+			+	+	
Служба в армии							
Где	+						
Когда	+						
Знакомства							
Одноклассники/коллеги	+						
Хобби							
Вид	+				+		
Круг знакомств	+				+		
Спорт							
Вид	+				+		
Круг знакомств	+				+		
Адреса	+						
Собственность							
Недвижимость	+				+		
Автомобиль	+				+		
Прочее							
Гражданская (политическая, социальная, философская, религиозная) позиция	+	+	+			+	
Личные проблемы	+	+	+			+	+

Таблица 2.

Пример использования косвенных признаков

Вид данных	Косвенные признаки
Ник	Ссылка на страницу в социальной сети или на сайте знакомств, как правило, содержит ник человека
Место работы	Пользователем могут быть размещены фотографии, сделанные в самом офисе, рядом со зданием офиса, на выставке, в которой участвовала компания, и по элементам, попавшим в кадр, возможно, можно будет определить организацию
Отношения	Невербальные сигналы на фотографиях
Место нахождения	Фотографии, временные метки и геометки
Круг знакомств	Подписчики/списки друзей
Текущие проблемы	Статусы в мессенджерах/социальных сетях
Психотип	Записи в блоге, фотографии, медиа-контент, комментарии

Таблица 3.

Пример государственных информационных ресурсов

№	Вид информации	Интернет-источники
1	Наличие долгов	Банк данных исполнительных производств http://fssprus.ru/iss/ip/
2	Наличие судебного производства	Сайт суда по месту прописки
3	Участие в арбитражном судебном производстве	http://kad.arbitr.ru/
4	Нахождение в розыске по подозрению в совершении преступления	http://fssprus.ru/iss/suspect_info
5	Нахождение в розыске по исполнительному производству	http://fssprus.ru/iss/ip_search
6	Сведения о доходах и расходах (в случае если человек является госслужащим или членом семьи госслужащего).	Сайты государственных служб
7	Действительность паспорта гражданина	http://services.fms.gov.ru/info-service.htm?sid=2000

1. Постановка проблемы, в рамках чего формулируется набор вопросов (задач), на которые необходимо найти ответы;
2. Планирование, в ходе которого определяются методические приемы к поиску и анализу данных;
3. Сбор данных, в процессе которого осуществляется непосредственное получение первичной информации относительно тех объектов исследования, которые были определены на предыдущем этапе;
4. Обработка данных, которая позволяет извлечь полную информацию для анализа (например, по ранее выявленным фотографиям можно определить временные метки и геометки);
5. Анализ информации, который нацелен на поиск ответов на поставленные вопросы (решение задач);
6. Разработка отчета, в рамках которого формулируются выводы, рекомендации, приводятся подтверждающие свидетельства;
7. Распространение информации, которая предполагает презентацию результатов заинтересованным лицам (заказчику).

Концептуальная модель цикла может быть следующей:

$$M = \langle S1, S2, S3, S4, S5, S6, S7, R1, R2, R3 \rangle,$$

где: *S1* – множество вопросов, *S2* – упорядоченное множество источников, *S3* – множество полученных данных, *S4* – множество данных, извлеченных в результате декомпозиции, *S5* – множество ответов на вопросы, *S6* – множество выводов, *S7* – множество выводов в визуальном виде, $R1: S1 \cup S2 \rightarrow S3$, $R2: S3 \rightarrow S4$, $R3: S1 \cup S3 \cup S4 \rightarrow S5$.

Данная модель предполагает дальнейшее структурирование всех указанных процессов. Рассмотрим их подробнее.

2.1. Формулировка проблемы

В зависимости от имеющихся условий проблема может включать комплекс вопросов или задач.

В рамках методологии конкурентной разведки проблему часто трактуют как разрыв между желаниями и реальностью [10-13]. Например, проблема может являться результатом сложившейся ситуации, в которой есть начальная точка T_0 (с которой проблема возникла), а также событие T_X, нарушающее равновесие, которое показывает нам, что ситуация движется по нежелательному для нас сценарию. Также есть некий желаемый результат *P2* и есть нежелаемый результат *P1*. Для решения проблемы следует определить причины расхождения и меры для его устранения. В процессе сбора и анализа информации необходимо найти ответы на следующие вопросы [10]:

1). Что происходит? (Ситуация $[T_0 + T_X]$);
2). Что нас не устраивает? (*P1*);
3). Что мы хотим получить вместо этого? (*P2*)?

Правильно сформулировать проблему зачастую означает наполовину решить задачу поиска ее решения. Для уточнения формулировок могут использоваться следующие методические приемы:

1). Перефразировать: сформулировать проблему другими словами без потери изначального значения. Пример: *как можно не допустить рост затрат? → как можно ограничить рост затрат?*

Рис. 2. Для каждой фазы есть свои методы структурирования

2). Перевернуть на 180 градусов: сформулировать обратную проблему. Пример: *как нам убедить сотрудников поехать всем вместе в театр? → как нам отговорить сотрудников от поездки в театр?*

3). Расширить фокус проблемы. Пример: *стоит ли поменять мобильный телефон? → как стать более представительным?*

4). Перевести фокус на что-то другое. Пример: *как нам уменьшить затраты? → как нам увеличить продажи?*

5). Задавать рекурсивно вопрос «почему?» и переформулировать проблему, добираясь до сути. Например: *человек плохо себя чувствует; почему? → болит голова; почему? → накануне праздновал день рождения.*

2.2. Планирование

На этапе планирования необходимо определиться, какие данные и для чего следует искать. Из-за большого объема информации о человеке заниматься сбором всей имеющейся информации не всегда целесообразно, т.е. необходимо фокусироваться на тех данных, которые позволят решить сформулированную проблему и найти ответы на поставленные вопросы (или решить задачи).

Для структурирования подхода к планированию (а также последующего анализа) удобно воспользоваться методикой, известной как «тестирование альтернативных гипотез» [14]. Алгоритм тестирования альтернативных гипотез в данном случае можно представить состоящим из восьми шагов, а именно:

1). Зафиксировать гипотезы для анализа;

2). Подготовить список важных свидетельств и аргументов, поддерживающих и противоречащих каждой гипотезе;

3). Подготовить матрицу, сопоставляющую гипотезы со свидетельствами, а также провести анализ, насколько свидетельства позволяют определить относительную вероятность каждой гипотезы;

4). Пересмотреть матрицу, причем при необходимости смены формулировки гипотез удалить свидетельства и аргументы, не представляющие ценности;

5). Сделать предварительные заключения об относительной вероятности каждой из гипотез. Продолжить анализ, стараясь опровергнуть каждую гипотезу, а не доказать ее;

6). Оценить, насколько чувствительно полученное заключение относительно немногих свидетельств, в том числе оценить последствия для проведенного анализа в случае, если свидетельство ошибочное, ведущее к заблуждению или неправильно интерпретированное;

7). Подготовить отчет с заключениями, а также обсудить относительную вероятность всех гипотез, а не только самой вероятной;

8). Определить ключевые моменты для дальнейшего наблюдения, которые могут привести к переоценке ситуации.

Фиксируя определенные гипотезы и формулируя вопросы, ответы на которые позволят выявить аргументы и факты, которые бы подтверждали или опровергали бы выбранные гипотезы, можно подготовить своеобразную карту для поиска искомой информации (рис.3а).

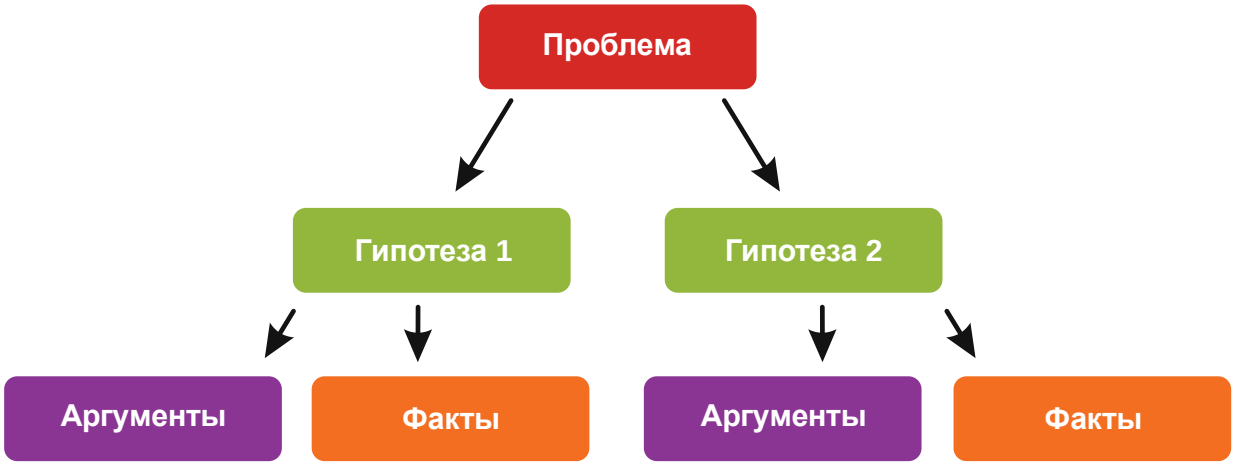

Рис. 3а. Связи между формулировкой проблемы, гипотезами, аргументами и фактами

Сформулировав гипотезы, можно создать таблицу с перечнем возможных вопросов, ответы на которые и позволят подготовить перечень свидетельств (фактов) и аргументов, поддерживающих и противоречащих каждой гипотезе. При формировании таблицы следует придерживаться принципа полноты взаимно исключающих гипотез .

Также надо понимать, что сами источники данных потребуется еще найти. Например, чтобы получить перечень друзей данного человека в социальных сетях, скорее всего, потребуется обладать следующей информацией: имя и фамилия, примерный возраст или месяц рождения, город проживания, фотография с изображением лица. Эти данные изначально можно получить в процессе знакомства с человеком, и, соответственно, можно осуществить поиск по соответствующим сайтам.

2.3. Сбор данных

Имея четкое представление о том, какая информация нам необходима, можно уже заняться технической стороной дела – поиском информации в сети Интернет.

Как правило, присутствие человека во всевозможных социальных сетях и Интернет-сервисах можно легко обнаружить, зная его ник, то есть его некий уникальный идентификатор [15, 16]. Часто люди его формируют при создании личной электронной почты, а затем используют в различных Интернет-сервисах. Автоматически проверить наличие пользователя в различных ресурсах можно, например, с помощью такого сервиса, как https://namechk.com/.

Для сбора данных активно используются поисковики, такие как: Google и Yandex. Современные поисковые машины обладают развитым языком запросов, позволяющим сузить критерии поиска информации. Так, например, у Google есть операторы, позволяющие:

- осуществлять поиск по определенному словосочетанию, например, взяв в кавычки полное имя «Иван Иванович Иванов», мы быстрее найдем нужную информацию о конкретном человеке;
- исключать из выдачи нерелевантные результаты, например, поставив «минус» перед словом «программист»: «Иван Иванович Иванов» –программист, мы уберем ссылки на ненужные нам страницы о программисте с таким же именем.
- осуществлять поиск на сайтах в определенном домене; например, используя -site:gov.ru мы сузим поиск государственными сайтами и др.

Не все знают, что зачастую можно найти сохраненную информацию, даже если ее нет на сайте. Данные удаленных страниц могут быть в кэше поисковой машины Google и тогда необходимо воспользоваться оператором cache:. Также есть специализированные сайты, сохраняющие «слепки» страниц, например, https://archive.org/web/ .

Можно сформулировать следующий типовой алгоритм первоначального поиска информации о человеке:

1). Поиск страниц, содержащих полное имя человека. Для уточнения необходимо пользоваться какой-либо известной информацией (сфера деятельности, место работы и т.п.);
2). Поиск учетных записей в социальных сетях и Интернет-сервисах;
3). Определение ника человека, адресов электронной почты, телефонов, фотографий;
4). Поиск страниц, содержащих значимые части полученных данных, например: только имя электронной почты (без доменного имени), специфические фразы (из анкет);

5). Поиск данных из среды окружения, например, фотографий, размещаемых друзьями, и т.д.

Дальнейшие шаги зависят от конкретной задачи поиска.

2.4. Обработка

В ходе обработки из полученных данных извлекается информация для анализа.

В качестве примеров такой обработки можно привести:

- извлечение из обнаруженных файлов скрытых данных: например, гео-меток, информации о моделях фотоаппаратов, именах пользователей, редактировавших изображение и т.п.;
- восстановление информации в случае, например, низкого разрешения фото-документа и т.п.;
- установление личности людей, знакомых с данным человеком, по полученным фотографиям;
- установление круга общих знакомых с конкретным человеком.

2.5. Анализ

Логичным и неотделимым продолжением обработки полученной информации является ее анализ. На данном этапе полученные данные могут быть сведены в таблицу с указанием того, поддерживают ли они ту или иную гипотезу.

Поэтому необходимо понимать, что польза от структурированного подхода к поиску и анализу информации будет ощутима при больших объемах информации и большом наборе гипотез, выбранных для тестирования (см. рис.4).

Рис. 3б. Извлечение гео-метки из фотографии и сопоставление с картой

	Гипотеза 1	Гипотеза 2	Гипотеза 3
Свидетельство 1	−	−	−
Свидетельство 2	+	−	+
Свидетельство 3	+	−	+
Свидетельство 4	−	+	+
ИТОГО:	2	3	1

Рис. 4. Сопоставление гипотез и свидетельств

2.6. Подготовка и распространение отчета

Так как наша статья сфокусирована на возможных методах сбора и анализа информации о человеке в сети Интернет, авторы оставят без рассмотрения вопросы написания отчета и его распространения заинтересованным лицам, что достаточно хорошо освещено в литературе [17].

3. Примеры применения структурированного анализа

В качестве примеров применения структурированных техник приведем два реальных примера из работы партнерских компаний [18] и один пример из сферы расследования преступлений [19].

3.1. Пример выявления «черного» блогера

Первый случай – о черном PR в Интернет. Ситуация: анонимный блогер написал негативную статью о компании. Алгоритм работы был следующий:

1). Была определены задачи вычислить человека и добиться извинений. Все мы знаем, как поведение человека резко меняется, когда его действия перестают быть анонимными.

2). В качестве гипотез были определены следующие:

- блогер связан с конкурирующей компанией (*H1*);
- блогер связан с компанией-клиентом (*H2*).

3). В качестве первичных источников информации выбрали его блог, социальные сети и выдачу поисковых машин.

4). В результате анализа содержимого блога обнаружены его публикации с фотографиями с профессиональной выставки, с помощью которых было определено, что он работает в конкурирующей организации.

Зная название компании, был осуществлен поиск страниц пользователей социальных сетей, указавших в качестве работодателя данную компанию. На странице одного блога была дана ссылка на личный блог, который оказался тем сайтом, на котором была опубликована информация. ФИО автора было установлено. Это позволило собрать дополнительную информацию, не только из Интернет-источников.

Взаимосвязи данных проиллюстрированы на следующем рисунке:

5). Была собрана вся информация, демонстрирующая предвзятое отношение к компании.

В итоге, блогер принес свои извинения.

3.2. Пример проверки потенциального партнера

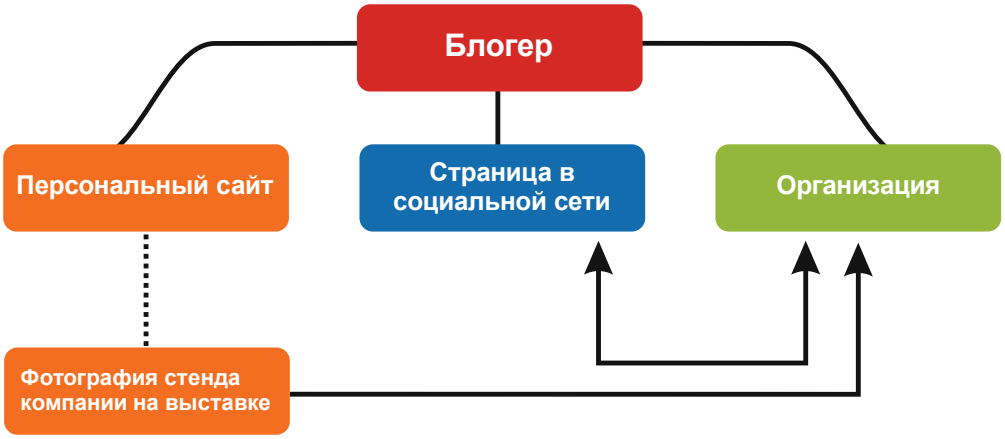

Рис.5. Пример ассоциативной карты злобного блогера

Следующая ситуация заключалась в том, что в потенциальные партнеры в регионе навязывался человек, на первый взгляд, с подозрительным прошлым, т.е. возникла потребность проверить насколько человеку можно доверять. Алгоритм проверки состоял в следующем:

1). В открытых источниках были обнаружены материалы уголовного дела, а также фотографии человека в сомнительном окружении. Описанный выше прием структурирования формулировки проблемы позволил определить направление работы по сбору и анализу информации (табл.4).

2). Для упрощения изложения материала авторы ограничились двумя взаимоисключающими гипотезами:

- потенциальный бизнес-партнер имеет криминальное прошлое/настоящее (H1);
- потенциальный бизнес-партнер не имеет криминального прошлого/настоящего (H2).

3). Сформулировав гипотезы, легко было получить таблицу с перечнем возможных вопросов, ответы на которые позволили подготовить перечень

Таблица 4.

Пример структурированного разбора ситуации

Вопрос	Ответ
Что происходит? (Ситуация [T0 + TX])	Появился человек, инициирующий партнерские отношения и озвучивающий заманчивые предложения по развитию бизнеса в данном регионе
Что нас не устраивает? (P1)	В ходе личного общения появились подозрения, что человек: › имеет криминальное прошлое; › не достаточно искренен
Что мы хотим получить вместо этого? (P2)	Объективную информацию, позволяющую принять решение о целесообразности сотрудничества с данным человеком

Таблица 5.

Пример структурированного разбора ситуации

Вопрос	Источник информации
Есть ли факты о наличии прямых знакомств с представителями криминальной среды?	Списки друзей в социальных сетях и других Интернет-сервисах
Есть ли свидетельства, что человек ведет стиль жизни, свойственный представителям криминальной среды?	Фотографии из социальных сетей
Есть ли информация, что человек был вовлечен в судебные разбирательства?	Сайты судов, СМИ

Таблица 6.

Пример анализа информации по конкретному лицу

Факт/аргумент	H1 (криминал)	H2 (отсутствие криминала)
Против человека было возбуждено уголовное дело, материалы которого были обнаружены.	Да	Нет
Уголовное дело было прекращено.	?	Да
На ряде фотографий, размещенных на страницах в социальных сетях, присутствуют люди криминального вида.	Да	Нет
На сайте суда по предположительному месту проживания человека отсутствует какая-либо информация о судебном производстве в отношении его.	Нет	?
Итого	**Да – 2, ? – 1, Нет -1**	Да –1, ? – 1, Нет -2

свидетельств и аргументов, поддерживающих и противоречащих каждой гипотезе. В данном примере для простоты изложения авторы ограничились тремя вопросами (табл.5).

4). В рассматриваемом примере удалось найти копии нескольких листов материалов уголовного дела, в свое время возбужденного против потенциального партнера. Так как материалы были убраны с сайта, их удалось извлечь из кэша Google (применялся оператор cache:).

5). В табл. 6 представлены найденные ответы и оценка гипотез.

Проведенный пример показал, что, видимо, наиболее правдива гипотеза «Потенциальный бизнес-партнер имеет криминальное прошлое/настоящее» (*H1*).

Конечно, рассматриваемый пример, предельно упрощен, в связи с чем может сложиться впечатление, что данный подход добавляет личную формалистику.

3.3. Крушение авиалайнера А321

Данный пример авторы привели, чтоб показать, что предлагаемый подход не ограничен исключительно областью безопасности кадров, но может быть полезен в других ситуациях, подлежащих анализу. Не смотря на уровень трагичности события, связанного с катастрофой А321, авторы взяли ответственность провести исследование инцидента с крушением самолета в Египте, так как это событие очень широко освещено в российской и зарубежной прессе. Алгоритм проведения структурированного анализа ситуации (который опирался только на открытые источники) с авиалайнером был следующий:

1). Следующие гипотезы сразу пришли на ум, когда произошла трагедия:
- ошибка пилота (Н1);
- технический сбой (Н2);
- запуск портативной ракеты террористами (Н3);
- взрыв бомбы (*Н4*).

2). Для понимания произошедшего была нужна следующая информация:
- информация об обломках самолета;
- информация о том, что зарегистрировали бортовые самописцы;
- информация о полете с радаров и т.п.

Вся эта информация просачивалась в Интернет от экспертов, участвовавших в расследовании.

3). Ниже приведены наиболее значимые свидетельства и аргументы в уже известную матрицу (см. рис.4). Так, самолет потерпел крушение, когда был на высоте 10 000 метров [19], это сразу же исключает гипотезу о применении переносного зенитно-ракетного комплекса. Пилоты не успели активировать сигнал SOS, что говорит о том, что все произошло мгновенно. То же самое показали записи из черных ящиков. Был зафиксирован резкий обрыв записи. На записи речевого самописца был обнаружен странный шум, но не было обнаружено сигналов сигнализации о какой-либо неисправности. Обломки самолета были разбросаны на площади более 20 квадратных километров, вытянутой в виде эллипса, что говорит о том, что он развалился/взорвался высоко в воздухе.

4). Таким образом, была получена наиболее вероятная версия: взрыв на борту некоего взрывного устройства (табл.7).

Заметим, что официальные заявления уполномоченных лиц совпали с результатами, полученными в рамках первичного (может показаться, наивного) применения предлагаемого структурированного анализа.

Таблица 7.

Пример анализа информации по крушению самолета

Событие/аргумент	Ошибка пилота (Н1)	Технический сбой (Н2)	Запуск портативной ракеты (Н3)	Взрыв бомбы (Н4).
Самолет потерпел крушение, когда был на высоте 10 000 метров	?	?	- (Н3 исключена)	?
Пилоты не успели активировать сигнал SOS	?	-		+
Обломки самолета были разбросаны на площади 8×4 км	-	?		+
Самописцы зафиксировали только обрыв связи	-	-		+
Не зафиксировано сигнала тревоги, только странный шум	-	-		+
Итого:	? – 2 «-» - 3	? - 2 «-» - 3		? - 1 «+» - 4

Заключение

Развитие сети Интернет, особенно ее мультисервисов, предоставляет потенциальную возможность получить важную информацию о конкретном человеке. В то же время динамичность данных в Интернет и сверхстремительный рост их общих объемов обуславливают необходимость использования различных оптимизационных экспертных техник сбора и анализа информации. В работе предложен методический подход к проведению структурированного анализа, позволяющий путем пошаговой декомпозиции данных и связей между данными обеспечить максимальное получение значимых сведений о человеке в приемлемое время. В основу метода предложено использовать так называемый разведывательный цикл сбора и анализа информации.

К достоинству структурированного анализа следует отнести то, что он не налагает ограничений на применение любых экспертных техник анализа и использования каких-либо дополнительных информационных источников, а также может применяться в различных областях безопасности персонала, включая менеджмент информационной безопасности и менеджмент персональных данных для решения прямых и обратных задач.

Предложенный подход имеет непосредственное прикладное значение, в работе показаны реальные примеры получения целевой информации как о субъектах, так и объектах исследования. Данный подход может использоваться для реализации комплекса контрмер по проверке персонала, внедряемого в рамках реализации системы менеджмента информационной безопасности в соответствии с требованиями стандарта ISO/IEC 27001:2013 [9]

Подход прошел неоднократную апробацию на практике, в настоящее время действует учебный курс по тематике.

Дальнейшее развитие структурированного метода видится в комплексировании различных техник экспертного и объективного анализа о конкретном человеке, а также формализации в целях максимальной автоматизации процессов сбора и анализа информации о конкретном человеке.

Литература

1. Херлберт Д., Воас Д. Большие данные и сетевые миры // Открытые системы. СУБД. 2014. № 4. С. 50-51.
2. Bershadskaya L., Chugunov A., Trutnev D. Information Society development in Russia: Measuring progress and gap. In Proceedings of the 1st Conference on Electronic Governance and Open Society: Challenges in Eurasia (St Petersburg, Russia, November 18-20, 2014). EGOSE 2014. ACM New York, NY, USA, pp. 7-13 DOI= http://dx.doi.org/10.1145/2729104.2729122.
3. Пилькевич С.В., Еремеев М.А. Модель социально значимых Интернет-ресурсов // Труды СПИИРАН. 2015. № 2. С. 62-83.
4. Bazzell M. Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information. 4 ed. CreateSpace Independent Publishing Platform, 2015. 432 p.
5. Семашко К.В., Шеремет И.А. Математическое моделирование информационно-психологических отношений в социумах. М: Наука, 2007. 157 с.
6. Марков А.С., Цирлов В.Л., Барабанов А.В. Методы оценки несоответствия средств защиты информации. М.: Радио и связь, 2012. 192 с.
7. Intelligence Cycle. FBI. 2015. https://www.fbi.gov/about-us/intelligence/intelligence-cycle .
8. The Intelligence Cycle. CIA. 2013. https://www.cia.gov/kids-page/6-12th-grade/who-we-are-what-we-do/the-intelligence-cycle.html .
9. Дорофеев А.В., Марков А.С. Менеджмент информационной безопасности: основные концепции // Вопросы кибербезопасности. 2014. № 1 (2). С. 67-73.
10. Jones M.D. The Thinker's Toolkit: 14 Powerful Techniques for Problem Solving Crown Publishing Group, 2009. 384 p.
11. Minto B. The Pyramid Principle: Logic in Writing and Thinking 3rd Edition. 3rd edition. Prentice Hall, 2010. 177 p.
12. Pherson K.H., Pherson R.H. Critical Thinking For Strategic Intelligence. CQ Press, 2012. 312 p.
13. Resseler S. Social Network Analysis as an Approach to Combat Terrorism: Past, Present, and Future Research. Homeland Security Affairs. 2006. Vol. 2, № 2 (Jul. 2006), pp. 1-10.
14. Heuer Jr. R J., Pherson R.H. Structured Analytic Techniques for Intelligence Analysis. Spi Edition. CQ Press, 2014. 280 p.
15. Bowman M., Debray S. K., Peterson L. L. Reasoning about naming systems. ACM Trans. Program. Lang. Syst. 1993, Vol. 15, № 5 (Nov. 1993), pp. 795-825. DOI= http://dx.doi.org/10.1145/161468.161471 .
16. Layton R., Perez C., Birregah B., Watters P., Lemercier M. Indirect Information Linkage for OSINT through Authorship Analysis of Aliases. Lecture Notes in Computer Science. 2013. № 7867 (Apr. 2013), pp. 36-46. DOI = http://dx.doi.org/10.1007/978-3-642-40319-4_4.
17. Clark R. M. Intelligence Analysis: A Target-Centric Approach. 4th Edition. CQ Press, 2012. 432 p.
18. Dorofeev A., Markov A., Tsirlov V. Structured Approach to the Social Network Analysis of Information About a Certain Individual. In Proceedings of the 2nd Conference on Electronic Governance and Open Society: Challenges in Eurasia (St Petersburg, Russia, November 24-25, 2015). EGOSE 2015. ACM New York, NY, USA, pp. 1-5.
19. Crash of Metrojet Flight 7K9268 (2015-10-31). Flightradar24 AB. 2015. URL: http://www.flightradar24.com/blog/crash-of-metrojet-flight-7k9268/.

Structured monitoring of open personal data on the Internet

Aleksandr Dorofeev, Director for Development of ZAO [CJSC] "NPO Eshelon", Russian Federation, Moscow.

Aleksei Markov, Doctor of Science in Technology, Senior Research Scientist, Professor at the Bauman Moscow State Technical University, Russian Federation, Moscow.

Abstract. *The paper is devoted to issues of competitive intelligence and monitoring of information security of the organisation's personnel. With a view to optimise the search of information on a certain person on the Internet, it is proposed to use a structured information collection and analysis method based on the decomposition of data and data links related to the person. A classification of open sources of information in the Internet global network is presented. Internet resources that are directly or indirectly related to information on certain people are briefly considered. A cyclic information collection and analysis model allowing to make the data and relations between data items more precise is considered. A conceptual model that enables the expert to carry out well-founded consecutive information search is developed. The main processes of information collection and analysis as well as methodological devices and structured analysis algorithms are examined in detail. Separate attention is paid to the method of forming and checking hypotheses at different stages of information collection and analysis. Real-life examples proving the effectiveness of the proposed methodological approach to analysing information on a certain person present on the Internet including social networks are presented.*

Keywords: *security monitoring, personnel security, structured analysis, intelligence cycle, social networks, Internet resources, business intelligence, dossier.*

References

1. Kherlbert D., Voas D. Bol'shie dannye i setevye miry, Otkrytye sistemy. SUBD, 2014, No. 4, pp. 50-51.
2. Bershadskaya L., Chugunov A., Trutnev D. Information Society development in Russia: Measuring progress and gap. In Proceedings of the 1st Conference on Electronic Governance and Open Society: Challenges in Eurasia (St Petersburg, Russia, November 18-20, 2014). EGOSE 2014. ACM New York, NY, USA, pp. 7-13, DOI= http://dx.doi.org/10.1145/2729104.2729122.
3. Pil'kevich S.V., Eremeev M.A. Model' sotsial'no znachimykh Internet-resursov, Trudy SPIIRAN, 2015, No. 2, pp. 62-83.
4. Bazzell M. Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information. 4 ed. CreateSpace Independent Publishing Platform, 2015. 432 pp.
5. Semashko K.V., Sheremet I.A. Matematicheskoe modelirovanie informatsionno-psikhologicheskikh otnoshenii v sotsiumakh, M., Nauka, 2007, 157 pp.
6. Markov A.S., Tsirlov V.L., Barabanov A.V. Metody otsenki nesootvetstviia sredstv zashchity informatsii, M., Radio i sviaz', 2012, 192 pp.
7. Intelligence Cycle, FBI, 2015, URL: https://www.fbi.gov/about-us/intelligence/intelligence-cycle .
8. The Intelligence Cycle, CIA, 2013, URL: https://www.cia.gov/kids-page/6-12th-grade/who-we-are-what-we-do/the-intelligence-cycle.html .
9. Dorofeev A.V., Markov A.S. Menedzhment informatsionnoi bezopasnosti: osnovnye kontseptsii, Voprosy kiberbezopasnosti, 2014, No. 1 (2), pp. 67-73.
10. Jones M.D. The Thinker's Toolkit: 14 Powerful Techniques for Problem Solving Crown Publishing Group, 2009. 384 p.
11. Minto B. The Pyramid Principle: Logic in Writing and Thinking 3rd Edition. 3rd edition. Prentice Hall, 2010. 177 p.
12. Pherson K.H., Pherson R.H. Critical Thinking For Strategic Intelligence. CQ Press, 2012. 312 p.
13. Resseler S. Social Network Analysis as an Approach to Combat Terrorism: Past, Present, and Future Research. Homeland Security Affairs. 2006. Vol. 2, No. 2 (Jul. 2006), pp. 1-10.
14. Heuer Jr. R J., Pherson R.H. Structured Analytic Techniques for Intelligence Analysis. Spi Edition. CQ Press, 2014. 280 p.
15. Bowman M., Debray S. K., Peterson L. L. Reasoning about naming systems. ACM Trans. Program. Lang. Syst. 1993, Vol. 15, No. 5 (Nov. 1993), pp. 795-825. DOI= http://dx.doi.org/10.1145/161468.161471 .
16. Layton R., Perez C., Birregah B., Watters P., Lemercier M. Indirect Information Linkage for OSINT through Authorship Analysis of Aliases. Lecture Notes in Computer Science. 2013. No. 7867 (Apr. 2013), pp. 36-46. DOI = http://dx.doi.org/10.1007/978-3-642-40319-4_4.
17. Clark R. M. Intelligence Analysis: A Target-Centric Approach. 4th Edition. CQ Press, 2012. 432 p.
18. Dorofeev A., Markov A., Tsirlov V. Structured Approach to the Social Network Analysis of Information About a Certain Individual. In Proceedings of the 2nd Conference on Electronic Governance and Open Society: Challenges in Eurasia (St Petersburg, Russia, November 24-25, 2015). EGOSE 2015. ACM New York, NY, USA, pp. 1-5.
19. Crash of Metrojet Flight 7K9268 (2015-10-31). Flightradar24 AB. 2015. URL: http://www.flightradar24.com/blog/crash-of-metrojet-flight-7k9268/.