

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ
ИНСТИТУТ ПРОБЛЕМ РЕГИСТРАЦИИ ИНФОРМАЦИИ

**А.Г. Додонов, Д.В. Ландэ,
В.В. Цыганок, О.В. Андрейчук,
С.В. Каденко, А.Н. Грайворонская**

**РАСПОЗНАВАНИЕ
ИНФОРМАЦИОННЫХ ОПЕРАЦИЙ**

Киев – 2017

УДК 004.5

ББК 22.18, 32.81, 60.54

Додонов А.Г. Распознавание информационных операций/ А.Г. Додонов, Д.В. Ландэ, В.В. Цыганок, О.В. Андрейчук, С.В. Каденко, А.Н. Грайворонская. – К.: ООО «Инжиниринг», 2017. – 282 с.

Книга посвящена рассмотрению вопросов распознавания информационных операций на основе анализа информационного пространства, прежде всего веб-ресурсов сети Интернет, социальных сетей, блогов. В этом случае технология анализа открытых источников OSINT выступает как средство сбора и первичного анализа современных информационных потоков. В книге подробно рассматриваются математические основы распознавания информационных операций, которые базируются на математической статистике, нелинейной динамике, теории сложных сетей (Complex Networks), информационном и математическом моделировании, социологии. Отдельный раздел посвящен применению подходов теории экспертных оценок, систем поддержки принятия решений при распознавании информационных операций.

Для широкого круга специалистов в области информационных технологий и безопасности.

Рекомендовано к изданию ученым советом Института проблем регистрации информации НАН Украины (протокол № 5 от 12 декабря 2017 года)

Рецензенты:

д.т.н., с.н.с. *И.Ю. Субач*

д.ю.н., профессор *К.И. Беляков*

© А.Г. Додонов, 2017

© Д.В. Ландэ, 2017

© В.В. Цыганок, 2017

© О.В. Андрейчук, 2017

© С.В. Каденко, 2017

ISBN 978-966-2344-60-8

© А.Н. Грайворонская, 2017

Оглавление

Введение	5
1. Информационные операции	10
1.1. Информационное влияние	13
1.2. Этапность информационных операций	19
1.3. Моделирование информационных операций	23
1.4. Задача распознавания информационных операций	41
1.5. Задача противодействия информационным операциям	54
1.6. Анализ документов	55
2. OSINT – разведка по открытым источникам	62
2.1. OSINT как дисциплина разведки	62
2.1.1 Области применения OSINT	65
2.1.2. Международный опыт	68
2.2. Мониторинг информационного пространства	69
2.3. Источники информации	71
2.3.1. Веб-пространство	73
2.3.2. Глубинный веб	78
2.3.3. Социальные медиа	90
2.3.4. Базы данных открытых источников	102
2.4. Технологии OSINT	112
2.5. Правовые вопросы	115
3. Элементы нелинейной динамики при распознавании информационных операций	128
3.1. Временные ряды	129

3.2. Корреляционный анализ	134
3.3. Анализ Фурье.....	141
3.4. Вейвлет-анализ	145
3.5. Корреляция с шаблоном	157
3.6. Фрактальный анализ	159
3.7. Мультифрактальный анализ	170
4. Анализ сетевых структур при выявлении информационных операций.....	191
4.1. Параметры сетей	191
4.2. Сетевые признаки информационных операций ...	200
5. Поддержка принятия решений при выявлении информационных операций.....	209
5.1. Иерархическая декомпозиция и целевое динамическое оценивание альтернатив	212
5.2. Особенности работы с экспертами	215
5.3. Методика применения СППР при выявлении информационных операций	220
5.4. Методика построения баз знаний СППР при выявлении информационных операций	228
5.5. Построение ранжирования объединенного множества альтернатив с учётом важности источников информации	231
5.6. Построение стратегии противодействия информационным операциям	249
5.7. Концепция создания информационно-аналитической системы выявления информационных операций.....	259
Заключение	238
Литература	266

Введение

Информационная операция является компонентой информационной войны. Сегодня это понятие актуально как никогда. Практически все мы невольно становимся свидетелями и участниками различных этапов информационных противоборств. В классическом понимании информационная война – это одна из форм информационного противоборства, комплекс мероприятий по информационному воздействию на массовое сознание для изменения поведения людей и навязывания им целей, которые не соответствуют их интересам, а также, естественно, защита от подобных воздействий.

Как известно, информационная война – это действия, начатые для достижения информационного преимущества путем нанесения вреда информации, процессам, которые базируются на информации и информационных системах неприятеля при одновременной защите собственной информации, процессов, которые базируются на информации и информационных системах. Основные методы информационной войны - блокирование или искажение информационных потоков и процессов принятия решений неприятелем.

Заметим также, что войны в информационной среде в современной науке и военных доктринах, в отличие от журналистской практики, в настоящее время принято называть информационными операциями [Ландэ, 2009], подчеркивая, что они являются лишь элементами реальных многоаспектных противостояний. Информационные операции являются компонентами и сопровождением более общих процессов. Вместе с тем ареной информационных операций является информационное пространство, которое, с одной стороны, является местом информационных сражений, а с другой - средой отображения реальных боевых действий.

Информационные операции определяются как «акции, направленные на воздействие на информацию и информационные системы противника, и защиту собственной информации и информационных систем» [Roadmap, 2003]. Проявления информационных операций

встречаются во многих сферах – военной, социальной, экономической.

Основная задача информационных операций состоит в манипулировании массовым сознанием с различными целями, среди которых можно назвать:

- внесение в общественное сознание и сознание отдельных людей определенных идей и взглядов;
- дезориентация людей и их дезинформация;
- ослабление определенных убеждений людей, устоев общества;
- запугивания масс.

Нет нужды доказывать, что информационные аспекты многих социальных явлений важны для понимания, проведения и противодействия информационным операциям. Действительно, трудно себе представить, например, избирателей, которые голосуют вне информационного контекста избирательной кампании или продвижение продукции без активного влияния на потенциальных покупателей.

Основой технологии современных информационных операций являются принципы синергетики, концепции эмерджентности, учет «системных эффектов». Предполагается, что запущенные в результате специальных кампаний информационные воздействия должны саморазвиваться, лавинообразно расширяться, приводя их инициаторов к желаемым результатам. Синергетические подходы базируются на рассмотрении общества как сложной системы, каждый элемент которой имеет множество степеней свободы, и поэтому гарантируют корректность результатов моделирования лишь на качественном уровне.

Информационные операции различаются по своей природе, при этом обладают крайне сложной природой, и поэтому, с трудом поддаются анализу и моделированию. Сложность связана с двумя группами факторов:

- субъективные, связанные с сознательной, целенаправленной деятельностью участвующих в информационных операциях людей;

- объективные, связанные с тем, что в социальной системе, состоящей из большого числа элементов действуют «системные эффекты», статистические закономерности.

Сегодня математическое моделирование широко применяется во многих областях науки и техники, вместе с тем, моделирование социальных процессов (именно к таким относятся информационные операции) остается открытой для исследователей проблемой.

В социальных системах, среди многих других характеристик, наиболее отчетливо проявляется целостность, т.е. наличие таких свойств, которые не присущи ни одному из составляющих систему элементов, взятому в отдельности. Это свойство, называемое «эмергентностью», является результатом возникновения между элементами системы особых синергетических связей. Под термином «эмергентность», впервые введенном Дж. Г. Луисом в 1875 г., понимается, что в физических системах целое является большим, чем сумма частей, т.е. на каждом уровне сложности возникают новые качества, которые не присущи составляющим частям. Эмергентность социальной системы не дает возможность ограничиться изучением ее элементов и связей между ними, а предполагает целостный анализ всей системы. До конца XX-го века при анализе сложных, в том числе и социальных систем, в основном использовался редукционистский подход, который стремился объяснить множество свойств сложных систем свойствами их элементов – атомов или молекул. Вследствии развития системного анализа, появления науки о сложности, технологического прорыва в вычислительных возможностях ситуация резко изменилась. Сегодня развились такие направления, как теории хаоса, сложных сетей, нелинейных и самоорганизующихся систем. В результате компьютерного моделирования сложных систем оказалось, что многие их свойства не могут быть выведены из заранее определенного набора динамических уравнений. Наоборот, уравнения могут быть получены только в результате численного моделирования.

Вместе с тем, очевидно, что невозможно разработать и применять на практике некоторую универсальную

методику моделирования информационных операций. Это, в первую очередь, связано со слабой формализуемостью многих понятий и факторов. В каждом отдельном случае приходится доверять информированности и интуиции аналитиков, профессионально занимающихся вопросами организации и противодействия информационным операциям. Иногда им удается прогнозировать отдельные закономерности процессов, отчетливо проявляющиеся себя на уровне общественной практики.

С объективными факторами дело обстоит совершенно иначе. Они вполне поддаются анализу на статистическом уровне и допускают количественные оценки, которые могут использоваться для построения обоснованных прогнозов. Современная прикладная статистика включает большой арсенал детально разработанных и апробированных методов. Однако статистика позволяет описывать лишь формальные аспекты изучаемых явлений, оставляя за бортом содержательные аспекты. Поэтому возникает необходимость расширения набора инструментальных средств, используемых при анализе и моделировании информационных операций. Одним из наиболее интересных и перспективных направлений в этом плане является, безусловно, математическое моделирование. Главное его достоинство – возможность одновременного учета как формальных, так и содержательных аспектов, определяющих динамику исследуемых процессов. Действительно, структура обоснованной модели всегда отражает смысл процессов, доступный пониманию ее авторов. Кроме того, адекватные модели строятся с помощью апробированных методов, обеспечивающих их формальную строгость.

Применительно к информационным операциям перспективным является моделирование, обусловленное некоторыми реалистичными правилами поведения отдельных элементов системы, уточняющимися некоторыми изменяемыми при моделировании параметрами. В этом случае наибольшую ценность приобретает обратная задача - по реальному поведению некой зависимости оценить величину параметров модели.

Попытки моделирования информационных операций предпринимались давно, но они тормозились

вычислительными трудностями, особенно в случае необходимости описания динамики систем с обратными связями. В настоящее время имеется достаточно много возможностей для эффективной компьютерной обработки данных, что позволяет, с одной стороны, подготавливать наборы входных параметров на основании анализа результатов статистических исследований, а с другой стороны – решать формализованные задачи с достаточной степенью точности и в допустимое время. Все это дает основания полагать, что в ближайшее время математическое моделирование станет основным инструментальным средством планирования информационных операций и противодействия им.

В работе рассмотрены некоторые модели информационных потоков в сети Интернет, которые, с одной стороны влияют на информационные операции, а с другой – все чаще становятся средой их проведения. При этом большее внимание уделено таким частным вопросам как технологии и моделирование информационных операций, информационных влияний, информационных потоков, и как результат – распознавание информационных операций.

При изучении информационных операций необходимо учитывать объективные критерии, и в качестве одного из таковых, можно рассматривать динамику распространения информационных сюжетов в соответствующем фрагменте информационного пространства.

При этом выбор тематики информационных потоков, позволяющих детектировать информационные операции, является содержательно сложной, неоднозначной задачей.

Монография содержит результаты исследований, проведенных при грантовой поддержке Государственного фонда фундаментальных исследований по конкурсному проекту № Ф73/23558 «Разработка методов и средств поддержки принятия решений при выявлении информационных операций».

1. 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» [Roadmap, 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, искажать 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 они – существенная и традиционная составляющая боевых операций. Несмотря на то, что формальное определение в документах Департамента обороны США ориентированно на военные аспекты, оно вполне применимо практически для любой области жизни.

Ниже будут рассматриваться такие информационные операции, 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 (ЛПР). Ценность информации включает ее своевременность, точность и «аналитичность». С практической точки 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 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, что 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.

Одной из целей 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. 1.1).

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.

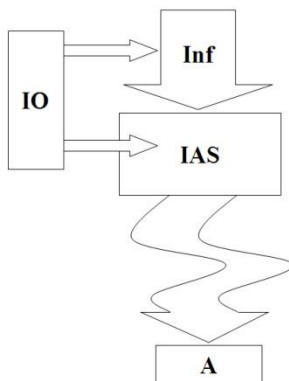


Рис. 1.1 – Воздействие на информационно-аналитическую систему противника: Inf – информационное пространство; IAS – информационно-аналитическая система; А — абонент системы (ЛПР); IO — информационные воздействия

Для достижения эффективности влияния на процессы принятия решения противником иногда необходимо предпринимать определенные действия в течение длительного времени, прежде чем они вступят в силу.

1.1. Информационное влияние

Одна из основных компонент информационных операций – социальное влияние, охватывающее все многообразие процессов влияния. Существенные изменения в убеждениях или отношении людей к некоторой проблеме или явлению, как ожидается, будут вести к изменению в поведении, связанном с этой проблемой.

В 1948 году Харольд Д. Лассвел [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 Лассвел, 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 динамического социального воздействия Латэйна [Latane, 1981], [Latane, 1997], развитой многими другими авторами, прежде всего, в работах [Nowak,1990], [Lewenstein, 1993], [Kacperski,2000], [Sobkowicz,2003].

Пытаясь обосновать механизм социального влияния сообщений Латэйна [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.

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 индивида i ($-\infty < I_i < \infty$). O_i представляет мнение индивидуума (± 1) по актуальному вопросу, где $+1$ и -1 представляют поддержку или возражение относительно данного вопроса, соответственно. S_i представляет силу индивида i или влияние ($S_i > 0$), β — сопротивление индивидуума к изменениям ($\beta > 0$), $d_{i,j}^\alpha$ — расстояние между индивидуумами i и j ($d_{i,j}^\alpha \geq 1$), α — показатель сокращения расстояния ($\alpha \geq 2$), N — общее количество агентов (индивидуумов, составляющих сообщество). Значение β , 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 α 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 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 популяции, входящее 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.

Внешнее воздействие при применении системы СМИ реализуется с помощью совокупности информационных воздействий, то есть манипуляциями информацией (распространение, сокрытие, изменение).

Для современных информационных операций характерные особенности, которые:

- позволяют скрыть факт их проведения, но при этом получить целевой эффект;
- является признаком целенаправленной деятельности более широкого характера;
- имеют существенно меньшую стоимость достижения целей по сравнению с традиционными средствами.

Использование сетевых СМИ в цикле управления, в частности, позволяют выполнять функции доведения принятых органами управления решений до широкой аудитории. Кроме того, в этих СМИ отображается реакция общества на принимаемые решения. Согласно исследованиям Е. Noelle-Newman [Noelle, 1973], многие люди пытаются скрывать свои взгляды и убеждения, если они противоречат мнению большинства общества. Соответственно и этой теории СМИ – эффективный инструмент влияния на формирование общественного мнения. Многие исследователи информационного противоборства в открытых системах придерживаются именно такой мысли. В частности, S. Ball-Rokeach и M. DeFleur [Ball-Rokeach, 1984] подчеркивают сильную связь между общественной системой, СМИ и аудиторией. Зависимость аудитории от СМИ определяется индивидуальными различиями получателей, масштабом освещаемых событий, количеством и централизацией информационных функций, выполняемых СМИ. Новостные сообщения часто включают в себя не только информацию о некотором событии, но и отношение автора к ней [Лозовский, 2000], [Лозовский, 2006], [Лозовский, 2003], [Лозовский, 2000], [Урсу, 2011], [Урсу, 2012]. Например, при освещении событий на востоке Украины, отношение редакции СМИ к ним можно определить по тому, как упоминаются противоборствующие стороны: «ополченцы» - «боевики», «каратели» – «силы антитеррористической операции» и тому подобное.

Для осуществления эффективного информационного воздействия на сознание людей необходимо обеспечить возможность прочтения сообщения как можно большим количеством людей [Brosius, 1992] [Wanta, 1994]. С учетом

того, что в настоящее время в день появляется более 100 тыс. новостных сообщений только в электронных СМИ украинского сегмента сети Интернет, даже последние новости о довольно значимом событии могут оказаться незамеченными.

Исследование S. Ball-Rokeach, M. Rokeach і J. Grube [Ball, 1976] показало, что люди часто пересматривают свои убеждения и модели поведения, если в них оказываются определенные противоречия. Результаты данного исследования используются для осуществления перманентного, постепенного изменения мировоззрения аудитории с помощью последовательности информационных воздействий.

С точки зрения проведения информационных операций сетевые СМИ имеют ряд преимуществ по сравнению с традиционными [Манойло, 2007]:

- оперативность, доступность, экономичность распространение информации;
- потенциально неограниченная аудитория;
- комплексность предоставления информации, и ее восприятия;
- доступность создания СМИ любой организацией или человеком.

В настоящее время информационные операции широко применяются [Почепцов, 2001], в ряде стран созданы специальные подразделения информационных операций, разработанные руководящие документы по их применению в мирное и военное время, регламенты взаимодействия подразделений и т.д. Созданы также специальные негосударственные структуры, которые имеют возможности эффективного проведение информационных операций.

1.2. 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 модель. При этом окончательное решение принимается с определенным оперативным риском.

Процесс информационного воздействия охватывает такие основные этапы [Чхартишвили, 2004]:

- предварительная разведка;
- выявление текущей обстановки, состояния противника;
- управление противником (информационное воздействие на противника с целью передачи ему сведений соответствующих замыслу управляющего);
- оперативная разведка (проверка результатов рефлексивного управления);
- оперативное управление – 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 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 (информационных 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), раскрутки этого повода, т.е. пропаганды (в отличие от мер контрпропаганды при оборонительных информационных операциях), а также принятия мер по препятствованию информационному противодействию.

Оперативное управление информационными операциями с использованием информационно-аналитических систем можно проиллюстрировать с помощью диаграммы, представленной на рис. 1.2.

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 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 (P), данные, которые определяют меры информационного воздействия – информационные операции (IO) на информационное пространство и непосредственно на объекты реального мира (людей, окружающую среду, компьютерные системы и т. д.).

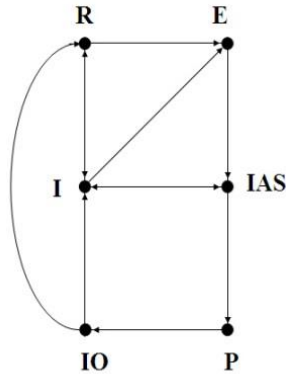


Рис. 1.2. Диаграмма оперативного управления с использованием информационно-аналитических систем

Примерный алгоритм оценки информационной операции был представлен в центре информационных операций ВС США [Roadmap, 2003] в виде восьми последовательных шагов с выполнением конкретных действий во время каждого из них:

- 1 – характеристика информационного пространства;
- 2 – включение оценки информационной операции в план (планирование) военной операции;
- 3 – изучение потребностей в информации для оценки информационной операции, и планирование сбора данных;
- 4 – составление исходных данных оценки информационной операции;
- 5 – выполнение мероприятий информационной операции и разведывательного обеспечения;
- 6 – наблюдение и сбор данных для оценки информационной операции;
- 7 – анализ полученных данных; выполнение оценки информационной операции;
- 8 – доклад полученных результатов и представление рекомендаций командиру.

Информационные операции, являясь по большей части мероприятиями, где применяется оружие нелетального воздействия, позволяют добиться тех же результатов, что и в случае применения традиционного оружия, минимизировав при этом или сведя на нет возможные негативные последствия.

По словам аналитика МО США Лары М. Дадкхах, изучавшей войну в Афганистане, возможности боевиков движения «Талибан» вести информационные операции были порой самой «эффективной противовоздушной обороной». В конечном итоге совершенно не важно, каким способом ликвидирован военный самолет – физически сбит ракетой или «посажен» в результате действия вражеской пропаганды.

Развивая данную тему, командование ВС США определило, что «целью органов информационных операций станет более точное установление соотношения применения оружия летального и нелетального воздействия для повышения эффективности применения информационных возможностей войск и тем самым более быстрого достижения цели военной операции». При этом оно рекомендовало учесть вышеназванные элементы в последующей редакции наставления КНШ ВС США JP 3-13 «Информационные операции», а также во входящих в серию JP 3-13.XXX наставлениях, имеющих прямое отношение к информационным операциям. Со своей стороны специалисты ВС отметили крайнюю необходимость подготовки и выпуска нового документа, посвященного оценке информационных операций.

1.3. 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 существует 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 состояния. При этом процесс моделирования заключается в наблюдении эволюции (согласно данным правилам) системы во времени, 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 охватывает 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 и сложных

сетей, синергетика, фрактальный и волновой (вейвлет) анализ, многоагентное моделирование, теория самоорганизованной критичности (изучающей динамическое развитие до критического состояния, характеризуемого сильными пространственно-временными флуктуациями, без внешнего управления [Вак, 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.

Для выявления информационного влияния в социальных сетях можно использовать разные модели. Рассмотрим некоторые из них: [Wasserman, 1994]

- модели с порогами, в том числе - с линейными, где агент может находиться в активном и пассивном состояниях, причем возможен переход - только из пассивного состояния в активное (обратный переход не допускается)
- модели независимых каскадов, относящихся к категории моделей так называемых «систем взаимодействующих частиц»;
- модели просачивания и заражения;
- модели Изинга;

- модели на основе клеточных автоматов;
- модели на основе цепей Маркова.

В частности, по преимущественному характеру информационных воздействий, информационные операции в сети Интернет можно условно разделить на:

- «пропагандистские», когда информационные воздействия имеют преимущественно пропагандистский характер, например, в ходе проведения и подготовки выборов, политической борьбы;
- «дезинформационные», когда основной целью является введение противника в заблуждение; – такие информационные операции используются как в политической и экономической сфере, так и при вооруженных конфликтах;
- «манипулятивные», когда главная задача - с использованием различных технологий информационных воздействий (внушения, манипулирования, модификации установок) поставить поведение конкурента под контроль;
- оборонительные («контроперация»), в которых ставится цель нейтрализовать информационное воздействие противника (например - контрпропаганда), защищаться от этого влияния, дать ответ на информационное воздействие конкурента [Манойло, 2003].

Если информационное воздействие зависит только от информированности и взаимосвязей между агентами, можно использовать классическую теоретико-игровую модель. Выигрыш, получаемый агентом зависит от действий «друзей». К теоретико-игровым моделям информационных воздействий относятся модели [Губанов, 2009]:

- взаимной информированности;
- согласованных коллективных действий;
- коммуникаций и задачи поиска минимально достаточной сети;
- стабильности сети;

- информационного влияния и управления;
- информационного противоборства.

В большинстве моделей в основном рассматриваются правила взаимодействия агентов, но что касается самой сети влияния в целом, ее свойств, взаимосвязей и процессов взаимодействия, то имеющийся анализ этих моделей предоставляет очень немного результатов. Поэтому в [Əliquliyev, 2010] предлагается сначала определить так называемый «информационный архетип» пользователей социальных сетей, а также состояния агентов. Агент, как и во многих других моделях, при этом может находиться в активном и пассивном состоянии, причем способность активного агента распространять определенную идею по сети зависит от многих факторов.

Приведенные выше различные подходы к классификации моделей информационных воздействий свидетельствуют, что известные на сегодняшний день модели адекватно отражают многие свойства и эффекты, которые имеют место в социальных сетях. При этом все модели информационного влияния и противостояния должны формулироваться с учетом специфики решаемой практической задачи и возможности идентификации моделируемой, параметров агентной среды, возможных действий активных агентов, их преимуществ и информированности.

В целом можно сделать вывод, что моделирование влияния в социальных сетях в настоящее время уже становится самостоятельной дисциплиной, такие модели станут самостоятельной ветвью исследований.

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 потребуется выбор значений параметров, которые бы в некотором разумном приближении соответствовали изучаемой ситуации, причем в большинстве случаев продуктивным оказывается использование относительных величин. Так,

конечно, нельзя 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 авиационные 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 перепечатавающие 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 средним специалистам в предметной области делать более-менее точные выводы.

Получить данные подобной динамики можно, например, ежедневно заходя на сайты интеграторов новостей (news.yandex.ru, webground.su, 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.

При моделировании и осуществлении информационных операций необходимо учитывать значение ценности информации для лиц, принимающих решения. Ценность информации включает ее своевременность, точность, и «аналитичность». С практической точки зрения ценность информации также может быть определена как ее значимость или применимость, пригодность к использованию. Под применимостью информации понимают обеспечение лицам, принимающим решения, доступа к готовой к использованию информации.

Для создания сетевой мультиагентной модели распространения информации, прежде всего, формируется реалистичное виртуальное информационное пространство, населенное виртуальными агентами. С ними ассоциируются отдельные сообщения в социальной сети,

которые, помимо всего прочего, инкапсулируют в себя ссылки на информационные ресурсы сети Интернет.

Ниже предложена модель формирования тематического информационного потока (ТИП) на базе сетевой мультиагентной модели [Додонов, 2015а], в рамках которой отдельные документы, образующие информационный сюжет, ассоциируются с агентами, жизненный цикл агентов - с жизненным циклом документов в информационном пространстве. Предполагается, что с течением времени происходит эволюция популяции агентов.

В данной работе рассматривается мультиагентная модель распространения информации, ключевым элементом в которой является сообщение. Каждое сообщение может вызывать разные типы реакции, такие как положительные/отрицательные комментарии, выражение поддержки или отторжения (like/dislike – лайк/дизлайк), текст сообщения может быть скопирован, размещен на другой странице (repost – репост), также в одном сообщении может быть размещена ссылка на другое сообщение (link – линк). Поэтому в качестве агента в модели будем рассматривать сообщение, а эволюция агента будет связана с событиями, которые с ним происходят. В качестве основной характеристики агента введем «энергию» (E), которая отображает актуальность сообщения и степень интереса к нему. Естественно, что старение информации или отрицательная реакция уменьшает энергию сообщения, а положительная реакция или появление ссылки на сообщение увеличивает его энергию.

Правила эволюции агента в модели

Агент появляется с начальной энергией E_0 и с каждым дискретным отсчетом времени его энергия уменьшается на 1. Будем рассматривать события типичные для социальных сетей: like, dislike, repost, link (предоставление ссылки на одного агента другим). Эти события влияют на энергию агента следующим образом: like повышает энергию на 1, dislike уменьшает на 1, repost повышает на 2, link повышает на 1. С другой стороны, вероятность того, что какое-то из этих событий произойдет, зависит от актуальности сообщения, интереса к информации в нем,

что в терминах модели выражается энергией. В связи с этим определим вероятности того, что с сообщением с энергией E произошло определенное событие, следующим образом:

$$P_{like}^{(E)} = p_{l_0} \varphi(E); \quad P_{dislike}^{(E)} = p_{d_0} \varphi(E); \quad P_{repost}^{(E)} = p_{r_0} \varphi(E),$$

где p_{l_0} , p_{d_0} , p_{r_0} – параметры модели, а φ – это некоторая монотонно неубывающая функция от текущей энергии агента со значениями в $[0, 1]$. При падении энергии до 0, агент «умирает» и больше не рассматривается.

На рис. 1.3 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.

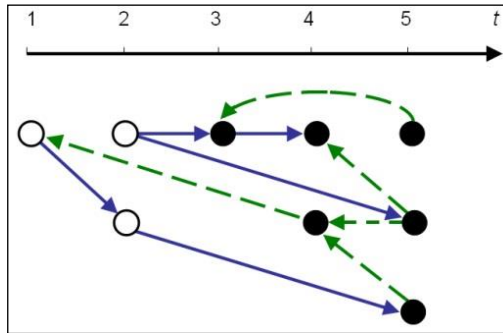


Рис. 1.3 – Фрагмент мультиагентного пространства

Моделирование динамики информационного потока

На Рис. 1.4 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.

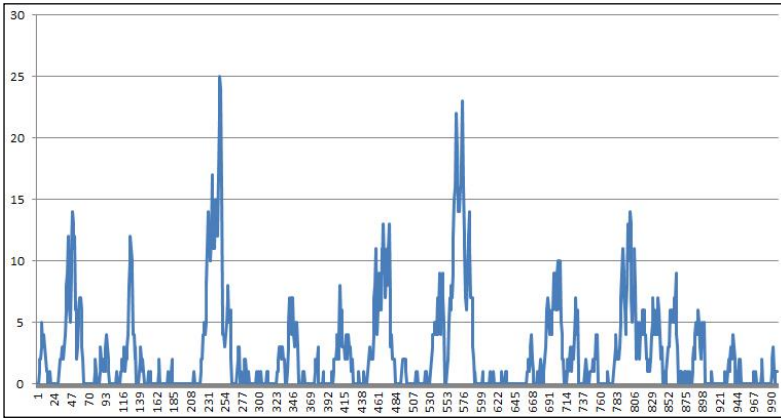


Рис. 1.4 – Динамика изменения количества агентов в модели

Моделирование динамики всего информационного потока начинается с одного агента. Появление нового агента возможно двумя способами. Первый заключается в копировании существующего агента с помощью операции репост. Также возможно самозарождение агента, что отвечает публикации нового сообщения. Таким образом, в каждый момент времени с определенными вероятностями, с каждым из агентов, может произойти любое из событий. Также в любой момент времени с вероятностью p_s может появиться новый агент в результате самозарождения.

Рассмотрим жизненный путь одного агента [Ландэ, 2016a]. Агент появляется с начальным значением энергии E_0 и далее его энергия изменяется в зависимости от событий, которые с ним происходят. Будем считать, что возможны два события: лайк и репост. За единицу времени может произойти одно из этих событий, оба одновременно или не произойти ни одного.

Обозначим ε_t значение энергии агента в момент времени t . Тогда значение энергии в следующий момент времени можно записать следующим образом

$$\varepsilon_{t+1} = \varepsilon_t + \delta_t,$$

где δ_t является случайной величиной со значениями в $\{-1, 0, 1, 2\}$. Согласно с правилами изменения энергии, введенными выше, увеличение энергии на 2 соответствует тому, что произошли одновременно лайк и репост; увеличение на 1 – произошел только репост; энергия не меняется, если был лайк; и уменьшается на 1, если не произошло ни одно из событий. Следовательно, можно указать условное распределение δ_t при известной энергии ε_t :

$$P(\delta_t = 2 | \varepsilon_t = E) = p_{like}^{(E)} p_{repost}^{(E)};$$

$$P(\delta_t = 1 | \varepsilon_t = E) = (1 - p_{like}^{(E)}) p_{repost}^{(E)};$$

$$P(\delta_t = 0 | \varepsilon_t = E) = p_{like}^{(E)} (1 - p_{repost}^{(E)});$$

$$P(\delta_t = -1 | \varepsilon_t = E) = (1 - p_{like}^{(E)}) (1 - p_{repost}^{(E)}).$$

Данные формулы справедливы при $E > 0$. Далее будем использовать обозначение $P_{\Delta}^{(E)} = P(\delta = \Delta | \varepsilon = E)$.

Процесс изменения энергии агента можно рассматривать как целочисленное случайное блуждание с переходными вероятностями

$$p_{ij} = \begin{cases} P_{j-i}^{(i)}, & (j-i) \in \{-1, 0, 1, 2\}, \quad i > 0; \\ 1, & i = j = 0; \\ 0, & \text{иначе.} \end{cases}$$

Так как значение энергии в следующий момент времени зависит только от значения энергии в предыдущий момент времени, то стохастическая последовательность $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_t, \dots)$ является марковской цепью с переходными вероятностями p_{ij} . Граф, соответствующий такой марковской цепи показан на рис. 1.5.

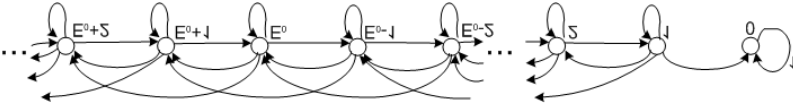


Рис. 1.5 – Граф, соответствующий случайному блужданию энергии агента

Можно явно указать вероятностное распределение последовательности $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_t, \dots)$ через переходные вероятности:

$$P((\varepsilon_0, \varepsilon_1, \dots, \varepsilon_t, \dots) = (E_0, E_1, \dots, E_t, \dots)) = \prod_{i=1}^{\infty} p_{E_{i-1}E_i}.$$

Или аналогичное распределение для $(\delta_0, \delta_1, \dots, \delta_t, \dots)$:

$$P((\delta_1, \delta_2, \dots, \delta_t, \dots) = (\Delta_1, \Delta_2, \dots, \Delta_t, \dots)) = \prod_{i=1}^{\infty} P_{\Delta_j}^{(E_0 + \sum_{j=1}^{i-1} \Delta_j)},$$

при условии, что $\Delta_i \in \{-1, 0, 1, 2\}$ и если для некоторого

$k: \sum_{i=1}^k \Delta_i = -E_0$ (энергия агента упала до 0 на k -ом шаге), то $\Delta_l = 0, \forall l > k$. Для последовательностей $(\Delta_1, \Delta_2, \dots, \Delta_t, \dots)$, которые не удовлетворяют этим условиям, вероятность равна 0. Далее будем рассматривать только те последовательности, которые удовлетворяют указанным условиям.

Теоретически существуют последовательности $(\Delta_1, \Delta_2, \dots, \Delta_t, \dots)$, которые реализуются с положительной вероятностью, и при этом $k: \sum_{i=1}^k \Delta_i = -E_0$ может быть сколь угодно большим. Заметим, что k соответствует длине жизни агента и имеет смысл времени, на протяжении которого информационное сообщение остается актуальным. Следовательно, длина жизни агента должна быть конечной с большой вероятностью.

Обозначим τ_{E_0} длину жизни агента с начальным значением энергии E_0 или, что то-же самое, время, за которое из E_0 попали в 0. В реалистичной модели хотелось бы иметь оценку $P(\tau_{E_0} > T_{max}) < \varepsilon$ для малого ε и не очень

большого значения T_{max} , для того, чтобы можно было вместо бесконечных последовательностей $(\Delta_1, \Delta_2, \dots, \Delta_t, \dots)$ рассматривать конечные.

Рассмотри функцию $\rho_T(E) = P(\tau_E > T)$. Справедливо рекуррентное соотношение:

$$\rho_T(E) = P_2^{(E)} \rho_{T-1}(E+2) + P_1^{(E)} \rho_{T-1}(E+1) + P_0^{(E)} \rho_{T-1}(E) + P_{-1}^{(E)} \rho_{T-1}(E-1).$$

Систему таких рекуррентных соотношений можно решить, используя начальные условия:

$$\rho_0(E) = \begin{cases} 0, & E = 0, \\ 1, & E \neq 0. \end{cases}$$

При начальных параметрах $p_{l_0} = 0.4$, $p_{r_0} = 0.1$ из решения рекуррентного уравнения можно получить оценку $P(\tau_{E_0} > 1.5E_0) < 10^{-3}$. То есть время жизни агента ограничено $1.5E_0$ с большой вероятностью и, следовательно, для получения достаточно точных оценок распределения для $(\delta_0, \delta_1, \dots, \delta_t, \dots)$ можно рассматривать вектора конечной длины $T_{max} = 1.5E_0$.

Представляет интерес распределение количества событий, которые происходят с агентами. Рассмотрим распределение количества лайков. Заметим, что если в момент времени t произошел like, то $\delta_t \in \{0, 2\}$, если же нет, то $\delta_t \in \{-1, 1\}$. Обозначим $(\Delta'_1, \dots, \Delta'_{T_{max}})$ вектор, который удовлетворяет условию, что $\Delta'_t \in \{0, 2\}$, при $t = t_1, \dots, t_n$ и $\Delta'_t \in \{-1, 1\}$ иначе, где $0 < t_1 < \dots < t_n < T_{max}$. Тогда для агента, получившего лайк справедлива формула:

$$P\{like\} = \sum_{t_1 < \dots < t_n} \sum_{(\Delta_1, \dots, \Delta_{T_{max}})} \prod_{i=1}^{T_{max}} P_{\Delta'_i}^{(E_0 + \sum_{j=1}^{i-1} \Delta_j)}.$$

Численные результаты

На рис. 1.6 показана плотность полученного распределения, при начальных параметрах $p_{l_0} = 0.4$, $p_{r_0} = 0.1$. Точками, соединенными отрезками прямых линий, обозначены полученные значения P (агент получил n лайков)

Гладкая кривая соответствует плотности распределения Вейбулла:

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Параметры распределения Вейбулла k и λ были получены методом максимального правдоподобия. При указанных начальных параметрах, полученные значения $k = 1.9$, $\lambda = 3.8$.

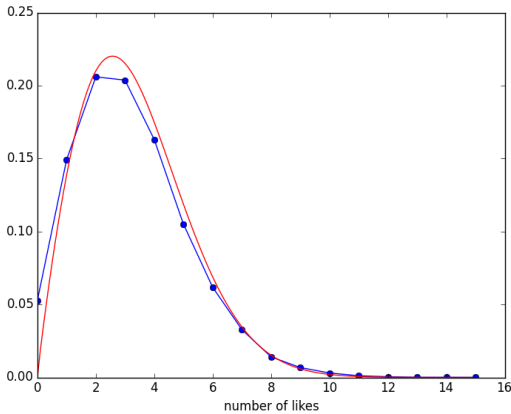


Рис. 1.6 – Плотность распределения количества лайков, полученных агентом при начальных параметрах $p_{l_0} = 0.4$, $p_{r_0} = 0.1$

На рис. 1.7 представлен аналогичный результат при начальных параметрах $p_{l_0} = 0.3$, $p_{r_0} = 0.1$. В данном случае полученные параметры распределения $k = 1.9$, $\lambda = 3.0$.

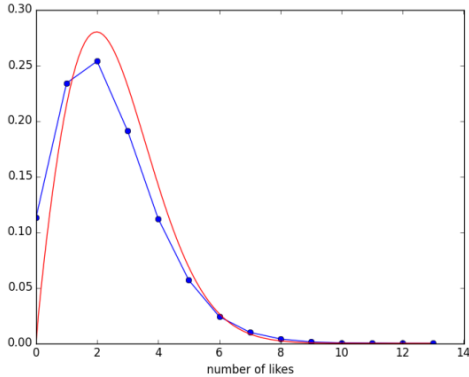


Рис. 1.7 – Плотность распределения количества лайков, полученных агентом при начальных параметрах $p_{10} = 0.3$, $p_{r0} = 0.1$

Исследование реальных информационных потоков

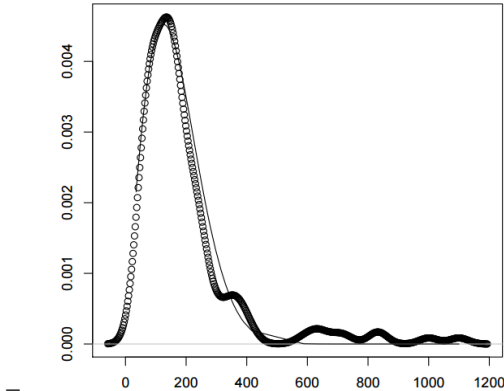
Полученные результаты моделирования сравнивались с результатами проведенными авторами исследования жизненного цикла новостных сообщений в сети микроблогов Twitter, где, в частности, анализировались характеристики роста количества специальных ретвостов (ретвиттов) выбранных сообщений [Li, 2012]. Распределение лайков и ретвиттов в этом случае, как и в модели, соответствовало стандартному распределению Вейбулла, причем параметр k с высокой точностью совпал с модельным (рис. 1.8).

Программные средства, разработанные с помощью языка программирования R для статистических расчетов, состояли из трех частей:

- Средства сканирования и накопления данных о росте количества ретвиттов для сообщений отдельного пользователя сети в режиме он-лайн (для сообщений от газеты Нью-Йорк Таймс был выбран цикл сканирования в 15 с).
- Обработка накопленной информации путем аппроксимации данных о росте количества ретвиттов функцией Вейбулла, получение соответствующих

коэффициентов масштаба и формы (нелинейная аппроксимация с помощью метода наименьших квадратов), а также вычисление производной для оценки скорости роста количества ретвитов сообщений и построения необходимых графиков.

- Накопление полученных результатов для дальнейшего анализа. С этой целью данные, полученные с помощью программных средств на языке R, импортируются во внешние базы данных.



- Рис. 1.8 – Плотность распределения количества ретвитов полученных из реальной сети (аппроксимация распределением Вейбулла при $k = 1.9$, $\lambda = 180$)

Таким образом, накапливались записи, включающие текст каждого сообщения, время его публикации в сети, значения коэффициентов масштаба и формы, графики роста количества ретвитов и лайков, графики аппроксимации данных о росте числа ретвитов функцией Вейбулла, график скорости роста количества ретвитов и т.д.

В результате описанных исследований построена мультиагентная модель «жизненного цикла» новостей в информационных сетях.

В результате моделирования выявлены статистические закономерности, относящиеся к количеству лайков и репостов отдельных сообщений, распределение которых, как видно по результатам моделирования, соответствуют распределению Вейбулла.

Данные моделирования проверены путем исследования реальной сети микроблогов Twitter. Совпадение результатов моделирования и параметров распределения реальной сети позволяют говорить о закономерности, присущей реальным сетям, а также об адекватности модели.

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.

1.4. Задача распознавания 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. 1.9), 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 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 информационные потоки, соответствующие тематикам информационных операций – тематические информационные потоки.

Исследованию динамики информационных потоков посвящены многочисленные научные работы [Corso, 2005], [Kleinberg, 2006], [Ландэ, 2006], [Rakesh, 2014] показано, что в типовых ситуациях динамике распространения новостей, информационного сюжета присущ характер «всплеска», волны с явным периодом возрастания его влияния и плавным спадом.

В результате анализа многочисленных диаграмм поведения тематических информационных потоков (ТИП), были выявлены наиболее типичные, базовые профили их поведения [Ландэ, 2012]. Некоторые сюжеты развиваются следующим образом: после быстрого информационного всплеска подготовки идет плавный спад (например, публикации о стихийных бедствиях), некоторые, напротив, предполагают длительную плавную информационную подготовку, после чего идет резкий спад (например,

публикации об планируемых заранее мероприятиях). Существуют также тематические потоки, характеризующиеся симметричной кривой динамики, как узкие, кратковременные, так и растянутые во времени.

В случае информационных потоков, которые ассоциируются с конкретными тематическими потоками, необходимо описывать динамику каждого из таких потоков отдельно, принимая во внимание то, что рост одного из них может автоматически приводить к уменьшению других и наоборот. Поэтому ограничение на объемы информации по всем тематикам распространяется и на совокупность всех информационных сюжетов. В случае изучения общего информационного потока наблюдается явление «перетекания» объемов публикаций из одних, теряющих актуальность информационных сюжетов, в другие.

Приведенные в [Горбулін, 2009] тренды сообщений, соответствующих этапам информационной операции показаны на рис. 1.9. 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.

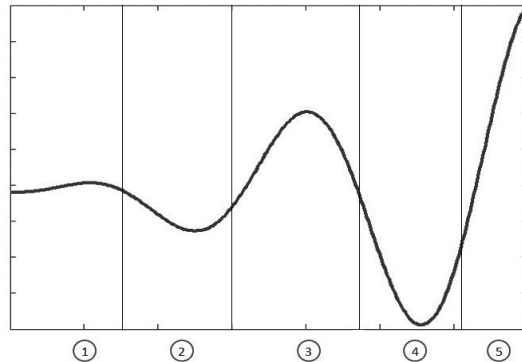


Рис. 1.9 – Динамика количества тематических сообщений во время проведения информационной операции: 1 – фон; 2 – затишье; 3 – «артподготовка»; 4 – затишье; 5 – атака/триггер роста

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. 1.10 приведена обоснованная в [Хорошевский, 2012] диаграмма количества публикаций, соответствующая тренду инновационной деятельности.

Объединяя графики, соответствующие началу информационной операции (рис. 1.9) и тренду инновационной деятельности (рис. 1.10), можно получить полный график, соответствующий отображению информационных операций в информационном пространстве (рис. 1.11).

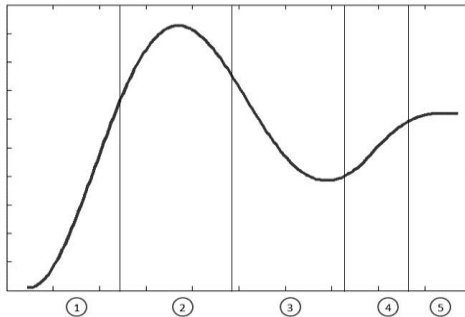


Рис. 1.10 – Диаграмма количества публикаций, соответствующих тренду инновационной деятельности: 1 – атака/триггер роста; 2 – пик завышенных ожиданий; 3 – утрата иллюзий; 4 – общественное осознание; 5 – продуктивность/фон

Предполагается, что системное нарушение типовой динамики некоторых информационных сюжетов в открытом информационном пространстве может свидетельствовать как об информационных операциях [Горбулін, 2009], так и о существовании информационной резервации. При исследовании информационных операций также большое внимание уделяется анализу динамики информационных сюжетов [Додонов, 2013], 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», то уже по первым трем компонентам можно с большой вероятностью предсказать будущие события.

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 путем 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].

Отметим, что предложенная модель позволяет отличать информационные потоки, поведение которых определяется естественными закономерностями информационного пространства, от потоков, вызванных влиянием внешних факторов. В частности, в случае информационных резерваций, в качестве индикатора может рассматриваться отклонение трендов динамики некоторых информационных сюжетов от характерных форм распределения, появление периодических зон

нестабильности значений, или, наоборот, удивительная локальная стабильность этих значений.

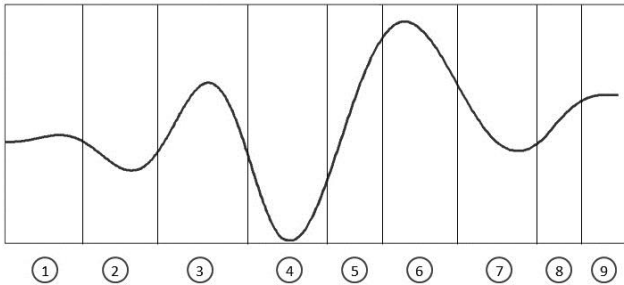


Рис. 1.11 – Обобщенная диаграмма, соответствующая всем этапам жизненного цикла информационных операций: 1 – фон; 2 – затишье; 3 – «артподготовка»; 4 – затишье; 5 – атака/триггер роста; 6 – пик завышенных ожиданий; 7 – утрата иллюзий; 8 – общественное осознание; 9 – продуктивность/фон

В качестве примера, на рис. 1.12 показана динамика публикаций в RUNet (тематических информационных потоков) по запросам «Банки, Кипр», «Оффшор», «Вирджинские острова» за март-апрель 2013 года, в период известных кризисных событий, полученная с помощью системы InfoStream. Как видно из рис. 1.12, 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 «Оффшор» и «Банки Кипра» (0,73), а также «Оффшор» и «Вирджинские острова» (0,77).

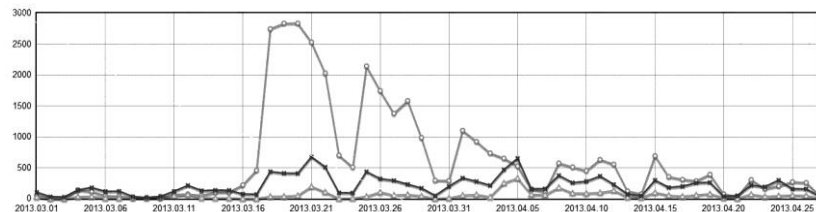


Рис. 1.12 – Диаграмма динамики тематических информационных потоков по запросам: о – «Банки Кипра»; Δ – «Вирджинские острова»; х – «Оффшор»

По-видимому, проявления информационных операций в области оффшорных банков в данном случае лучше всего увидеть при анализе более общей тематики – «Оффшоры». На графике соответствующего числового ряда четко видны две области локальных экстремумов, соответствующие кризисным ситуациям на Кипре и на Вирджинских островах, а также фазы, соответствующие «затишьям» и «артподготовкам».

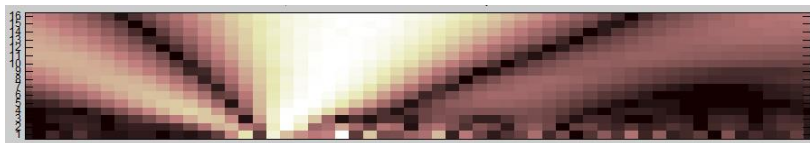
Можно высказать предположение, что если динамика частного информационного потока в какой-то момент начинает существенно отличаться от динамики потока, соответствующего более общей тематике (как в рассматриваемом случае, «Банки Кипра» и «Оффшор»), то возможно проявление признаков начала информационной операции, относящейся узкой тематике.

При проведении вейвлет-анализа [Астафьева, 1996], [Buckheit, 1995] (рис. 1.13) было принято решение использования вейвлета «Мексиканская шляпа», как близкого по форме к диаграмме, приведенной на рис. 1.11.

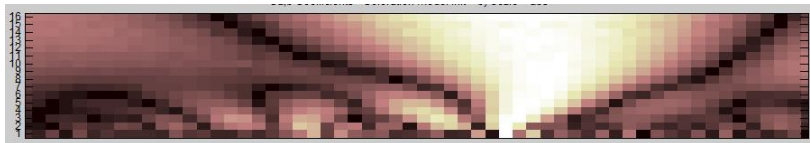
Рассматриваемые процессы четко просматриваются как на вейвлет-спектрограммах, так и на соответствующих им скелетонах (графиках линий экстремумов).

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

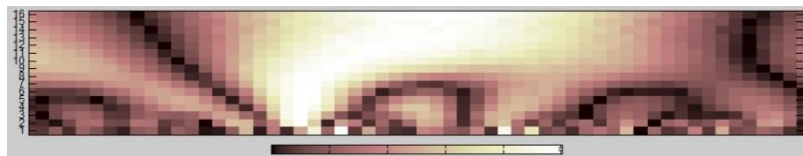
соответствующие коррективы, например, искусственно моделировать случайные отклонения.



а)



б)



в)

Рис. 1.13 – Вейвлет-скейлограммы, соответствующие динамике тематических информационных потоков по запросам: а – «Банки Кипра»; б – «Вирджинские острова»; в – «Оффшор»

Ниже приведены результаты экспресс-анализа тематического информационного потока, соответствующего объекту – Национальной академии наук Украины, с учетом рисков, которым она подвергалась во второй половине 2015 года. В результате анализа средствами системы контент-мониторинга InfoStream был сформулирован запрос за период с 01.07.2015 по 31.12.2015, в результате отработки которого был получен тематический информационный поток объемом 1932 документа с украинского сегмента веб-пространства. Для выявления информационных операций доступными аналитическими средствами анализировалась динамика публикаций по целевой тематике (рис. 1.14).

Понятия в динамике :

• (нан] | нану] | (нац-академ-наук)) & (реорган | расформ | реформиру | реформируван | скасуван | заборгован | задолж |

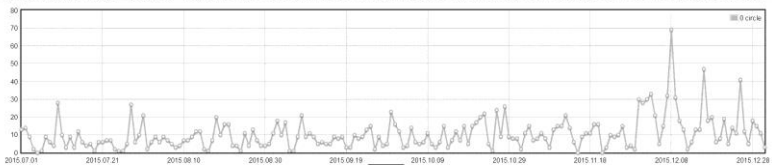


Рис. 1.14 - Динамика тематического информационного потока

Для выявления степени «близости» фрагментов исследуемого временного ряда диаграмме информационной операции в различных масштабах предлагается использовать «вейвлет-анализ», который в настоящее время нашел применение как в естественных науках, так и в социологии.

Вейвлет-коэффициенты показывают, насколько поведение процесса в данной точке аналогично вейвлета в данном масштабе. На вейвлет-скейлограмме (рис. 1.15) видно все характерные особенности исходного ряда: масштаб и интенсивность периодических изменений, направление и значение трендов, наличие, расположение и продолжительность локальных особенностей.

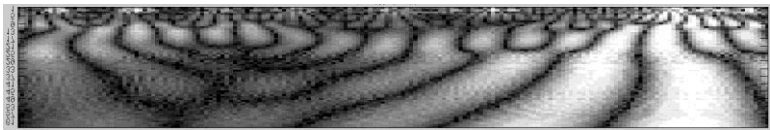


Рис. 1.15 – Вейвлет-скейлограмма (вейвлет Морле) рассматриваемого информационного потока

После определения критических точек, с помощью системы контент-мониторинга выполняется построение основных сюжетных цепочек, из сообщений запроса по выбранным датам. Таким образом, определяются основные события за указанные даты: *Национальную академию наук хотят пустить по ветру ... (Vector News) 2015.12.03*

- *Под Радой митингуют молодые ученые (Левый берег) 2015.12.08*
- *Украинское правительство разрушает украинскую науку (2000.ua) 2015.12.08*
- *"В мире науки - связаны руки": у АП митингуют сотрудники НАН Украины (УНИАН) 2015.12.16*

Проведенный анализ, в частности, показывает, что в отношении НАН Украины ведется целенаправленная информационная операция, противодействие которой широко отражается в информационном пространстве страны.

Для моделирования информационных операций в рамках этой работы создан набор шаблонов информационных операций в различных масштабах, который соответствует этапам 1-6 формы, приведенной на Рис. 1.11. Примеры шаблонов разной длины приведены на Рис. 1.16.

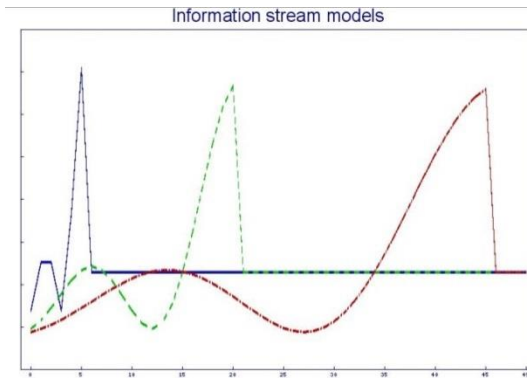


Рис. 1.16 – Примеры шаблонов длиной 5, 20 и 45

Сопоставление разработанной системы шаблонов с временным рядом, соответствующем реальному тематическому информационному потоку (Рис. 1.17) в построенной модели осуществляется путем корреляционного анализа.

Соответствующая корреляционная диаграмма, в которой горизонтальная ось соответствует времени, вертикальная – масштабу (длине шаблонов), а окрашивание – коэффициентам корреляции, приведена на рис. 1.18. Эта диаграмма позволяет четко визуализировать этапы проведения информационных операций, соответствующие определенным шаблонам при различных масштабах наблюдения.

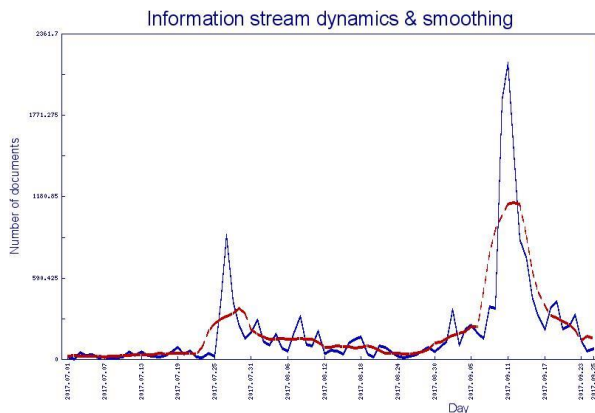


Рис. 1.17 - Временной ряд, соответствующий реальному тематическому информационному потоку и кривая сглаживания

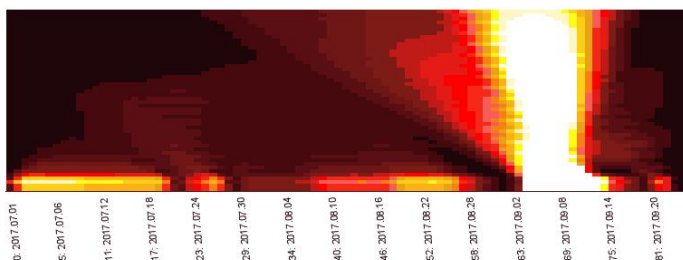


Рис. 1.18 - Корреляционная диаграмма, соответствующая корреляции тематического информационного потока и приведенной выше системы шаблонов

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.

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:

$$\sum_{k=1, k \neq s}^n c_k F_{i,k} = F_{i,s}, \quad i = 1, \dots, n, \quad i \neq s,$$

где центрированные коррелянты $F_{i,k}$ 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].

В соответствии с [Расторгуев, 2014] выявление информационных угроз, как и любых других угроз, целесообразно начинать с анализа подходов и методов целенаправленного перепрограммирования информационных систем, будь то технические системы, или социальные структуры, или люди. Выделение последовательности присущих информационной операции событий, а затем выявление этих событий в общем «жизненном шуме» — это классический путь решения подобных задач. Данный путь состоит из следующих этапов:

1. Последовательность событий, составляющих угрозу, всегда связана с целью. Достижение цели предполагает перепрограммирование субъектов информационного противоборства. Чтобы понимать, насколько достигнута цель, желательно уметь оценивать результаты перепрограммирования и собственные возможности,

говоря другими словами — степень возможного поражения информационным оружием той или иной системы. Одни объекты уничтожить и перепрограммировать не составляет труда, с другими — задача может оказаться неразрешимой в рамках выделенных для этой цели ресурсов из запаса времени.

2. Оценка и обоснование частоты появления событий, при которой уже можно говорить о начале информационной операции по перепрограммированию 8 .

3. Оперативное и доступное представление информации пользователю в случае выявления опасных тенденций.

1.5. Задача противодействия 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», неафелированных) 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 реалий 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.

1.6. Анализ документов

Вооруженные силы США входят в число лидеров в области теории и практики информационного противоборства.

Идеи информационных операций рассматривались с учетом положений таких документов американских вооруженных сил, как «Стратегия национальной безопасности» 2010 года, «Национальная военная стратегия» 2011 года, «Удерживая глобальное американское лидерство. Приоритеты военного строительства в XXI веке»

Суть глобальных информационных операций изложена в концепции строительства и применения национальных ВС США «Единые силы-2020». Согласно ее положениям предусматривается координация в вопросе применения войск (сил), рассредоточенных земному шару, создание объединенных группировок, точно соответствующих решаемой задаче и способных действовать как единое целое. Согласно американской военной доктрине, в настоящее время ВС США ведутся 14 направлений деятельности в сфере информационных операций (Табл. 1). За последние несколько лет в Вооруженных силах США введены в действие следующие документы (в скобках дата официальной публикации):

- новая редакция доктрины JP 3-13 "Информационные операции" (январь 2006);
- Доктрина совместных ПсО JP 3-53 (сентябрь 2003);
- Полевой устав СВ США FM 3-05.30 "Психологические операции" (апрель 2005);
- Полевой устав СВ США FM 3-05.301 "Тактика, приемы, способы и порядок проведения психологических операций" (декабрь 2003, в настоящее время уже пересматривается);
- Полевой устав СВ США FM 3-05.302 "Тактика, приемы, способы и порядок проведения тактических психологических операций" (октябрь 2005);
- Карманный справочник "Технические средства ПсО: виды, ТТХ и возможности" (апрель 2005);
- Карманное "Руководство командира по планированию ПсО" (август 2005), которое содержит выдержки из уставов FM 3-05.301 и FM 3-05.302. Оба справочника направлены в подразделения ПсО;
- Целая серия новых программ боевой подготовки (ARTEP), в том числе:
 - ARTEP 33-712-MTP "Задачи боевого планирования для штаба и штабной роты группы ПсО и роты штабной и обслуживания батальона ПсО" (апрель 2006);

- ARTER 33-715-МТР "Задачи боевого планирования для батальона подготовки и распространения материалов ПсО" (сентябрь 2006);
 - ARTER 33-737-30-МТР "Задачи боевого планирования для роты тактических ПсО" (в планах 2007);
 - ARTER 33-727-МТР "Задачи боевого планирования для региональной роты ПсО" (в планах 2007);
- Серия документов, регламентирующих вопросы боевой подготовки личного состава сил ПсО, в том числе:
- полный справочник офицера ПсО STP 33-37II-OFS "Положение об основных стандартах офицера ПсО" (2006), рассчитанный на офицеров среднего звена (капитан-подполковник) и содержащий основные задачи, вопросы боевой подготовки и требования к уровню знаний, которым должны соответствовать офицеры ПсО данных категорий;
 - наставление по подготовке специалистов ПсО рядового состава STP 33-37F14-SM-TG (апрель 2004).

Под руководством управления боевой подготовки и научных исследований ВС США ведется активная научно-исследовательская деятельность в интересах ПсО.

Таблица 1. Направления деятельности ВС США в сфере информационных операций. Источник: Joint Publication 3-08. Interorganizational Coordination During Joint Operations. 24 June 2011. p. 324 (D-10, JP 3-08)

(1) Strategic Communication (SC)	Стратегические коммуникации
(2) Joint Interagency Coordination Group	Совместная межведомственная координационная группа
(3) Public Affairs	Связи с общественностью
(4) Civil-Military Operations	Совместные мероприятия с участием военных и гражданских лиц
(5) Cyberspace Operations	Операции в киберпространстве
(6) Information Assurance	Обеспечение безопасности информационной инфраструктуры
(7) Space Operations	Космические операции
(8) Military Information Support Operations	Информационные операции военной поддержки
(9) Intelligence (information operations intelligence integration)	Разведка (интеграция информационных операций в разведывательную деятельность)
(10) Military Deception	Военный обман
(11) Operations Security	Безопасность операций
(12) Special Technical Operations (STO).	Специальные технические операции
(13) Joint Electromagnetic Spectrum Operations	Объединённые операции в электромагнитном диапазоне
(14) Key Leader Engagement (KLE)	Ключевые обязательства

Серьезные материалы, посвященные различным темам информационных и психологических операций, постоянно публикуются в ведущих военно-научных журналах ВС США Military Review, Parameters, Special Warfare и других.

В 2010 г. было принято решение заменить термин психологические операции, использовавшийся ранее, термином информационные операции военной поддержки. Это связано с тем, что новая терминология в большей степени отражает деятельность Министерства обороны в сфере информирования и влияния на вражеское, нейтральное и дружественное общественное мнение, направленное на достижение стратегических целей американского военного командования. В распространенном меморандуме отмечается, что изменение терминологии не сильно повлияет на проводимые, на тот момент операции. Военные подразделения, обеспечивающие психологические операции продолжают усиливать возможность «убеждать-изменять-влиять» в операциях любого вида, в любом месте в любое время [Memorandum, 2017].

Термин «информационные операции» введен специалистами оборонных ведомств США в полевых уставах FM 100-23 «Миротворческие операции» и FM 33-5 «Ведение психологической войны».

В указанных инструктивных документах одним из принципиальных подходов к определению сущности понятия «информационная операция» является возможность использования методов и средств этой формы информационного противоборства не только во время боевых действий, но и в мирное время.

В Доктрине Воздушных сил США 2-5 «Информационные операции» (Air Force Doctrine Document 2-5 «Information operations») от 2005 информационные операции определяются как меры, применяемые для воздействия на вражескую информацию или информационную систему при одновременной защите собственной информации и информационных систем.

Согласно Руководству Вооруженных сил США 3-13 «Информационные операции» (Joint publication 3-13 «Information operations») от 2006 информационные операции - это интегрированное применение ключевых возможностей электромагнитных средств, компьютерных сетей, психологических операций, военного искусства и безопасности операций вместе со специальной поддержкой и соответствующими возможностями с целью влияния,

разрушения, нанесения ущерба, захвата процесса принятия решений (человеком или техническими средствами) при одновременной защите собственного. Это же наставление от 2012 содержит уже более общее определение: «Информационные операции – это интегрированное применение во время военных операций возможностей, касающихся информации, вместе с другими средствами операции, с целью воздействия, разрушение, причинение вреда, захвата процесса принятия решения враждебной стороной или потенциальным врагом при одновременной защите собственного». Политика стратегических коммуникаций НАТО 2009 года (NATO Strategic Communications Policy) определяет «информационные операции» как военные рекомендации и координацию информационных военных мероприятий НАТО с целью создания желаемого воздействия на намерения, понимание (представление) и способность вражеской стороны и других субъектов (потенциального врага, лиц, принимающих решения, культурных авторитетов, представителей международного сообщества и т.д.) в поддержку операций, миссий и целей Альянса.

В 2010 году Объединенный комитет начальников штабов ВС США опубликовал доктрину информационных операций военной поддержки, в которой говорилось, что такие операции являются важнейшим элементом американской внешней политики. В ходе конфликта такие операции усиливают эффективность боевой мощи (*force multiplier*), позволяют понизить эффективность вооруженных сил противника, ограничить вмешательство гражданских лиц, сократить сопутствующий ущерб, увеличить поддержку проводимой операции со стороны местного населения [Military, 2011].

В России 14 января 2014 года министр обороны подписал приказ о создании в составе Генерального штаба ВС России кибернетического командования, основная задача которого заключается в защите от несанкционированного вмешательства в электронные системы управления России. 22 февраля 2017 года, выступая в Госдуме, он подтвердил, что в Вооружённых силах России созданы войска информационных операций.

В Украине информационные операции до сих пор не определены официальными доктринами и концепциями, однако их возможное проведение базируется на ряде основополагающих документов:

Закон Украины «О разведывательных органах Украины» от 22 марта 2001 года №2331-III к основным задачам разведывательных органов относит «осуществление специальных мероприятий, направленных на поддержку национальных интересов и государственной политики Украины в экономической, политической, военной, военно-технической, экологической и **информационной сферах** ...».

Дополнительно в Законе Украины «О Службе внешней разведки Украины» от 01.12.2005 года № 3160-IV в статье 3 отдельно определены задачи Службы внешней разведки Украины, в том числе по «осуществлению специальных мер воздействия, направленных на поддержку национальных интересов и государственной политики Украины в экономической, политической, военно-технической, экологической и **информационной сферах**...».

Законы Украины «О Службе безопасности Украины» от 25.03.1992 года № 2229-XII и «О борьбе с терроризмом» от 20.03.2003 года № 638-IV возлагают на Службу безопасности Украины целый ряд задач по обеспечению **информационной безопасности государства**, особенно в условиях проведения антитеррористической операции».

В Законе Украины «О Вооруженных Силах Украины» (в связи с принятием Закона Украины «О внесении изменений в некоторые законы Украины относительно Сил специальных операций Вооруженных Сил Украины» от 07.06.2016 года № 1437-VIII) части четвертая и пятая статьи 1 изложены в следующей редакции: «Соединения, воинские части и подразделения Вооруженных Сил Украины в соответствии с законом могут привлекаться к осуществлению мер правового режима военного и чрезвычайного положения, организации и поддержанию действий движения сопротивления, проведению военных **информационно-психологических операций**...».

2. OSINT – разведка по открытым источникам

2.1. OSINT как дисциплина разведки

В качестве одного из важнейших инструментов проведения информационных операций выступает так называемая «разведка по открытым источникам» (Open source intelligence, OSINT) — одно из направлений разведки, которое включает в себя поиск, выбор и сбор разведывательной информации, полученной из общедоступных источников, и анализ этой информации.

Концепция OSINT базируется на двух основных понятиях:

- открытый источник – это источник информации, который предоставляет её без требования сохранения её конфиденциальности, т.е. предоставляет информацию, не защищенную от публичного раскрытия. Открытые источники относятся к среде общедоступной информации, и не имеют ограничения в доступе для физических лиц;
- общедоступная информация – это информация, опубликованная или размещенная для широкого использования; доступная для общественности.

По утверждениям аналитика ЦРУ Шермана Кента 1947 года, политики получают из открытых источников до 80 процентов информации, необходимой им для принятия решений в мирное время. Позднее генерал-лейтенант Самуэль Уилсон, который был руководителем Разведывательного управления Министерства обороны США в 1976—1977 годах, отмечал, что «90 процентов разведанных приходит из открытых источников и только 10 — за счёт работы агентуры».

OSINT обычно осуществляется с помощью мониторинга, анализа и исследования информации из сети Интернет. Материалы, составленные на основе информации из открытых источников, поддерживают все методы ведения разведки и разведывательные мероприятия путем накопления разведывательных знаний, анализа и их распространения.

В соответствии с [АТР, 2012] разведка в открытых источниках OSINT является также одним из способов ведения разведки, вносит значительный вклад при планировании боевых действий, а также предоставляет всю необходимую информацию при их проведении. Также определяется:

1) Разведка в открытых источниках (OSINT) является одним из методов ведения разведки путем сбора информации из открытых источников, ее анализа, подготовки и своевременного предоставления конечного продукта вышестоящему руководству в целях решения определенных разведывательных задач.

2) OSINT является методом ведения разведки, разработанным на основе сбора и анализа общедоступной информации, и не находящимся под непосредственным контролем правительства США. OSINT является результатом систематизированного сбора, обработки и анализа необходимой общедоступной информации.

Американский исследователь по вопросам безопасности Марк М. Ловенталь определяет открытую информацию как «любую информацию, которая может быть получена из открытых коллекций: все типы СМИ, правительственные отчеты и другие документы, научные исследования и отчеты, коммерческие поставщики информации, Интернет и т. д. Основная характеристика открытой информации – это то, что для ее получения не требуются нелегальных методов сбора и что она может быть получена с помощью средств, полностью соответствующих авторским правам и коммерческим условиям поставщиков.

Мировое сообщество все больше использует информацию из открытых источников в целях решения широкого спектра задач. В частности, роль OSINT при проведении информационных операции определяется рядом аспектов, среди которых оперативность поступления, объем, качество, ясность, легкость дальнейшего использования, стоимость получения и т.д. Следующие факторы влияют на процесс планирования и подготовки ведения OSINT:

- Эффективное информационное обеспечение. Большая часть необходимых справочных материалов об

объектах информационных операций добывается из открытых источников. Эта основа достигается путем сбора информации из СМИ. Накопление данных из открытых источников является основной функцией OSINT.

- Релевантность. Доступность, глубина и масштабы публично доступной информации позволяют находить необходимую информацию без привлечения специализированных человеческих и технических средств разведки.
- Упрощение процессов сбора данных. OSINT предоставляет необходимую информацию, исключая потребность в привлечении излишних технических и человеческих методов ведения разведки.
- Глубина анализа данных. Являясь частью разведывательного процесса, OSINT позволяет руководству осуществлять глубокий анализ общедоступной информации в целях принятия соответствующих решений.
- Оперативность. Резкое сокращение времени доступа к информации в сети Интернет. Сокращение человеко-часов, связанных с поиском информации, людей и их взаимоотношений на основе открытых источников. Быстрое получение ценной оперативной информации. Стремительно меняющаяся обстановка во время кризисов полнее всего отражается в текущих новостях, поэтому (как достоверно известно) за падением Берлинской стены и в Вашингтоне, и в штаб-квартире ЦРУ в Лэнгли следили не по сводкам разведслужб, а прикинув к телеэкранам и смотря репортажи CNN с места событий.
- Объем. Возможность массового мониторинга определенных источников информации, с целью поиска интересующего контента, людей и событий. Как показывает опыт, грамотно собранные фрагменты информации из открытых источников в совокупности могут быть эквивалентны или даже более значимы, чем профессиональные разведывательные отчеты.
- Качество. По сравнению с отчетами специальных агентов информация из открытых источников оказывается предпочтительнее уже потому, что лишена субъективизма, не разбавлена ложью.

- Ясность. Так что если в случае использования OSINT надежность открытых источников бывает как ясной, так и неясной, то в случае с тайно добытыми данными степень их надежности всегда вызывает сомнения.
- Легкость использования. Любые тайны принято окружать барьерами из грифов секретности, особых режимов доступа. Что же касается данных OSINT, то их можно легко передавать в любые заинтересованные инстанции. Возможно проведение комплексного расследования на основании данных из Интернета
- Стоимость. Стоимость добычи данных в OSINT минимальна, определяется лишь стоимостью используемого сервиса.

В частности, сегодня, предлагаемые для OSINT программно-технологические решения обеспечивают:

- сбор данных из социальных сетей, таких как Facebook, Twitter или Youtube, анализ собранных данных;
- нахождение в собранном контенте сути событий;
- агрегирование информации полученной из сети Интернет;
- информационное влияние в сети Интернет;
- оценку достоверности информации;
- мониторинг и распознавание идентичности в сети Интернет, в том числе с помощью геолокации;
- работу с информацией, полученной из невидимого веб-пространства (dark web, hidden web, deep web)

2.1.1 Области применения OSINT

Существует множество применений OSINT, среди которых можно назвать:

Разведка

Открытые источники содержат огромное количество информации, необходимой и соответствующей требованиям разведывательных органов, обеспечивающей понимание объективных и субъективных факторов, связанных, например, с проведением информационных операций. При этом, безусловно, для повышения эффективности разведывательной деятельности используется открытая информация в комплексе с агентурными ресурсами.

Инициатива Американского сообщества разведки по открытым источникам (известная как National Open Source Enterprise) выражена Директивой сообщества разведки 301, которая обнародована директором Национальной разведки [DNI, 2006]. Директива устанавливает полномочия и обязанности помощника заместителя директора Национальной разведки по открытым источникам (ADDNI / OS), Центра открытых источников DNI и Национального комитета по работе с открытыми источниками.

OSINT в вооруженных силах

Ниже приведены подразделения вооруженных сил США, которые участвуют в деятельности OSINT:

- Unified Combatant Command
- Defense Intelligence Agency
- National Geospatial-Intelligence Agency
- US Army Foreign Military Studies Office
- EUCOM JAC Molesworth
- Foreign Media Monitoring in Support of Information Operations, U.S. Strategic Command

Национальная безопасность

Департамент внутренней безопасности включает активное разведывательное подразделение для работы с открытыми источниками. 14 февраля 2007 г. был учрежден «Domestic Open Source Enterprise» для поддержки департамента OSINT и работы с государственными, местными и племенными партнерами.

Юстиция

Сообщество правоохранительных органов OSINT применяет разведку по открытым источникам для прогнозирования, предотвращения, расследования преступлений и преследования преступников, включая террористов. Кроме того, центры фьюжн вокруг США все чаще используют OSINT для поддержки их разведки и расследований.

Примерами успешных правоохранительных органов OSINT являются Scotland Yard OSINT; Королевская канадская конная полиция (RCMP) OSINT.

Полицейский отдел Нью-Йорка (NYPD) включает подразделение OSINT как отдел шерифа округа Лос-

Анджелес, расположенный в Бюро по чрезвычайным операциям и связанный с Объединенным региональным разведывательным центром Лос-Анджелеса.

В плане правоохранительной деятельности OSINT может применяться при борьбе с такими явлениями, как:

- Организованная преступность и банды
- Педофилия
- Кража персональных данных и вымогательство
- Отмывание денег
- Преступность в сфере нарушения интеллектуальной собственности
- Деятельность экстремистских организаций

При этом с помощью OSINT обеспечивается выявление вовлеченности и усиление влияния в Интернете:

- Идентификация ключевых фигур и активистов
- Мониторинг сети противника в режиме реального времени
- Ограничение распространения информации
- Формирование общественного мнения
- Идентификация и разработка экстремистских организаций
- Риски для общественного транспорта
- Санкции и правовые требования
- Анализ баз данных противников (HME, IED, TTPs)
- Геолокация целей
- Поддержка для военных операций

Кибернетическая безопасность

В рамках OSINT обеспечивается поддержка процессов обеспечения кибербезопасности, в частности, могут быть даны ответы на такие вопросы из области защиты телекоммуникационных сетей посредством получения информации:

- Кто атакует вашу организацию?
- Каковы их мотивы?
- Как они организованы?
- Какие инструменты используют?

Бизнес

OSINT в бизнесе включает в себя коммерческую разведку, интеллектуальную разведку и бизнес-аналитику, и часто является основной областью практики частных разведывательных агентств.

Предприятия могут использовать информационных брокеров и частных следователей для сбора и анализа соответствующей информации для деловых целей, которые могут включать в себя средства массовой информации, глубокую сеть, веб-2.0 и коммерческий контент.

2.1.2. Международный опыт

Ведение разведки в открытых источниках повышает эффективность деятельности всего разведывательного сообщества США, начиная национальным и заканчивая тактическим уровнями. Ниже приводится список некоторых организаций, которые занимаются в США сбором, приобретением, использованием, анализом, распространением информации из открытых источников.

- Совет по защите открытых источников (DOISC);
- Командование разведки и безопасности ВС США (INSCOM);
- Служба разведывательной информации Департамента сухопутных войск (DA IIS).
- Директор национальной разведки центра открытых источников (DNI OSC).
- Академия открытых источников;
- Департамент передовых систем (ASD).
- ФБР.
- Федеральный научно-исследовательский отдел (FRD), библиотека Конгресса.

Наряду с широким применением OSINT в США, приведем еще примеры применения этой технологии в других странах.

Служба внешней разведки Германии, Федеральная разведывательная служба, также использует преимущества Open Source Intelligence в подразделениях Abteilung Gesamtlage/FIZ и Unterstützende Fachdienste (GU).

В Австралии экспертом по открытым источникам является Управление национальных оценок (Office of

National Assessments), которое является одной из разведывательных госструктур. В Великобритании существует информационная служба BBC Monitoring, сосредоточенная на сборе открыто доступной информации силами журналистов. Анализом же собранных в BBC данных занимаются подписчики этого сервиса, в том числе и сотрудники секретных британских спецслужб.

2.2. 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 распространению компьютерного вируса Petya в середине 2017 года. Для этого был составлен запрос «*вирус&Petya*», введенный через веб-интерфейс системы (рис. 2.1).

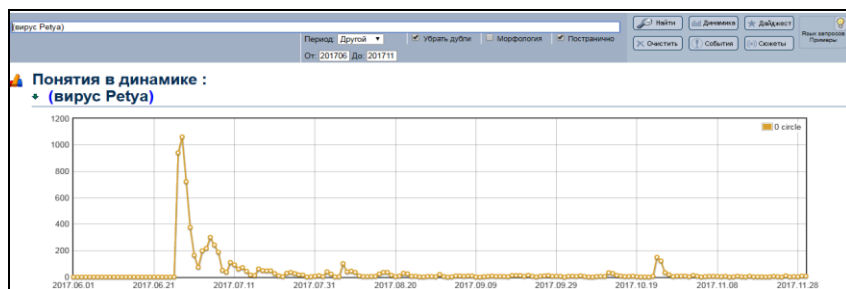


Рис. 2.1 – Диаграмма динамики понятия во времени

На этой диаграмме видно, что пик кризиса пришелся на конец июня 2017 года и заметна последующая повторная вспышка в конце октября предположительно клона этого вируса.

На примере проблем кибербезопасности рассмотрим, как из массивов текстовой информации из сети Интернет могут быть выявлены документы, содержащие максимальное количество материалов по компьютерным вирусам.

Для получения списка основных сюжетов, относящихся к рынку нефтепродуктов, был введен запрос «**вирус & кибер**», который уточнялся специальным признаком **lang.RUS & country.UA**, означающие в системе InfoStream поиск документов на русском языке в украинском сегменте Сети (Рис. 2.2). После этого достаточно перейти в режим просмотра и проанализировать документы, ссылки на которые выданы системой (рис. 2.3).

Кроме того, можно перейти в режим «Сюжеты», в котором предусмотрена кластеризация результатов поиска с учетом весовых критериев, что позволяет выдавать пользователю лишь наиболее весомые цепочки документов. Поэтому обеспечивается достаточно высокий уровень соответствия выдаваемых документов и потребности, выраженной запросом.

(*) Обзор основных сюжетов

В виде отчета (RTF)

(вирус кибер) & (русский язык) & Украина);
документов - 265, сюжетов - 45

США обвинили КНДР в распространении вируса WannaCry

[ТСН.ua](#) 2017.12.19 11:14
 Администрация президента Соединенных Штатов Америки Дональда Трампа обвинила Северную Корею в распространении компьютерного вируса WannaCry, который весной поразил тысячи систем в больницах, банках, госучреждениях во всем мире. Об этом сообщает Reuters. "Эта атака (кибератака - ред.) была широко распространена и стоила миллиарды, и Северная Корея несет непосредственную ответственность."
[Дубли - Похожие документы - Оригинал](#)
 Всего в сюжете сообщений: 158
 Первое сообщение: УНИАН, 2017.12.19 01:47
 Ключевые слова: [WANNACRY](#) [США](#) [ВИРУС](#) [КИБЕРАТАК](#) [КНДР](#) [БОССЕРП](#) [АТАК](#) [СЕВЕРН](#) [КОМПАН](#) [КОРЕ](#) [ПХЕНЬЯН](#) [ОБВИНЕН](#) [ПРЕЗИДЕНТ](#) [MICROSOFT](#) [МИР](#) [УКРАЇН](#) [ХАКЕР](#) [БЕЗОПАСНОСТ](#) [ОТВЕТСТВЕН](#) [КОР](#)

2017.12.23 13:04 [США назвали ответственных за кибератаку при помощи вируса-вымогателя WannaCry Минфин](#)

2017.12.22 13:14 [В США официально обвинили КНДР в атаке вируса WannaCry Народный Корреспондент](#)

2017.12.21 17:45 [Это провокация США: в КНДР ответили на обвинения в причастности к вирусу WannaCry УкронNews.com](#)

2017.12.21 17:19 [В КНДР отреагировали на обвинения в организации кибератак УСБН Одесса](#)

2017.12.21 17:16 [В КНДР отреагировали на обвинения в организации кибератак УСБН Вінниця](#)

2017.12.21 17:16 [В КНДР отреагировали на обвинения в организации кибератак УСБН Чернігіє](#)

2017.12.21 17:16 [В КНДР отреагировали на обвинения в организации кибератак УСБН Івано-Франківськ](#)

2017.12.21 17:16 [В КНДР отреагировали на обвинения в организации кибератак УСБН Житомир](#)

2017.12.21 17:14 [В КНДР отреагировали на обвинения в организации кибератак УСБН Запоріжжя](#)

2017.12.21 16:47 [Это провокация США: в КНДР ответили на обвинения в причастности к вирусу WannaCry УНИАН](#)

2017.12.21 16:02 [В КНДР отреагировали на обвинения в организации кибератак Data.ua](#)

Рис. 2.2 – Основная сюжетная цепочка по запросу

Активная база данных: Система интеграции интернет-ресурсов

Главная | Помощь | Кабинет | Источники | Статистика | Новости проекта

(вирус кибер)&(lang_RUS&&country_UA)

Период: 17 дней | Убрать дубли | Морфология | Постранично

Найти | Динамика | Даджест | Язык запросов | Примеры

Обзор основных сюжетов

(вирус кибер) & (русский язык) & (Украина); документов - 265, сюжетов - 45

В виде отчета (RDF) | Сеть подсетей (Эвэ) | Распечатать

1. США обвинили КНДР в распространении вируса WannaCry Администрация президента Соединенных Штатов Америки Дональда Трампа обвинила Северную Корею в распространении компьютерного вируса WannaCry, который весной поразил тысячи систем в больницах, банках, государственных во всем мире. Об этом сообщает Reuters. "Эта атака (кибератака – ред.) была широко распространена и стоила миллиарды, и Северная Корея несет непосредственную ответственность." Сюжет полностью (158)	2017.12.19 01:47	Российские хакеры готовят новую масштабную кибератаку на энергосеть Украины - The Atlantic утинан	158
2. Передается через Facebook: по Украине распространился новый компьютерный вирус Передается через Facebook: по Украине распространился новый компьютерный вирус. Однако есть рекомендации, которые помогут сохранить свой компьютер. На этой неделе по Украине и по другим странам распространился опасный компьютерный вирус, заключающийся в распространении киберреструктуризации кавера для криптовалюты Monero. Сюжет полностью (11)	2017.12.22 19:44	Передается через Facebook: по Украине распространился новый компьютерный вирус ТСН.ua	11
3. Российские хакеры несколько лет преследовали 200 журналистов, в том числе из Украины Журналисты оказались третьей по численности группой, которую атаковали	2017.12.19 23:02	Эксперты США бьют тревогу из-за необычной активности кремлевских хакеров: на электросети Украины готовится масштабная кибератака Dialog.ua	7

Рис. 2.3 – Фрагмент цепочки основных сюжетов

2.3. Источники информации

К открытым источникам информации относятся такие традиционные СМИ, как газеты, журналы, радио, телевидение, публичные отчеты правительств, профессиональные и академические отчеты, официальные данные о бюджетах, демографии, материалы пресс-конференций, различные публичные заявления, результаты наблюдений – радиомониторинг, использование общедоступных данных дистанционного зондирования земли и аэрофотосъемок (например, Google Earth), конференции, доклады, статьи. Кроме того, к открытым источникам относятся современные интернет-ресурсы, в частности веб-сообщества и контент, созданные пользователями (enduser generated content) – социальные сети, видеохостинги, вики-справочники, блоги, веб-форумы.

При проведении исследований с помощью OSINT обеспечивается доступ к огромным объемам информации. Правильное целеуказание и понимание того, где следует искать необходимую информацию, является важным фактором эффективной работы экспертов-аналитиков.

Различают два основных вида источников информации – основные и дополнительные.

Основным источником является документ или физический объект, содержащий информацию, которая была написана или создана в результате исследований и анализа. Данные источники являются непосредственными свидетелями того или иного события и содержат «инсайдерскую информацию». Как правило, такая информация является фрагментарной, неоднозначной и сложной для анализа. Основные источники упоминаются в качестве первоисточников при последующем анализе. Среди таких источников можно назвать:

- оригиналы документов, выписки из них, переводы, научные журналы, выступления, письма, интервью, новостные и видео кадры, официальные отчеты и др.;
- творческие работы: поэзия, драматургия, романы, музыка и др.;
- артефакты: керамика, мебель, одежда, исторические здания и пр.;
- рассказы и воспоминания людей.

Дополнительные или вторичные источники, включают правительственные пресс-службы, новостные коммерческие организации, информацию от пресс-секретарей и неправительственных организаций.

В процессе ведения разведки в открытых источниках необходимо учитывать присутствие дезинформации. Сотрудники подразделения, ведущего разведку в открытых источниках, обычно не получают информацию путем прямого наблюдения. Они опираются в основном на вторичные источники, которые могут преднамеренно или непреднамеренно добавлять, удалять, изменять или иным образом фильтровать информацию, поступающую для широкой общественности. Важно знать историю открытых источников и определить цель опубликования информации для того, чтобы распознавать фактическую информацию, выявить предвзятость, попытки оказания влияния на людей.

Интернет представляет собой динамичную информационную среду, состоящую из неподвижных и движущихся виртуальных и специальных веб-сайтов,

содержащих смесь старой и новой информации. Существует множество виртуальных баз данных, которые увеличиваются с геометрической прогрессией.

2.3.1. Веб-пространство

Веб-ресурсы индексируются стандартными информационно-поисковыми системами, которые, как правило, неспособны получить доступ или заиндексировать всю информацию, необходимую для эффективного проведения разведки в открытых источниках. Считается, что при ведении разведки в открытых источниках из всего объема информации, находящейся в сети Интернет, можно увидеть лишь ее четвертую часть. Три четверти содержатся в так называемом Deep Web.

Поисковые системы являются основными инструментами, которые используют сотрудники OSINT при проведении разведки и сбора общедоступной информации из открытых источников. Сотрудники OSINT активно используют поисковые системы и различные критерии поиска, чтобы найти текст, изображение и информацию на страницах тысяч веб-сайтов. Технически, поисковые системы осуществляют поиск по индексу веб-сайтов. Коммерческие и государственные поисковые системы отличаются друг от друга в зависимости от параметров поиска, осуществления процесса поиска и способов предоставления результатов поиска. Большинство поисковых систем используют так называемые программы-пауки, которые сканируют веб-ресурсы и формируют индексированные базы данных поисковых систем.

Иногда в результате поиска находится ссылка на документ, но при попытке открыть его, возникает ошибка доступа. В таких случаях, сотрудники OSINT могут найти необходимое содержание на таких сайтах, как www.archive.org.

Веб-пространство, основанное на физической инфраструктуре сети Интернет и протоколе передачи данных HTTP, объединяет сотни миллионов веб-серверов, подключенных к сети Интернет (рис. 2.4). 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 вебу 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), в настоящее время количество веб-серверов превышает 1 млрд (Рис. 2.4).

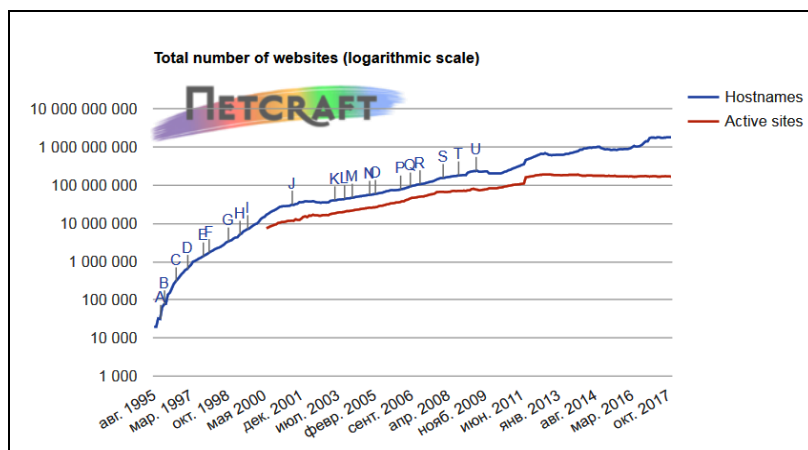


Рис. 2.4. Динамика роста количества веб-серверов в логарифмической шкале (Netcraft, октябрь 2017 года)

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

Возможности доступа к интернет-ресурсам, которые привлекают 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) – 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. 2.5).

Avalanche (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. 2.5 – Фрагмент веб-сайта программы Copernic Agent

Newprosoft Web Content Extractor (www.newprosoft.com) – программа сканирования и извлечения данных из веб-сайтов.

Portable Offline Browser от MetaProducts Corporation (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) – программная надстройка, объединяющая результаты информационно-поисковых систем Google и Яндекс, которая позволяет осуществлять конкурентный анализ, базирующийся на оценке эффективности сайтов и контекстной рекламы.

WebSite-Watcher (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) – бесплатный сервис, позволяющий автоматически собирать новую информацию с веб-ресурсов, поставленных на мониторинг.

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

Newspaper Map (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) – 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 (aggregator.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) – 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. 2.6).

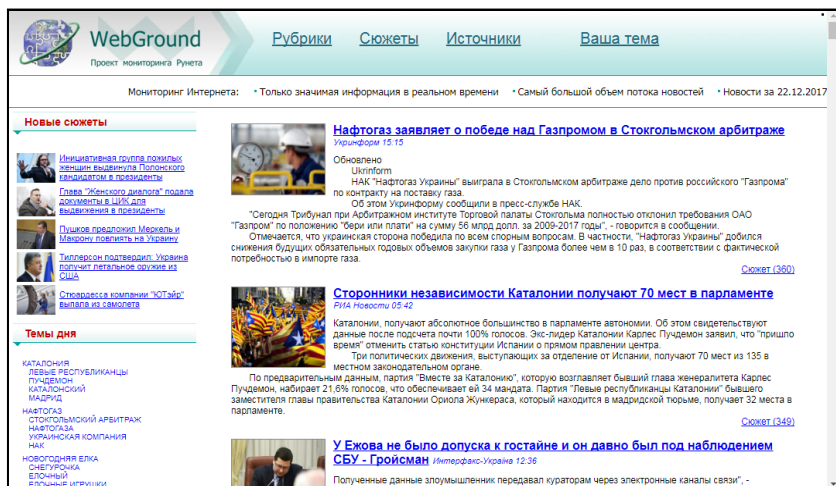


Рис. 2.6 – Фрагмент агрегатора новостей WebGround

2.3.2. 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.

Deep Web является частью веб-пространства и содержит, по некоторым оценкам в два раза больше информации, чем в остальной его части. В отличие от обычных веб-ресурсов, Deep Web не индексируется обычными поисковыми системами. Веб-сайты и базы данных, расположенные на нем не доступны при осуществлении поиска через традиционные поисковые системы. Deep Web предлагает огромное количество ресурсов следующего содержания:

- динамического. Веб-страницы, появляющиеся при осуществлении поиска;

- ответ на представленный запрос;
- несвязанного. Веб-страницы, которые не связаны с другими веб-страницами и блокирующие доступ к своему содержанию при использовании стандартных программ поиска информации;
- скрытого. Веб-страницы без обратных и входящих ссылок;
- ограниченного доступа. Веб-сайты, которые ограничивают доступ к информации;
- не содержащие гипертекстовых документов;
- текстовое содержание, закодированное в формате мультимедийных файлов, которые не обнаруживаются поисковыми системами.

Совокупность источников в веб-пространстве, недоступных пользователям традиционных поисковых систем, образует так называемый «глубинный веб» – понятие, введенное Дж. 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 путать 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) 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. 2.7). По мнению разных авторов к видимому веб относится лишь 20-30% веб-пространства.

The screenshot shows the WIPO IP Services website interface. At the top, the WIPO logo and 'IP SERVICES' are displayed. Below this is a navigation bar with links for 'ABOUT WIPO', 'IP SERVICES', 'PROGRAM ACTIVITIES', 'RESOURCES', and 'NEWS & EVENTS'. A breadcrumb trail reads 'Home > IP Services > PATENTSCOPE® > Patent Search'.

On the left side, there is a 'PATENTSCOPE®' menu with various options like 'About Patents', 'PCT Resources', 'Database Search', etc. The main content area features a notice about a system transition and a search result for patent WO/2006/108069, titled 'SEARCHING THROUGH CONTENT WHICH IS ACCESSIBLE THROUGH WEB-BASED FORMS'. The result includes bibliographic data, a list of applicants (GOOGLE, INC., HALEVY, Alon Y., MADHAVAN, Jayant, KO, David H.), inventors (HALEVY, Alon Y., MADHAVAN, Jayant, KO, David H.), an agent (PARK, A. Richard), and the priority date (06.04.2005 US).

Рис. 2.7 – Фрагмент веб-ресурса WIPO с описанием патента Google на поиск в глубинном веб

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. Общепризнано, 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 либо удаляют сами ресурсы, либо удаляют гиперссылки, переводя ресурсы в глубинный веб. Например, недавно бизнес-каталоги Auto.ru и Drom.ru отказались отдавать свои объявления «Яндексу», т.е. защищая свои информационные активы, компании перевели свои ресурсы в глубинный веб.

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 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 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.

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/nts/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) 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. Метапоисковый пакет DeepQueryManager (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); криминалогические 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» (Рис. 2.8).

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) – сервис, обеспечивающий оперативный поиск в 5000 отобранных веб-сайтах политической направленности, предоставление новостей из нескольких десятков авторитетных источников.

Infomine (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.

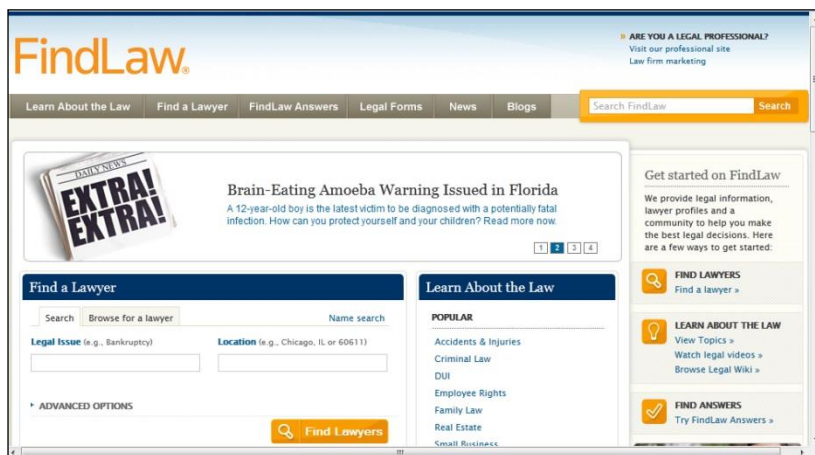


Рис. 2.8 – Фрагмент веб-сайта сервиса FindLaw

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 содержаться в ней. Типичный пример – информационно-поисковые системы по украинскому (zakon.rada.gov.ua) или российскому законодательству (www.kodeks.ru). Тысячи документов из баз данных становятся доступны только после входа в систему, а роботы стандартных поисковых систем не в состоянии проиндексировать 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). Сегодня объем базы данных Alexa превышает 350 млрд. веб-страниц (рис. 2.9).

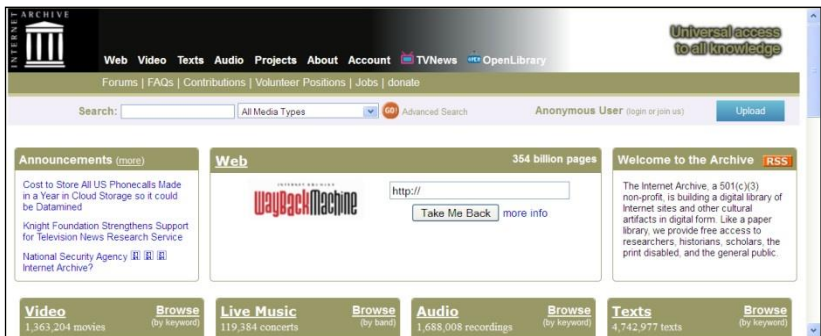


Рис. 2.9 – Заглавная страница веб-сайта www.archive.org

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.

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

Российская компания «Р-Техно» создала систему «it2b.интернетошпионаж 3000+», предназначенную для выгрузки данных из невидимого сегмента сети Интернет. На основе этой системы построен поисковый сервис Web Insight (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 приводится информация о реальном количестве документов на исследуемом веб-сайте, представленном в RUNet, и о количестве документов, проиндексированных различными поисковыми системами, в том числе, Google и Яндекс. Получив репрезентативную выборку по сайтам, например, по рейтингу Рамблера top100 (top100.rambler.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.

2.3.3. Социальные медиа

Социальные медиа представляют собой 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 основываются на идеологической и технологической базе веб 2.0, позволяющей 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 базируются 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, его основное содержание – 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.

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 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 присущий социальные связи. Сегодня термин «социальная сеть» обозначает понятие, оказавшееся шире своего социального аспекта, оно включает, например, многие

информационные сети, в том числе и веб. Рассматривают не только статистические, но и динамические сети, для понимания структуры которых необходим учет принципов их эволюции.

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.

Основные социальные медиа

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) – 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 миллиардов «дружеских связей». Ежемесячное количество просмотров страниц *Facebook* превышает 1 триллион.

Twitter (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.

Sina Weibo (кит. 新浪微博) (weibo.com) – китайский сервис микроблогов, запущенный компанией Sina Corp 14 августа 2009 года. Сервис является самым популярным в Китае и используется более чем 30% пользователей Интернета, с проникновением на рынок, похожим на Twitter в США. По состоянию на 3 квартал 2017 года в Sina Weibo насчитывается более 376 млн. активных пользователей в месяц. Ежедневно на Сину Вайбо отправляется около 100 миллионов сообщений. 8 сентября 2017 г. Weibo потребовал от своих пользователей до 15 сентября согласовать свои учетные записи с их настоящими именами.

LinkedIn (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; позволяет рекомендовать пользователей (позволяет им быть рекомендованными), публиковать вакансии, создавать группы по интересам. Социальная сеть LinkedIn также позволяет публиковать информацию о деловых поездках и конференциях.

«ВКонтакте» (vk.com) – крупнейшая в России социальная сеть, крупнейшая в Европе, созданная в 2006 г., позиционирующая себя как «современный, быстрый и эстетичный способ общения в сети». По данным на август 2017 среднесуточная аудитория составляет более 80 миллионов посетителей, зарегистрировано более 410 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 24 мая 2013 г. Роскомнадзор внёс домен vk.com и его IP-адрес в «Единый реестр запрещённых сайтов», однако уже через несколько часов удалил его оттуда, обосновав ошибку человеческим фактором. 16 мая 2017 года Президент Украины подписал указ, которым ввел в действие решение СНБО об обновлении списка санкций против ряда российских компаний, среди которых и соцсеть «ВКонтакте». Доступ к ним в Украине полностью ограничен с 1 июня 2017 г.

«Одноклассники» (OK.ru) – российская социальная сеть, принадлежащая Mail.Ru Group. Проект запущен 4 марта 2006 года. На 1 января 2016 года в сети зарегистрировано более 290 млн. пользователей. Посещаемость сайта – более 42 млн. посетителей в сутки (на август 2017). 16 мая 2017 года Президент Украины подписал указ, которым ввёл в действие решение СНБО об обновлении списка санкций против ряда российских компаний, среди которых соцсеть «Одноклассники». Ограничения к сайту «Одноклассники» вступили в действие в Украине 1 июня 2017 г.

Google+ (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 компонент: *Круги*, *Темы*, *ВидеоВстречи*, *Мобильная версия*. Основополагающими 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 своё 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 уникальных функции: мгновенная загрузка фото и Чат (*Huddle*). Социальная сеть Google+ позволяет получать хорошие позиции в поиске Google.

«Мой Круг» (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 и rabota.mail.ru. В 2011 году социальная сеть «Мой круг» завершила процесс интеграции с социальными платформами – появилась возможность поиска друзей через аккаунты Facebook, Twitter, LiveJournal и LinkedIn.

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 возможности их комментирования; указание расширенных сведений о пользователе; друзья и лента друзей; картинки пользователей; функции безопасности аккаунта.

Мониторинг социальных медиа

Мониторинг социальных медиа – важнейший инструмент OSINT. 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 партнером Twitter. Обеспечивает постинг (posting) по расписанию, возможность отслеживать сообщения по ключевым словам и упоминаниям. HootSuite также предоставляет полноценную интеграцию с Facebook. Система HootSuite является условно-платной, есть бесплатная версия (аналитика, 5 социальных профилей, 2 RSS/Atom ленты). Она 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 аналитическом

интерфейсе с функциями одновременной работы нескольких сотрудников. Она предоставляет 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) – сервис компании Google, позволяющий в режиме реального времени анализировать данные по темам, тенденциям, событиям, имеющим отношение к личности или бизнесу.

Topsy (topsy.com) – бесплатный сервис поиска в режиме реального времени по социальным медиа.

HowSociable (www.howsociable.com) – бесплатный инструмент мониторинга брендов и ключевых слов в 32 социальных сетях.

Twitalyzer (www.twitalyzer.com) – аналитическая программа-клиент для Твиттера, позволяющая отслеживать количество переходов, анализировать позитивные и негативные комментарии, сегментировать аудиторию. Она интегрирована с системой Google Analytics, выводит интерактивные диаграммы и графические инструменты.

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

Kurrently (www.kurrently.com) – бесплатная поисковая система по социальным сетям Twitter и Facebook, позволяющая отслеживать, распространять целевую информацию по социальным сетям.

Trackur (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) – 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) – 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 для оценки успешности проведенных кампаний в сети Интернет.

SemanticForce (www.semanticforce.net) – сервис, обеспечивающий мониторинг неструктурированных источников – комментариев в сетевых СМИ и интернет-магазинах (рис. 2.10). Выдает более 20 видов аналитических отчетов. Сервис SemanticForce интегрирован с внешними системами: Klout, Copiny, GoogleAnalytics.

Brandspotter (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) – 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 фильтрует сообщения, в которых бренд упоминается лишь вскользь. Данные начинают отображаться в системе через 2–4 часа после их публикации.

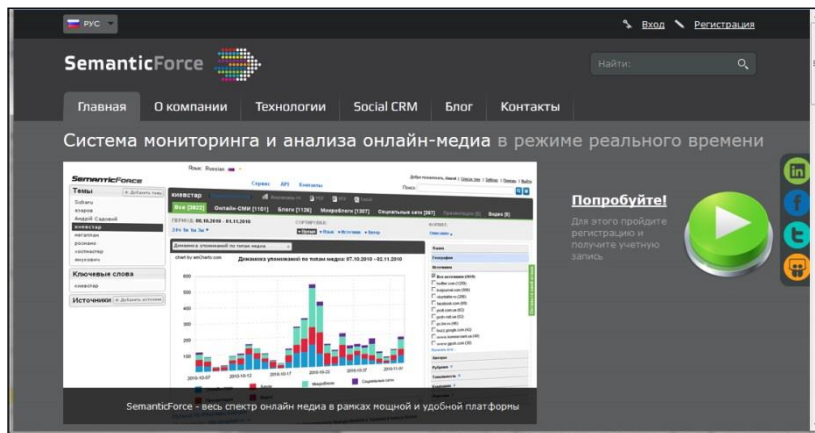


Рис. 2.10 – Фрагмент веб-сайта службы SemanticForce

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

TweetDeck (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.

2.3.4. Базы данных открытых источников

Для ведения OSINT обеспечивается доступ к базам данных и другим ресурсам сети Интернет в целях облегчения, эффективной обработки, хранения, поиска и обмена открытой информацией. Эти базы данных размещены в различных вычислительных сетях, как локальных, так и глобальных, веб-ресурсах, «Deep Web» и др. При этом в системах OSINT используется проанализированная информация из открытых источников, и сопровождаются следующие информационные базы данных:

- база данных с оперативной информацией, которая обеспечивает взаимосвязь между информационными потребностями и запросами, показывает статус процесса сбора материалов, обработки, и содержит статистические данные (графики, таблицы);
- база данных с технической информацией, которая поддерживает операции по сбору информации и состоит из необработанных текстовых, аудио- и видео материалов, переводов и записей стенограмм.

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), 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. 2.11). 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.

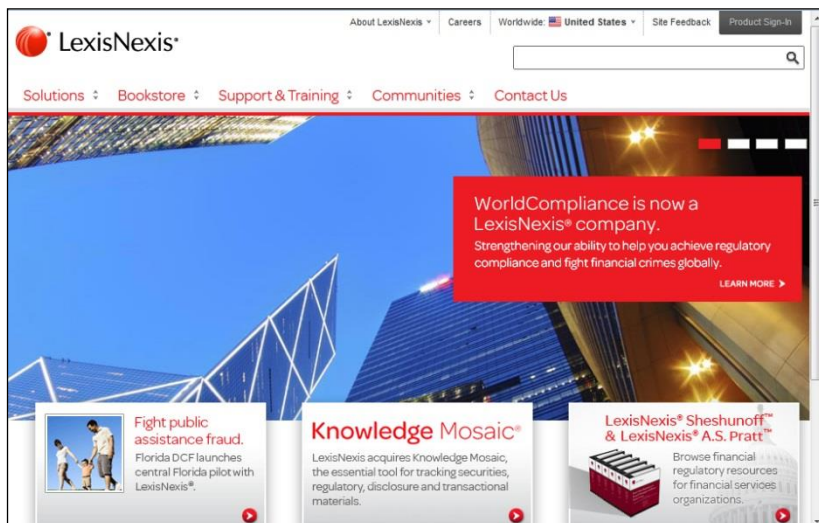


Рис. 2.11 – Фрагмент веб-сайта службы LexisNexis

Служба **Factiva** (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), 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 рынкам (рис. 2.12). Основные продукты ISI Emerging Markets: CEIC Data, Emerging Market Information Service (EMIS), Islamic Finance Information Service (IFIS), IntelliNews, ISI Compliance Edition, ISI DealWatch.



Рис. 2.12 – 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.

Украинская информационно-аналитическая платформа для бизнес аналитики, конкурентной разведки и проверки контрагентов YouControl (<https://youcontrol.com.ua/>) на конец 2017 года позволяла получить актуальную информацию о компании из 42 украинских государственных реестров и ежедневно узнавать об изменениях, происходящих в них. С помощью автоматического мониторинга YouControl следит за каждой компанией, которая добавляется в список, проверяет информацию по особо важным реестрам, собирает найденные изменения в единый документ. Сервис YouControl позволяет определить, как компания связана с другими, при этом учитывается, что у компаний может быть общий учредитель, директор, адрес и т. д. В интерфейсе пользователя можно пошагово в графическом режиме раскрывать корпоративные связи организации, досье которой анализируется на текущий момент.

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). 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) – содержит личные характеристики и сведения о способностях и компетентности граждан США.

В России и Украине популярны такие базы данных, как:

«Лабиринт» (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) – 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;

«**KAPE**» (kare.pulsцен.com.ua) – БД предприятий Украины – 384000 компаний, БД агропромышленного комплекса Украины – 218000 компаний;

Базаданных **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);

Базы данных международной корпорации **Creditreform**, представленной в России компанией «Кредитреформ РУС» (www.creditreform-rus.ru), а в Украине информационным агентством «Кредитреформ Украина» (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), содержащей данные о 30 млн. компаний.

Europages (www.europages.eu) – Европейская бизнес-директория – информационно-поисковая B2B-система, охватывающая свыше 2 млн. поставщиков, производителей и дистрибьютеров 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 исполнения – 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.

Существуют большое количество приложений коммерческих программ, инструменты и базы данных в сети Интернет с возможностями полнотекстового поиска. Приведем некоторые из них:

- Академия Google. Академия Google (scholar.google.com) предоставляет простые способы обширного поиска научной литературы. Данная система поиска может из одного места осуществлять поиск по всем существующим научным дисциплинам и источникам, таким, как статьи, монографии, учебники, диссертации, рефераты, препринты и др. Академия Google помогает пользователю найти необходимую для

него работу из всех доступных мировых научных исследований.

- Spokeo. Данный сервис специализируется на сборе информации о людях (их имена, адреса, номера телефонов) из телефонных книг, социальных сетей, маркетинговых списков, бизнес-сайтов и других открытых источников.
- Blog Pulse. Представляет собой автоматизированную систему определения тенденции в блогах путем применения компьютерного обучения и методов обработки естественного языка.
- Pipl. Система запросов помогает найти Deep Web, которые не могут быть найдены при поиске стандартными поисковыми системами.
- Monitter. Поисковая система, созданная на основе браузера Twitter. Monitter отображает в вашем браузере результаты поиска сразу в поисковых системах по ключевым словам.
- Maltego. Данная система является аналитическим приложением, которое предоставляет сбор и анализ данных, а также формирование смысловых групп из полученных сведений. Maltego позволяет определить существующие связи между информационными объектами, а также выявить ранее неизвестные.

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 «Р-Техно», «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.

2.4. Технологии OSINT

OSINT - это очень разнообразная форма сбора и анализа информации. При работе OSINT часто следует принимать меры предосторожности при сборе информации из Интернета. Это может быть сделано в форме использования VPN для анонимности и незаметного сбора информации. Оценка источников становится важной для общего процесса сбора и анализа OSINT. Аналитик OSINT нуждается в интеллектуальном анализе для выявления истинного или ложного процесса, который повлияет на прогнозирование будущего. Наконец, аналитики должны найти использование оценочного анализа для того, чтобы его результаты можно было включить в готовый классифицированный, неклассифицированный или запатентованный интеллектуальный продукт.

Сбор информации в OSINT, как правило, отличается от сбора данных в других разведывательных дисциплинах, где получение необработанной информации, подлежащей анализу, может быть основной трудностью. В OSINT основной трудностью является определение релевантных, надежных источников из огромного количества общедоступной информации. Однако это не так сложно для тех, кто знает, как получить доступ к местным знаниям и как использовать экспертов, которые могут создавать новые индивидуальные знания «на лету».

Этапы OSINT

Процесс OSINT состоит из четырех этапов: планирования, подготовки, сбора и производства конечного материала – аналитики и четырех основных процессов: анализа, накопления разведанных, оценки и распределения по направлениям. Процесс ведения разведки, также как и процессы подготовки ответных информационных операций (планирование, подготовка, выполнение и подведение итогов), пересекаются и повторяются в соответствии с требованиями практики.

Как указано в «Инструкции по ведению разведки в полевых условиях», OSINT повышает эффективность и оказывает поддержку процессу ведения разведки и других операций.

На Рис. 2.13 представлена схема процесса ведения OSINT.

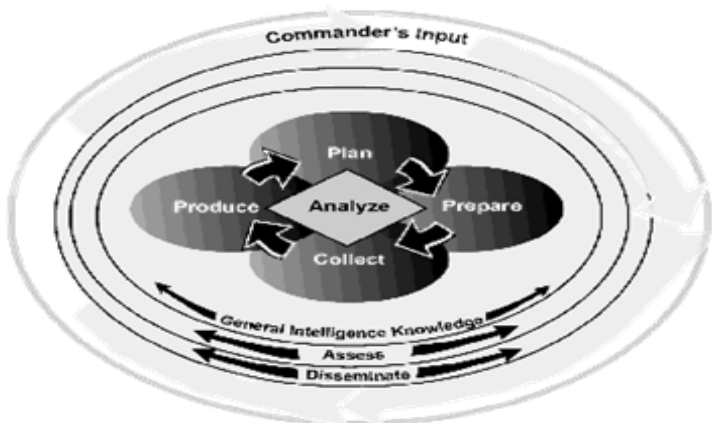


Рис. 2.13 – Сбор информации путем OSINT позволяет принимать правильные решения [АРТ, 2012]: План ⇒ Подготовка ⇒ Сбор ⇒ Производство. Общие знания разведки. Оценка. Распространение

Сбор разведанных синхронизирует и интегрирует процессы планирования, использования сил и средств, обработки и распределения элементов системы для

поддержки боевых операций, что является объединенной разведывательной и оперативной функцией.

После анализа, информация, полученная из различных источников, становится разведанными, которые содержат необходимую информацию о противнике, угрозах, климате, погодных условиях, рельефе местности и др.

Установлено, что такие элементы структуры OSINT, как постоянный поток информации, технические средства, программное обеспечение, безопасность средств коммуникации и базы данных охватывают средства:

- обеспечения доступности разведывательных данных. Обеспечение доступности разведанных является процессом, благодаря которому разведывательные организации активно и быстро получают доступ к разведанным;
- разработки и ведения автоматизированной сети разведки. Главной задачей является предоставление информационных систем, которые обеспечивают связь, совместный анализ и обработку, распространение материалов и создание условий доступности разведывательных данных;
- создания и поддержания доступа. Эта задача влечет за собой установление, обеспечение и поддержание доступа к секретным и несекретным программам, базам данных, сетям, системам и другим Интернет ресурсам для войск союзных государств, объединенных сил, национальных агентств и международных организаций;
- создания и ведения баз данных. Эта задача предполагает создание и поддержание несекретных и секретных баз данных. Создание и ведение базы данных способствует быстрому анализу, подготовке отчетов, обработке, распространению, ведению длительных боевых действий.

2.5. Правовые вопросы

Безусловно, OSINT 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 OSINT под запрет.

В Украине «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, в Украине правовое регулирование в информационной сфере, к которой, безусловно, относится и OSINT, основывается на 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 узаконенного понятия «OSINT» в Украине сегодня не существует, деятельность по сбору, хранению, обработке и распространению информации регулируется целым рядом законодательных и нормативных актов:

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

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

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

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 OSINT иногда может быть воспринято как проведение оперативно-розыскной деятельности, проводить которую, согласно Закону Украины «Об оперативно-розыскной деятельности» от 18.02.1992 г. № 2135-XII 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 утвержденной Указом Президента Украины №96/2016 от 27 января 2016 года Стратегии кибербезопасности Украины декларируются основные задачи силовым органам, среди которых: «на разведывательные органы Украины – реализация разведывательной деятельности по выявлению угроз национальной безопасности Украины в киберпространстве, других событий и обстоятельств, касающихся сферы кибербезопасности», а также предусматривается «создание системы своевременного выявления, противодействия и нейтрализации киберугроз, в том числе с привлечением волонтерских организаций», всё это, безусловно, относится к применению средств OSINT в этой области.

В то же время действующими нормативно-правовыми документами Украины предусмотрена уголовная 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 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. Большую часть сведений можно получить только путем консультаций с соответствующими экспертами.

При этом сегодня к деятельности компаний, занимающихся разведкой в открытых источниках, 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 OSINT.

Также, можно выделить три класса основных проблем авторского права, имеющих отношение к конкурентной разведке: проблемы, связанные с правомерностью использования входной информации (источников информации), на основании которой формируются отчеты – результаты OSINT; проблемы, связанные с авторскими правами на результаты OSINT; проблемы, связанные с правами на применение (использование) специализированного программного обеспечения, необходимого для проведения OSINT.

Кроме того, одна из проблем, стоящая перед службами OSINT в 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 OSINT.

Ситуация может измениться, если будет создана четкая правовая база для деятельности, связанной со сбором и аналитической обработкой информации и, в частности, для разведки по открытым источникам.

Важное значение для становления OSINT имеет ряд статей Закона Украины «О защите от недобросовестной конкуренции» № 236/96-ВР от 07.06.1996, где (ст. 15-1), запрещается «Неправомерный сбор коммерческой информации», «Разглашение коммерческой информации» «Неправомерное использование коммерческой информации» (гл 4, ст. 16, 17, 19, соответственно).

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 лиц в других компаниях и т.п.

Зачастую усилия бизнес-разведки направлены на получение коммерческой тайны конкурентов. Можно сказать, что деятельность в области OSINT иногда направлена на добычу информации, которая не является общедоступной и охраняется законом. Эти действия нарушают огромное количество статей Уголовного Кодекса Украины, в частности, статью 231 «Незаконный сбор с целью использования или использование сведений, составляющих коммерческую или банковскую тайну».

Таким образом, разведка по открытым источникам может легитимно использовать лишь те методы и способы сбора и обработки информации, которые не противоречат законодательству, т.е. основные функции OSINT — 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.

Гражданский кодекс Украины 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 является 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 лицом, которое 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»).

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, и нашей страной. 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, использование и распространение именно критичных данных без согласия субъекта данных.

Право на приватность гарантируется Конституцией Украины. Статья 32 Конституции Украины гласит: «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).

В соответствии с украинским законодательством персональными данными 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 украинском законодательстве 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 (как украинский, так и российский) имеют прямое отношение к сфере информационных технологий и телекоммуникаций, оба содержат спорные, противоречащие сложившейся практике, казались 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 тщательным 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-компании 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 OSINT приходится сталкиваться с многочисленными противоречиями и казусами в существующем законодательстве, например, в Законе Украины «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 OSINT должны выглядеть примерно так: «Персона А. провела переговоры с персоной Б.», «Флаг над рейхстагом подняли персоны I, II и III». 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 бизнес-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 их заявок, но сопутствующая цель любой уважающей себя организации – это наработка базы клиентов. И эта база часто имеет собственное коммерческое значение. Известны многочисленные случаи совершенно легальной перепродажи баз данных клиентов, например, при прекращении деятельности фирмы-владельца. Однако, например, в украинском законодательстве предусмотрены условия уничтожения персональных данных, среди которых (ст. 15), «прекращение правоотношений между субъектом персональных данных и владельцем или распорядителем базы...». А это означает, оператор, например, 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 OSINT 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 OSINT. Многие службы безопасности сегодня пользуются базами данных с информацией о лицах. 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.

В США основным правовым механизмом ведения разведки в открытых источниках министерства обороны является Совет по защите открытых источников (DOISC). Он служит форумом для координации и содействия мероприятиям и программам ведения разведки в открытых источниках для всех служб и боевых команд. Данный совет консультирует и докладывает заместителю министра обороны по разведке о вопросах ведения разведки в открытых источниках, о новых инициативах по улучшению эффективности работы подразделения OSINT и деятельности министерства обороны в целом. В обязанности совета входят:

- координирует деятельность подразделения OSINT и утверждает его план ведения разведки в открытых источниках;
- определяет последовательность требований к процессу ведения разведки в открытых источниках.

Армейский стандарт США «АТР 2-22.9» устанавливает общие понятия, основные концепции и методы сбора разведывательных данных из открытых источников для Армии США. В этом документе подчеркивается характеристика OSINT как разведывательной дисциплины, его связи с другими разведывательными дисциплинами, и возможности его применения в ходе объединенных операций.

Использование общедоступной информации является важным аспектом технической разведки (TECHINT). Несмотря на то, что намерения, возможности и факторы уязвимости противников и потенциальных угроз подлежат засекречиванию, результаты OSINT (в частности, открытого сервиса «Google Earth») способствуют получению информации о самых скрытных государствах и организациях. Такие примеры свидетельствуют об ответственности деятельности в этой области.

Авторское право является одной из форм защиты, опубликованных и неопубликованных работ, предусмотренных главой 17 Кодекса США, что определяет авторов «оригинальных работ авторов», в том числе литературных, драматических, музыкальных и художественных произведений.

Национальные законы об авторских правах являются ограничениями OSINT. Нарушение прав, в частности, предусмотренных главой 17 Кодекса США, законами об авторских правах, все же оставляют возможность правомерного использования OSINT, что определяется четырьмя факторами:

- целью и характером использования;
- свойствами, используемых авторских работ;
- количеством и частями авторской работы, которые используются;
- воздействием использования авторских работ на потенциальный рынок или ценность этих работ.

3. Элементы нелинейной динамики при распознавании информационных операций

В данной главе основное внимание уделено распознаванию информационных операций на основе изучения динамических свойств информационных потоков в глобальных компьютерных сетях, в частности, в сети Интернет.

Для исследования информационных потоков в Интернете, т.е. потока сообщений, которые публикуются на страницах веб-сайтов, в социальных сетях, блогах, и т.п., должен применяться современный инструментарий. Так известные методы обобщения информационных массивов (классификация, фазовое укрупнение, кластерный анализ и т.д.) уже не всегда пригодны даже для адекватного количественного отражения процессов, происходящих в информационном пространстве [Lande, 2007].

Количественный анализ динамики информационных потоков, которые генерируются в Интернете, становится сегодня одним из наиболее информативных методов исследования актуальности тех или иных тематических направлений. Эта динамика обусловлена разнообразными качественными факторами, многие из которых не поддаются точному описанию. Однако общий характер временной зависимости количества тематических публикаций в сети Интернет все же допускает построение математических моделей, их исследование, прогнозирование. Наблюдения временных зависимостей объемов сетевых информационных потоков убедительно свидетельствуют о том, что механизмы их генерации и распространения, очевидно, связаны со сложными нелинейными процессами. Именно этой теме посвящена данная глава.

Для анализа временных рядов, которые отображают зависимость объемов информационных потоков от времени, используют разнообразные методы и подходы. При этом оказывается, что все эти подходы взаимосвязаны и более того, ключевую роль играет понятие корреляции.

Изложение построено вокруг схемы показанной на Рис. 3.1, причем особое внимание уделено взаимосвязям.



Рис. 3.1 – Взаимосвязи между подходами к анализу временных рядов

3.1. Временные ряды

Временной ряд – это набор наблюдаемых значений упорядоченных по времени. Далее будут рассматриваться дискретные временные ряды, значения которых фиксировались через равные промежутки времени. Будем обозначать такой временной ряд x_1, x_2, \dots, x_T или коротко $\{x_t\}_{t=1}^T$ подразумевая, что фиксирование значений ряда происходило через равный промежуток времени h : $t_0, t_0 + h, t_0 + 2h, \dots, t_0 + (T - 1)h$.

Если значения временного ряда однозначно задаются некоторым математическим соотношением (таким как, например, $x_t = A \cdot \sin(vt)$), то такой ряд является детерминированным. Если значения временного ряда можно описать только в терминах вероятностного распределения, то речь идет о статистическом временном ряде. Такие ряды и будут рассматриваться далее. Анализируя временные ряды, мы будем рассматривать их как реализацию стохастического процесса.

В качестве примеров далее будут использоваться три временных ряда, которые были получены с помощью популярного сетевого сервиса GoogleTrends. Эти временные

ряды отображают уровень интереса к Дональду Трампу, Хилари Клинтон и «русским хакерам» с августа 2016 года по апрель 2017 года. Временные ряды, получаемые с помощью GoogleTrends, показывают динамику популярности поискового запроса. Максимальная точка на графике равна 100 и соответствует дате, когда запрос был наиболее популярен, а остальные точки на графике определяются в процентном соотношении к максимуму. Все три временных ряда показаны на Рис. 3.2. Для простоты ссылок на данные ряды в дальнейшем обозначим их Т (Д. **Т**рамп), К (Х. **К**линтон), Х («русские **х**акеры»).

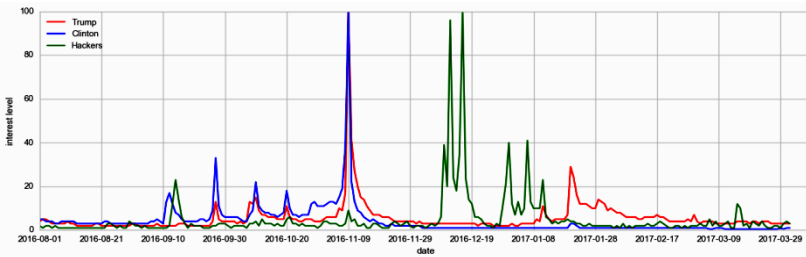


Рис. 3.2 – Временные ряды, которые отображают интерес к Дональду Трампу (Т), Хилари Клинтон (К) и «русским хакерам» (Х) с 1 августа 2016 года по 1 апреля 2017 из GoogleTrends.

В некоторых случаях полезно рассмотреть более гладкую версию исходного временного ряда. Сглаживание помогает выявить существенные тенденции в динамике ряда, скрыв при этом шум и различные особенности, которые проявляются при небольших масштабах. Существуют разнообразные методы сглаживания. Наиболее простой способ сглаживания – это вычисление скользящего среднего. Простое скользящее среднее равно среднему арифметическому значению элементов ряда из интервала заданной длины, а именно

$$SMA_t = \frac{1}{w} \sum_{i=0}^{w-1} x_{t-i},$$

где w – ширина сглаживающего интервала (количество элементов, по которым рассчитывается среднее), SMA_t – значение простого скользящего среднего в точке t .

Полученное значение SMA_t относится к середине сглаживающего интервала, поэтому сглаженный ряд y_t может быть определен как $y_t = SMA_{t + \lfloor \frac{w}{2} \rfloor}$.

При использовании сглаживания скользящим средним, чем больше ширина сглаживающего интервала, тем более гладкой получится функция. На Рис. 3.3 показано как выглядит сглаженный ряд T при увеличении значения w .

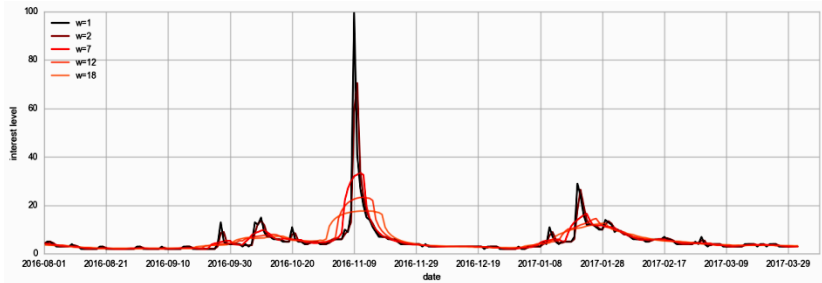


Рис. 3.3 – Исходный ряд T и сглаженный простым скользящим средним с шириной сглаживающего интервала 2, 7, 12, 18.

Результаты сглаживания ряда можно продемонстрировать на графике, у которого ось абсцисс соответствует временной оси, а вдоль оси ординат отложена ширина сглаживающего интервала. На графике показаны значения $y_t^{(w)}$ – то есть элементы сглаженного ряда в точке t при использовании интервала ширины w (Рис. 3.4).

При вычислении простого скользящего среднего все точки, которые попали в сглаживающий интервал, имеют одинаковый вес. Естественно, что можно использовать не равные веса. Таким образом, приходим к определению взвешенного скользящего среднего

$$WMA_t = \frac{1}{w} \sum_{i=0}^{w-1} a_i x_{t-i},$$

где $\sum_{i=0}^{w-1} a_i = 1$.

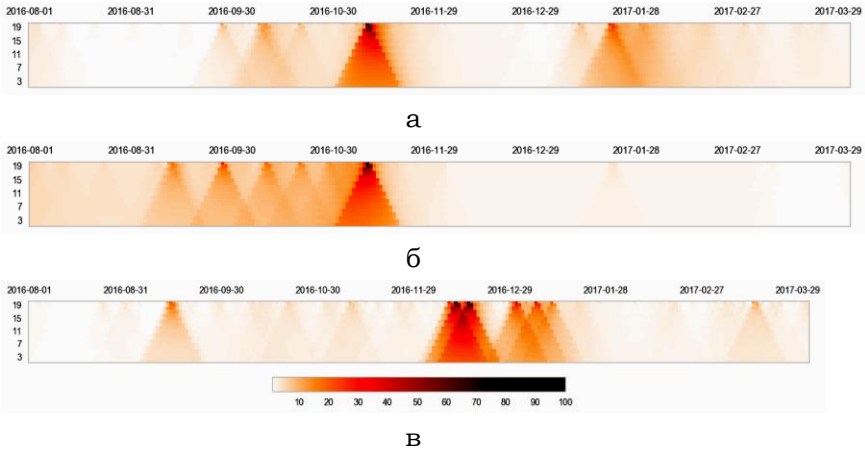


Рис. 3.4 – Значения сглаженных простым скользящим средним временных рядов Т (а), К (б) и Х (в) в зависимости от ширины сглаживающего интервала. Вдоль оси абсцисс отложено время, а вдоль оси ординат – ширина интервала.

Другой часто используемый метод сглаживания рядов – это **экспоненциальное сглаживание**. Предыдущие значения ряда учитываются с экспоненциально убывающими весами. Будем обозначать элементы сглаженного ряда y_t , и сразу определим $y_0 = x_0$. Следующие элементы ряда y_t получают по рекурсивной формуле

$$y_t = \alpha x_t + (1 - \alpha)y_{t-1},$$

где $0 < \alpha < 1$ – коэффициент сглаживания. Очевидно, что при $\alpha = 1$ получаемый ряд y_t совпадает с исходным x_t . Таким образом, если значение α близко к 1, то наибольший вес при определении y_t присваивается соответствующему x_t , а предыстория ряда «мало значит». С другой стороны, если бы α равнялось 0, то весь ряд y_t сгладился бы до одного значения $y_t = y_0$. То есть при α близком к 0 предыстория ряда учитывается с большим весом, чем текущее значение.

На Рис. 3.5 показан ряд Т, а также соответствующие сглаженные ряды при различных значениях параметра α .

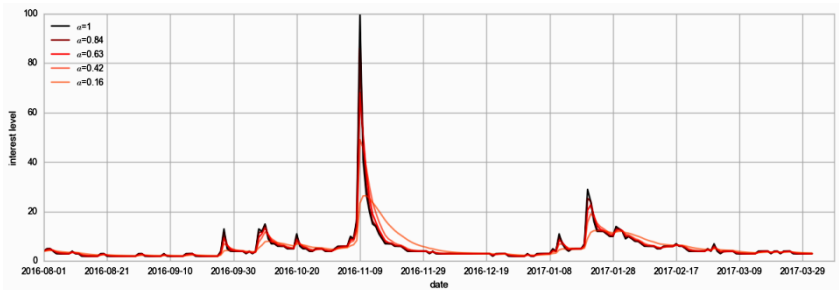


Рис. 3.5 – Исходный ряд T и сглаженный с помощью экспоненциального сглаживания с параметром, равным 0.84, 0.63, 0.42, 0.16

Как и в случае с простым скользящим средним продемонстрируем результаты сглаживания ряда на графике. В данном случае вдоль оси ординат отложим параметр α (Рис. 3.6). На графике показаны значения $y_t^{(\alpha)}$ – значение в точке t сглаженного с параметром α исходного ряда.

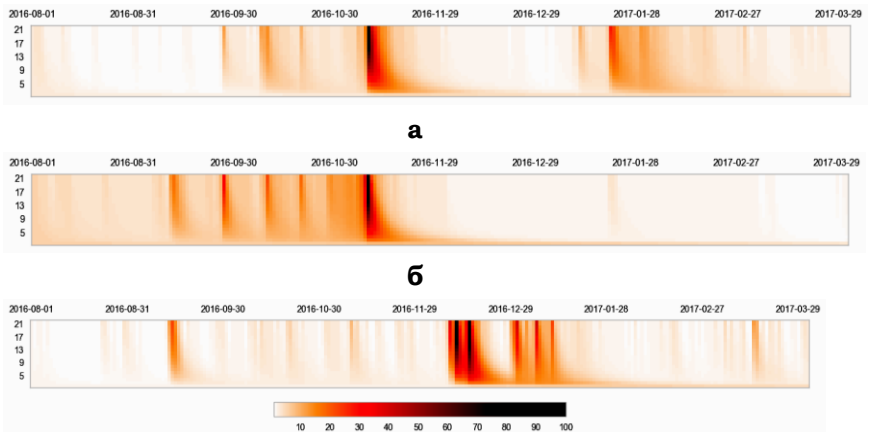


Рис. 3.6 – Значения экспоненциально сглаженных временных рядов T (а), K (б) и X (в) в зависимости от параметра α . Вдоль оси абсцисс отложено время, а вдоль оси ординат – параметр α .

В качестве примеров мы рассматриваем временные ряды Т, К, и Х, у которых есть недельная периодичность. Это характерное свойство многих процессов в информационном пространстве. Известно, что публикация новостных сообщений часто происходит с недельной периодичностью, а также активность пользователей разнится в будни и выходные дни. Для того чтобы исключить периодическую компоненту из рядов, сгладим их с помощью простого скользящего среднего с интервалом шириной 7 (число дней в неделе) в соответствии с формулой:

$$x_t^{New} = \frac{x_{t-3} + x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2} + x_{t+3}}{7},$$

где x_t – исходные значения ряда, x_t^{New} – новое значение ряда в момент времени t . На Рис. 3.7 показаны сглаженные временные ряды.

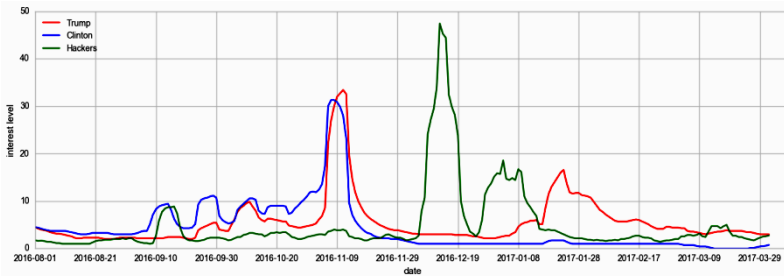


Рис. 3.7 – Временные ряды Т, К и Х, сглаженные с помощью простого скользящего среднего с интервалом длины 7

3.2. Корреляционный анализ

Многие методы исследования временных рядов базируются на некотором предположении о статистическом равновесии или постоянстве. Одним из таких полезных предположений является стационарность [Вох 2015].

Временной ряд называется *строго стационарным* или *стационарным в узком смысле*, если его статистические свойства не изменяются со временем. Формально, если совместное распределение случайных величин $x_t, x_{t+1}, \dots, x_{t+n}$ совпадает с распределением $x_{t+k}, x_{t+k+1}, \dots, x_{t+k+n}$ при любых целых значениях сдвига k , то временной ряд $\{x_t\}_{t=1}^T$

называется строго стационарным. У стационарных временных рядов постоянное математическое ожидание

$$\mu = Ex_t$$

и дисперсия

$$\sigma^2 = Var(x_t) = E(x_t - Ex_t)^2.$$

При этом значения μ и σ^2 можно оценить как выборочное среднее

$$\hat{\mu} = \bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$$

и выборочную дисперсию

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^2.$$

Свойство стационарности также имеет большое значение при сравнении временных рядов. Линейная зависимость между двумя случайными величинами измеряется ковариацией. Для временных рядов определяют кроссковариационную функцию. По определению, **кросс-ковариация** с временной задержкой k между случайными процессами $\{x_t\}_{t=1}^T$ и $\{y_t\}_{t=1}^T$ равна

$$\gamma_{xy}(k, t) = Cov(x_t, y_{t+k}) = E[(x_t - \mu_x)(y_{t+k} - \mu_y)].$$

Из предположения о стационарности в узком смысле следует, что распределение пар величин x_t, y_{t+k} одинаково для произвольного значения t . Следовательно, ковариация между величинами x_t и y_{t+k} не зависит от t , а зависит только от значения k , то есть $\gamma_{xy}(k, t) = \gamma_{xy}(k), \forall t$. Набор значений $\{\gamma_{xy}(k)\}$ образует кроссковариационную функцию.

Нормировав кроссковариационный коэффициент, получим **кросскорреляционный коэффициент**

$$\rho_{xy}(k) = \frac{Cov(x_t, y_{t+k})}{\sigma_x \sigma_y} = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y}.$$

Кросскорреляционная функция является мерой подобия между двумя временными рядами.

Чаще всего кроссковариационные и кросскорреляционные коэффициенты оценивают по формулам

$$\hat{\gamma}_{xy}(k) = \frac{1}{T} \sum_{t=1}^{T-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), \quad \hat{\rho}_{xy}(k) = \frac{\hat{\gamma}_{xy}(k)}{\hat{\gamma}_{xy}(0)}.$$

Заметим, что такие оценки справедливы для стационарных в узком смысле рядов, так как соответствующие коэффициенты не зависят от времени, а в общем случае это может и не выполняться. Часто используют более слабое требование, чем стационарность в узком смысле, – стационарность в широком смысле.

Временной ряд $\{x_t\}_{t=1}^T$ стационарен в широком смысле, если его математическое ожидание не изменяется со временем, то есть $\forall t \exists E x_t = \text{const}$ и ковариационная функция зависит только от разности аргументов $\text{Cov}(x_t, x_s) = K(t - s)$.

Так как в определении указано, что математическое ожидание постоянно и, легко заметить, что дисперсия также не изменяется со временем $\text{Var}(x_t) = \text{Cov}(x_t, x_t) = K(0) = \text{const}$, то в этом случае, как и для строго стационарных рядов, справедливы оценки (1) и (2).

На Рис. 3.8 показана иллюстрация к вычислению корреляции. Рассматриваются два центрированных временных ряда. Для вычисления коэффициента корреляции нужно умножить соответствующие элементы рядов и вычислить их среднее значение. Результат умножения на Рис. 3.8 показан линией. Площадь затемненной области под линией с учетом знака равна коэффициенту ковариации между двумя рядами.

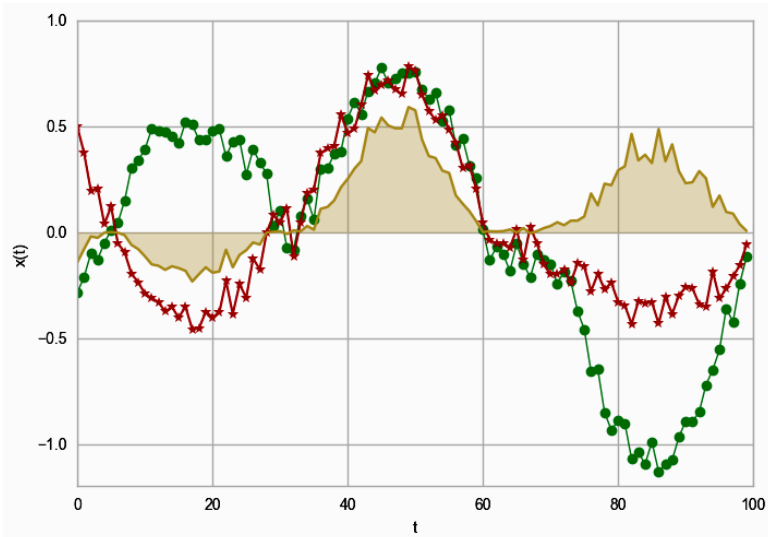
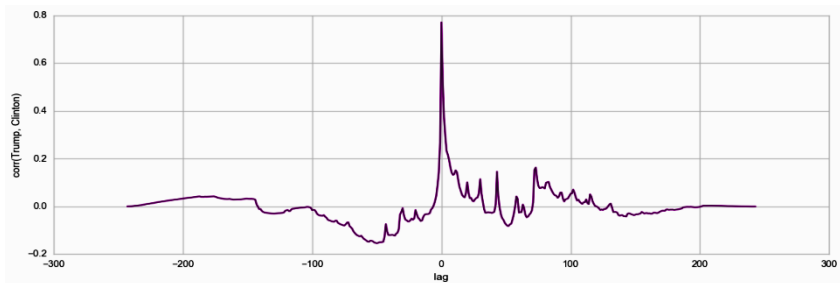


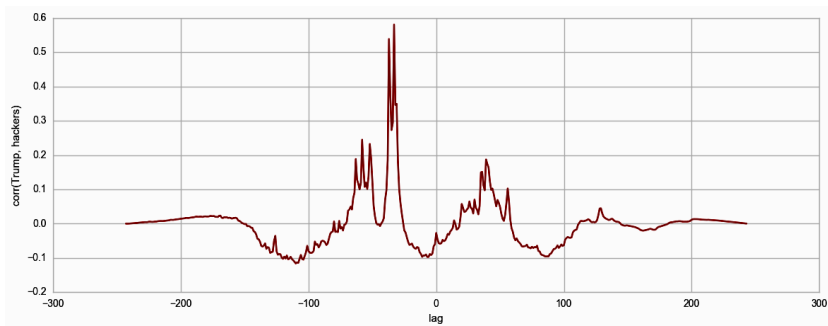
Рис. 3.8 – Иллюстрация к определению корреляции. Показаны два временных ряда. Затемненная область показывает вклад в значение корреляции между этими двумя рядами

Для примера приведем оценку кросскорреляционных функций для рядов Т, К, Х. На Рис. 3.9 а показана корреляционная функция для рядов Т и К. Максимальное значение (приблизительно равное 0.8) функция достигает при временной задержке 0. То есть два временных ряда, связанные с интересом к Дональду Трампу и Хилари Клинтон, сильно коррелированы.

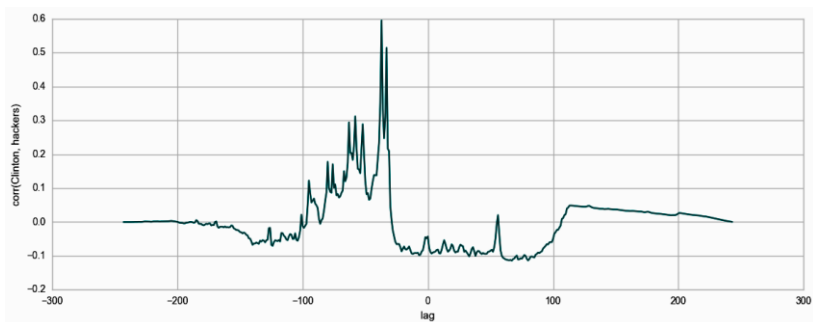
На Рис. 3.9б показана корреляционная функция для рядов Т и Х. Максимальное значение (приблизительно равное 0.7) функция достигает при временной задержке 34 дня. Это отвечает тому факту, что начиная с 13 декабря 2016 года (34 дня после выборов в США 8 ноября) резко возросло количество новостных сообщений про «русских хакеров».



а



б



в

Рис. 3.9 – Корреляционные функции для пар рядов Т и К (а), Т и Х (б), К и Х (в). Вдоль оси абсцисс отложена временная задержка (lag), вдоль оси ординат – оценка корреляционного коэффициента

Автокорреляция

Можно подсчитать ковариацию не для двух различных рядов, а для одного ряда. Такая ковариация называется **автоковариацией** с временной задержкой или лагом k

$$\gamma_k = Cov(x_t, x_{t+k}) = E[(x_t - \mu)(x_{t+k} - \mu)].$$

Набор величин $\gamma_k, k = 0, 1, 2, \dots$ называется автоковариационной функцией, а их нормированное значение $\rho_k, k = 0, 1, 2, \dots$ – **автокорреляционной функцией**

$$\rho_k = \frac{E[(x_t - \mu)(x_{t+k} - \mu)]}{\sqrt{E(x_t - \mu)^2 E(x_{t+k} - \mu)^2}} = \frac{Cov(x_t, x_{t+k})}{Var(x_t)} = \frac{\gamma_k}{\gamma_0}.$$

Автокорреляционная функция описывает зависимость между значениями случайного процесса в различные моменты времени (Рис. 3.10).

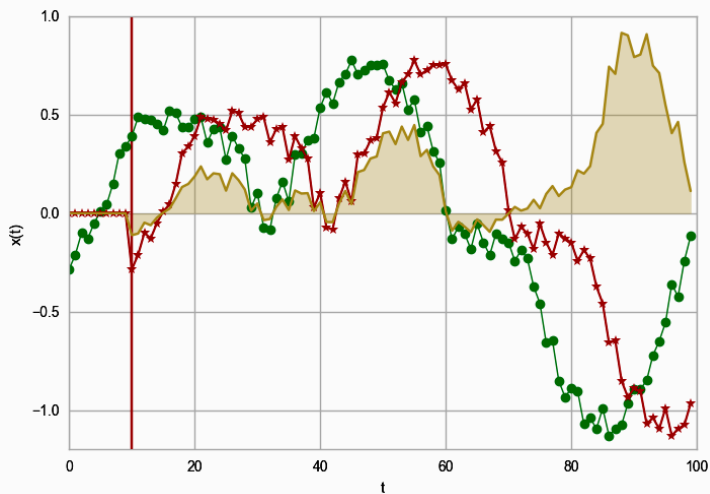


Рис. 3.10 – Иллюстрация к определению автокорреляции. Показан временной ряд и тот же самый ряд, сдвинутый на 10 значений вправо. Затененная область показывает вклад в значение автокорреляционного коэффициента с лагом 10.

Чаще всего автоковариационные и автокорреляционные коэффициенты оценивают по формулам

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x}), \quad \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}.$$

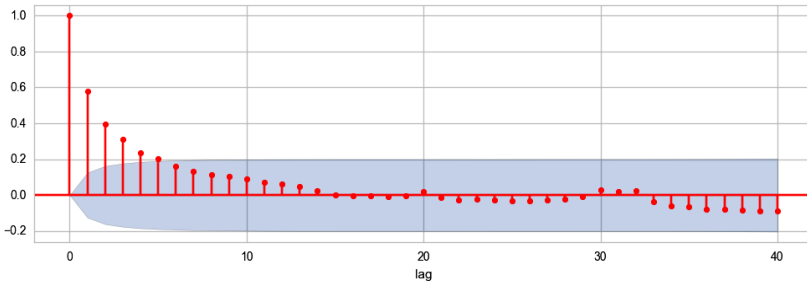
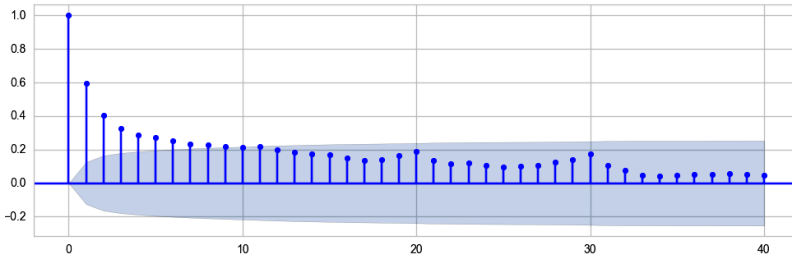
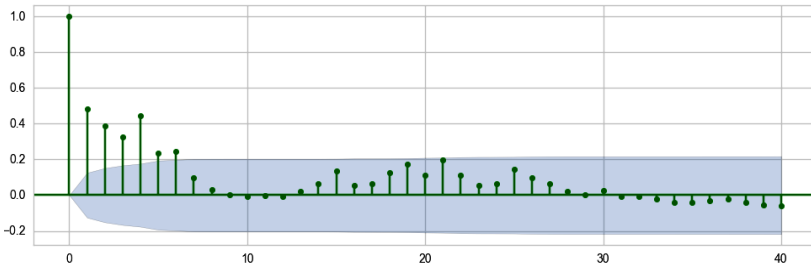
**а****б****в**

Рис. 3.11 – Автокорреляционные функции для рядов Т (а), К (б), Х (в). Вдоль оси абсцисс отложена временная задержка (лаг), вдоль оси ординат – автокорреляционный коэффициент. Затемненная область показывает стандартное отклонение для оценки автокорреляционного коэффициента

После вычисления оценок для автокорреляционных коэффициентов возникает вопрос: равны ли

коэффициенты ρ_k нулю начиная с некоторого значения k ? Для ответа на этот вопрос нужно сравнить значение оценки $\hat{\rho}_k$ с его стандартным отклонением. Если мы принимаем предположение, что $\rho_k = 0$, то стандартное отклонение оценки $\hat{\rho}_k$

$$se(\hat{\rho}_k) \cong \frac{1}{\sqrt{T}}$$

На практике часто используют эмпирическое правило, согласно которому автокорреляционные коэффициенты оценивают для временной задержки не более чем $T/4$. На Рис. 3.11 показаны автокорреляционные функции для рядов T, K, X.

Определение автокорреляционной функции вводилось для стационарных временных рядов, но оценить ее значение можно для произвольного временного ряда. Для нестационарных временных рядов такая автокорреляционная функция убывает очень медленно.

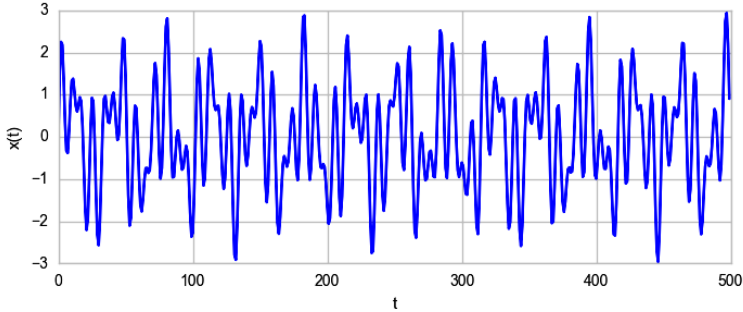
3.3. Анализ Фурье

Классический анализ Фурье предоставляет возможность исследовать функцию во временной и частотной области. Суть перехода в частотную область состоит в том, что функция раскладывается на составляющие, которые являются гармоническими колебаниями с разными частотами. При этом каждой частоте соответствует коэффициент, который отображает амплитуду колебания на данной частоте. Если представить функцию графически во временной области, то получим информацию о том, как функция изменяется со временем. Если изобразить функцию в частотной области, то получим информацию о частотах, колебания на которых она содержит. Для этого используют прямое и обратное преобразование Фурье

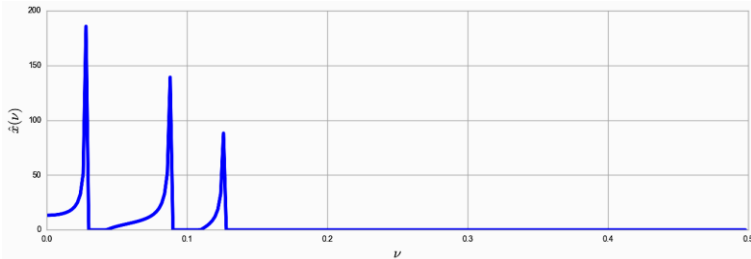
$$\hat{x}(\nu) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi\nu t} dt,$$

$$x(t) = \int_{-\infty}^{\infty} \hat{x}(\nu)e^{i2\pi\nu t} d\nu.$$

На Рисунке 3.12а показан пример функции, которая на самом деле является суммой трёх синусоид с разными периодами. Глядя только на график функции во временной области достаточно трудно понять, что она состоит из трёх гармонических колебаний и определить их периоды. На Рисунке 3.12б показано преобразование Фурье для этой функции. Из графика в частотной области наглядно видно, что функция содержит колебания на трёх разных частотах.



а



б

Рис. 3.12 – Функция, которая является суммой трёх синусоид с различными периодами (а) и оцененный спектр Фурье для этой функции (б)

Сегодня преобразование и спектры Фурье находят разнообразные применения в системах машинного обучения. Часто спектры Фурье используются в качестве обучающих параметров. Например, в [Rodrigues 2014] предложена модель прогнозирования временного ряда, в которой спектр Фурье вместе с некоторыми другими параметрами подается на вход нейронной сети.

Преобразования и спектры Фурье часто используются при распознавании речи. В [Alam 2014] специальные признаки, сформированные на основе преобразования Фурье, используются в системе распознавания речи с различными условиями обучения. Спектры Фурье также используются в качестве обучающих параметров для нейронных сетей в системах автоматического детектирования определенных событий в речи или на фоне шума [Sazonov 2010, Wang 2014]. Другой задачей в области распознавания является определение эмоциональной окраски речи. В [Wang 2015] предложена модель распознавания, в основе которой лежат определенные параметры Фурье, и демонстрируется эффективность использования таких параметров для идентификации различных эмоциональных состояний в голосовых сигналах.

В алгоритмах машинного обучения основанных на применении ядра, таких как машина опорных векторов (support vector machine), для аппроксимации ядер высокой размерности часто используют случайные признаки Фурье (random Fourier features). Такой подход был предложен в [Rahimi 2008] и основан на теореме Бохнера из гармонического анализа, которая гарантирует, что при некоторых свойствах ядра его преобразование Фурье будет вероятностным распределением.

Преобразование Фурье можно воспринимать как определение корреляции между исходным сигналом и гармоническими функциями с различными частотами колебания. На Рис. 3.13 показана иллюстрация аналогичная с Рис. 3.8 и Рис. 3.10.

Несмотря на свои преимущества и многочисленные приложения, преобразование Фурье является плохим методом для исследования функций, которые эволюционируют со временем. Для таких функций нужен некоторый способ оценивания спектра не по всей длине временного ряда, а по его различным частям. Примером такого подхода является оконное преобразование Гэбора

$$G(v, \tau, s) = \int_{-\infty}^{\infty} x(t) e^{-\frac{(t-\tau)^2}{s^2}} e^{-i2\pi vt} dt.$$

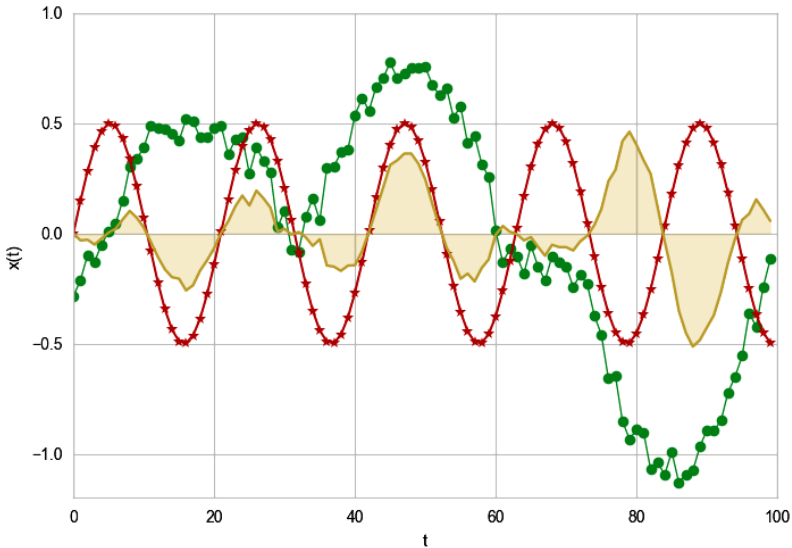


Рис. 3.13 – Иллюстрация к определению преобразования Фурье как вычисление корреляции между исходным сигналом и гармоническим колебанием. Затемненная область показывает вклад в значение преобразования Фурье или амплитуды, которая соответствует данной частоте колебания

Временное окно $e^{-\frac{(t-\tau)^2}{s^2}}$ выделяет отрезок временного ряда с центром в точке τ и имеет ширину, которая определяется параметром s , что позволяет выделить часть исследуемого ряда.

При использовании преобразования Гэбора возникает проблема выбора ширины окна. Сделать оконную функцию зависящей от частоты так, чтобы для низких частот окно становилось шире, а высоких – уже, позволяет следующий класс преобразований, а именно вейвлет преобразование. Основное преимущество вейвлет преобразования состоит в том, что выделенный из временного ряда кусок анализируется с той степенью детальности, которая соответствует его масштабу.

3.4. Вейвлет-анализ

Вейвлет преобразование имеет корреляционную природу. В данном случае рассматривается корреляция исходной функции с функцией вейвлетом на разных масштабах. Для того чтобы такую процедуру всегда можно было выполнить и корреляционные коэффициенты были информативными, вейвлет должен обладать определенными математическими свойствами. Буквально слово вейвлет переводится как «маленькая волна» или «всплеск», и, как следует из названия, вейвлет хорошо локализован во времени. С математической точки зрения, вейвлет – это функция $\psi(t)$, которая удовлетворяет следующим свойствам:

1. Функция $\psi(t)$ квадратично интегрируема ($\psi \in L^2(\mathbb{R})$) или, другими словами, имеет конечную энергию

$$E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty.$$

2. Обозначим $\hat{\psi}(\lambda)$ преобразование Фурье от функции $\psi(t)$, тогда

$$\int_0^{\infty} \frac{|\hat{\psi}(\lambda)|^2}{\lambda} d\lambda < \infty.$$

На Рис. 3.14 показаны примеры вейвлетов, которые часто используются на практике.

Непрерывное вейвлет преобразование

Вейвлет $\psi(t)$, свойства которого были описаны выше, часто называют материнским или базовым вейвлетом. На основании материнского вейвлета строят семейство функций с помощью растяжения/сжатия и параллельного переноса. Это необходимо, чтобы исследовать различные области исходного сигнала и с различной степенью детальности.

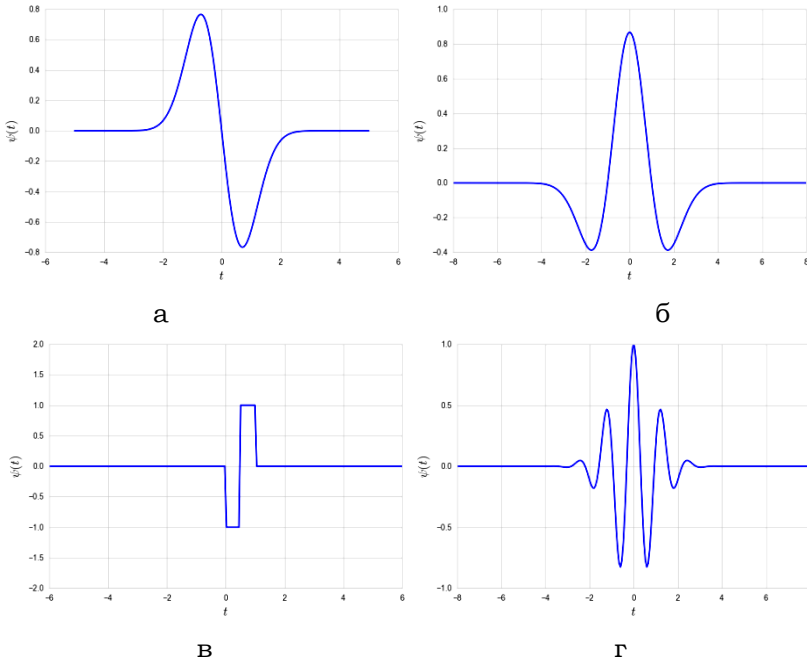


Рис. 3.14 – Примеры вейвлетов, которые часто используются в приложениях: (а) гауссова волна (первая производная гауссовой функции), (б) мексиканская шляпа, (в) вейвлет Хаара, (г) вейвлет Морле (действительная часть).

Введем параметры масштаба s (scale) и сдвига l (location), тогда преобразованная версия материнского вейвлета будет следующей

$$\psi_{s,l}(t) = \frac{1}{\sqrt{|s|}} \psi\left(\frac{t-l}{s}\right).$$

Непрерывным вейвлет преобразованием функции $x(t) \in L^2(\mathbb{R})$ называется выражение

$$W(s, l) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-l}{s}\right) dt = \int_{-\infty}^{\infty} x(t) \psi_{s,l}^*(t) dt,$$

где $l, s \in \mathbb{R}$, $s \neq 0$; ψ^* - функция комплексно сопряженная с ψ , величины $\{W(s, l)\}_{l, s \in \mathbb{R}}$ называются коэффициентами вейвлет преобразования.

Из формулы в определении непрерывного вейвлет преобразования непосредственно видно, что суть такого преобразование состоит в вычислении корреляционных коэффициентов специального вида. На Рис. 3.15 показано как на исходный ряд накладывается вейлет мексиканская шляпа и определяется корреляция между частью ряда и “шаблоном”, который представляет собой вейвлет.

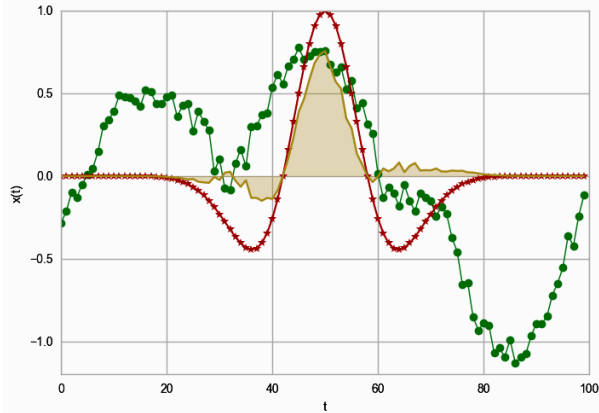


Рис. 3.15 – Иллюстрация к вычислению вейвлет преобразования как вычислению корреляции между исходным сигналом и функцией вейвлетом. Затемненная область показывает вклад в значение вейвлет преобразования для данных сдвига и масштаба

Стоит заметить, что непрерывное вейвлет преобразования является обратимой операцией. Обратное вейвлет преобразование выполняется следующим образом:

$$x(t) = \frac{1}{C_g} \int_{-\infty}^{\infty} \int_0^{\infty} W_x(s, l) \psi_{s, l}(t) \frac{ds dl}{s^2}.$$

Проиллюстрируем результаты вейвлет преобразования на нескольких простых примерах (Рис. 15). Первый пример – сумма двух колебательных процессов. Под графиком сигнала показаны коэффициенты вейвлет преобразования, которые были получены при использовании вейвлета мексиканская шляпа. Вдоль горизонтальной оси изменяется время (параметр сдвига l), вдоль вертикальной оси – масштаб (параметр s). На

графике вейвлет коэффициентов можно увидеть два периодических процесса. На Рис. 3.16 можно увидеть, как отражаются на значениях вейвлет коэффициентов периодические процессы с разной амплитудой и частотой, а также отдельные пики в сигнале.

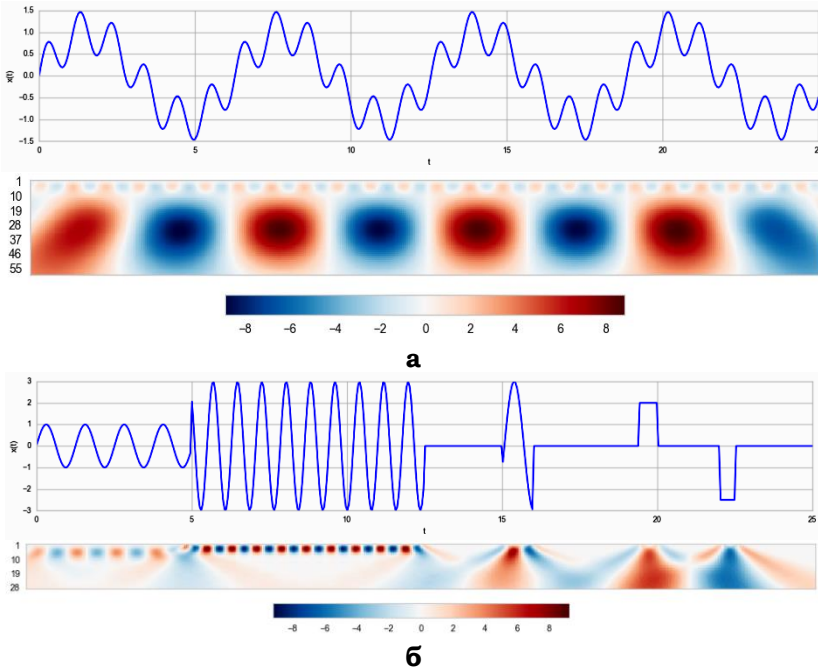


Рис. 3.16 – Примеры вейвлет преобразования. Сумма двух синусоид и ее вейвлет-преобразование (а). Функция, составленная из колебательных процессов разной частоты и амплитуды, а также отдельных пиков (б).

Посмотрим как выглядят коэффициенты вейвлет преобразования для временных рядов T , K , X . На Рис. 3.17а, б показаны результаты вейвлет преобразования для ряда T с использованием вейвлета мексиканская шляпа (а) и гауссовой волны (б). На Рис. 3.17в, г показаны результаты вейвлет преобразования для рядов K и X с использованием вейвлета «мексиканская шляпа».

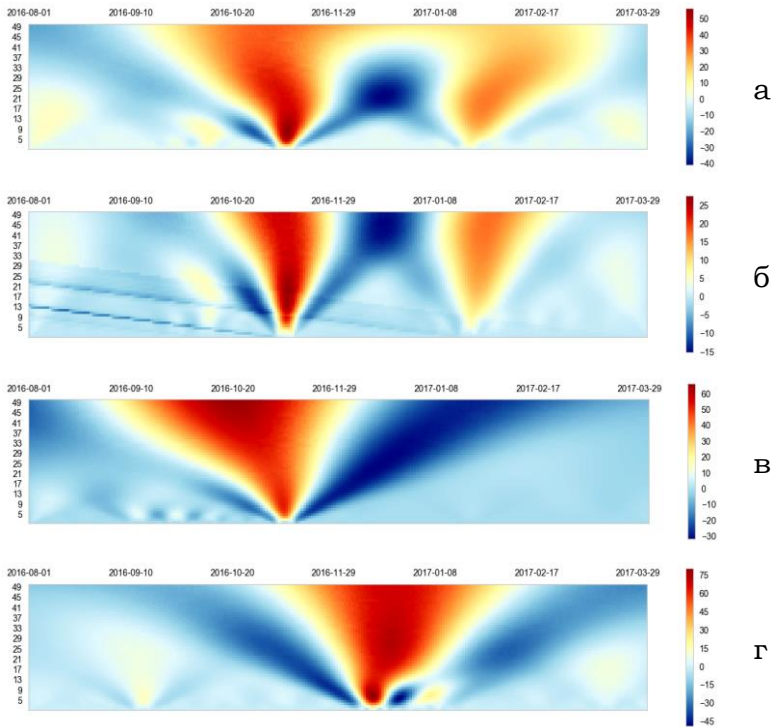


Рис. 3.17 – Вейвлет-преобразование для ряда Т с использованием вейвлета мексиканская шляпа (а); действительная часть вейвлет-преобразования для ряда Т с использованием вейвлета Морле (б); вейвлет преобразование для ряда К (в) и ряда Х (г) с использованием вейвлета мексиканская шляпа

Еще раз заметим, что непрерывное вейвлет преобразование, как и преобразование Фурье можно рассматривать в терминах корреляции. Преобразование Фурье – это корреляция между исходным временным рядом и волной $\varphi(t) = e^{-i2\pi\nu t}$. Волна покрывает всю временную ось и характеризуется только частотой ν , поэтому преобразование Фурье зависит только от частоты. Вейвлет преобразование – это корреляция между исходным временным рядом и вейвлетом $\psi(t)$. Таким образом, вейвлет преобразование зависит от положения вейвлета на

временной оси и его масштаба, которые определяются параметрами l и s соответственно.

Энергия сигнала

Полная энергия сигнала $x(t)$ по определению равна

$$E = \int_{-\infty}^{\infty} |x(t)|^2 dt.$$

Используя коэффициенты вейвлет преобразования можно определить энергию сигнала, которая соответствует определенному сдвигу и масштабу

$$E(s, l) = |W(s, l)|^2.$$

Значения $E(s, l)$ можно изобразить на графике, точно также как и коэффициенты вейвлет преобразования. Такой график обычно называют скейлограммой (scalogram). Также можно определить относительный вклад энергии, которая соответствует определенному масштабу, в полную энергию, или другими словами распределение энергии в зависимости от масштаба

$$E(s) = \frac{1}{C_g} \int_{-\infty}^{\infty} |W(s, l)|^2 dl.$$

Заметим, что на скалограмме, в отличие от графика с коэффициентами вейвлет преобразования, все значения положительны, и наиболее ярко выделяются области с наибольшей энергией (Рис. 3.18).

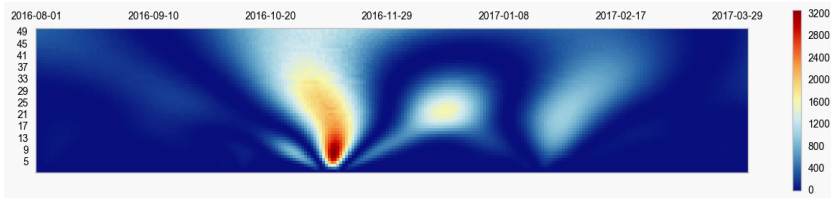


Рис. 3.18 – Скалограмма для временного ряда T

Сравнение временных рядов с помощью вейвлет преобразования

Ранее был показан способ выявления зависимости между двумя временными рядами с помощью кросс-корреляции.

Сейчас мы рассмотрим некоторые способы сравнения временных рядов с помощью коэффициентов вейвлет преобразования. Эти способы также можно применять для того, чтобы выявить некоторый тип отношения или взаимосвязи между временными рядами. Метрики сравнения коэффициентов вейвлет преобразования, а также примеры их применения к реальным практическим задачам, подробно описаны в [Addison 2017].

Рассмотрим два временных ряда x_t и y_t , и обозначим коэффициенты вейвлет преобразования этих рядов $W_x(s, l)$ и $W_y(s, l)$. Начнем с простейшего способа сравнения – возьмем разность модулей соответствующих коэффициентов

$$DiffMOD_{x,y}(s, l) = |W_x(s, l)| - |W_y(s, l)|.$$

На Рис. 3.19 показаны значения $DiffMOD_{x,y}(s, l)$ для двух пар рядов – сверху для рядов Т и К, а снизу для рядов Т и Х. Таким простым способом можно выделить области в которых коэффициенты вейвлет преобразования схожи, а значит и в исходных временных рядах есть похожие участки.

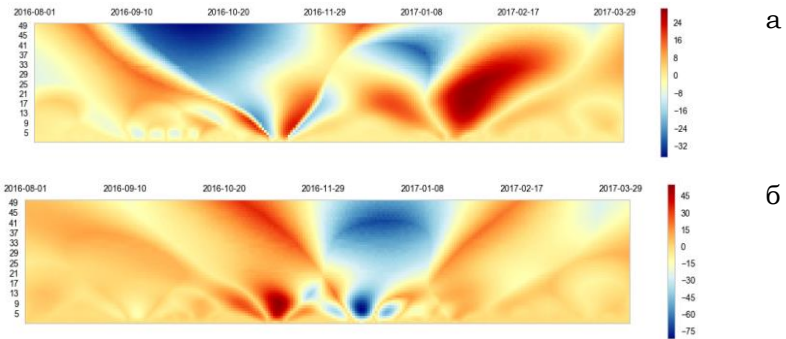


Рис. 3.19 – $DiffMOD_{x,y}(s, l)$ для рядов Т и К (а), Т и Х (б)

Другим простым способом сравнения является отношение модулей коэффициентов вейвлет преобразования

$$RatioMOD_{x,y}(s, l) = \frac{|W_x(s, l)|}{|W_y(s, l)|}.$$

Такую метрику нужно использовать с осторожностью, так как $W_y(s, l)$ принимать нулевые или близкие к нулю значения.

Дополнительную информацию можно получить, если использовать комплексный вейвлет (например, вейвлет Морле). Тогда кроме абсолютного значения вейвлет коэффициентов появляется фаза. Комплексный коэффициент всегда можно представить в виде

$$W(s, l) = |W(s, l)|e^{i\phi(s, l)},$$

где, как известно, модуль числа равен

$$|W(s, l)| = \sqrt{\operatorname{Re}(W(s, l))^2 + \operatorname{Im}(W(s, l))^2},$$

а фаза

$$\phi(s, l) = \tan^{-1} \left[\frac{\operatorname{Im}(W(s, l))}{\operatorname{Re}(W(s, l))} \right].$$

Следовательно, можно также сравнить фазы коэффициентов

$$\Delta\phi_{x,y}(s, l) = \phi_x(s, l) - \phi_y(s, l)$$

Кросс-вейвлет преобразование используется для выделения областей одинаковой энергии между сигналами в области преобразования, а также определения относительной фазы

$$CrWT_{x,y}(s, l) = W_x^*(s, l)W_y(s, l).$$

На рисунках обычно отображают значение $|CrWT_{x,y}(s, l)|$, по аналогии со скейлограммой. В таком случае, если временной ряд x идентичен ряду y , то мы получим скейлограмму для ряда x .

Представляет особый интерес вычисление кросс-вейвлет преобразования в случае, когда используется комплексный вейвлет (например, вейвлет Морле). Тогда

$$\begin{aligned} CrWT_{x,y}(s, l) &= W_x^*(s, l)W_y(s, l) \\ &= |W_x(s, l)|e^{-i\phi_x(s, l)}|W_y(s, l)|e^{i\phi_y(s, l)} = \\ &= |W_x(s, l)||W_y(s, l)|e^{i(\phi_y(s, l) - \phi_x(s, l))}. \end{aligned}$$

Таким образом, вычисляя кросс-вейвлет преобразование можно извлечь значение разности фаз между коэффициентами вейвлет преобразования для двух временных рядов.

На Рис. 3.20 показаны значения $CrWT_{x,y}(s, l)$ для рядов Т и К. Для рядов Т и К выделяется область соответствующая пику интереса во время выборов, что означает, что это область высокой энергии для обоих рядов.

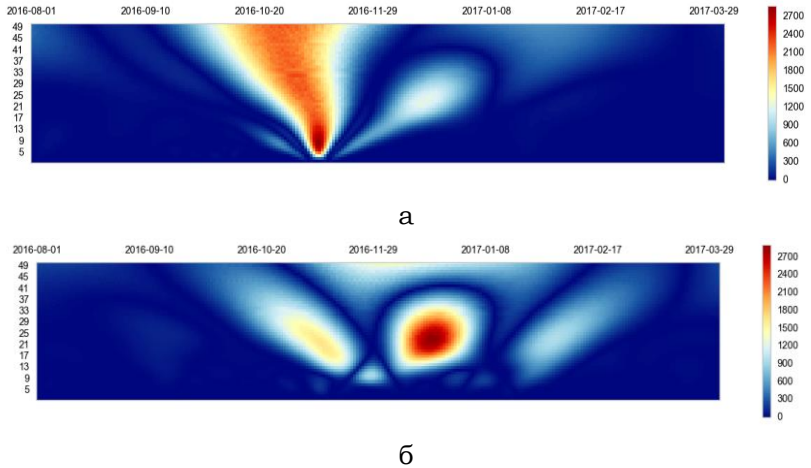


Рис. 3.20 – Кросс-вейвлет преобразование с использованием вейвлета мексиканская шляпа для рядов Т и К (а) и рядов Т и Х (б)

Если проинтегрировать коэффициенты вейвлет преобразования по времени, то получим вейвлет-кросс-корреляционную меру (Wavelet Cross-Correlation Measure), которая зависит от масштаба

$$\begin{aligned}
 W_{x,y}(s) &= \frac{|\int W_x^*(s, l) W_y(s, l) dl|}{\sqrt{\int |W_x(s, l)|^2 dl \int |W_y(s, l)|^2 dl}} \\
 &= \frac{|\int CrWT_{x,y}(s, l) dl|}{\sqrt{\int |W_x(s, l)|^2 dl \int |W_y(s, l)|^2 dl}}
 \end{aligned}$$

Такая мера помогает обнаружить корреляцию между сигналами, которые содержат колебания с разной амплитудой или фазой, но, тем не менее, коррелированы между собой (3.21).

Также можно расширить определение вейвлет кросс-корреляционной меры, если ввести зависимость от сдвига между рядами

$$W_{x,y}(s, k) = \frac{|\int W_x^*(s, l)W_y(s, l - k)dl|}{\sqrt{\int |W_x(s, l)|^2 dl \int |W_y(s, l)|^2 dl}}$$

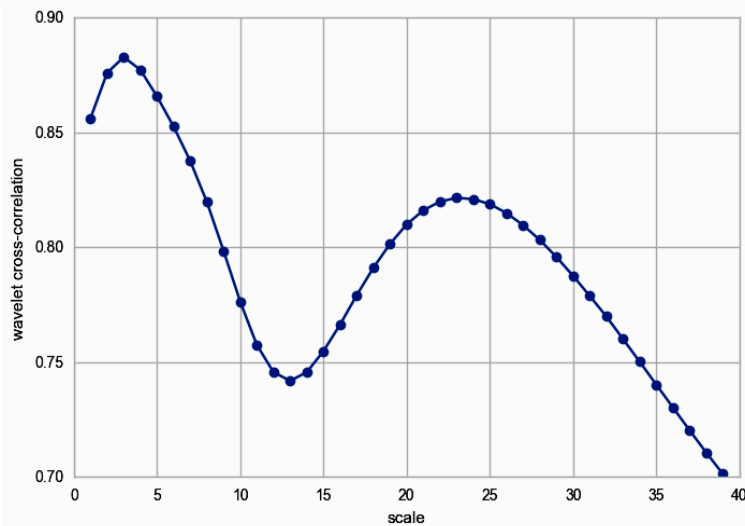


Рис. 3.21 – Зависимость вейвлет-кросс-корреляционной меры от масштаба для временных рядов T и K

Для исследования связи между компонентами временных рядов локально в плоскости преобразования, мы можем определить квадратичную оценку **когерентности вейвлетов** следующим образом:

$$WCH_{x,y}^2(s, l) = \frac{|\langle W_x^*(s, l)W_y(s, l) \rangle|^2}{\langle |W_x(s, l)|^2 \rangle \langle |W_y(s, l)|^2 \rangle},$$

где $\langle \cdot \rangle$ обозначает локальную операцию сглаживания, как во временной шкале, так и в масштабной, при этом сглаживание выполняется на компонентах преобразования.

Методы кросс-вейвлет анализа используют при исследовании свойств нескольких временных рядов, зависимых между собой нетривиальным образом. Например, в геофизике возникает задача выявления причинно-следственных связей или корреляции между метеорологическими или другими явлениями окружающей среды, которые происходят на большом расстоянии друг от друга. В [Maraun, 2004] анализируются особенности применения кросс-вейвлет преобразования и оценки когерентности вейвлетов для двумерных временных рядов такого типа. Также методы вейвлет анализа, в том числе и кросс-вейвлет преобразование, были использованы в [Adamowski, 2008] для изучения метеорологических временных рядов и данных об уровне потока реках. Из временных рядов двух типов выделяли компоненты, которые далее использовались в модели прогнозирования наводнений. В работе показано, что использование кросс-вейвлет анализа полезно в том случае, когда существует относительно стабильный сдвиг фазы между потоковым и метеорологическим временными рядами. С помощью кросс-вейвлет преобразования определялась разность фаз между потоковыми и метеорологическими данными, что улучшило качество модели прогнозирования наводнений.

Другим примером является [Labat, 2010], где кросс-вейвлет анализ проводился для климатических индексов и показателей сброса пресной воды в Африке. В этом случае, кросс-вейвлет преобразование и оценка когерентности использовались для визуализации и анализа периодических колебаний в данных длиной в 2-8 лет. В [Kelly, 2003] подтверждается, что методы на основе кросс-вейвлет анализа могут быть эффективным инструментом в поиске квазипериодичности временного ряда.

Методы кросс-вейвлет анализа также нашли применение в медицине. В [Li, 2007] описывается использование кросс-вейвлет преобразования, оценки когерентности и некоторых других методов на основе вейвлетов для исследования динамики взаимодействия между колебаниями, генерируемыми двумя анатомически различными группами нейронов. Результаты исследования могут быть использованы для анализа и количественного определения временного взаимодействия между

нейронными осцилляторами, а также для исследования механизмов эпилепсии.

В [Aguilar-Contraria, 2008] используются инструменты кросс-вейвлет анализа, чтобы показать, что связь между переменными денежной политики и макроэкономическими переменными со временем изменилась, причем эти изменения не являются однородными на разных частотах.

Данные, полученные с помощью кросс-вейвлет преобразования, также могут использоваться как исходные данные для алгоритмов классификации. В [Dey, 2010] коэффициенты кросс-вейвлет преобразования подавались на вход искусственной нейронной сети и классификатора Fuzzy.

Дискретное вейвлет преобразование и приближение функций с помощью ряда

Дискретное вейвлет преобразование определяется таким образом, чтобы можно было полностью восстановить исходный сигнал, используя бесконечные суммы дискретных вейвлет коэффициентов. Такой подход также приводит к быстрому вычислению вейвлет преобразования и его обратного [Addison, 2017].

Пусть $x(t)$ принадлежит пространству 2π -периодических квадратично интегрируемых функций. Тогда $x(t)$ можно представить в виде ряда Фурье

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{int},$$

где коэффициенты c_n имеют вид

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} x(t) e^{-int} dt.$$

Набор функций $w_n(t) = e^{int}$ – это ортонормированный базис в пространстве $L^2(0, 2\pi)$, построенный с помощью масштабного преобразования $w_n(t) = w(nt)$ из базовой функции $w(t) = e^{int}$.

Пусть теперь $x \in L^2(\mathbb{R})$. Базисной функций в пространстве $L^2(\mathbb{R})$ должна быть функция, которая достаточно быстро убывает к 0 на $\pm\infty$. Поэтому, для построения базиса используются вейвлеты – хорошо локализованные солитоноподобные функции. Для того чтобы покрыть вейвлетами всю действительную ось, используют перенос вдоль оси. Для простоты можно использовать целые сдвиги k и аналоги синусоидальной частоты, как степени двойки $\psi_{jk} = \psi(2^j t - k)$. Вейвлет $\psi \in L^2(\mathbb{R})$ называется ортогональным, если семейство функций $\{\psi_{jk}\}$ образует ортонормированный базис в $L^2(\mathbb{R})$.

3.5. Корреляция с шаблоном

С помощью непрерывного вейвлет-преобразования выявляются участки исследуемого ряда, которые по форме наиболее похожи на вейвлет (Рис. 3.22). Идея состоит в том, чтобы сравнить части ряда с некоторым шаблоном на разных масштабах (Рис. 3.23). При этом вейвлет как функция должен обладать определенными математическими свойствами, в частности быстро убывать к нулю на бесконечности. В некоторых случаях полезно использовать шаблон, который не соответствует требованиям к вейвлету. Для этого вместо вейвлет-преобразования будем вычислять корреляцию между частью временного ряда и некоторым шаблоном p

$$C(l, k) = \frac{\sum_{i=1}^k (x_{l+i} - \bar{x})(p_i - \bar{p})}{\sqrt{\sum_{i=1}^k (x_{l+i} - \bar{x})^2 \sum_{i=1}^k (p_i - \bar{p})^2}}$$

Полученный коэффициент $C(l, k)$ зависит от значений x_{l+1}, \dots, x_{l+k} . То есть параметр l отвечает сдвигу шаблона, а параметр k соответствует количеству точек в шаблоне и в рассматриваемом отрезке ряда. Параметр k в данном случае является аналогом масштаба s , который использовали при вейвлет-преобразовании.

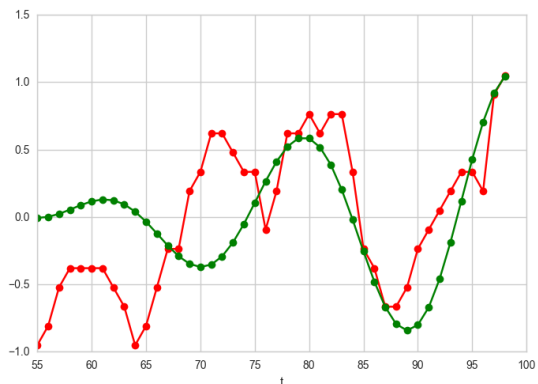


Рис. 3.22 – Отрезок временного ряда с наложенным шаблоном

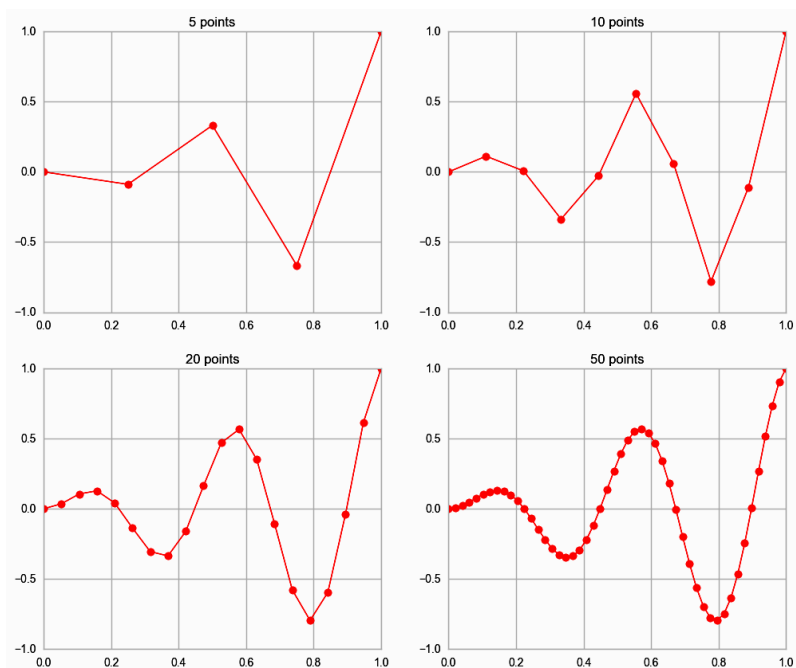


Рис. 3.23 – Шаблон «змея» с разным количеством точек.

Если при вычислении коэффициента вейвлет-преобразования всегда использовался весь временной ряд,

то в данном случае для вычисления $C(l, k)$ используются k точек ряда и шаблон длины k .

Полученные корреляционные коэффициенты $C(l, k)$ представим на графике, который похож на скейлограмму (Рис. 3.24).

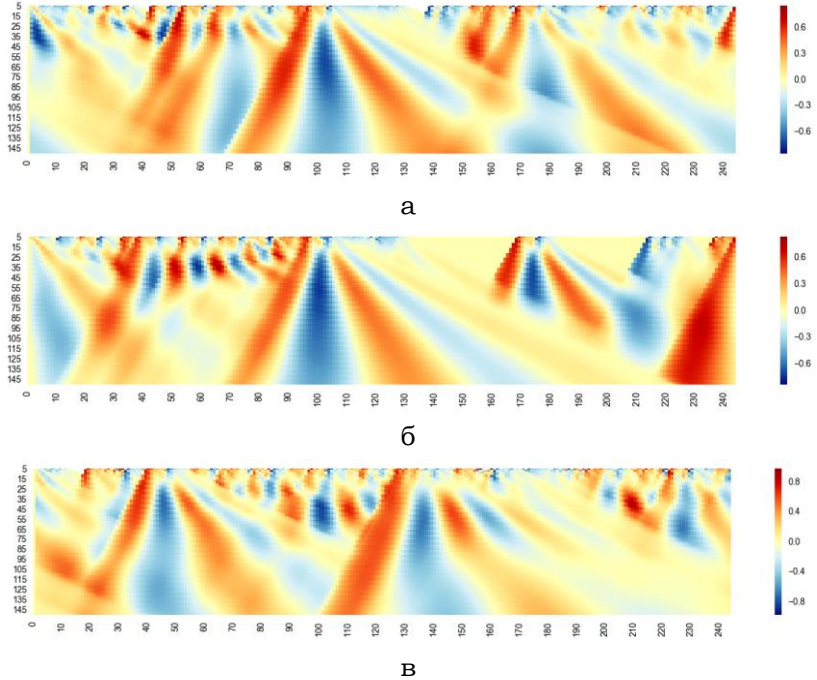


Рис. 3.24 – Корреляционные коэффициенты $C(l, k)$ вычисленные для рядов Т (а), К (б) и Х (в) с использованием шаблона, показанного на Рис. 3.22

3.6. Фрактальный анализ

Термин фрактал ввел и популяризировал Бенуа Мандельброт. Чаще всего фракталами называют геометрические объекты, которые имеют сильно изрезанную форму и обладают свойством самоподобия.

Строго и общепринятого определения фрактала в данный момент не существует, хотя Бенуа Мальдеброт использовал несколько пробных определений. Одно из них, введенное в [Mandelbrot, 1982], звучит так:

Фракталом называется множество, размерность Хаусдорфа-Безиковича которого строго больше его топологической размерности.

Строгое определение размерности Хаусдорфа-Безиковича или фрактальной размерности будет введено позже. Суть такого определения сводится к тому, чтобы выделить класс сильно изрезанных объектов, для описания которых недостаточно топологической размерности. Например, существуют кривые, топологическая размерность которых всегда равна 1, но они изогнуты таким сложным образом, что заполняют плоскость или пространство. Так кривые Пеано, проходят через любую точку единичного квадрата. Другой пример – траектория броуновской частицы, которая не является гладкой ни в одной точке.

Первое определение, хотя и является строгим, но исключает многие физические фракталы, и поэтому не используется. Было предложено следующее определения фрактала:

Фракталом называется структура, состоящая из частей, которые в каком-то смысле подобны целому.

Второе определение подчеркивает, что отличительным признаком для фрактала является самоподобие. Приведем строгое определение самоподобного множества, которое используется в математике. Для этого потребуется несколько предварительных определений.

Пусть множество $E \subset \mathbb{R}^d$ замкнуто. Тогда отображение $S: E \rightarrow E$ называется **отображением подобия** (similarity) на E , если $\exists t: 0 < t < 1: |S(x) - S(y)| = t|x - y|, \forall x, y \in E$.

То есть отображение подобия S превращает множество E в геометрически подобное множество.

Рассмотрим набор отображение подобия S_1, \dots, S_m . Множество $F \subseteq E$ является **инвариантным** относительно преобразований S_i , если

$$F = \bigcup_{i=1}^m S_i(F).$$

Множество, которое является инвариантным относительно набора отображений подобия, называется самоподобным.

Определение самоподобного множества можно понять интуитивно. Действительно, по определению множество самоподобно, если его можно «собрать» из кусочков, которые подобны целому множеству. Тогда простейшим примером самоподобного множества будет отрезок $[0,1]$. Возьмем, например $S_1 = \frac{x}{2}$ и $S_2 = \frac{1}{2} + \frac{x}{2}$. Тогда $[0,1] = S_1([0,1]) + S_2([0,1])$.

Очевидно, что простого самоподобия не достаточно для того, чтобы назвать объект фракталом. В самом деле, не будем же мы считать фракталами отрезок прямой или листок бумаги в клеточку. Мы будем рассматривать фрактальные объекты, которые обладают свойством самоподобия, а также сложной структурой.

Приведем один базовый простейший пример фрактального множества, который будет удобно использовать для демонстрации основных идей в дальнейшем. Это множество Кантора или канторова пыль. Классический процесс построения канторова множества начинается с единичного отрезка $C_0 = [0,1]$. Удалим из C_0 среднюю треть, останется множество $C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$. Множество C_1 состоит из двух отрезков, из каждого из которых теперь удалим среднюю треть, получим множество C_2 . Продолжая повторять эту процедуру, получим последовательность множеств $\{C_i\}_{i=1}^{\infty}$. Множество Кантора – это пересечение

$$C = \bigcap_{i=1}^{\infty} C_i.$$

Заметим, что множество C самоподобно. Возьмем отображения подобия $S_1(x) = \frac{x}{3}$ и $S_2(x) = \frac{x}{3} + \frac{2}{3}$, тогда $C = S_1(C) \cup S_2(C)$. С другой стороны известно, что мера Лебега множества Кантора равна 0, точно также как и для точки, или любого счетного множества. Но очевидно, что структура множества Кантора намного более сложная, и здесь уже возникает идея, что для описания такого

множества нужна специальная мера, которой и станет фрактальная размерность множества.

Фрактальная размерность

Вернемся к определению размерности Хаусдорфа-Безиковича. Для этого нам понадобится понятие покрытия множества.

Пусть U – непустое множество в \mathbb{R}^d . **Диаметр множества U** по определению равен

$$|U| = \sup\{|x - y|: x, y \in U\}.$$

Если $F \subset \bigcup_{i=1}^{\infty} U_i$ и $0 < |U_i| \leq \delta$ для любого i , то набор множеств $\{U_i\}$ называется **δ -покрытием** для множества U .

Пусть F – подмножество некоторого замкнутого множества в \mathbb{R}^d . Для произвольных $s \geq 0$ и $\delta > 0$ определим

$$\mathcal{H}_{\delta}^s(F) = \inf \left\{ \sum_i |U_i|^s : \{U_i\} - \delta \text{-покрытие для } F \right\},$$

где инфимум берется по всем возможным δ -покрытиям множества F . По определению s -размерная мера Хаусдорфа

$$\mathcal{H}^s(F) = \lim_{\delta \rightarrow 0} \mathcal{H}_{\delta}^s(F).$$

Такой предел существует для любого множества $F \subset \mathbb{R}^d$, с той оговоркой, что часто он равен нулю или бесконечности.

Размерность Хаусдорфа-Безиковича множества F определяется как

$$D_H(F) = \inf\{s: \mathcal{H}^s(F) = 0\} = \sup\{s: \mathcal{H}^s(F) = \infty\}.$$

Или, что то-же самое

$$\mathcal{H}^s(F) = \begin{cases} 0, & s < D_H(F); \\ \infty, & s > D_H(F). \end{cases}$$

Для примера вычислим размерность множества Кантора C . Выше была описана процедура построения, и согласно с ней на n -м шаге имеется 2^n отрезков длины $1/3^n$ каждый и далее множество только уменьшается. Поэтому в качестве

диаметра покрытия δ можно взять величину $1/3^n$ и использовать 2^n множеств в покрытии. По определению

$$\mathcal{H}^s(C) = \lim_{\delta \rightarrow 0} \mathcal{H}_\delta^s(C),$$

и теперь можно перейти от предела по $\delta \rightarrow 0$, к пределу $n \rightarrow \infty$

$$\lim_{\delta \rightarrow 0} \mathcal{H}_\delta^s(C) = \lim_{n \rightarrow \infty} \sum_{i=1}^{2^n} \left(\frac{1}{3^n}\right)^s.$$

Остается определить значение такого предела

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{2^n} \left(\frac{1}{3^n}\right)^s = \lim_{n \rightarrow \infty} \left(\frac{2}{3^s}\right)^n = \begin{cases} 0, & \frac{2}{3^s} < 1, \\ 1, & \frac{2}{3^s} = 1, \\ \infty, & \frac{2}{3^s} > 1. \end{cases}$$

Следовательно, предел $\mathcal{H}_\delta^s(C)$ при $\delta \rightarrow 0$ не равен нулю или бесконечности при $\frac{2}{3^s} = 1$, поэтому

$$D_H(C) = \frac{\ln 2}{\ln 3} \approx 0,63.$$

Статистически самоподобные процессы и показатель Херста

Многие объекты в окружающем нас мире статистически самоподобны (классический пример – береговые линии), это означает, что части таких объектов имеют одинаковые статистические характеристики при изменении масштаба. При изучении эволюции информационных потоков, структуры массивов документов в Интернет и исследовании процессов в информационном пространстве часто возникают самоподобные структуры, и в частности временные ряды.

Дадим определение самоподобного процесса.

Действительнозначный процесс $\{x(t), t \in \mathbb{R}\}$ является самоподобным с показателем Херста $H > 0$, если для всех $a > 0$ конечномерные распределения $\{x(at), t \in \mathbb{R}\}$

идентичны конечномерным распределениям $\{\alpha^H x(t), t \in \mathbb{R}\}$, что можно кратко записать

$$\{x(at), t \in \mathbb{R}\} =^d \{\alpha^H x(t), t \in \mathbb{R}\}.$$

То есть, по определению, для самоподобного процесса изменение временного масштаба эквивалентно изменению масштабу значений процесса. Это означает, что реализации такого процессы выглядят одинаково на разных масштабах. При этом, естественно, что процесс не является точной копией себя на разных масштабах, сохраняются только статистические свойства.

Показатель Херста представляет собой меру персистентности — склонности процесса к трендам. Значение $H = 0.5$ соответствует некоррелированному поведению значений ряда, как у броуновского движения. Значения в диапазоне $0.5 < H < 1$ означают, что направленная в определенную сторону динамика процесса в прошлом, вероятнее всего, повлечет продолжение движения в том же направлении. Если же $H > 0.5$, то прогнозируется, что процесс изменит направленность.

Опишем некоторые свойства самоподобных процессов, которые важны для приложений. Во-первых, у таких процессов автоковариационная функция гиперболически затухает, и имеет вид

$$\rho_k \approx k^{(2H-2)} L(t) \text{ при } k \rightarrow \infty,$$

где $L(t)$ — медленно меняющаяся на бесконечности функция, то есть такая, что

$$\forall x > 0: \lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1.$$

Следовательно, для самоподобных процессов ряд из ковариационных коэффициентов расходится

$$\sum_{k=1}^{\infty} \rho_k = \infty.$$

Такая бесконечная сумма говорит о долговременной зависимости в ряде.

Во-вторых, дисперсия выборочного среднего убывает медленнее, чем величина, обратная размеру выборки

$$\sigma^2 \left(x_t^{(m)} \right) \sim m^{2H-2},$$

где последовательность $\{x_t^{(m)}\}$ получили, разбив исходную последовательность $\{x_t\}$ на непересекающиеся блоки длины m и взяв среднее в каждом из блоков.

Методы оценивания показателя Херста

Метод оценивания показателя Херста, предложенный им самим, называется методом нормированного размаха или R/S анализом. Для временного ряда $\{x_t\}_{t=1}^T$ стандартное отклонение S определяется по формуле

$$S = \sqrt{\frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^2}, \quad \text{где } \bar{x} = \frac{1}{T} \sum_{t=1}^T x_t,$$

а величина размаха ряда

$$R = \max_{1 \leq t \leq T} x^{(t)} - \min_{1 \leq t \leq T} x^{(t)}, \quad \text{где } x^{(t)} = \sum_{i=1}^t (x_i - \bar{x}).$$

Отношение R/S и есть нормированным размахом. Херст обнаружил, что для многих наблюдаемых временных рядов нормированный размах хорошо описывается эмпирическим соотношением

$$\frac{R}{S} = \left(\frac{T}{2} \right)^H.$$

Значения показателя Херста можно оценить, если вычислить значения статистики R/S в зависимости от T и построить график такой зависимости в двойной логарифмической шкале. Оценкой показателя Хёрста будет оценка наклона прямой, которая наилучшим образом аппроксимирует зависимость $\log R/S$ от $\log T$.

Используем метод R/S для вычисления показателя Хёрста для рядов Т, К и Х. На Рис. 3.25 показаны результаты оценивания для рядов Т и К. Полученные значения показателя Хёрста – 0.62 и 0.68 соответственно – свидетельствуют о склонности данных процессов к трендам, хотя и не очень высокую.

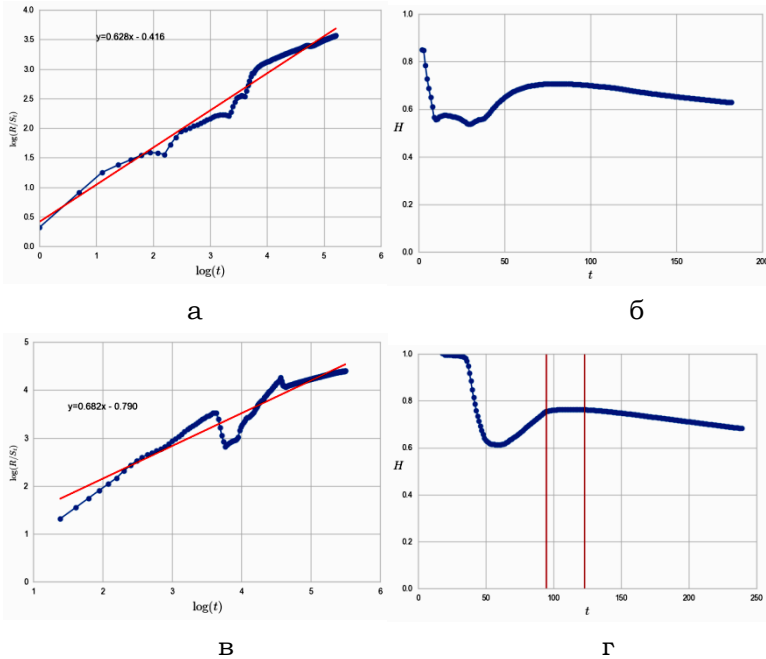


Рис. 3.25 – Оценивание показателя Херста для рядов Т и К. Зависимость статистики R/S для ряда Т (а) и ряда К (в) от времени в логарифмической шкале. Зависимость показателя Херста от времени для ряда Т (б) и ряда К (г).

В случае ряда К на Рис. 3.25 видно, что зависимость $\log R/S$ от $\log t$ плохо аппроксимируется линейной зависимостью, так как график имеет сильный излом. Если построить зависимость показателя Хёрста от времени (Рис. 3.26), то можно определить момент времени, начиная с которого значение показателя начинает убывать. Отметив этот момент времени на графике временного ряда X , можно увидеть, что это момент резкого возрастания значений ряда, до которого значения ряда имели значительно меньшую дисперсию.

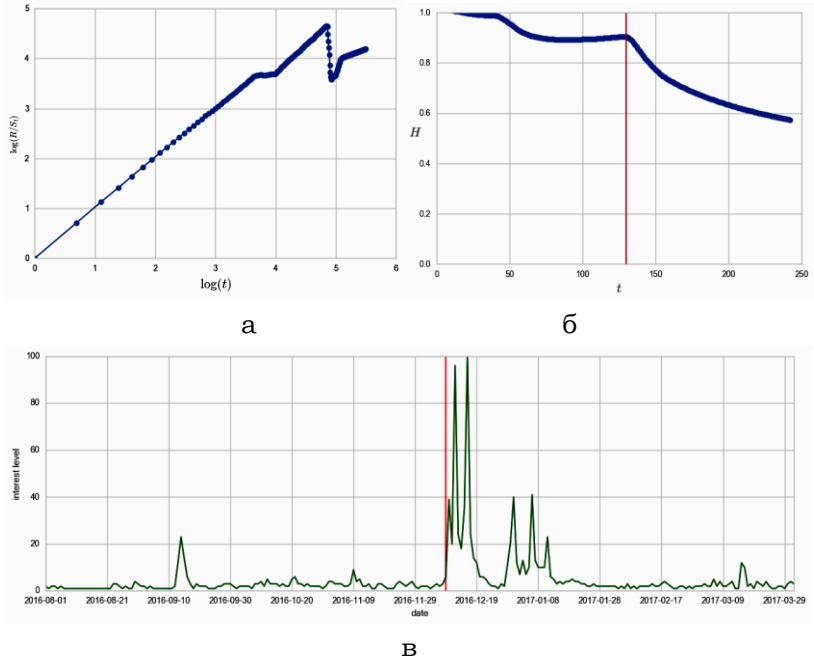
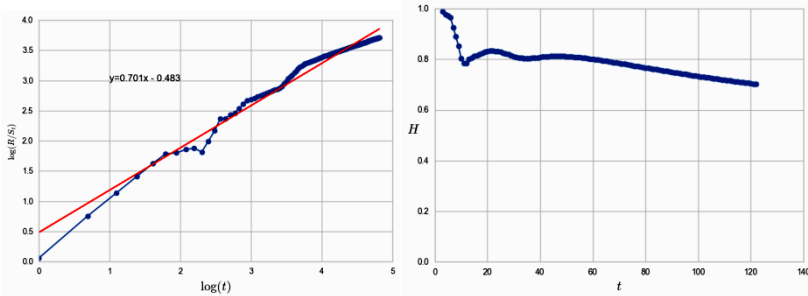


Рис. 3.26 – Оценивание показателя Хёрста для ряда X .
 Зависимость статистики R/S от времени в логарифмической шкале (а). Зависимость показателя Хёрста от времени (б). Вертикальной линией отмечен момент времени, после которого значение показателя Хёрста начинает убывать. Временной ряд X с отмеченным моментом времени, после которого происходит изменение показателя Хёрста (в)

Поведение ряда X , начиная с начала декабря 2016 (начало наибольшего пика в значениях ряда) можно рассмотреть отдельно. Оценка показателя Хёрста для второго участка ряда показана на Рис. 3.27. При этом стоит учесть, что в данном примере временной ряд становится слишком коротким, так как для R/S анализа используют ряды с не менее чем 200 элементами. Тем не менее, резкие изменения в зависимости показателя Хёрста от времени, которые имеют вид «ступеньки», свидетельствуют о том, что исследуемый процесс состоит из различных процессов, которые имеет смысл рассмотреть отдельно.



а б

Рис. 3.27 – Оценивание показателя Херста для второй части ряда X . Зависимость статистики R/S от времени в логарифмической шкале (а). Зависимость показателя Хёрста от времени (б).

ΔL-метод

Скейлограммы, полученные с помощью непрерывного вейвлет-преобразования, используют для визуализации особенностей временного ряда. В [Ландэ, 2009] предложен другой метод визуализации, который также помогает выявить тренды, периодичности и локальные особенности ряда. Предложенный подход значительно проще в реализации, чем вейвлет анализ.

Метод, который авторы назвали Δ -метод, базируется на методе DFA (Detrended Fluctuation Analysis), который также будет рассмотрен далее. Суть подхода состоит в определении и отображении абсолютного отклонения точек ряда накопленных значений от соответствующих значений линейной аппроксимации.

Опишем Δ -метод более подробно. Для начала зафиксируем некоторую ширину окна s (масштаб на котором рассматривается ряд). Рассмотрим точку x_l и выберем для нее окно ширины s так, чтобы точка l была в центре этого окна (или смещена на 1, если s чётное). Построим линейную аппроксимацию по точкам окна и обозначим $L_{l,j,s}$ значение локальной аппроксимации в точке j для отрезка с центром в l . Далее вычислим абсолютное отклонение x_j (Рис. 3.28) от линии аппроксимации $\Delta_{l,j,s} = |x_j - L_{l,j,s}|$.

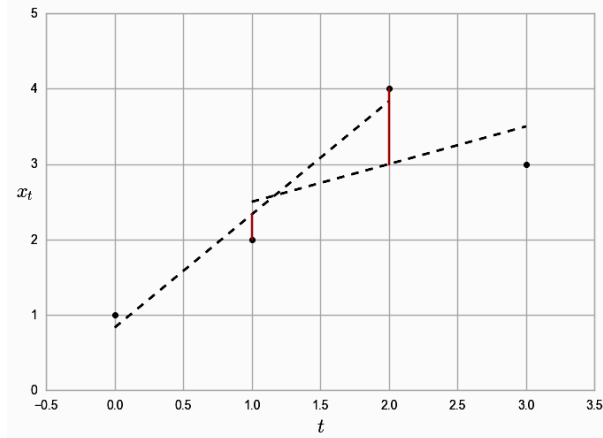
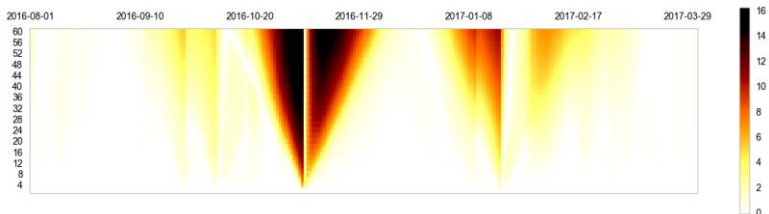


Рис. 3.28 – Четыре точки временного ряда с линейной аппроксимацией для двух окон с шириной три. Также показано отклонение $\Delta_{t,j,s}$ центральной точки окна от соответствующей линейной аппроксимации.

Метод предполагает вычисление значений $\Delta_{l,j,s}$ для всех точек $l = 1, \dots, T$ и окон шириной $s = 1, \dots, [T/4]$. Для фиксированной ширины окна вычисляется среднее квадратичное отклонение

$$E(l, s) = \sqrt{\frac{1}{s} \sum_j |x_j - L_{l,j,s}|^2} = \sqrt{\frac{1}{s} \sum_j \Delta_{l,j,s}^2}.$$

Далее полученные значения демонстрируются на диаграмме, похожей на скейлограмму. Примеры таких диаграмм показаны на Рис. 3.29.



а

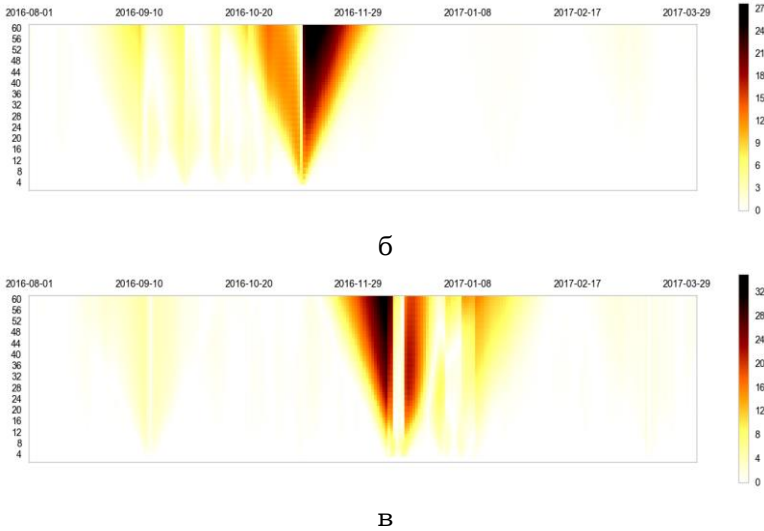


Рис. 3.29 – Коэффициенты полученные с помощью ΔL -метода для рядов Т (а), К (б) и X (в).

Предложенный метод визуализации абсолютных отклонений ΔL , как и метод вейвлет-преобразований, позволяет выявлять единичные и нерегулярные «всплески», резкие изменения значений количественных показателей в разные периоды времени, а также гармонические составляющие в ряде.

3.7. Мультифрактальный анализ

Для описания самоподобных объектов, которые возникают в природе, часто не хватает одной фрактальной размерности, так как во многих случаях такие объекты не являются однородными. Наиболее общее описание природы таких объектов дает теория мультифракталов, согласно которой объект характеризуется бесконечной иерархией размерностей, что позволяет отличить однородные объекты от неоднородных.

Мультифрактальное множество (сигнал) можно понимать как некое объединение различных однородных фрактальных подмножеств (сигналов), каждое из которых имеет собственное значение фрактальной размерности. Значения таких фрактальных размерностей отображаются

в мультифрактальном спектре, формальное определение которого будет введено позже. Важно, что мультифрактальный спектр может использоваться как мера подобия. Такой подход может использоваться, например, для формирования репрезентативных выборок из массивов документов, как дополнение традиционных методов, базирующихся на выявлении содержательного подобия документов. Практические приложения такого подхода: предъявление пользователю обозримых результатов поиска, отражающих весь спектр документального массива или выделение подмножеств документов для дальнейших исследований [Ландэ, 2009a; Ландэ, 2009b].

Для того, чтобы детально рассмотреть идею мультифрактального множества и мультифрактального спектра понадобится несколько дополнительных понятий и определений. Будем использовать в качестве примера обобщенное множество Кантора, а также меру на этом множестве. Мультифрактальный спектр в этом примере появляется достаточно естественным и простым образом, поэтому помогает получить интуитивное понимание. Также для анализа мультифрактальности ключевое значение имеет показатель Гёльдера, определение которого будет введено чуть позже.

Начнем с построение обобщенного множества Кантора и определения на нем меры. Классическое множество Кантора можно обобщить несколькими способами. Например, можно использовать функцию сжатия $F_i(x) = rx + (1 - r)i$, где $i = 0, 1$ и $r \in \left(0, \frac{1}{2}\right)$, вместо $F_i(x) = \frac{1}{3}x + \frac{2}{3}i$. Обозначим такое множество $C(r)$. На обобщенном множестве Кантора можно вести равномерно распределенную меру, а можно определить p -меру с такой же функцией сжатия, но вероятностями p и $1 - p$, где $p \in [0, 1]$.

Более интересным обобщением является множество Кантора с переменными коэффициентами разделения. Пусть есть последовательность $\{r_j\}$, $r_j \in \left(0, \frac{1}{2}\right)$. Множество будем строить с помощью следующей итерационной процедуры. Пусть $C_0 = [0, 1]$. Удалим из середины отрезка $[0, 1]$ открытый отрезок длины $1 - 2r_1$. Останется два

замкнутых отрезка длины r_1 . Объединение этих двух отрезков обозначим C_1 . На j -ой итерации множество C_j будет состоять из объединения 2^j замкнутых отрезков длины $r_1 \cdot r_2 \cdot \dots \cdot r_j$. Таким образом, получим множество

$$C(\{r_j\}) = \bigcap_{j=1}^{\infty} C_j.$$

Замкнутые отрезки, которые появляются при итерационной конструкции множества Кантора, можно закодировать с помощью конечных слов из алфавита $\{0,1\}$. На первом шаге получаем левый отрезок I_0 и правый отрезок I_1 . На n -ом шаге название отрезка w имеет длину n и на следующем шаге отрезок I_w будет разделен на отрезки I_{w0} и I_{w1} .

На множестве $C(\{r_j\})$ можно ввести p -меру, которая будет обладать следующим свойством

$$\mu(I_{w0}) = p\mu(I_w), \quad \mu(I_{w1}) = (1-p)\mu(I_w).$$

Еще более обобщенную меру на $C(\{r_j\})$ получим, если использовать последовательность весов $\{p_j\}, p_j \in [0,1]$ и определить меру с помощью следующего правила $\mu(I_{w_1 w_2 \dots w_n}) = p_{w_1 1} \cdot p_{w_2 2} \cdot \dots \cdot p_{w_n n}$, где $p_{0j} = p_j$, а $p_{1j} = 1 - p_j$.

Для того чтобы сделать некоторые выводы о структуре меры понадобится показатель Гёльдера. Сначала рассмотрим определение и смысл показателя Гёльдера для функций и мер, а потом применим их к построенному только что обобщенному множеству Кантора и мере на нём.

Показатель Гёльдера и мультифрактальный анализ для мер

Характеристикой гладкости функции является показатель Гёльдера, который содержит информацию о поведении функции в окрестности точки. Чем меньше значение показателя Гёльдера, тем менее гладкой является функция.

Пусть x – ограниченная функция на \mathbb{R} и $t_0 \in \mathbb{R}$, тогда **локальный показатель Гёльдера** функции x в точке t_0 определяется как

$$h_x(t_0) = \sup_{\Delta t \rightarrow 0} \{\alpha \geq 0: |x(t + \Delta t) - x(t)| = O(\Delta t^\alpha)\}.$$

Другими словами, локальный показатель Гёльдера характеризует поведение функции в окрестности точки следующим образом

$$|x(t + \Delta t) - x(t)| \sim \Delta t^{h_x(t)}.$$

Последнее соотношение стоит сравнить с похожим соотношением для монофрактальных процессов

$$|x(t + \Delta t) - x(t)| \sim \Delta t^H,$$

где H – это показатель Херста. То есть для мультифрактальных процессов локальный показатель Гёльдера $h_x(t)$ по сути является «локальным показателем Херста», который может меняться в зависимости от t .

Для того чтобы измерить регулярность меры в окрестности точки также вводят показатель Гёльдера.

Показатель Гёльдера или локальная размерность меры μ на \mathbb{R} определяется следующим образом

$$h_\mu(x) = \lim_{r \rightarrow 0^+} \frac{\log[\mu(B(x, r))]}{\log r},$$

где $B(x, r)$ – шар с центром в точке x и радиусом r .

Вернемся к примерам с множеством Кантора. Для равномерной распределенной меры на классическом множестве Кантора:

$$h_\mu(x) = \frac{\log 2}{\log 3}, x \in C.$$

Если же мы рассмотрим множество $C(\{r_i\})$ с мерой μ , определенной с помощью весов p и $1 - p$, то локальная размерность меры будет отличаться в разных точках. Например,

$$h_\mu(0) = \lim_{j \rightarrow \infty} \frac{\log p^j}{\log \prod_{i=1}^j r_i} = \frac{\log p}{\lim_{j \rightarrow \infty} \frac{1}{j} \sum_{i=1}^j \log r_i} = \frac{\log p}{\log r_0}, h_\mu(1) = \frac{\log(1-p)}{\log r_0},$$

где ввели обозначение $\lim_{j \rightarrow \infty} \frac{1}{j} \sum_{i=1}^j \log r_i = \log r_0$. Если предположить, что $\sup_i r_i < 1/2$, и без ограничения общности считать, что $p < 1 - p$, то можно показать, что

$$h_\mu(x) \in \left[\frac{\log p}{\log r_0}, \frac{\log(1-p)}{\log r_0} \right], x \in C(\{r_i\}).$$

(доказательство [Aldroubi 2016]). Таким образом, для p -Канторовой меры μ известны возможные значения локальной размерности. Для того чтобы описать некоторые свойства меры μ нужно рассмотреть множества уровня значений локальной размерности, а именно множества вида

$$E_h = \{x \in \mathbb{R} : h_\mu(x) = h\}.$$

Далее можно сравнить размеры множеств E_h при разных значениях h . Во многих практически важных случаях для сравнения таких множеств нужно будет использовать фрактальную размерность. Таким образом, приходим к определению мультифрактального спектра.

Мультифрактальным спектром меры μ на \mathbb{R} называется отображение $d_\mu(h) = D_H(E_h)$.

То есть с помощью мультифрактального спектра отображается, какие значения показателя Гёльдера присутствуют в неоднородном объекте (мере, множестве, сигнале), и в каком соотношении между собой. Каждому значению показателя Гёльдера соответствует фрактальная размерность множества точек, в которых значение показателя Гёльдера равно данному (Рис. 3.30).

Для p -Канторовой меры μ можно показать, что (доказательство [Aldroubi 2016])

$$d_\mu(h) = \inf_{q \in \mathbb{R}} \left(qh - \frac{\log(p^q + (1-p)^q)}{\log r_0} \right).$$

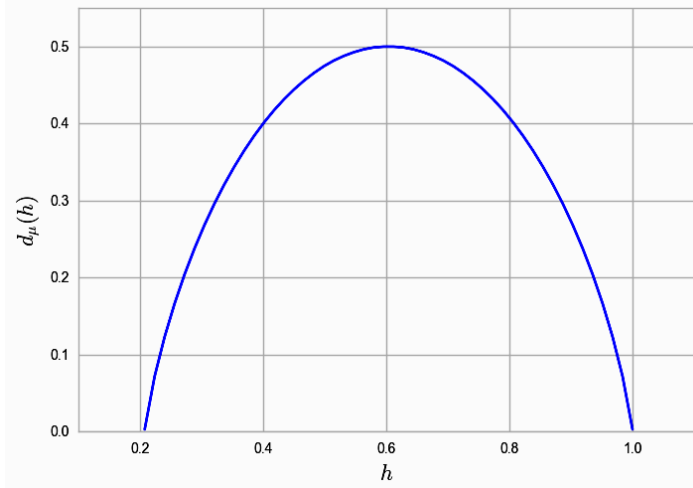


Рис. 3.30 – Мультифрактальный спектр для p -Канторовой меры.

Подход к оцениванию мультифрактального спектра

Выше был описан теоретический подход к определению мультифрактального спектра для меры. Для практических целей прямое вычисление показателя Гельдера в каждой точке и вычисление фрактальных размерностей множеств уровня данного показателя не осуществимо. Ключ к практическому подходу дает следующее определение структурной функции (structure function) меры μ

$$Z(q, j) = \frac{1}{2^j} \sum_i \mu \left(B \left(\frac{i}{2^j}, \frac{1}{2^{j+1}} \right) \right)^q,$$

где сумма берется только по таким отрезкам, где мера не равна 0. Также определим масштабную функцию (scaling function)

$$\tau(q) = \liminf_{j \rightarrow +\infty} \left(\frac{\log(Z_\mu(q, j))}{\log 2^{-j}} \right).$$

Известно, что для того, чтобы покрыть множество E_h нужно приблизительно $2^{-d_\mu(h)j}$ шаров и из определения

$h_\mu(x) = \lim_{j \rightarrow \infty} \frac{\log[\mu(B(x, 2^{-j}))]}{\log 2^{-j}}$ следует, что $\mu(B(x, 2^{-j})) \sim 2^{-h_\mu(x)j}$, поэтому масштабную функцию можно оценивать следующим образом

$$Z(q, j) \sim 2^{-j} \sum 2^{-h_\mu q j} \sim 2^{-j} 2^{-d_\mu(h)j} 2^{-hqj} = 2^{-(1+d_\mu(h)+hq)j}.$$

С другой стороны, масштабную функцию определили таким образом, что $Z(q, j) \sim 2^{-\tau(q)j}$, поэтому

$$2^{-(1+d_\mu(h)+hq)j} = 2^{-\tau(q)j}.$$

И при $j \rightarrow \infty$

$$d_\mu(h) = \inf_{q \in \mathbb{R}} (1 - \tau(q) + hq).$$

Таким образом, получили выражения для мультифрактального спектра через масштабную функцию. Данный подход позволяет численно определять мультифрактальный спектр для временных рядов. Сначала определяется структурная функция, с помощью нее – масштабная функция, и далее через преобразование Лежандра происходит переход к мультифрактальному спектру.

Мультифрактальный анализ для функций

Для функций справедливы определения и утверждения аналогичные соответствующим определениям и утверждениям для мер.

Мультифрактальным спектром локально ограниченной функции $x: \mathbb{R} \rightarrow \mathbb{R}$ называется отображение

$$d_x: \mathbb{R}^+ \cup \{+\infty\} \rightarrow [0; d] \cup \{-\infty\},$$

такое, что

$$d_x(h) = D_H(E_x(h))$$

(принимается, что $D_H(\emptyset) = -\infty$).

Для функций справедливо следующее утверждение о показателе Гельдера, которое похоже на введенное определение показателя Гельдера для меры. Вспомним, что

$$|x(t + \Delta t) - x(t)| \sim C_t \Delta t^{h_x(t)},$$

тогда из этого выражения не сложно вывести следующее утверждение.

Лемма. Пусть $x: \mathbb{R} \rightarrow \mathbb{R}$ ограниченная функция, для которой $h_x(t) = H \in [0,1]$, тогда

$$h_x(t) = \liminf_{j \rightarrow +\infty} \left(\frac{\log \left(R_x \left(B(t, 2^{-j}) \right) \right)}{\log 2^{-j}} \right),$$

где $R_x(A) = \max_{t \in A} x(t) - \min_{t \in A} x(t)$ – размах функции x на множестве A .

Из леммы следует, что мультифрактальный формализм для функций может быть основан на структурной функции

$$Z(q, j) = \frac{1}{2^j} \sum_i R_x \left(B \left(\frac{i}{2^j}, \frac{1}{2^{j+1}} \right) \right)^q \quad (1)$$

и соответствующей масштабной функции

$$\tau(q) = \lim_{j \rightarrow +\infty} \inf_{+} \left(\frac{\log(Z_x(q, j))}{\log 2^{-j}} \right) \quad (2)$$

что приводит к определению мультифрактального спектра

$$d_\mu(h) = \inf_{q \in \mathbb{R}} (1 - \tau(q) + hq).$$

Мультифрактальные процессы

Стохастический процесс называется мультифрактальным, если он обладает стационарными приращениями и удовлетворяет равенству

$$E[|x(t)|^q] = c(q)t^{\tau(q)+1},$$

для некоторого положительного q , где $\tau(q)$ – масштабная функция.

Если масштабная функция $\tau(q)$ линейно зависит от q , то процесс называется монофрактальным. Если процесс $x(t)$ самоподобный с показателем Херста H , то $\tau(q) = Hq - 1$.

Метод DFA и его применение к оцениванию мультифрактального спектра

В [Peng, 1994] предложен метод Detrended Fluctuation Analysis (DFA) для определения длительных корреляций в зашумленных и нестационарных временных рядах. Ключевая особенность метода DFA состоит в том, что он основан на теории случайных блужданий. Временной ряд

не анализируется в исходном виде, вместо этого выполняется центрирование ряда и переход к накопленным суммам

$$y_t = \sum_{k=1}^t x_k.$$

В таком случае можно рассматривать y_t как положение случайного блуждания после t шагов. Далее метод DFA предполагает анализ среднеквадратического отклонения значений ряда от тренда на различных непересекающихся кусках ряда.

Для метода DFA было предложено множество модификаций, а также вариантов применения для различных практических задач. Обзор таких методов приводится, например, в [Kantelhardt 2009]. Важным шагом стала разработка подхода к численному оцениванию мультифрактального спектра на основе метода DFA. Такой метод называется Multifractal Detrended Fluctuation Analysis (MF-DFA) и был предложен в [Kantelhardt 2002]. Эффективность метода MF-DFA была проанализирована для различных модельных временных рядов (броуновское движение, дробное броуновское движение, биномиальные каскады) [Oswiecimka 2012]. Также метод активно используют для анализа реальных временных рядов, часто экономических [Suarez-Garcia 2013].

Подробное пошаговое описание алгоритма MF-DFA приведено в [Thompson 2016]. Опишем все эти шаги.

Шаг 1. Приведение временного ряда к агрегированному виду. Различают агрегированные и дисагрегированные наборы данных. Пример агрегированных данных – количество новых сообщений в сети Интернет по некоторой теме в день. Соответственные дисагрегированные данные – приращение количества сообщений по сравнению с предыдущим днем. Если исходный временной ряд $\{z_t\}_{t=1}^{T+1}$ агрегированный, то нужно перейти к дисагрегированному $\{y_t = z_{t+1} - z_t\}_{t=1}^T$. Временной ряд, который будет использоваться в алгоритме, при центрировании и подсчете накопленных сумм – следующий:

$$x_t = \sum_{k=1}^t (y_k - \bar{y}), \quad t = 1, \dots, T.$$

Данный этап обработки ряда необходим для корректной работы метода, так как он основан на теории случайных блужданий.

Шаг 2. Зададим множество $\mathcal{S} = \{3, \lfloor N/4 \rfloor\}$. Для каждого значения $s \in \mathcal{S}$ разделим временной ряд $\{x_k\}_{k=1}^N$ на $N_s = \lfloor \frac{N}{s} \rfloor$ непересекающиеся части длины s . Если N не делится нацело на s , то нужно повторить процедуру, начиная с другой стороны временного ряда, и в результате получится $2N_s$ частей.

Шаг 3. Для каждого значения $j = 1, \dots, N_s$ j -ая часть временного ряда состоит из значений $\{x_{(j-1)s+i}\}_{i=1}^s$, и аналогично при $j = N_s + 1, \dots, 2N_s$ — $\{x_{N-(j-N_s)s+i}\}_{i=1}^s$. Для каждой j -ой части временного ряда нужно определить тренд $X_j(i)$. Во многих случаях достаточно использовать линейную аппроксимацию значений ряда, полученную с помощью метода наименьших квадратов. Для достаточно длинных рядов используют полиномиальную аппроксимацию степени m (далее будет приведен алгоритм подбора оптимального параметра m). Теперь вычислим среднеквадратичное отклонение части временного ряда от тренда

$$F^2(j, s) = \frac{1}{s} \sum_{i=1}^s [x_{(j-1)s+i} - X_j(i)]^2, \quad \text{при } j = 1, \dots, N_s,$$

и аналогично при $j = N_s + 1, \dots, 2N_s$.

Шаг 4. Обозначим \mathcal{Q} множество значений порядков моментов. Множество \mathcal{Q} должно содержать 0, положительные и отрицательные значения. Обычно выбирают симметричное относительно 0 множество. Для фиксированных значений $s \in \mathcal{S}$ и $q \in \mathcal{Q}$ вычислим норму l_q для вектора, состоящего из оцененных дисперсий $\{F^2(j, s): j = 1, \dots, 2N_s\}$

$$F_q(s) = \left(\frac{1}{2N_s} \sum_{j=1}^{2N_s} [F^2(j, s)]^{q/2} \right)^{1/q}, \quad q \in \mathcal{Q} \setminus \{0\},$$

$$F_0(s) = \exp \left\{ \frac{1}{4N_s} \sum_{j=1}^{2N_s} \ln[F^2(j, s)] \right\}.$$

Шаг 5. Для каждого $q \in \mathcal{Q}$ нужно выполнить линейную аппроксимацию зависимости $\ln[F_q(s)]$ от $\ln(s)$. При этом наклон полученной линейной функции – это оценка для $h(q)$. Масштабную функцию $\tau(q)$ получаем из выражения $\tau(q) = qh(q) - 1$.

Шаг 6. Оценим производную полученной оценки функции $\tau(q)$

$$\alpha_0 = \left. \frac{d\tau(q)}{dq} \right|_{q=q_0}, \quad q_0 \in \mathcal{Q},$$

и в результате получаем оценку мультифрактального спектра

$$f(\alpha) = q_0 \alpha_0 - \tau(q_0).$$

Примеры масштабных функций и мультифрактальных спектров для рядов T, K и X полученные с помощью метода MF-DFA приведены на Рис. 3.31.

Для корректной работы приведенного алгоритма нужно заранее выбрать множество \mathcal{S} длин отрезков, на которые делится ряд. Выясним, каким должно быть множество \mathcal{S} . С одной стороны, если использовать значения $s \geq N/4$, то мы разделим ряд на малое количество частей, и оценка $F_q(s)$ будет посчитана по малому количеству оценок дисперсий $F^2(j, s)$. С другой стороны, если использовать полиномиальную регрессию, то при $s \leq 10$ на шаге 3 будет использоваться мало точек для полиномиальной регрессии. Таким образом, разумное ограничение: $10 \leq s \leq N/4$. При этом в случае линейной регрессии, можно использовать и небольшие значения $s = 3, 4, \dots$

Для длинных временных рядов, которые содержат более двух тысяч значений, рекомендуется использовать ограничения

$$s_{min} = \max\{10, N/100\}, \quad s_{max} = \min\{20s_{min}, N/10\},$$

и выбирать шаг таким образом, чтобы во множестве \mathcal{S} было не более 100 значений. Также степень m нужно выбрать таким образом, чтобы регрессионный полином $X_j(i)$ адекватно отображал тренд в каждом куске временного ряда.

В случае коротких временных рядов, таких как T, K и X достаточно использовать линейную регрессию.

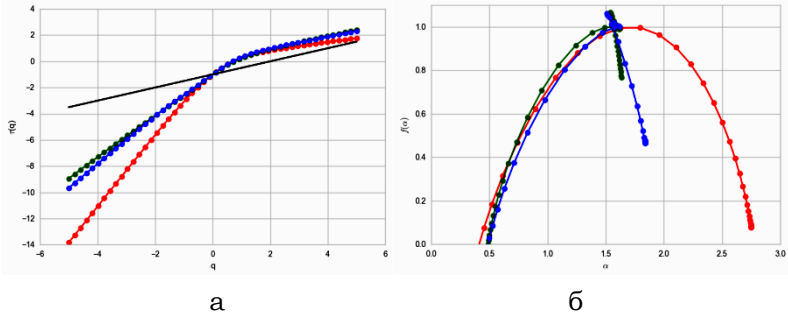


Рис. 3.31 – Масштабные функции (а) и мультифрактальные спектры (б) для рядов T, K и X полученные с помощью метода MF-DFA

Фильтрация

При оценивании мультифрактального спектра периодическая компонента в ряде может существенно повлиять на результат. Влияние синусоидального тренда на оценку спектра показано в [Nagarajan 2004]. Также предложен сглаживающий фильтр для минимизации влияния тренда, который при этом не влияет на мультифрактальные свойства ряда. Стоит заметить, что обычные методы сглаживания ряда (например, простое скользящее среднее) существенно влияют на мультифрактальные свойства.

Рассмотрим сглаживающий фильтр:

1. Вычислим преобразование Фурье для исследуемого ряда x_t и обозначим его $f(v)$.

2. Предположим, что синусоидальный тренд соответствует частоте v_k . Определим модифицированное преобразование Фурье

$$|\tilde{f}(v_k)| = 0.5(|f(v_{k-1})| + f(v_{k+1}))|$$

$$|\tilde{f}(v)| = |f(v)|, v \neq v_k$$

3. С помощью обратного преобразования Фурье перейдем к отфильтрованному ряду \tilde{x}_t

Продемонстрируем использование такого фильтра на некотором временном ряде x_t , а также его модифицированной версии $y_t = x_t + s_t$, где s_t – синусоида (Рис. 3.32). Вычислим преобразование Фурье для ряда y_t и обозначим его $f(v)$. Преобразование Фурье для рядов x_t , y_t , а также \tilde{y}_t отличается в одной точке $v_k = 0.01$, которая соответствует частоте колебания s_t (Рис. 3.34).

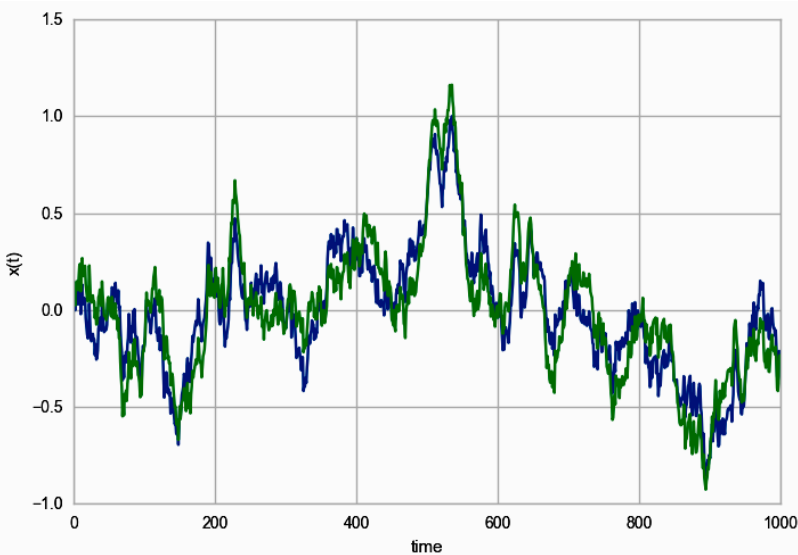


Рис. 3.32 – Мультифрактальный временной ряд (синий) и тот же самый ряд с наложенным синусоидальным трендом (зеленый).

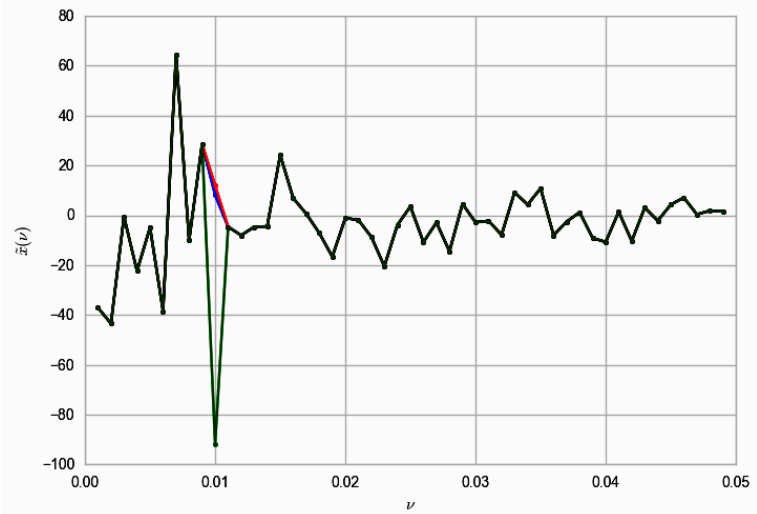


Рис. 3.33 – Преобразование Фурье для исходного ряда x_t (синий), ряда $y_t = x_t + s_t$ (зеленый) и отфильтрованного ряда \tilde{y}_t (красный)

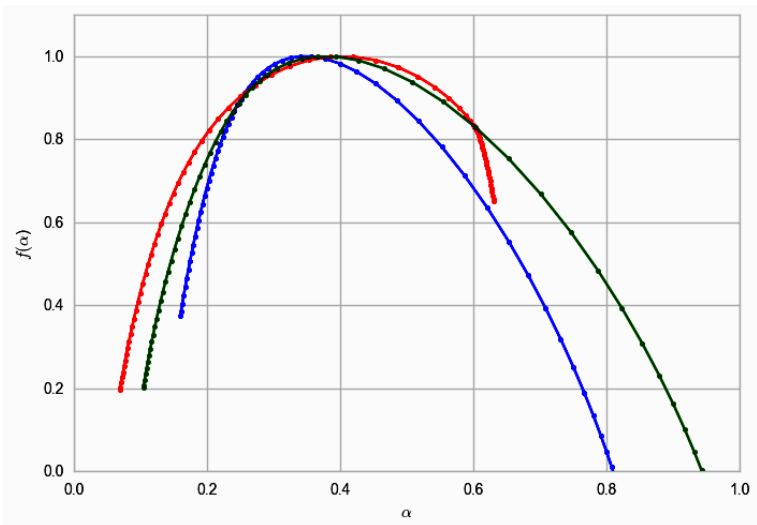


Рис. 3.34 – Мультифрактальные спектры для исходного ряда x_t (синий), ряда $y_t = x_t + s_t$ (красный) и отфильтрованного ряда \tilde{y}_t (зеленый)

Использование вейвлетов для оценивания мультифрактального спектра

Для исследования фрактальных характеристик объекта естественным инструментом являются вейвлеты [Aldroubi, 2016]. В первую очередь потому, что точечный показатель Гёльдера можно оценить через вейвлет-коэффициенты. Справедливо следующее утверждение, которое связывает вейвлет-преобразования с показателем Гёльдера.

Утверждение. Пусть функция x в точке t имеет показатель Гёльдера $h_x(t)$. Предположим также, что в точке t у функции x нет осциллирующей особенности (the singularity is not oscillating). Тогда

$$W_\psi[x](s, t) \sim s^{h_x(t)}, \quad s \rightarrow 0^+,$$

при условии, что у вейвлета ψ первые n моментов равны 0 и $n > h(t_0)$. С другой стороны, если выбрать вейвлет такой, что $n < h(t_0)$, то

$$W_\psi[x](s, t) \sim s^n, \quad s \rightarrow 0^+.$$

Из утверждения следует, что можно охарактеризовать поведение функции x в окрестности точки t следующим образом: чем более гладкой является функция, тем быстрее убывает значение вейвлет-преобразования $W_\psi[x](s, t)$ при уменьшении масштаба s . Например, если функция x является непрерывно дифференцируемой в точке t (это означает, что $h_x(t) = +\infty$), то вейвлет-преобразования введет себя по вышеуказанной формуле. То есть значения $W_\psi[x](s, t)$ зависят только от формы вейвлета (параметра n). Принципиально другой случай получаем, если функция x в точке t имеет показатель Гёльдера в пределах от 0 до 1, что часто встречается на практике. Тогда асимптотическое поведение вейвлет-коэффициентов задается соответствующим выражением (см. выше), т.е. зависит от показателя Гёльдера, для произвольного вейвлета с нулевым первым моментом. Указанная связь между асимптотическим поведением коэффициентов вейвлет преобразования и показателем Гёльдера лежит в основе

методов оценивания мультифрактального спектра с помощью вейвлетов.

Простой и наивный способ оценить мультифрактальный спектр, используя вышеуказанные выражения, может начинаться с определения структурной функции следующим образом

$$Z(q, s) = \int |W_\psi[x](s, t)|^q dt, \quad q \in \mathbb{R}.$$

Проблема такого определения состоит в том, что вейвлет-преобразование может принимать очень маленькие по модулю значения и обращаться в 0, соответственно при возведении в отрицательную степень q можно получить очень большие значения. Таким образом, с вычислительной точки зрения такой способ очень нестабильный. Для решения этой проблемы широко используются два подхода: первый на основе линий максимума (WaveletTransformModulusMaxima) и второй на основе вейвлет-лидеров.

Оценивание спектра с помощью линий максимума. WTMM

Подход к оцениванию мультифрактального спектра с помощью линий максимума активно разрабатывался, начиная с работы [Mallat, 1992]. Оказывается, что выражения из последнего подпункта будут справедливы, если вместо последовательности вейвлет-преобразований $W_\psi[x](s, t)$ при постоянном значении t_0 и уменьшающемся масштабе s , рассмотреть кривую локальных максимумов модуля $W_\psi[x](s, t)$. Для начала введем необходимые определения.

Точкой максимума модуля (*modulus maximum*) называют точку (s_0, t_0) такую, что $W_\psi[x](s_0, t) \leq W_\psi[x](s_0, t_0)$, где значения t – это соседние (правое и левое) значения для t_0 , при этом хотя бы для одного значения t (правого или левого) неравенство должно быть строгим $W_\psi[x](s_0, t) < W_\psi[x](s_0, t_0)$.

Линией максимумов называют связную кривую в пространстве (s, t) , которая состоит из точек максимума

модуля. Совокупность линий максимумов, полученных из вейвлет-преобразований функции, называют **скелетомом**.

На Рис. 3.35 показаны коэффициенты вейвлет-преобразования для рядов Т, К и Х с выделенными скелетонами.

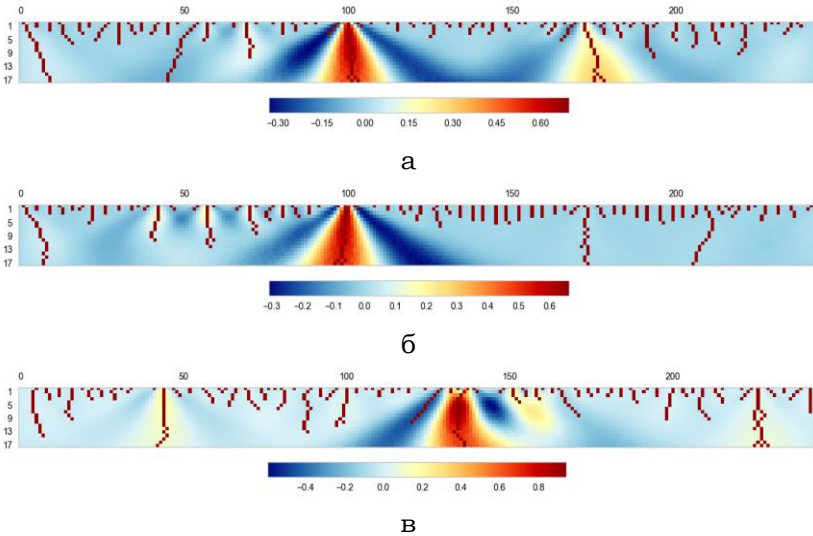


Рис. 3.35 – Коэффициенты вейвлет-преобразования для рядов Т (а), К (б) и Х (в) с выделенными скелетонами.

Теперь вернемся к определению структурной функции. Вместо интегрирования по всем вейвлет-преобразованиям при фиксированном масштабе s , перейдем к дискретной сумме по локальным максимумам

$$Z(q, s) = \sum_{l \in \mathcal{L}(s)} \left(\sup_{\substack{(\tau, s') \in l \\ s' < s}} |W_\psi[x](t, s)| \right)^q, \quad q \in \mathbb{R}.$$

Далее по соответствующим формулам можно перейти к масштабной функции и мультифрактальному спектру.

На Рис. 3.36 показаны масштабная функция и мультифрактальный спектр для рядов Т, К и Х. На Рис. 3.36 а, кроме масштабных функций для исследуемых рядов, показана масштабная функция для броуновского движения.

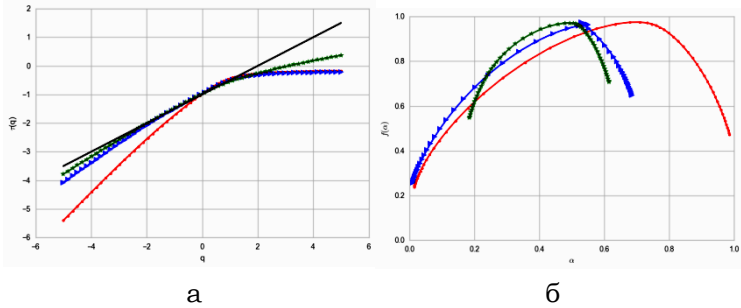


Рис. 3.36 – Масштабные функции (а) и мультифрактальные спектры (б) для рядов Т, К и Х полученные с помощью метода WTMM.

Вейвлет лидеры

Понятие вейвлет-лидеров и их применение для вычислительных целей начал разрабатывать [Jaffard 2004]. Цель применения вейвлет-лидеров такая же, как и у линий максимума – стабилизация суммы вейвлет коэффициентов при возведении в отрицательную степень. Кроме того, в отличие от линий максимума, подход на основе вейвлет-лидеров естественным образом обобщается для d -мерного пространства [Jaffard 2006].

Вейвлет-лидеры определяются через коэффициенты дискретного вейвлет преобразования, поэтому вместо произвольного параметра масштаба s используются только значения $s = 2^{-j}$.

Пусть x – ограниченная функция, тогда **вейвлет-лидеры** этой функции определяются по формулам

$$d_j(t) = \max_{t' \in \{t-1, t, t+1\}} |W_\psi[x](t', 2^{-j})|.$$

Определив вейвлет-лидеры можно ввести формулу для структурной функции

$$Z(q, j) = 2^{-j} \sum_t (d_j(t))^q.$$

Точно так же, как и в случае с методом WTMM, получили структурную функцию, которая более стабильная с вычислительной точки зрения, чем структурная функция, определенная непосредственно через полный набор

вейвлет-коэффициентов. Далее оценивание масштабной функции и мультифрактального спектра происходит точно таким же способом, как и ранее по изначальным формулам.

На Рис. 3.37 показаны масштабная функция и мультифрактальный спектр для рядов Т, К и Х, оцененные с помощью метода вейвлет лидеров

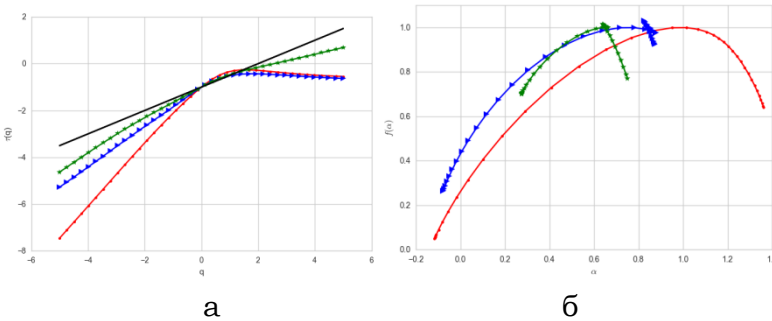


Рис. 3.37 – Масштабные функции (а) и мультифрактальные спектры (б) для рядов Т, К и Х полученные с помощью вейвлетов лидеров

Разные методы оценивания мультифрактального спектра могут приводить к разным результатам. В качестве примера рассмотрим работу методов для броуновского движения. Известно, что броуновское движение является монофрактальным с показателем Херста 0.5. Поэтому его масштабная функция имеет вид $\tau(q) = 0.5q - 1$ и спектр сосредоточен в одной точке $f(0.5) = 1$. Подробное сравнение методов MF-DFA и WTMM для различных примеров представлено в [Oswiecimka 2006]. Ниже представим результаты оценивания спектра методами MF-DFA, WTMM и с помощью вейвлетов лидеров для броуновского движения.

На Рис. 3.38 показано 50 сгенерированных траекторий броуновского движения длиной 1000. Для каждого процесса была оценена масштабная функция $\tau(q)$ каждым из трех методов. На Рис. 3.39а, в, д показаны усредненные функции $\tau(q)$ со стандартным отклонением. Мультифрактальные спектры, которые соответствуют усредненным масштабным функциям показаны на Рис. 3.39б, г, е.

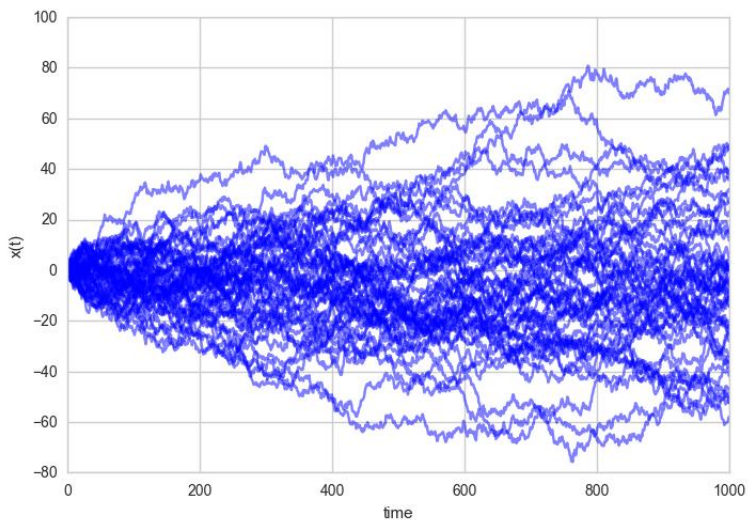
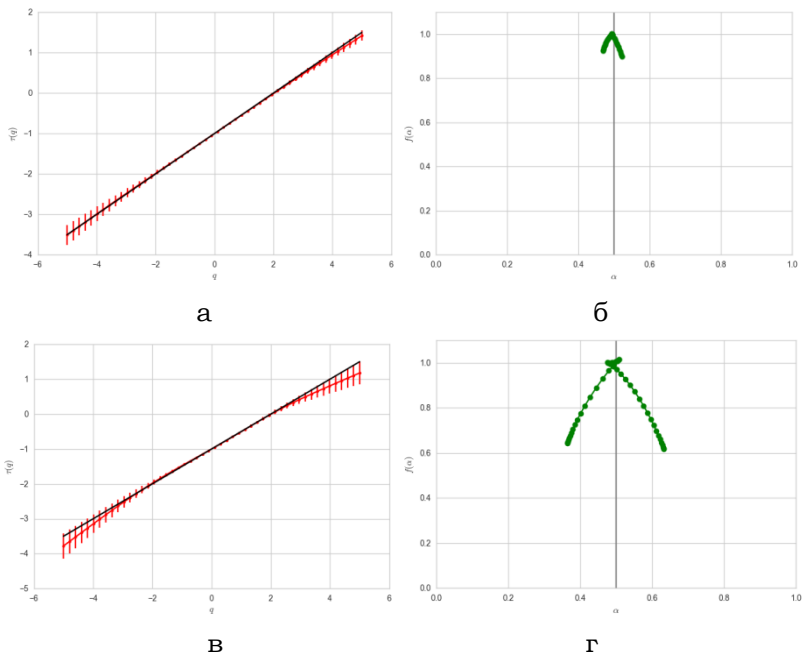


Рис. 3.38 – Сгенерированные траектории броуновского движения



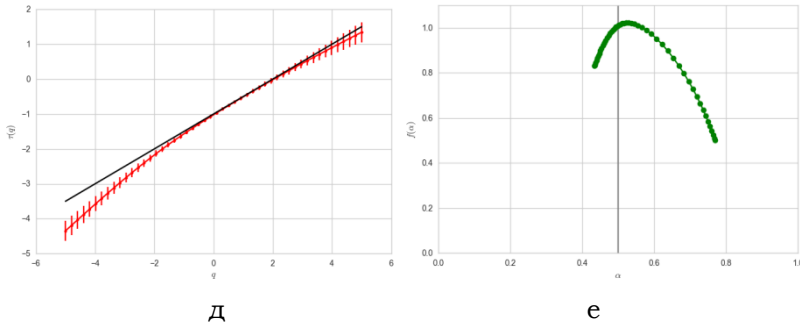


Рис. 3.39 – Усредненная масштабная функция для сгенерированных броуновских процессов, оцененная методом MF-DFA (а), вейвлет-лидеров (в) и WTMM (д) и соответствующие мультифрактальные спектры (б, г, е)

Мультифрактальный спектр может использоваться для сравнения временных рядов. Мультифрактальные спектры одинаковой формы свидетельствуют о схожести рядов, в то время как различная форма спектра означает, что в природе рядов есть принципиальные различия. Это свойство используется в различных исследованиях. Например, в [Ivanov, 1999] анализируются сигналы физиологической природы, такие как сердцебиение. Показано, что мультифрактальные спектры временных рядов, которые соответствуют здоровому сердцебиению значительно шире, чем в противном случае. Таким образом, мультифрактальный спектр может быть индикатором при анализе динамики сердечного ритма. Похожий подход для различения здоровых пациентов и пациентов с нарушением баланса представлен в [Shimizu, 2002].

В [Sun, 2001] рассматривается ежедневный индекс HangSeng на фондовом рынке Гонконга. При этом статистически анализируется, как варьирование определенных параметров влияет на мультифрактальный спектр. В этом случае различие в мультифрактальных спектрах также отображает различие в природе экономических процессов. Результаты исследований в области экономики, которые основаны на сравнении мультифрактальных спектров, представлены в [Zhou, 2009]. Используя пример индекса Dow Jones Industrial Average, выявлены экономические факторы, которые влияют на форму спектра.

4. Анализ сетевых структур при выявлении информационных операций

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.

4.1. Параметры сетей

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 среднего 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.Строгатц (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 составляло бы $1/2 \cdot k(k-1)$. 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 большую 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. 4.1). Оказалось, что именно слабые связи являются тем феноменом, который связывает сеть в единое целое.

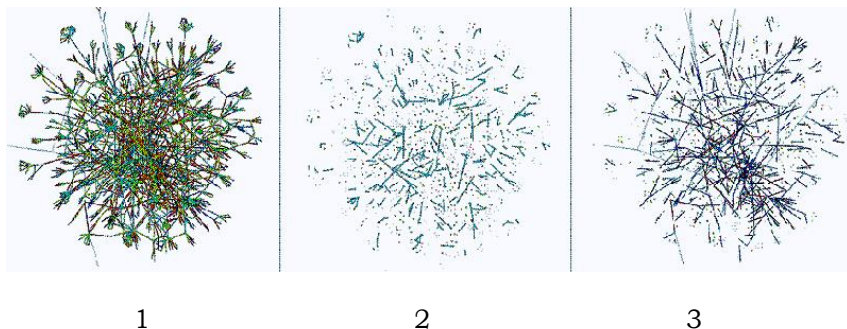


Рис. 4.1 – Структура сети:

- 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. Милграм 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.Строгатц 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 представлены на рис. 4.2: 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.

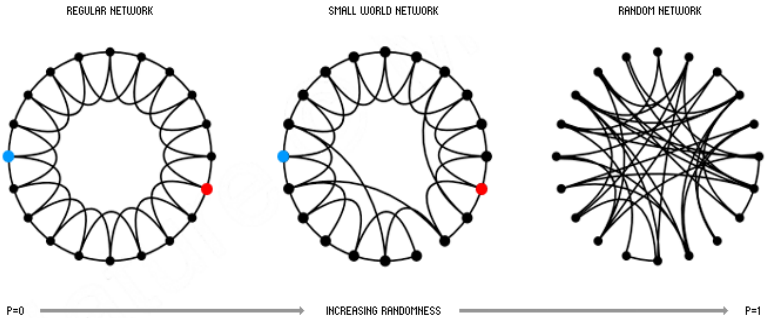


Рис. 4.2 – Модель Уаттса-Строгатца

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.

На Рис. 4.3 приведены графики изменения средней длины пути и коэффициента кластеризации искусственной сети Д. Уоттса и С. Строгатца от вероятности установления «далеких связей» (в полулогарифмической шкале).

Например, WWW является сетью, для которой также подтвержден феномен малых миров. Анализ топологии веб, проведенный Ши Жоу (S.Zhou) и Р.Дж.Мондрагоном (R.J.Mondragon) из Лондонского университета, показал, что узлы с большой степенью исходящих гиперссылок имеют больше связей между собой, чем с узлами с малой степенью, тогда как последние имеют больше связей с узлами с большой степенью, чем между собой. Этот феномен был назван "клубом богатых" (rich-club phenomenon). Исследование показало, что 27% всех соединений имеют место между всего 5% наибольших узлов, 60% приходится на соединение других 95% узлов с 5% наибольших и только 13% - это соединение между узлами, которые не входят в лидирующие 5%.

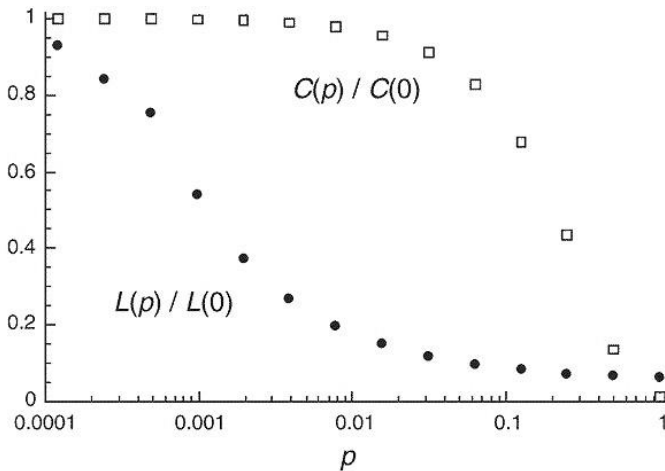


Рис. 4.3 – Динамика изменения длины пути и коэффициента кластерности в модели Уоттса-Строгатца в полулогарифмической шкале (ось Ox – вероятность замены ближних связей далекими)

Эти исследования дают основания полагать, что зависимость WWW от больших узлов значительно существеннее, чем предполагалось ранее, т.е. она еще более чувствительна к злонамеренным атакам. 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.

Практикой доказано [Rothenberg, 2002], что террористические сети чаще всего не только безмасштабные, но также демонстрируют свойства малых миров, т.е. то, что наличие тесно связанных кластеров (групп тесно связанных узлов) обеспечивает локальная связь даже в случаях удачных атак, когда концентраторы (наибольшие посредники) выходят из строя.

При изучении «малых миров» определился интересный подход, логически связанный с понятием перколяции (протекания) [Broadbent, 1957], [Снарский, 2007].

Оказывается, что многие вопросы, которые возникают при анализе сетевой безопасности в Интернет, непосредственно относятся к этой теории. Самая простая, очищенная от всех физических и математических наслоений формулировка задачи теории перколяции имеет такой вид: «Данна сеть, случайная часть ребер которой проводит сигнал, а другая часть - не проводит. Основной вопрос - чему равна минимальная концентрация проводящих связей, при которой еще существует путь через всю сеть?». К задачам, которые решаются в рамках теории перколяции и анализа сетей относятся такие, как определение предельного уровня проводимости, изменения длины пути и его траектории при приближении к предельному уровню проводимости, количества узлов, которые необходимо вывести из строя, чтобы нарушить связность сети.

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.

4.2. Сетевые признаки информационных операций

Как расширение приведенной выше мультиагентной модели распространения информации можно рассматривать модель, в которой учитывается структура формируемой сети [Путачев, 2015]. В рамках этой модели каждый агент – источник информации – обладает не «потенциалом», а некоторым рейтингом (которому на схемах соответствует размер соответствующего узла). Связями в рассматриваемой сети являются факты перепечатки или «пересказа».

В основе модели лежит предположение, что при проведении информационных операций наиболее рейтинговые источники перепечатавают информацию у наименее рейтинговых, или образуются кластеры низкорейтинговых изданий, перепечатавающих одну и ту же новость.

На Рис. 4.4 приведен пример типичных информационных операций, выявляемых в рамках данной модели.

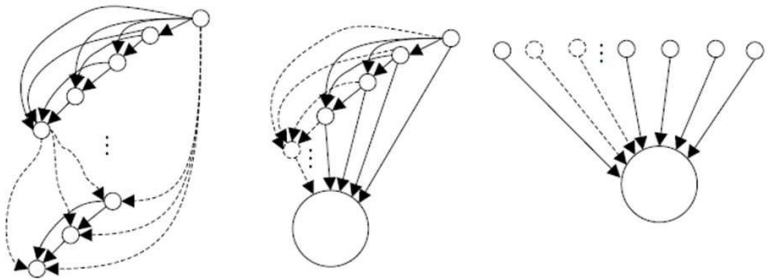


Рис. 4.4 – Примеры сетей распространения информации, имеющих признаки информационных операций

В рамках формализации этой же модели выбирается несколько десятков параметров топологии сетей распространения информации, таких как диаметр, плотность, кластеризация, посредничество и т.п., которые сравниваются с некоторыми эталонными значениями.

К достоинствам данной модели следует отнести ее формальную строгость и соответствие активно развивающемуся в последнее время направлению Complex Networks, что позволяет ожидать ее дальнейшего развития. К недостаткам следует, по-видимому, отнести малую корреляцию с содержательной стороной распознаваемых информационных операций, а также определенную вычислительную сложность при выявлении нечетких информационных дубликатов документов.

Технологические этапы исследования взаимного влияния источников информации

Для эффективного исследования взаимного влияния источников информации из сети Интернет (веб-ресурсов, социальных медиа) предлагается последовательность шагов,

этапов обработки информации, каждый из которых сам по себе обеспечивает получение аналитического продукта. Совокупность таких этапов, которые базируются на использовании необходимых и доступных инструментальных средств, специальных приемов, можно рассматривать как процедуру проведения действий, направленных на получение аналитических материалов, включающих построение и анализ сети их взаимного влияния.

При проведении данных информационно-аналитических исследований на базе контент-мониторинга к таким задачам можно отнести:

- Нахождение релевантных публикаций на заданную тематику.
- Выявление взаимных контекстных ссылок и перепечаток в документах, представленных различными информационными источниками.
- Построение сети влияния, анализ и визуализация взаимосвязей информационных источников, в том числе ранжирование узлов построенной сети по степени влияния.
- Выявление возможных информационных операций и построение сценария противодействия информационным операциям в сетевой среде.

Получение репрезентативного массива публикаций

Для получения репрезентативного массива публикаций по выбранной тематике необходимо выбрать систему контент-мониторинга, которая предоставляет поток информационных сообщений по определенной тематике. Тематика может выражаться запросом на языке информационно-поисковой системы.

В качестве системы контент-мониторинга авторами была выбрана система InfoStream, которая в настоящее время охватывает 10 тыс. источников информации на русском и украинском языках. В базы данных системы ежедневно поступает более 100 тыс. документов. Система InfoStream обеспечивает поиск, а также просмотр списка и полных текстов релевантных документов.

В приведенном на Рис. 4.5 примере показан фрагмент интерфейса системы, через который обрабатывался запрос, относящийся к обсуждению в январе 2016 года вопрос отставки премьер-министра Украины А. Яценюка.

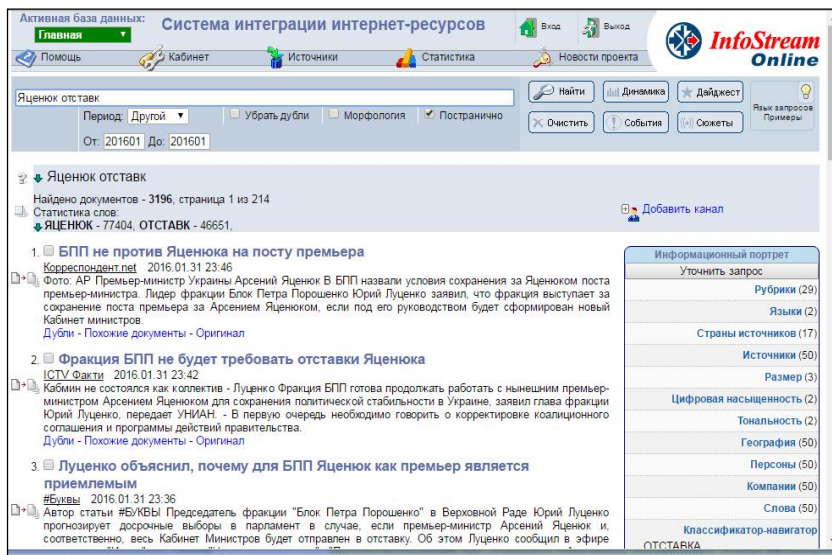


Рис. 4.5 – Фрагмент интерфейса системы контент-мониторинга

В результате был сформирован тематический информационный массив, который охватывает 3196 документов.

Определение контекстных ссылок

Основой построения сети влияния источников информации является контекстные ссылки и перепечатки в тематическом информационном потоке. Контекстные ссылки выявляются путем идентификации шаблонов (Табл. 1) в документах выбранного информационного массива и признаков точных перепечаток, определяемых методами выявления плагиата. В свою очередь, сами шаблоны периодически определяются/дополняются экспертами в автоматизированном режиме путем анализа контекста потока документов системы контент-мониторинга методами Text Mining.

Таблиця 1. Шаблони названий інформаційних ресурсів (фрагмент)

№	Код источника	Шаблон 1	Шаблон 2
1	srd06193	Деро	"Деро"
2	srd03176	Українські національні новини	УНН
3	srd00045	Сегодня.ua	"Сегодня"
4	srd03076	ТСН.ua	"ТСН"
5	srd07509	112.ua	"112"
6	srd02348	Gazeta.ua	
7	srd00069	Корреспондент.net	"Корреспондент "
8	srd07487	Еспресо TV	"Еспресо ТВ"
9	srd02535	Телеканал новин "24"	"24"
10	srd06453	Телеграф.com.ua	"Телеграф"
11	srd01351	ЗІК	"ЗІК"
12	srd02514	РБК-Україна	РБК-Украина
13	srd04508	Українські Новини	
14	srd07686	"Антикор"	
15	srd00253	"Обозреватель"	
16	srd00057	Інтерфакс	Інтерфакс
17	srd00404	ICTV Факти	ICTV
18	srd04125	РІА Новини Україна	
19	srd02732	УКРІНФОРМ	УКРІНФОРМ
20	srd00095	УНІАН	УНІАН
21	srd00094	Українська правда	
22	srd01408	Цензор.Нет	
23	srd00064	ЛІГАБізнесІнформ	
24	srd00039	Газета День	
25	srd07038	Вести.ua	

Построение сети влияния источников информации

Найденные в текстах контекстные ссылки и перепечатки позволяют сформировать матрицу цитирования, транспонируя которую в соответствии с приведенной выше гипотезой, можно сформировать матрицу влияния. Данной матрицы соответствует сеть влияния источников, пример визуализации которой для рассмотренного выше тематического информационного массива с помощью системы Gephi приведен на Рис. 4.6.

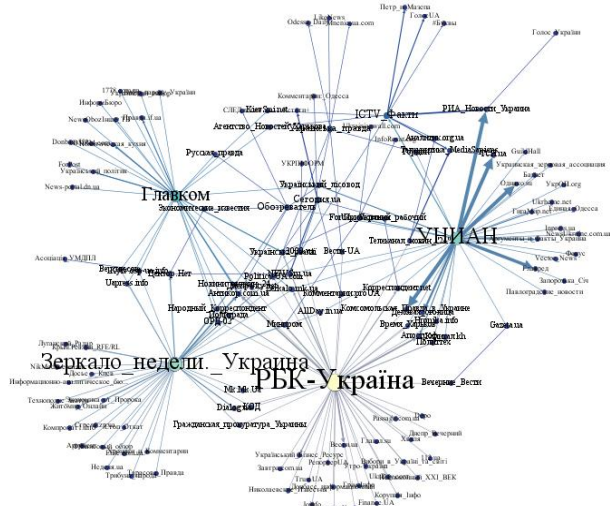


Рис. 4.6 – Фрагмент сети связей источников по выбранной тематике

Исследование сети влияния источников информации

Построенную сеть влияния источников информации можно исследовать с помощью общепринятых инструментальных средств (например, с помощью системы Gephi были получены следующие параметры построенной сети, как количество узлов: 141, ребер 196, плотность графа: 0.01, средний коэффициент кластеризации: 0,026, средняя длина пути: 1,26 и т.д.).

Для содержательного анализа большое значение имеет вес узлов сети, список важных узлов по критерию выходной мощности приведены в Таблице 2.

Таблица 2. Наиболее влиятельные узлы по количеству цитирования

№	Веб-ресурс	Выходная степень
1	РБК-Україна	50
2	Зеркало недели	38
3	УНИАН	35
4	Главком	28
5	ICTV-Факты	10
6	Сегодня.ua	7
7	Українська правда	7
8	Обозреватель	6
9	Forbes-Украина	4
10	Цензор.Нет	3

Перспективным подходом к ранжированию источников по уровню влияния является алгоритм HITS, предложенный Дж. Кляйнбергом [Kleinberg, 2006].

Алгоритм HITS обеспечивает выбор из сети лучших «авторов» (узлов, на которые введут ссылки) и «посредников» (узлов, от которых идут ссылки включения).

В соответствии с алгоритмом HITS для каждого узла сети рекурсивно вычисляется его значимость как автора $a(v_j)$ и посредника (хаба) $h(v_j)$ по формулам:

$$a(v_j) = \sum_{i \rightarrow j} h(v_i);$$

$$h(v_j) = \sum_{i \rightarrow j} a(v_i).$$

В этих формулах суммирование производится по всем узлам, которые ссылаются на (или на которые ссылается) данный узел.

Парфразируя обозначения, приведенные в [Kleinberg, 2006], а именно заменяя «авторство» на «подверженность

воздействию», а «посредничество» на «влиятельность» можно с небольшими вычислительными затратами вычислять соответствующие характеристики узлов сети влияния.

Также для выявления информационных воздействий большое значение имеет определение «скрытых» связей. Методика определения скрытых связей, скрытых воздействий, в частности, приведена в работе [Snarskii, 2016].

Определение возможных информационных операций

Сеть информационного воздействия источников информации позволяет оперативно идентифицировать возможные информационные операции в соответствии с подходами, предложенными в [Потемкин, 2015]. Предполагается, что вероятность наличия информационной операции мала, если информация о происшествии сначала зарождается во влиятельном информационном источнике, а затем перепечатывается (со ссылками или без них) менее влиятельными источниками (Рис. 4.7). Обратные явления, когда более влиятельные издания перепечатывают информацию в менее влиятельных, пусть и многочисленных, может быть признаком информационной операции, атаки (Рис. 4.8). Именно такие картины наблюдались при сетевом анализе реальных тематических информационных потоков (Рис. 4.9).

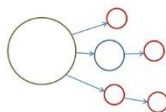


Рис. 4.7 – Типовой сценарий распространения информации

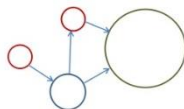


Рис. 4.8 – Сценарий распространения информации, присущий информационной операции

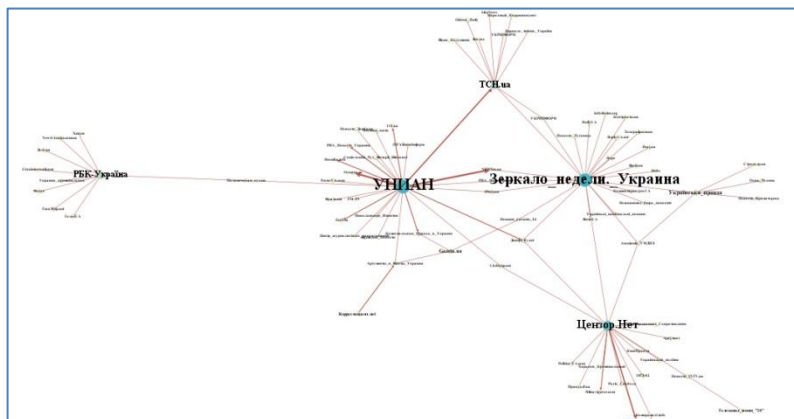


Рис. 4.9 - Фрагмент сети связей источников по специальной тематике

К преимуществам данной модели следует отнести ее формальную строгость и соответствие активно развивающемуся в последнее время направлению Complex Networks, что позволяет ожидать дальнейшего развития. К недостаткам следует, по-видимому, отнести малую корреляцию с содержательной стороной распознаваемых информационных операций, а также определенную вычислительную сложность при обнаружении нечетких информационных дубликатов документов.

5. Поддержка принятия решений при выявлении информационных операций

Планирование мероприятий по усилению информационной безопасности, предупреждению негативных/вражеских информационных воздействий и информационных операций (а также планированию успешных операций в процессе информационной борьбы) требует четкого понимания и точного знания предметной области. Впрочем, сфера информационной безопасности представляет собой слабо структурированную предметную область, которая трудно поддается описанию, в частности, формальному, количественному. В работах [Андрейчук, 2016; Andriichuk, 2017; Каденко, 2016; Kadenko, 2016] показано, что информационные операции (ИО) относятся к слабо структурированным предметным областям. Технологии экспертной поддержки принятия решений как раз представляют собой средство решения задач в слабо структурированных предметных областях [Тоценко, 2002]. Поэтому, вопросы использования экспертных технологий поддержки принятия решений в области информационной безопасности и конкретный контекст применения отдельных методов заслуживают отдельного рассмотрения.

Информационная безопасность как слабо структурированная предметная область

В работе [Таран, 2006] приводятся следующие признаки слабо структурированных предметных областей (рис. 5.1): отсутствие цели функционирования, которую можно формализовать, отсутствие оптимальности, уникальность, динамичность, неполнота описания, наличие субъективного человеческого фактора, невозможность построения аналитической модели, отсутствие эталонов, большая размерность.

В разделе 1 отмечалось, что на информационные операции влияет много чисто качественных (в частности, социально-психологических) критериев, факторов и параметров. Эти факторы трудно поддаются формальному математическому (аналитическому) описанию.



Рис. 5.1 – Признаки слабо структурированных предметных областей

В [Горбулін, 2009] речь шла о невозможности разработки и практического применения определенной универсальной методики моделирования информационных операций, прежде всего, вследствие слабой формализации понятий и факторов. Авторы отмечают, что в каждом случае следует консультироваться с аналитиками, то есть экспертами по анализу информационных операций, полагаться на их компетентность. При этом иногда аналитики в состоянии построить точные прогнозы определенных закономерностей, которые затем подтверждаются практикой. К аналитикам-экспертам следует обращаться для описания субъективных факторов. Когда же речь идет о факторах объективных, то их описание и анализ предлагается осуществлять с помощью известных методов, которые оперируют детерминированными данным, в частности, методов математической статистики и анализа временных рядов. Впрочем, эти методы направлены лишь на описание формальных аспектов и не касаются аспектов содержательных. С учетом этого, в [Горбулін, 2009] констатируется необходимость расширения

технологического инструментария, который может использоваться для анализа и моделирования информационных операций.

Таким образом, информационные операции (как, кстати, и любые операции, предусматривающие существенное человеческое участие) представляют собой яркий пример слабо структурированной предметной области. Одним из технологических средств анализа и моделирования информационных операций, должны стать экспертные технологии поддержки принятия решений. На целесообразность использования экспертных знаний в слабо структурированных предметных областях указывают также исследования Delphi Group по составу знаний организаций [Тузовский, 2005]: электронная документация - 20%, бумажная документация - 26%, электронные базы знаний - 12%, в головах специалистов - 42%.

В результате исследований выяснилось, что значительная часть знаний находится не в базах знаний, или на бумажных или электронных носителях, а именно в головах экспертов (аналитиков, специалистов). Итак, в контексте описания и анализа информационных операций, безусловно, имеет смысл привлекать экспертные знания, особенно, когда речь идет о качественных факторах субъективного характера.

5.1. Иерархическая декомпозиция и целевое динамическое оценивание альтернатив

В работе [Горбулін, 2009] отмечено, что информационная операция представляет собой междисциплинарный набор методов и технологий, который охватывает многие сферы, от военной науки до социологии. При этом не существует универсальной, стандартной технологии проведения информационных операций (которая могла бы послужить не только военным, но и руководителям крупных правительственных или коммерческих организаций). Итак, как отмечают авторы работы, актуальным вопросом остается разработка научной базы информационных операций.

С учетом вышеупомянутого междисциплинарного характера описания информационных операций и воздействий удобным инструментом должна стать экспертная иерархическая декомпозиция. В частности, данный подход составляет основу метода целевого динамического оценивания альтернатив (МЦДОА), предложенного в [Тоценко, 2002] и усовершенствованного в [Цыганок, 2013]. Метод как раз позволяет объединить в единую иерархию большое количество критериев и мероприятий различной природы (из различных дисциплин), которые влияют на определенную главную цель.

В зависимости от типа конкретной информационной операции (наступательного или оборонительного [Горбулін, 2009]) аналитик-эксперт (или же ЛПР) может соответствующим образом сформулировать главную цель. Любая информационная операция декомпозируется на определенные этапы или шаги. В [Горбулін, 2009] приводятся их подробные перечни. Содержание этих шагов может варьироваться в зависимости, опять-таки, от типа операции и специфики контекста. Например, если речь идет о моделировании и декомпозиции (в рамках МЦДОА) информационной атаки на академию наук (главная цель), то цель «Дискредитация научного учреждения в СМИ» может декомпонироваться на цели более низкого уровня

«Дискредитация научных трудов и достижений» и «Дискредитация научных сотрудников».

В МЦДОА декомпозиция производится до уровня атомарных подцелей (факторов, критериев), на которые может влиять ЛПР. Эти цели называются проектами, и, как правило, могут быть охарактеризованы определенной величиной (абсолютной или относительной численной, логического/булевского или порогового типа).

Общей принципиальной задачей методов поддержки принятия решений, которые предусматривают иерархическую декомпозицию задачи, в частности МЦДОА и методов анализа иерархий и сетей (АНР/АНП) [Saaty, 2008], является построение рейтинга или ранжирования нескольких объектов (альтернатив, проектов). На основе такого рейтинга ЛПР может сделать обоснованный выбор наилучшей альтернативы (варианта решения) из заданного множества, или расставить приоритеты в своей деятельности (то есть выяснить, какие факторы или мероприятия являются важнейшими для достижения заданной главной цели). Для построения такого рейтинга необходимо определить относительную важность всех целей, входящих в граф иерархии целей, построенной экспертами (или инженерами по знаниям на основе диалогов с экспертами). Чтобы определить относительную важность подцелей отдельной цели (ее «потомков» в графе иерархии), эксперты должны попарно сравнить их между собой. Оценки воздействий (весов) могут осуществляться экспертами в разных шкалах парных сравнений. Недавние исследования [Цыганок, 2011; Цыганок, 2012] показали, что эксперту следует предоставлять возможность вводить значение каждого отдельного парного сравнения в наиболее удобной для него шкале. Когда экспертами оценены все относительные веса целей в графе иерархии, можно рассчитать относительное влияние каждой цели и проекта на главную цель (относительную эффективность), как показано в [Тоценко, 2002].

Если оценки задаются несколькими экспертами, то следует учитывать несколько важных аспектов. Первым из них является компетентность экспертов. Если известно, что эксперты, которые оценивают критерии, проекты, или альтернативы, имеют разную компетентность, то ее следует

определять на основе нескольких составляющих: самооценки, взаимной оценки компетентности членов экспертной группы, и объективной компоненты (как показано в [Тоценко, 2002]). Разницей в относительной компетентности экспертов можно пренебречь лишь в случае, когда размер экспертной группы является достаточно большим [Tsyganok, 2012]. Вторым важным аспектом является согласованность оценок, полученных от разных членов экспертной группы. Оценки следует проверять на согласованность: ведь рекомендации, выданные ЛПР, на основе несогласованных экспертных данных, закономерно будут вызывать недоверие. Для оценки согласованности экспертных оценок предлагается применять спектральные методы, описанные, в частности, в [Zgurovsky, 2004] и [Olenko, 2016].

Преимущество спектральных методов над другими подходами к оценке согласованности (в частности, теми, что предложены Саати и коллегами [Forman, 1998]) – следующее. При необходимости (если уровень согласованности экспертных оценок в группе низкий), спектральные методы позволяют организовать конструктивную пошаговую обратную связь с экспертами. То есть, экспертам предлагается изменить соответствующие оценки («выбросы») таким образом, чтобы уровень согласованности повысился до необходимой величины. Когда оценки, задаваемые различными членами экспертной группы, достигают достаточно высокого уровня согласованности, их можно агрегировать (получать на их основе обобщенные групповые оценки). Для агрегации экспертных оценок целесообразно применять комбинаторный метод агрегации [Циганок, 2000]. Среди многочисленных преимуществ этого метода следует назвать возможность его использования для агрегации неполных матриц парных сравнений и максимальное использование избыточности экспертной информации.

В контексте МЦДОА «взвешенная» иерархия критериев (целей) называется базой знаний (БЗ) о предметной области. В данной работе внимание сосредоточено на предметных областях, касающихся информационной безопасности и, в частности, информационных операций. По содержанию, такие БЗ представляет собой один из типов моделей

предметных областей. БЗ строится экспертами (или инженерами по знаниям в ходе диалогов с экспертами) с помощью специальных программных средств – автоматизированных систем поддержки принятия решений (СППР).

Отметим, что МЦДОА не требует, чтобы все данные, которые вводятся в БЗ СППР, были исключительно экспертными оценками. Например, если речь идет о сравнении нескольких альтернатив по определенному критерию, то значение оценок не обязательно должны выражаться в шкале парных сравнений. Иногда это могут быть абсолютные значения, доступные из открытых источников. Допустим, для анализа информационной политики или кампании часто используются абсолютные показатели, такие как «количество публикаций с негативной окраской в течение месяца». Такой показатель вполне правомерно может быть включен в иерархию критериев, которая описывает информационную политику организации.

5.2. Особенности работы с экспертами

В результате исследований выяснилось, что значительная часть знаний находится не в базах знаний, или на бумажных или электронных носителях, а именно в головах экспертов (аналитиков, специалистов). Итак, эксперт (аналитик, специалист), как правило, является представителем узко-ориентированной предметной области, и, в общем случае, не ориентируется в технологиях и методах поддержки принятия решений. Поэтому процесс получения информации от эксперта должен быть максимально комфортным для него. Формальную сторону процесса следует делегировать инженеру по знаниям, а математические вычисления – автоматизированной СППР. Для достижения этой цели в процессе экспертизы целесообразно учитывать несколько особенностей.

Во-первых, если эксперт не знаком с технологиями поддержки принятия решений, его имеет смысл ознакомить хотя бы с общим ходом экспертизы. В

идеальном случае, участникам экспертизы следует объяснить всю технологию, с помощью которой будут обрабатываться оценки и формироваться рекомендации по выбору варианта решения для АПР. Это повысит уровень доверия экспертов к процессу и даст им возможность вводить данные в СППР в приемлемом формате. Итак, перед тем, как начинать сбор информации у экспертов, целесообразно провести с ними ознакомительные тренинги (coaching sessions).

Во-вторых, иерархия критериев (целей) (рис. 5.3) должна быть максимально сбалансированной. Проекты желательно располагать на одном уровне (поскольку, как показано в [Saaty, 2008], их веса должны лежать в пределах одного порядка). Следует избегать появления большого (более 7 ± 2) количества «потомков» у одной цели в графе иерархии, учитывая психофизиологические ограничения человека [Miller, 1956]. При построении иерархии нежелательно ставить эксперту много однотипных вопросов (в частности, относительно положительного или отрицательного характера влияния целей на общего «предка» в графе, или по попарно совместимости целей-потомков).

В-третьих, при оценке влияний, следует отдавать предпочтение не численным значениям, а их вербальным эквивалентам (например, «1» - равнозначность, «2» - слабое преимущество, ..., «5» - очень сильное преимущество).

В общем, интерфейс автоматизированной СППР должен быть максимально удобным для пользователя и интуитивно понятным ему.

Более детальный анализ особенностей работы с экспертом приведен в работе [Каденко, 2016]. Остановимся подробнее на проблемах, возникающих на разных этапах диалога с экспертом.

Ознакомительные тренинги перед экспертизой

Первое, о чем следует помнить организатору экспертизы: в общем случае АПР и эксперты не знакомы с математическими методами, с помощью которых будут вводиться и обрабатываться экспертные данные. Если АПР и эксперты будут иметь представление об общем ходе

экспертизы, то им будет легче вводить данные в формате, необходимом для ее проведения. Поэтому еще перед началом экспертизы имеет смысл провести соответствующие ознакомительные тренинги (coaching session). Например, автор метода анализа иерархий, Том Саати [Saaty, 1996], и его последователи, уделяют таким тренингам очень большое внимание.

Особенности работы с экспертами на этапе построения иерархии

На этапе построения иерархии следует учесть несколько особенностей работы с экспертами.

1) Следует помнить об уравновешенности (сбалансированности) иерархии: проекты (атомарные, неделимые подцели главной цели) желательно располагать на одном уровне. На этапе оценки относительных влияний проект, расположенный на высшем уровне иерархии, вероятнее всего, будет иметь на порядок больший вес, чем проекты, расположенные на более низких уровнях. Этот вес будет еще больше, если проект является подцелью многих целей высших уровней, ведь влияние проекта на главную цель исчисляется как сумма произведений относительных воздействий по всем путям, ведущим от проекта к главной цели в графе иерархии критериев [Тоценко, 2002; Каденко, 2008]. Добавление уровня иерархии критериев, фактически, уменьшает на порядок величины весов целей (проектов), которые расположены на этом уровне.

Проблеме поиска оптимальной структуры иерархии посвящены, в частности, работы С. Липовецкого [Lipovetsky, 2006; Lipovetsky, 2006].

2) Необходимо, чтобы эксперты сосредоточились именно на декомпозиции главной цели, абстрагируясь от посторонних, более общих понятий.

3) Следует предотвращать появление в иерархии одноименных подцелей или проектов, чтобы избежать неоднозначности [Андрійчук, 2014].

4) Следует обеспечить четкое и последовательное введение в иерархию связей и влияний – в этом смысле эксперту удобно выбирать подцели и проекты из списка

уже введенных, или вводить новые формулировки в готовый список.

5) В диалоге с экспертом следует избегать многократного повторения однотипных процедур (таких, как, например, определение характера влияния (положительного или отрицательного), и попарная проверка совместимости целей). Например, для того, чтобы попарно проверить совместимость m целей, следует поставить эксперту $m(m - 1) / 2$ однотипных вопросов.

6) Следует акцентировать внимание на проектах, подчеркивая их семантическую атомарность (для этого имеет смысл ставить эксперту уточняющий вопрос: «данная цель может считаться проектом или она предполагает дальнейшую декомпозицию?»).

7) Следует помнить о психофизиологических ограничениях человека, и при построении иерархии не допускать, чтобы количество непосредственных подцелей каждой цели превышало 7 ± 2 [Miller, 1956].

Таблица 5.1. Соответствие делений шкалы Саати вербальным эквивалентам

Numerical values	Verbal term	Explanation
1	Equally important	Two elements have equal importance
3	Moderately more important	Experience or judgment slightly favors one element
5	Strongly more important	Experience or judgment strongly favors one element
7	Very strongly more important	Dominance of one element proved in practice
9	Extremely more important	The highest order dominance of one element over another
2,4,6,8	Important intermediate values	Compromise is needed

Уход от количественных значений на этапе оценивания

На этапе оценки влияний желательно избегать количественных значений, предлагая эксперту называть их

вербальные эквиваленты («лучший», «самый плохой», «лучше», «хуже», «выберите лучшую альтернативу», «преобладает», «уступает», вместо «ранг 1», «ранг 5», «проранжируйте альтернативы»; «сильно преобладает», «очень сильно уступает», вместо «в 5 раз лучше», «в 9 раз хуже»).

Ранги и кардинальные значения (в т.ч. парные сравнения) должны приводиться в соответствие вербальным эквивалентам автоматически (таблица 5.1).

Рассмотрим в качестве примера соответствие делений шкалы Лайкерта «согласен - не согласен» [Likert, 1932] вербальным эквивалентам: 1 – "абсолютно не согласен", 2 – "не согласен", 3 – "не определился", 4 – "согласен", 5 – "полностью согласен". Эта шкала используется во многих современных методиках анкетирования, например, в методике оценки городских локаций SpaceShaper [St. James Park, 2007]. Респонденты вообще не имеют дела с числами. Их ответы типа «согласен - не согласен» должным образом обрабатываются, определяются общие индексы удовлетворенности, и на их основе строятся наглядные паутинные / лепестковые (spider) или столбчатые (bar) диаграммы (рис. 5.2).

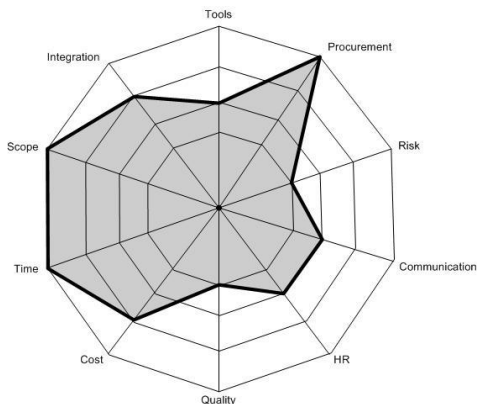


Рис. 5.2– Пример лепестковой (паутинной) диаграммы

Современные методики сбора и обработки экспертной информации, а также интерфейсы автоматизированных СППР должны учитывать указанные особенности диалога с

экспертами в процессе экспертизы. В контексте работы с экспертами общей целью разработчиков методик и СППР должно быть обеспечение максимально «комфортной» для экспертов среды экспертизы и делегирование максимального количества логических и математических процедур автоматизированной СППР. Некоторые из современных методов (таких как, например, комбинаторный метод, описанный, в частности, в [Цыганок, 2000]) позволяют максимально полно использовать экспертную информацию. (При условии связности графа предпочтений на множестве альтернатив, комбинаторный метод позволяет за счет избыточности экспертной информации (даже если она задана в виде неполной матрицы парных сравнений, введенных только одним экспертом в вербальной шкале) рассчитать веса этих альтернатив).

Итак, хотя процедуры получения данных от экспертов и ввода этих данных в СППР могут казаться тривиальными, они характеризуются определенными существенными особенностями. Проанализированы основные этапы экспертизы, в которых задействованы эксперты. Предложен ряд рекомендаций для инженеров по знаниям, организаторов экспертиз, и разработчиков программного обеспечения СППР. Эти рекомендации позволяют сделать процесс получения экспертных данных и их ввод в СППР более конструктивным, и, таким образом, повысить уровень доверия АПР к результатам экспертизы.

5.3. Методика применения СППР при выявлении информационных операций

Источники информации оказывают существенное влияние на людей. За последние годы стало очевидно, что средства массовой информации можно эффективно использовать для распространения дезинформации. Кроме того, социальные эксперименты показывают, что многие люди верят в неподтвержденные новости и сами продолжают их распространять. Например, в работе [Lewandowsky, 2012] представлен обзор известных ложных убеждений и дезинформации в американском обществе. В

работе [Berinsky, 2017] описаны проведенные социальные эксперименты, в которых исследовалась подверженность людей политическим слухам, связанным с реформой здравоохранения, принятой Конгрессом США в 2010 году. В зависимости от способа представления информации, 17-20% участников эксперимента верили в них, 24-35% участников не имели определенного мнения, а 45-58% опрошенных отклоняли слухи.

Под информационной операцией (ИО) [Горбулін, 2009; Додонов, 2016; Ландэ, 2016b] подразумевается комплекс информационных мероприятий (новостные статьи в интернете и газетах, новости по телевизору, комментарии в социальных сетях, форумах и т.п.), направленных на изменение общественного мнения об определенном объекте (личность, организация, институт, страна и т.п.). Например, распространив слухи о проблемах в банке, можно спровоцировать его вкладчиков на возврат вкладов, что, в свою очередь, может привести к его банкротству. В основном – это мероприятия дезинформированного характера.

В работах [Горбулін, 2009; Додонов, 2016; Ландэ, 2016b] показаны методы выявления ИО, которые базируются на анализе временных рядов, построенных на основе контент-мониторинга тематического информационного потока. Отметим ряд проблемных ситуаций, которые могут возникнуть при идентификации ИО из-за недостатков соответствующих методов и технологий:

- 1) На фоне достаточно большого общего количества публикаций об объекте ИО количество публикаций (информационных вбросов) об одном отдельном его компоненте может быть весьма незначительным и, как следствие, не будут выявлены соответствующие системные нарушения типичной динамики информационных сюжетов (такие как, например, выявленные вейвелеты “мексиканская шляпа” и Морле на соответствующей вейвелет-скейлограмме). Некоторые ИО могут быть комплексными и соответствующие информационные вбросы могут быть поэтапными, касаться различных компонентов объекта ИО в разные периоды времени. Если их количество будет размыто на фоне общего количества публикаций об объекте ИО (“информационного шума”) и

соответствующие информационные атаки не будут идентифицированы, то может быть пропущено начало информационной кампании по дискредитации объекта ИО, а определенный информационный ущерб его имиджу не буде учтен.

2) Средства контент-мониторинга обрабатывают запросы, состоящие из ключевых слов, в результате чего будут найдены соответствующие публикации. Ключевые слова формулируются, исходя из названия объекта ИО. Но сложный объект ИО может иметь значительное количество компонент с соответствующими названиями, которые не учтены в запросах и, как следствие, не все публикации по тематике будут найдены.

3) Запросы, касающиеся объекта ИО, имеют разную степень важности по отношению к компонентам ИО, которых они касаются. Отсутствие информации о значении этих степеней важности (то есть их равнозначность) приводит к снижению адекватности модели ИО.

Для предотвращения вышеописанных недостатков предлагается использовать следующую методику применения СППР при выявлении ИО.

Содержание методики

Содержание предлагаемой методики применения инструментария экспертной поддержки принятия решений при идентификации ИО заключаются в следующем:

1) Проводится предварительное исследование объекта ИО, выбираются его целевые параметры (показатели). Далее предполагается, что ранее (в ретроспективе) уже имели место ИО против объекта и его состояние (соответствующие целевые показатели) в результате ухудшалось.

2) Проводится групповая экспертиза по определению и декомпозиции целей информационной операции, а также оценке степени влияния. Таким образом, объект ИО декомпозируется как сложная слабо структурированная система. Для этого используются средства системы распределенного сбора и обработки экспертной информации (СРСОЭИ). Для получения экспертной

информации в полной мере и без искажений используется система экспертного оценивания СЭО.

3) Строится соответствующая база знаний (БЗ) средствами СППР на основании результатов проведенной средствами СРСОЭИ групповой экспертизы, а также имеющейся объективной информации.

4) Проводится анализ динамики тематического информационного потока средствами системы контент-мониторинга (СКМ). Дополняется БЗ СППР.

5) Рассчитываются рекомендации средствами СППР на основании построенной БЗ. Для этого вычисляются степени достижения целей ИО в ретроспективе и сопоставляются с соответствующими изменениями состояния объекта ИО. Вычисляются средние значения степеней достижения целей ИО, при которых происходило ухудшение значений целевых показателей объекта ИО. Таким образом, путем мониторинга состояния объекта ИО за текущий период времени, можно предугадать ухудшение значений целевых показателей объекта ИО на основании сравнения вычисленных за текущий период времени значений степеней достижения целей ИО с вышеуказанным средним значением. В случаях наличия достаточного для статистики объема выборки, а также достаточной корреляции между значениями степеней достижения целей ИО и ухудшением значений целевых показателей объекта ИО, можно даже прогнозировать количественные значения ухудшения целевых показателей объекта ИО на текущий период времени.

Преимуществами предложенной методики являются:

1) Большая детализация модели – на фоне большого количества публикаций об объекте ИО вообще, изменения динамики количества публикаций, вызванные вбросом информации об одном из компонентов ИО, будут незначительными и, как следствие, не будут выявлены.

2) Увеличивается объем найденных тематических публикаций, так как запросов и ключевых слов будет больше.

3) Взвешенность компонентов ИО позволяет избежать ситуации, когда все компоненты имеют одинаковую

важность. Построенная таким образом модель ИО будет более адекватной.

4) Построенная один раз БЗ может использоваться в дальнейшем на протяжении значительного периода времени без необходимости заново проводить экспертизы.

5) Использование инструментария СРСОЭИ позволяет экспертам работать через глобальную сеть, что обеспечивает экономию времени и средств.

Недостатки предложенной методики заключаются в следующем:

1) Применение экспертных технологий требует временных и финансовых затрат на проведение групповой экспертизы. Кроме того, необходимо проводить своевременную актуализацию БЗ для ее повторного использования в будущем.

2) Сложность и, порой, неоднозначность представления некоторых достаточно сложных формулировок компонентов ИО в виде запросов в системе контент-мониторинга.

Пример использования методики

Продемонстрируем предложенную в предыдущем разделе методику на примере информационной операции против Национальной академии наук (НАН) Украины. Как известно, НАН Украины сейчас переживает не лучшие времена. В последние годы дела с финансированием все хуже и хуже: уменьшается бюджет НАН Украины и уменьшается доля бюджета НАН Украины в бюджете страны. Это видно из данных о распределении расходов Государственного бюджета Украины, например, для 2014 – 2016 годов [Закон України, 2014; Закон України, 2015; Закон України, 2016]. Предположим, что это уменьшение финансирования является результатом информационной операции против НАН Украины.

Групповая экспертная декомпозиция проходила в СРСОЭИ «Консенсус-2» [Циганок, 2017]. Далее была построена БЗ в СППР «Солон-3» [Тоценко, 2003]. Для экспертного оценивания использовалась СЭО «Уровень» («Рівень») [Циганок, 2012b].

В результате групповой экспертизы было получено 15

экспертных формулировок составляющих или компонентов ИО против НАН Украины, а именно:

- 1) Бюрократия в НАН Украины;
- 2) Неэффективная кадровая политика НАНУ;
- 3) Коррупция в НАН Украины;
- 4) Принижение уровня научных результатов НАН Украины;
- 5) Отсутствие внедрений научных разработок в производство;
- 6) Принижение уровня международного сотрудничества;
- 7) Нецелевое и неэффективное использование недвижимости НАНУ;
- 8) Нецелевое и неэффективное использование земельных ресурсов НАНУ;
- 9) Дискредитация президента НАН Украины;
- 10) Дискредитация управляющего делами НАН Украины;
- 11) Дискредитация других известных личностей НАН Украины;
- 12) Противопоставление научных результатов МОН и НАН;
- 13) Противопоставление научных результатов НАН Украины и других академических организаций;
- 14) Противопоставление достижений украинских фирм и НАН Украины;
- 15) Противопоставление научных результатов зарубежных организаций и НАН Украины.

Средствами СКМ InfoStream [Григорьев, 2007] проведен анализ динамики тематического информационного потока. Для этого, согласно каждому из вышеперечисленных компонентов ИО, на специализированном языке формируются запросы, по которым в дальнейшем и будет происходить анализ динамики публикаций по целевой тематике.

Проведён экспресс-анализ [Ландэ, 2016b] тематического информационного потока, который соответствует объекту ИО – НАН Украины. В результате экспресс-анализа средствами СКМ InfoStream был получен соответствующий тематический информационный поток из украинского

сегмента веб-пространства. Для выявления информационных вбросов с помощью имеющихся аналитических средств анализировалась динамика публикаций по целевой тематике.

Для выявления степени подобия фрагментов соответствующего временного ряда диаграмме ИО в разных масштабах был использован "вейвлет-анализ". Вейвлет коэффициенты показывают, насколько поведение процесса в определенной точке подобно вейвлету в определенном масштабе. На соответствующей вейвлет-спектрограмме видны все характерные особенности исходного ряда: масштаб и интенсивность периодических изменений, направление и значение трендов, наличие, расположение и продолжительность локальных особенностей.

Динамику ИО наиболее точно отражают вейвлеты "мексиканская шляпа" и Морле [Додонов, 2015b]. Поэтому анализируются временные ряды в соответствии с каждым из 15 компонентов ИО на протяжении 4-х периодов (01.01.2013 – 31.12.2013, 01.01.2014 – 31.12.2014, 01.01.2015 – 31.12.2015 и 01.01.2016 – 15.12.2016) и идентифицируется наличие вышеупомянутых вейвлетов.

На основании выявленных информационных вбросов (атак) и их параметров (расположение и продолжительность) инженер по знаниям дополнил БЗ СППР «Солон-3». В частности, был идентифицирован вброс по компоненту объекта ИО "Понижение научных результатов НАН Украины", осуществленный 30.11.2015, продолжительностью 14 дней. Соответственно, введен в качестве характеристики проекта "Понижение научных результатов НАН Украины" параметр «продолжительность выполнения проекта» сроком в 14 дней, а также введен в качестве характеристики влияния проекта "Понижение уровня научных результатов НАН Украины" на цель "дискредитации научных результатов НАН Украины" параметр задержки в распространении влияния на срок 10 месяцев. Для остальных выявленных информационных вбросов характеристики проектов и воздействий были введены аналогичным образом.

Таким образом, для периода 01.01.2015 – 31.12.2015 дополнена БЗ, которая имеет структуру, показанную детально на рис. 5.3, рис. 5.4 и рис. 5.5, соответственно.

Следует отметить, что для некоторых компонент ИО, а именно: "Коррупция в НАН Украины", "Бюрократия в НАН Украины", "Неэффективная кадровая политика НАНУ", "Нецелевое и неэффективное использование земельных ресурсов НАНУ" и "Нецелевое и неэффективное использование недвижимости НАНУ" были обнаружены на протяжении 2015 по 2 информационных вброса, поэтому для расчёта соответствующие проекты вводились в БЗ по 2 раза. Например: для компоненты ИО "Бюрократия в НАН Украины" – проекты "Бюрократия в НАН Украины 1" и "Бюрократия в НАН Украины 2", но каждый из них имеет различные характеристики продолжительности выполнения (9 и 15 дней) и соответствующие влияния имеют разные характеристики задержки в распространении (9 и 11 месяцев).



Рис. 5.3 – Фрагмент сети декомпозиции целей: "Дискредитация структуры НАН Украины" и "Дискредитация известных личностей НАН Украины"

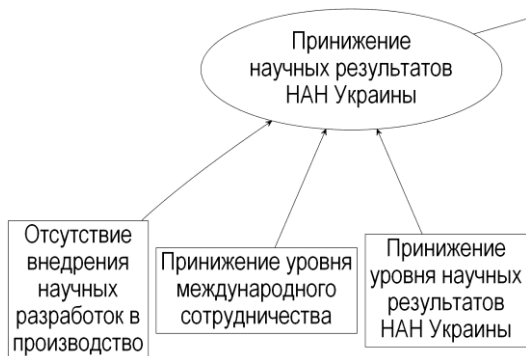


Рис. 5.4 – Декомпозиция целей "Принижение научных результатов НАН Украины" и "Дискредитация организационной структуры НАНУ"

Далее в СППР «Солон-3» вводятся степени выполнения проектов. Если для некоторых компонент ИО не было обнаружено никаких информационных вбросов, как, в частности, для "Противопоставления достижений украинских фирм и НАН Украины" и "Дискредитации деятельности Управления делами НАНУ", то для соответствующих проектов устанавливаются степени выполнения 0%. Для всех остальных проектов – 100%.

5.4. Методика построения баз знаний СППР при выявлении информационных операций

При обнаружении ИО следует учитывать влияние различных тематик публикаций на формирование информационного пространства.

Применение описанного в [Андрейчук, 2016; Andriichuk, 2017] подхода и соответствующей методики требует наличия группы экспертов. Работа экспертов стоит достаточно дорого и требует значительного времени. Поэтому актуальной проблемой является уменьшение использования экспертной информации в процессе построения БЗ СППР при обнаружении ИО.

Итак, целью данного исследования является разработка такой методики построения БЗ СППР при обнаружении

ИО, которая позволяет определять рейтинг информационного воздействия тематик публикаций и при этом как можно меньше использует экспертную информацию.

Сущность предлагаемой методики построения БЗ СППР при обнаружении ИО состоит в следующем:

1) Проводится групповая экспертиза по определению и декомпозиции целей информационной операции. Таким образом, происходит декомпозиция ИО как сложной слабо структурированной системы. Для этого используются средства СРСОЭИ.

2) Средствами СППР строится соответствующая БЗ на основании результатов проведенной средствами СРСОЭИ групповой экспертизы, а также имеющейся объективной информации. Для уточнения запросов к СКМ, а также дополнения БЗ СППР недостающими объектами и связями используется сеть ключевых слов предметной области соответствующей ИО.

3) Средствами СКМ проводится анализ динамики тематического информационного потока. БЗ СППР дополняется частичными коэффициентами влияния.

4) Средствами СППР на основании построенной БЗ рассчитываются рекомендации.

Пример использования методики

Предложенная методика продемонстрирована на примере темы Brexit, которая исследовалась научным сообществом [Bachmann, 2016].

На основе интерпретации базы данных, сформированной в результате работы СРСОЭИ «Консенсус-2», инженер по знаниям сформировал соответствующую БЗ СППР «Солон-3». Структура иерархии целей этой БЗ приведена на рис. 5.7.

Далее средствами СКМ InfoStream был проведен анализ динамики тематического информационного потока. Для этого, согласно каждой из компонент ИО, на специализированном языке были сформулированы запросы, по которым и происходил анализ динамики публикаций по целевой тематике.

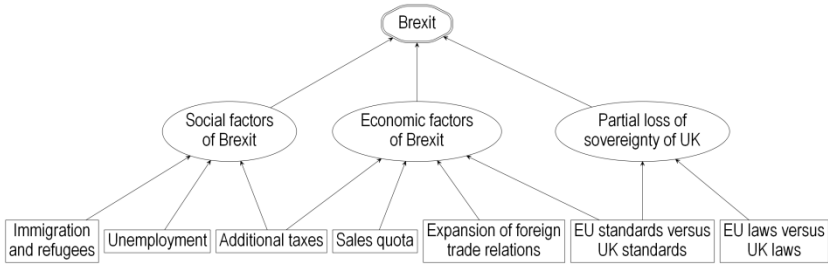


Рис. 5.7 – Структура иерархии целей БЗ

При формировании запросов, в соответствии с особенностями структуры иерархии целей (рис. 5.7), были использованы следующие правила:

- при прохождении сверху вниз, соответствующие запросы низших компонент ИО через знак "&" дополняются запросами более высоких компонент для уточнения;
- в случае абстрактности, неконкретности компонент ИО, осуществлялось прохождение снизу вверх, а соответствующие запросы через "|" дополнялись запросами более низких компонент;
- в случае конкретности компонент ИО запрос делался уникальным.

Для уточнения запросов СКМ и дополнения БЗ СППр недостающими объектами и связями использовалась сеть ключевых слов для Brexit, которая показана на рис. 5.8.

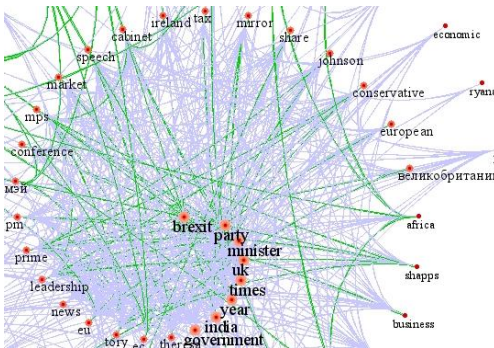


Рис. 5.8 – Сеть ключевых слов для тематики Brexit

На основе результатов выполнения запросов к системе InfoStream, а именно количества найденных документов для каждого из них, были вычислены соответствующие частичные коэффициенты влияния (ЧКВ). При вычислении значений ЧКВ использовалось следующее предположение – степень влияния компоненты ИО пропорциональна количеству соответствующих найденных документов.

Полученные значения ЧКВ были введены в БЗ. Таким образом, удалось избежать процедуры обращения к экспертам для оценки степени влияния компонент ИО.

На основе построенной и дополненной вышеописанным образом БЗ средствами СППР «Солон-3» были рассчитаны рекомендации. Получен следующий рейтинг информационного воздействия тематик публикаций: "Immigration and refugees" – 0.364, "EU standards versus UK standards" – 0.177, "EU laws versus UK laws" – 0.143, "Expansion of foreign trade relations" – 0.109, "Additional taxes" – 0.084, "Sales quota" – 0.08, "Unemployment" – 0.043.

5.5. Построение ранжирования объединенного множества альтернатив с учётом важности источников информации

В процессе информационно-аналитических исследований аналитики часто сталкиваются с задачей ранжирования множества объектов или альтернатив (продуктов на рынке, кандидатов или политических партий, участвующих в выборах, и так далее) по определенному критерию. Критерий может быть явно задан ЛПР или представлен в форме некоего тематического запроса (например «ведущие обувные магазины», «лучшие комические актеры» и так далее). Даже такие случайно выбранные формулировки демонстрируют, что проблематично подобрать альтернативы, удовлетворяющие соответствующим запросам, исключительно на основе количественных данных, поскольку нет количественного критерия, в соответствии с которым альтернативы можно было бы измерить. Распространенный подход, наиболее

явно подытоженный, вероятно, Томом Саати [Saaty, 1994], состоит в следующем: если вы не можете измерить альтернативы, то наилучший способ численно их описать – сравнить их друг с другом. Концептуально, сравнения альтернатив по определенному критерию можно разделить на кардинальные и ординальные оценки. Кардинальные парные сравнения альтернатив несут информацию о численном соотношении между ними, согласно заданному критерию, в то время как ординальные сравнения указывают лишь на порядок следования альтернатив в ранжировании. Кардинальные парные сравнения подходят для относительно небольшого числа альтернатив, расположенных в пределах одного порядка (согласно заданному критерию). Ординальные парные сравнения легче получать и обрабатывать, и они могут использоваться для составления ранжирований большого количества альтернатив.

Количество данных, используемых в информационно-аналитических исследованиях, обычно, велико. Поэтому чаще используются именно порядковые оценки альтернатив. В частности, это является одной из возможных причин, по которым поисковые сайты оперируют именно ранжированиями (отношениями порядка) ссылок, а не численно выраженными рейтингами какого-либо вида.

Для того чтобы составить ранжирование альтернатив на основании всей доступной информации, нужно принять во внимание ранжирования, предоставленные несколькими информационными источниками (ИИ), и агрегировать их каким-то образом. Чтобы моделировать слабо структурированную предметную область, следует учитывать данные из всех доступных источников. В данном разделе мы попытаемся рассмотреть наиболее общий случай, когда источники информации могут включать бумажную и электронную документацию, результаты поиска данных в Интернет, а также – экспертные оценки.

Проблема агрегации данных из разных источников состоит в том, что, в общем случае, источники имеют разные веса, зависящие, опять-таки, от ряда факторов. Например, разумно предположить, что более достоверному

источнику, предоставляющему большее количество информации, следует присвоить больший вес перед тем, как данные будут агрегироваться. Существует около десятка методов агрегации ранжирований (предложенных Борда, Кондорсе, Кемени, и другими), которые широко известны и описаны в массе публикаций, но проблема определения весов ИИ (голосующих, жюри, экспертов, поисковых сайтов и так далее) при агрегации данных не утратила актуальности, и по-прежнему является предметом открытой дискуссии.

Итак, в последующем изложении будут рассмотрены методы агрегации данных, представленных в форме ранжирований альтернатив, построенных разными ИИ, при чем особое внимание мы уделим проблеме определения относительной важности ИИ.

Описание проблемы

Проблема агрегации ранжирований стала актуальной еще в античные времена, когда начали возникать демократические сообщества и системы голосования. Подробный анализ правил голосования и их эволюции приведен, например, в [Rzazewski, 2014].

Первые методы агрегации ранжирований, которые возникли в процессе развития демократических процедур голосования, и не утратили актуальности до сегодняшнего дня, были предложены в конце 18-го века Борда [Borda, 1781] и Кондорсе [Condorcet, 1785]. С тех пор было разработано еще множество подходов к агрегации индивидуальных предпочтений на основе различных функций коллективного блага (social welfare). В период 1950-х и 1960-х годов была сформулирована теорема невозможности Эрроу [Arrow, 1970] и предприняты попытки преодолеть её ограничения [Kemeny, 1959; Copeland, 1951]. Ф. Алескеров в работе [Aleskerov, 1999] подытоживает правила агрегации, учитывающие требования теоремы невозможности Эрроу.

В 2000-х годах, когда широкое распространение получили поисковые и метапоисковые сайты, проблема агрегации ранжирований стала еще более актуальной, поскольку стало очевидным, что помимо данных, полученных от экспертов и аналитиков, необходимо

учитывать также и электронные источники (при этом участие человека часто сводится к минимуму). В данном контексте имеет смысл упомянуть несколько работ, напечатанных в этот период, в частности, [Dwork, 2001; Renda, 2003; Аборок, 2005]. Авторы этих публикаций предлагают и сравнивают несколько подходов к агрегации данных, полученных из электронных источников (из сети).

Определение весов ИИ и взвешенная агрегация ранжирований альтернатив представляют отдельную проблему. Методы Борда и Кондорсе можно без труда экстраполировать на случай, когда ИИ обладают разными весами, как показано Тоценко в [Totsenko, 2005]. Медиану Кемени более проблематично обобщить на случай разной весомости источников. Методы ординального факторного анализа, описанные в [Kadenko, 2008; Kadenko, 2013] позволяют вычислять веса критериев и источников на основе доступных наборов ранжирований, уже построенных экспертами ранее. Данный подход можно обобщить на случай вычисления весов электронных источников, если модераторы/аналитики (evaluators) априорно задают некое глобальное ранжирование альтернатив.

Дворк и соавторы в [Dwork, 2001] подчеркивают важность вовлеченности аналитиков в процесс предварительного взвешивания ИИ. Впрочем, указанная публикация посвящена, преимущественно, сравнительному анализу методов агрегации ранжирований, и не касается напрямую вычисления весов ИИ.

Говоря о весах ИИ, мы можем, опять-таки, упомянуть Тома Саати [Saaty, 1994] и его академическую школу (в том числе, Раманатана и Ганеша [Ramanathan, 1994], Формана и Пенивати [Forman, 1998], а также Янга и других [Yang, 2017]). Но их методы направлены, в первую очередь, на агрегацию экспертных оценок, представленных в форме матриц парных сравнений (МПС) в неких шкалах отношений (а не ранжирований).

В данном подразделе мы предложим подходы к определению весов ИИ на основе их предварительной оценки, а также – количества и качества информации, предоставленной ИИ.

Постановка задачи

Дано: n ИИ ($IS_1 - IS_n$). Каждый ИИ предоставляет ранжирование R_i , включающее m_i альтернатив ($i=1...n$). В роли ИИ могут выступать эксперты, поисковые сайты, аналитики, занимающиеся мониторингом контента, бумажная или электронная документация, и так далее. Подчеркнем, что число альтернатив, выдаваемых разными информационными источниками, в общем случае – различно ($m_i \neq m_j$; $i, j=1...n$).

Найти: агрегированное (глобальное) ранжирование альтернатив.

Идея решения для случая равной весомости источников информации

Предлагается строить алгоритм решения на основе методов агрегации индивидуальных ранжирований (ординальных оценок). Известно около десятка методов агрегации индивидуальных ранжирований (в общем случае выдающих различные результаты). Самые распространенные из них – правила Борда, Кондорсе, медиана Кемени. Напомним, что если в роли ИИ выступают эксперты, то число альтернатив, которые эксперт может сопоставить за один сеанс не должно превышать 7 ± 2 . Поскольку поисковые сайты выдают в качестве результатов поиска сотни (до 1000) ссылок, использование методов агрегации, которые оперируют матрицами доминирования, нецелесообразно в силу большой размерности (если речь идет, например о 10 ИИ, то анализировать придется 10 матриц 1000×1000). Поэтому наиболее простой и приемлемый вариант – использование метода агрегации Борда, в котором фигурируют не матрицы доминирования, а сами вектора ранжирований. Если все ИИ обладают одинаковыми весами, то можно использовать следующий алгоритм.

1) Найти длину ранжирования, включающего все уникальные альтернативы, входящие в индивидуальные

ранжирования $M = \text{card} \left(\bigcup_{i=1}^n \{A_j^{(i)}; j=1...m_i\} \right)$.

2) Сформировать унифицированное множество альтернатив. Для этого – к каждому ранжированию R_i m_i альтернатив нужно добавить все остальные альтернативы:

$\bigcup_{k=1}^n \{A_j^{(k)}; j=1..m_k\} / \{A_j^{(i)}; j=1..m_i\}$ с рангом m_i+1 . При унификации альтернатив, предоставленных разными ИИ, следует проверять их на предмет семантического подобия. Если ИИ – эксперты, то для этой цели можно использовать методы определения семантического подобия, предложенные в [Андрійчук, 2014].

3) В результате получим M альтернатив в каждом ранжировании. Каждое ранжирование R_i будет включать m_i альтернатив с разными рангами и $(M - m_i)$ альтернатив с одинаковым рангом (m_i+1) (аналогичный подход упоминался Дворк и соавторами в [Dwork, 2001]).

4) Просуммировать ранги каждой альтернативы по всем ИИ.

5) Отсортировать альтернативы в порядке увеличения сумм индивидуальных рангов. Результатом этой сортировки будет искомое ранжирование.

$$\sum_{j=1}^n r_{kj} > \sum_{j=1}^n r_{lj} \Rightarrow r_k < r_l, k, l = 1..M, \quad (5.1)$$

где r_{kj} и r_{lj} – ранги альтернатив A_k и A_l в ранжировании, заданном ИИ с номером j .

Использование других методов при меньших мощностях множества альтернатив

Если число альтернатив – на порядок меньше (несколько десятков), или ЛПР/пользователя интересуют лишь первые несколько альтернатив, выданных ИИ, то для обобщения результатов индивидуальных ранжирований имеет смысл применять другие методы агрегации.

Допустим, есть 10-15 ИИ, каждый из которых выдает ранжирование альтернатив; из этих альтернатив интерес представляют первые 10-20. Если все альтернативы при этом – разные, то после унификации множества альтернатив, мы получим до 300 (15×20) различных объектов. В действительности многие альтернативы,

фигурирующие в индивидуальных ранжированиях, предоставленных разными ИИ, совпадают (особенно если речь идет о поисковых сайтах). Итак, если у нас есть 15 ИИ и каждый из них выдает ранжирование 20 альтернатив, то мощность множества уникальных альтернатив будет составлять порядка 100 элементов.

Следовательно, в случае применения методов агрегации, оперирующих не векторами ранжирований (как описанный выше метод Борда), а матрицами ординальных парных сравнений (как метод Кондорсе и медиана Кемени), придется анализировать 15 квадратных матриц размерностью 100×100 . Поскольку отношение строгого порядка (ранжирования) антисимметрично (если альтернатива A_1 превосходит альтернативу A_2 , то A_2 уступает A_1), при агрегации анализируются только элементы матриц, стоящие выше главной диагонали $\{d_{ik}, i < k\}$. Поэтому, даже если в общем (итоговом) ранжировании будет 100 альтернатив, то анализировать в каждой матрице придется $(100 \times 99)/2$ или 4950 элементов. Итак, очевидно, что при малой мощности множества альтернатив можно использовать для агрегации результатов работы нескольких ИИ метод агрегации Кондорсе и/или медиану Кемени.

Недостаток метода Борда – в том, что при агрегации индивидуальных ранжирований происходит явный эвристический переход от ординальной шкалы к шкале отношений. Этот переход – неизбежен, поскольку, как доказывает Б.Г. Литвак в известной работе [Литвак, 1982], именно выражение оценок альтернатив в шкале отношений является необходимым и достаточным условием существования линейного обобщенного критерия. Так, например, альтернатива с рангом 2 не обязательно превосходит альтернативу с рангом 4 именно в 2 раза (напомним, что доминирующим альтернативам в ранжировании присваивается меньший ранг).

Метод Кондорсе, Марковские цепи и медиана Кемени позволяют избежать столь явного эвристического перехода от рангов альтернатив к шкале отношений.

Краткий обзор существующих методов

Метод Кондорсе. Основная идея метода Кондорсе состоит в том, что если в большинстве индивидуальных ранжирований альтернатива A_i превосходит альтернативу A_k , то и в итоговом ранжировании это отношение доминирования должно сохраняться.

Для построения итогового ранжирования необходимо сначала построить индивидуальные матрицы ординальных сравнений.

Если в ранжировании R_j альтернатива A_i превосходит A_k , то соответствующий элемент матрицы ординальных парных сравнений равен 1, если A_i уступает A_k – (-1), если альтернативы эквивалентны – 0:

$$d_{ik}^{(j)} = \begin{cases} -1, A_i < A_k \\ 0, A_i \equiv A_k \\ 1, A_i > A_k \end{cases} \quad (5.2)$$

Чтобы агрегировать ранжирования, предоставленные несколькими ИИ, по правилу Кондорсе, мы должны ограничить число альтернатив в итоговом ранжировании фиксированным числом M и построить ординальные МПС на основе ранжирований, предоставленных всеми ИИ: $\{D^{(j)} = \{d_{ik}^{(j)}; i, k = 1 \dots M\}; j = 1 \dots n\}$. После этого нужно вычислить суммы всех соответствующих элементов индивидуальных МПС и заполнить ординальную МПС, соответствующую глобальному (итоговому) отношению порядка (турниру).

$$D = \{d_{ik}; i, k = 1 \dots M\}, \quad (5.3)$$

где

$$d_{ik} = \begin{cases} 0, \sum_{j=1}^n d_{ik}^{(j)} = 0 \\ 1, \sum_{j=1}^n d_{ik}^{(j)} > 0 \\ -1, \sum_{j=1}^n d_{ik}^{(j)} < 0 \end{cases} .$$

После этого следует подсчитать строчные суммы элементов матрицы доминирования итогового отношения D и проранжировать полученные суммы.

$$\sum_{k=1}^M d_{ik} > \sum_{l=1}^M d_{lk} \Rightarrow r_i < r_l \quad (5.4),$$

где r_i и r_l – ранги альтернатив A_i и A_l в глобальном ранжировании R .

Недостатки метода Кондорсе – возможность появления равных рангов ссылок и нарушение транзитивности в итоговом отношении порядка вследствие так называемого «парадокса Кондорсе». Парадокс Кондорсе является частным случаем нарушения условий, предъявленных к правилам агрегации индивидуальных предпочтений Кеннетом Эрроу [Arrow, 1970]. Если возможно организовать обратную связь с ИИ, то транзитивность итогового отношения может быть восстановлена с помощью алгоритмов, описанных в [Tsyganok, 2010b; Kadenko, 2012].

В контексте агрегации результатов работы поисковых сайтов, преимуществом метода Кондорсе (о котором упоминают Дворк и соавторы в [Dwork, 2001]) является возможность «отфильтровывания» спама.

С точки зрения теоремы невозможности Эрроу, более «справедливым» результатом агрегации ранжирований является медиана Кемени.

Медиана Кемени. Медиану Кемени можно считать аналогом среднего арифметического для случая ординальных (порядковых, ранговых), а не кардинальных (количественных, численных) оценок.

Поскольку ординальные оценки не несут информации о количественном соотношении между альтернативами, обобщение евклидового расстояния и понятия центра масс на случай векторов ранжирований – нецелесообразно.

Расстояние между двумя ранжированиями заданного множества альтернатив зависит от числа перестановок, необходимого для того, чтобы получить из одного ранжирования – другое. Расстояние Кемени между двумя ранжированиями строится на основе метрики Хеминга. Если на множестве альтернатив $A=\{A_1..A_m\}$ заданы два ранжирования R_1, R_2 , и им соответствуют матрицы доминирования D_1 и D_2 , построенные по формуле (5.2), то

расстояние Кемени K будет рассчитываться следующим образом.

$$K(R_1, R_2) = \sum_{i=1}^m \sum_{k=1}^m |d_{ik}^{(1)} - d_{ik}^{(2)}| \quad (5.5),$$

Если задано множество из n ранжирований $\{R_1..R_n\}$ m альтернатив $A=\{A_1..A_m\}$, то, по определению [Кемени, 1959], медианой Кемени данного множества ранжирований будет ранжирование R :

$$R = \arg \min_{R \in T} \sum_{j=1}^n K(R, R_j) \quad (5.6),$$

где T – множество всех возможных ранжирований альтернатив. Очевидно, что процедура поиска медианы Кемени – весьма трудоемкая. Некоторые оценки сложности алгоритма ее расчета приведены в [Dwork, 2001]. Один из наиболее «простых» алгоритмов построения медианы Кемени описан в вышеупомянутой работе Литвака [Литвак, 1982].

Обобщение предлагаемых подходов на случай разных весов источников информации

Допустим, экспертным или каким-либо другим методом (например, на основе опыта, как показано в [Kadenko, 2008; Kadenko, 2013]) заданы разные веса ИИ $\{w_1..w_n\}$, соответствующие их «авторитетности». Тогда формулы для построения итогового ранжирования альтернатив приобретут несколько иной вид. Все ранжирования, выданные отдельными ИИ, будут учитываться с соответствующими весовыми коэффициентами. Так, формула (5.1) (метод Борда) будет выглядеть следующим образом.

$$\sum_{j=1}^n w_j r_{kj} > \sum_{j=1}^n w_j r_{lj} \Rightarrow r_k < r_l, k, l = 1..M \quad (5.7)$$

Формула (5.3) (метод Кондорсе) приобретет следующий вид.

$$D = \{d_{ik}; i, k = 1..M\}$$

где

$$d_{ik} = \begin{cases} 0, & \sum_{j=1}^n w_j d_{ik}^{(j)} = 0 \\ 1, & \sum_{j=1}^n w_j d_{ik}^{(j)} > 0 \\ -1, & \sum_{j=1}^n w_j d_{ik}^{(j)} < 0 \end{cases} . \quad (5.8)$$

В формуле (5.6) весовые коэффициенты станут множителями при соответствующих значениях расстояний Кемени.

$$R = \arg \min_{R \in T} \sum_{j=1}^n w_j K(R, R_j) \quad (5.9).$$

Обобщение методов, основанных на применении Марковских цепей, описанных в [Dwork, 2001], на случай разной весомости источников, представляет собой отдельную задачу, которую следует рассматривать в рамках отдельного исследования.

Опыт и многочисленные публикации показывают, что когда в процессе принятия решений принимает участие группа экспертов, необходимо принимать во внимание индивидуальные коэффициенты компетентности экспертов (если группа достаточно мала) (см. [Циганок, 2011], [Tsyganok, 2011], [Tsyganok, 2012]). Аналогично, если для построения ранжирования альтернатив по определенному информационному запросу (или в рамках некоего информационно-аналитического исследования) используются данные из нескольких ИИ, то их веса также следует учитывать при агрегации.

Теперь предлагается обратиться непосредственно к проблеме вычисления весов ИИ. В следующем пункте будет предложено несколько концептуальных подходов к ее решению.

Подходы к вычислению относительных весов источников информации

1) Определение весов на основе опыта. Данный подход был изложен в [Kadenko, 2008; Kadenko, 2013]. По сути, эксперты или аналитики задают собственное ранжирование альтернатив, а затем, на основе этого ранжирования и ранжирований, предоставленных ИИ, рассчитываются веса этих источников (предполагается, что агрегация производится методом Борда или Кондорсе).

2) Определение относительной весомости ИИ на основе количества и качества предоставленной информации. Разумно предположить, что относительный вес ИИ должен зависеть от количества и качества информации, предоставленной источником, в контексте заданного информационного запроса (в заданной предметной области). Критерии, отражающие качество и количество информационного поиска, соответствующего заданному запросу, следующие.

- Число альтернатив (ссылок, в случае поиска информации в сети), которые выдает ИИ в ответ на конкретный запрос;
- Релевантность этих альтернатив (ссылок), т.е., их соответствие конкретному запросу.

Релевантность результатов работы ИИ может определяться на основе оценок пользователей (модераторов, аналитиков, экспертов) из данной предметной области. Эти оценки ИИ могут основываться на предыдущем опыте обращения к ИИ с поисковыми запросами.

Можно рассматривать релевантность результатов работы ИИ в рамках поисковых запросов определенного типа (поиск информации на определенном языке, формул, изображений и т. п.). Соответственно, если возможно выделить тип запроса, то экспертная оценка для каждого типа запросов будет соответствовать опыту работы ИИ с данным типом запросов. В противном случае (выделить тип

запроса проблематично) имеет смысл использовать среднее значение экспертной оценки.

В контексте данного пункта мы предлагаем вычислять весомость ИИ аналогично компетентности эксперта. За основу метода определения весомости ИИ предлагается избрать подход к определению компетентности экспертов, предложенный Тоценко ([Тоценко, 2002], [Totsenko, 2002]). Согласно данному подходу, компетентность эксперта зависит от самооценки, объективной компоненты и взаимной оценки.

$$k = s(x_1b + x_2v), \quad (5.10)$$

где k – коэффициент компетентности эксперта, s – самооценка, b – объективная оценка, v – взаимооценка, x_1 , x_2 – коэффициенты относительной важности объективной и взаимной оценок соответственно.

В случае определения весомости обобщенных ИИ (экспертов, поисковых систем, и т. п.) в роли *взаимооценки* предлагается использовать взвешенное количество альтернатив, предоставленных i -м ИИ относительно суммарного количества альтернатив предоставленный всеми ИИ V_i :

$$V_i = \frac{m_i}{\sum_{k=1}^n m_k}, \quad (5.11)$$

где m_i – число альтернатив, предоставленных i -м ИИ, n – количество ИИ.

В роли *самооценки* будет выступать значение экспертной оценки E_i . Оно зависит от предыдущего опыта использования i -го ИИ. Определяется тип запроса (например: русский, английский, картинка, видео, формула, другой) и, в зависимости от него, берётся соответствующее относительное значение E_i .

В роли *объективной оценки* будет выступать взвешенное количество альтернатив, предоставленных i -м ИИ относительно общего числа *различных (уникальных)*

альтернатив, предоставленных всеми ИИ O_i (эффективность нахождения новой информации ИИ):

$$O_i = \frac{m_i}{P}, \quad (5.12)$$

где P – общее число *различных (уникальных)* альтернатив, предоставленных всеми ИИ, то есть мощность объединения результатов нахождения альтернатив по всем ИИ

$P = \text{card} \left(\bigcup_{i=1}^n \{A_j^{(i)}; j=1..m_i\} \right)$. В случае если ИИ являются экспертами, для определения совпадения найденных альтернатив по смыслу можно использовать метод смысловой идентификации, описанный в [Андрійчук, 2014].

Таким образом, весомость ИИ, по аналогии с компетентностью эксперта, будет вычисляться по следующей формуле:

$$w_i^* = E_i(x_1 O_i + x_2 V_i), \quad (5.13)$$

где w_i^* – ненормированный вес ИИ, x_1, x_2 – соответствующие коэффициенты относительной важности, E_i – нормированное значение экспертной оценки. Далее относительные значения весовостей ИИ нормируются:

$$w_i = \frac{w_i^*}{\sum_{i=1}^n w_i^*}, \quad (5.14)$$

где w_i – нормированный относительный вес ИИ.

Мы предлагаем исходить из эвристического предположения о том, что «объективная» и «взаимная» оценка ИИ представляют выпуклую комбинацию ($x_1 + x_2 = 1$), так что x_1 и x_2 будут лежать в диапазоне $[0;1]$ и меняться в зависимости от конкретного информационного запроса.

В результате поиска каждый ИИ выдает несколько альтернатив (в общем случае результаты, предоставленные разными ИИ, различны). На рис. 5.9 показан пример работы пяти ИИ, которые выдали 9 различных

альтернатив. При этом 1-я и 6-я альтернативы были предоставлены двумя ИИ, 2-я и 5-я альтернативы – тремя ИИ, 3-я альтернатива – пятью ИИ, остальные альтернативы – одним ИИ.

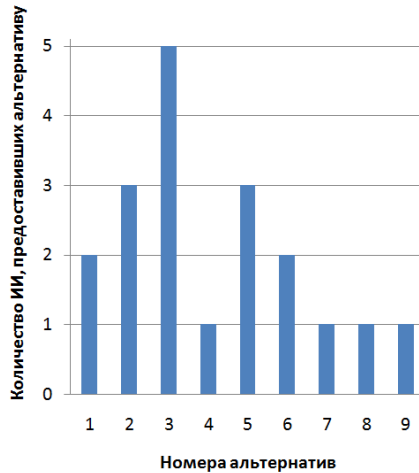


Рис. 5.9 – Пример объединения множеств альтернатив, предоставленных различными ИИ

Для иллюстрации зависимости коэффициентов χ_1 и χ_2 от результатов информационного поиска, рассмотрим несколько крайних случаев.

Крайний случай 1. Предположим, все ИИ предоставили полностью различные множества альтернатив. Например, обобщенные результаты поиска выглядят так, как показано на рис. 5.10: 5 ИИ выдали 9 различных альтернатив, причем каждую из альтернатив выдал только один ИИ (допустим: $IS_1: a_1, a_2$; $IS_2: a_3, a_4$; $IS_3: a_5$; $IS_4: a_6, a_7, a_8$; $IS_5: a_9$).

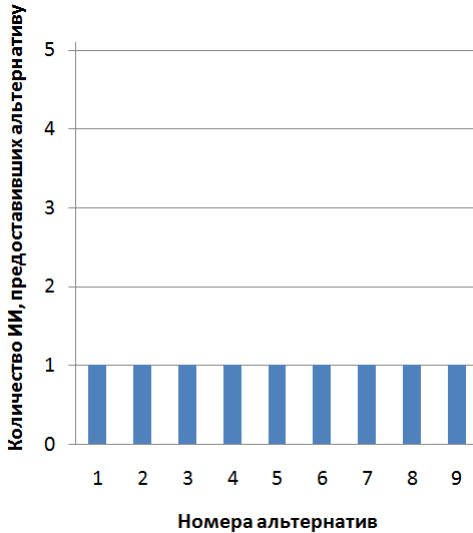


Рис. 5.10 – Пример объединения множества альтернатив в случае, когда все ИИ предоставили абсолютно разные альтернативы

Поскольку в данном информационном запросе все ИИ предоставили абсолютно разные альтернативы (все нашли различную, уникальную информацию), релевантность предоставленных альтернатив в контексте поискового запроса вызывает сомнение. Следовательно, в весомости ИИ целесообразно полагаться, преимущественно, на экспертную оценку, т.к. в ней отражена релевантность результатов работы ИИ, основанная на опыте предыдущих поисков, а также на эффективность нахождения источником новой информации.

Крайний случай 2. Предположим, все ИИ предоставили одни и те же альтернативы. Например, обобщенные результаты поиска выглядят так, как показано на рис. 5.11: 5 ИИ выдали по 9 одинаковых альтернатив, то есть каждую из альтернатив выдали (нашли) все ИИ (порядок, в котором альтернативы располагают различные ИИ, в общем случае, разный).

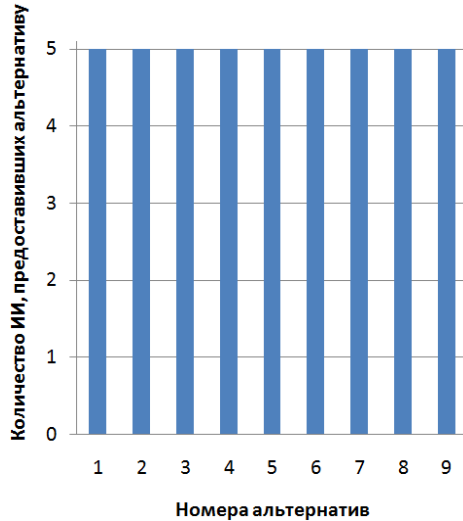


Рис. 5.11 – Пример объединения множеств альтернатив в случае, когда все ИИ предоставили одни и те же альтернативы

В данном случае весомость ИИ должна равняться E_i , поскольку по данному информационному запросу все ИИ предоставляют альтернативы одинаково эффективно и, соответственно, весомость может определяться (опять-таки) только по предыстории информационных поисков, содержащихся в экспертной оценке.

Для формального задания коэффициентов x_1 и x_2 с учетом изложенных соображений введем величину ρ , характеризующую плотность пересечения результатов предоставления альтернатив по всем ИИ.

$$\rho = \frac{\sum_{j=1}^P h_j}{nP}, \quad (5.15)$$

где h_j – количество ИИ, предоставивших j -ю альтернативу, n – количество ИИ.

В «крайнем случае 1» ρ принимает минимальное значение $\rho = \frac{1}{n}$.

В «крайнем случае 2» ρ принимает максимальное значение $\rho = 1$.

С учетом вышеизложенных требований, коэффициент x_2 предлагается вычислять по следующей формуле:

$$x_2 = \rho; \quad x_1 = 1 - x_2. \quad (5.16)$$

В «крайнем случае 1» нормированный относительный вес ИИ принимает значение $w_i = \frac{E_i((1 - \frac{1}{n})O_i + \frac{1}{n}V_i)}{\sum_{j=1}^n E_j((1 - \frac{1}{n})O_j + \frac{1}{n}V_j)}$, а в

«крайнем случае 2» $w_i = E_i$.

3) Статистический подход является, по сути, модификацией предыдущего. Для учета количества информации, предоставленной ИИ («объективная» компонента веса в формуле (5.13)), можно нормировать количества альтернатив по всем ИИ (см. формулу (5.11)), а затем – вычислить важность соответствующей компоненты (т.е. значение x_1) как дисперсию данного показателя (нормированного количества выданных источником альтернатив).

$$x_1 = D(V); \quad x_2 = 1 - x_1 \quad (5.17)$$

В этом случае, если все ИИ выдают равное число альтернатив, то соответствующей компонентой относительной весомости ИИ можно будет пренебречь, так как он не меняется от источника к источнику. Данная компонента начнет «играть роль» только в случае, когда количества альтернатив, выданные разными ИИ, будут отличаться.

5.6. Построение стратегии противодействия информационным операциям

Вероятно, что термин ИО стал популярен после рассекречивания множества документов Департамента обороны США, где ИО упрощенно определялись как «действия, направленные на влияние на информацию и информационные системы противника, на защиту собственной информации и информационных систем». Затем, в «Дорожной карте информационных операций» [1], термин был уточнен как «Интегрированное применение основных средств радиоэлектронной борьбы, операций в компьютерных сетях, психологических операций, военной маскировки и операций по обеспечению безопасности, в концепции со связанными с ними возможностями, с целью оказания влияния, разрушения, уничтожения или захвата у противника управления процессом принятия решений (как личностного, так и автоматизированного) при одновременной защите своих средств». Закладываемый в термин ИО смысл охватывает и раскрывает информационное влияние на массовое сознание (как на враждебное, так и на дружеское), влияние на информацию, доступную противнику и необходимую ему для принятия решений, а также на информационно-аналитические системы противника [2]. В современных условиях ИО, как неотъемлемая часть информационной войны, рассматривается в качестве нового вида боевых действий, активного противодействия в информационном пространстве, а информация при этом – в качестве потенциального оружия и цели для нанесения удара.

Принято различать два основных типа ИО – наступательные и оборонительные. Однако на практике большинство ИО являются смешанными, и большинство составляющих их процедур относятся одновременно к наступательным и оборонительным. Особенностью наступательных ИО (информационных атак) является то, что объекты влияния таких операций определены и планирование основывается на довольно точной информации об этих объектах. Информационная атака, чаще всего, требует нахождения или создания

информационного повода (для оборонительных ИО поводом может быть сама информационная атака противника), раскручивания этого повода, т.е. пропаганды (контрпропаганды при оборонительных ИО), а также мероприятий по противостоянию информационному воздействию. Таким образом, ИО, вне зависимости от ее типа, можно разделить на следующие этапы: оценка, планирование, выполнение и завершающая фаза. Далее, в соответствии с целью данного исследования, мы более детально рассмотрим оборонительную ИО, соответствующую доктрине большинства прогрессивно развивающихся государств.

Типичная оборонительная ИО охватывает такие основные этапы:

- Оценка:
 - Анализ возможных уязвимостей (целей);
 - Сбор информации о возможных операциях;
 - Определение возможных «заказчиков» информационных влияний:
 - определение сфер общего интереса объекта и потенциальных «заказчиков»;
 - ранжирование потенциальных заказчиков по их интересам;
- Планирование:
 - Стратегическое планирование оборонительной операции (явное или неявное):
 - определение критериев информационных влияний;
 - моделирование информационных влияний с учетом связей объекта, динамики влияния, «особых» (критических) точек влияния
 - прогнозирование следующих шагов;
 - расчет последствий;
 - Тактическое планирование контр-операций;
- Выполнение – реализация информационного влияния:
 - Выявление и «сглаживание» информационного повода;

- Контрпропаганда;
- Оперативная разведка;
- Оценка информационной среды;
- Корректирование информационного противодействия;
- Завершающая фаза:
 - Анализ эффективности;
 - Использование положительных результатов информационного влияния;
 - Противодействие отрицательным результатам.

Как можно видеть из предложенной детализации оборонительной ИО, основополагающей компонентой является стратегическое планирование. Очевидно, не существует единого «стандартного» плана проведения ИО. Можно лишь рассмотреть образцовую, полученную путем обобщения некоторых уже реализованных ИО, последовательность действий при их осуществлении. При чем выбор оптимального набора таких мероприятий в определенный момент времени зависит, в первую очередь, от наличия ресурсов для их проведения в этот текущий момент, а так же от результатов выполнения ранее выбранных мероприятий. Оптимальность здесь следует рассматривать с точки зрения эффективности достижения целей той, или иной оборонительной ИО.

Целью данного исследования является усовершенствование имеющегося аппарата поддержки принятия решений (ППР) с учетом особенностей процесса стратегического планирования в слабо структурированных предметных областях. В контексте данной работы была поставлена более конкретная цель: разработать комплексную методику ППР, позволяющую повысить качество процесса стратегического планирования оборонительной ИО. Предлагается методика формального построения стратегии ИО с привлечением группы специалистов, компетентных в данной области. На основе современных методов экспертной ППР предлагается возможность наиболее полного получения не искаженных знаний от специалистов, и использования их для построения адекватной модели предметной области. Работу

методики предлагается продемонстрировать на иллюстративном примере.

Содержание и общие этапы процесса построения стратегии

Как известно, в общем понимании стратегия представляет собой не детализированный план действий, рассчитанный на продолжительный период времени и направленный на достижение определенной главной цели. В то же время, план должен быть гибким, конструктивным, стойким к неопределенности условий среды, и предусматривать конкретизацию путем декомпозиции главной цели.

В слабо структурированных предметных областях, к которым относятся управление, охрана окружающей среды, производство, социальная сфера и др., присутствует актуальная проблема построения долгосрочных не детализированных планов деятельности. Не возникает сомнений в том, что при создании таких стратегических планов нужно опираться на все имеющиеся знания в определенной предметной области. Поскольку, знания в каждой такой области не являются полностью формализованными и, поэтому, большей частью, находятся лишь в головах специалистов, то было бы неразумно при планировании не использовать информацию, полученную от экспертов. Тем более, было бы неосмотрительным сводить оценки вариантов планирования только лишь к количественным (например, финансовым) показателям. Чтобы иметь реалистичные долгосрочные планы, их нужно адаптировать к неминуемым изменениям текущей ситуации и учитывать наличие ресурсов для их осуществления, необходимых в каждый определенный момент. Поэтому стратегические планы могут быть рациональными лишь на определенном интервале времени.

Цель данного исследования – создать технологию, которая бы включала формальные механизмы построения стратегических планов в слабо структурированных предметных областях с привлечением групп экспертов и инженеров по знаниям.

Учитывая перечисленные требования к стратегиям, а именно необходимость в реалистичных и динамических

планах, предлагается при их построении использовать инструментарий распределения ограниченных ресурсов между определенными предложенными мероприятиями. Ресурсы распределяются на заданный момент времени, в зависимости от потенциального вклада определенного мероприятия в достижение стратегической цели. Фактически, результаты проведенной работы должны давать ответ на вопрос: «какие мероприятия должны быть выполнены при текущих условиях для наиболее эффективного достижения стратегической цели?».

С учетом вышеуказанного, разработанная технология построения стратегии предусматривает несколько этапов.

1) Построение базы знаний (БЗ).

Этот этап реализован в виде веб-ориентированной программной системы, которая позволяет АПР, инженерам по знаниям и экспертам работать удаленно для создания БЗ без необходимости собираться вместе.

Этап включает ряд под-этапов:

а. Подбор групп экспертов для проведения экспертизы.

Задача выбора экспертов в общем случае возлагается на АПР и на инженеров по знаниям. Причем, в рамках экспертизы при решении разных вопросов формируются разные группы специалистов, наиболее компетентных в каждой определенной области.

б. Построение (в ходе диалога с экспертами) иерархии целей, которая описывает предметную область.

На этом под-этапе АПР формулирует стратегическую цель, которая, в ходе проведения экспертиз инженерами по знаниям, подлежит декомпозиции на локальные цели (факторы), которые существенно влияют на ее достижение. В процессе декомпозиции, эксперты согласовывают свои мнения относительно состава множеств факторов влияния на ту или иную цель и приходят к консенсусу в каждом вопросе. Инженеры по знаниям, наделенные функциями организаторов экспертиз, для каждой декомпозиции локальной цели формируют отдельную группу экспертов.

Решение о достаточном уровне детализации и прекращение дальнейшей декомпозиции стратегической

цели принимают организаторы экспертизы в случае, когда нижний уровень иерархии целей будут составлять только лишь цели (факторы), представляющие собой готовые к реализации конкретные мероприятия (проекты).

Результатом данного этапа построения стратегии является иерархическая структура, которая, в соответствии с мнением данной экспертной группы, в полной мере описывает предметную область.

в. Оценка экспертами относительных влияний целей в иерархии.

Относительное влияние каждой цели в графе иерархии определяется инженером по знаниям в случае наличия достоверных знаний об уровне этого влияния на достижение заданной цели, или же, в противном случае, – группой экспертов путем парных сравнений влияний целей (факторов).

Результатом данного под-этапа являются относительные величины взаимных влияний целей, полученные в результате агрегации индивидуальных экспертных оценок, выполненных в шкалах разной детальности и представленных в виде неполных матриц парных сравнений (МПС) влияний в рамках группы. Агрегацию предлагается производить комбинаторным методом [7], преимущества которого в эффективности по сравнению с другими методами, подтверждены соответствующим экспериментальным исследованием [8].

Этот метод агрегации имеет несколько преимуществ над имеющимися подходами к обработке МПС:

- В методе максимально полно используется избыточность информации.

- Метод позволяет определять весомость альтернатив в случаях, когда некоторые элементы МПС отсутствуют (не заданы). Т.е., для определения весов альтернатив не является обязательным требованием наличие всех парных сравнений в матрицах. Необходимым условием является лишь связность графа, соответствующего обобщенной МПС.

- Метод является одноэтапным (в отличие от известных подходов, применяемых для вычисления весов в групповых методах оценивания [9]). Агрегация парных сравнений в таких групповых методах ППР представляет собой

двухэтапную процедуру: или (1) сначала агрегируются индивидуальные МПС, а потом – на основе обобщенной матрицы вычисляется вектор весов альтернатив, или (2) сначала по каждой МПС вычисляется вектор весов, а потом все векторы агрегируются. В случае (1) согласованность всех индивидуальных МПС не гарантирует согласованности итоговой МПС. В случае (2) согласованность индивидуальных МПС не гарантирует согласованности векторов весов, вычисленных по каждой МПС. Если уровень согласованности не достаточен для корректного выполнения агрегирования и расчета весов, то в таком случае двухэтапность процедур делает невозможной организацию обратной связи с экспертами для повышения согласованности. При использовании комбинаторного метода нет надобности в поэтапном достижении желательного уровня согласованности и, поэтому, не имеет места конфликт между двумя последовательными процессами согласования. Если нужно повысить согласованность парных сравнений, то определенные элементы индивидуальных МПС корректируются, с согласия экспертов, сформировавших соответствующие МПС.

Заметим, что агрегацию допустимо выполнять лишь при достаточной согласованности [10] экспертных суждений. Для оценки уровня согласованности парных сравнений предлагается использовать Дважды-Энтропийный индекс согласованности [11], который на основе сформированного спектра экспертных оценок весов каждого из влияний определяет степень согласованности и отвечает всем поставленным требованиям, что выгодно отличает его от других известных индексов. В случае недостаточной согласованности метод предусматривает возможность повышения ее уровня путем обратной связи с экспертами.

На этом построение БЗ заканчивается и начинается этап построения оптимальной стратегии на основе знаний, заложенных в БЗ.

2) Определение оптимальной стратегии.

Очевидно, что чем больше весомость определенного проекта, или мероприятия, тем существеннее он влияет на достижение стратегической цели. Поэтому, направление ресурсов на этот проект будет приносить более весомые и

ощутимые результаты. В тот же время, не следует выделять на проект ресурсов меньше, чем необходимо для его старта и выполнения. Следовательно, в качестве оптимальной стратегии предлагается избирать оптимальный вариант распределения ресурсов между проектами (т.е. тот, который обеспечивает наиболее эффективное достижение стратегической цели).

Задача выбора оптимального распределения ресурсов между проектами является предметом отдельного исследования. Следует отметить, что поскольку проекты могут характеризоваться разными сроками реализации и, кроме того, цели могут иметь разные временные задержки влияния на главную цель, то оптимальное распределение ресурсов имеет место только для определенного заданного момента времени. Благодаря применению метода целевого динамического оценивания альтернатив [12, 13] в рамках заданного стратегического плана есть возможность оценить и сравнить совершенно разноплановые проекты/мероприятия: те, что дают моментальный результат, с теми, эффект от выполнения которых может проявиться в далекой стратегической перспективе. Другой важный параметр, который характеризует проекты – диапазон необходимых объемов ресурсов. Например, если минимальный необходимый объем финансирования для осуществления проекта составляет 1 млн. грн., а запрашиваемый – 2 млн. грн., то нет смысла выделять на этот проект сумму, которая не принадлежит этому диапазону.

Учитывая указанные особенности, наиболее рациональным способом решения задачи распределения ресурсов между проектами на заданный период является целенаправленный перебор всех возможных распределений ресурсов с заданной точностью (предположим, до 10 тыс. грн.), например, с использованием Генетического алгоритма [14].

В зависимости от сложности предметной области и сформулированной цели, которая должна быть достигнута, процесс построения стратегического плана может быть более простым, или более сложным. Впрочем, предложенный математический аппарат и разработанные программные средства ППР дают возможность, опираясь

на все доступные знания о предметной области, создавать довольно масштабные и содержательные, а главное, реалистичные, перспективные планы.

Пример построения стратегии

Далее предлагается гипотетический пример, который показывает заключительные этапы процесса построения оптимального стратегического плана по обеспечению противодействия информационным операциям на 5-летнюю перспективу, при условии наличия финансовых ресурсов в объеме 200 млн. грн.

В рамках примера считаем, что иерархия с главной целью «Обеспечить достаточный уровень противодействия информационным атакам» уже построена, и продолжается под-этап в) построения стратегического плана – оценка относительных влияний проектов на некоторую цель из графа иерархии целей.

Допустим, что оценивание на этом этапе примера выполняется группой из трех равно-компетентных экспертов. Каждому эксперту формально предоставляется возможность определить наличие предпочтения в каждой паре из 4-х проектов – выполнить ординальное сравнение („>” – больше; „<” – меньше), определиться с вербальной шкалой оценивания, выбрать количество делений для этой шкалы и, собственно, избрать номер конкретного деления.

В таблице 5.2 приведены данные экспертного оценивания важности мероприятий, входящих в состав цели «Реализация информационного влияния»: C_1 – Выявление и «сглаживание» информационного повода; C_2 – Контрпропаганда; C_3 – Оперативная разведка; и C_4 – Оценка информационной среды. Символом «*» в матрицах обозначены элементы, по которым эксперты, по той или иной причине, не предоставили информации.

Таблица 5.2. Пример экспертного оценивания относительных влияний проектов

	Эксперт 1				Эксперт 2				Эксперт 3						
		C ₁	C ₂	C ₃	C ₄		C ₁	C ₂	C ₃	C ₄		C ₁	C ₂	C ₃	C ₄
Ординальное сравнение	C ₁	1	>	>	>	C ₁	1	>	<	>	C ₁	1	*	>	>
	C ₂		1	<	>	C ₂		1	<	>	C ₂		1	<	<
	C ₃			1	>	C ₃			1	*	C ₃			1	>
	C ₄				1	C ₄				1	C ₄				1
Количество делений шкалы		C ₁	C ₂	C ₃	C ₄		C ₁	C ₂	C ₃	C ₄		C ₁	C ₂	C ₃	C ₄
	C ₁	1	5	9	9	C ₁	1	5	5	9	C ₁	1	*	8	9
	C ₂		1	5	9	C ₂		1	7	9	C ₂		1	9	3
	C ₃			1	7	C ₃			1	*	C ₃			1	9
Номер деления		C ₁	C ₂	C ₃	C ₄		C ₁	C ₂	C ₃	C ₄		C ₁	C ₂	C ₃	C ₄
	C ₁	1	2	3	5	C ₁	1	3	2	5	C ₁	1	*	4	8
	C ₂		1	2	2	C ₂		1	4	5	C ₂		1	2	2
	C ₃			1	5	C ₃			1	*	C ₃			1	3
Номер шкалы		C ₁	C ₂	C ₃	C ₄		C ₁	C ₂	C ₃	C ₄		C ₁	C ₂	C ₃	C ₄
	C ₁	1	1	3	1	C ₁	1	3	4	1	C ₁	1	*	2	3
	C ₂		1	2	1	C ₂		1	4	2	C ₂		1	3	5
	C ₃			1	2	C ₃			1	*	C ₃			1	1
Унифицированные значения парных сравнений		C ₁	C ₂	C ₃	C ₄		C ₁	C ₂	C ₃	C ₄		C ₁	C ₂	C ₃	C ₄
	C ₁	1	2.500	1.732	5.000	C ₁	1	2.615	0.833	5.000	C ₁	1	*	2.011	6.839
	C ₂		1	0.739	2.000	C ₂		1	0.574	2.333	C ₂		1	0.760	0.726
	C ₃			1	3.138	C ₃			1	*	C ₃			1	3.000
C ₄				1	C ₄				1	C ₄				1	

На основе унифицированных значений парных сравнений (нижний ряд матриц таблицы 5.2) вычисляются относительные веса влияний проектов (таблица 5.3).

Таблица 5.3. Рассчитанные относительные веса влияний проектов

Пометка проекта	Нормализованное значение веса
C ₁	0.4455
C ₂	0.1743
C ₃	0.2919
C ₄	0.0883

Для построения оптимальной стратегии на 5-летнюю перспективу была использована СППР «Солон-3» (см. рис. 5.12).

Загальна кількість ресурсів		200 000,00		з точністю до		10,00		(значення не менше 3,05)	
№	Назва проекту	R min	% min	R max	% max	виділено			
1	Моделирование информационных влияний с учетом "особых" (критических) точек влияния	10000	0	0	100	0,000			
2	Прогнозирование следующих шагов	15000	0	0	100	0,000			
3	Расчет следствий	12000	40	25000	100	0,000			
4	Тактическое планирование контр-операций	10000	30	30000	100	0,000			
5	Выявление и "сглаживание" информационного повода	12000	55	20000	100	20000,000			
6	Контрпропаганда	50000	50	75000	100	0,000			
7	Оперативная разведка	10000	40	25000	100	25000,000			
8	Оценка информационной среды	5000	40	12000	100	12000,000			
9	Корректирование информационного противодействия	5000	20	15000	100	15000,000			
10	Анализ эффективности противодействия информационной атаке с юга	2000	20	8000	100	8000,000			
11	Использование положительных результатов информационного влияния	5000	80	7000	100	7000,000			
12	Противодействие отрицательным результатам информационной атаки с востока	15000	55	25000	100	17990,000			
13	Противодействие отрицательным результатам информационной атаки с севера	20000	25	20000	100	0,000			

Мінімальна кількість ресурсів необхідна для виконання

Рис. 5.12 – Рассчитанное распределение ресурсов между проектами

Из рис. 5.12 можно увидеть, что для каждого проекта, который претендует на финансирование, вводятся экспертные оценки: минимально необходимое количество ресурсов для существования проекта (R_{min}), процент выполнения проекта при минимальном финансировании ($\%_{min}$), количество ресурсов, которое запрашивается (R_{max}) и запланированный процент выполнения при этом ($\%_{max}$ – обычно равняется 100%). После выполнения расчетов, количества выделенных ресурсов отображаются в колонке «выделено».

Список рекомендованных действий для ЛПР в виде набора проектов с рассчитанными объемами финансирования будет базисом для оптимального стратегического плана по обеспечению достаточного уровня противодействия информационным атакам в 5-летней перспективе при условии обеспечения определенным количеством финансовых ресурсов.

5.7. Концепция создания информационно-аналитической системы выявления информационных операций

На рис. 5.13 представлена Use Case диаграмма концепции информационно-аналитической системы

обнаружения информационных операций, представляющая собой реализацию приведенных выше методик путем системной интеграции СРСОЭИ, СЭО, СППР и СКМ.

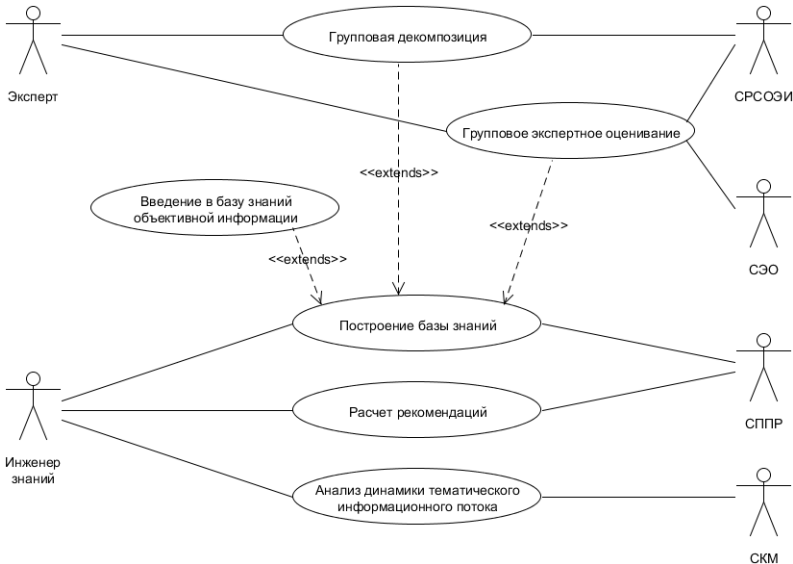


Рис. 5.13– Use Case диаграмма концепции информационно-аналитической системы обнаружения информационных операций

Актеры:

- Эксперт – специалист, которого приглашают или нанимают за денежное вознаграждение для предоставления квалифицированного заключения, оценки или суждения по некоторым вопросам.

- Инженер по знаниям – специалист, который формирует БЗ и рассчитывает на ее основе рекомендации с помощью инструментария СППР.

- СРСОЭИ – система для дистанционной групповой работы экспертов в глобальной сети.

- СЭО – комплекс программных средств, который гибко адаптируется к уровню компетентности экспертов и позволяет получать знания от них в полной мере и без искажения.

– СППР – система, которая рассчитывает и выдает рекомендации, основываясь на введенных в ее БЗ знаниях (как объективных, так и экспертных).

– СКМ – система, обеспечивающая автоматизированный сбор информации с веб-сайтов в режиме реального времени, ее структурирование, группировку по семантическим признакам, а также тематическое избирательное распределение, предоставление доступа к информационным базам данных в поисковых режимах и анализ динамики тематического информационного потока.

Варианты применения:

– Построение базы знаний. Процесс инициируется инженером по знаниям. В рамках данного процесса средствами СППР строится проблемно-ориентированная модель предметной области на основании знаний, полученных от экспертов, и результатов контент-мониторинга глобальной сети.

– Групповая декомпозиция. Процесс инициируется инженером по знаниям. Эксперты принимают участие в групповой декомпозиции целей. Согласно сформулированной главной цели ИО, инженер по знаниям формирует группу экспертов-специалистов, компетентных в вопросах, которые будут рассматриваться. Процесс групповой декомпозиции происходит в ходе поэтапного диалога с экспертами из группы путем последовательного разбиения главной цели на составляющие. На каждом этапе декомпозиции эксперту из группы предлагается сформулировать множество целей (выбрать цели среди существующих или ввести новые), которые непосредственно влияют на раскрываемую цель. Далее, аналогичным образом, происходит декомпозиция других целей. Процесс декомпозиции завершается, когда в качестве целей выступают конкретные компоненты ИО, обладающие формулировками, на основании которых можно сделать соответствующие запросы для СКМ.

– Групповая экспертная оценка. Процесс инициируется инженером по знаниям, когда закончилась групповая декомпозиция. В ходе этого процесса для каждой группы подцелей определяется характер их воздействий

(качественный или количественный, положительный или отрицательный) и величина их степеней влияния.

– Ввод в БЗ объективной информации. Процесс инициируется инженером по знаниям. В ходе этого процесса в БЗ может вводиться информация, которая может быть получена путем измерения определенных величин.

– Расчет рекомендаций. Процесс инициируется инженером по знаниям. В ходе этого процесса, на основе построенной БЗ, производится расчет относительной эффективности информационных вбросов по каждой из компонент ИО, то есть «относительного вклада» каждого из таких информационных вбросов в достижение главной цели ИО. Также рассчитывается интегральная оценка эффективности информационных вбросов и степени достижения целей. Учитывается временная динамика. На основании анализа ретроспективы предполагается изменение значений целевых показателей объекта на текущий период.

– Центральной компонентой концепции информационно-аналитической системы обнаружения информационных операций является СППР. А поскольку основной компонентой СППР является БЗ, рассмотрим ее структуру, которая приведена на рис. 5.14. Основными элементами БЗ являются объекты и связи между ними. Объектами БЗ могут быть цели и проекты. В нашем случае проектам соответствуют конкретные компоненты ИО.

– Объект БЗ называется в виде короткой формулировки. Может быть количественным или качественным, пороговым или квазилинейным. Для проектов указываются продолжительность выполнения и необходимые объемы ресурсов.



Рис. 5.14 – Структура БЗ экспертной СППР

Связь между объектами БЗ может: быть положительной или отрицательной, иметь временную задержку, определять принадлежность объекта к группе совместимости. Также она характеризуется частным коэффициентом влияния.

Заключение

Для эффективного анализа современных информационных процессов на основе мониторинга информационных потоков из глобальных компьютерных сетей должны применяться современные методы, базирующиеся на нелинейном анализе, многие из которых нашли успешное применение в естественных науках. Современные подходы позволяют применять для анализа и моделирования даже общественных и информационных систем методы, апробированные в первую очередь в естественных науках. Анализ информационных потоков выступает фундаментом таких направлений как моделирование, проектирование и прогнозирование.

В работе предложено и обосновано применение методики и информационной технологии обнаружения информационных кампаний и операций на основе анализа статистических параметров распределения характеристик «жизненного цикла» информационных сообщений.

Рассмотренные в работе подходы позволяют описывать информационные процессы, процессы информационного влияния, они пригодны для описания общих тенденций в динамике информационных процессов. При этом продвижение в освоении современного информационного пространства невозможно без общих представлений о структуре и свойствах динамики сетевых информационных процессов, что в свою очередь требует выявления и учета их устойчивых закономерностей. Следует отметить, что часто подходы, которые базируются на применении точных методов и математическом формализме, а также методов компьютерного моделирования, в действительности, могут давать преимущественно качественные выводы, что обуславливается многопараметричностью реальных моделей. Вместе с тем, даже такие результаты могут объяснить реальность во многих случаях лучше, чем традиционные качественные методы.

Методы, алгоритмы, аналитические инструменты, которые рассматриваются в данной работе, выступают не только в качестве демонстрационной основы для объяснения реально происходящих событий и процессов, но

и как необходимые компоненты при их планировании и прогнозировании.

В то же время, приведенные модели и методы пригодны для описания общих тенденций динамики информационных процессов, распознавания информационных операций, 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, составляет основу научной методологии. В настоящее время только ретроспективный анализ уже реализованных информационных операций остается относительно надежным способом их верификации.

Литература

[Addison, 2017] Addison, Paul S. The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance, Second edition / Addison, Paul S. – Boca Raton, FL : CRC Press, Taylor & Francis Group, 2016. – 446 p.

[Aguiar-Conraria, 2008] Aguiar-Conraria, L. Using Wavelets to Decompose the Time–Frequency Effects of Monetary Policy / Luis Aguiar-Conrariaa, Nuno Azevedob, Maria Joana Soares // Physica A: Statistical Mechanics and its Applications. – 2008. – Vol. 387. – Issue 12. –P. 2863–2878.

[Aldroubi, 2016] Aldroubi, A. New Trends in Applied Harmonic Analysis: Sparse Representations, Compressed Sensing, and Multifractal Analysis / Akram Aldroubi, Carlos Cabrelli, Stephane Jaffard, Ursula Molter (eds.). – Birkhäuser Basel, 2016. – 334 p. – (Applied and Numerical Harmonic Analysis)

[Aleskerov, 1999] Aleskerov, Fuad T. Arrovian Aggregation Models / Fuad T. Aleskerov. Springer, 1999 – 226 p.

[Andriichuk, 2017] Andriichuk O.V., Kachanov P.T. Usage of expert decision-making support systems in information operations detection / Proceedings of the International Symposium for the Open Semantic Technologies for Intelligent Systems (OSTIS 2017), Minsk, Republic of Belarus, 16-18 February, 2017, pp. 359-364.

[Arrow, 1970] Arrow, K. J. Social Choice and Individual Values / K.J. Arrow – 2nd edition. Yale University Press – 1970 – 144 p.

[ATP, 2012] ATP 2-22.9. Army Techniques Publication No. 2-22.9 (FMI 2-22.9). Headquarters Department of the Army Washington, DC, 10 July 2012.

[Bachmann, 2016] Veit Bachmann, James D. Sidaway Brexit geopolitics / Geoforum 77 (2016) pp. 47–50.

[Ball, 1976] Ball-Rokeach, S.J. & DeFleur, M.L. (1976). A dependency model of mass-media effects. Communication Research, 3, 3–2.

[Berinsky, 2017] Berinsky A. J. Rumors and Health Care Reform: Experiments in Political Misinformation, British

Journal of Political Science, Volume 47, Issue 2 April 2017, pp. 241-262.

[Borda, 1781] J.C. Borda. Memoire sur les elections au scrutin. Histoire de l'Academie Royale des Sciences, 1781.

[Box, 2015] Box, G. E. P. Time Series Analysis, Forecasting and Control / George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung. – New Jersey: John Wiley & Sons, 2015. – 712 p.

[Braichevsky, 2007] Braichevsky S., Lande D., Snarskii A. On the fractal nature of mutual relevance sequences in the Internet news message flows // arXiv preprint arXiv: 0710.0228, 2007.

[Broadbent, 1957] Broadbent S.R., Hammersley J.M. Percolation processes // I. Crystals and mazes, Proc Cambridge Philos. Soc. – pp. 629-641. – 1957.

[Brosius, 1992] Brosius, H.B. & Kepplinger, H.M. (1992). Beyond agenda-setting: The influence of partisanship and television. Journalism Quarterly, 69, 893–902.

[Cayley, 1889] Cayley A. A Theorem on Trees. Quarterly Journal of Mathematics. 1889. Vol. 23. P. 376–378.

[Chatfield, 2004] Chatfield, C. The analysis of time series: an introduction 6th ed. / Chris Chatfield. – CHAPMAN & HALL/CRC, 2004. – 333 p.

[Condorcet, 1785] M.-J. Condorcet. Essai sur l'application de l'analyse a la probabillite des decisions rendues a la pluralite des voix, 1785.

[Copeland, 1951] Copeland A. H. A reasonable social welfare function. / A. H. Copeland. Mimeo, University of Michigan, 1951.

[Corso, 2005] G.M. Del Corso, A. Gulli, F. Romani. Ranking a stream of news. In Proceedings of 14th International World Wide Web Conference, pp. 97-106, Chiba, Japan, 2005.

[Dey, 2010] Dey, D. Cross-wavelet Transform as a New Paradigm for Feature Extraction from Noisy Partial Discharge Pulses / D. Dey, B. Chatterjee, S. Chakravorti, S. Munshi // IEEE Transactions on Dielectrics and Electrical Insulation. – 2010. – Vol. 17. – No. 1. – P. 157–166.

[DNI, 2006] DNI Intelligence Community Directive 301 – "National Open Source Enterprise". July 2006.

[Dreyfus, 1980] Dreyfus S.E., Dreyfus H.L. A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition (PDF). Washington, DC: Storming Media, 1980. URL: https://www.researchgate.net/publication/235125013_A_Five-Stage_Model_of_the_Mental_Activities_Involved_in_Directed_Skill_Acquisition.

[Dreyfus, 2004] Dreyfus S.E. The Five-Stage Model of Adult Skill Acquisition. *Bulletin of Science Technology & Society* 2004 24: 177. DOI: 10.1177/0270467604264992.

[Dwork, 2001] Dwork C. Rank aggregation methods for the Web / C. Dwork, R. Kumar, M. Naor, D. Sivakumar // WWW '01 Proceedings of the 10th international conference on World Wide Web – 2001 – Pages 613-622.

[Feder, 1991] Feder J. *Fractals*. – Springer Science + Business Media, LLC, 1988. — 305 p.

[Forman, 1998] Forman E. Aggregating individual judgments and priorities with the analytic hierarchy process/ Forman, E. & Peniwati, K // *European Journal of Operational Research*, Vol. 108, 1998 – pp.131-145.

[Harte, 2001] Harte D. *Multifractals. Theory and applications* / David Harte. – Chapman and Hall/CRC , 2001. – 264 p.

[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.

[Hartley, 1928] Hartley R.V.L. Transmission of information. *Bell System Technical J.* 1928. 7. P. 535–563.

[Holland, 1994] Holland J.H. Adaptation in natural and artificial systems. An introductory analysis with application to biology, control, and artificial intelligence / J.H.Holland; London: Bradford book edition. – 1994. – 211 p.

[Hwang, 1981] Hwang C.L., Yoon K. Multiple attribute decision making: methods and applications: a state-of-the-art survey. Berlin, New York: Springer-Verlag, 1981. 259 p.

[Jaffard, 2004] Jaffard, S. Wavelet Techniques in Multifractal Analysis / Jaffard, S // *Fractal Geometry and*

Applications: a Jubilee of Benoit Mandelbrot – Multifractals, Probability and Statistical Mechanics, Applications. – San Diego: 2004. – Vol. 72. – Part 2. – P. 91-151.

[Kadenko, 2008] Kadenko S.V. Determination of Parameters of Criteria of "Tree" Type Hierarchy on the Basis of Ordinal Estimates / S.V.Kadenko // Journal of Information and Automation Sciences – 2008 – vol. 40, i.8 – pp. 7-15.

[Kadenko, 2010] Kadenko S.V. Personnel-related decision making using ordinal expert estimates / S.V.Kadenko// OR52 Keynotes and Extended Abstracts - 52nd Conference of the Operational Research Society – 2010 – pp. 178-184.

[Kadenko, 2012] Kadenko S.V. A Method for Improving the Consistency of Individual Expert Rankings during Their Aggregation / S.V.Kadenko, V.V.Tsyganok // Journal of Automation and Information Sciences. – 2012. – v.44, issue 4. – p.23-31.

[Kadenko, 2013] Kadenko S.V. Defining the Relative Weights of the Alternatives Estimation Criteria Based on Clear and Fuzzy Rankings / S.V. Kadenko // Journal of Information and Automation Sciences – 2013 – vol. 45, i. 2 – pp. 41-49.

[Kadenko, 2016] Kadenko S.V. Prospects and Potential of Expert Decision-making Support Techniques Implementation in Information Security Area / S.V.Kadenko //CEUR WorkshopProceedings – 2016 – pp. 8-14.

[Kantelhardt, 2002] Kantelhardt, Jan W., Zschiegner Stephan A., Koscielny-Bunde Eva et al. Multifractal detrended fluctuation analysis of nonstationary time series /– Physica A: Statistical Mechanics and its Applications. V.316, 2002. – P. 87-114.

[Kantelhardt, 2009] Kantelhardt, Jan W. Fractal and multifractal time series / Kantelhardt, Jan W. // Encyclopedia of Complexity and Systems Science. – Springer, 2009. – P. 3754–3779.

[Kemeny, 1959] Kemeny, J. G. Mathematics without Numbers / John G. Kemeny // Daedalus Vol. 88, No. 4, Quantity and Quality (Fall, 1959), pp. 577-591.

[Kleinberg, 2006] Kleinberg J. Temporal dynamics of on-line information streams // *Data Stream Management: Processing High-Speed Data Streams*. – Springer, 2006.

[Kruger, 1999] Kruger J., Dunning D. Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*. 1999. 77(6). P. 1121–1134.

[Lande, 2007] Dimitri Lande, Sergei Braichevski und Dimitri Busch. Informationsfluesse im Internet // *IWP - Information Wissenschaft & Praxis*, 59(2007), Heft 5, Seiten 277-284.

[Lande, 2009] Lande D.V., Snarskii A.A. Diagram of measurement series elements deviation from local linear approximations // *arXiv preprint arXiv:0903.3328*, 2009.

[Lewandowsky, 2012] Lewandowsky S., Ecker U. K. H., Seifert C. M., Schwarz N., Cook J. Misinformation and Its Correction Continued Influence and Successful Debiasing, *Psychological Science in the Public Interest*, Department of Human Development, Cornell University, USA, 2012, Volum e 13 issue 3, pp. 106-131.

[Li, 2012] Li R., Lei K.H., Khadiwala R., Chang K.C. TEDAS: A Twitter-based Event Detection and Analysis System // *Data Engineering (ICDE)*, 2012 IEEE 28th International Conference, 2012. — P. 1273–1276.

[Likert, 1932] Likert R. A Technique for the Measurement of Attitudes / R.Likert// *Archives of Psychology* – 1932, # 140. – pp. 1-55.

[Lipovetsky, 2006] Lipovetsky S. Optimal Hierarchy Structures for Multi-Attribute-Criteria Decisions / S. Lipovetsky // *Journal of Systems Science and Complexity*. – 2006 – 22, pp. 228-242.

[Lipovetsky, 2014] Lipovetsky S. Parameters of Optimum Hierarchy Structure in AHP / S.Lipovetsky // *Proceedings of the International Symposium for the Analytic Hierarchy Process ISAHP 2014*. [Электронный ресурс]. Режим доступа: <http://www.isahp.org/proceedings/symposium/?year=2014&page=papers>.

[Mallat, 1992] Mallat, S. Singularity Detection and Processing with Wavelets / Stephane Mallat, Wen Liang Hwang // IEEE Transactions on Information Theory. – 1992. – Vol. 38. – No. 2. – P. 617–643.

[Mallat, 2009] Mallat, S. A Wavelet Tour of Signal Processing The Sparse Way / Stephane Mallat. – Academic Press, 2009. – 805 p.

[Mandelbrot, 1982] Mandelbrot, B. B. The fractal geometry of nature / Benoit B. Mandelbrot. – W. H. Freeman and Company, 1982. – 468 p.

[Memorandum, 2010] Memorandum for all soldiers and Civilians associated with the Psychological Operations Regiment. 23 June 2010. Major General, USA Commanding Thomas R. Csrnko <http://www.heise.de/forum/Telepolis/Kommentare/MISO-statt-PSYOP/UNITED-STATES-ARMY-MEMORANDUM-FOR-All-Soldiers-and-Civilians/posting-1327431/show/>

[Military, 2011] Military Information Support Operations. Joint Publication 3-13.2. 07 January 2010 Incorporating Change 1 20 December 2011. P. 10. <http://fas.org/irp/doddir/dod/jp3-13-2.pdf>

[Miller, 1956] Miller G. A. The Magical Number Seven, Plus or Minus Two / G. A. Miller // The Psychological Review, - 1956, - vol. 63 – pp. 81—97.

[Montgomery, 2008] Montgomery, Douglas C. Introduction to time series analysis and forecasting / Douglas C. Montgomery, Cheryl L. Jennings, Murat. – New Jersey: John Wiley & Sons, 2008. – 441 p. – (Wiley series in probability and statistics).

[Noelle, 1973] Noelle-Neumann, E. (1973). Return to the concept of powerful mass media. Studies of Broadcasting. – pp. 67–112.

[Noelle, 1984] Ball-Rokeach, S.J., Rokeach, M. & Grube, J.W. (1984, November). The great American values test. Psychology Today, 34–41.

[Olenko, 2016] Olenko A. Double Entropy Inter-Rater Agreement Indices / Olenko Andriy & Tsyganok Vitaliy // Applied Psychological Measurement 40(1), 2016 – pp. 37–55.

[Oswiecimka, 2012] Effect of detrending on multifractal characteristics / Oswiecimka P., Drozd S., Kwapien J., Gorski A.Z. – eprint arXiv:1212.0354, 2012.

[Peng, 1994] Mosaic organosation of DNA nucleotides / [Peng C.-K., Buldyrev S. V., Havlin S. et al.] – Physical Review E, V. 49 E. 2, 1994.

[Rakesh, 2014] Rakesh V., Singh D., Vinzamuri B., Reddy C.K. Personalized Recommendation of Twitter Lists Using Content and Network Information // Association for the Advancement of Artificial Intelligence (Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, 2014.

[Ramanathan, 1994] Ramanathan R., Ganesh L.S. Group Preference Aggregation Methods Employed in AHP: An Evaluation and Intrinsic Process for Deriving Members' Weightages. European Journal of Operational Research. 1994. 79. P. 249–265.

[Renda, 2003] Renda M.E. Web metasearch: rank vs. score based rank aggregation methods / M.E. Renda, U. Straccia // Proceedings of the 2003 ACM symposium on Applied computing – 2003 – Pages 841-846.

[Rodriguez, 2014] Rodriguez N., Bravo G., Rodriguez N., Barba L. Haar Wavelet Neural Network for Multi-step-ahead Anchovy Catches Forecasting // Polibits. - 2014. - Issue 50. - P. 49-53.

[Roadmap, 2003] Information operations roadmap. – DoD US. – Washington, D.C.: GPO, 2003.

[Rothenberg, 2002] Rothenberg R. From whole cloth: Making up the terrorist network // Connections, vol. 24, no. 3, pp. 36-42, 2002.

[Rzążewski, 2014] Rzążewski, K. Każdy głos się liczy! Wędrownica przez krainę wyborów / K.Rzążewski, W.Słomczyński, K.Życzkowski–Wydawnictwo Sejmowe Warszawa 2014 – 420 p.

[Saaty, 1994] Saaty T.L. 1994. Fundamentals of Decision Making and Priority Theory with The Analytic Hierarchy Process. / T.L.Saaty; RWS Publications, Pittsburgh PA, 204220.

[Saaty, 1996] Saaty T. L. Decision Making with Dependence and Feedback: The analytic Network Process / T. L. Saaty – Pittsburgh, RWS Publicaitons, 1996 – 370 p.

[Saaty, 2008] Saaty T.L. Relative measurement and its generalization in decision making. Why pairwise comparisons are central in mathematics for the measurement of intangible factors. The Analytic Hierarchy/Network Process / T.L.Saaty // Statistics and Operations Research, Vol. 102(2), 2008 – pp.251-318.

[Snarskii, 2016] Snarskii A.A., Zorinets D.I., Lande D.V. "Conjectural" links in complex networks // Physica A: Statistical Mechanics and its Applications, 2016. – Vol. 462. – pp. 266-273.

[St. James Park, 2007] St. James Park, Southampton. CABE Spaceshaper Workshop Facilitators Report. [Электронний ресурс]. Режимдоступу: <http://www.westleydesign.co.uk/what-we-do/Downloads/WestleyDesign-StJamesSpaceshaper.pdf>.

[Suarez-Garcia, 2013] Suarez-Garcia P. Multifractality and long memory of a financial index / Suarez-Garcia P., Gomez-Ullate D. – Eprint arXiv:1306.0490, 2013.

[Sun, 2001] Sun, X. Multifractal analysis of Hang Seng index in Hong Kong stock market / X. Sun, H. P. Chen, Z. Q. Wu, Y. Z. Yuan // Physica A-Statistical Mechanics and its Applications.– 2001. – Vol. 291. – Issue 1–4. – P. 553–562.

[Thompson, 2016] Thompson, James. R. Multifractal Detrended Fluctuation Analysis: Practical Applications to Financial Time Series / Thompson, James. R., Wilson, James R. // Mathematics and Computers in Simulation. – 2016. – Vol. 126. – Issue C. – P. 63–88.

[Totsenko, 2005] Totsenko V.G. Method of Determination of Group Multicriteria Ordinal Estimates with Account of Expert Competence / V.G. Totsenko // Journal of Automation and Information Sciences. – 2005. – vol. 37. -N10. - P.19-23.

[Tsyganok, 2010a] Tsyganok V.V. Investigation of the aggregation effectiveness of expert estimates obtained by the pair-wise comparison method / V.V. Tsyganok // Mathematical and Computer Modelling. – 2010. – Vol.52(3-4). – pp. 538-544.

[Tsyganok, 2010b] Tsyganok V.V. On Sufficiency of the Consistency Level of Group Ordinal Estimates / V.V.Tsyganok, S.V.Kadenko // Journal of Automation and Information Sciences. – 2010. – v.42, issue 8.– p.42-47

[Tsyganok, 2012] Tsyganok, V. Significance of expert competence consideration in group decision making using AHP. / V. Tsyganok, S. Kadenko, O. Andriichuk//International Journal of Production Research. Vol. 50, Issue 17, 2012 – pp. 4785-4792.

[Tsyganok, 2015] Tsyganok V.V. Using Different Pair-wise Comparison Scales for Developing Industrial Strategies / V.V.Tsyganok, S.V.Kadenko, O.V.Andriichuk // Int. J. Management and Decision Making. – 2015. – Vol. 14, No.3. – pp.224-250.

[Wanta, 1994] Wanta, W. & Foote, J. The president-news media relationship: A time series analysis of agenda-setting. Journal of Broadcasting & Electronic Media. 1994. 38, 437–449.

[Wasserman, 1994] Wasserman S. and Faust K. Social Network Analysis: Methods and Applications. New York: Cambridge University Press, 1994. – 825 p.

[Yang, 2017] Yang Q, Du P-a, Wang Y, Liang B. A rough set approach for determining weights of decision makers in group decision making. PLoS ONE 12(2). 2017: e0172679. doi:10.1371.

[Zgurovsky, 2004] Zgurovsky, M.Z. Group incomplete paired comparisons with account of expert competence / Zgurovsky, M.Z., Totsenko, V.G., & Tsyganok,V.V.// Mathematical and Computer Modelling, 39(4), 2004 – pp.349–361.

[Zhou, 2009] Zhou, W.-X. The components of empirical multifractality in financial returns / Wei-Xing Zhou // EPL. – 2009. – Vol. 88. – Issue 2. – Article Number 28004.

[Аборок, 2005] Аборок, Ш.К. Метапоисковая система с использованием показателя обобщенной релевантности документа для формирования результатов поиска / Ш.К. Аборок, В.И. Костин // Информатика и компьютерные технологии-2005: сб. тр. 1-ой Междунар. студ. науч.-техн. конф., 15 дек. 2005 г. - Донецк, 2006. - С. 268-269.

[Аверченков, 2011] Аверченков В.И., Ерохин В.В. Системы организационного управления: учеб. пособ. 3-е изд., стереотипное. — Москва: Изд-во «ФЛИНТА» 2011. 208 с.

[Андрійчук, 2014] Андрійчук О.В. Метод змістової ідентифікації об'єктів баз знань систем підтримки прийняття рішень / О.В. Андрійчук // Реєстрація, зберігання і обробка даних. — 2014. — Т. 16, № 1. — С. 65-78.

[Андрійчук, 2015] Андрійчук О.В. Метод визначення змістової подібності об'єктів баз знань експертних систем підтримки прийняття рішень: дис. ... канд. техн. наук: 01.05.04. – Київ, 2015. – 162 с.

[Андрійчук, 2016] Андрейчук О.В., Качанов П.Т. Методика применения инструментария экспертной поддержки принятия решений при идентификации информационных операций / Информационные технологии и безопасность: Материалы международной научно-технической конференции 1 декабря 2016 года/ – К: ИПРИ НАН Украины, 2016. – С. 141-155.

[Волошин, 1993] Волошин О.Ф., Гнатієнко Г.М. Процедури визначення компетентності експертів. Вісник Київ. ун-ту. Фіз.-мат. науки. 1993. № 3. С. 102–111.

[Гнатієнко, 2008] Гнатієнко Г.М., Снитюк В.С. Експертні технології прийняття рішень. Київ: ТОВ «Маклаут», 2008. 444 с.

[Горбов, 2013] Горбов І.В. Визначення потенційних експертних груп науковців у мережі співавторства з використанням методів підтримки прийняття рішень / І.В. Горбов, С.В. Каденко, І.В. Балагура, Д.Ю. Манько, О.В. Андрійчук // Реєстрація, зберігання і обробка даних. - 2013. - 15, № 4. - С. 86-96.

[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 – К.: Старт-98, 2007. – 40 с.

[Губанов, 2009] Губанов Д. А., Новиков Д. А., Чхартишвили А. Г. Модели репутации и информационного управления в социальных сетях // Управление большими системами. Выпуск 26.1. – М.: ИПУ РАН, 2009. – С. 209-234.

[Додонов, 2009] Додонов О.Г. Інформаційні потоки в глобальних комп'ютерних мережах / Додонов О.Г., Ланде Д.В., Путятін В.Г. – К: Наукова думка, 2009. – 295 с.

[Додонов, 2013] Додонов А.Г., Ландэ Д.В. Мультиагентная модель поведения тематических информационных потоков // Материалы VI Всероссийской мультikonференции по проблемам управления (30 сентября – 5 октября 2013 г.) – Т. 4. – Ростов-на-Дону: Издательство Южного федерального университета, 2013. – С. 102-107.

[Додонов, 2015a] Додонов В.А., Ландэ Д.В. Мультиагентные модели информационного влияния // Информационные технологии и безопасность. Материалы XV Международной научно-практической конференции ИТБ-2015. – К.: ИПРИ НАН Украины, 2015. – С. 75-83.

[Додонов, 2015b] Додонов А.Г., Ландэ Д.В., Бойченко А.В. Сценарный подход при исследовании динамики информационных потоков в сети Интернет // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2015): материалы V междунар. науч.-техн. конф. (Минск 19-21 февраля 2015 года) - Минск: БГУИР, 2015. - С. 225-230.

[Додонов, 2016] Додонов А.Г., Ландэ Д.В., Коваленко Т.В. Модели предметных областей в системах поддержки принятия решений на основе мониторинга

информационного пространства // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2016): материалы VI междунар. науч.-техн. конф. (Минск 18-20 февраля 2016 года) - Минск: БГУИР, 2016. - С. 171-176.

[Ельмеев, 1999] Ельмеев В.Я., Овсянников В.Г. Прикладная социология. Очерки методологии. 2-е изд., испр. и доп. Санкт-Петербург: Изд-во. СПГУ, 1999. 296 с.

[Закон України, 2014] Закон України Про Державний бюджет України на 2014 рік [Електрон. ресурс]. Режим доступу: <http://zakon4.rada.gov.ua/laws/show/719-18>.

[Закон України, 2015] Закон України Про Державний бюджет України на 2015 рік [Електрон. ресурс]. Режим доступу: <http://zakon4.rada.gov.ua/laws/show/80-19>.

[Закон України, 2016] Закон України Про Державний бюджет України на 2016 рік [Електрон. ресурс]. Режим доступу: <http://zakon4.rada.gov.ua/laws/show/928-19>.

[Каденко, 2008] Каденко С.В. Определение параметров иерархии критериев типа "дерево" на основе ординальных оценок / С. В. Каденко // Проблемы управления и информатики. - 2008 . - № 4. - С. 84-91.

[Каденко, 2016] Каденко С.В. Проблеми представлення експертних даних у системах підтримки прийняття рішень/ С.В. Каденко // Реєстрація, зберігання і обробка даних, Т. 18 № 3, 2016 – с. 67-74.

[Каденко, 2016] Каденко С.В. Можливості та перспективи використання експертних технологій підтримки прийняття рішень у сфері інформаційної безпеки / Информационные технологии и безопасность: Материалы международной научно-технической конференции 1 декабря 2016 года/ – К: ИПРИ НАН Украины, 2016. – С. 31-42.

[Ландэ, 2006] Ландэ Д.В., Фурашев В.Н., Брайчевский С.М., Григорьев А.Н. Основы моделирования и оценки электронных информационных потоков: Монография. – К.: Инжиниринг, 2006. – 176 с.

[Ландэ, 2007] Ландэ Д.В. Модель диффузии информации // Информационные технологии и безопасность. Менеджмент информационной безопасности. Сборник научных трудов Института проблем регистрации информации. – Вып. 10. – 2007. – С. 51-67.

[Ландэ, 2009] Ландэ Д.В. Как организовать оборону: 12 шагов противодействия // Телеком, 2009. – № 6. – С. 46-51.

[Ландэ, 2012] Ландэ Д.В., Брайчевский С.М. Моделирование поведения тематических сюжетов новостей в веб-пространстве // Тринадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2012 (16-20 октября 2012 г., г. Белгород, Россия): Труды конференции. – Т. 1. – Белгород: Изд-во БГТУ, 2012. – С. 197-204.

[Ландэ, 2016а] Ландэ Д.В., Грайворонская А.Н., Березин Б.А. Мультиагентная модель распространения информации в социальной сети // Реєстрація, зберігання і обробка даних, 2016. – Т. 18, № 1. – С. 70-77.

[Ландэ, 2016б] Ландэ Д.В., Додонов В.А., Коваленко Т.В. Информационные операции в компьютерных сетях: моделирование, выявление, анализ // Моделирование-2016: материалы пятой Международной конференции Моделирование-2016, Киев, 25-27 мая 2016 г. / ИПМЭ НАН Украины, 2016. – С. 198-201.

[Лозовский, 2000] Лозовский Б.Н. Методика сбора информации // Основы творческой деятельности журналиста. СПб : 2000. – С. 79-95.

[Лозовский, 2003] Лозовский Б.Н. Средства массовой информации как объект манипуляции // Известия Уральского государственного университета. 2003. № 27. – С. 74-85.

[Лозовский, 2006] Лозовский Б.Н. Журналистика как специфическая репрезентация действительности // Известия Уральского государственного университета. 2006. № 40. – С. 27-38.

[Лозовский, 2008] Лозовский Б.Н. Аудитория и СМИ: взаимодействие и манипулирование // Известия Уральского государственного университета. Серия 1.

Проблемы образования, науки и культуры. 2008. № 56. – С. 277–291.

[Манойло, 2007] Манойло А.В., Государственная информационная политика в особых условиях: Монография. – М.: МИФИ, 2003. – 388 с.

[Потемкин, 2015] Потемкин А.В. Выявление информационных операций в средствах массовой информации сети Интернет: диссертация ... кандидата технических наук: 05.13.10 / Потемкин Алексей Владимирович; [Место защиты: Брянский государственный технический университет]. – Брянск, 2015. – 144 с.

[Почепцов, 2001] Почепцов Г.Г. Теория коммуникации – М.: «Рефл-бук», К.: «Ваклер» – 2001. – 656 с.

[Расторгуев, 2014] Расторгуев С.П., Литвиненко М.В. Р24 Информационные операции в сети Интернет / Под общ. ред. А.Б. Михайловского. — М.: АНО ЦСОиП, 2014. — 128 с.

[Снарский, 2007] Снарский А.А., Безсуднов И.В., Севрюков В.А. Процессы переноса в макроскопических неупорядоченных средах: От теории среднего поля до перколяции. – М.: УРСС, Изд-во АКИ, 2007. – 304 с.

[Таран, 2006] Таран Т.А. Искусственный интеллект. Теория и приложения / Т.А. Таран, Д.А. Зубов; Восточноукр. нац. ун-т им. Владимира Даля. — Луганск: ВНУ им. В. Даля, 2006. — 239 с.

[Тоценко, 2001] Тоценко В.Г. Об одном подходе к поддержке принятия решений при планировании исследований и развития. Часть 2. Метод целевого динамического оценивания альтернатив / В.Г.Тоценко // Проблемы управления и информатики. – 2001. – №2. – С. 127-139.

[Тоценко, 2002] Тоценко В.Г. Методы и системы поддержки принятия решений. Алгоритмический аспект [Текст] / В. Г. Тоценко; ИПРИ НАНУ. – К.: Наукова думка, 2002. – 382 с.

[Тоценко, 2003] Свідоцтво про держ. реєстрацію автор. права на твір №8669. МОН України Держ. деп. інтелект. власності. Комп'ютерна програма "Система підтримки

прийняття рішень СОЛОН-3" (СППР СОЛОН-3) / В.Г.Тоценко, П.Т.Качанов, В.В.Циганок // зареєстровано 31.10.2003.

[Урсу, 2011] Урсу Н.С. Манипулирование в политике: новые каналы коммуникации в Интернете. / Н.С. Урсу // Вестник Российской нации. – 2011. – №3. – С. 148-160.

[Урсу, 2012] Урсу Н.С. Технологии манипулирования в сети Интернет. / Н.С. Урсу // Социально-гуманитарные знания. – 2012. – №1. – С. 303-310.

[Циганок, 2000] Циганок В.В. Комбінаторний алгоритм парних порівнянь зі зворотним зв'язком з експертом / В.В.Циганок // Реєстрація, зберігання і обробка даних. – 2000. – Т.2, №2. – С.92-102.

[Циганок, 2008] Циганок В.В. Метод обчислення ваг альтернатив на основі результатів парних порівнянь, проведених групою експертів. Реєстрація, зберігання і обробка даних. 2008. Т. 10. № 2. С. 121–127.

[Циганок, 2010] Циганок В.В. Проблема розподілу ресурсів, як розширення можливостей систем підтримки прийняття рішень Реєстрація, зберігання і обробка даних. 2010, т.12, №2 – С.232-237.

[Циганок, 2011] Циганок В.В. Агрегація групових експертних оцінок, що отримані у різних шкалах / В.В.Циганок // Реєстрація, зберігання і обробка даних. – 2011. – т.13. – №4.– С.74-83.

[Циганок, 2012a] Циганок В.В. Експериментальний аналіз технології експертного оцінювання / В.В.Циганок, П.Т.Качанов, С.В.Каденко, О.В.Андрійчук, Г.А.Гоменюк // Реєстрація, зберігання і обробка даних. – 2012. – Т. 14, № 1. – С. 91-100.

[Циганок, 2012b] Свідоцтво про реєстрацію авторського права на твір № 44521 Державної служби інтелектуальної власності України. Комп'ютерна програма "Комплекс програмних засобів для експертного оцінювання шляхом парних порівнянь «Рівень»" / В.В. Цыганок, О.В. Андрійчук, П.Т. Качанов, С.В. Каденко // від 03.07.2012.

[Циганок, 2013] Циганок В.В. Удосконалення методу цільового динамічного оцінювання альтернатив та особливості його застосування / В. В. Циганок // Реєстрація, зберігання і обробка даних. - 2013. - Т. 15, № 1. - С. 90-99.

[Циганок, 2015] Циганок В.В. Інструментарій підтримки прийняття рішень як засіб стратегічного планування / В.В.Циганок, С.В.Каденко, О.В.Андрійчук, П.Т.Качанов, П.Д.Роїк // Озброєння та військова техніка – 2015, № 3(7). – С. 59-66.

[Циганок, 2017] Свідоцтво про реєстрацію авторського права на твір №75023 Міністерство економічного розвитку і торгівлі України. Комп'ютерна програма „Система розподіленого збору та обробки експертної інформації для систем підтримки прийняття рішень – «Консенсус-2»” / В.В.Циганок, П.Т.Качанов, О.В.Андрійчук, С.В.Каденко // від 03/10/2012.

[Чепков, 2014] Чепков І.Б., Ланецкий Б.М., Леонтьєв О.Б., Лук'янчук В.В. Методичний підхід до обґрунтування раціонального співвідношення обсягів розробки, закупівлі та ремонту озброєння й військової техніки. Озброєння та військова техніка. 2014. № 3. С. 9–14.

[Шелухин, 2011] Шелухин О. И. Мультифракталы. Инфокоммуникационные приложения / О. И. Шелухин. – М.: Горячая линия – Телеком, 2011. – 576 с.

[Əliquliyev , 2010] Əliquliyev R.M., İmamverdiyev Y.N., Abdullayeva F.C. Sosial şəbəkələr // Bakı: «İnformasiya Texnologiyaları» nəşriyyatı, 2010. – 287 s.

Научное издание

А.Г. Додонов
Д.В. Ландэ
В.В. Цыганок
О.В. Андрейчук
С.В. Каденко
А.Н. Грайворонская

**РАСПОЗНАВАНИЕ
ИНФОРМАЦИОННЫХ ОПЕРАЦИЙ**

Монография

Підписано до друку 05.12.2017

Формат 60 x 84/8. Офсетний друк.

Умов. др. арк. 12,5. Наклад 200 прим.

Віддруковано з оригінал-макета у видавництві ТОВ “Інжиніринг”
м. Київ, вул. Федорова, 9

ISBN 978-966-2344-60-8