

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ
ИНСТИТУТ ПРОБЛЕМ РЕГИСТРАЦИИ ИНФОРМАЦИИ
НАН УКРАИНЫ
НАЦИОНАЛЬНАЯ АКАДЕМИЯ ПРАВОВЫХ НАУК
УКРАИНЫ
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ
ИНФОРМАТИКИ И ПРАВА
НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
УКРАИНЫ «КПИ»
УЧЕБНО-НАУЧНЫЙ ЦЕНТР ИНФОРМАЦИОННОГО
ПРАВА И ПРАВОВЫХ ВОПРОСОВ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ ФСП

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
И БЕЗОПАСНОСТЬ:
ОСНОВЫ ОБЕСПЕЧЕНИЯ
ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ**

ВЫПУСК 14

КИЕВ – 2014

*Рекомендовано к печати ученым советом
Института проблем регистрации информации НАН Украины
(протокол № 4 от 24 июня 2014 г.)*

Информационные технологии и безопасность: основы обеспечения информационной безопасности. Материалы международной научной конференции ИТБ-2014. – К.: ИПРИ НАН Украины, 2014. – 180 с. ISBN: 978-966-2344-34-9

В сборник вошли материалы, представленные на Международной конференции «Информационные технологии и безопасность: основы обеспечения информационной безопасности» (ИТБ-2014), которая проходила 28 мая 2014 года в г.Киев, Украина.

Сборник охватывает широкий круг актуальных проблем обеспечения информационной безопасности при использовании современных сетевых информационных технологий, правового обеспечения информационной безопасности, методического обеспечения систем организационного управления.

Редакционная коллегия:

А.Г. Додонов, д.т.н., профессор; В.Г. Пилипчук, д.ю.н., профессор, член-корр. НАПрН Украины; А.М. Богданов, д.т.н., профессор; Д.В. Ландэ, д.т.н., с.н.с.; В.В. Мохор, д.т.н., профессор; Н.А. Ожеван, д.ф.н., профессор; В.Н. Фурашев, к.т.н., с.н.с.; Е.С. Горбачик, к.т.н., с.н.с.; М.Г. Кузнецова, к.т.н., с.н.с.

ISBN 978-966-2344-34-9

- © Институт проблем регистрации информации НАН Украины, 2014
- © Научно-исследовательский институт информатики и права НАПрН Украины, 2014
- © Учебно-научный центр информационного права и правовых вопросов информационных технологий ФСП НТУУ «КПИ», 2014
- © Коллектив авторов

МЕТОДИКА АНАЛИТИЧЕСКОГО ИССЛЕДОВАНИЯ ДИНАМИКИ СОБЫТИЙ НА ОСНОВЕ МОНИТОРИНГА ВЕБ-РЕСУРСОВ СЕТИ ИНТЕРНЕТ

Додонов А.Г., Ландэ Д.В.

*Институт проблем регистрации информации
НАН Украины*

Введение, постановка задачи

В настоящее время контент сети Интернет образует значимый динамический сегмент информационного пространства, информационные потоки, содержание и объемы которых необходимо учитывать при проведении аналитических исследований практически в любой области. Основным объектом анализа при этом являются событийные или тематические срезы этих потоков – массивы информационных сообщений, документов, соответствующих определенным событиям или тематикам. Динамика информационных потоков определяется комплексом как внутренних, так и внешних нелинейных механизмов, которые отражаются, возможно, в неявном виде. Зачастую удовлетворительным оказывается упрощенное понимание информационного сюжета как некоторой зависимой от времени величины $n(t)$, поведение которой описывается нелинейными уравнениями [1, 2]. Таким информационным потокам можно ставить в соответствие временные ряды, для анализа которых все чаще обоснованно применяются формальные методы: статистического, фрактального, Фурье или вейвлет-анализа.

Для эффективного проведения информационно-аналитических исследований на основе анализа контента сети Интернет (а точнее ее веб-ресурсов) авторами предлагается последовательность шагов, этапов обработки информации, каждый из которых сам

по себе обеспечивает получение продукта. Совокупность таких этапов, базирующихся на использовании необходимых и доступных инструментальных средств, специальных приемов, можно рассматривать как методику, процедуру проведения действий, нацеленных на получение аналитических материалов, которые могут использоваться для поддержки принятия решений.

Любая методика рассчитана на решение конкретных задач. При проведении информационно-аналитических исследований на базе интернет-контента к таким задачам можно отнести:

1. Нахождение релевантных публикаций по тематике.
2. Определение динамики тематических публикаций.
3. Определение критических точек в динамике тематических публикаций.
4. Выявление объектов мониторинга.
5. Выявление и визуализация взаимосвязей событий и объектов мониторинга, а также объектов мониторинга между собой.
6. Прогнозирование развития событий.

Этапы информационно-аналитического исследования

В соответствии с приведенными выше задачами предлагается выделить следующие этапы информационно-аналитического исследования:

1. Выбор системы интеграции интернет-документов.
2. Формирование запроса в среде выбранной системы. Нахождение тематических публикаций по запросу с помощью систем контент-мониторинга.
3. Определение динамики тематических публикаций по запросу.
4. Определение критических точек в динамике тематических публикаций.
5. Определение основных событий в критических точках.

6. Выявление объектов мониторинга.
 7. Выявление и визуализация взаимосвязей.
 8. Прогноз развития событий.
- Рассмотрим эти этапы подробно.

Выбор системы интеграции интернет-документов

Для получения репрезентативной информации об объекте исследования необходимо воспользоваться поисковой системой с аналитическими функциями, охватывающей достаточный объем информации, относящейся к исследуемому объекту/событию. Для анализа динамики информационных потоков необходимо каким-то образом получить соответствующую статистику, представленную в виде временных рядов. Многие современные информационно-аналитические системы содержат в своем составе средства отображения статистики вхождения в базы данных понятий, соответствующих пользовательским запросам. В настоящее время существует несколько открытых информационных сервисов, в рамках которых можно наблюдать временную динамику объемов публикаций по тематикам, определяемым запросами. Так Google books Ngram Viewer (<http://ngrams.googlelabs.com/>), предоставляет визуализацию динамики количества книг, в которых упоминаются слова. Сервис «Яндекс пульс блогосферы» (<http://blogs.yandex.ru/pulse/>) также позволял отображать динамику публикаций в блогах, содержащих заданные пользователем ключевые слова, однако был закрыт ввиду малой посещаемости. Сегодня этот сервис доступен лишь по специальному разрешению компании «Яндекс». На сайте Национального корпуса русского языка (НКРЯ) в бета-режиме запущен сервис N-грамм (<http://www.ruscorpora.ru/ngram.html>), близкий по функциональности сервису Google books Ngram Viewer.

Безусловно, самыми эффективными для решения задач анализа динамического контента являются специализированные системы интеграции сетевого контента. В частности, в рамках исследований

авторами использовалась система контент-мониторинга веб-ресурсов InfoStream (www.infostream.ua), реализующая необходимую функциональность и охватывающая около 100 тыс. документов в сутки с 7000 веб-сайтов.

Формирование запроса в среде выбранной системы

Массив тематических документов (тема – события, связанные с Евромайданом в Киеве 2013-2014 гг.) выбирается с помощью системы InfoStream путем ввода запроса на языке данной системы:

(майдан | евромайдан)&(избиен | разгон | штурм | беркут | молотов | титушк | погиб)&lang.RUS,

по которому в период с ноября 2013 года по март 2014 года было опубликовано свыше 200 тысяч **тематических публикаций** на веб-сайтах (рис. 1).

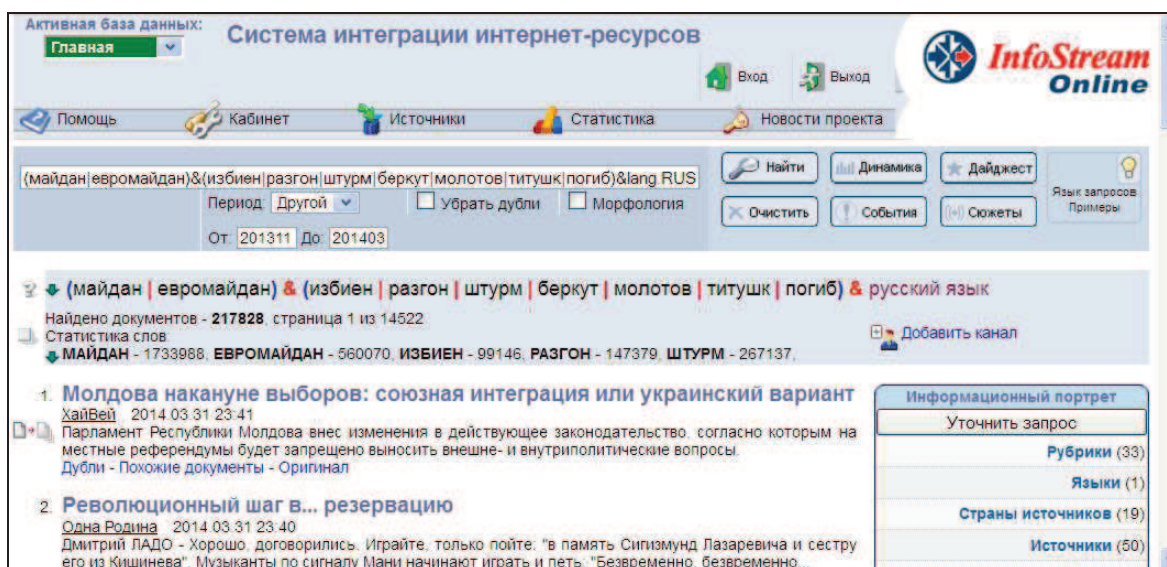


Рис. 1 – Поисковый интерфейс системы InfoStream

Система InfoStream обеспечивает поиск, а также просмотр списка и полных текстов релевантных документов.

Определение динамики тематических публикаций по запросу

Режим динамики событий системы интеграции интернет-ресурсов позволяет получить данные о количестве публикаций по заданному запросу за указанный промежуток времени. Эти данные могут быть загружены в настольную систему обработки данных и отображаются в виде графика (рис. 2). В интерфейсе пользователя обеспечивается переход к просмотру релевантных документов по выбранной дате.

После этого данные временной динамики за каждые сутки нормируются, т.е. в среде системы Excel формируется временной ряд, содержащий относительные значения, равные отношению количества тематических сообщений к общему потоку сообщений за сутки.



Рис. 2 – Режим «Динамика событий» системы интеграции

Это, в частности, позволяет избавиться от недельной периодичности в количестве тематических публикаций. Затем происходит переход к процедуре определения критических точек в данном временном ряде.

Определение критических точек в динамике сообщений

Критические точки как локальные максимумы временного ряда динамики публикаций можно определить, например, визуально по графику, представленному на рис. 2. Вместе с тем, существуют несколько научно-обоснованных методик, базирующихся на методах цифровой обработки сигналов.

Кривая цикла информационных операций

В результате анализа многочисленных диаграмм поведения ТИП, были выявлены наиболее типичные, базовые профили их поведения [3, 4]. Предложенные модели полностью соответствуют реальным данным, которые экстрагируются системами контент-мониторинга. Поэтому приведенные зависимости могут быть использованы как шаблоны, например, для выявления информационных операций – как путем анализа ретроспективного фонда сетевых публикаций, так и для оперативного мониторинга появления некоторых их признаков в реальном времени.

В частности, для выявления информационных операций [5] 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.

На рис. 3 приведен обобщенная диаграмма, соответствующая всем этапам жизненного цикла информационных операций, обоснованная авторами в [2].

Для выявления степени «близости» фрагментов исследуемого временного ряда приведенной диаграмме в различных масштабах предлагается использовать так называемый «вейвлет-анализ» [6], который в настоящее время нашел широкое применение как в естественных науках, так и в социологии [7].

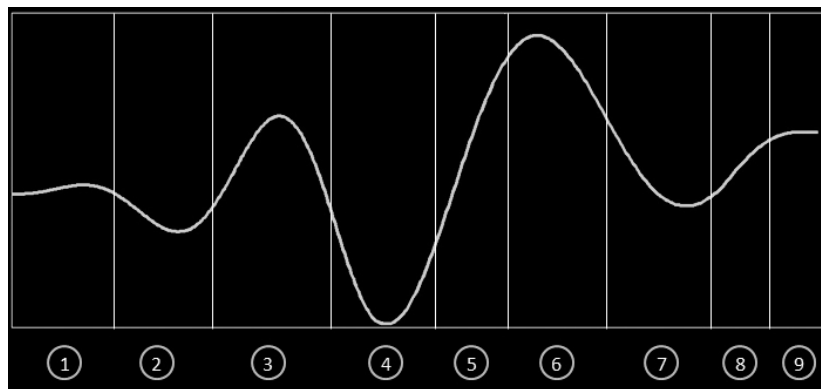


Рис. 4 – Жизненный цикл информационных операций: 1 – фон; 2 – затишье; 3 – «артподготовка»; 4 – затишье; 5 – атака/триггер роста; 6 – пик завышенных ожиданий; 7 – утрата иллюзий организаторов; 8 – общественное осознание; 9 – продуктивность/фон

Главная идея вейвлет-преобразования заключается в том, что нестационарный временной ряд разделяется на отдельные промежутки (так называемые «окна наблюдения»), и на каждом из них вычисляется величина, показывающая степень близости закономерностей исследуемых данных с разными сдвигами некоторого вейвлета (специальной функции) в разных масштабах. Вейвлет-преобразование генерирует набор коэффициентов, которые являются функциями двух переменных: времени и частоты, и потому образуют поверхность в трехмерном пространстве.

Непрерывное вейвлет-преобразование для функции $f(t)$ строится с помощью непрерывных масштабных преобразований и переносов выбранного вейвлета $\psi(t)$ с произвольными значениями масштабного коэффициента a и параметра сдвига b :

$$W(a,b) = (f(t), \psi(t)) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi^* \left(\frac{t-b}{a} \right) dt.$$

Полученные вейвлет-коэффициенты можно представить в графическом виде, если по одной оси отложить сдвиг вейвлета (ось времени), а по другой – масштабы (ось масштабов), и раскрасить точки полученной схемы в зависимости от величины

соответствующих коэффициентов (чем больше коэффициент, тем ярче цвета).

Эти коэффициенты, которые показывают, насколько поведение процесса в данной точке аналогично вейвлету в данном масштабе. Чем ближе от анализируемой зависимости в окрестности данной точки к виду вейвлета, тем большую абсолютную величину имеет соответствующий коэффициент. Применение этих операций, с учетом свойства локальности вейвлета в частотно-временной области, позволяет анализировать данные на разных масштабах и точно определять положение их характерных особенностей во времени.

На скейлограмме видны все характерные особенности исходного ряда: масштаб и интенсивность периодических изменений, направление и значение трендов, наличие, расположение и продолжительность локальных особенностей.

В работе [2] показано, что вейвлеты «мексиканская шляпа» и Морле (рис. 4) наиболее точно отражает динамику информационных операций, результаты применения этого вейвлета приведены на рис. 5, благодаря чему выбраны три даты (2013.11.30, 2014.01.22, 2014.02.19), соответствующие критическим точкам исследуемого процесса.

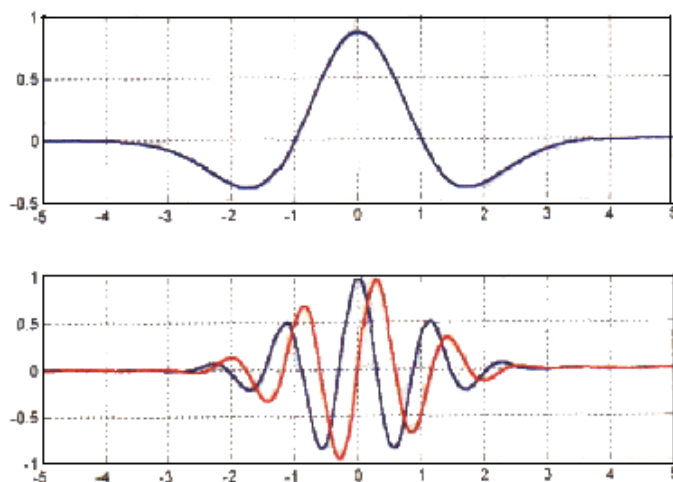


Рис. 4. Вейвлеты mexh, Морле

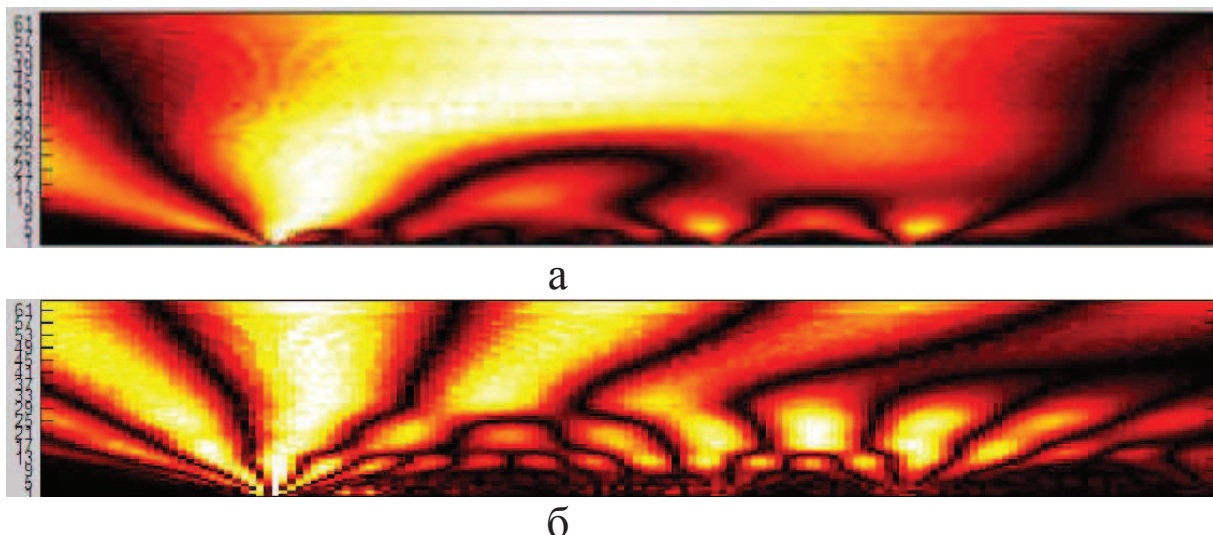


Рис. 5 – Вейвлет-спектограммы исследуемого временного ряда
(а – «мексиканская шляпа», б – вейвлет Морле)

Следует отметить, что инструменты построения вейвлет-спектограмм доступны как в ряде пакетов математических программ, например, в Matlab, так и через Интернет в режиме онлайн (<http://ion.researchsystems.com/cgi-bin/ion-p?page=wavelet.ion>).

Определение основных событий в критических точках

После определения критических точек с помощью системы контент-мониторинга выполняется построение основных сюжетных цепочек из сообщений, соответствующих запросу за выбранные даты. Таким образом определяются основные события за указанные даты (рис. б).

Для последующего анализа отбирается три массива сообщений, соответствующие трем выбранным датам, объекты из которых (в простейшем случае – персоны и веб-источники) могут рассматриваться как объекты мониторинга.

2013.11.30: Разгон демонстрантов на Майдане

<p>1. Азаров считает разгон демонстрантов на Майдане в Киеве провокацией</p> <p>Премьер-министр Украины Николай Азаров считает разгон демонстрантов на Майдане Незалежности в Киеве провокацией и обещает, что ситуация будет тщательным образом расследована. Об этом УНИАН сообщил пресс-секретарь премьер-министра Виталий Лукьяненко. Позиция премьера заключается в том, что необходимо провести в сжатые сроки тщательное и объективное расследование, и для этого создана оперативно-</p>	<p>2013.11.30 14:52 Пятеро участников Евромайдана госпитализированы из Шевченковского райотдела милиции Suds.info 336</p> <p>2013.11.30 23:53 Янукович приказал Генпрокуратуре наказать виновных в разгоне Евромайдана korabelov.info</p>
---	---

2014.01.22: Штурм на ул. Грушевского

<p>1. В центр Киева стягивают бронетехнику</p> <p>КИЕВ. 22 января. В центре Киева сосредотачиваются крупные силы бойцов внутренних войск МВД. Известно, что к стадиону "Динамо", где собрались протестующие, прибыл БТР. Значительное количество силовиков стоят рядами, прикрывшись щитами, перегородив улицу Грушевского, передает "Интерфакс-Украина". 22 января в Киеве произошли очередные столкновения радикальной оппозиции с милицией.</p>	<p>2014.01.22 13:11 "Беркут" разогнал протестующих на Грушевского: в центре Киева дрожи Глязед 479</p> <p>2014.01.22 23:58 В Киеве объявлена эвакуация Гуляй-Поле</p>
---	---

2014.02.19: Штурм правительственного квартала

<p>1. Кровавая ночь в Киеве: сможет ли Янукович удержать власть?</p> <p>Ситуацию на Украине в интервью ИА "Медиафакс" оценивают ведущие украинские эксперты. ПОЧЕМУ УКРАИНА НЕ ИЗРАИЛЬ? Минувшей ночью в столице Киева вспыхнувшая драма перешла в трагедию: в боях между силовиками и сторонниками Майдана погибли по меньшей мере 36 человек, из которых 25 - активисты оппозиции, а 11 - милиционеры.</p>	<p>2014.02.19 14:51 ПР и оппозиция готовы провести экстренное заседание парламента НОВОСТИ Bigmir.net 543</p> <p>2014.02.19 23:59 Украина на краю пропасти и в трауре. Ежедневник</p>
---	---

Рис. 6 – Основные сюжетные цепочки за выбранные даты

Выявление объектов мониторинга

С помощью методов экстрагирования фактических данных, применяющихся в системах интеграции интернет-ресурсов, в интерфейсе пользователя формируются так называемые «информационные портреты», охватывающие списки персон, топонимов, языков, компаний и т.п., содержащиеся в документах, релевантных некоторому заданному запросу.

В нашем случае из «информационного портрета», соответствующего тематическому запросу выбираются наиболее упоминаемые персоны и/или веб-ресурсы за выбранные даты (рис. 7 и 8). Эти списки могут агрегироваться, в результате чего возможно определение взаимосвязей событий и объектов (рис. 9).

В качестве системы визуализации авторами выбрана система анализа и отображения сетей Gephi (www.gephi.org).

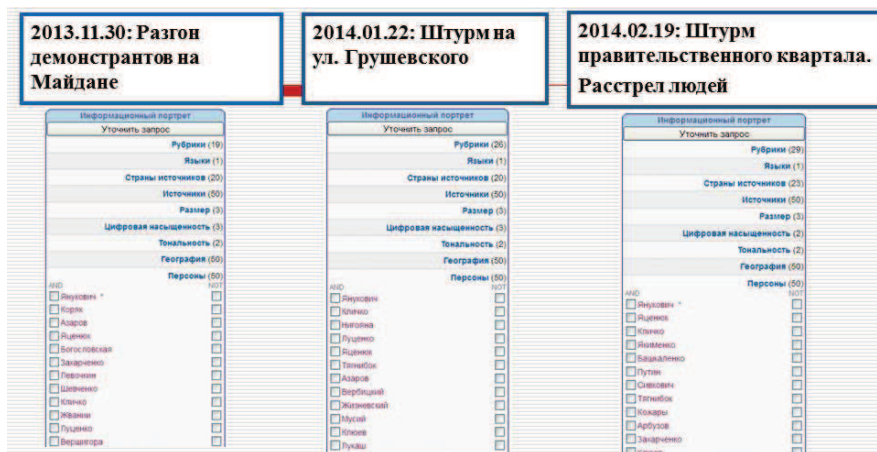


Рис. 7. Списки «доминирующих» персон

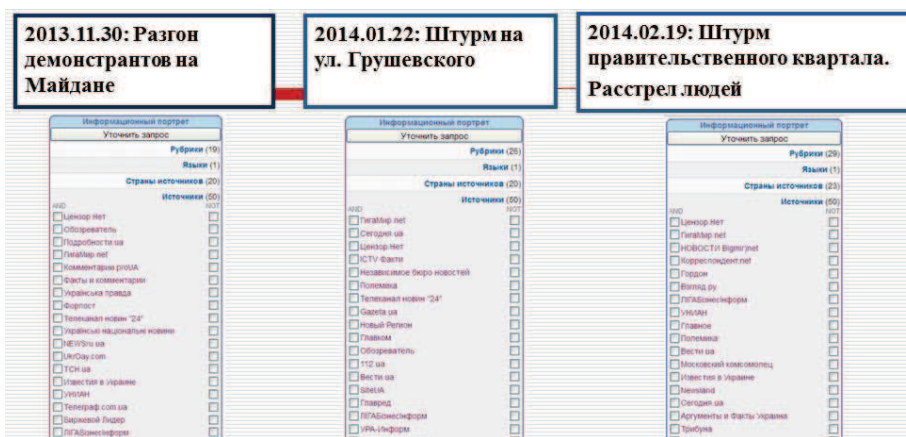


Рис. 8. Списки «доминирующих» веб-ресурсов

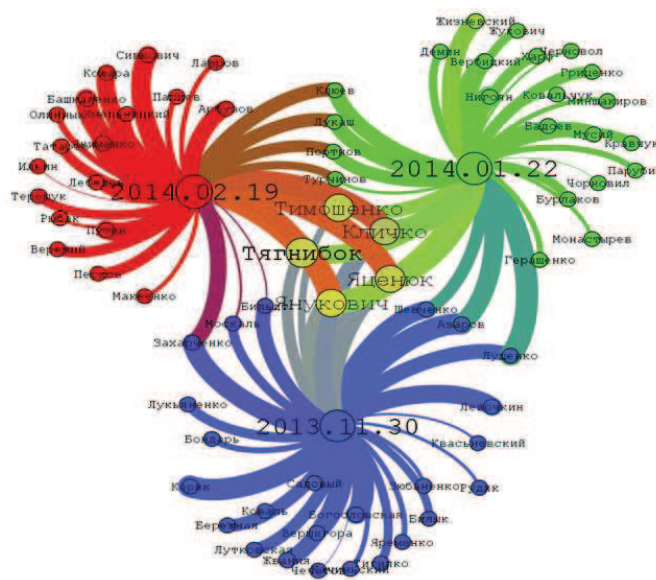


Рис. 9. Пример визуализации связей событий и объектов

Эти же данные позволяют выявлять взаимосвязи между объектами, например, между указанными аналитиками веб-ресурсами и персонами (рис. 10).

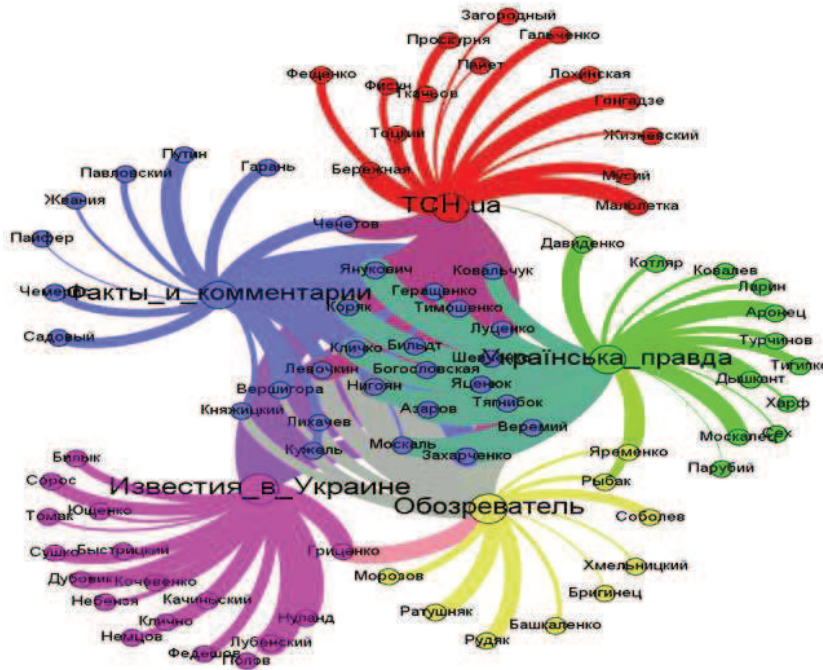


Рис. 10. Визуализация взаимосвязей веб-ресурсов и персон

На рис. 9 можно видеть, что каждому массиву (узлы, идентифицированные датами) соответствуют объекты. При этом в центральной части сети располагаются объекты, общие для нескольких событий (О-зона), а «гребешки» на периферии соответствуют специальным объектам, отражающим специфику конкретных событий (С-зоны).

Также можно предложить критерий релевантности события, связанного с конкретной датой, общей тематике: чем большая часть объектов из него попадает в О-зону, тем он более релевантен тематике. Формально значение этого критерия $k_{i,N}$ для сюжета i тематике s может быть записано следующим образом:

$$k_{i,N} = \frac{|T_{i,N} \cap T_{s,N}|}{N},$$

где N – количество объектов, $T_{i,N}$ – множество значимых объектов события i , – $T_{s,N}$ – множество значимых объектов для всей тематики.

Подход к прогнозу: R/S-анализ

Для решения задач прогнозирования перспективным представляется применение теории фракталов при анализе информационного пространства. Фрактальный анализ самоподобия информационных массивов может рассматриваться как технология, предназначенная для осуществления аналитических исследований с элементами прогнозирования, пригодная к экстраполяции полученных зависимостей.

Важнейшей характеристикой рядов, которые имеют хаотичное поведение, является, как известно, показатель Херста (H), определяемый в результате так называемого R/S -анализа [8]. Этот показатель базируется на анализе нормированного разброса – отношения разброса значений исследуемого ряда R к стандартному отклонению S .

Достаточно часто, когда соотношение R/S имеет постоянный тренд, можно говорить о соотношении:

$$R/S = (N/2)^H,$$

где H – показатель Херста, который для достаточно широкого класса рядов связан с хаусдорфовой (фрактальной) размерностью D постой формулой: $D + H = 2$.

На рис. 11 представлено соотношения R/S для рассматриваемого в этой работе временного ряда. Как можно видеть, кривая нормированного размаха удовлетворительно аппроксимируется прямой в двойном логарифмическом масштабе. Численные значения H характеризуют разные типы коррелированной динамики (персистентности). При

$H = 0,5$ наблюдается некоррелированное поведение значений ряда, а значение $0,5 < H < 1$ соответствует степени автокорреляции ряда.

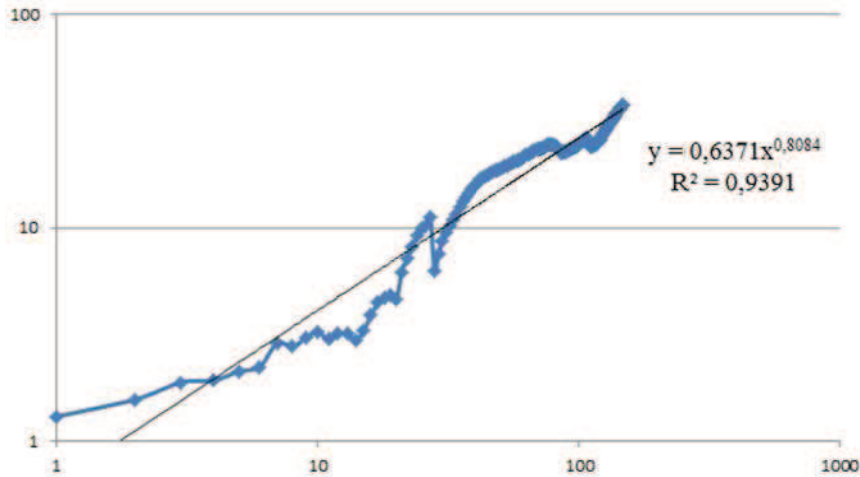


Рис. 11. Кривая R/S в двойной логарифмической шкале

Как можно видеть, значение показателя Херста для исследуемого информационного потока соответствует величине 0,81, что подтверждает предположение о самоподобии и итеративности рассматриваемых процессов в информационном пространстве. Это означает, что общая информационная напряженность остается на большом уровне, как только исчезает «шлейф» одного сюжета по выбранной тематике информационных, ему на смену возникает новый сюжет, т. е. его поведение в дальнейшем будет близко к предшествующему поведению.

Выводы

В докладе представлена методика аналитического исследования, которая базируется на использовании современных инструментальных средствах анализа и визуализации информационных потоков и временных рядов.

Предложенную методику можно использовать в качестве основы для проведения аналитической и

прогнозной деятельности на основе исследования контента современных компьютерных сетей.

Литература

1. Додонов А.Г., Ландэ Д.В. Моделирование и анализ тематических информационных потоков // Информационное противодействие угрозам терроризма, 2013. – № 20. – С. 52-59.

2. Ланде Д.В. Новітні підходи й технології інформаційно-аналітичної підтримки прийняття рішень // Національна безпека: український вимір: щокв. наук. зб. / Рада нац. безпеки і оборони України, Ін-т пробл. нац. Безпеки, 2008. – Вип. 1-2 (20-21). – С. 87-105.

3. Ланде Д.В. Тренди відображення інформаційних операцій в інформаційному просторі // Інформація і право, 2013. – N 1 (7). – С. 82-88.

4. Додонов А.Г., Ландэ Д.В., Прищепя В.В., Путятин В.Г. Конкурентная разведка в компьютерных сетях.– К.: ИПРИ НАН Украины, 2013. – 248 с.

5. Горбулін В.П., Додонов О.Г., Ланде Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія. – К.: Інтертехнологія, 2009. – 164 с.

6. Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения // Успехи физических наук, 1996. – 166. – № 11. – Р. 1145-1170.

7. Давыдов А.А. Системная социология. – М.: Издательство АКИ, 2008. – 192 с.

8. Федер Е. Фракталы. – М., Мир, 1991. – 261 с.

СОДЕРЖАНИЕ

<i>Додонов А.Г., Ландэ Д.В.</i>	
МЕТОДИКА АНАЛИТИЧЕСКОГО ИССЛЕДОВАНИЯ ДИНАМИКИ СОБЫТИЙ НА ОСНОВЕ МОНИТОРИНГА ВЕБ-РЕСУРСОВ СЕТИ ИНТЕРНЕТ.....	3
<i>Архипов О.Є, Архипова Є.О.</i>	
ОСОБЛИВОСТІ РОЗУМІННЯ ПОНЯТЬ «ІНФОРМАЦІЙНА БЕЗПЕКА» ТА «БЕЗПЕКА ІНФОРМАЦІЇ».....	18
<i>Горбачик О.С.</i>	
ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ЯК ФАКТОР ТРАНСФОРМАЦІЇ СИСТЕМ ОРГАНІЗАЦІЙНОГО УПРАВЛІННЯ.....	31
<i>Березін Б.О., Ланде Д.В., Шиховець О.В.</i>	
ЖИВУЧІСТЬ ІНФОРМАЦІЙНИХ ОБ'ЄКТІВ ПРИ ДОВГОТЕРМІНОВОМУ ЗБЕРІГАННІ ІЗ ВИКОРИСТАННЯМ АРХІВНИХ НОСІЇВ.....	40
<i>Брайчевський С.М.</i>	
ЕЛЕКТРОННІ ЕНЦИКЛОПЕДИЧНІ РЕСУРСИ ЯК ЕЛЕМЕНТ ТЕХНОЛОГІЇ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ.....	46
<i>Кузнецова М.Г.</i>	
ПРОБЛЕМИ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ В СИСТЕМАХ ОРГАНІЗАЦІЙНОГО УПРАВЛІННЯ.....	50
<i>Левченко О.В.</i>	
СИСТЕМА ІНДИКАТОРІВ ОЦІНКИ СТАНУ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ ДЕРЖАВИ.....	57
<i>Сенченко В.Р.</i>	
КОНЦЕПЦІЯ ПОВУДОВИ ІНФОРМАЦІЙНО-АНАЛІТИЧНОЇ СИСТЕМИ УПРАВЛІННЯ МІГРАЦІЙНИМИ ПРОЦЕСАМИ В УКРАЇНІ.....	60
<i>Мельник К.С.</i>	
АКТУАЛЬНІ ПИТАННЯ ЗАХИСТУ ПЕРСОНАЛЬНИХ ДАНИХ В ІНФОРМАЦІЙНИХ РЕСУРСАХ.....	68