# Архитектура системы информационной поддержки на основе мониторинга информационного пространства и метода сценарного анализа

Додонов А.Г., Ландэ Д.В.
Институт проблем регистрации информации НАН Украины

Создание автоматизированной системы информационной поддержки процессов противодействия враждебным информационным компаниям, выполнения мероприятий по отражению деструктивных внешних и внутренних информационных воздействий – актуальная проблема, особенно в условиях ведения информационных войн [1], [2].

Как основные задачи такой системы можно рассматривать:

- построение сценариев противодействия информационным деструктивным воздействиям на основе созданной онтологии понятий;
- контент-мониторинг (непрерывный во времени содержательный анализ) информационного пространства с учетом знаний экспертов;
- выявления закономерностей (трендов) путем анализа динамики изменения значений отдельных факторов безопасности;
- выявление информационных воздействий, информационных операций;
- прогнозирование развития информационных сюжетов, ситуаций;
- оценка эффективности процедур поддержки информационной безопасности.

Соответственно, для реализации такой системы информационной поддержки процессов, в частности, связанных с национальной безопасностью необходимо:

- создать онтологию понятий предметной области (узлов - факторов безопасности и соответствующих причинно-следственных связей – зависимостей факторов), определить вид целевой функции безопасности объектов-узлов этой онтологии в зависимости от значений факторов безопасности;
- постоянно актуализировать значения факторов безопасности и связей в зависимости от результатов мониторинга информационного пространства и знаний экспертов;
- определять возможные сценарии на основе анализа онтологии и выявления соответствующих частичных онтологий;
- анализировать динамику изменения значений отдельных факторов и связей с целью выявления закономерностей, прогнозирования;
- постоянно проводить оценку эффективности проводимой информационной поддержки.

В соответствии с этими задачами, предполагается, что система информационной поддержки должна состоять из трех основных подсистем и интерфейсов с администраторами и пользователями (рис. 1).


Рис. 1. Архитектура системы

**Онтологии понятий предметной области**

Онтология в данном случае представляет собой функциональный аналог базы знаний, отражающей знания экспертов о предметной области, т.е. в качестве узлов графа онтологии выбираются важнейшие факторы предметной области обеспечения безопасности, а в качестве связей – причинно-следственные связи между факторами (с математической точки зрения – граф с направленными ребрами) [3]. Узлам и связям приписываются числовые значения, которые в дальнейшем могут корректироваться. Связи также могут иметь различные веса (сила влияния) и быть как положительными (увеличение значения первого фактора приводит к увеличению значения второго фактора), так и отрицательными (увеличение значения первого фактора приводит к уменьшению значения второго фактора). Онтологии, как правило, создаются экспертами, однако, возможно автоматизированное создание онтологий на основе анализа текстовых корпусов соответствующего содержания. На рис. 2 приведен набор основных модулей, составляющих подсистему ведения онтологий.

Инструменты ведения онтологий – современные онтологические редакторы, а также специальные СУБД, ориентированные на хранение сетевых структур.

Выбор целевой функции для расчета заданного уровня обеспечения безопасности объектов – узлов онтологии – осуществляется экспертами. Чаще всего целевая функция, зависящая от значений факторов, линейная, отличаются лишь коэффициенты, которые подбираются экспертно, а затем итеративно уточняются во время эксплуатации автоматизированной системы. Линейная функция позволяет упрощать вычисления в случае большой размерности пространства факторов, однако, в некоторых случаях ввиду зависимости факторов, целевая функция должна принимать нелинейный вид.

Предполагается постоянная актуализация значений факторов безопасности и связей между ними в зависимости от объемов и содержания сообщений,

появляющихся в целевом фрагменте информационного пространства, и знаний экспертов. Для мониторинга информационного пространства необходимо использовать специализированные системы контент-мониторинга веб-пространства, социальных медиа, СМИ и т.п.


Рис. 2. Архитектура подсистемы ведения онтологий

**Определение возможных сценариев на основе анализа онтологии**

Сценарии информационной поддержки, как правило, связываются с определенными факторами безопасности (чаще всего объектами и уязвимостями). После выбора целевых факторов сценария в графе онтологии выявляются подграфы (частичные онтологии), наиболее тесно связанные с выбранными факторами. Далее решается задача частичной оптимизации целевой функции на выбранных подграфах, т.е. вычисляется целевая функция в зависимости от изменений факторов безопасности, соответствующих выбранным сценариям. После этого выбираются значения факторов безопасности, изменяемых в данных подграфах, соответствующие наилучшим значениям целевой функции.

Корректировка целевой функции может выполняться экспертами в зависимости от изменения приоритетов безопасности, изменения структуры графа онтологии. Кроме того, в результате использования данной системы возможно устранение некоторых возможных уязвимостей в информационном пространстве, выявление «скрытых» связей, изменение существующих связей между факторами, создание дополнительных связей (в том числе, путем информационного противодействия, воздействия на реальные объекты информационной инфраструктуры как собственной, так и вероятного противника).

**Динамика изменения значений отдельных факторов и связей**

Анализ динамики изменения значений отдельных факторов безопасности и связей во времени (как по их отражению в информационном пространстве, так и внесенных экспертами) позволяет выявлять некоторые закономерности изменения этих факторов (периодичности, тренды, аномалии) путем применения современных средств цифровой обработки сигналов (дисперсионный-, вейвлет-анализ и т.п.), выявлять возможные информационные операции путем сравнения с соответствующими шаблонами их динамики, а также выполнять прогнозирование [4].

Существует свыше 100 методов прогнозирования на основе анализа временных рядов, в частности, рядов динамики объемов тематических публикаций. Среди них, методы, которые базируется на статистическом и структурном анализе. К их числу относятся: экспоненциальное сглаживание, методы скользящего окна, авторегрессионный анализ, методы нейронных сетей.

Один из подходов к прогнозированию базируется на фрактальном анализе. Для данного ряда строится зависимость R/S от номера элемента ряда. Если эта зависимость близка к степенной, то показатель степени (коэффициент Херста), позволяет говорить о том, будет ли поведение ряда персистентным, т.е. в дальнейшем будет близко к предшествующему поведению.

Оценка эффективности информационной поддержки принятия решений с точки зрения поставленных целей позволяет реализовывать обратную связь, т.е. корректировать целевую функцию, модифицировать онтологию, расширять информационную и экспертную поддержку, планировать дальнейшее развитие системы.

**Контент-мониторинг информационного пространства**

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 [5]. 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.

Для эффективного проведения информационно-аналитических исследований на основе анализа контента сети Интернет в рамках рассматриваемой системы предлагается последовательность шагов, этапов обработки информации. 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.

Применение систем контент-мониторинга информационного пространства позволяет проводить содержательный ретроспективный анализ, выходить на соответствующие публикации, события, сюжеты, выявлять новые, ранее неизвестные факторы безопасности и информационные взаимосвязи.

Информационная база системы, базирующейся на использовании, среди прочего, сетевого контента формируется средствами контент-мониторинга. Эти средства охватывают огромные объемы информации (Big Data) из динамически возрастающих информационных потоков при наличии шумовой информации, слабо доступных ресурсов (Hidden Web, Deep Web), так называемого «скрытого Интернета» [6]. Задачи подсистемы мониторинга информационного пространства (рис. 3) следующие:

- мониторинг целевых объектов;
- контроль медиаприсутствия и медиаактивности целевых объектов;
- выявление новых объектов мониторинга;
- выявление взаимосвязей объектов;
- формирование ретроспективных фондов.


Рис. 3. Архитектура подсистемы мониторинга информационных ресурсов

Подсисистема аналитической обработки

Подсистема аналитической обработки представляет собой аналитический блок системы информационной поддержки приятия решений (рис. 4), обеспечивающий решение следующих задач:

- нахождение релевантных тематических сообщений в информационном пространстве;
- определение динамики тематических сюжетов;
- определение критических точек в динамике тематических сюжетов;
- отслеживание сюжетных цепочек, соответствующих событиям, процессам;
- выявление основных событий и объектов из тематического сюжета;
- визуализация взаимосвязей событий и объектов мониторинга, а также объектов мониторинга между собой.


Рис. 4. Подсистема аналитической обработки

В соответствии со своим назначением данная подсистема, вместе с подсистемой мониторинга информационного пространства, позволяет реализовать следующие этапы информационно-аналитического исследования [7]:

- выбор системы контент-мониторинга интернет-ресурсов;
- формирование запроса в среде выбранной системы. Нахождение тематических публикаций по запросу с помощью систем контент-мониторинга;
- определение динамики тематических публикаций по запросу;
- определение критических точек в динамике тематических публикаций;
- определение основных событий в критических точках;
- выявление объектов мониторинга;

- выявление и визуализация взаимосвязей;

- прогноз развития событий.

Для получения репрезентативной информации об объекте исследования необходимо воспользоваться системой контент-мониторинга, 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.

Эффективными для решения задач анализа динамического контента являются специализированные системы интеграции сетевого контента. В частности, в Украине доступны такие системы контент-мониторинга: InfoStream (www.infostream.ua), S-monitor (s-monitor.com), YouScan (youscan.ru) и др. реализующие необходимую функциональность.

Далее на языке выбранной Формирование запроса в среде выбранной системы контент-мониторинга формируется запрос для формирования массива тематических документов, релевантных выбранной предметной области (пример приведен на рис. 1).

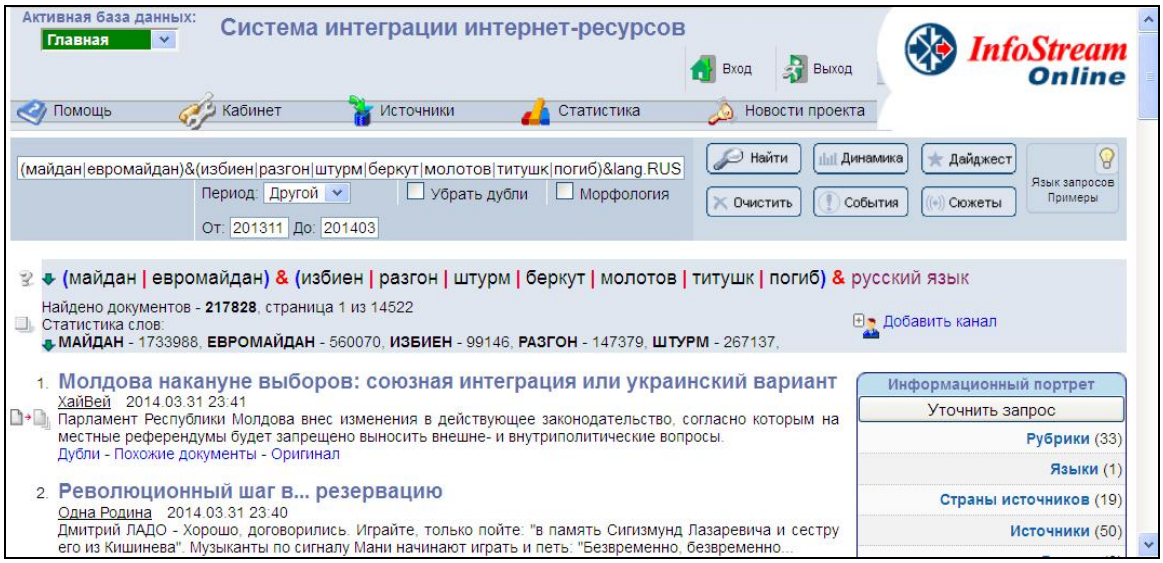



Рис. 5. Интерфейс системы контент-мониторинга

Подсистема мониторинга информационного пространства должна обеспечить получения данных о количестве публикаций по заданной предметной области за указанный промежуток времени (рис. 6). Эти данные могут быть загружены в настольную систему обработки данных и отображаются в виде графиков. В интерфейсе пользователя обеспечивается переход к просмотру релевантных документов по выбранной дате.


Рис. 6. Динамика публикаций по предметной области

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

В результате анализа многочисленных диаграмм поведения информационных потоков были выявлены наиболее типичные, базовые профили их поведения [1, 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. На рис. 7 приведен обобщенная диаграмма, соответствующая всем этапам жизненного цикла информационных операций [8]. Для выявления степени «близости» фрагментов исследуемого временного ряда приведенной диаграмме в различных масштабах предлагается использовать так называемый «вейвлет-анализ» [9], который в настоящее время нашел широкое применение, как в естественных науках, так и в социологии [10].

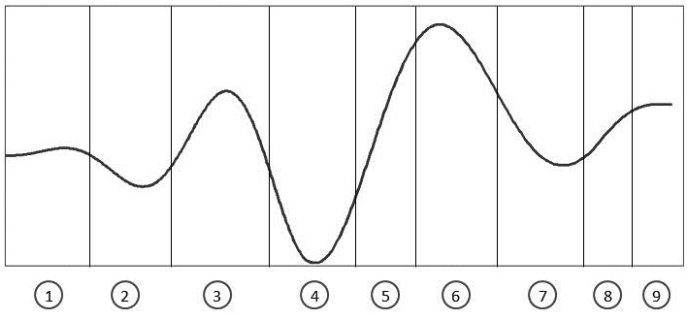



Рис. 7. Жизненный цикл информационных операций: 1 – фон; 2 – затишье; 3 – «артподготовка»; 4 – затишье; 5 – атака/триггер роста; 6 – пик завышенных ожиданий; 7 – утрата иллюзий организаторов; 8 – общественное осознание; 9 – продуктивность/фон

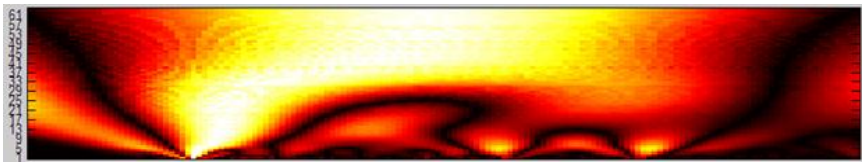
Главная идея применения вейвлет-преобразования в этом случае заключается в том, что нестационарный временной ряд разделяется на отдельные промежутки (так называемые «окна наблюдения»), и на каждом из них вычисляется величина, показывающая степень близости исследуемых данных с разными сдвигами некоторого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

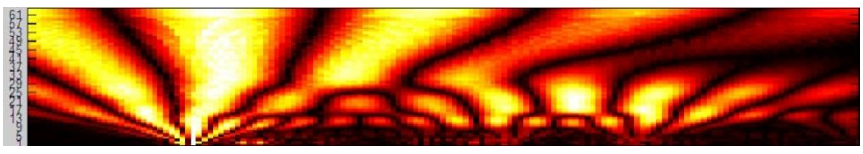
$$W(a,b) = (f(t),\psi(t)) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi^* \left( \frac{t-b}{a} \right) dt .$$

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 показывают, насколько поведение процесса в данной точке близко к вейвлету в данном масштабе. Чем ближе от анализируемой зависимости в окрестности данной точки к виду вейвлета, тем большую абсолютную величину имеет соответствующий коэффициент.

Вейвлеты «мексиканская шляпа» и Морле наиболее точно отражает динамику информационных операций, результаты применения этого вейвлета приведены на рис. 8, благодаря чему могут быть выбраны три даты, соответствующие критическим точкам исследуемого процесса.



а



б

Рис. 8. Вейвлет-спектограммы исследуемого временного ряда

(а – «мексиканская шляпа», б – вейвлет Морле)

Следует отметить, что инструменты построения вейвлет-спектограмм также 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).

После определения критических точек выполняется построение основных сюжетных цепочек из сообщений, соответствующих запросу за выбранные даты. Таким образом определяются основные события за указанные даты (рис. 9).

Для последующего анализа отбираются массивы сообщений, соответствующие выбранным датам, объекты из которых могут рассматриваться как объекты мониторинга.


Рис. 9. Основные сюжетные цепочки за выбранные даты

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. Эти списки могут агрегироваться, в результате чего возможно определение взаимосвязей событий и объектов (рис. 10).

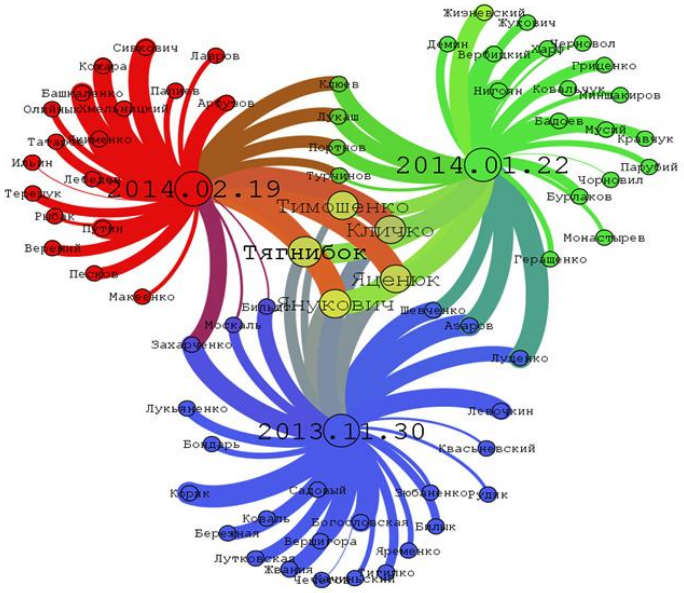



Рис. 10. Пример визуализации связей событий и объектов с помощью программы Gephi (www.gephi.org).

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

На рис. 10 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) [13].

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 $i$ тематики $s$ может быть записано следующим образом:

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

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

Для решения задач прогнозирования перспективным является 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 [12]. 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.

На рис. 11 представлено соотношения $R/S$ для рассматриваемого в этой работе примера. В данном случае 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.

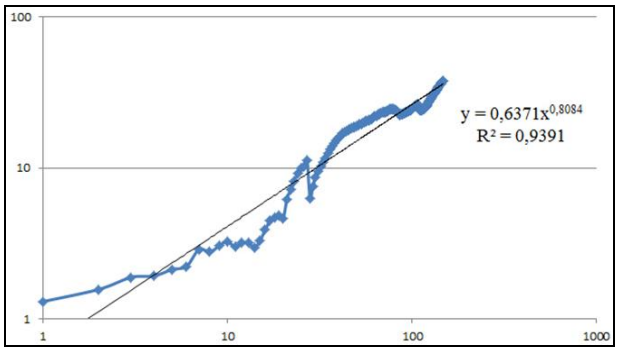



Рис. 11. Кривая *R/S* в двойной логарифмической шкале

Значение показателя Херста, превышающее 1/2, подтверждает предположение о самоподобии и итеративности рассматриваемых процессов в информационном пространстве. Это, в частности, для рассматриваемого примера означает, что общая информационная напряженность остается на высоком уровне – как только исчезает «шлейф» одного сюжета по выбранной тематике, ему на смену возникает новый сюжет, т.е. поведение исследуемого процесса в дальнейшем будет близко к предшествующему поведению.

***Выводы***

В статье представлена архитектура системы информационной поддержки принятия решений, идеология создания и использования онтологий для построения сюжетов информационного противодействия, детально рассмотрена методика аналитического исследования, которая базируется на использовании инструментальных средствах анализа и визуализации информационных потоков и временных рядов.

Предложенные архитектурные решения можно использовать при реалтизации систем информационной поддержки принятия решений, базирующихся на контент-мониторинге информационного пространства и сценарном анализе, а так же в качестве базы для проведения аналитической и прогнозной деятельности.

***Литература***


1. Горбулін В.П., Додонов О.Г., Ланде Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія. – К.: Інтертехнологія, 2009. – 164 с.
2. Шульц В.Л., Кульба В.В., Шелков А.Б., Чернов И.В. Сценарный анализ эффективности управления информационной поддержкой государственной политики России в Арктике. // Национальная безопасность / nota bene. – 2011. – № 6. – С. 104-137.
3. Боргест Н.М. Научный базис онтологии проектирования // Онтология проектирования. – 2013. №1(7). – С.7-25.
4. Шульц В.Л., Кульба В.В., Шелков А.Б., Чернов И.В. Структурно-динамический подход к сценарному анализу процессов информационного противоборства в Арктике // Труды XII Всероссийского совещания по проблемам управления (ВСПУ 2014). – М.: ИПУ РАН, 2014. – С. 8889-8901.
5. Додонов А.Г., Ландэ Д.В. Моделирование и анализ тематических информационных потоков // Информационное противодействие угрозам терроризма, 2013. – № 20. – С. 52-59.
6. Додонов А.Г., Ландэ Д.В., Путятин В.Г. Компьютерные сети и аналитические исследования. – К.: ИПРИ НАН Украины, 2014. – 486 с.
7. Додонов А.Г., Ландэ Д.В., Прищепа В.В., Путятин В.Г. Конкурентная разведка в компьютерных сетях. – К.: ИПРИ НАН Украины, 2013. – 248 с.
8. Додонов А.Г., Ландэ Д.В. Методика аналитического исследования динамики событий на основе мониторинга веб-ресурсов сети Интернет // Информационные технологии и безопасность: основы обеспечения информационной безопасности: Материалы международной научной конференции ИТБ-2014. – К.: ИПРИ НАН Украины, 2014. – С. 3-17.
9. Ланде Д.В. Тренди відображення інформаційних операцій в інформаційному просторі // Інформація і право, 2013. – N 1 (7). – С. 82-88.
10. Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения // Успехи физических наук, 1996. – 166. – № 11. – Р. 1145-1170.
11. Давыдов А.А. Системная социология. – М.: Издательство ЛКИ, 2008. – 192 с.
12. Федер Е. Фракталы. – М., Мир, 1991. – 261 с.
13. 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.