

ПРИМЕНЕНИЕ ГРАФОВ ГОРИЗОНТАЛЬНОЙ ВИДИМОСТИ В ИНФОРМАЦИОННОЙ АНАЛИТИКЕ

Д.В. Ландэ^{1,2}, д.т.н., с.н.с. (dwlande@gmail.com),

А.А. Снарский^{2,1}, д.ф.-м.н., профессор (asnarskii@gmail.com)

*¹Институт проблем регистрации информации НАН Украины,
Киев, Украина*

*²НТУУ «Киевский политехнический институт им. Игоря
Сикорского», Киев, Украина*

Описывается применение метода графов горизонтальной видимости (Horizontal Viability Graph – HVG) в информационной аналитике, при формировании модели предметной области, соответствующей специальному запросу к поисковой системе, а также при построении сети связей информационных источников, отражающих общую тему.

Введение

В настоящее время в естественных науках и технологиях получили широкое распространение методы исследования временных рядов, в основе которого лежит их преобразование в граф (сложную сеть). При таком отображении объединяются две развитые области исследований – нелинейные методы анализа временных рядов и методы теории сложных сетей. Появляется возможность применить развитые методы анализа сложных сетей [1, 2] к анализу временных рядов. Предложено несколько методов построения сетей на основе временных рядов [3, 4], в частности, так называемый граф горизонтальной видимости (Horizontal Visibility Graph – HVG) [5, 6].

Алгоритм построения графов видимости по временному ряду проиллюстрирован на Рис. 1. При построении графа горизонтальной видимости на горизонтальной оси (ось времени) отмечаются точки, соответствующие индексу моменту времени t_i , от которых в перпендикулярном направлении строятся отрезки высотой, равной значениям ряда в этих точках – $x(t_i)$. Узлами графа горизонтальной видимости являются внешние вершины построенных отрезков. Связь между вершинами в HVG считается существующей, если горизонтальная прямая, проведенная из одной из вершин пересекает отрезок другой вершины, не пересекая ни одного из построенных отрезков, находящихся между ними. Этот геометрический критерий можно записать следующим образом: два узла (элемента ряда),

например, t_m и t_n соединены связью, если (см. рис. 2), $x(t_m), x(t_n) > x(t_p)$ для всех $p: n < p < m$.

Работа посвящена применению концепции Horizontal Viability Graph к двум областям информационной аналитики – построению моделей предметных областей из ключевых слов, отражающих наиболее важные понятия в тематических информационных потоках, а также построению сети связей источников информации.

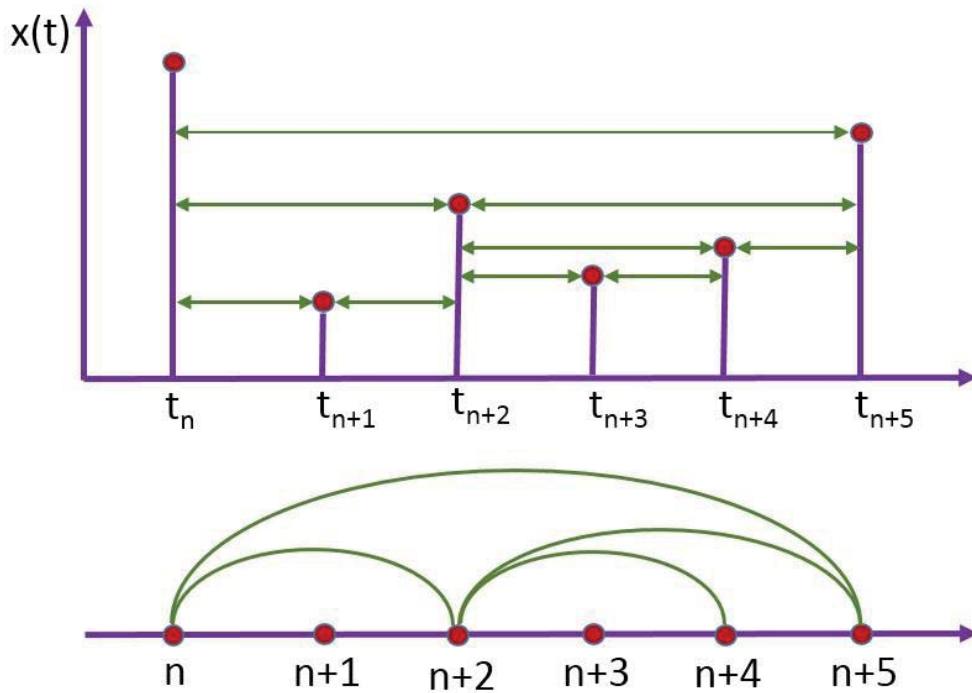


Рисунок 1 – Пример построения графа горизонтальной видимости

Сеть из ключевых слов – модель предметной области

Известно несколько подходов к построению так называемых сетей слов (Language Network) из текстов и способов интерпретации узлов и связей. Например, узлы могут быть соединены между собой, если соответствующие им слова стоят рядом в тексте, принадлежат одному предложению или абзацу, соединены синтаксически или семантически [7, 8]. В [9] был предложен метод построения сети слов путем анализа текстовых корпусов с применением графов видимости.

Поисковая система выдает по тематическому запросу документы, в каждом из которых выделены M ключевых слов, ранжированных в порядке важности, причем ранг определяется весовым значением ключевого слова, например, его относительная частота в документе (слова из стоп-словаря, естественно, не учитываются), TFIDF (TF –

частота слова в документе, IDF – величина, обратная количеству документов из массива, в котором встретился данный термин) и его модификации. Рассмотрим последовательность весовых значений ключевых слов из документов по некоторой узкой тематике, определяемых, например, запросом, следующих друг за другом в порядке их появления в информационном потоке (времени публикации). На Рис. 2 представлена цепочка документов и характеризующих их ключевых слов и их «видимость справа налево». Предполагается, что первый по времени документ в информационном потоке – D_1 , за ним следуют D_2 , D_3 и т.д. В пределах одного документа ввиду ранжированности по весу соседние ключевые слова связываются друг с другом (справа налево), и лишь лидирующие по весу ключевые слова могут связываться с ключевыми словами других документов (не обязательно с самими большими весовыми значениями).

Следует отметить, что одно и то же ключевое слово может присутствовать в разных документах и иметь в них различные весовые значения. При этом одно и то же ключевое слово в каждом документе может «видеть» различные ключевые слова, т.е. выходная степень его как узла графа горизонтальной видимости может превышать 1. Связь между некоторыми лидирующими по весу ключевыми словами из различных документов может ассоциироваться с «клубом богатых» [10, 11], феномену, присущему большинству сетей языка [12]. Традиционно связи в сетях языка строятся путем соединения тех узлов (ключевых слов), которые соответствуют одному фрагменту текста (документу, абзацу или предложению), т.е. в рассматриваемом примере степень каждого из узлов не может быть меньше M .

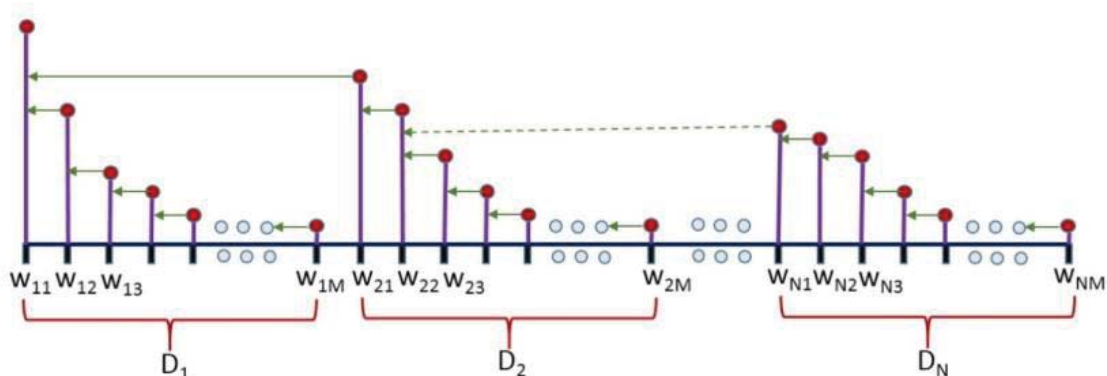


Рисунок 2 – Горизонтальная видимость ключевых слов w_{ij} из документов D_i

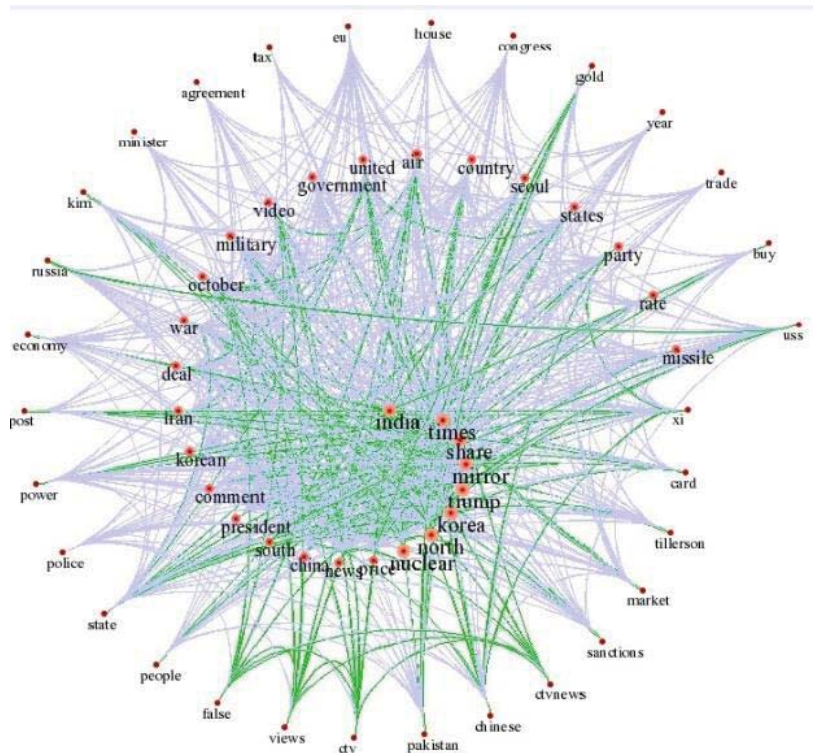
Оценить качество сети, построенной в соответствии с концепцией HVG могут лишь пользователи – аналитики информационной системы, но данный метод обеспечивает значительное сокращение количества связей между узлами (Рис. 3), оставляя наиболее значимые (между ближайшими по весу, что важно для ассортативных сетей [13, 14], в которых наряду с «феноменом клуба богатых» существенны связи между узлами, имеющими близкие степени). На рис. 3 представлены две сети связи терминов, первая – традиционная, а вторая построенная с применением предложенного метода. Как можно видеть, второй подход обеспечивает создание более наглядных сетевых структур.

Сеть взаимосвязи источников информации

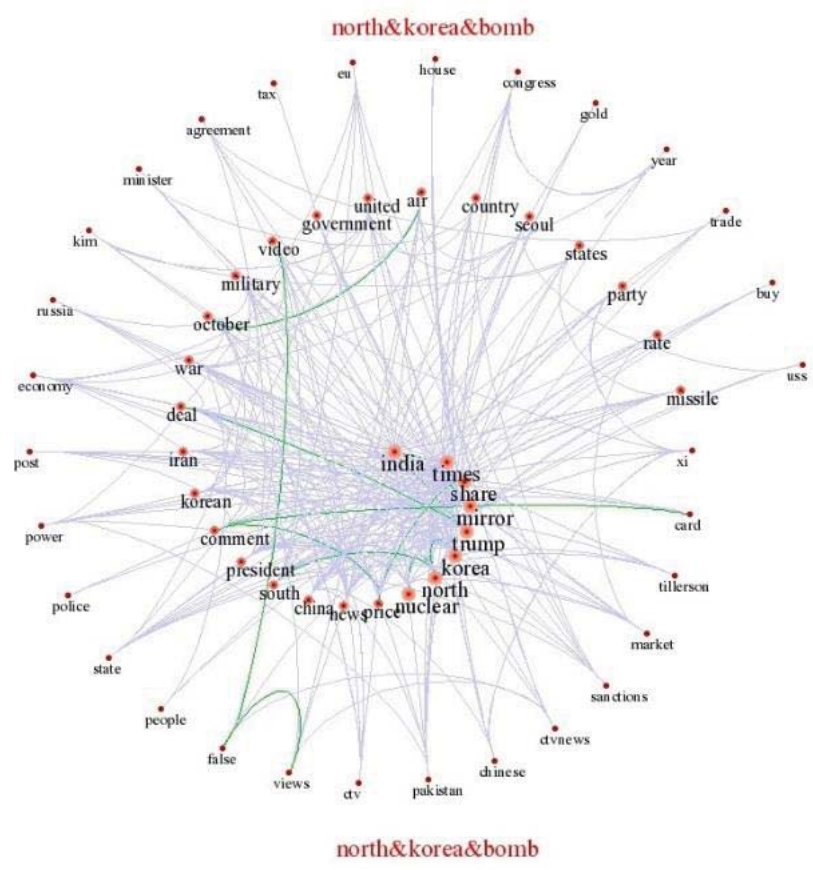
При анализе тематических информационных потоков из различных источников, формируемых системами контент-мониторинга, существует проблема построения сети связей источников информации. При этом основанием для связи между двумя источниками может служить тот факт, что они часто публикуют совпадающие или близкие по теме документы. Построение такой сети позволяет определять, какие из информационных источников по данной тематике являются основными, наиболее влиятельными, какие подвержены определенным информационным влияниям.

Предлагаемый подход базируется на предположении, что источники информации заранее ранжированы по объемам публикаций исходя из опыта наблюдения за ними в течение продолжительного времени, т.е. им уже приписаны некоторые весовые значения.

Предполагается построение временного ряда, элементы которого будут соответствовать документам из тематического информационного потока в порядке их появления в системе контент-мониторинга сетевых ресурсов (публикации на веб-ресурсах, в социальных сетях). Значения ряда будут соответствовать весовому значению информационного источника, опубликовавшему соответствующий документ. Предполагается, что весовое значение, степень важности источника определены заранее. По данному ряду, как и в предыдущем случае строится граф горизонтальной видимости, причем «направление взгляда» как и в предыдущем примере направлено в прошлое.



a)



b)

Рисунок 3 – Сети языка из ключевых слов: а) традиционная; б) Horizontal Visibility Graph

Таким образом устанавливается связь источника информации с другим, более рейтинговым, если он существует и опубликовал информацию раньше. Построенная таким образом сеть (Рис. 5) отражает связи источников по заданной тематике, позволяет определять лидеров среди них, делать предположения относительно первоисточников информации. Например, на Рис. 5 показана сеть связей международных источников информации, отражающих тематику ядерных испытаний в Северной Корее в октябре 2017 г. по данным системы контент-мониторинга.

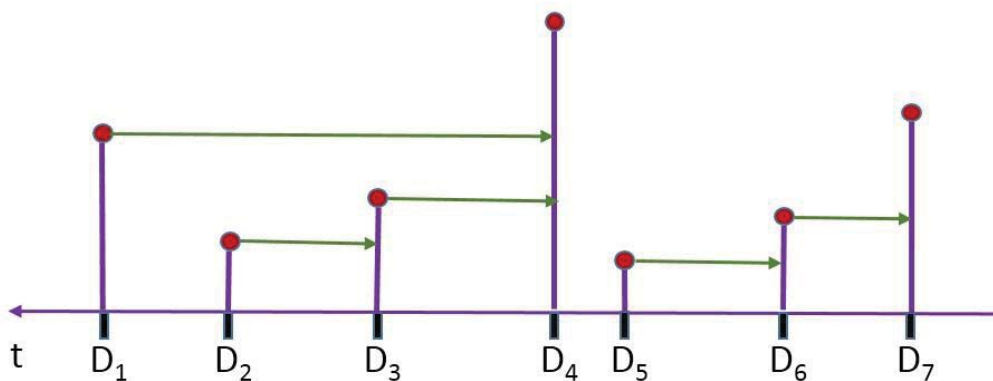


Рисунок 4 – Горизонтальная видимость документов D_i тематического информационного потока

Выводы

Предложены реализации концепции Horizontal Viability Graph при анализе тематических информационных потоков, которые, с одной стороны, расширяют сферу применения данного подхода, а, с другой стороны, позволяют решать задачи формирования и наглядной визуализации моделей предметных областей как сети терминов, соответствующих основным понятиям и связям между ними и сетей взаимосвязей источников информации.

Публикация содержит результаты исследований, проведенных при грантовой поддержке Государственного фонда фундаментальных исследований Украины по конкурсному проекту Ф73

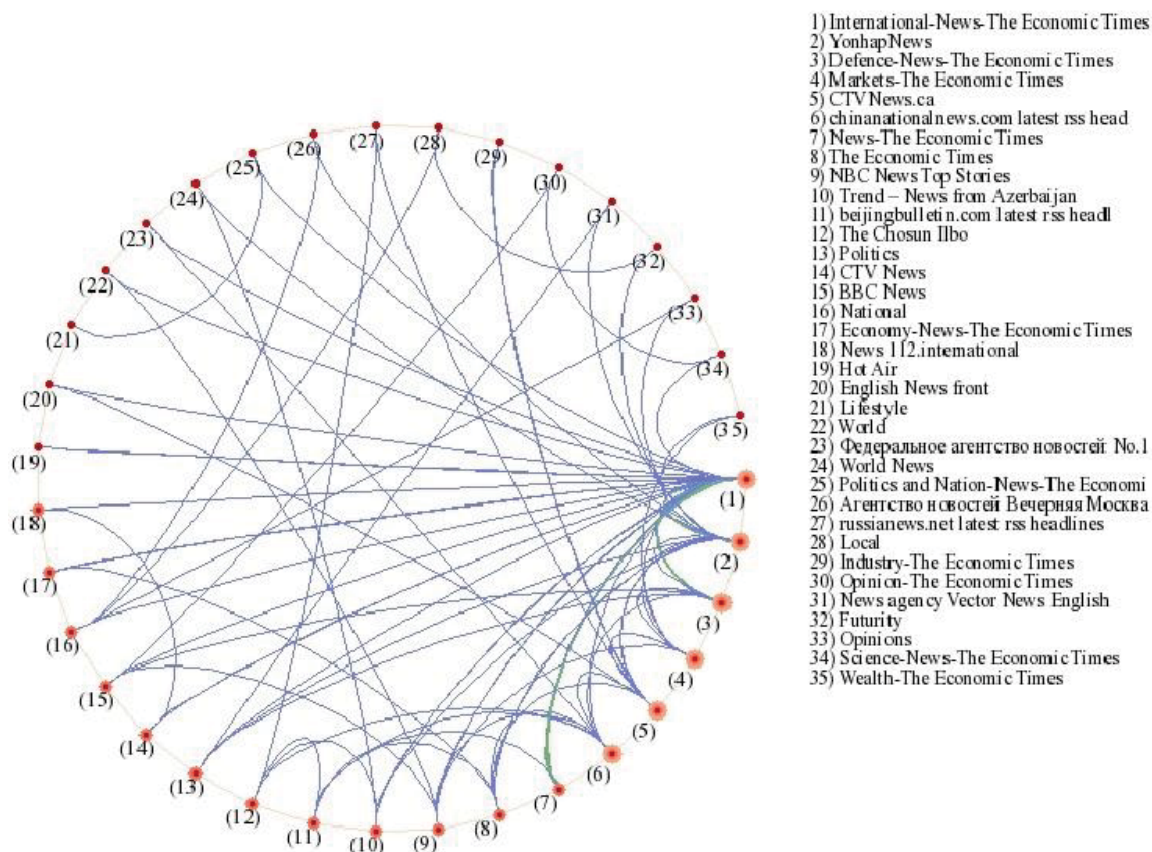


Рисунок 5 – Пример графа горизонтальной видимости, отражающего связи источников информации по заданной теме

Литература

1. Albert R., Barabási A.-L. Statistical mechanics of complex networks // *Rev. Mod. Phys.*, 2002. – 74. – pp. 47-97.
2. Newman M.E.J. The structure and function of complex networks // *SIAM Rev.*, 2003. – 45. – pp. 167-256.
3. Nunez A. M., Lacasa L., Gomez J. P., Luque B. Visibility algorithms: A short review // *New Frontiers in Graph Theory*, Y. G. Zhang, Ed. Intech Press, ch. 6. – pp. 119-152 (2012).
4. Bezsudnov I.V., Snarskii A.A. From the time series to the complex networks: The parametric natural visibility graph // *Physica A: Statistical Mechanics and its Applications*, 2014, 414, 53-60.
5. Luque B., Lacasa L., Ballesteros F., Luque J. Horizontal visibility graphs: Exact results for random time series // *Physical Review E*, – pp. 046103-1–046103- 11 (2009).
6. Gutin G., Mansour T., Severini S. A characterization of horizontal visibility graphs and combinatoris on words // *Physica A*, – 390 – pp. 2421-2428 (2011).

7. Ferrer-i-Cancho R., Sole R. V. The small world of human language // Proc. R. Soc. Lond. – B 268, 2261 (2001).
8. Caldeira S. M. G., Petit Lobao T. C., Andrade R. F. S., Neme A., Miranda J. G. V. The network of concepts in written texts // Preprint physics/0508066 (2005). [6] Ferrer-i-Cancho R., Sole R.V., Kohler R. Patterns in syntactic dependency networks // Phys. Rev. E 69, 051915 (2004).
9. D. V. Lande, A. A. Snarskii, E. V. Yagunova, E. V. Pronoza The Use of Horizontal Visibility Graphs to Identify the Words that Define the Informational Structure of a Text // 12th Mexican International Conference on Artificial Intelligence, 2013. – pp. 209-215.
10. Colizza, V. and Flammini, A. and Serrano, M. A. and Vespignani, A. Detecting rich-club ordering in complex networks // *Nature Physics*. 2, 2006. – 2: 110–115.
11. Julian J. McAuleya, Luciano da Fontoura, Tibério S. Caetan. Program Rich-club phenomenon across complex network hierarchies // *Applied physics letters*, 2007, -V 91 (8), 084103
12. Heuvel M.P., Sporns O. Rich-Club Organization of the Human Connectome // *Journal of Neuroscience*, 2011. – 31 (44). – pp. 15775-15786.
13. M.E.J. Newman. Assortative mixing in networks // *Phys. Rev. Lett.* **89**, 208701 (2002).
14. Piraveenan, M., Prokopenko M., and A. Y. Zomaya. Local assortativeness in scale-free networks // *EPL (Europhysics Letters)* 84.2, 28002 (2008).

**НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ
ИНСТИТУТ ПРОБЛЕМ РЕГИСТРАЦИИ ИНФОРМАЦИИ
НАН УКРАИНЫ**

**ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ
И БЕЗОПАСНОСТЬ**

**МАТЕРИАЛЫ XVII МЕЖДУНАРОДНОЙ
НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ**

ВЫПУСК 17

Киев – 2017

*Рекомендовано к печати Ученым советом
Института проблем регистрации информации НАН Украины
(протокол № 3 от 5 декабря 2017 г.)*

**Информационные технологии и безопасность. Материалы XVII
Международной научно-практической конференции ИТБ-2017.** – К.:
ООО "Инжиниринг", 2017. – 292 с. ISBN 978-966-2344-59-2

В сборник вошли материалы докладов, представленных на XVI
Международной научно-практической конференции «Информационные
технологии и безопасность» (ИТБ-2017, 30 ноября 2017 года, г. Киев,
Украина).

В сборнике представлены статьи, посвященные вопросам
кибернетической безопасности критических инфраструктур,
моделированию и противодействию информационным операциям,
технологиям информационно-аналитических исследований на основе
открытых источников информации, онтологическому подходу,
семантическим сетям, сценарному анализу при обеспечении
информационной поддержки принятия решений, компьютерному
моделированию процессов и систем, актуальным проблемам
технологического и правового обеспечения информационной и
кибернетической безопасности.

Для специалистов в области информационных технологий,
информационной безопасности, информационного права а также для
аспирантов и студентов старших курсов высшей школы соответствующих
специальностей.

Редакционная коллегия:

*А.Г. Додонов, д.т.н., профессор; А.М. Богданов, д.т.н., профессор;
В.В. Голенков, 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.В. Хаджинов, д.т.н.,
профессор; В.В. Циганок, д.т.н., с.н.с.; 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., О.В. Андрейчук,
к.т.н., Гулякина Н.А., к.т.н., профессор*

© Институт проблем регистрации
информации НАН Украины,
2017

ISBN 978-966-2344-59-2

© Коллектив авторов, 2017

СОДЕРЖАНИЕ

<i>Додонов О.Г., Горбачик О.С., Кузнєцова М.Г.</i> ОРГАНІЗАЦІЯ УПРАВЛІННЯ ГРУПОЮ МОБІЛЬНИХ ТЕХНІЧНИХ ОБ'ЄКТІВ	3
<i>Андрущенко В.Б., Балагура І.В.</i> АНАЛІЗ ПУБЛІКАЦІЙНОЇ АКТИВНОСТІ ЗА НАПРЯМКОМ КОМП'ЮТЕРНОЇ БЕЗПЕКИ НА БАЗІ РЕСУРСІВ WEB OF SCIENCE ТА SCOPUS	8
<i>Баранов О.А.</i> ІНТЕРНЕТ РЕЧЕЙ (ІОТ): ІНТЕГРАЛЬНА БЕЗПЕКА	18
<i>Березин Б., Ландэ Д., Павленко О.</i> РАЗРАБОТКА, ОЦЕНКА И ИСПОЛЬЗОВАНИЕ АЛГОРИТМА СЕГМЕНТАЦИИ СЛОВ ДЛЯ СИСТЕМ МОНИТОРИНГА НАЦИОНАЛЬНЫХ ИНТЕРНЕТ- РЕСУРСОВ	22
<i>Бойченко А.В.</i> ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ СЦЕНАРНОГО АНАЛІЗУ НА БАЗІ ДОСЛІДЖЕННЯ МОДЕЛЕЙ ПРЕДМЕТНИХ ОБЛАСТЕЙ	32
<i>Galata L., Korniyenko B., Yudin A.</i> RESEARCH OF THE SIMULATION POLYGON FOR THE PROTECTION OF CRITICAL INFORMATION RESOURCES	35
<i>Губська Д.О.</i> ПРАВОВІ ПИТАННЯ ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНОЇ І КІБЕРНЕТИЧНОЇ БЕЗПЕКИ ПЛАТЕЖІВ	52
<i>Humennyi D., Khlaponin Yu., Parkhomey I., Rudnitska O.</i> STRUCTURAL MODEL OF ROBOT-MANIPULATOR FOR CAPTURE OF NO-COOPERATION CLIENT SPACESCRAFT	63
<i>Добровська С.В.</i> ВИЗНАЧЕННЯ ПУБЛІКАЦІЙНОЇ АКТИВНОСТІ В НАУКОВИХ НАПРЯМКАХ, ЯКІ СПРИЯЮТЬ ОБОРОНОЗДАТНОСТІ КРАЇНИ (ЗАХИСТ ІНФОРМАЦІЇ, КОМП'ЮТЕРНА БЕЗПЕКА)	77
<i>Довгополий А., Олег Білобородов О.</i> АНАЛІЗ ТЕХНОЛОГІЙ ІНФОРМАЦІЙНО- ПСИХОЛОГІЧНИХ ВІЙН ТА ІНФОРМАЦІЙНО- ПСИХОЛОГІЧНОЇ ЗБРОЇ РОСІЙСЬКОЇ ФЕДЕРАЦІЇ	83

<i>Зубок В. Ю., Захарченко О.І., Беланов Ю.О.</i>	
РОЗПІЗНАННЯ АНОМАЛЬНИХ СТАНІВ В ІНФОРМАЦІЙНО-ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМАХ ПРИ НЕЧІТКОМУ ОПИСІ ПОДІЙ	92
<i>Kadenko S.V.</i>	
DEFINING RELATIVE WEIGHTS OF DATA SOURCES DURING AGGREGATION OF PAIR-WISE COMPARISONS	97
<i>Кузнєцова Н.В.</i>	
ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ АНАЛІЗУ КЛІЄНТСЬКОЇ БАЗИ АБОНЕНТІВ ТА ПРОГНОЗУВАННЯ ЇХ ПОВЕДІНКИ	114
<i>Kuzminykh V.A., Koval A.V., Osipenko M.V.</i>	
METHOD OF MACHINE LEARNING BASED ON STOCHASTIC AUTOMATA IN PROBLEMS OF DATA CONSOLIDATION FROM OPEN SOURCES	121
<i>Кузьмичов А.І.</i>	
АНАЛІЗ ВІДКРИТИХ НАБОРІВ ДАНИХ НА ПРИКЛАДІ ОЦІНЮВАННЯ СВІТОВОЇ ЕКОЛОГІЧНОЇ ПРОБЛЕМИ ІЗ КРИТИЧНИМ РІВНЕМ CO₂ В АТМОСФЕРІ	131
<i>Кунченко-Харченко В., Огірко І., Огірко О.</i>	
ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ ТА МОДЕЛЮВАННЯ КІБЕРБЕЗПЕКИ	139
<i>Кучеров Д.</i>	
КЕРУВАННЯ ПЕРЕВАНТАЖЕННЯМ КОМП'ЮТЕРНОЇ МЕРЕЖІ	149
<i>Lande D.V., Andriichuk, O.V., Hraivoronska A.M., Guliakina N.A.</i>	
APPLICATION OF DECISION-MAKING SUPPORT, NONLINEAR DYNAMICS, AND COMPUTATIONAL LINGUISTICS METHODS DURING DETECTION OF INFORMATION OPERATIONS	159
<i>Ландэ Д.В., Снарский А.А.</i>	
ПРИМЕНЕНИЕ ГРАФОВ ГОРИЗОНТАЛЬНОЙ ВИДИМОСТИ В ИНФОРМАЦИОННОЙ АНАЛИТИКЕ	175
<i>Мохор В., Цуркан О., Цуркан В., Герасимов Р.</i>	
ОЦІНЮВАННЯ ЗАХИЩЕНОСТІ ІНФОРМАЦІЇ В КОМП'ЮТЕРНИХ СИСТЕМАХ ЗА СОЦІОІНЖЕНЕРНИМ ПІДХОДОМ	183