

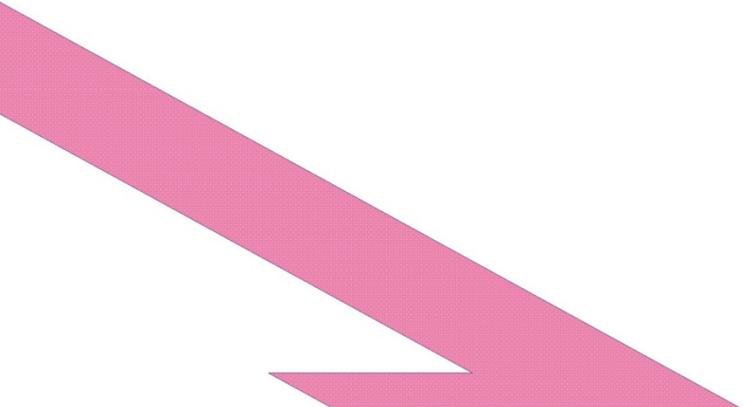


**I T H E A**



**International Journal**

**INFORMATION** **TECHNOLOGIES**  
**&**  
**KNOWLEDGE**



**2018** **Volume 12** **Number 1**



---

---

**International Journal**  
**INFORMATION TECHNOLOGIES & KNOWLEDGE**  
**Volume 12 / 2018, Number 1**

**EDITORIAL BOARD**

Editor in chief: **Krassimir Markov** (Bulgaria)

<b>Abdelmgeid Amin Ali</b>	(Egypt)	<b>Larissa Zaynutdinova</b>	(Russia)
<b>Aleksey Voloshin</b>	(Ukraine)	<b>Levon Aslanyan</b>	(Armenia)
<b>Alexander Kuzemin</b>	(Ukraine)	<b>Luis F. de Mingo</b>	(Spain)
<b>Alexander Palagin</b>	(Ukraine)	<b>Natalia Ivanova</b>	(Russia)
<b>Alexey Petrovskiy</b>	(Russia)	<b>Nataliia Kussul</b>	(Ukraine)
<b>Alfredo Milani</b>	(Italy)	<b>Natalia Pankratova</b>	(Ukraine)
<b>Arnold Sterenharz</b>	(Germany)	<b>Nelly Maneva</b>	(Bulgaria)
<b>Avram Eskenazi</b>	(Bulgaria)	<b>Nikolay Lyutov</b>	(Bulgaria)
<b>Darina Dicheva</b>	(USA)	<b>Olena Chebanyuk</b>	(Ukraine)
<b>Ekaterina Solovyova</b>	(Ukraine)	<b>Orly Yadid-Pecht</b>	(Israel)
<b>George Totkov</b>	(Bulgaria)	<b>Rafael Yusupov</b>	(Russia)
<b>Givi Bedianashvili</b>	(Georgia)	<b>Rumyana Kirkova</b>	(Bulgaria)
<b>Gocha Tutberidze</b>	(Georgia)	<b>Stoyan Poryazov</b>	(Bulgaria)
<b>Hasmik Sahakyan</b>	(Armenia)	<b>Tatyana Gavrilova</b>	(Russia)
<b>Iliia Mitov</b>	(Bulgaria)	<b>Vadim Vagin</b>	(Russia)
<b>Irina Petrova</b>	(Russia)	<b>Vasil Sgurev</b>	(Bulgaria)
<b>Ivan Popchev</b>	(Bulgaria)	<b>Velina Slavova</b>	(Bulgaria)
<b>Jeanne Schreurs</b>	(Belgium)	<b>Vitaliy Lozovskiy</b>	(Ukraine)
<b>Juan Castellanos</b>	(Spain)	<b>Vladimir Ryazanov</b>	(Russia)
<b>Karola Witschurke</b>	(Germany)	<b>Martin P. Mintchev</b>	(Canada)
<b>Koen Vanhoof</b>	(Belgium)	<b>Yuriy Zaychenko</b>	(Ukraine)
<b>Krassimira B. Ivanova</b>	(Bulgaria)	<b>Zurab Munjishvili</b>	(Georgia)

International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" (IJ ITK)  
is official publisher of the scientific papers of the members of the ITHEA International Scientific Society

IJ ITK rules for preparing the manuscripts are compulsory.

The rules for the papers for IJ ITK are given on [www.ithea.org](http://www.ithea.org)

Responsibility for papers published in IJ ITK belongs to authors.

**International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Volume 12, Number 1, 2018**  
Edited by the **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria, in collaboration with:  
University of Telecommunications and Posts, Bulgaria, V.M.Glushkov Institute of Cybernetics of NAS, Ukraine,  
Universidad Politécnica de Madrid, Spain, Hasselt University, Belgium,  
University of Perugia, Italy, Institute for Informatics and Automation Problems, NAS of the Republic of Armenia  
St. Petersburg Institute of Informatics, RAS, Russia,

**Printed in Bulgaria**  
**Publisher ITHEA®**

Sofia, 1000, P.O.B. 775, Bulgaria. [www.ithea.org](http://www.ithea.org), e-mail: [info@foibg.com](mailto:info@foibg.com)  
Technical editor: Ina Markova

Издател: ИТЕА®, София 1000, ПК 775, България, [www.ithea.org](http://www.ithea.org), e-mail: [info@foibg.com](mailto:info@foibg.com)

**Copyright © 2018 All rights reserved for the publisher and all authors.**

® 2007-2018 "Information Technologies and Knowledge" is a trademark of ITHEA®

© ITHEA® is a registered trademark of FOI-Commerce Co.

**ISSN 1313-0455 (printed)**

**ISSN 1313-048X (online)**

## МЕТОДОЛОГИИ СИСТЕМ И СИСТЕМНОГО АНАЛИЗА КАК ОСНОВА РАЗВИТИЯ ИНФОРМАТИКИ

Наталья Д. Панкратова, Виолетта Н. Волкова

**Abstract:** Рассматривается история возникновения и развития междисциплинарных научных направлений систем и системного анализа. Приводятся основные результаты осмысления этапов развития и состояния теории систем и системного анализа в научных работах двух школ системного анализа, в развитии которых принимают участие авторы данной статьи. Предлагается идея возрождения информатики как обобщающей науки об информации.

**Keywords:** закономерности, информатика, информационная система, информационно-управляющий комплекс, классификации систем, метод, модель, системный анализ, теория систем.

**ACM Classification Keywords:** H.4.2. Information System Application: type of system strategy

---

### Введение

В истории развития информатики было несколько периодов [Волкова и Черный, 2016]. Термин информатика (нем. Informatik, франц. Informatique, англ. Informatics) впервые появился в 1957 г. в Германии **К. Штейнбух**, затем пять лет спустя в 1962 г. независимо друг от друга слово «информатика» предложили **Ф. Дрейфус** во Франции и **У. Бауэр** в США, а в 1963 г. в журнале «Известия вузов. Электромеханика» вышла небольшая статья профессора Московского энергетического института (МЭИ) **Ф.Е. Темникова** [Темников, 1963], в которой «информатика» определялась как наука об информации, состоящая из трёх частей: теории *информационных элементов IE*, теории *информационных процессов IP* и теории *информационных систем*

$$IS_{\text{def}} = \langle IE, IP, IS \rangle. \quad (1)$$

Информатика (или инфология) по **Ф. Е. Темникову** — это расширенная теория информации, которая инициирована потребностями радиолокации и радиосвязи, развиваемых с 1948 г. **К. Шенноном**, **В. А. Котельниковым**, **А. Н. Колмогоровым**, **А. А. Харкевичем**, **Н. Винером**, **С. Райсом**, **Л. Бриллюэном**, **Д. Миддлтоном** и др. В концепции **Ф.Е. Темникова** информатика

должна была стать обобщающей теорией о принципах обработки информации и моделирования интеллектуальной деятельности в техно-, био- и социосфере, понималась не в узком смысле как «компьютерная наука», а выступала в качестве фундамента для понимания информационных процессов, в том числе протекающих и внутри аналоговых и цифровых вычислительных машин.

Однако проект информатики **Ф. Е. Темникова** остался практически неизвестен и не реализован. Вскоре возникли более частные дисциплины, которые стали называться информатикой.

В 1966 г. термин «информатика» стал использоваться **А. И. Михайловым, А. И. Чёрным и Р. С. Гиляревским** [Михайлов и др., 1968] в качестве наименования науки о научной информации. В этом значении термин приобрёл широкую известность, как в Советском Союзе, так и за рубежом.

Затем, с начала 1970-х гг., под влиянием западной традиции и концепции **А. П. Ершова** термин «информатика» стал означать технические и программные средства хранения и обработки информации при помощи ЭВМ — т. е. то, что в США и Великобритании называется *компьютерной наукой* (computer science).

В настоящее время ряд учёных считает, что следовало бы возродить и принять в качестве обобщающего определение **Ф. Е. Темникова**, что особенно актуально в условиях бурного развития инновационных технологий третьей и четвертой промышленных революций. При этом основой обобщающей информатики должна являться именно *информационная система*, включающая *информационные элементы* и *информационные процессы*.

Естественно, что основой такой обобщающей информатики должна стать теория систем. Поэтому в данной статье поставлена цель – осмыслить становление и развитие теории систем и ее прикладного направления – системного анализа, а также привести сферы их применения.

---

### **Возникновение междисциплинарных научных направлений**

---

В конце XIX в. по мере развития производственных процессов, науки и технологий, исследований функционирования и развития живых организмов стало резко увеличиваться число комплексных проектов и проблем, которые не решаются с помощью традиционных математических методов, требуют объединения разных методов и участия специалистов различных областей знаний. Возник интерес к системным представлениям. Понятие «система», ранее употреблявшееся в быденном смысле, превратилось в специальную общенаучную категорию, начали появляться обобщающие научные направления, которые исторически иногда возникали параллельно на разной прикладной или теоретической основе и носили различные наименования.

---

---

Важный вклад в становление системных представлений внес в начале XIX века А. А. Богданов (Малиновский), предложивший всеобщую организационную науку *тектологию* и сделавший попытку найти и обобщить организационные законы, проявления которых прослеживаются на неорганическом, органическом, социальном, культурном и др. уровнях.

Затем возникло междисциплинарное направление «**исследование операций**» (англ. *operations research*), инициированное потребностями военной сферы и занимающееся постановкой задач с учетом многокритериальности, отбора релевантных факторов, разработкой вариантов решения на основе использования методов теории оптимизации, статистического моделирования и эвристических подходов. В конце 1943–1944 гг. идея общности процессов в различных средах и науках появилась на объединенном совещании инженеров, физиологов и математиков, организованном Н. Винером совместно с Дж. фон Нейманом в Принстоне, на котором была предпринята попытка создания единой для различных наук терминологии. Фактически можно считать, что на этом совещании формировались идеи прикладных системных представлений. Но выбран был термин «*кибернетика*», который использовался в Древней Греции Платоном (*κῤῆτηνρεβῦ* – кормчий, рулевой), а в XIX веке – А. М. Ампером, Б.Ф. Трентовским (1843, <http://www.ecolife.ru>). Однако Н. Винер, выбирая название новой науки, образовал его от термина Дж. Максвелла «*governor*» – «регулятор». Термин стал широко распространяться после публикации в 1948 г. книги Н. Винера «*Cybernetics*» одновременно в Париже и американском Кембридже.

Понятия «система» и «теория систем» стали использоваться в различных областях знаний после того, как стали известны работы Л. фон Берталанфи, предложившего организмический подход к биологическим и социальным объектам и явлениям и понятие открытой системы – система, постоянно обменивающаяся веществом, энергией и информацией с внешней средой. Эта идея была впервые высказана Берталанфи в 1937 г. на семинаре по философии в Чикагском университете. Однако первые его публикации на эту тему появились только после Второй мировой войны. (Bertalanffy L. von., *General System Theory*. Изд.1.-е, ФРГ, 1945 [Исследования, 1969]).

Значимыми для становления теории систем являются работы К. Боулдинга, У. Росс Эшби, Р. Л. Акоффа, М. Месаровича, А. Д. Холла и др., которые в последующем были переведены и опубликованы в СССР [Садовский и др., 1969]. Большую роль в пропаганде идей теории систем в СССР играли инициаторы и редакторы переводов Э. Л. Наппельбаум, который читал публичные лекции по теории систем в московском Политехническом музее, В. Н. Садовский, Э. Г. Юдин, И. В. Блауберг, С. П. Никаноров, А. И. Уёмов, Ю. А. Урманцев. На развитие теории систем существенное влияние оказали работы П. К. Анохина, В. С. Тюхтина, В. Г. Афанасьева, В. А. Лефёвра, В. А. Лекторского; М. И. Сетрова; Б. В. Бирюкова; А. А. Малиновского,

Г. П. Щедровицкого, на развитие исследований в области кибернетики и информатики, а в последующем – на становление прикладной теории систем оказали влияние работы математиков (в частности, работы А. Н. Колмогорова и А. А. Ляпунова), и специалистов по кибернетике, теории информации и теории автоматического управления.

Первоначально в 1960-1970-е годы при постановке и исследовании сложных проблем проектирования и управления, при разработке автоматизированных информационных систем широкое распространение получил термин «*системотехника*», чему способствовало издание в 1957 г. монографии Г. Х. Гуда и Р. Э. Макола «Systems Engineering». При издании книги в СССР в 1961 г. в редакции издательства не понравился буквальный перевод «системная инженерия» или «инженерия систем», и был изобретён термин «*системотехника*», который предложил Ф. Е. Темников, создавший в 1969 г. в МЭИ первую в СССР кафедру системотехники [Волкова, 2001]. Однако в последующем, поскольку в термине в явном виде звучала «техника», термин стал использоваться в основном в приложениях системных методов к техническим направлениям, и возник термин «*системология*» (от др.-греч. σύνθεσις – целое, составленное из частей; λόγος – «слово», «мысль», «смысл», «понятие»). Термин предложен в 1965 г. И. Б. Новиком и широко использовался Б. С. Флейшманом, В. Т. Куликом, В. В. Дружининым и Д. С. Конторовым и др. для исследования систем как технической, так и биологической природы [Волкова, 2001].

Постепенно между философией и математикой сложился спектр прикладных междисциплинарных научных направлений (табл. 1).

Поскольку теорию систем исходно развивали философы и до сих пор в библиотечных классификациях относят к философским наукам, для прикладных исследований стали возникать другие термины. Вначале применяли термин *системный подход*. В последующем и в настоящее время прикладным направлением теории систем считается *системный анализ*.

---

### **Этапы развития системного анализа как прикладной научной методологии**

---

Далее ограничимся рассмотрением тех процессов 20 века, которые оказали непосредственное воздействие на становление и тенденции развития системной проблематики и методологии системного анализа. Выделим с прагматической точки зрения четыре этапа формирования системного анализа как прикладной научной методологии [Панкратова, 2002, 1999].

Таблица 1

## Междисциплинарные научные направления

Направления	Дата возникновения	Наиболее известные ученые
<b>Философия</b>		
Тектология	1903-1924 гг.	А. А. Богданов (Малиновский)
Теория систем Системный подход	1930-1950-е гг. В СССР – С 1960-х гг.	Л. фон Берталанфи, К. Боулдинг, Дж. ван Гиг, М. Месарович; В. Г. Афанасьев, И. В. Блауберг, В. Н. Садовский, Э. Г. Юдин, В. С. Тюхтин, А. И. Уёмов, Ю. А. Урманцев, и др.
Синергетика	С 1960-х гг.	И. Пригожин, И. Стенгерс, Г. Хакен, А. П. Руденко и др.
Системология	С 1970-х гг.	В. Т. Кулик, И. Б. Новик, Б. С. Флейшман, Б. Ф. Фомин и др.
Системный анализ (прикладная теория систем)	1960-е гг.  С 1970-х гг.	RAND-corporation, Э. Квейд, В. Кинг, Д. Клиланд, С. Оптнер, С. Янг, Э. Янч;  Е. П. Голубков, Н. Н. Моисеев, Ю. И. Черняк, С. П. Никаноров, Л. Е. Садовский, Г. Б. Клейнер, Ф. И. Перегудов, В. Н. Сагатовский, Ф. П. Тарасенко, В. З. Ямпольский, С. А. Валуев, В. Н. Волкова, А. А. Денисов, В. Н. Козлов, Д. Н. Колесников, Ю. И. Дегтярев, А. А. Емельянов и др.
Системотехника System Engineering	С 1962 г. С 1970-х гг.	Г. Гуд, Р. Макол, А. Холл, Ф. Е. Темников, Г. Честнат, В. В. Дружинин, Д. С. Конторов, В. И. Николаев, и др.
Кибернетика	1947 г. С 1960-х гг.	Н. Винер, У. Р. Эшби; А. И. Берг, Л. П. Крайзмер, М. Б. Игнатъев, Н. Е. Кобринский, Л. Т. Кузин, Е. З. Майминас, Л. А. Растрингин и др.
Исследование операций	С 1940-1950-х гг.	Ф. Морз, Дж. Кимбелл, Р. Акофф, Е. С. Вентцель, Т. Саати, М. Сасени, У. Черчмен и др.
<b>Специальные дисциплины</b>		

Первый этап становления системного анализа относится к первой половине XX века и является периодом появления и формирования основных идей системного мышления. Принимая во внимание введенное определение системного мышления [Панкратова, 2002], полагаем целесообразным за основу важнейших, основополагающих работ использовать перечень, который предложен в работе [Садовский и др., 1996], но дополнить его трудами В.И. Вернадского (1863–1945), положившего начало новому системному пониманию процессов, происходящих на нашей Планете. В результате научных исканий и размышлений о перспективе биосферы ученый обосновал гипотезу о естественной насущной необходимости трансформации биосферы в сферу разума, - ноосферу, обеспечивающую глобальный гомеостаз Природы, посредством рационального, гармоничного, уважительного сотрудничества человека с Природой [Вернадский, 1989]. В итоге получаем следующий перечень основополагающих монографий: «Биосфера и ноосфера» [Вернадский, 1989]; «Всеобщая организационная наука, или тектология» [Богданов, 1913–1929]; «Теория систем» [Bertalanffy, 1945]; «Кибернетика или управление и связь в животном и машине» [Wiener, 1983]; «Праксеология» [Котарбинский, 1983].

Роль, место и значение трудов А. А. Богданова, Л. фон Берталанфи, Н. Винера, Т.Котарбинского в становлении и развитии идей системности и системного мышления достаточно детально проанализированы в работе [Садовский, 1996].

Важная роль и практическое значение фундаментальных работ Л. фон Берталанфи [Bertalanffy, 1945] и Н. Винера [Wiener, 1983] как системных первоисточников в становлении и развитии системного мышления являются общепризнанными. Особо необходимо отметить, что в данных работах независимо предложена новая идея, сущность которой состоит в переходе к исследованию общих свойств, которые характерны для различных типов объектов. Л. фон Берталанфи рассматривает вопрос с позиции общности принципов построения и структурных свойств различных типов систем, а Н. Винер — с позиции общности принципов и свойств управления различными типами сложных объектов, в т.ч. субъектов живого мира и объектов техники различного назначения. И эти подходы достаточно долго развивались независимо. Вместе с тем, оба подхода имеют непосредственное отношение к системным исследованиям.

Выполненные независимо труды В.И. Вернадского, Л. фон Берталанфи и Н. Винера создали единую, идейную базу для формирования принципиально новой фундаментальной парадигмы в науке, концептуальная новизна которой заключается в переходе:

- от исследования конкретных свойств объектов определенного типа (физические, химические, биологические, экономические и т.д.) к исследованию общих свойств, которые характерны для различных типов объектов;
- от исследования свойств и особенностей процессов определенного вида к исследованию структуры, свойств, и особенностей взаимосвязей, взаимозависимости и взаимодействия разнородных процессов;
- от исследования свойств отдельных объектов определенного типа к исследованию свойств и структуры взаимосвязей, взаимозависимости и взаимодействия разнотипных объектов.

Эти признаки новизны в определенной степени позднее реализованы в форме основных принципов теории системного анализа, и поэтому приведенную парадигму можно характеризовать как «теоретическая парадигма системной методологии».

Этот период характеризуется независимым появлением разнесенных во времени публикаций философских и методологических идей, принципов, подходов. Важнейшим итогом первого этапа следует считать создание идейной базы для формирования новой фундаментальной парадигмы в науке. Основные идеи данной парадигмы, отражающие ее различные аспекты, независимо представлены в трудах *В. И. Вернадского*, а в последующем – *Л. фон Берталанфи* и *Н. Винера*.

---

---

*Второй этап* становления системного анализа формировался в чрезвычайных условиях с начала 30-х до конца 40-х годов XX века. Это был период появления практической необходимости оперативного решения реальных сложных системных задач государственного значения, создания различных технических систем военного назначения в условиях жесткого лимита времени. Разработка методологического аппарата выполнялась эмпирически и независимо в различных организациях разных стран. Обмен опытом был исключен условиями военного времени. В результате были созданы эмпирические предпосылки формирования парадигмы системного анализа как методологии решения реальных системных задач в практически допустимые сроки с практически приемлемой погрешностью в условиях концептуальной неопределенности. Поэтому данный этап можно считать этапом эмпирического формирования системной методологии.

Именно в этот период в 1948 г. появился термин «*системный анализ*» в работах корпорации RAND, которая является одной из так называемых "думающих", неприбыльных корпораций, занимающихся разработкой военных доктрин, рекомендаций для выбора проектов новых систем оружия, исследованием военного и научного потенциала противника, рынков сбыта оружия и т. п. проблемами анализа и прогнозирования развития военного потенциала США. Первой методикой системного анализа, в которой были определены порядок, методы формирования и оценки приоритетов элементов структур целей (названных в методике *деревьями целей*), была методика ПАТТЕРН (PATTERN – Planning Assistance Through Technical Evaluation from Relevans Number – помощь планированию посредством относительных показателей технической оценки) [Лопухин, 1971]. Считается, что инициатором создания методики является Ч. Дэвис, вице-президент фирмы "Хониуэлл" корпорации RAND. Назначением, конечной целью создания системы ПАТТЕРН была подготовка и реализация планов обеспечения военного превосходства США над всем миром. Перед разработчиками методики была поставлена задача – связать военные и научные планы правительства США. После того, как методика была заслушана в Конгрессе США, и сенатор Г. Хемфри выступил в 1964 г. с предложением создать на базе идеи ПАТТЕРН Бюро помощи президенту в подготовке решений научно-информационными методами (PASSIM – President Advisory Staff on Scientific Information Management), открытые публикации о развитии методики практически отсутствуют. Были лишь сообщения о вариантах методики – ПАТТЕРН-МО, НАСА-ПАТТЕРН, и др. зарубежных методиках – ПРОФИЛЕ, ППБ и т. п., [Янч, 1974].

*Третий этап* становления и развития системного анализа формировался в послевоенных условиях с середины 40-х до конца 70-х годов XX века.

Этот этап принципиально отличается от предыдущих периодов качественно новыми задачами, принципиальными социально-политическими изменениями в мире, уникальными теоретическими и практическими научно-техническими достижениями. Был создан теоретический базис математического и методологического инструментария формализации и автоматизации на базе ЭВМ процедур решения реальных сложнейших организационных и технических системных проблем в различных сферах практической деятельности. Созданы принципиально новые отрасли — космонавтика и атомная энергетика. Сложившаяся в рассматриваемый период ситуация в развитии общества характеризовалась, с одной стороны, непрерывно растущей потребностью решения практически важных прикладных системных проблем многопрофильного характера, а с другой стороны, появлением качественно новых возможностей их решения, которые предоставляла вычислительная техника. Данные обстоятельства инициировали процессы формирования и теоретического обоснования методологии системного анализа и непосредственно связанных с ним научных направлений и дисциплин: общая теория систем, системотехника, компьютерная математика, прикладная математика, имитационное моделирование систем, теория вычислительных систем, проектирование вычислительных машин, теория программирования, теория автоматической обработки цифровой информации и другие. Разработаны, созданы и введены в эксплуатацию сложные и большие уникальные технические системы различного назначения, что способствовало развитию системного анализа.

В создание теории системного анализа и системной методологии весомый вклад внесли переводы научных трудов этого периода, авторами которых являются *К. Боулдинг, Дж. Клир, М. Месарович, С. Оптнер, Т. Саати, Г. Саймон, А. Холл, У. Р. Эшби*. Важный вклад в разработку теории, в решение сложнейших, междисциплинарных технических и организационных проблем и в создание сложных и больших систем различного назначения в СССР внесли труды ведущих научных школ, основателями которых являются *Н. П. Бусленко, А. А. Вавилов, В. М. Глушков, Д. М. Гвишиани, А. А. Дородницын, А. П. Ершов, М. В. Келдыш, Г. В. Кисунько, С. П. Королев, Г. И. Марчук, А. Л. Минц, Н. Н. Моисеев* и др. Труды этих школ создан теоретический базис, математический и методологический инструментарий формализации и автоматизации на базе ЭВМ решения реальных системных проблем; практически реализована фундаментальная теоретическая парадигма системного анализа, выполнены и реализованы проекты сложных технических систем различного назначения.

Этот этап является периодом синхронного развития теории системного анализа и практики системных исследований. Одновременно к концу периода появились глобальные системные проблемы, которые были неразрешимы на основе имевшегося на то время арсенала математических и методологических средств системного анализа. Особенности и динамика

---

---

этого кризиса определялись многими причинами. Одной из главных причин явилось быстрое увеличение темпов роста сложности и масштабов реальных системных проблем, обусловленное глобализацией мировых процессов. Взаимосвязи, взаимозависимости, взаимодействия экономических, социальных, экологических и других глобальных и региональных процессов становились определяющими факторами мирового развития. В результате появился новый эффект развития, который французский экономист М. Годе четко и полно охарактеризовал короткой фразой: «будущее перестало походить на прошлое» [Godet, 1983]. В этих условиях глобальные процессы мировой системы оказались под воздействием сложно структурированного, многоуровневого, иерархического множества почти непрогнозируемых, непрерывно изменяющихся позитивных и негативных взаимосвязей, взаимозависимостей и взаимодействий.

В 1976 г. был создан Всесоюзный научно-исследовательский институт системных исследований (ВНИИСИ) Государственного комитета Совета Министров СССР по науке и технике и Академии наук СССР (директором на протяжении 17 лет был академик Д. М. Гвишиани [Гвишиани, 1987]). На Институт, как филиал Международного института прикладного системного анализа (МИПСА) при Римском Клубе, были возложены функции головной организации по научно-методическому обеспечению прогнозирования и планирования по аналогии с RAND Corporation. Основной тематикой исследований ВНИИСИ была разработка методов системных исследований по важнейшим прикладным междисциплинарным и межотраслевым проблемам, включая разработку методов управления созданием и функционированием сложных организационных и технических систем. В 1992 г. Институт получил новое название – Институт системного анализа (ИСА) РАН. Издаваемый институтом Ежегодник «Системные исследования» внес важный вклад в развитие и распространение идей теории систем и системного анализа. .

*Четвертый этап* развития системного анализа продолжается с начала 80-х годов XX века до настоящего времени и является периодом глобализации мировых процессов и угроз. Мировые процессы развития характеризуются высокими темпами глобализации и непрерывным возрастанием сложности взаимозависимости и взаимодействия. Определяющим принципом системных исследований становится глобальное видение исследуемых проблем с учетом возрастающих взаимосвязей и взаимозависимостей всех стран и народов мира. Главной целью исследований становится достижение такого системно согласованного, взаимозависимого развития всех компонентов цивилизации, при котором ни один элемент мировой системы не может расти за счет других. Для достижения этой цели необходимо сосредоточить усилия на ускорении преодоления методологического кризиса, проявившегося в конце 1970-х годов. Одна из важнейших причин этого кризиса: современная методология системного анализа не соответствует глобальной, многоуровневой, иерархической,

---

---

многодисциплинарной, многофакторной структуре современной глобальной проблематике и глобальной, многоуровневой, иерархической структуре информационных компьютерных систем и сетей, которые являются потенциальным инструментарием исследования глобальной проблематики. Одним из направлений устранения недостатка можно считать последовательную разработку концепций, стратегий и программ решения наиболее важных проблем современности. К их числу, в первую очередь следует отнести проблемы: предвидения качественных и количественных изменений в различных сферах практической деятельности [Згуровский и Панкратова, 2005], управления рисками и безопасностью сложных технических систем, техногенно и экологически опасных процессов [Панкратова и Курилин, 2000, 2001; Pankratova, 2017], развития интеллектуальных информационных технологий и сетей поддержки научных исследований, взаимоотношений Природы и Общества на основе глобального экологического мониторинга, оценивания тенденций и управления развитием мировой экологической обстановки [Pankratova, 2014] и другие. И на этой основе требуется обеспечить формирование многоуровневой, иерархической структуры метаметодологии системного анализа глобальной проблематики. Таким образом, четвертый этап является этапом глобализации системной проблематики.

В этот период в 1997 г. в Украине, г. Киев был создан Институт прикладного системного анализа НАН Украины и МОН Украины, инициатором создания и директором которого был академик НАН Украины М.З.Згуровский. Создание Института двойного подчинения было связано с воплощением концепции интеграции науки и образования с целью проведения в Украине передовых научных исследований и целевой подготовки высококвалифицированных кадров для НАН Украины и других отраслей народного хозяйства в сфере прикладного системного анализа, новейших информационных технологий и компьютерных наук. С целью повышения качества обучения, повышения образовательного и научного уровня аспирантов и молодых ученых, развития и реализации их творческого потенциала, укрепления и расширения связей между отечественными и зарубежными учебными и научными заведениями с 1998 г. проводится международная научная конференция «Системный анализ и информационные технологии» [sait.kpi.ua](http://sait.kpi.ua). С 2002 года издается международный журнал "System research and information technologies", труды которого индексируются во многих наукометрических базах. Под руководством Н.Д. Панкратовой работает научная школа «Разработка теории и методов системного анализа и их применение к решению междисциплинарных задач различной природы», которая подготовила ряд монографий, учебников, пособий по системному анализу [Zgurovsky and Панкратова., 2007; Згуровский и Панкратова, 2011; Панкратова, 2018 и др. ].

В 1994 г. В Санкт-Петербургском политехническом университете *В. Н. Козлов* переименовал возглавляемую им кафедру Технической кибернетики в кафедру «Системный анализ и

управление» и открыл одноименное направление подготовки бакалавров и магистров. С 1998 г. началось становление Научно-педагогической школы «Системный анализ в проектировании и управлении» [Козлов и Волкова, 2015]. Школа считает себя преемницей школы *Ф. Е. Темникова* (1906–1993), создавшего в 1969 г. первую кафедру Системотехники, и школы Ленинградского политехнического института (ЛПИ), в котором с 1973 г. на факультете технической кибернетики *А.А. Денисов* (1934–2010) исследовал общность процессов в системах различной физической природы и развивал теорию информационного поля и ее дискретный вариант – информационный подход к анализу систем.

Школа проводит ежегодную конференцию, объединяющая ученых, развивающих теорию систем и системный анализ в вузах и научных организациях России, Украины, Польши, Норвегии, Финляндии и др. стран, и подготовила ряд коллективных монографий и учебников по теории систем и системному анализу [Козлов, 2010; Волкова и Денисов, 2014], способствующих подготовке специалистов для решения прикладных задач с применением идей и методов теории систем и системного анализа.

---

### **Заключение**

В статье рассмотрена краткая история становления и развития теории систем и ее прикладного направления – системного анализа. Развитие теории систем во многом было инициировано развитием технологий автоматизации управления, и, в свою очередь, способствовало становлению теории проектирования информационных систем. Методология системного анализа применялась при решении междисциплинарных задач разной природы, обеспечивая в условиях концептуальной неопределенности формирование целостных, междисциплинарных знаний об исследуемом объекте как о совокупности взаимосвязанных процессов различной природы для последующего принятия решений относительно его дальнейшего развития и поведения с учетом множества конфликтующих критериев и целей, наличия факторов риска, неполноты и недостоверности информации.

Проведенный краткий анализ позволяет сделать вывод, что системный анализ как прикладная научная дисциплина позволяет с единой позиции заданных целей исследования изучать на основе системных принципов и критериев свойства объектов, свойства отношений между объектами, свойства отношений между процессами, свойства отношений между объектами и процессами, при этом изучить как отдельно любое из отношений, так и любое сочетание этих отношений. Следовательно, системный анализ принципиально отличается от аксиоматических научных дисциплин и подходов, которые изучают автономно, или только отдельные объекты или только процессы, или только отношения. Поэтому системный анализ в соответствии с

классификацией Дж. Клира [Клир, 1990] следует рассматривать как многомерную научную дисциплину.

В то же время анализ современного состояния теории систем и системного анализа позволяет осознать, что современная методология прикладного системного анализа не вполне соответствует глобальной, многоуровневой, иерархической структуре разнородных, многофакторных, многофункциональных взаимосвязей, взаимозависимостей и взаимодействий объектов исследования современной глобальной проблематики и недостаточно использует потенциальные возможности глобальной, многоуровневой, иерархической системы информационных компьютерных сетей. Одним из перспективных направлений устранения недостатков можно считать оперативную, последовательную разработку концепций, стратегий и программ исследования наиболее важных глобальных проблем современности. И на этой основе целесообразно выполнять формирование многоуровневой, иерархической структуры метаметодологии системного анализа глобальной проблематики, обеспечив ее системное, функциональное согласование с иерархической структурой взаимосвязей, взаимозависимостей и взаимодействий объектов исследования и соответственно с иерархической структурой информационных компьютерных систем и сетей как инструментальной основы ее реализации.

В настоящее время становится все более очевидным, что внедрение активно развивающихся технологий третьей и четвертой промышленных революций окажет существенное влияние и на информационные элементы, и на информационные процессы, и на развитие информационно-управляющих систем. Поэтому в соответствии с целью, поставленной во введении, целесообразно возродить первоначальное определение информатики (1), предложенное *Ф. Е. Темниковым* [Темников, 1963].

Для приближения к возвращению информатике статуса обобщающей науки об информации перспективным представляется более глубокое исследование закономерностей строения, функционирования и развития систем применительно к информационно-управляющим комплексам, применение методов и моделей теории систем и прикладного системного анализа.

---

## **Bibliography**

---

- [Bertalanffy, 1960] Bertalanffy L. von. General System Theory. Foundations, Development, Applications. N.Y. Braziler 1971. (Изд.1.-е, ФРГ, 1945). Берталанфи Л. фон. Общая теория систем. Изд.2-е. М.: Мир, 1960.
- [Godet, 1983] Godet M. Reducing the Blunders in Forecasting. Futures, 15, №3, 1983. P. 181–192.
- [Kotarbinski, 1983] Kotarbinski T. Praxiology. An Introduction to the Sciences of Efficient Action. Oxford, 1965. Котарбинский Т. Праксеология. М.: Экономика, 1983.

- [Pankratova et al., 2014] Pankratova N.D., Bidyuk P.I., Selin Y. M., Savchenko I.O., Malafeeva L.Y., Makukha M.P., Savastiyarov V.V. Foresight and Forecast for Prevention, Mitigation and Recovering after Social, Technical and Environmental Disasters. Improving Disasters Resilience and Mitigation – IT Means and Tools. Springer, 2014. P.119-134.
- [Pankratova, 2017] Pankratova N.D. The integrated system of safety and survivability complex technical objects operation in conditions of uncertainty and multifactor risks. Proceedings of conference IEEE (№50). May 29 – June 2, 2017, Kyiv, Ukraine. P. 1135-1140.
- [Wiener, 1983] Wiener N, Cybernetics.M. I. T. Press, Cambridge, Massachusetts, 1948. Винер Н. Кибернетика или управление и связь в животном и машине. М.: Наука, 1983.
- [Zgurovsky and Панкратова, 2007] Zgurovsky M.Z., Pankratova N.D. System analysis: Theory and Applications. Springer.2007.
- [Богданов, 1913-1929] Богданов А.А. Всеобщая организационная наука, или тектология. Т. 1–3, М.: 1913-1929.
- [Вернадский, 1989] Вернадский, В. И. Биосфера и ноосфера. М : Наука, 1989.
- [Волкова и Денисов, 2014] Волкова В.Н.. Денисов А.А. Теория систем и системный анализ: учебник для академического бакалавра. 2-е издание, перераб. и дополненное. М.: Изд-во Юрайт, 2014.
- [Волкова и Черный, 2016] Волкова В. Н., Черный Ю. Ю. Вклад Ф.Е. Темникова в развитие информатики. Прикладная информатика. Т. 11. № 5 (65), 2016. С. 122-143.
- [Волкова, 2001] Волкова В. Н. Из истории развития системного анализа в нашей стране . Экономическая наука современной России. М.: ЦЭМИ. № 2, с. 138-152, № 3, 2001 с.127-138.
- [Гвишиани, 1987] Гвишиани Д.М. Международный институт прикладного системного анализа: цели, основные перспективы. Системные исследования. Методологические проблемы. Ежегодник 1987. М.: Наука, 1987. С. 7–25.
- [Згуровский и Панкратова, 2005] Згуровский М.З., Панкратова Н.Д. Технологическое предвидение. Киев: Изд-во Политехника. 2005.
- [Згуровский и Панкратова, 2011] Згуровский М.З., Панкратова Н.Д. Системный анализ: проблемы, методология, приложения. Киев: Изд-во Наук.думка, 2011.
- [Клир, 1990] Клир Дж. Системология. Автоматизация решения системных задач. М.: Радио и связь, 1990.
- [Козлов и Волкова, 2015 ] Козлов В. Н. Волкова В. Н. Научно-педагогическая школа «Системный анализ в проектировании и управлении». Системный анализ в проектировании и управлении:

- Сб. науч. трудов XIX Междунар. науч.-практич. конф. Ч. 1. СПб.: Изд-во Политехн. ун-та, 2015. 488 с. С. 5–16.
- [Козлов, 2010] Козлов В. Н. Системный анализ, оптимизация и принятие решений М.: Проспект, 2010.
- [Лопухин, 1971] Лопухин М. М. ПАТТЕРН – метод планирования и прогнозирования научных работ. М.: Сов. радио, 1971.
- [Михайлов и др., 1968] Михайлов А.И., Черный А.И., Гиляревский Р.С. Основы информатики. М.: Наука, 1968.
- [Панкратова и Курилин, 2000] Панкратова Н.Д., Курилин Б.И. Концептуальные основы системного анализа рисков в динамике управления безопасностью сложных систем. Ч. 1. Основные утверждения и обоснования подхода. Проблемы управления и информатики. № 6, 2000. С. 110–132.
- [Панкратова и Курилин, 2001] Панкратова Н.Д., Курилин Б.И. Концептуальные основы системного анализа рисков в динамике управления безопасностью сложных систем. Ч. 2. Общая задача системного анализа рисков и стратегия ее решения. Проблемы управления и информатики. № 2, 2001. С. 108–126.
- [Панкратова, 1999] Панкратова Н.Д. Общие тенденции и системные проблемы развития информационных технологий. Проблемы управления и информатики. № 1, 1999. С. 58–68.
- [Панкратова, 2002 ] Панкратова Н.Д. Становление и развитие системного анализа как прикладной научной дисциплины. System research&Information technologies. №1, 2002. С.65—94.
- [Панкратова, 2018] Панкратова Н.Д. Системний аналіз. Теорія та застосування. Київ: Вид-во Наук.думка, 2018.
- [Садовский и др., 1969] Исследования по общей теории систем: Сб. переводов / Общ. ред. и вступит. статья В. Н. Садовского и Э. Г. Юдина. М.: Прогресс, 1969.
- [Садовский, 1996] Садовский В.Н. Смена парадигм системного мышления // Системные исследования. Методологические проблемы. Ежегодник 1992–1994. М.: Эдиториал УРСС. 1996. С. 64–78.
- [Темников, 1963] Темников Ф.Е. Информатика. Известия вузов. Электромеханика. № 11 М., 1963.
- [Янч, 1974] Янч Э. Прогнозирование научно-технического прогресса. М.: Прогресс, 1974.

---

---

## Authors' Information

---



**Nataliya D. Pankratova** – DTs, Professor, Corresponding member of NASU, Depute director of Institute for applied system analysis, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Av. Pobedy 37, Kiev 03056, Ukraine;  
e-mail: natalidmp@gmail.com

**Major Fields of Scientific Research:** System analysis, Theory of risk, Decision support systems, Applied mathematics, Applied mechanics, Foresight, Scenarios, Strategic planning, Information technology



**Violetta N. Volkova** –DEc, Professor, Peter the Great St. Petersburg Polytechnic University, Polytechnicheskaya str., 29, St. Petersburg, Russia. 195251.  
e-mail: violetta\_volkova@list.ru

**Major Fields of Scientific Research:** Control in technical systems, Informatics. Information technology, Information systems, Information theory, Management in social and economic systems, System analysis, Strategic planning, Technical cybernetics, Theory of Systems

## METHODOLOGY OF SYSTEMS AND SYSTEM ANALYSIS

### AS THE BASIS OF INFORMATICS DEVELOPMENT

**Nataliya D. Pankratova, Violetta N. Volkova**

**Abstract:** The history of origin of interdisciplinary scientific directions is considered. The main results of understanding the stages of development and the state of the theory of systems and system analysis in the two scientific schools of system analysis are presented. The authors of this article take part in their development. The idea of the revival of informatics as a general science of information is proposed.

**Keywords:** regularities, informatics, information system, information-control complex, system classifications, method, model, system analysis, systems theory.

**ACM Classification Keywords:** H.4.2. Information System Application: type of system strategy

## О ТЕОРЕТИЧЕСКИХ ОСНОВАХ МОДЕЛИ CVP-АНАЛИЗА

В.П.Галушко, Т.Н. Болотова, В.И Макаров, В.Ю.Величко

**Аннотация.** Построена последовательная теория линейной модели CVP-анализа, которая дополнила существующую теорию принципиально новыми результатами, обратив внимание на ошибочную формулу для определения запаса прочности прибыли (*margin of safety*), которая является базовой и растиражирована во всей многочисленной научной и учебной литературе, посвященной модели CVP анализа. Дана классификация двух типов особых точек, в которых прибыль обращается в нуль. Первый тип особых точек является критическим объемом производства. Второй тип особых точек является точками окупаемости общих затрат, которые зависят от объема произведенной продукции и определяют часть произведенного объема продукции, реализация которой компенсирует общие затраты всего объема произведенной продукции. Предложены аналитические критерии принятия управленческих решений для планирования и организации производства с учетом рисков, обусловленных конкуренцией на рынках реализации продукции.

**Ключевые слова:** модель CVP анализа, критический объем производства, точка окупаемости общих затрат производства, прибыль, запас прочности по прибыли, аналитические критерии принятия решений.

**ITHEA ключевые слова:** J. Computer Applications: J.4 SOCIAL AND BEHAVIORAL SCIENCES

---

### Вступление

---

Экономический анализ коммерческих предприятий в основном базируется на модели, изучающей соотношение между тремя группами экономических показателей: «затраты – объем произведенного товара – прибыль». Сокращенное название – модель CVP анализа (Cost-Volume-Profit Analysis). Использованию модели CVP-анализа для организации и планирования бизнеса посвящено большое количество научной и учебной литературы [1-6].

Модель CVP-анализа определяет следующие, основные объективно обусловленные оценки (параметры) экономики предпринимательской деятельности: критический объем производства (break even point)  $Q = Q_c$ , прибыль (profit)  $P(Q)$  и запас прочности по прибыли (*margin of safety*)  $MS(Q)$ .

Аналитические выражения для этих объективно обусловленных оценок предполагают определенный вид зависимости функций выручки  $R(Q)$  и общих затрат  $C_{TC}(Q)$  в диапазоне деловой активности предприятия с заданным технологическим процессом, объемом производства и спросом на продукцию.

В цитированной литературе [1-6] рассмотрен простой случай. Функция выручки  $R(Q_{RL})$  линейно зависит от **реализованного** объема продукции  $Q_{RL}$ :

$$R(Q_{RL}) = P_R Q_{RL}, \quad (1)$$

где  $P_R$  - цена реализации единицы продукции.

Функция общих затрат  $C_{TC}(Q_{PR})$  линейно зависит от **произведенного** объема продукции  $Q_{PR}$ :

$$C_{TC}(Q_{PR}) = C_V(Q_{PR}) + C_{FC} = S_V Q_{PR} + C_{FC}, \quad (2)$$

где  $S_V$  – себестоимость единицы продукции, отнесенная на условно–переменные затраты;

$C_{FC}$  – фиксированные, условно–постоянные затраты, которые не зависят от количества произведенного товара  $Q_{PR}$ .

Используя формулы (1) и (2), объективно обусловленные оценки экономической эффективности определяются следующим образом [1-6]:

$$\text{критический объем производства} \quad Q_C = C_{FC} / (P_r - S_V); \quad (3)$$

$$\text{прибыль производства} \quad P(Q) = (R(Q) - C_{TC}(Q)); \quad (4)$$

$$\text{запас прочности прибыли (margin of safety)} \quad MS(Q) = (Q - Q_C) / Q. \quad (5)$$

**Цель статьи.** Дополнить существующую теорию CVP-анализа классификацией особых точек функции прибыли в линейной модели CVP анализа. Обратить внимание на ошибочное определение запаса прочности (надежности, margin of safety – формула (5)) [1-6] и продемонстрировать на конкретном примере практическое использование полученных новых результатов как аналитических критериев для принятия правильных управленческих решений по тактике и стратегии бизнеса.

---

### Линейная модель CVP-анализа

Рассмотрим пример анализа экономических показателей фермерского хозяйства площадью 500 га. Хозяйство за последние 5 лет имеет следующие показатели: средняя урожайность озимой пшеницы 3,2 т/га, объем полученной продукции 1600 т., условно-постоянные затраты составляют 1600 тыс.грн., себестоимость продукции по условно-переменным затратам  $S_V = 2000$

грн./т. Финансово-экономические показатели, которые были использованные в модели CVP анализа, приведены в табл.1 и их геометрическая интерпретация представлена на рис.1-2.

**Таблица 1. Финансово-экономические показатели производства озимой пшеницы фермерским хозяйством**

Наименование показателей	Обозначения	Значения показателей		
Объем реализации	$Q$ (т)	640	1066,7	1600
Цена реализации	$P_R$ (грн./т)	4500	4500	4500
Себестоимость продукции по условно-переменным затратам	$S_V$ (грн./т)	2000	2000	2000
Маржинальный доход	$M_P(Q) = R(Q) - S_V Q$ (т. грн.)	1600	2666.75	4000
Условно-постоянные затраты	$C_{FC}$ (т. грн.)	1600	1600	1600
<b>Критический объем производства</b>	$Q_C = C_{FC} / (P_R - S_V)$ (т)	640	640	640
<b>Объем окупаемости общих затрат</b>	$Q_{BEP} = Q_C + S_V(Q - Q_C) / P_R$ (т)	640	829,6	1066,7
Выручка от реализации	$R(Q) = P_R Q$ (тыс. грн.)	<b>2880</b>	<b>4800</b>	7200
Условно-переменные затраты	$C_V(Q) = S_V Q$ (тыс. грн.)	1280	2133.3	3200
Общие затраты на продукцию	$C_{TC}(Q) = C_V(Q) + C_{FC}$ (тыс.грн)	<b>2880</b>	3733.3	<b>4800</b>
Средняя себестоимость единицы произведенной продукции	$S_{AV}(Q) = C_{TC}(Q) / Q$ (грн./т)	4500	3499,9	3000
Функция прибыли	$P(Q) = R(Q) - C_{TC}(Q)$ (тыс.грн)	0	1066,75	2400

Функция прибыли в рамках линейной модели CVP-анализа обращается в нуль при реализации определенных объемов произведенной продукции в двух типах особых точек.

Первый тип особых точек - это точки критического объема производства (break-even point) известные ранее [1-6].

Второй тип особых точек определяют часть произведенного объема продукции, реализация которого компенсирует общие затраты всего произведенного объема товарной продукции и является точками окупаемости общих затрат производства.

Этот тип особых точек обусловлен пересечением двух зависимостей: функции выручки  $R(Q)$  и функции общих затрат  $C_{TC}(Q) = C_{TC}(Q_I)$  при некотором фиксированном значении  $Q = Q_I$ , которое зависит от объема произведенной продукции  $Q$ .

Точка пересечения этих двух зависимостей  $Q_{BEP}(Q_I)$  определяется из решения уравнения:

$$R(Q_{BEP}(Q_I)) = C_{TC}(Q_I) \quad (6)$$

Это новое уравнение в теории модели CVP анализа ранее не рассматривалось.

Финансовой основой экономики коммерческой деятельности предприятия в модели CVP-анализа является маржинальная прибыль  $M_p(Q)$  – разность между выручкой  $R(Q)$  и условно-переменными затратами  $C_v(Q)$ :

$$M_p(Q) = R(Q) - C_v(Q) \quad (7)$$

Если выполняется неравенство  $M_p(Q) > 0$ , то зависимости функций выручки  $R(Q)$  и общих затрат  $C_{TC}(Q)$  пересекаются в некоторой точке  $A = Q_C$  (рис.1).

В этой точке выполняются три равенства:

$$R(Q_C) = C_{TC}(Q_C) \quad (8)$$

$$M_p(Q_C) = C_{FC} \quad (9)$$

$$P_R = S_{AV}(Q_C) \quad (10)$$

где величина  $S_{AV}(Q) = C_{TC}(Q)/Q$  – средняя себестоимость единицы произведенной продукции объема  $Q$ .

Точка  $A = Q_C$  имеет простую экономическую интерпретацию – маржинальная прибыль  $M_p(Q_C)$ , полученная при реализации продукции  $Q = Q_C$ , **компенсирует** условно-постоянные затраты производителя  $C_{FC}$ .

В этом случае при реализации продукции объема  $Q_C$  выручка  $R(Q_C)$  компенсирует общие затраты производства этого объема продукции  $C_{TC}(Q_C)$  и средняя себестоимость единицы продукции объема  $Q = Q_C$  равна цене реализации  $P_R$  (формула 10).

Величина  $Q_C$  определяется из уравнений (8-10) с учетом формул (1,2) и табл.1:

$$Q_C = C_{FC} / (P_r - S_v) \quad (11)$$

Из равенства 11 следует, что объем продукции  $Q_C$  зависит только от величины условно-постоянных затрат  $C_{FC}$ , от величины себестоимости единицы продукции  $S_v$  по условно-переменным затратам технологии производства товарной продукции и от цены реализации  $P_R$  данного вида продукции.

Следовательно, точка  $Q_C$  является интегральной характеристикой системы "сфера производства - сфера потребления".

На рис.1. изображены следующие элементы: прямая  $R(Q)$  – зависимость выручки от объема производства  $Q$ ; прямая  $C_{TC}(Q)$  – зависимость общих затрат от объема производства  $Q$ ; прямая  $C_{FC}$  – условно-постоянные затраты; прямая  $C_{TC}(Q) = 4800$  грн. – общие затраты объема производства  $Q = 1600$  т; прямая  $C_{TC}(Q) = 3733,3$  грн. – общие затраты объема производства  $Q = 1066,7$  т; прямая  $OF$  – зависимость  $S_{AV}(Q=1066,7)Q$ ; прямая  $OE$  – зависимость  $S_{AV}(Q=1600)Q$ ; точки  $B$  и  $C$  – точки окупаемости затрат  $Q_{BEP} = 829,6$  т и  $Q_{BEP} = 1066,7$  т, соответствующие объемам производства  $Q = 1066,7$  т и  $Q = 1600$  т; точка  $A$  – точка критического объема производства  $Q_C = 640$  т,  $MP1(Q) = 426,7$  т – область маржинальной прибыли и  $MS1(Q) = 237,1$  т – запас прочности прибыли при  $Q = 1066,7$  т,  $MP2(Q) = 960$  т – область маржинальной прибыли и  $MS2(Q) = 533,3$  т – запас прочности прибыли при  $Q = 1600$  т.

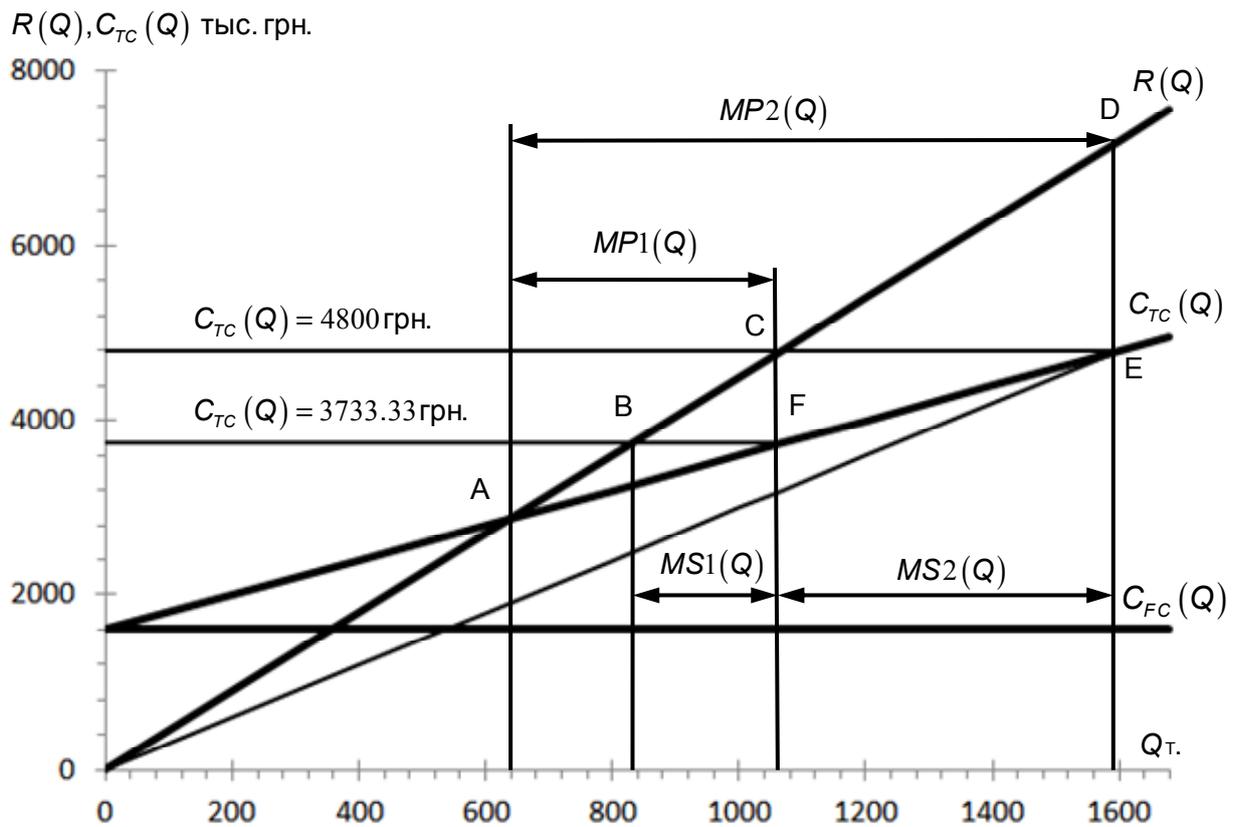


Рис. 1. Зависимость выручки  $R(Q)$  и общих затрат  $C_{TC}(Q)$  от объема производства  $Q$

Из графиков рис.1 видно, что точка  $A = Q_C$  делит область, заключенную между зависимостями функций  $R(Q)$  и  $C_{TC}(Q)$ , на две части: на область, где разность этих функций принимает отрицательные значения при реализации объема товара равного  $Q < Q_C$ , и на область, где разность этих функций принимает положительные значения при реализации объема товара  $Q > Q_C$ .

Тогда функцию прибыли  $P(Q)$  можно представить в виде:

$$P(Q) = [R(Q) - C_{TC}(Q)]\theta(Q - Q_C), \quad (12)$$

где тета-функция  $\theta(x) = 1$  при значениях  $x \geq 0$  и  $\theta(x) = 0$  при  $x < 0$ .

Значения функции прибыли  $P(Q)$  ограничены площадью (рис.1, точки A, D, E), которая характеризуется двумя размерами: вертикальным и горизонтальным.

Вертикальный размер определяют отрезки прямых параллельных оси ординат, величина которых есть разность между выручкой  $R(Q_i)$  и общими затратами  $C_{TC}(Q_i)$  (рис.1, точки C, F и

D, E). Точка  $Q = Q_i$  произвольная точка, которая больше  $Q_C$ . Экономическая интерпретация этих геометрических отрезков – это величина прибыли:  $P(Q_i) = R(Q_i) - C_{TC}(Q_i)$ .

Горизонтальный размер определяют отрезки прямых параллельных оси абсцисс  $C_{TC}(Q) = C_{TC}(Q_i)$ , заключенные между точками  $Q_{BEP}(Q_i)$  и  $Q_i$  (рис. 1, точки В, F и точки С, Е). Экономическая интерпретация этих геометрических отрезков следующая.

Величины  $Q_{BEP}(Q_i)$ , расположенные на зависимости  $R(Q)$ , являются корнями уравнения (6). Эти точки делят прямые параллельные оси абсцисс  $Q$  в интервале  $[0 - Q_i]$  на две части: на отрезки прямых в интервале  $[0 - Q_{BEP}(Q_i)]$  и на отрезки прямых в интервале  $[Q_{BEP}(Q_i) - Q_i]$  (рис.1,  $Q_{BEP}(Q = 829,6 \text{ т}) = В$  и  $Q_{BEP}(Q = 1066,7 \text{ т}) = С$ ).

Величины  $Q_{BEP}(Q_i)$  имеют простую экономическую интерпретацию. Выручка  $R(Q_{BEP}(Q_i))$  компенсирует полные затраты  $C_{TC}(Q_i)$  на производство продукции в объеме  $Q_i$  (табл.1). Величина  $Q_{BEP}(Q_i)$  является частью общего произведенного объема  $Q_i$ .

Согласно экономической интерпретации, точка  $Q_{BEP}(Q_i)$  – это точка окупаемости затрат (издержек) на производство продукции объема  $Q_i$ , которая определяется из уравнения 6:

$$Q_{BEP}(Q_i) = \left( Q_C + \frac{S_V}{P_r} (Q_i - Q_C) \right) \theta(Q_i - Q_C) \quad (13)$$

Из выражения (13) следует, что функция  $Q_{BEP}(Q)$  является неаналитической функцией в точке  $Q_i = Q_C$  типа  $f(x) = \theta(x)$ . В этой точке функция  $Q_{BEP}(Q)$  имеет скачок равный  $Q_{BEP}(Q_C) = Q_C$  (рис.2. кривая 1).

Такие изменения в функциональной зависимости  $Q_{BEP}(Q)$  можно интерпретировать как индикатор своеобразного «фазового» перехода в свойствах экономической системы в точке  $Q = Q_C$ , которую определим как точку критического объема производства для данного типа «фазового» перехода.

Действительно, при дальнейшем увеличении объема производства  $Q$  в области  $Q > Q_C$  возникает новая экономическая характеристика – точка окупаемости затрат (издержек)  $Q_{BEP}(Q)$ , которая зависит от объема произведенной продукции  $Q$ .

Функцию разности  $F(Q) = Q - Q_{BEP}(Q)$ , определяющая объем продукции, реализация которого приносит прибыль, по аналогии с фазовыми переходами второго рода в физических свойствах

(магнитных, сверхпроводящих), можно рассматривать как параметр порядка фазового перехода в экономической системе (рис.2, прямая 2).

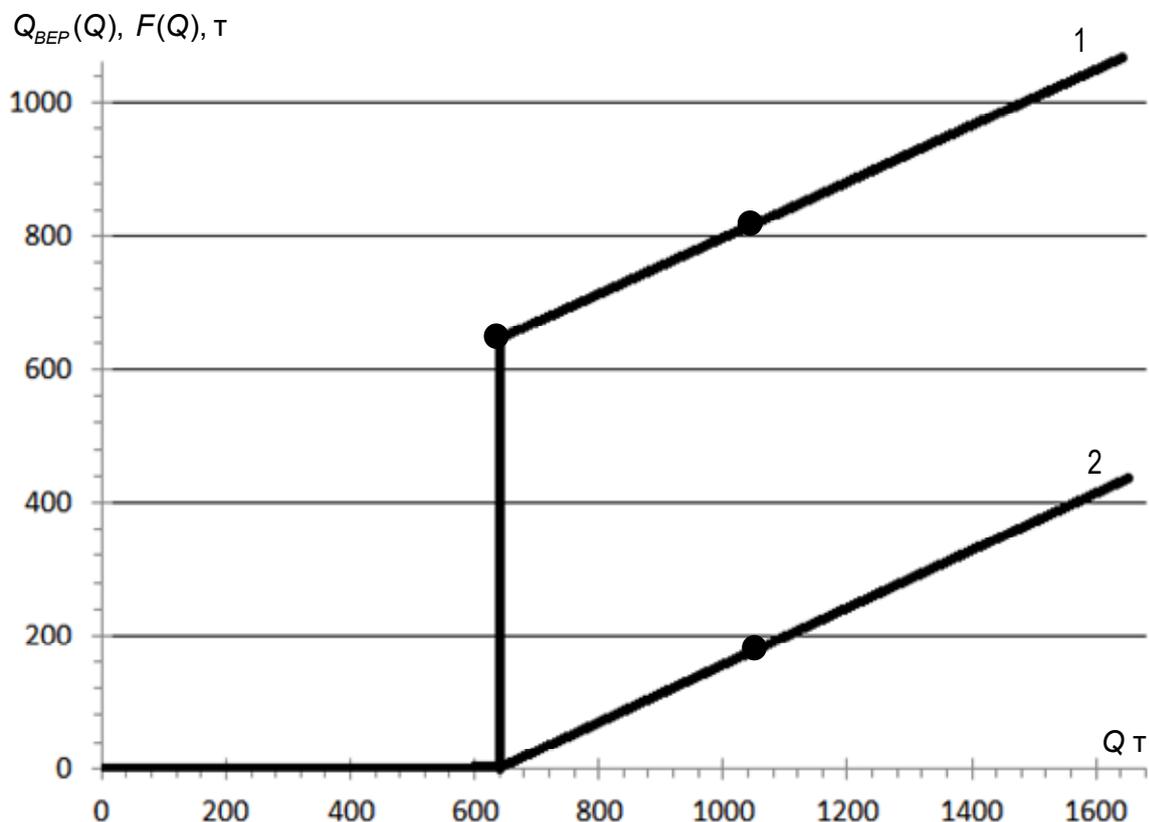


Рис.2. Зависимости точки окупаемости затрат  $Q_{BEP}(Q)$  (кривая 1) и части произведенного объема  $F(Q)$  (прямая 2), реализация которой приносит прибыль, от объема производства  $Q$ .

Функции  $F(Q)$  имеет вид:

$$F(Q) = \frac{(P_R - S_V)}{P_R} (Q - Q_C) \theta(Q - Q_C) \quad (14)$$

Таким образом, в модели CVP-анализа есть два типа объективно обусловленных оценок (точек), в которых функция прибыли обращается в нуль. По-видимому, впервые, на этот факт обратили внимание авторы работ [7,8]. В этих работах не было получено уравнения зависимости точки окупаемости общих затрат от объема произведенной продукции.

В модели CVP-анализа [1-6] рассматривалась только особая точка, определяемая уравнением (11). Поэтому строгость в определении названия такой точки была не обязательна: точка безубыточности, точка нулевой рентабельности, предел рентабельности и др. Если есть два типа особых точек, которые определяются разными аналитическими выражениями, то

необходимо их классифицировать, в соответствии с экономико-математической интерпретацией аналитических выражений (11) и (13).

Первый тип особой точки – это точка критического объема производства, в которой функция прибыли обращается в нуль, маржинальная прибыль компенсирует условно-постоянные затраты и скачком возникает зависимость  $Q_{BEP}(Q)$ .

Второй тип особых точек, в которых функция прибыль обращается в нуль, – это особые точки, зависящие от объема произведенной продукции и определяют ту часть продукции, реализация которой компенсирует затраты всего объема производства – это точки окупаемости затрат (издержек) общих затрат, расположенные на кривой  $R(Q)$ .

Найдем зависимость функции прибыли  $P(Q)$  от объема произведенной продукции. На основе данных приведенных в табл.1 и их геометрической интерпретации (рис.1), функцию  $P(Q)$  можно вычислять с помощью трех аналитических выражений:

$$P(Q) = (P_R - S_{AV}(Q))\theta(P - S_{AV}(Q)) \quad (15)$$

$$P(Q) = P_R(Q - Q_{BEP}(Q))\theta(Q - Q_{BEP}(Q)) \quad (16)$$

$$P(Q) = P_R F(Q) = (P_R - S_V)(Q - Q_C)\theta(Q - Q_C) = M_P(Q) - M_P(Q_C). \quad (17)$$

Функция прибыли  $P(Q)$  обращается в нуль в точках  $Q = Q_C$  и  $Q = Q_{BEP}(Q)$ . В этих точках функция прибыли  $P(Q)$  является неаналитической функцией типа  $f(x) = x\theta(x)$  со скачком производной  $\partial P(Q)/\partial Q$ :

$$\partial P(Q) / \partial Q \Big|_{Q=Q_{BEP}(Q)} = P_R \theta(Q - Q_{BEP}(Q)) \quad (18)$$

$$\partial P(Q) / \partial Q \Big|_{Q=Q_C} = (P_R - S_V) \theta(Q - Q_C) \quad (19)$$

С экономической точки зрения величина скачков равная  $P_R$  или  $(P_R - S_V)$  – это прибыль или маржинальная прибыль единицы реализованного товара.

---

### Критерии принятия управленческих решений планирования производства

---

При принятии управленческих решений о стратегии и тактики планирования и организации производства необходимо учитывать риски, обусловленные конкуренцией на рынках производства и реализации продукции. Поэтому производителю необходимо знать, какие изменения допустимы в цене и объемах реализации продукции, чтобы предприятие при реализации продукции было бы прибыльным (неубыточным).

Запас прочности (надежности) прибыли  $MS(Q)$ , естественно, определять как отношение прибыли  $P(Q)$  к выручке  $R(Q)$ :

$$MS(Q) = P(Q) / R(Q). \quad (20)$$

Выражение (20), используя выражения (15-17), позволяет определить как допустимые изменения в цене, так и в объемах реализации продукции. Запишем выражение (20) в трех эквивалентных вариантах, которые зависят от разных экономических показателей:

$$MS(Q) = P(Q) / R(Q) = (P_R - S_{AV}(Q)) / P_R, \quad (21)$$

$$MS(Q) = P(Q) / R(Q) = (Q - Q_{BEP}(Q)) / Q, \quad (22)$$

$$MS(Q) = P(Q) / R(Q) = ((P_R - S_V) / P_R)(Q - Q_C) / Q. \quad (23)$$

Формулы (21-23) имеют простой экономический смысл: реализация продукции со средней ценой меньшей, чем  $S_{AV}(Q)$  убыточна и предприниматель будет иметь прибыль при реализации продукции  $Q > Q_{BEP}$ .

Согласно данным (табл.1), критический объем производства  $Q_C = 640$  т, точки окупаемости затрат  $Q_{BEP}(Q) = 829,6$  т и  $Q_{BEP}(Q) = 1066,7$  т при объеме продукции 1066,7 т и 1600 т соответственно. Из данных, приведенных в табл.1, видно, что выручка от реализации объема продукции 1066,7 т равна общим затратам объема продукции 1600 т. Цены  $S_{AV}(1600) = 3000$  грн./т,  $S_{AV}(1066.7) = 3500$  грн./т,  $P_R = 4500$  грн./т и  $S_V = 2000$  грн./т, позволяют определить запас прочности прибыли в зависимости от объема производства  $Q$ :  $MS1(Q) = 22.2\%$  при  $Q = 1066.7$  т,  $MS2(Q) = 33.3\%$  при  $Q = 1600$  т. ((формулы (21-23) и рис.1)).

Зона безопасности  $(Q - Q_C) / Q$ , которая используется в литературе [1-5] для оценки области риска, с экономической точки зрения является областью маржинальной прибыли, которая в  $P_R / (P_R - S_V)$  раз больше области запаса прочности прибыли  $MS(Q)$ :

$$(Q - Q_C) / Q = (P_R / (P_R - S_V)) MS(Q) \quad (24)$$

В литературе [1-6] величина  $(Q - Q_C) / Q$  является базовой для определения области риска бизнеса. Согласно данным таб.1 величина  $MP(Q)$  равна 60%, которая в 1.8 раз больше, чем реальный запас прочности прибыли  $MS2(Q) = 33.3\%$  (рис.1).

Ошибка в определении запаса прочности прибыли (области риска)  $MS(Q)$  в модели CVP анализа является типичной и растиражирована в многочисленной научной и учебной литературе [1-6].

Например, согласно презентации [4] (табл.2), критический объем продажи товара  $Q_C = 400$  шт., прибыль  $P(Q) = \$20\,000$  и запас прочности прибыли  $MS(Q) = 20\%$ , определенный по формуле (5).

**Таблица 2. Финансово-экономические показатели фирмы по продаже товара**

Наименование показателей	Список обозначений	Параметры
Объем реализации	$Q$ (шт.)	500
Цена реализации	$P_R$ (\$/шт.)	500
Себестоимость продукции по условно-переменным затратам	$S_V$ (\$/шт.)	300
Выручка от реализации	$R(Q) = P_R Q$ (\$)	250 000
Общие затраты объема $Q$	$C_{TC}(Q) = C_V(Q) + C_{FC}$ (\$)	230 000
Условно-переменные затраты	$C_V(Q) = S_V Q$ (\$)	150 000
Условно-постоянные затраты	$C_{FC}$ (\$)	80 000
Функция прибыль	$P(Q) = R(Q) - C_{TC}(Q)$ (\$)	20 000

Согласно данным табл.2, средняя себестоимость продукции  $S_{AV}(Q)$  объема 500 шт. (формула (10)) равна  $\$230\,000/500 = \$460/\text{шт.}$ , точка окупаемости (формула (13)) равна  $Q_{BEF}(Q) = 460$  шт. Запас прочности прибыли  $MS(Q)$  (формулы (21-23)), равен 8%, т.е. в 2,5 раза меньше величины, определенной в [4].

Таким образом, в линейной модели CVP анализа аналитическими критериями для принятия управленческих решений о стратегии и тактики бизнеса являются критический объем производства  $Q_C$  (формула (8)), точка окупаемости общих затрат  $Q_{BEF}(Q)$  (формула (13)),

прибыль  $P(Q)$  (формулы (15-17), средняя себестоимость единицы продукции  $S_{AV}(Q)$  (формула (10)) и запас прочности прибыли  $MS(Q)$  (формулы (21-23)).

## Выводы

Предложена классификация особых точек  $Q = Q_C$  и  $Q = Q_{БЕР}(Q_i)$  функции прибыли, которая при определенных объемах производства и реализации в линейной модели CVP анализа обращается в нуль. Первый тип особых точек является критическим объемом производства, величина которого определяется величиной маржинальной прибыли, которая компенсирует условно – постоянные затраты. Второй тип особых точек является точками окупаемости общих затрат, которые зависят от объема произведенной продукции и определяют часть произведенного объема продукции, реализация которого компенсирует общие затраты всего объема производства.

Предложен аналитический критерий для принятия управленческих решений о стратегии и тактики управления бизнесом: запас прочности прибыли  $MS(Q)$ , зависящий от относительной разности  $(Q - Q_{БЕР}(Q))/Q$ .

Показано, что при принятии управленческих решений необходимо учитывать риски на основе анализа запаса прочности прибыли  $MS(Q)$ , а не на анализе области маржинальной прибыли  $(Q - Q_C)/Q$ , как это рекомендуется в литературе [1-6].

## Литература

1. Г.В. Савицкая. Экономический анализ, Москва, ООО Новое знание, 2005 г. 650 с.
2. Irfanullah Jan. Cost Volume Profit Analysis. AccountingExplained.com [online] <https://accountingexplained.com/managerial/cvp-analysis/>
3. Irfanullah Jan. Margin of Safety (MOS). AccountingExplained.com [online] <https://accountingexplained.com/managerial/cvp-analysis/margin-of-safety>
4. CVP Analysis, ACCY 121 VIDEOS AND HANDOUTS, California State University, Sacramento, p. 1-28 [online] [http://www.csus.edu/indiv/p/pforsichh/documents/cvpanalysisvideoslides\\_000.pdf](http://www.csus.edu/indiv/p/pforsichh/documents/cvpanalysisvideoslides_000.pdf)
5. CHAPTER 3 Cost–Volume–Profit Analysis. Pearson Education Canada. p.57-96 [online] <http://www.pearsoncanada.ca/media/highered-showcase/multi-product-showcase/horngren-ch03.pdf>
6. Хорин А.Н. CVP-анализ: расчет маржинального дохода и точки безубыточности. Центр дистанционного образования «Элитариум» Санкт-Петербург [online] <http://www.elitarium.ru/cvp->

analiz-dohod-tochka-bezubytochnosti-zatraty-cena-vyruchka-pribyl-proizvodstvo-prodazhi-strategiya-predpriyatje/

7. В.П. Галушко Точка беззбитковості: теоретичні погляди на методологію розрахунків, Економіка АПК, 2006, №10 с. 3-7.
8. Т.Н. Болотова. Точки безубыточности в экономике растениеводства, Вістник ХНАУ, №9, 2009, с. 204-213.

---

### Информация об авторах

---



**Галушко Валерий Павлович** – доктор экономических наук, профессор, член-корреспондент НААН, заведующий кафедрой административного менеджмента и внешнеэкономической деятельности, Национальный университет биоресурсов и природоиспользования Украины, 03041, Украина, Киев, улица Героев Обороны, 11, учебный корпус №10, к. 302.

Основные области научных исследований: маркетинг рынка сельскохозяйственной продукции, стратегии развития сельского хозяйства Украины.



**Болотова Татьяна Николаевна** – кандидат экономических наук, доцент, доцент кафедры экономики и предпринимательства, Харьковский национальный автомобильно-дорожный университет, 61000, Украина, г. Харьков, ул. Ярослава Мудрого, 25, e-mail: [tatabolotova1975@gmail.com](mailto:tatabolotova1975@gmail.com)

Основные области научных исследований: экономика предприятий АПК и технологий растениеводства, математические модели в экономике.



**Макаров Владимир Иванович** – доктор физико-математических наук, профессор. E-mail: [v.i.makarov37@gmail.com](mailto:v.i.makarov37@gmail.com)

Основные области научных исследований: физика твердого тела, теоретическая физика, автоматизированные системы управления технологическими процессами, экономика предприятий АПК и технологий растениеводства, математические модели в экономике.



**Величко Виталий Юрьевич** – к.т.н., старший научный сотрудник, Институт кибернетики им. В. М. Глушкова НАН Украины, Украина, г. Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: [velychko@aduis.com.ua](mailto:velychko@aduis.com.ua)

Основные области научных исследований: индуктивный логический вывод, обработка естественно-языковых текстов

## THEORETICAL BASES OF THE MODEL OF CVP-ANALYSIS

V.P. Galushko, T.N. Bolotova, V.I. Makarov, V.Yu.Velychko

**Annotation.** *A consistent theory of the linear model of CVP analysis has been built. Proposed theory supplemented the existing theory with fundamentally new results, drawing attention to the erroneous formula for the margin of safety, which is basic and replicated in all the numerous scientific and educational literature on the CVP model. Classification of two types of singular points in which the profit is zero is given. The first type of singular points is the break even point. The second type of singular points are recoupment of total costs points, which depend on the volume of output and determine the part of the output produced, the realization of which compensates for the total costs of the entire volume of output. Analytical criteria for making management decisions for the planning and organization of production, taking into account the risks caused by competition in the markets for the sale of products, are proposed.*

**Keywords:** *CVP analysis model, break even point, the margin of safety, payback period of total production costs, profit, margin of profit, analytical criteria for decision-making.*

## ИДЕНТИФИКАЦИЯ СМЫСЛОВОЙ БЛИЗОСТИ ФРАГМЕНТОВ ТЕКСТОВ НАУКОМЕТРИЧЕСКИХ БАЗ

Светлана Петрасова, Нина Хайрова, Валерия Киселева

**Аннотация:** Сложность анализа текстовой информации, содержащейся в наукометрических системах, определяется многозначностью и синонимичностью, которые свойственны языку на всех уровнях его представления, что, прежде всего, влияет на определение смыслового единства языковых единиц. При этом решение задачи усложняется, если речь идет о смысловой близости крупных информационных фрагментов. Поэтому в связи со стремительным ростом объемов информационных ресурсов в наукометрических системах и существующими подходами и методами анализа слабоформализованных данных становятся перспективными задачи обработки текстовой информации на базе интеллектуального анализа. В работе рассматривается информационная технология идентификации смысловой близости фрагментов текстов наукометрических систем. Предложенная технология позволяет определять общие информационные пространства научного взаимодействия авторов за счет идентификации семантически эквивалентных коллокаций в текстах. Технология включает модель формального описания семантико-грамматических характеристик слов атрибутивных, глагольных и субстантивных коллокаций и определение предиката семантической эквивалентности двухсловных коллокаций на основе уравнений алгебры конечных предикатов. Программная имплементация модели представляет собой веб-приложение, определяющее семантически близкие текстовые фрагменты статей, индексируемых в наукометрических базах Google Scholar и Science Direct. В результате определяется эвристическая оценка эффективности разработанной технологии для каждого типа коллокаций.

**Ключевые слова:** наукометрические базы, смысловая близость фрагментов текстов, информационное пространство, синонимия коллокаций, алгебра конечных предикатов.

**ITHEA Keywords:** H.3.3 .Information Search and Retrieval, I.2.4. Knowledge Representation Formalisms and Methods

## **Введение**

---

В последние годы в информационной практике наблюдается возрастающий интерес к наукометрическим исследованиям, обусловленный развитием одноименных баз данных. Современные наукометрические системы формируют статистику, характеризующую динамику показателей востребованности, активности и индексов влияния деятельности ученых. Таким образом, разработка методов структурного анализа современной науки позволяет выявлять исследовательские фронты [King, 2016], ключевые публикации, их авторов, а также мониторить развитие научных направлений и науки в целом.

В настоящее время информационные пространства, представляющие фронты научных исследований, и обозначающие общность, как научных направлений, так и ресурсов, определяются на базе явно выраженных критериев – цитирования, авторства, ключевых слов, а так же критериев коцитирования, проспективных связей и др. Однако в большинстве случаев используемые статистические методы теряют часть знаний об общности информации и являются недостаточными при разработке средств информационного обеспечения библиотек, электронных каталогов, компьютерной библиографии, систем автоматизированного импорта документов и т.п.

Использование технологии идентификации неявно выраженного отношения смысловой близости текстовых фрагментов в работах отдельных авторов позволит выделять единые информационные пространства научных групп в наукометрических системах, что обеспечит релевантный поиск и доступ к научно-исследовательским работам, выполняемым по схожим темам.

---

## **Постановка задачи исследования**

---

Целью работы является разработка информационной технологии идентификации смысловой близости текстовых фрагментов в наукометрических системах для определения информационного пространства научного взаимодействия авторов или общих фронтов научных исследований.

Исследованиями научных фронтов или кластеров в науке занимались Ю. Гарфилд [Garfield, 2005], И.В. Маршакова [Маршакова-Шайкевич, 2013], Ю. Грановский [Грановский, 2013] и др.

Среди основных подходов формирования исследовательских фронтов выделяют анализ цитирования документов. Согласно подходу коцитирования [Чайковский та ін, 2013] документы, совместно процитированные в других документах, отражают основные направления современных исследований и создают «ядро» специальности или отрасли науки.

Подобный анализ связей отражен в методе анализа проспективной связи И.В. Маршаковой. В исследовании [Маршакова-Шайкевич, 2013] «близость документов» определяется числом работ, одновременно цитирующих эти документы. Проспективная связь в системе научных публикаций приводит, с одной стороны, к идентификации тематических групп, характеризующих отдельные направления исследуемой области знаний, позволяет проследить их динамику и развитие во времени и объяснить появление новых направлений этой области. С другой стороны, использование этого метода позволяет выявлять научные сообщества (проспективные коллективы). К достоинствам метода проспективной связи относят объективность и точность. К недостаткам метода отнесены трудоемкость его процедур и большое количество операций механического счета [Акоев и др., 2014].

В своих работах [Пенькова и др., 2001], [Евстигнеев, 2004] определяют такие методы статистического анализа документов при выявлении исследовательских фронтов в наукометрических БД как статистический метод, метод подсчета числа публикаций, метод «цитат-индекса». При формировании информационных пространств *статистический метод* использует, кроме показателей количества публикаций, ссылок и ключевых слов, показатели количества ученых, журналов, открытий и др. *Метод подсчета числа публикаций* по научным направлениям дает возможность получить представление об относительном уровне развития отдельных отраслей науки при формировании информационных пространств наукометрических систем. *Метод «цитат-индекса»* базируется на наукометрическими индикаторе – количестве ссылок в научных публикациях.

В работе [Девяткин и др., 2016] для выявления исследовательских фронтов используется гибридная мера близости публикаций. Согласно подходу мера вычисляется на основе трех компонентов: близость по тематическому сходству текстов, при наличии общего цитирования и при наличии общих авторов.

В связи с постоянными изменениями информационного сообщества для адекватного формирования информационных пространств научных сообществ недостаточно использование явно выраженных критериев. Для решения данной задачи необходимо повысить уровень автоматизации обработки естественно-языковой информации, в том числе за счет решения задачи идентификации близких по смыслу фрагментов текстов или словосочетаний.

При определении смысловой близости словосочетаний используются или статистические закономерности, или определяются их синтаксические характеристики. При этом зачастую семантическая информация не учитывается или дополнительно привлекаются тезаурусы.

Среди наиболее разработанных методов определения смысловой близости словосочетаний выделяют:

- метод определения синонимических коллокаций на основе сравнения их переводов. В работе [Hua Wu et al, 2003] кандидаты в синонимические коллокации определяют на основе одноязычного корпуса, а затем, используя их переводы на втором языке, выбирают подходящие пары кандидатов;
- метод выявления перефразирования за счет сходства фрагментов фраз. Метод Pasca и Dienes [Pasca et al, 2005] определяет множество наборов перефразирования из слов и словосочетаний с помощью попарного выравнивания (alignment) небольших фрагментов предложений по множеству предложений веб-документов;
- метод определения сходства контекста на основе анализа параллельных корпусов. Метод Barzilay и McKeown [Barzilay et al, 2001] определяет однословные лексические парафразы, а так же синтаксические парафразы на основе совокупности нескольких английских переводов одного и того же исходного текста.

Все перечисленные подходы работают или на текстах довольно узких предметных областях, или, при статистических подходах, имеют достаточно низкую точность определения эквивалентных словосочетаний. Оба недостатка не позволяют использовать данные подходы при выделении фрагментов информации единых информационных пространств научного взаимодействия авторов в наукометрических системах.

Таким образом, несмотря на достигнутые результаты, на сегодняшний день проблема идентификации смысловой близости фрагментов текста, в частности, в наукометрических системах для определения фронтов научных исследований остается не до конца решенной.

---

### **Используемая информационная технология**

---

Разработанная информационная технология [Petrasova et al, 2017] позволяет формировать единые информационные пространства научного взаимодействия авторов (рис. 1).

Технология идентификации смысловой близости текстовых фрагментов включает следующие этапы:

1. Отбор статей наукометрических баз Google Scholar и Science Direct.

На этапах 2, 4 применяется разработанная модель идентификации семантически близких коллокаций [Петрасова и др., 2015] на основе алгебры конечных предикатов.

2. За счет выделения семантико-грамматических характеристик слов коллокаций, рядом стоящих в предложении, идентифицируются субстантивные, атрибутивные и глагольные коллокации.

3. Использование WordNet<sup>1</sup> для определения синонимичных слов коллокаций.

---

<sup>1</sup> <https://wordnet.princeton.edu/>

WordNet позволяет выделять классы эквивалентности (синсеты), то есть классы синонимических в каком из своих смыслов терминов, содержащиеся в текстах статей.

4. Далее определяется семантическая близость выделенных фрагментов.

Синонимичные слова могут образовывать близкие по смыслу словосочетания, например, «хранить данные» и «содержать сведения», в то же время, могут формировать близкие по смыслу словосочетания, «хранение данных»  $\neq$  «информация репозитария».

Итак, коллокации могут считаться близкими по смыслу, если семантико-грамматические характеристики коллокатов словосочетаний удовлетворяют предикату семантической эквивалентности [Petrasova et al, 2015].

5. На последнем этапе разработанной технологии, используя найденные подобные эквивалентные фрагменты информации, выделяем единое информационное пространство научного взаимодействия авторов.



Рис. 1. Основные этапы разработанной информационной технологии

### Описание модели

Идентификация информационно-лингвистических сущностей, в частности коллокации и отношений, с характерными для естественного языка гибкостью и многозначностью требует интеллектуальных средств обработки естественно-языковых текстов.

В качестве формального аппарата для построения модели идентификации дискретного, конечного набора смысловых сущностей и отношений в текстовой информации наукометрических систем использовался аппарат алгебры конечных предикатов.

Для описания семантических и грамматических характеристик фрагментов информации, а именно слов коллокаций, были введены предметные переменные, определяющие:

- часть речи:  $a^N$  (noun),  $a^A$  (adjective),  $a^V$  (verb);
- синтаксическую роль существительного:  $a^{NSub}$  (subject),  $a^{NObj}$  (object) и прилагательного:  $a^{AAtt}$  (attribute),  $a^{APr}$  (predicative);
- транзитивность глагола:  $a^{VTr}$  (transitive),  $a^{VIntr}$  (intransitive);
- возможные семантические роли существительных:  $c^{Ag}$  – агент (активный участник ситуации),  $c^{Att}$  – атрибут (связь предмета и признака),  $c^{Pac}$  – пациент (пассивный участник ситуации или объект действия),  $c^{Adr}$  – адресат (получатель сообщения),  $c^{Ins}$  – инструмент (участник, с помощью которого осуществляется действие, или инструмент осуществления действия),  $c^M$  – место (местонахождение одного из участников ситуации).

Введенный на множестве словоформ предикат  $P(x)$  превращается в 1, если главная словоформа словосочетаний  $x$  имеет определенную семантико-грамматическую информацию. Множество допустимых семантико-грамматических характеристик зависимого слова словосочетания  $y$  описывается предикатом  $P(y)$ .

Предикат идентификации рядом стоящих сущностей, образующих коллокации английского языка:

$$P(x, y) = (x^{NSubAg} \vee x^{NSubOfAg} \vee x^{VTr})(y^{NObjAtt} \vee y^{NObjPac} \vee y^{AAtt} \vee y^{APr})$$

Введенный предикат семантической эквивалентности между коллокациями определяет семантико-грамматические характеристики коллокатов близких по смыслу словосочетаний. Отношение семантической эквивалентности двух двухсловных коллокаций может быть определено как:

$$P(x_1, y_1) \times P(x_2, y_2) = \gamma_i(x_1, y_1, x_2, y_2) \wedge P(x_1, y_1) \wedge P(x_2, y_2)$$

где знак  $\times$  обозначает операцию определения смысловой близости, знак  $\wedge$  определяет конъюнкцию, предикат  $\gamma_i(x_1, y_1, x_2, y_2)$  исключает коллокации, между которыми не может быть установлена смысловая эквивалентность.

Предикат  $\gamma_i(x_1, y_1, x_2, y_2) = x_1^{VTr} y_1^{NObjPac} x_2^{VTr} y_2^{NObjPac}$  показывает семантическую близость глагольных коллокаций ( $V_x N_y$ ), например, identify information  $\approx$  extract data.

Предикат  $\gamma_2(x_1, y_1, x_2, y_2) = x_1^{NSubOfAg} y_1^{NObjAtt} y_2^{NObjAtt} x_2^{NSubAg} \vee x_1^{NSubOfAg} y_1^{NObjAtt} x_2^{NSubOfAg} y_2^{NObjAtt} \vee y_1^{NObjAtt} x_1^{NSubAg} y_2^{NObjAtt} x_2^{NSubAg}$  показывает семантическую близость субстантивных коллокаций ( $N_x N_y$ ), например, figure means  $\approx$  pattern technique.

Предикат  $\gamma_3(x_1, y_1, x_2, y_2) = y_1^{AAtt} x_1^{NSubAg} x_2^{NSubAg} y_2^{APr} \vee y_1^{AAtt} x_1^{NSubAg} y_2^{AAtt} x_2^{NSubAg} \vee x_1^{NSubAg} y_1^{APr} x_2^{NSubAg} y_2^{APr}$  показывает семантическую близость между атрибутивными коллокациями ( $A_y N_x$ ), например, important topics  $\approx$  essential issues, key field  $\approx$  central area.

## Программная реализация

Программная имплементация модели представляет собой веб-приложение, анализирующее текстовую информацию для идентификации семантически близких текстовых фрагментов, а именно коллокаций (рис. 2).

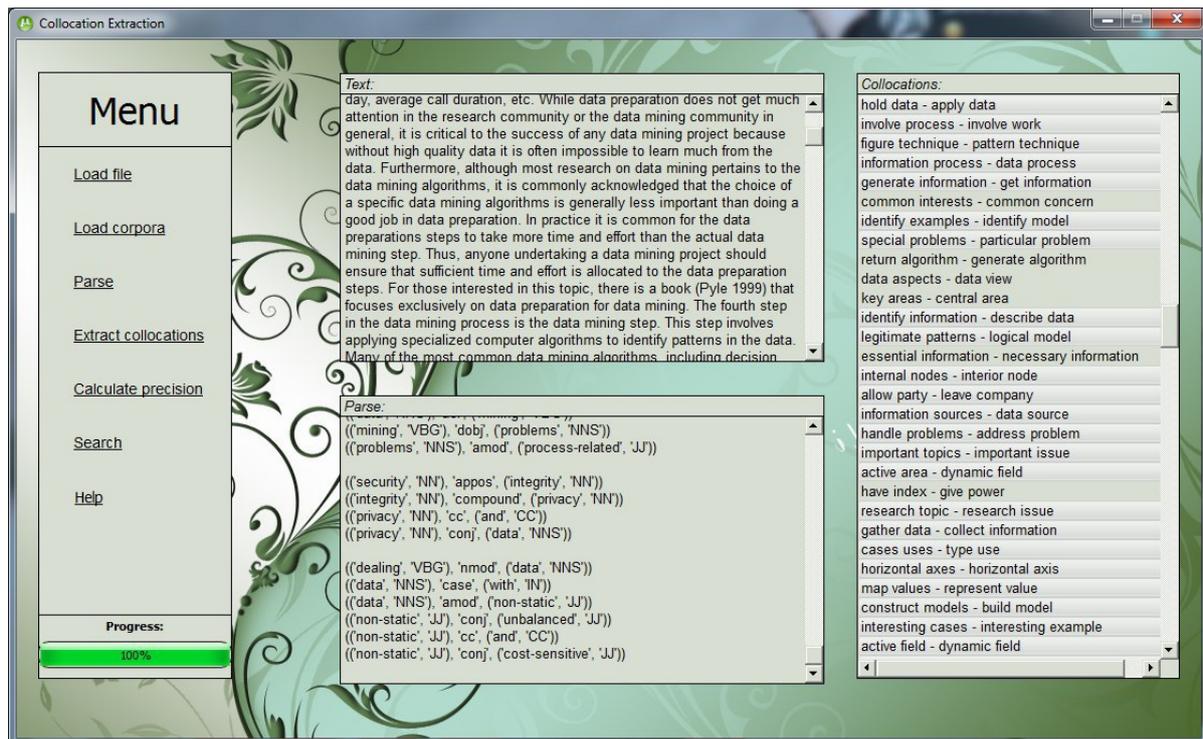


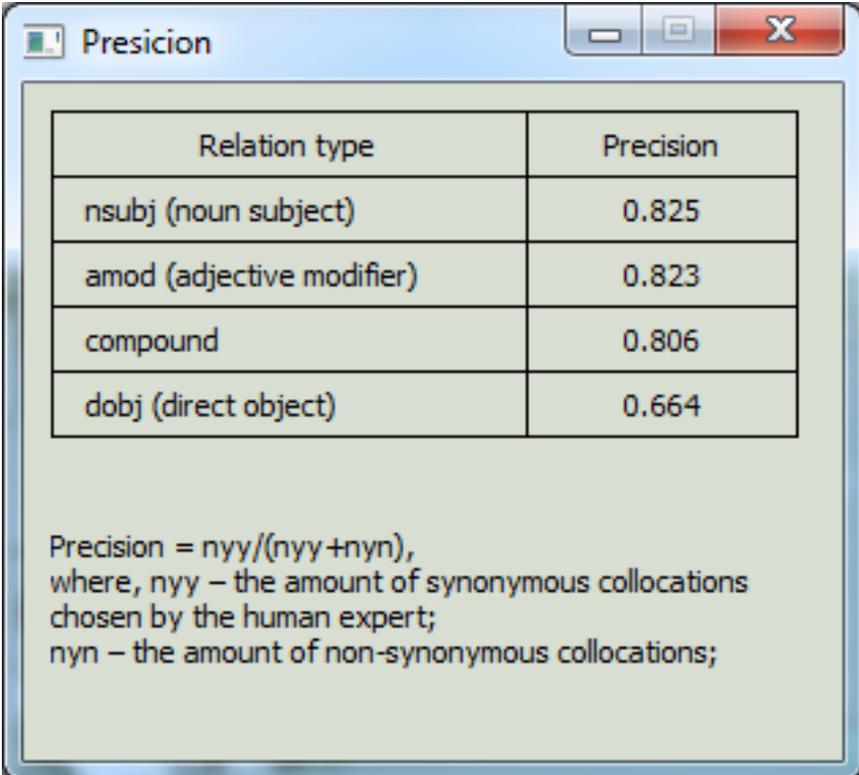
Рис. 2. Программная имплементация модели идентификации близких по смыслу коллокаций

Программа отображает извлеченную информацию в виде:

- перечня синонимичных коллокаций. Для извлечения кандидатов в коллокаты используются синсеты WordNet, содержащие синонимы слов коллокаций, которые отсортированы по частям речи и значениям;

- грамматической характеристики слов коллокаций, полученной с помощью модуля Universal Dependencies<sup>2</sup>. В атрибутивных, глагольных и субстантивных коллокациях Universal Dependencies определяет такие типы связей, как *amod: adjectival modifier*, *dobj: direct object*, *compound: compound*, *nsubj: nominal subject*;
- текстовой информации или ее источников, т.е. статей, индексируемых в Google Scholar и ScienceDirect, из которых данные коллокации были извлечены.

Для оценки эффективности работы технологии в меню программы было создано пункт Calculate Precision (рис. 3). Метрика Precision рассчитывается по формуле  $Precision = n_{yy} / (n_{yy} + n_{yn})$ .



Relation type	Precision
nsubj (noun subject)	0.825
amod (adjective modifier)	0.823
compound	0.806
dobj (direct object)	0.664

Precision =  $n_{yy} / (n_{yy} + n_{yn})$ ,  
where,  $n_{yy}$  – the amount of synonymous collocations  
chosen by the human expert;  
 $n_{yn}$  – the amount of non-synonymous collocations;

Рис. 3. Результаты Calculate Precision

## Выводы

Результатом данного исследования является разработка технологии идентификации близких по смыслу фрагментов текстовой информации в наукометрических системах. Программная реализация разработанной модели идентификации близких по смыслу коллокаций, основывающейся на использовании алгебры конечных предикатов, позволяет определить информационные пространства научного взаимодействия авторов статей.

<sup>2</sup> <http://universaldependencies.org/u/dep/index.html>

## Литература

---

- [Barzilay et al, 2001] Barzilay R., McKeown K.R. Extracting Paraphrases from a Parallel Corpus. In Proceedings of the 39th Annual Meeting on Association for Computational Linguistics (ACL '01). Stroudsburg, PA, USA, 2001. P. 50–57.
- [Garfield, 2005] Garfield E. The Agony and the Ecstasy – The History and the Meaning of the Journal Impact Factor. In Proc. of Inter. Congress on Peer Review and Biomedical Publication. Chicago, 2005. Access mode: <http://garfield.library.upenn.edu/papers/jifchicago2005.pdf>.
- [Hua Wu et al, 2003] Hua Wu, Ming Zhou Synonymous Collocation Extraction Using Translation Information. In Proceedings of the 41th Annual Meeting on Association for Computational Linguistics (ACL '03). Stroudsburg, PA, USA, 2003. Vol. 1. P. 120–127.
- [King, 2016] King C. Research Fronts 2016: The Hottest Areas in Science. Access mode: <http://stateofinnovation.com/research-fronts-2016-the-hottest-areas-in-science>
- [Pasca et al, 2005] Pasca M., Dienes P. Aligning Needles in a Haystack: Paraphrase Acquisition across the Web. In Proceedings of the Second International Joint Conference: Natural Language Processing (IJCNLP 2005). Korea, 2005. P. 119–130.
- [Petrasova et al, 2015] Petrasova S. and Khairova N. Automatic Identification of Collocation Similarity. In Proceedings of 10th International Scientific and Technical Conference: Computer Science & Information Technologies (CSIT'2015), Lviv, 2015. P. 136–138.
- [Petrasova et al, 2017] Petrasova S. and Khairova N. Using a Technology for Identification of Semantically Connected Text Elements to Determine a Common Information Space. Cybernetics and Systems Analysis, Springer. Vol. 53 (1). 2017. P. 115–124.
- [Акоев и др., 2014] Акоев М.А., Маркусова В.А. и др. Руководство по наукометрии: индикаторы развития науки и технологии : монография. – Екатеринбург: Изд-во Урал. ун-та, 2014. – 250 с.
- [Грановский, 2013] Грановский Ю.В. Наукометрия и управление научными коллективами // Наукоевческие исследования. – 2013. – С. 127–150.
- [Девяткин и др., 2016] Девяткин Д.А., Швец. А.В., Тихомиров И.А. Выявление направлений исследований и научных коллективов на основе анализа полнотекстовых коллекций научных публикаций. – Режим доступа : <http://www.gpntb.ru/win/inter-events/crimea2016/disk/2046.pdf>
- [Евстигнеев, 2004] Евстигнеев В.А. Наукометрические исследования в информатике // Новосибирская школа программирования. Переключка времен : сб. тр. – Новосибирск: Ин-т систем информатики им. А.П. Ершова СО РАН, 2004. – С. 203–217.
- [Петрасова и др., 2015] Петрасова С.В., Хайрова Н.Ф. Логико-лингвистическая модель идентификации семантически эквивалентных коллокаций // Вестник Национального

технического университета «Харьковский политехнический институт». – Харьков : НТУ «ХПИ», 2015. – № 58 (1167). – С. 14–17.

[Чайковський та ін., 2013] Чайковський Ю.Б., Сілікіна Ю.В., Потоцька О.Ю. Наукометричні бази та їх кількісні показники (Частина I. Порівняльна характеристика наукометричних баз) // Вісник Національної академії наук України. – 2013. – № 8. – С. 89–98.

[Маршакова-Шайкевич, 2013] Маршакова-Шайкевич И.В. Роль библиометрии в оценке исследовательской активности науки // УБС. – 2013. – № 44. – С. 210–247.

[Пенькова и др., 2001] Пенькова О.В., Тютюнник В.М. Информетрия, наукометрия и библиометрия: наукометрический анализ современного состояния // Вестник ТГУ. – 2001. – Т. 6. – № 1. – С. 86–88.

---

### Authors' Information

---



**Светлана Петрасова** – к.т.н., старший преподаватель кафедры интеллектуальных компьютерных систем Национального технического университета «Харьковский политехнический институт», ул. Кирпичева, 2, Харьков, 61002, Украина; e-mail: [svetapetrasova@gmail.com](mailto:svetapetrasova@gmail.com)

**Научные интересы:** искусственный интеллект, компьютерная лингвистика, Information Extraction, Natural Language Processing



**Нина Хайрова** – профессор, д.т.н., профессор кафедры интеллектуальных компьютерных систем Национального технического университета «Харьковский политехнический институт», ул. Кирпичева, 2, Харьков, 61002, Украина; e-mail: [nina\\_khajrova@yahoo.com](mailto:nina_khajrova@yahoo.com)

**Научные интересы:** искусственный интеллект, компьютерная лингвистика, экстракция знаний из текстов, Text Mining, Opinion Mining, Web Mining, Natural Language Processing



**Валерия Киселева** – магистр Национального технического университета «Харьковский политехнический институт», ул. Кирпичева, 2, Харьков, 61002, Украина; e-mail: [lerakiseleva@yahoo.com](mailto:lerakiseleva@yahoo.com)

**Научные интересы:** компьютерная лингвистика, идентификация семантической синонимии коротких фрагментов текстов.

## Identification of Semantic Similarity of Text Fragments in Scientometric Bases

Svitlana Petrasova, Nina Khairova, Valeriia Kysilova

**Abstract:** *This paper considers the information technology for identification of semantic similarity of text fragments in scientometric systems. The proposed technology allows determining common information spaces of authors' scientific interaction due to identification of semantic equivalence of collocations in texts. The technology includes a model for a formal description of the semantic and grammatical characteristics of words in attributive, verbal and substantive collocations and identification of the semantic equivalence predicate for two-word collocations based on the algebra of finite predicates. The developed software implementation is a web application that defines semantically connected text fragments of articles indexed in Google Scholar and Science Direct. As a result, an effectiveness estimate of the developed technology for each type of collocations is determined.*

**Keywords:** *scientometric bases, semantic similarity of text fragments, information space, collocation synonymy, algebra of finite predicates.*

## АНАЛИЗ ВЗАИМОЗАВИСИМОСТИ ВРЕМЕННЫХ РЯДОВ БИТКОИНА И АКТИВНОСТИ СООБЩЕСТВ В СОЦИАЛЬНЫХ СЕТЯХ

Людмила Кириченко, Тамара Радивилова, Виталий Булах, Вадим Чакрян

**Аннотация:** В работе проведен сравнительный кросскорреляционный и мультифрактальный анализ временных рядов курса биткоина и рядов активностей сообществ в социальных сетях связанных с этой криптовалютой. Была обнаружена значительная корреляция между курсом Биткоина и активностью сообществ временных рядов, когда интерес к биткоину и системе блокчейн возрастал вместе с ростом курса биткоина. Фрактальный анализ временных рядов выявил наличие самоподобных и мультифрактальных свойств для рядов активностей сообществ. Результаты исследований показывают, что ряды, которые имеют сильную корреляционную зависимость, имеют близкие мультифрактальные свойства.

**Ключевые слова:** самоподобные и мультифрактальные временные ряды, обобщенный показатель Херста, кросскорреляция, биткоин, социальные сети

**ACM Classification Keywords:** G.3 Probability and statistics - Time series analysis, Stochastic processes, G.1 Numerical analysis, G.1.2 Approximation - Wavelets and fractals.

---

### Введение

Биткоин является основной криптовалютой и производным финансовым инструментом, который пользуется большим спросом в последние годы. Биткоин - это децентрализованная электронная платежная система, созданная в 2008 году, рыночная капитализация которой неуклонно растет. Биткоин использует сложные криптографические методы для подписи транзакций и определения средств управления. Биткоин — это одноранговая система, пользователи которой проводят транзакции непосредственно без необходимости посредников, и в этих транзакциях записывается информация для последующей проверки всеми узлами в открытой распределенной книге называемой Блокчейн. [Nakamoto, 2008].

В настоящее время стало общепризнанным, что финансовые рынки обладают фрактальными свойствами [Peters, 2003]. В последние годы был проведен анализ цены биткоина на наличие фрактальных свойств, который выявил сильно выраженные фрактальные свойства (см, например [Meléndez, 2014]).

---

---

Анализ финансовых рынков при инвестиционных решениях полагается на традиционные источники информации, такие как раскрытие информации о компаниях, новости на рынке, отчеты, и так далее. Данные социальных сетей использовались для коммерческих целей для автоматического получения отзывов клиентов о товаре или бренде, поиска новых клиентов, анализа предпочтения и влияния на пользователей социальных сетей при голосовании, выборке, на рынке и т.д. За прошедшие несколько лет социальные сети быстро проникли в различные аспекты нашей жизни, и содержат огромный объем информации; данные социальных сетей представляют альтернативный источник информации для аналитиков и инвесторов финансовых рынков [Arnaboldi, 2015, Do, 2015, Bollen, 2011, Mittal, 2013, Matta, 2016]. Одно из направлений исследований данных в социальных сетях и финансовом секторе – это анализ взаимосвязи между информацией о биткойне и блокчейне, содержащейся в социальных сетях, и цене биткойна [Feng Mai, 2016, Peng Xie, 2017, Greaves, 2015].

Социальные сети и взаимодействия между участниками рынка могут существенно влиять на курс биткойна. В [Garcia, 2014] показано, что растущая популярность биткойна вызывает увеличение объема поиска, что, в свою очередь, приводит к более высокой активности в социальных сетях, связанной с биткойном. Покупка биткойнов пользователями приводит к большему интересу, что приводит к более высоким ценам, которые в конечном итоге будут возвращены в объем – образуется цикл социальной обратной связи.

Объектом данного исследования являются процессы взаимосвязи и взаимовлияния между динамикой финансовых рынков и социальных сетей. Субъектом исследования является корреляционная зависимость и фрактальная структура финансовых рядов и активности социальных групп. Целью работы является анализ взаимозависимости фрактальных характеристик временных рядов курса криптовалюты биткойн и активности сообществ в социальной сети фейсбук, связанных с биткойном.

---

### Самоподобные и мультифрактальные временные ряды

---

Самоподобие случайных процессов заключается в сохранении вероятностных характеристик при изменении масштаба времени. Стохастический процесс  $X(t)$  является самоподобным с параметром  $H$ , если процесс  $a^{-H} X(at)$  описывается теми же законами конечномерных распределений, что и  $X(t)$ :

$$\text{Law}\{a^{-H} X(at)\} = \text{Law}\{X(t)\}, \quad \forall a > 0. \quad (1)$$

Параметр  $H$ ,  $0 < H < 1$ , называемый показателем Херста, представляет собой степень самоподобия процесса. Наряду с этим свойством, показатель  $H > 0.5$  характеризует меру долгосрочной зависимости стохастического процесса. Это означает, что если временной ряд в течении какого-то времени возростал (убывал), то с вероятностью, близкой к показателю Херста, ряд сохранит эту тенденцию в течение аналогичного промежутка времени [Feder, 1988]. Начальные моменты самоподобного случайного процесса можно выразить как

$$M\left[|X(t)|^q\right] = C(q) \cdot t^{qH}, \quad (2)$$

где величина  $C(q) = M\left[|X(1)|^q\right]$ .

В сравнении с самоподобными процессами (1-2) мультифрактальные процессы проявляют более сложное скейлинговое поведение:

$$\text{Law}\{X(at)\} = \text{Law}\{\mathcal{M}(a) \cdot X(t)\}, \quad a > 0, \quad (3)$$

где  $\mathcal{M}(a)$  – случайная функция, независимая от  $X(t)$ . В случае самоподобия (монофрактальности) процесса  $\mathcal{M}(a) = a^H$ .

Для моментов мультифрактальных процессов выполняется отношение:

$$M\left[|X(t)|^q\right] = c(q) \cdot t^{qh(q)}, \quad (4)$$

где  $c(q)$  – некоторая детерминированная функция;  $h(q)$  – обобщенный показатель Херста, являющийся в общем случае нелинейной функций. Значение  $h(q)$  при  $q = 2$  совпадает со значением степени самоподобия  $H$ . Для монофрактальных процессов обобщенный показатель Херста не зависит от параметра  $q$ :  $h(q) = H$ . [Reidi, 2002].

Мультифрактальные объекты являются статистически неоднородными самоподобными объектами. Для мультифрактальных временных рядов статистическая неоднородность объекта выражается в неоднородности распределения данных ряда, т.е. наличии тяжелых хвостов функции плотности распределения вероятностей временного ряда.

---

### Метод мультифрактального детрендрованного флуктационного анализа

---

Одним из наиболее популярных методов исследования мультифрактальной структуры временных рядов является метод мультифрактального детрендрованного флуктационного анализа (МФ ДФА) [Kantelhardt, 2002, 2008], который основан на обработке нестационарных временных рядов с трендами.

Согласно МФ ДФА для начального временного ряда  $x(t)$  формируется кумулятивный временной ряд  $y(t) = \sum_{i=1}^t x(i)$ , который в дальнейшем делится на  $N$  непересекающихся сегментов длиной  $\tau$ , и для таких сегментов  $y(t)$  в дальнейшем рассчитывается флуктуационная функция:

$$F^2(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} (y(t) - Y_m(t))^2, \quad (5)$$

где  $Y_m(t)$  - локальный  $m$ -полиномиальный тренд на выбранном сегменте. Усредненная по всему временному ряду  $y(t)$  функция  $F(\tau)$  имеет масштабирование на участке длиной  $\tau$ :

$$F(\tau) \propto \tau^H \quad (6)$$

При изучении мультифрактальных свойств зависимость флуктуационной функции от параметра  $\tau$  рассматривается следующим образом:

$$F_q(\tau) = \left\{ \frac{1}{N} \sum_{i=1}^N [F^2(\tau)]^{\frac{q}{2}} \right\}^{\frac{1}{q}} \quad (7)$$

Если исследуемая серия является мультифрактальной и имеет долгосрочную зависимость, флуктуационная функция  $F_q(\tau)$  представлена степенным законом:

$$F_q(\tau) \propto \tau^{h(q)}, \quad (8)$$

где  $h(q)$  – обобщенный показатель Херста.

Очевидно, что оценивание показателя Херста по экспериментальным данным играет важнейшую роль в изучении процессов, обладающими мультифрактальными свойствами. Однако, прежде чем проводить анализ мультифрактальных свойств, необходимо убедиться, что временной ряд действительно является самоподобным, что соответствует наличию линейного участка флуктуационной функции  $F(\tau)$  в (6). При оценивании показателя Херста методом ДФА (5), необходимо сначала провести прикидочное оценивание, используя локальные полиномы увеличивающейся степени и определить наименьшую степень полинома, начиная с которой оценка показателя Херста перестает изменяться. После этого для оценивания самоподобия временного ряда необходимо удалять локальный полиномиальный тренд найденной степени.

---

### Методика расчета медианной кросскорреляции

---

При анализе данных, в которых присутствуют существенные выбросы, что характерно для мультифрактальных временных рядов, возникает вопрос о степени влияния данных выбросов на стандартные статистические оценки. Появление в выборке нескольких выбросов может сильно ухудшить оценку математического ожидания, дисперсии или автокорреляционной функции, но очень слабо повлиять на значение медианы. Поэтому использование медианы вместо математического ожидания в формуле коэффициента корреляции приводит к тому, что он становится менее восприимчив к сильным выбросам и такая оценка является более не смещенной:

$$r_{median} = \frac{\text{Median}[(x - \text{Median}[x]) (y - \text{Median}[y])]}{\sqrt{\text{Median}[(x - \text{Median}[x])^2] \text{Median}[(y - \text{Median}[y])^2]}} \quad (9)$$

Проиллюстрируем применение данного подхода на примере оценивания корреляционной функции процесса авторегрессии. На рис. 1 вверху слева представлен ряд авторегрессии первого порядка с коэффициентом авторегрессии  $\phi = 0.9$ , для которого, как известно, теоретическая нормированная автокорреляционная функция  $k(\tau) = \phi^\tau$ . Выборочная автокорреляционная функция, полученная с помощью среднего значения, представлена под временным рядом, а выборочная корреляционная функция, полученная с помощью медианного оценивания, показана внизу слева. Вверху справа изображен тот же авторегрессионный временной ряд, но с добавлением сильных выбросов в случайные моменты времени, а под ним соответствующие выборочные автокорреляционные функции. Очевидно, что оценка автокорреляционной функции, полученная медианным способом практически не изменилась, в то время как обычная выборочная корреляционная функция для ряда с выбросами не соответствует первоначальному ряду.

---

### Результаты исследований

---

Для проведения исследований в сети Facebook было выбрано три больших сообщества, интересы которых связаны с криптовалютой: Bitcoin Product/service (около 200 тыс. участников), Bitcoin Finance company (70 тыс. пользователей) и группа Blockchain (62 тыс. пользователей). Для периода с октября 2016 по октябрь 2017 были выбраны данные по лайкам и репостам для этих групп и данные курса криптовалюты Биткоин.

Таким образом, были сформированы семь временных рядов ежедневных данных: *Bitcoin* – курс криптовалюты Биткоин относительно доллар США, *BitPr\_likes* и *BitPr\_reposts* – соответственно количество лайков и репостов в сообществе Bitcoin Product/service, *BitFin\_likes* и *BitFin\_reposts* – количество лайков и репостов в сообществе Bitcoin Finance company, *Blchain\_likes* и *Blchain\_reposts* – количество лайков и репостов в сообществе Blockchain. На рис. 2 показаны временные ряды значений курса криптовалюты *Bitcoin* (вверху) и количества лайков (посредине) и репостов (внизу) сообщества Bitcoin Product/service (*BitPr\_likes* и *BitPr\_reposts*).

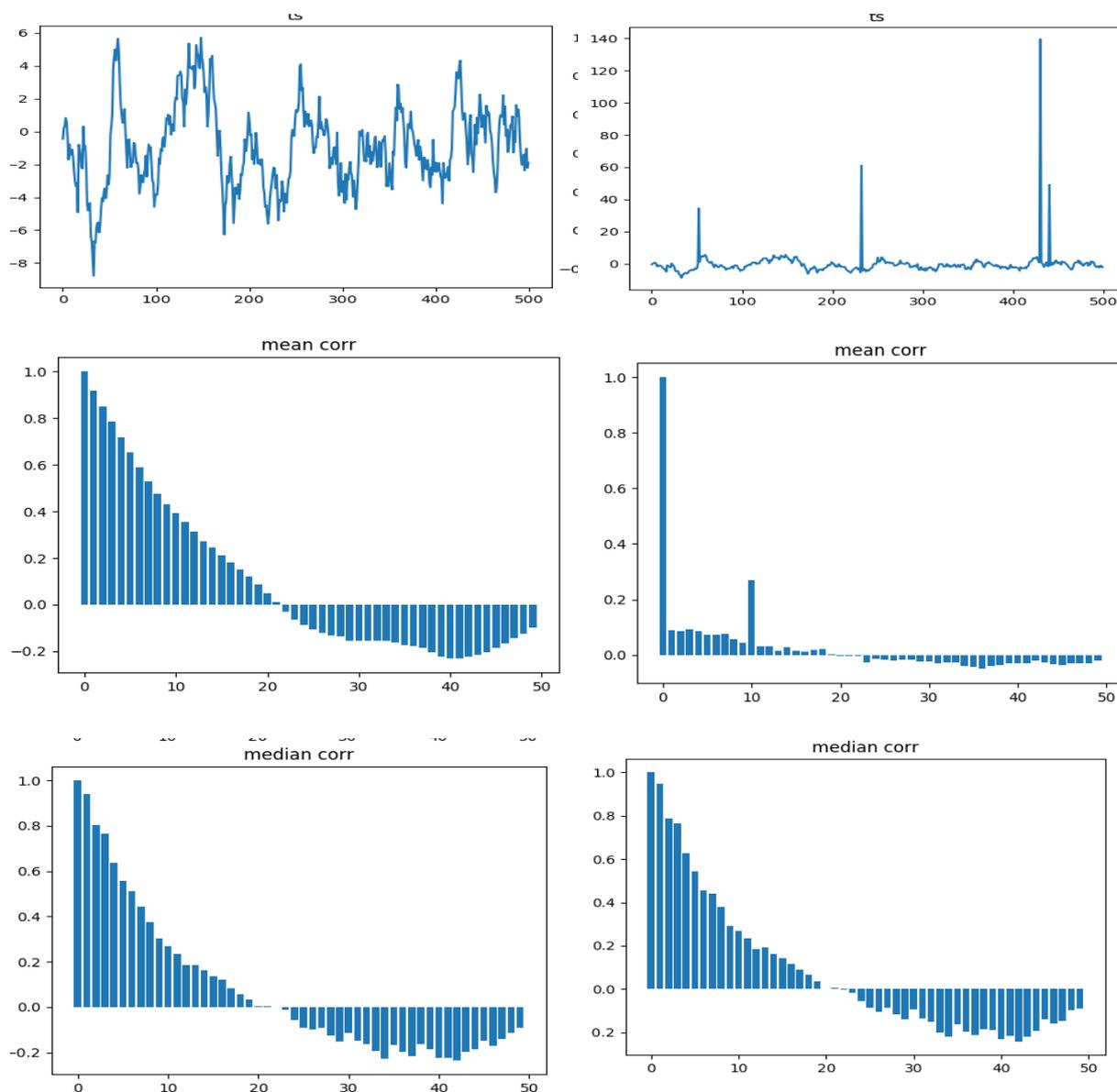


Рисунок 1. Ряд авторегрессии (вверху слева) и этот же ряд с выбросами (вверху справа) и их автокорреляционные функции: в средней части на основе математического ожидания, в нижней - на основе медианы.

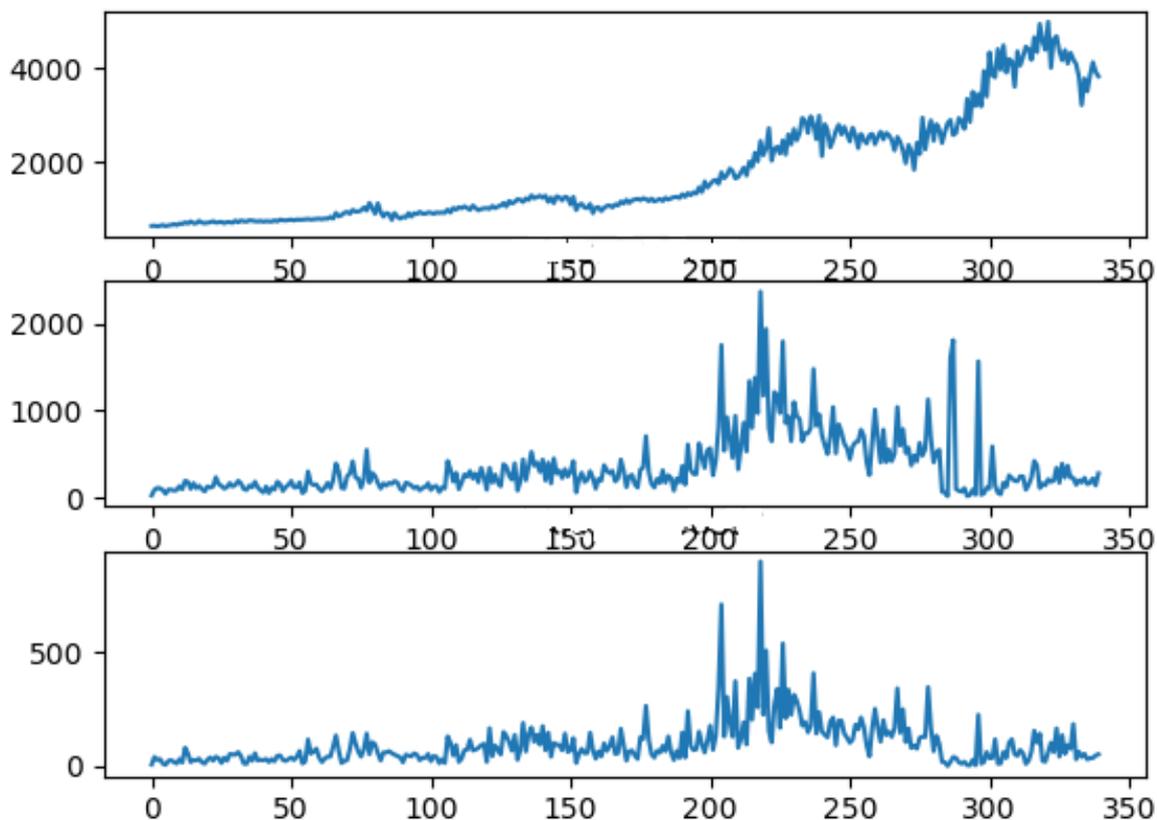


Рисунок 2. Временные ряды *Bitcoin*, *BitPr\_likes* и *BitPr\_reposts*

В работе проведен кросскорреляционный анализ между рядом курса криптовалюты Биткоин и рядами активностей сообществ. Рассчитаны выборочные коэффициенты кросскорреляции для первых 100 значений временных рядов, когда интерес к биткоину и системе блокчейн в сообществах возрастал вместе с ростом курса биткоина (см. рис. 2) и для последних 100 значений, когда пик интереса постепенно начал спадать. Также были рассчитаны коэффициенты кросскорреляции для близких по структуре рядов лайков и репостов в одних и тех же группах. Поскольку ряды активностей обладают значительными выбросами, при расчете коэффициентов корреляции был использован медианный способ расчета по формуле (9).

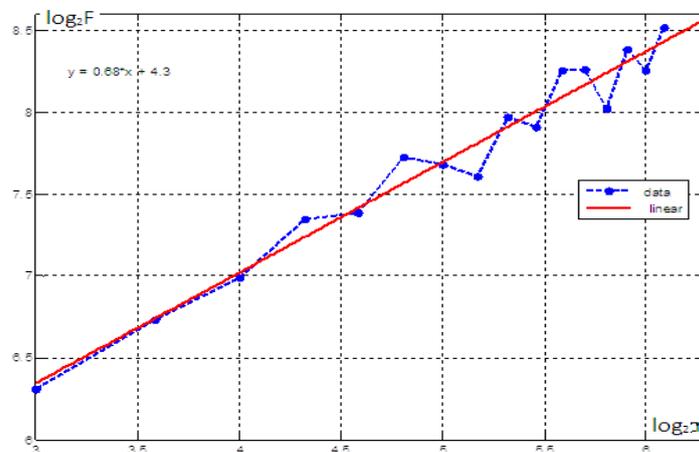
Анализ показал значимую корреляцию между курсом биткоина и активностью сообществ. Была обнаружена значительная корреляция между курсом Биткоина и активностью сообществ временных рядов, когда интерес к биткоину и системе блокчейн возрастал вместе с ростом курса биткоина. Для последних значений временных рядов, когда пик интереса постепенно начал спадать, корреляция стремится к нулю.

В табл.1 представлены максимальные коэффициенты корреляции  $r(\tau)$  между рядом *Bitcoin* и рядами лайков выбранных сообществ. Между рядами лайков и репостов одного сообщества существует сильная кросс-корреляция в диапазоне значений  $0.85 \leq r(0) \leq 0.95$

Таблица 1. Максимальные коэффициенты кросс-корреляции

Ряды	$r(\tau)$	
	Начальные значения ряда	Конечные значения ряда
<i>BitPr_likes</i>	0.81	0.18
<i>BitFin_likes</i>	0.58	0.12
<i>Blchain_likes</i>	0.43	-0.08

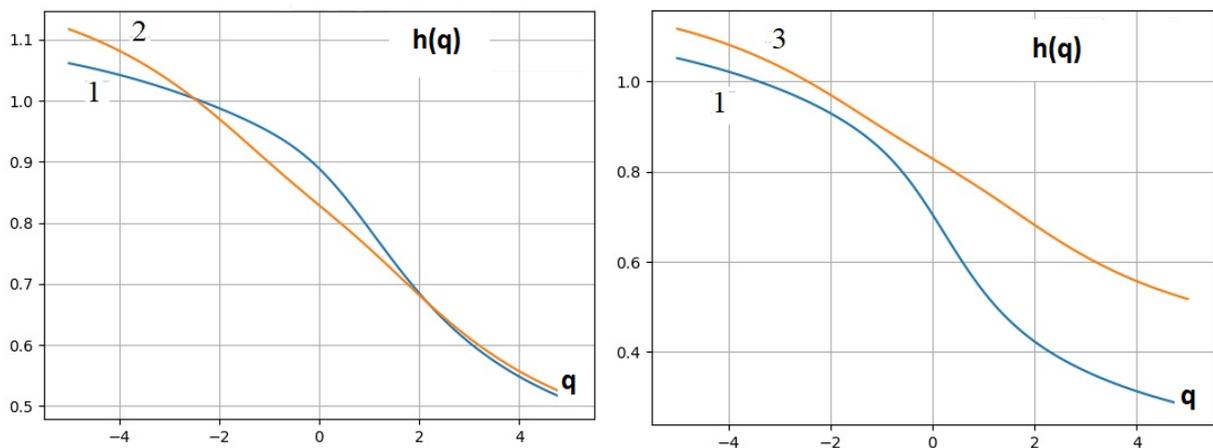
Проведен мультифрактальный анализ всех рассмотренных временных рядов. Особое внимание уделено проверке наличия свойства самоподобия (2) методом ДФА с использованием формулы (5) и подбором степени полиномиального тренда. Для проверки наличия самоподобия были построены и исследованы флуктуационные функции  $F(\tau)$  для каждого временного ряда. Самоподобное поведение соответствует наличию участка функции  $F(\tau)$  с аппроксимированной линейной зависимостью (6). На рис. 3 показан график  $F(\tau)$  для временного ряда *BitPr\_likes*, который был представлен на рис.1 (посередине). Полученная кривая достаточно хорошо аппроксимируется прямой линией с наклоном, определяющим показатель Херста. Исследования показали, что для данного ряда степень полиномиального тренда равна 2.

Рисунок 3. Флуктуационная функция  $F(\tau)$  для ряда *BitPr\_likes*

Исследование флуктуационных функций для временных рядов групп активности показало наличие самоподобных свойств. После этого был проведен мультифрактальный анализ методом МФДФА по формулам (7-8). Исследования показали наличие мультифрактальных свойств для всех рядов групповой активности.

Особое внимание было уделено фрактальному анализу первых значений временных рядов, при возрастающем интересе к биткоину и для последних значений, когда интерес постепенно начал спадать. Рис. 4 представляет функции обобщенного показателя Херста для рядов *Bitcoin* и *BitPr\_likes* (в левой части), *Bitcoin* и *Blchain\_likes* (в правой части) на интервалах времени соответствующим значениям 100-200 (значимая корреляция) и значениям 250-350 (незначимая корреляция) временных рядов. В верхней части рисунка показаны функции обобщенного показателя Херста для начальных значений рядов, в нижней- для последних значений.

Для начальных значений рядов



Для последних значений рядов

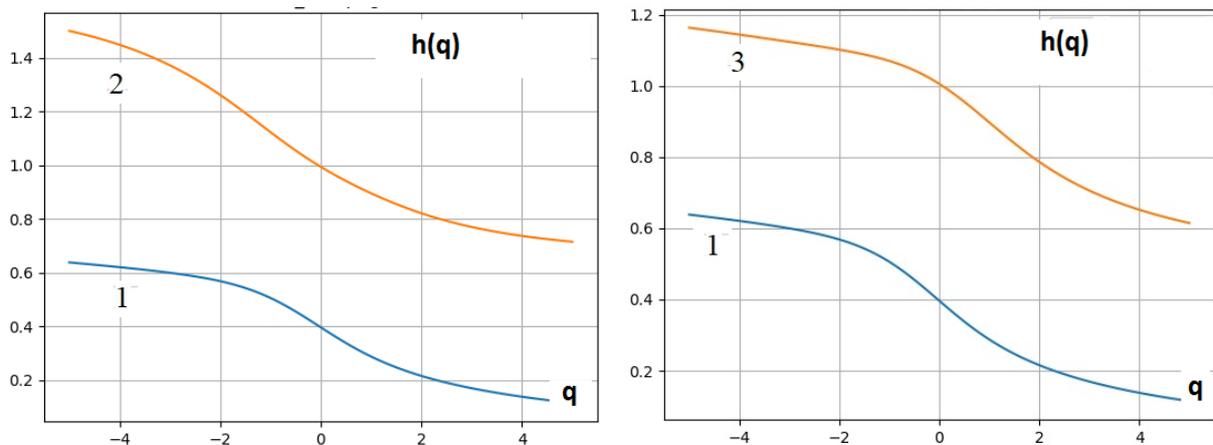


Рисунок 4 Обобщенный показатель Херста  $h(q)$  для *Bitcoin* (линия 1) и *BitPr\_likes* (линия 2), *Bitcoin* и *Blchain\_likes* (линия 3). Вверху значимая корреляция, внизу незначимая корреляция между рядами.

Результаты исследования показали, что ряды с сильной корреляционной зависимостью имеют сходную мультифрактальную структуру. Ряды *Bitcoin* и *BitPr\_likes* имеют сильную корреляцию в начальных значениях рядов и близкие значения мультифрактальной характеристики. Слабая корреляция между *Bitcoin* и *Blockchain\_likes* проявляется в различии между мультифрактальной структурой этих рядов. При незначимой корреляции между рядами функции обобщенного показателя Херста не имеют сходства.

Поскольку почти всегда ряды лайков и репостов сильно связаны друг с другом, они имеют близкие значения обобщенного показателя Херста. Ряды *BitPr\_likes* и *BitPr\_reposts* показаны на рис. 2. Их сильная корреляционная зависимость очевидна. На рис. 5 представлены функции обобщенного показателя Херста для рядов *BitPr\_likes* и *BitPr\_reposts* в начале и конце исследуемого периода времени.

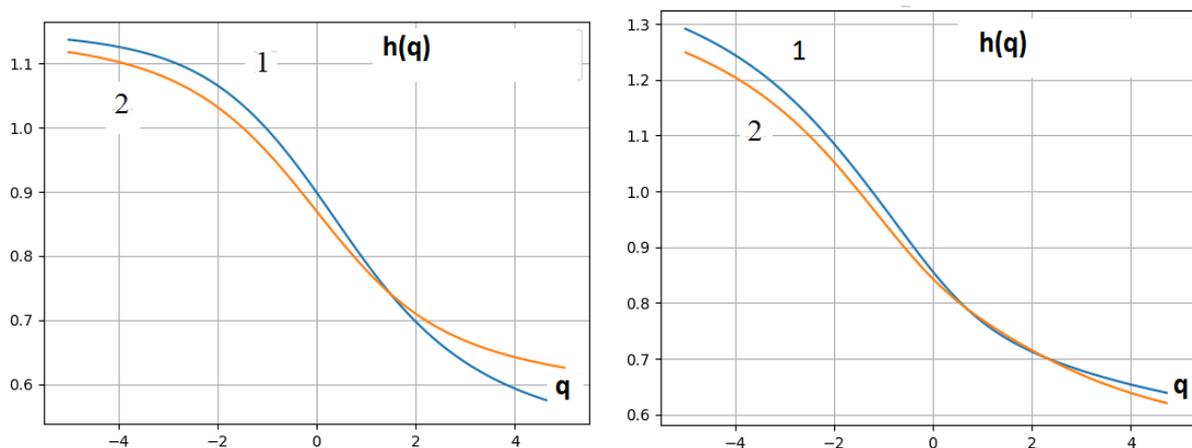


Рисунок 5. Обобщенный показатель Херста  $h(q)$  для рядов *BitPr\_likes* (линия 1) и *BitPr\_reposts* (линия 2), в начале (слева) и конце (справа) исследуемого периода времени.

Аналогичные результаты справедливы и для других исследованных рядов. Чем больше корреляция между фрактальными временными рядами, тем более близки их фрактальные характеристики.

---

### Благодарности

Работа опубликована при поддержке общества ITHEA ISS ([www.ithea.org](http://www.ithea.org)) и ADUIS ([www.aduis.com.ua](http://www.aduis.com.ua)).

## **Заключение**

---

В работе был проведен корреляционный анализ временных рядов, который выявил значимую корреляцию между курсом Биткоина и активностью связанных с ним социальных сообществ. При расчете коэффициентов корреляции был использован медианный способ расчета, который показал хорошие результаты. Фрактальный анализ временных рядов выявил наличие самоподобных и мультифрактальных свойств как для ряда цен биткоина, так и для рядов активностей сообществ.

Результаты исследований показали, что ряды, имеющие корреляционную зависимость, имеют сходную мультифрактальную структуру: чем больше корреляция между фрактальными рядами, тем более сходны их фрактальные характеристики.

Таким образом, фрактальная структура временных рядов криптовалюты и активности социальных сообществ и, как следствие, динамика их развития взаимосвязаны. Это указывает на необходимость при прогнозировании курсов криптовалют учитывать данные, полученные из анализа социальных сетей.

## **Литература**

---

- [Arnaboldi, 2015] Valerio Arnaboldi, Andrea Passarella, Marco Conti, Robin I.M. Dunbar, "Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs", Elsevier, 2015, p.116.
- [Bollen, 2011] Bollen, Johan, Huina Mao, and Xiaojun Zeng, "Twitter mood predicts the stock market." Journal of Computational Science, vol. 2.1, p.1-8, 2011.
- [Do, 2015] Do, Rotimi Opeke, James Webb, "CS224W Final Project: Predicting Yelp Ratings From Social Network Data", Department of Computer Science, Stanford University, p.8, December 9, 2015.
- [Feder, 1988] Feder J.: Fractals. Plenum, New York 1988.
- [Feng Mai, 2016] Feng Mai, Qing Bai, Zhe Shan, Xin (Shane) Wang, Roger H. L. Chiang, "From Bitcoin to Big Coin: The Impact of Social Media on Bitcoin Performance", 2016. Electronic copy available at: <http://ssrn.com/abstract=2545957>
- [Garcia, 2014] Garcia D, Tessone CJ, Mavrodiev P, Perony N, "The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy", J. R. Soc. Interface 11, 2014. Electronic copy available at: <http://dx.doi.org/10.1098/rsif.2014.0623>
- [Greaves, 2015] Alex Greaves, Benjamin Au, "Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin", Department of Computer Science, Stanford University, p.8, December 8, 2015.

- [Kantelhardt, 2002] J.W. Kantelhardt, E. Koscielny-Bunde, H.H.A. Rego, S. Havlin, A. Bunde, "Detecting long-range correlations with detrended fluctuation analysis", *Physica A.*, vol. 295, pp. 441-454, 2001.
- [Kantelhardt, 2008] J.W. Kantelhardt, "Fractal and multifractal time series", *Mathematics of complexity and dynamical systems*, pp. 463-487, 2012.
- [Matta, 2016] Martina Matta, "The Predictor Impact of Web Search and Social Media", Theses doctoral research. University of Cagliari, p.95, 2016.
- [Meléndez, 2014] Guillermo Romero Meléndez, "The fractal nature of bitcoin: evidence from wavelet power spectra", *Fundacion Universidad de las Americas Puebla*, 2014, p.39. Electronic copy available at: <http://ssrn.com/abstract=2547281>.
- [Mittal, 2013] A. Mittal, A. Goel, "Stock Prediction Using Twitter Sentiment Analysis" in proceeding of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, p.1-5, 2013.
- [Nakamoto, 2008] Nakamoto, Satoshi, "Bitcoin: A Peer-to-Peer Electronic Cash System", 31 October 2008. Electronic copy available at: <https://bitcoin.org/bitcoin.pdf>
- [Peng Xie, 2017], Hailiang Chen, Yu Jeffrey Hu, "Network Structure and Predictive Power of Social Media for the Bitcoin Market", *Georgia Tech Scheller College of Business Research Paper No. 17-5*, 2017. Electronic copy available at: <https://ssrn.com/abstract=2894089>.
- [Peters, 1994], Peters E.E.: *Fractal Market Analysis: applying chaos theory to investment and economics*. Wiley, 2003.
- [Riedi, 2002] R.H.Riedi, "Multifractal processes", in Doukhan P., Oppenheim G., Taqqu M.S. (Eds.), *Long Range Dependence: Theory and Applications*: Birkhuser, pp. 625–715, 2002.
- [Sho Tsugawa, 2014] Sho Tsugawa, Hiroyuki Ohsaki, "Emergence of Fractals in Social Networks: Analysis of Community Structure and Interaction Locality", in *38th Annual Computer Software and Applications Conference*, 2014.

---

### Информация об авторах

---



**Людмила Кириченко** – д.т.н., профессор Харьковского национального университета радиозлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: [lyudmyla.kirichenko@nure.ua](mailto:lyudmyla.kirichenko@nure.ua).

Основные области научных исследований: самоподобные и мультифрактальные временные ряды, фрактальный анализ, вейвлет-анализ, детерминированные хаотические системы.



**Тамара Радивилова** – к.т.н., доцент Харьковского национального университета радиоэлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: [tamara.radivilova@gmail.com](mailto:tamara.radivilova@gmail.com).

Основные области научных исследований: самоподобные и мультифрактальные временные ряды, телекоммуникационные системы, управление трафиком, информационная безопасность.



**Виталий Булах** - аспирант Харьковского национального университета радиоэлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: [bulakhvitalii@gmail.com](mailto:bulakhvitalii@gmail.com).

Основные области научных исследований: самоподобные и мультифрактальные временные ряды, интеллектуальный анализ временных рядов.



Вадим Чакрян -, к.т.н., ассистент Харьковского национального университета радиоэлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: [vadym.chakrian@nure.ua](mailto:vadym.chakrian@nure.ua)

Основные области научных исследований: менеджмент безопасности, телекоммуникационные сети и системы, риск менеджмент, методы противодействия кибер-атакам

## ANALYSIS OF INTERDEPENDENCE OF BITCOIN TIME SERIES AND TIME SERIES OF SOCIAL NETWORKS ACTIVITIES

**Lyudmyla Kirichenko, Tamara Radivilova, Vitalii Bulakh, Vadim Chakrian**

*Abstract. The comparative fractal and crosscorrelation analysis for community groups in the social network and the Bitcoin currency rate is carried out. The time series of daily data of likes and reposts for three social communities and the Bitcoin rate were formed. Time series crosscorrelation analysis revealed a significant correlation between the crypto currency rate and the social communities activity. Fractal analysis indicated the presence of multifractal properties, both for the series of Bitcoin prices, and for ones of community activities. The results have shown that the series having a significant correlation dependence have a similar multifractal structure.*

*Keywords: self-similar and multifractal time series, generalized Hurst exponent, crosscorrelation, bitcoin, social networks.*

## ФОРМИРОВАНИЕ ОБУЧАЮЩИХ ВЫБОРОК ДЛЯ СИНТЕЗА АДАПТИВНОЙ ЛОГИЧЕСКОЙ СЕТИ ТИПА «ТРАПЕЦЕИДАЛЬНАЯ МАТРИЦА»

Владимир Опанасенко, Сергей Крывый, Станислав Завьялов

**Аннотация:** Рассматривается задача формирования обучающих выборок адаптивной логической сети типа «трапецеидальная матрица» на основе универсальных логических элементов для реализации задачи классификации входного множества двоичных векторов. Трапецеидальная матрица представлена набором треугольных матриц с разными структурами связей.

**Ключевые слова:** адаптивная логическая сеть, булева функция, обучающая выборка, трапецеидальная матрица.

**ITHEA Keywords:** [1.5 Pattern Recognition](#).

---

### Введение

Для задач распознавания образов [Опанасенко, 17] важным этапом является процедура классификации произвольного входного множества двоичных векторов на заданных структурах. В данной работе такими структурами являются *адаптивные логические сети (АЛС)*, которые позволяют адаптировать структуры под заданный алгоритм функционирования [Опанасенко, 16А]. Использование кристаллов FPGA, представленных функциональным полем универсальных логических элементов (ЛЭ), позволило решить вопросы аппаратной реализации алгоритмов классификации путем определения конфигурации соответствующей структуры кристалла [Palagin, 17].

С точки зрения топологии системы АЛС представляет собой матрицу ЛЭ. *Универсальным ЛЭ* будем называть комбинационный автомат [Опанасенко, 14]:

$$L = \langle n, F \rangle ,$$

где:  $n$  – количество двоичных входов или размерность входных переменных ЛЭ;

$F = \{f_{\rho}\}$ ,  $\rho = [1 \div 2^{2^n}]$  – множество булевых функций, реализуемых ЛЭ.

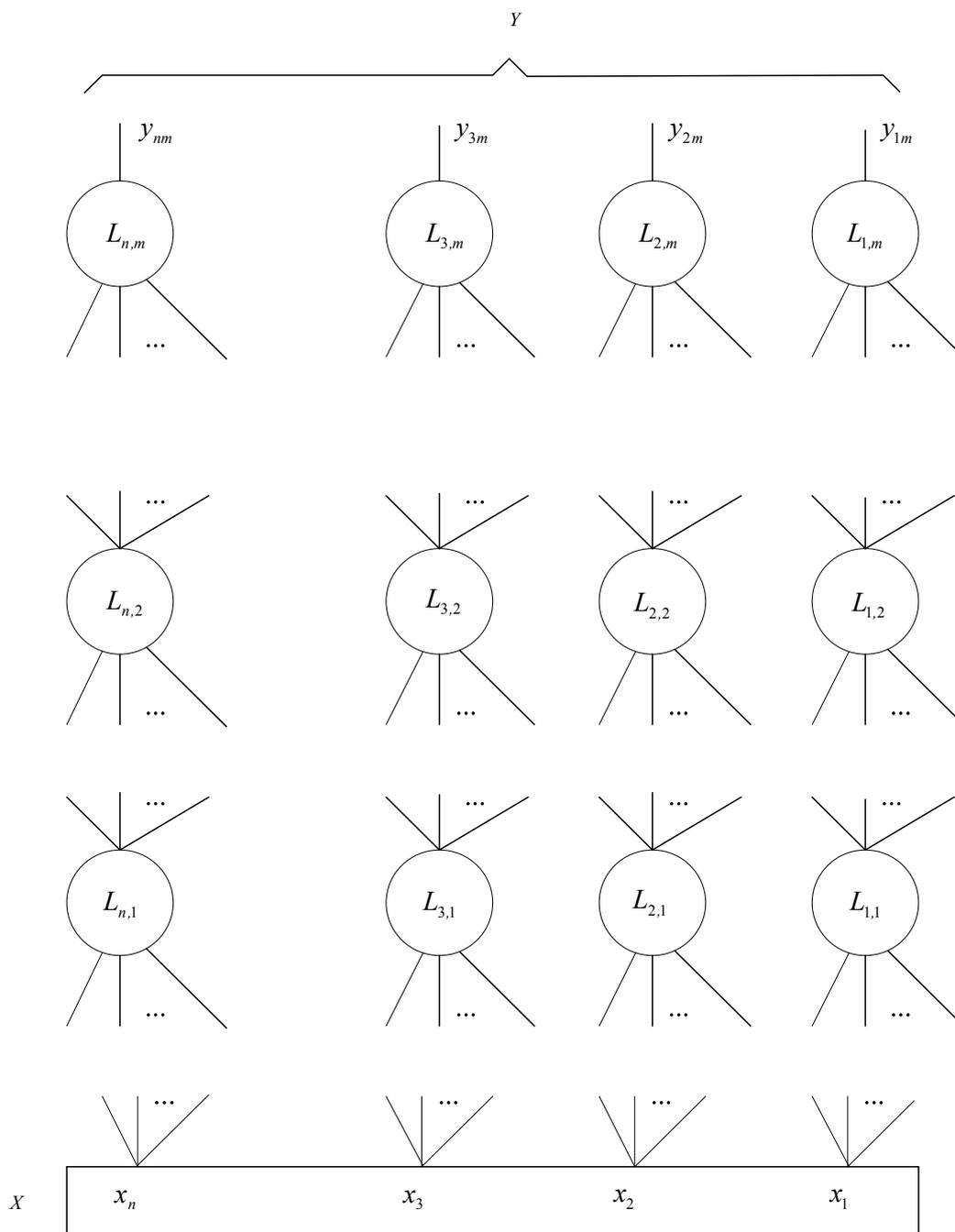
Универсальность ЛЭ заключается в возможности его настройки на реализацию произвольной булевой функции. Для случая ( $n = 2$ ) ЛЭ реализует одну из 16 логических функций, представляющих полный (базовый) набор функций двух переменных.

Введем некоторые определения.

*Вектор* – упорядоченное множество элементов, которые называются компонентами вектора.

Число компонент вектора определяет его длину (размерность).

Обобщенная структура АЛС представлена на рис. 1.



**Рис. 1.** Топология адаптивной логической сети

Структура АПС может быть описана следующей системой:

$$A = \langle n, h, F, S, L, m, D, X, Y \rangle, \quad (1)$$

где:

$n$  – разрядность входных двоичных векторов (размерность АПС по входу);

$h$  – выходная разрядность ( $h = \overline{1 \div n}$ ), размерность АПС по выходу;

$F = \{F_{ij}\}$  – множество логических функций системы;

$S$  – структура связей между ЛЭ;

$L = \{L_{ij}\}$  – множество ЛЭ ( $i$  – порядковый номер элемента ЛЭ;  $j$  – номер уровня обработки);

$m$  – количество уровней обработки;

$D = \{d\}$  – множество  $n$ -мерных двоичных векторов (обучающая выборка);

$X$  – полное множество входных двоичных векторов;

$Y = \{Y_{ij}\}$  – обобщенная функция системы,  $Y_{ij} = f_{ij}(Y_{v,(j-1)}, Y_{w,(j-1)})$  – значение функции  $f_{ij}$ , реализуемой элементом  $L_{ij}$ ,  $Y \in [0 \vee 1]$ , структура которого приведена на рис. 2 ( $v, w$  – значение индекса  $i$  для входов ЛЭ).

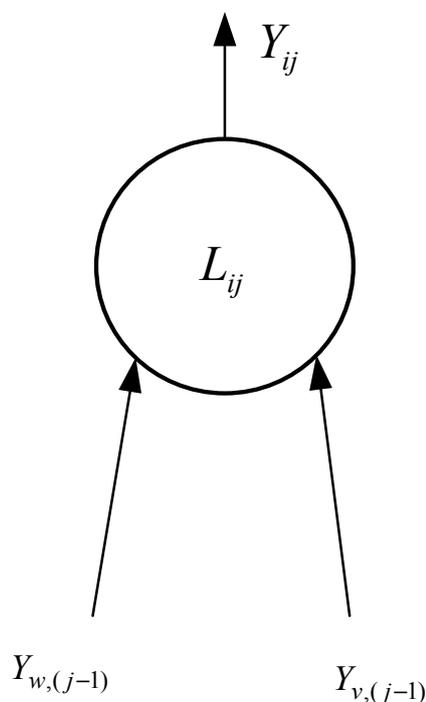


Рис. 2. Структура универсального ЛЭ

Каждый уровень АЛС представляет линейку  $\rho$ -входных ЛЭ, каждый из которых может быть настроен на выполнение любой из полного набора  $(2^{2^\rho})$  булевых функций его входных переменных и реализует отображения  $l$ -мерных ( $l \leq n$ ) двоичных векторов в  $u$ -мерные ( $l \geq u$ ) двоичные вектора.

Ограничимся рассмотрением двух типов АЛС, различаемых по топологическому признаку:

- ◆ структура типа «треугольная матрица» (ТМ), для которой  $h = 1$ ;
- ◆ структура типа «трапециевидальная матрица» (ТпМ), для которой ( $h = \overline{2 \div (n-1)}$ ), т.е. состоит из набора ТМ.

Задача синтеза логической сети сводится к задаче определения типов булевых функций для всех ЛЭ сети для заданной топологии сети по заданной обучающей выборке.

### 1. Структуры типа «треугольная матрица» (ТМ)

Функциональный блок типа ТМ предназначен для разбиения полного множества  $n$ -разрядных векторов  $X$  на два подмножества векторов  $X_1$  и  $X_2$ , заданных посредством обучающих выборок  $D_1$  и  $D_2$  ( $X_1 \cup X_2 = X$ ;  $X_1 \cap X_2 = \emptyset$ ;  $D_1 \subseteq X_1, D_2 \subseteq X_2$ ) [Ornassenko-16-2]. ТМ реализует отображение  $\mathfrak{S}: X \rightarrow Y$ ,  $Y \in [0 \vee 1]$ , такое что  $\mathfrak{S}: X_1 \rightarrow 1$ ,  $\mathfrak{S}: X_2 \rightarrow 0$  с учетом выполнения требуемых условий.

Задача настройки (адаптации) ТМ формулируется следующим образом. Пусть имеется полное множество  $n$ -мерных двоичных векторов  $X = \{x_p\}$ , где  $p = 1 \div 2^n$ . Задано подмножество  $D \subset X$ , которое является обучающей выборкой для алгоритма классификации. Для произвольного входного множества  $n$ -мерных двоичных векторов  $G = \{g\}$ , ( $G \subset X$ ) необходимо реализовать следующее отображение:

$$Y(g) = \begin{cases} 1, & \text{если } g \in D; \\ 0, & \text{если } g \notin D. \end{cases} \quad (2)$$

В общем случае задача адаптации структуры ТМ на реализацию функции (2) сводится к задаче построения универсального логического элемента произвольной разрядности на основе ЛЭ фиксированной разрядности и состоит в определении структуры связей  $S$  и типов логических функций  $f_{ij}$  для этих ЛЭ, что в совокупности реализует отображение  $\Psi: X \rightarrow Y$ . Для определения множества логических функций  $F = \{f_{ij}\}$  можно использовать известные подходы [Ornassenko, 15,16А], использующие полиномы Жегалкина для структурного описания ТМ. В

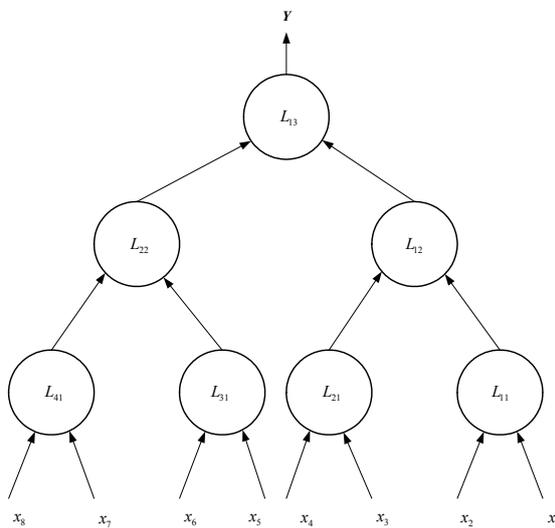
зависимости от структуры связей  $S$  в соответствии с (1) введем следующие определения типов структур ТМ.

**ТМ с логарифмической структурой связи**

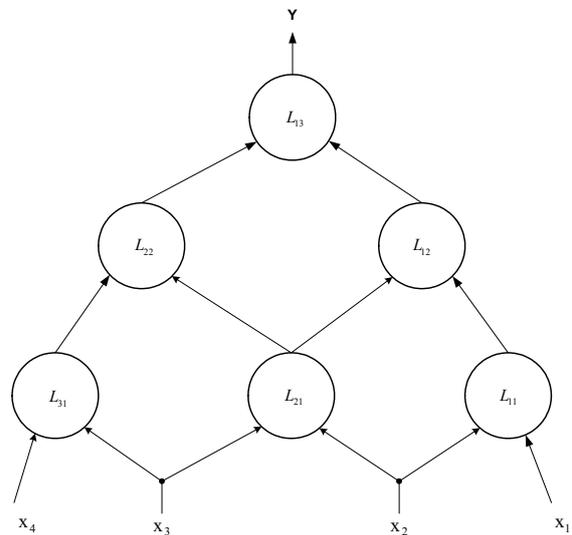
Структура ТМ (рис. 3) с логарифмической структурой связи (ЛСС) имеет следующие характеристики ( $n = 8$ ):

$$m = Ent\{\log_2 n\}; j = 1 \div \log_2 n, i = 1 \div (n/2^j);$$

$$w = (n/2^{j-1})^{i+1}; v = (n/2^{j-1})^i; Card\{L_{ij}\} = (n-1).$$



**Рис. 3.** Структура ТМ с логарифмической структурой связи



**Рис. 4.** Структура ТМ с сотовой структурой связи

**ТМ с сотовой структурой связи**

Структура ТМ (рис. 4) с сотовой структурой связи имеет следующие характеристики ( $n = 4$ ):

$$m = (n - 1); j = 1 \div (n - 1), i = 1 \div (n - j); v = i, w = (i + 1); Card\{L_{ij}\} = (n^2 - n)/2.$$

**2. Структуры типа «трапецидальная матрица»**

Для разбиения произвольного множества двоичных векторов на более чем два подмножества требуется выполнить усечение матрицы ТМ, глубина которого определяется числом

разделяемых подмножеств. В результате получаем «трапецеидальную матрицу» (ТпМ) [Oranasenko, 16С].

Рассмотрим далее синтез параметрического модуля в виде ТпМ для реализации задачи классификации или разбиения множества входных двоичных векторов на более чем два подмножества на основе универсальных ЛЭ. Структурно ТпМ состоит из  $z$  пересекающихся составных ТМ и имеет  $n$  входов и  $h$  выходов. Для каждой ТМ необходимо определить свою обучающую выборку, которая будет напрямую зависеть от топологии матрицы.

Архитектура ТпМ описывается системой (1), где:

$D = \{ D^\gamma \}$  – множество прообразов ( $\bigcap_\gamma D^\gamma = \emptyset, \forall \gamma = 1 \div k$ ), в котором каждый прообраз

$D^\gamma = \{ D_\sigma^\gamma \}$  задан множеством двоичных векторов  $D_\sigma^\gamma$  ( $\forall \sigma = 1 \div \phi_\gamma$ , где:

$\phi_\gamma$  – мощность множества  $D_\sigma^\gamma$ ). Вектор  $D_\sigma^\gamma = (d_{\sigma,i}^\gamma)$  является  $n$ -мерным двоичным вектором ( $i = 1 \div n; d_{\sigma,i}^\gamma \in \{ 0, 1 \}$ );

$h$  – выходная размерность ТпМ (разрядность выходных двоичных векторов  $y \in Y$ ), которая определяется величиной  $h = (\text{Ent} \{ \log_2 k \} + 1)$ , так как входные векторы, не относящиеся к  $\gamma$ -му образу, объединяются в одно кодовое представление.

Множеству двоичных векторов  $D_\sigma^\gamma$  прообраза  $D^\gamma$  ставится в соответствие образ  $G^\gamma$  ( $\bigcap_\gamma G^\gamma = \emptyset$ ). Образ  $G^\gamma = (g_z^\gamma)$  представляется двоичным вектором ( $z = 1 \div h$ ). Каждому образу

$G^\gamma$  соответствует заданный выходной двоичный вектор  $Y^\gamma = (y_z^\gamma)$ , ( $\bigcap_\gamma Y^\gamma = \emptyset, Y^\gamma \in Y$ ).

Сформулируем постановку задачи классификации для ТпМ.

Необходимо реализовать посредством структуры ТпМ отображение  $\mathfrak{X}: D \rightarrow G$  ( $G^\gamma = \mathfrak{S}(D^\gamma)$ ).

Выходами ТпМ является двоичный вектор  $y_z$ , причем  $\forall \gamma \mathfrak{X}: D^\gamma \rightarrow Y^\gamma$ .

Таким образом, алгоритм решения задачи классификации разбивается на следующие шаги:

- разбить ТпМ на  $z$  составляющих ТМ;
- для каждой ТМ определить множество обучающих выборок;
- на основе анализа обучающих выборок и соответствующих значений компонент выходного двоичного вектора выполнить синтез ТМ (определить множество логических функций для ЛЭ, которые в совокупности реализуют поставленную задачу классификации).

Представим обучающую выборку (прообраз)  $D^\gamma$  в виде матрицы:  $D^\gamma = \left\| d_{\sigma,i}^\gamma \right\|$  или

$$D^\gamma = \left\| \begin{array}{cccccc} d_{1,n}^\gamma & \dots & d_{1,i}^\gamma & \dots & d_{1,2}^\gamma & d_{1,1}^\gamma \\ d_{2,n}^\gamma & \dots & d_{2,i}^\gamma & \dots & d_{2,2}^\gamma & d_{2,1}^\gamma \\ \vdots & \dots & \vdots & \dots & \vdots & \vdots \\ d_{\sigma,n}^\gamma & \dots & d_{\sigma,i}^\gamma & \dots & d_{\sigma,2}^\gamma & d_{\sigma,1}^\gamma \end{array} \right\|. \quad (3)$$

В связи с тем, что при представлении двоичного вектора предполагается нумерация индекса  $i$  справа–налево (в сторону старших двоичных компонент вектора), то, в отличие от общепринятого представления матрицы, столбцы матрицы  $D^\gamma$  нумеруются справа–налево (в сторону старших значений индекса столбца).

Таким образом, в общем случае мы имеем  $k$  матриц  $D^\gamma$ , содержащих одинаковое количество столбцов, определяемое размерностью двоичных векторов обучающей выборки, и переменное количество строк, определяющее мощность  $\gamma$ -ой выборки.

При разбиении ТпМ на составляющие ТМ, каждая  $z$ -ая ТМ будет оперировать с  $i$ -ми ( $i = i_1(z) \div i_2(z)$ ) компонентами входных двоичных векторов, где  $i_1(z)$  и  $i_2(z)$  – соответственно верхняя и нижняя границы значений индекса  $i$  для  $z$ -ой ТМ при произвольной структуре связей ЛЭ. Представим обучающую выборку (матрицу (3)) в общем виде для произвольной  $z$ -ой ТМ ( $D_z = \{D_z^\gamma\}$ ):

$$D_z^\gamma = \left\| \begin{array}{cccccc} d_{1,i_2(z)}^\gamma & d_{1,(i_2(z)-1)}^\gamma & \dots & d_{1,i}^\gamma & \dots & d_{1,i_1(z)}^\gamma \\ d_{2,i_2(z)}^\gamma & d_{2,(i_2(z)-1)}^\gamma & \dots & d_{2,i}^\gamma & \dots & d_{2,i_1(z)}^\gamma \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{\sigma,i_2(z)}^\gamma & d_{\sigma,(i_2(z)-1)}^\gamma & \dots & d_{\sigma,i}^\gamma & \dots & d_{\sigma,i_1(z)}^\gamma \end{array} \right\|. \quad (4)$$

При этом обучающая выборка  $D_z$  для  $z$ -ой ТМ формируется на основе матриц (4) по правилам:

$$D_z = \bigcup_{\gamma} D_z^\gamma = \bigcup_{\gamma, \sigma} (d_{\sigma,i}^\gamma / i = i_1(z) \div i_2(z)), \quad (5)$$

$$\text{Card}(D_z) \leq \sum_{\gamma} (\text{Card}(D_z^\gamma)). \quad (6)$$

С учетом (5),( 6), матрицу (4) представим в виде

$$D_z = \begin{pmatrix} d_{1,i_2(z)} & d_{1,(i_2(z)-1)} & \dots & d_{1,i} & \dots & d_{1,i_1(z)} \\ d_{2,i_2(z)} & d_{2,(i_2(z)-1)} & \dots & d_{2,i} & \dots & d_{2,i_1(z)} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{\sigma,i_2(z)} & d_{\sigma,(i_2(z)-1)} & \dots & d_{\sigma,i} & \dots & d_{\sigma,i_1(z)} \end{pmatrix}. \quad (7)$$

При этом,  $\sigma = 1 \div \phi_z$ , где  $\phi_z = \sum_{\gamma} \phi_{\gamma}$ .

Объединение множеств (обучающих выборок  $D_z^{\gamma}$ ) в (5) выполняется для  $\gamma$ -ых множеств, удовлетворяющих условию

$$\bigcup_{\gamma} g_z^{\gamma} = 1. \quad (8)$$

ТпМ представляется совокупностью ТМ, выходом каждой из которых является компонента  $y_z$  вектора  $Y$ , а на вход поступают компоненты с индексом  $i$  ( $i = i_1(z) \div i_2(z)$ ) входного двоичного вектора  $x_z$ .  $Z$ -ая ТМ реализует преобразование  $\mathfrak{R}_z : (x_z) \rightarrow y_z$ . ТМ реализует разбиение произвольного множества двоичных векторов  $X = (x_z)$  на два подмножества при заданном множестве прообразов  $D_z \subset X_z$  ( $D_z$  – обучающая выборка). Таким образом, выход  $y_z$  определяется следующим образом:

$$y_z = \begin{cases} 1, & \text{если } x_z \in D_z; \\ 0, & \text{в противном случае.} \end{cases} \quad (9)$$

## 2.1. ТпМ с сотовой структурой связи

ТпМ с сотовой структурой связи, в соответствие с (3.1), характеризуется следующими параметрами:

- количество уровней ( $m$ ) определяется величиной ( $m = n - h$ );
- количество входов (двоичных)  $z$ -ой ТМ определяется величиной ( $n - h + 1$ );
- $v = i$ ,  $w = (i + 1)$ ;
- количество пересекаемых ТМ определяется значением величины  $\delta = (h - m) = (2h - n)$  и представляется рядом: 2, 3, ...,  $R$  ( $R = h$ , если  $\delta \leq 0$  и  $R = (h - \delta)$ , если  $\delta > 0$ );
- количество уровней ( $j = 1 \div s$ ) в ТпМ, имеющих логические элементы  $L_{i,j}$ , которые принадлежат одновременно всем ТМ, определяется значением  $\delta$ :

если  $\delta \leq 0$ , то  $s = |\delta| + 1$ ; если  $\delta > 0$ , то  $s = \delta$ ;

– количество ЛЭ на  $j$ -ом уровне ТпМ определяется величиной  $N_j = (n - j)$ . Общее число ЛЭ в ТпМ равно  $N = (n + h - 1)(n - h) / 2$ ;

– индексы ( $i$ ) элементов входного вектора  $z$ -ой ТМ находятся в диапазоне ( $i_1(z) = z$ ;  $i_2(z) = (z + n - h)$ ).

При разбиении ТпМ на составляющие ТМ каждая  $z$ -ая ТМ будет оперировать с  $i$ -ми ( $i = z \div (z + n - h)$ ) компонентами входных двоичных векторов. Поэтому представим матрицу (3)

в общем виде для произвольной  $z$ -ой ТМ ( $D_z = \{D_z^\gamma\}$ ):

$$D_z^\gamma = \begin{pmatrix} d_{1,(z+n-h)}^\gamma & d_{1,(z+n-h-1)}^\gamma & \dots & d_{1,i}^\gamma & \dots & d_{1,z}^\gamma \\ d_{2,(z+n-h)}^\gamma & d_{2,(z+n-h-1)}^\gamma & \dots & d_{2,i}^\gamma & \dots & d_{2,z}^\gamma \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{\sigma,(z+n-h)}^\gamma & d_{\sigma,(z+n-h-1)}^\gamma & \dots & d_{\sigma,i}^\gamma & \dots & d_{\sigma,z}^\gamma \end{pmatrix}. \quad (10)$$

Обучающая выборка (8) для  $z$ -ой ТМ формируется на основе матриц (10) по правилам (5, 6) и имеет вид:

$$D_z = \begin{pmatrix} d_{1,(z+n-h)} & d_{1,(z+n-h-1)} & \dots & d_{1,i} & \dots & d_{1,z} \\ d_{2,(z+n-h)} & d_{2,(z+n-h-1)} & \dots & d_{2,i} & \dots & d_{2,z} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{\sigma,(z+n-h)} & d_{\sigma,(z+n-h-1)} & \dots & d_{\sigma,i} & \dots & d_{\sigma,z} \end{pmatrix}. \quad (11)$$

Объединение множеств (обучающих выборок  $D_z^\gamma$ ) в (11) выполняется для  $\gamma$ -ых множеств, удовлетворяющих условию (8). Для произвольного входного двоичного вектора выход  $y_z$  определяется выражением (9).

Для трапецеидальной ТМ с сотовой структурой связи каждый логический элемент  $L_{i,j}$  необходимо настроить на одну из логических функций множества  $F = \{f_{ij}\}$ . ТпМ представляется множеством пересекаемых ТМ, поэтому логические элементы  $L_{i,j}$  с одинаковыми индексами ( $i$ ) для  $m$ -го уровня, принадлежащие одновременно двум и более ТМ должны настраиваться на один и тот же тип логической функции.

Логический элемент  $L_{i,j}$  в общем виде для  $j$ -го уровня ТпМ будет иметь значения индекса  $i$  для  $z$ -ой ТМ в диапазоне  $-(z \leq i \leq (z+m-j))$ .

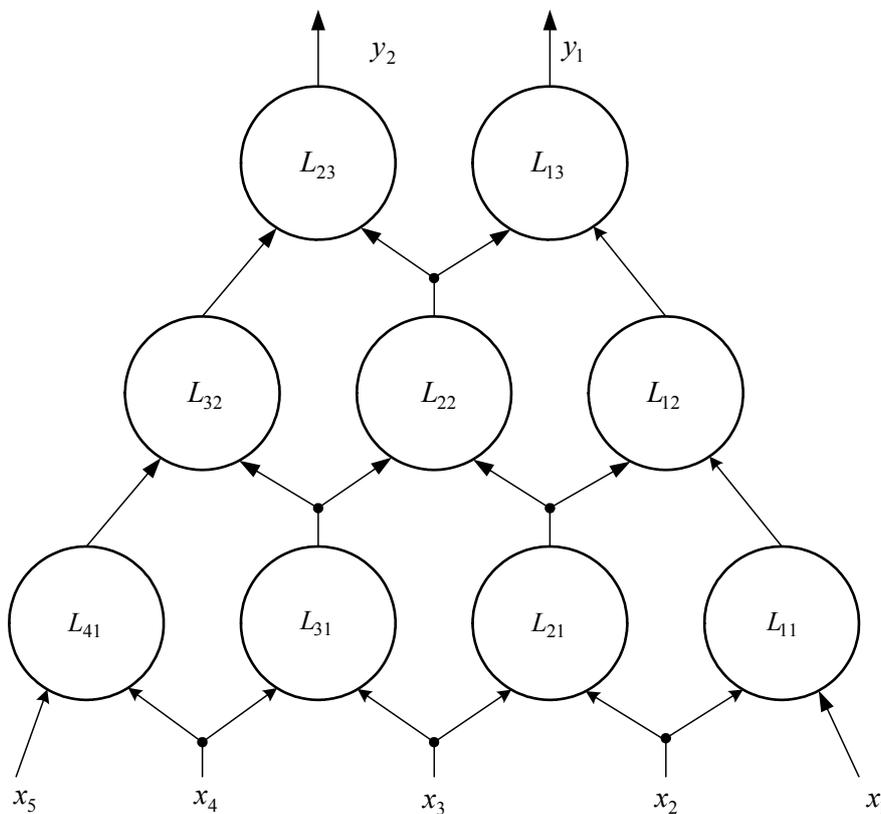
Таким образом, принадлежность логического элемента  $L_{i,j}$  к  $z$ -ой ТМ будет определяться корректностью ограничения вида:

$$0 \leq (i-z) \leq (m-j).$$

При определении типов логических функций для ТпМ последовательно определяются типы логических функций для каждой  $z$ -ой ТМ с учетом пересечений ЛЭ.

**Пример 1.** Рассмотрим описание ТпМ с сотовой структурой связи для  $n=5, h=2$  (рис. 5). В качестве обучающей выборки имеется три множества прообразов:

$$D^1 = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}; D^2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}; D^3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$



**Рис. 5.** ТпМ с сотовой структурой связи

Пусть образы  $G$  кодируются следующим образом:

$$G^y = \{g_z^y\}.$$

$$G^1 = \{g_1^1 = 1, g_2^1 = 0\}; G^2 = \{1, 0\}; y_1 = 1; y_2 = 0.$$

$$G^2 = \{g_1^2 = 0, g_2^2 = 1\}; G^3 = \{0, 1\}; y_1 = 0; y_2 = 1.$$

$$G^3 = \{g_1^3 = 1, g_2^3 = 1\}; G^4 = \{1, 1\}; y_1 = 1; y_2 = 1.$$

Так как мы имеем две ТМ, то каждая из них оперирует своими обучающими выборками.

Первая ТМ оперирует компонентами входного вектора, имеющего значения индекса ( $i$ ) от 1 до 4:

$$D_1^1 = \left\| \begin{array}{cccc} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{array} \right\|; D_1^2 = \left\| \begin{array}{cccc} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{array} \right\|; D_1^3 = \left\| \begin{array}{cccc} 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array} \right\|.$$

Вторая ТМ оперирует компонентами входного вектора, имеющего значения индекса ( $i$ ) от 2 до 5:

$$D_2^1 = \left\| \begin{array}{cccc} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{array} \right\|; D_2^2 = \left\| \begin{array}{cccc} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{array} \right\|; D_2^3 = \left\| \begin{array}{cccc} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{array} \right\|.$$

На основе матриц  $D_z^y$  по правилам (5), (6) формируются обучающие выборки для первой и второй ТМ.

Для первой ТМ, с учетом  $g_1^1 = g_1^3 = 1$ , объединяем  $D_1^1$  и  $D_1^3$ :

$$D_1 = \begin{vmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{vmatrix}.$$

Для второй ТМ, с учетом  $g_2^2 = g_2^3 = 1$ , объединяем  $D_2^2$  и  $D_2^3$ :

$$D_2 = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{vmatrix}.$$

Алгоритмы синтеза ТМ подробно рассмотрены в [Oranassenko, 15].

## 2.2. ТпМ с логарифмической структурой связи

Если ТпМ имеет логарифмическую структуру связи, то пересечение составных ТМ отсутствуют и составные ТМ реализуют преобразование данных независимо, а ТпМ описывается следующими параметрами:

– количество уровней ( $m$ ) определяется величиной:

$$m = \log_2 n - \log_2 h = \log_2(n/h);$$

– количество входов (двоичных)  $z$ -ой ТМ определяется величиной  $k = 2^m$ ;

– структура связей определяется значениями

$$v = (n/2^{j-1})^i; w = (n/2^{j-1})^{i+1};$$

– количество ЛЭ на  $j$ -м уровне ТпМ определяется величиной  $N_j = n / (2^j)$ , а общее число ЛЭ для ТпМ с логарифмической структурой связи равно  $N = (n - h)$ ;

–  $z$ -я ТМ оперирует с  $i$ -ми информационными компонентами (разрядами)  $n$ -мерного входного вектора –  $i_1(z) = ((z-1) \times 2^m + 1)$ ,  $i_2(z) = (z \times 2^m)$ .

Представим обучающую выборку (матрицу (4)) в общем виде для произвольной  $z$ -й ТМ ( $D_z = \{ D_z^\gamma \}$ ) с логарифмической структурой связи:

$$D_z^\gamma = \begin{pmatrix} d_{1,((z)2^m)}^\gamma & d_{1,((z)2^m-1)}^\gamma & \dots & d_{1,i}^\gamma & \dots & d_{1,((z-1)2^m)}^\gamma \\ d_{2,((z)2^m)}^\gamma & d_{2,((z)2^m-1)}^\gamma & \dots & d_{2,i}^\gamma & \dots & d_{2,((z-1)2^m)}^\gamma \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{\sigma,((z)2^m)}^\gamma & d_{\sigma,((z)2^m-1)}^\gamma & \dots & d_{\sigma,i}^\gamma & \dots & d_{\sigma,((z-1)2^m)}^\gamma \end{pmatrix}. \quad (12)$$

При этом обучающая выборка  $D_z$  для  $z$ -й ТМ формируется на основе матриц (12) по правилам (5, 6):

$$D_z = \begin{pmatrix} d_{1,(z2^m)} & d_{1,(z2^m-1)} & \dots & d_{1,i} & \dots & d_{1,((z-1)2^m)} \\ d_{2,(z2^m)} & d_{2,(z2^m-1)} & \dots & d_{2,i} & \dots & d_{2,((z-1)2^m)} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{\sigma,(z2^m)} & d_{\sigma,(z2^m-1)} & \dots & d_{\sigma,i} & \dots & d_{\sigma,((z-1)2^m)} \end{pmatrix}. \quad (13)$$

Объединение множеств (обучающих выборок  $D_z^\gamma$ ) в (5) выполняется для  $\gamma$ -х множеств, удовлетворяющих условию (8).

Для произвольного входного двоичного вектора, при заданном множестве (13) преобразов  $D_z$  ( $D_z$  – обучающая выборка), выход  $y_z$  определяется выражением (9).

**Пример 2.** Рассмотрим описание ТпМ с логарифмической структурой связи для  $n = 8$ ,  $h = 2$  (рис. 6). Пересечение составных ТМ отсутствует.

В качестве исходных данных имеем три множества преобразов:

$$D^1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix};$$

$$D^2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix};$$

$$D^3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

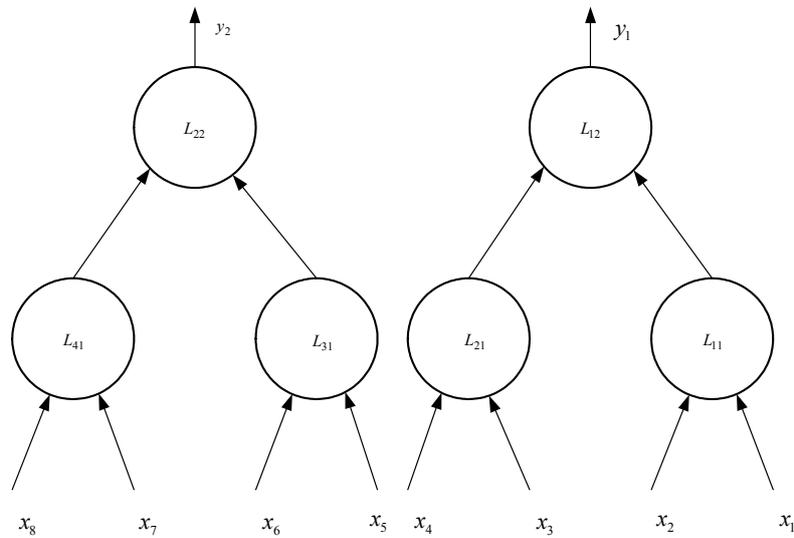


Рис. 6. ТпМ с логарифмической структурой связи

Пусть образы  $G$  кодируются следующим образом:

$$G^y = \{g_z^y\}.$$

$$G^1 = \{g_1^1 = 1, g_2^1 = 0\}; G^1 = \{1, 0\}; y_1 = 1; y_2 = 0.$$

$$G^2 = \{g_1^2 = 0, g_2^2 = 1\}; G^2 = \{0, 1\}; y_1 = 0; y_2 = 1.$$

$$G^3 = \{g_1^3 = 1, g_2^3 = 1\}; G^3 = \{1, 1\}; y_1 = 1; y_2 = 1.$$

Так как имеется две ТМ, то каждая из них оперирует своими обучающими выборками. Первая ТМ оперирует компонентами входного вектора с индексами  $i = 1 \div 4$ , вторая – ( $i = 2 \div 8$ ).

На основе матриц  $D_z^y$  по правилам (5), (6) формируются обучающие выборки для первой и второй ТМ.

Для первой ТМ, с учетом  $g_1^1 = g_1^3 = 1$ , объединяем  $D_1^1$  и  $D_1^3$ :

$$D_1 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Для второй ТМ, с учетом  $g_2^2 = g_2^3 = 1$ , объединяем  $D_2^2$  и  $D_2^3$ :

$$D_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Алгоритмы синтеза ТМ подробно рассмотрены в [Опанасенко, 15].

---

### Заключение

Разбиение произвольного множества двоичных векторов на более чем два подмножества требует усечения матрицы ТМ, глубина которого определяется числом разделяемых подмножеств. В результате получается «трапецеидальная матрица» (ТпМ).

Структурно ТпМ состоит из  $z$  пересекающихся составных ТМ, поэтому для каждой составной ТМ определена своя обучающая выборка в зависимости от топологии матрицы. Таким образом, в работе формализована процедура разбиения заданной обучающей выборки на фрагменты обучающих выборок для каждой составной ТМ, которые являются исходными данными для синтеза ТМ и ТпМ в целом.

---

### Библиография

- [Опанасенко, 14] Опанасенко В.Н., Крывый С.Л. Прямая задача синтеза адаптивных логических сетей. - International Journal "Information Technologies & Knowledge". – Изд-во: ИТЕА, София, Болгария – 2014, Vol. 8, N.1. – pp. 3–12.
- [Опанасенко, 15] Опанасенко V.N., Крыви S.L. Synthesis of Adaptive Logical Networks on the Basis of Zhegalkin Polynomials. - Cybernetics and Systems Analysis. Springer New York. – 2015, Vol. 51, N.6. – pp. 969–977. DOI: [10.1007/s10559-015-9790-1](https://doi.org/10.1007/s10559-015-9790-1).
- [Опанасенко, 16A] Опанасенко V., Крыви S. Method synthesis of the configurable logical blocks on basis of universal logical elements. - Radioelectronic and Computer Systems. KHAI Kharkiv. – 2016, Vol. 79, N.5. – pp. 93–97.
- [Опанасенко, 16B] Опанасенко V., Крыви S. Algorithms synthesis of the adaptive logical network on basis of universal logical elements. - Proceeding of 13<sup>th</sup> International Conference on Pattern recognition and information processing. – Minsk, Belarus. 3–5 October 2016. – pp.16–20.

- [Opanasenko, 16C] Opanasenko V., Kryvyi S. Synthesis of multilevel structures with multiple outputs. - CEUR Workshop Proceeding of 10<sup>th</sup> International Conference of Programming, UkrPROG 2016; Kyiv; Ukraine; 24 May 2016 . Volume 1631. – Code 122904. – pp.32–37.
- [Opanasenko, 17A] Opanasenko V.N., Kryvyi S.L. Synthesis of Neural-Like Networks on the Basis of Conversion of Cyclic Hamming Codes. - Cybernetics and Systems Analysis. Springer New York. – 2017, Vol. 53, N.4. – pp. 627–635. DOI: 10.1007/s10559-017-9965-z.
- [Palagin, 17B] Palagin A.V., Opanasenko V.N., Kryvyi S.L. Resource and Energy Optimization Oriented Development of FPGA-Based Adaptive Logical Networks for Classification Problem. [In book: Green IT Engineering: Components, Networks and Systems Implementation, V. Kharchenko, Y. Kondratenko, J. Kacprzyk \(Eds.\), Vol. 105. Berlin, Heidelberg: Springer International Publishing, pp.195-218 \(2017\), DOI: DOI 10.1007/978-3-319-55595-9\\_10.](#)

---

#### Сведения об авторах

---

**Опанасенко Владимир Николаевич** – профессор, доктор технических наук Института кибернетики им. В.М. Глушкова НАН Украины, Украина, Киев, 03187, просп. Глушкова, 40;

**e-mail:** [opanasenkovm@nas.gov.ua](mailto:opanasenkovm@nas.gov.ua)

**Крывий Сергей Лукьянович** – профессор, доктор физико-математических наук, профессор Киевского национального университета им. Тараса Шевченко, Украина, Киев, 03187, просп.

Глушкова, 4д, Факультет кибернетики; **e-mail:** [krivoi@i.com.ua](mailto:krivoi@i.com.ua)

**Завьялов Станислав Борисович** – кандидат технических наук, директор ООО «Радионикс», Украина, Киев; **e-mail:** [radionix13@gmail.com](mailto:radionix13@gmail.com).

#### Forming of learning samples for a synthesis of adaptive logic network type "Trapezoidal Matrix"

**Volodymyr Opanasenko, Sergei Kryvyi, Stanislaw Zavyalov**

**Abstract:** *The problem of forming of learning samples of an adaptive logical network of a "trapezoidal matrix" type is considered on the basis of universal logic elements for the implementation of the problem of classifying the input set of binary vectors. The trapezoidal matrix is represented by a set of triangular matrices with different topologies of structures.*

**Keywords:** *adaptive logical network, Boolean function, learning sample, trapezoidal matrix.*

## ОСНОВНЫЕ КОНЦЕПЦИИ РЕСУРСНОГО ТЕОРЕТИКО-ИГРОВОГО ПОДХОДА К МОДЕЛИРОВАНИЮ СБАЛАНСИРОВАННОГО РАЗВИТИЯ

Сергей Полумиенко, Сергей Горда

**Аннотация.** В работе предлагается объединение ресурсного и теоретико-игрового подходов для построения модели сбалансированного развития в виде кооперативной игры. Модель отображает взаимодействие участников исходной системы регионального или национального уровня и позволяет определить критерии оптимальности ее развития и найти соответствующую стратегию, удовлетворяющую разносторонние интересы участников системы.

**Ключевые слова:** ресурсный подход, кооперативная модель, сбалансированное развитие

**ITHEA Keywords:** I.6.1 Simulation Theory, I.6.5 Model Development, K.4.1 Public Policy Issues

---

### Введение

Проблема устойчивого, а главное, сбалансированного развития различных сфер жизнедеятельности человека, несмотря на свою более, чем 20-летнюю историю со времени принятия ООН «Повестки дня XXI столетия», так и не нашла своего решения, несмотря на существенное количество исследований. На то есть разные причины, которые отмечаются и Департаментом устойчивого развития ООН [United Nations, 2013], и многими другими авторами, например, [Bossel, 1999].

Одной из основных причин является отсутствие методов анализа всей многогранности общества, среды его обитания и развития, прежде всего, учета и сбалансирования интересов и действий всех его членов. В основном исследования и полученные решения направлены или на какую-либо отрасль деятельности, или на определенную задачу. Например, «Бедность» – является одним из разделов системы индикаторов устойчивости развития ООН, но направлена, прежде всего, на оценку уровня жизни населения разных стран и подходов к ее улучшению, но не затрагивает основы и причины устойчивости развития, что требует анализа экономических, природных и иных предпосылок. Для решения проблемы сбалансированности развития необходимы методы нахождения компромиссных решений, относящиеся к сфере достаточно сложных теоретико-игровых и близких к ним моделей, исследования которых также не получили

полноценного применения, а используются только в отдельных задачах общей проблемы устойчивого развития общества.

Данная публикация продолжает работы [Polumiienko, Rybakov, Trofymchuk, 2013], [Полумієнко, Рибаків, 2015], [Полумієнко, Горда, 2016], направленные на нахождение конструктивно достижимой концепции дальнейшего развития общества с учетом всех составляющих этого процесса. В основе исследования - определение и сравнительный анализ имеющихся и перспективных региональных ресурсов в совокупности всех национальных ресурсов и использование теоретико-игрового подхода, а именно, теории кооперативных игр для определения модели сбалансированного развития.

---

### **Индикаторы и агрегаты состояния ресурсов**

---

В качестве исходной системы выберем определенный регион страны. Состояние системы будем описывать на основе совокупности оценок, отображающих в определенном смысле статические характеристики ее элементов, которые рассматриваются как ресурсы, необходимые для обеспечения жизнедеятельности и развития - природные, социальные; экономические и другие ресурсы. Участники системы, исходя из наличия контролируемых ими ресурсов, выполняют свои действия, которые состоят в изменении - использовании имеющихся и создании новых ресурсов, что отображает динамику системы.

Для описания ресурсов вводятся показатели или индикаторы, которые отражают количественные оценки объема ресурса, а также его качественные признаки. Ресурсы региона рассматриваются как относительные величины - доля их объема в совокупном объеме по всей стране. Каждый из ресурсов описывается независимой от других ресурсов величиной, которая не является определенным отношением или функцией, например, не является традиционной величиной валового внутреннего продукта на душу населения, которая есть результатом расчетов. Перечень ресурсов и их оценок основывается на эффективно достижимой, регулярной и достоверной информации о них.

Исходя из этих условий, определяется перечень категорий ресурсов (табл. 1), каждая из которых включает совокупность видов ресурсов. В частности, категория природных ресурсов включает такие виды: земельные ресурсы; водные ресурсы; атмосфера; полезные ископаемые; лесные ресурсы; животный и растительный мир; отходы и опасные химические вещества; финансирование мероприятий по реновации.

Земельные ресурсы, в частности, описываются индикаторами приведенными в таблице 2.

**Таблица 1. Индикаторы, агрегаты и весовые коэффициенты по категориям и видам ресурсов**

Категории ресурсов (количество индикаторов)	Агрегаты	Весовые коэффициенты видов и категорий ресурсов	
		$u_{1,1} - u_{1,8}$	$w_1$
Природные ресурсы (76)	$Aq_{1,1} - Aq_{1,8}$	$u_{1,1} - u_{1,8}$	$w_1$
Социальные ресурсы (95)	$Aq_{2,1} - Aq_{2,8}$	$u_{2,1} - u_{2,8}$	$w_2$
Экономические ресурсы (100)	$Aq_{3,1} - Aq_{3,21}$	$u_{3,1} - u_{3,21}$	$w_3$
Ресурсы государственной и региональной власти (31)	$Aq_{4,1} - Aq_{4,8}$	$u_{4,1} - u_{4,8}$	$w_4$

**Таблица 2. Индикаторы состояния земельных ресурсов**

Земельные ресурсы (вид 1 категории 1)	$Aq_{1,1}$
Площадь	$r_{1,1,1}$
<i>Земли в естественном состоянии</i>	
Леса и другие покрытые лесом площади	$r_{1,1,2}$
Заповедные земли, заказники, парки, сады и т.п.	$r_{1,1,3}$
Непригодные для использования в сельском хозяйстве земли, возможные для лесоразведения	$r_{1,1,4}$
<i>Площадь территорий, покрытых поверхностными водами</i>	$r_{1,1,5}$
<i>Земли хозяйствования</i>	
Сельскохозяйственные угодья	$r_{1,1,6}$
Застроенные земли	$r_{1,1,7}$
Площадь военных объектов	$r_{1,1,8}$
Площадь оборонительных объектов	$r_{1,1,9}$
Площадь опасных объектов	$r_{1,1,10}$
Площадь разрушенных и опасных объектов и территорий вследствие военных конфликтов, терактов	$r_{1,1,11}$
Площадь месторождений горючих полезных ископаемых	$r_{1,1,12}$
Площадь месторождений металлических полезных ископаемых	$r_{1,1,13}$
Площадь месторождений неметаллических полезных ископаемых	$r_{1,1,14}$
Земли без или с незначительным растительным покровом	$r_{1,1,15}$
<i>Непригодные для использования или загрязненные земли</i>	
Площадь отвалов месторождений горючих полезных ископаемых	$r_{1,1,16}$
Площадь отвалов месторождений металлических полезных ископаемых	$r_{1,1,17}$
Площадь отвалов месторождений неметаллических полезных ископаемых	$r_{1,1,18}$
Площадь загрязнения, в том числе для организованного складирования твердых бытовых отходов	$r_{1,1,19}$
Площади составов, непригодных и запрещенных к использованию веществ и земель	$r_{1,1,20}$
Площадь мест неорганизованного складирования отходов	$r_{1,1,21}$
Критические земли	$r_{1,1,22}$
Открытые заболоченные земли	$r_{1,1,23}$
Другие земли	$r_{1,1,24}$
Длина распространения оползней	$r_{1,1,25}$
Площадь распространения подтопления	$r_{1,1,26}$

Будем обозначать через  $n$  – определенный регион страны, через  $l, l=1, \dots, 4$ , - категории, а через  $k, k=1, \dots, K$ , - виды ресурсов. Положим, что  $r_{n,l,k}$  - относительный объем ресурса вида  $k$  и категории  $l$  в регионе  $n$  ( $n, l, k$ -ресурс), -

$$r_{n,l,k} = \text{res}_{n,l,k} / \sum_n \text{res}_{n,l,k} .$$

где  $\text{res}_{n,l,k}$  - величина натурального объема  $n, l, k$ -ресурса.

На основе физических и других свойств видов ресурсов образуются агрегаты - их агрегированные количественные оценки  $A_{n,l,k}$ . В частности, в случае земельных ресурсов  $A_{n,l,k}$  имеем (для сокращения опущено индекс региона  $n$ ):

$$\begin{aligned} A_{1,1} &= a_{1,1,1} * (1 - a_{1,1,2}), \\ a_{1,1,1} &= (r_{1,1,2} + r_{1,1,3} + r_{1,1,4}) / r_{1,1,1}, \\ a_{1,1,2} &= (r_{1,1,6} + \dots + r_{1,1,24}) / ((r_{1,1,1} - r_{1,1,5}) * (1 - r_{1,1,25}) * (1 - r_{1,1,26})), \end{aligned}$$

где  $a_{1,1,1}$  - отношение площади земель в естественном состоянии к их общей площади.  $a_{1,1,2}$  – других земель с поправками на территорию, подверженную оползням и подтоплению, к общей площади без территории поверхностных вод. Агрегат  $A_{1,1}$  принимает большее значение в случае лучшего состояния ресурсов,  $0 \leq A_{1,1} \leq 1$ .

Пусть  $q(A_{n,l,k})$  - оценка свойства ресурсов  $r_{n,l,k}$ , которая сопоставляется агрегату  $A_{n,l,k}$ . Величина  $q(A_{n,l,k})$  выражается на основе набора элементарных качественных оценок  $q_{l,k,m}(A_{n,l,k})$  свойств ресурса  $m$ . Для определенности будем считать, что  $q_{l,k,m}(A_{n,l,k})$  принимает значение от 0 до 4 в зависимости от состояния вида ресурсов: 0 соответствует его уничтожению, 1 - частичному уничтожению, 2 – нейтральному состоянию, 3 – частичному восстановлению, 4 – полному восстановлению.

Для земельных ресурсов, в частности, можно выделить такие оценки:

- $q_{1,1,1}(A_{1,1})$  – доступность;
- $q_{1,1,2}(A_{1,1})$  – стабильность;
- $q_{1,1,3}(A_{1,1})$  – самостоятельное восстановление;
- $q_{1,1,4}(A_{1,1})$  – качество до использования;
- $q_{1,1,5}(A_{1,1})$  – качество после использования и восстановления;
- $q_{1,1,6}(A_{1,1})$  – стоимость (оценка ее уровня);

- $q_{1,1,7}(A_{1,1}) - q_{1,1,13}(A_{1,1})$ , – давление всех категорий ресурсов по отношению к данному виду ресурсов, учитывая влияние той же категории ресурсов;
- $q_{1,1,14}(A_{1,1}) - q_{1,1,19}(A_{1,1})$ , – необходимость вида ресурсов  $k$ ,  $k \leftrightarrow l_k$ , для существования других категорий ресурсов  $l_k$ ;
- $q_{1,1,20}(A_{1,1})$  – потенциал использования;
- $q_{1,1,21}(A_{1,1})$  – потенциал восстановления.

С учетом количества свойств и максимального значения, равного 4, имеем -

$$q(A_{1,1}) = \frac{1}{84} \sum_{m=1}^{21} q_{1,1,m}(A_{1,1}), \quad (1)$$

где  $q_{1,1,m}(A_{1,1})$  – оценка  $m$ -го свойства  $A_{1,1}$ .

На основе (1) определим совокупную оценку объема и качества ресурсов (см. табл. 2), -

$$Aq_{1,1} = A_{1,1} * q(A_{1,1}), \quad (2)$$

$Aq_{1,1}$  принимает значения от 0 до 1, его большее значение соответствует лучшему состоянию ресурсов.

По аналогии с (1) – (2), но при других количественных и качественных оценках и способах агрегации (см. [Полумієнко, Рибаків, 2015]), образуются перечисленные в табл. 1 агрегаты  $Aq_{1,2} - Aq_{4,8}$ . Эти агрегаты будем обозначать через  $Aq_{n,l,k}$  и сопоставлять всем  $n,l,k$ -ресурсам.

Данные оценки не учитывают сравнительную важность видов и категорий ресурсов для региона. В отдельных международных индексах (см. [Полумієнко, Горда, 2016]), в частности, в индексе национального могущества [Ваїва, 2014], для учета этого вводятся их весовые коэффициенты.

Определим величины  $u_{n,l,k}$  и  $w_{n,l}$ , которые соответственно будем сопоставлять видам ресурсов  $k$  категории  $l$  и отдельно категориям  $l$  (см. табл. 1), -

$$\sum_k u_{n,l,k} = 1, 0 \leq u_{n,l,k} \leq 1, \sum_{l=1}^4 w_{n,l} = 1, 0 \leq w_{n,l} \leq 1$$

Сформируем величины:

$$B_n = \sum_{l=1}^4 Au_{n,l} * w_{n,l} \quad Au_{n,l} = \sum_{k \leftrightarrow l} Au_{n,l,k} = \sum_{k \leftrightarrow l} Aq_{n,l,k} * u_{n,l,k} \quad (3)$$

где сумма берется по всем видам ресурсов  $k$  категории  $l$  и по категориям ресурсов  $l$ .

Будем определять  $B_n$  (3) как индекс уровня регионального развития.

По построению, власть региона должна быть заинтересована в максимизации индекса  $B_n$ . Также имеют смысл другие условия. Проблеме устойчивого развития соответствует ситуация, когда

агрегаты, характеризующие природные и социальные ресурсы находятся на том же уровне или возрастают.

Величину

$$B = \min_n B_n \quad (4)$$

будем определять как индекс уровня национального развития.

Но, индексы (3) и (4) не определяют сбалансированности развития и использования имеющихся ресурсов. Решение этой задачи требует дополнительных средств.

### Модель сбалансированного развития

Будем сопоставлять видам ресурсов  $k$  категории  $l$  региона страны  $n$  игроков  $i_{n,l,k} \in I_n$ , - участников системы, которые контролируют и изменяют  $n,l,k$ - ресурс. Объем ресурсов игрока будем определять с помощью агрегатов  $Aw_{n,l,k}$ . Считается, что игроки заинтересованы в увеличении  $Aw_{n,l,k}$ , на что и направляются их стратегии изменения контролируемых ими  $n,l,k$ - ресурсов.

Будем определять, что стратегия  $s(i_{n,l,k}, t)$  игрока  $i_{n,l,k}$  в момент  $t$  разбиения отрезка времени  $[t_0, T]$  является вектором

$$s(i_{n,l,k}, t) = (sc(i_{n,l,k}, t), sd(i_{n,l,m(k)}, t)), k, m=1, \dots, K, m \neq k,$$

состоящим из конструктивной  $sc(i_{n,l,k}, t)$  и деструктивной  $sd(i_{n,l,m(k)}, t)$  компонент,  $s(i_{n,l,k}, t) \in S(i_{n,l,k}, t)$ ,  $S(i_{n,l,k}, t)$  – множество всех стратегий игрока. Введение конструктивной и деструктивной составляющей стратегий связано и с тем, что созданный ресурс  $sc(i_{n,l,k}, t)$  вообще отличается от израсходованных ресурсов  $sd(i_{n,l,k}, t)$ . Индекс  $m$  определяет набор ресурсов, которые контролируются другими игроками. Тогда -

$$\begin{aligned} sc(i_{n,l,k}, t) &= Aw(i_{n,l,k}, t) cw(i_{n,l,k}, t), \\ sd(i_{n,l,k}, t) &= \sum_l \sum_{m=1, m \neq k}^K Aw(j_{n,l,m(k)}, t) dw(j_{n,l,m(k)}, t), \end{aligned} \quad (5)$$

где  $cw(i_{n,l,k}, t)$  и  $dw(i_{n,l,k}, t)$  – доля полученных и израсходованных игроком ресурсов. При этом игрок заинтересован в увеличении разности

$$scd(i_{n,l,k}, t) = sc(i_{n,l,k}, t) - sd(i_{n,l,k}, t). \quad (6)$$

На  $s(i_{n,l,k}, t)$  влияют стратегии других игроков, направленные на тот же ресурс. Для повышения своего влияния, иногда вообще не имея возможности индивидуально реализовать собственные интересы, игрок должен кооперировать с другими игроками, входя в коалиции.

Коалиция  $c$  - любое объединение игроков  $i \in I$ , имеющих согласованную при создании коалиции (одинаковую) коалиционную стратегию  $s_c$ ,  $c \in C$ ,  $C$  – множество всех коалиций [Данилов, 2002]. Стратегии  $s_c$  заключаются в создании и/или использовании определенного вида ресурсов, например, в создании строительных конструкций и использовании необходимых для этого водных ресурсов.

Игрок  $i_{n,l,k}$  входит в коалицию  $c_{n,l,k}$  с долей  $sw(i_{n,l,k}, t)$  своего  $n,l,k$ -ресурса, которая прибавляется к коалиционному ресурсу, необходимому коалиции для выполнения стратегии  $sc(c_{n,l,k}, t)$ . Далее, коалиция при создании коалиционного  $n,l,k$ -ресурса тратит совокупность других  $n,l,m$ -ресурсов, зависимых от стратегии  $sc(c_{n,l,k}, t)$ . По аналогии с (5), положим, что

$$sc(c_{n,l,k}, t) = \sum_{i \in c} sc(i_{n,l,k}, t), \quad sd(c_{n,l,k}, t) = \sum_{i \in c} sd(i_{n,l,k}, t). \quad (7)$$

Из (5) - (7) имеем величину

$$scd(c_{n,l,k}, t) = sc(c_{n,l,k}, t) - sd(c_{n,l,k}, t), \quad (8)$$

которую коалиции  $c_{n,l,k}$  стараются увеличить.

Заметим, что (8) отображает текущие тенденции развития общества, ориентированного на максимальную эффективность. Нахождение же совокупного баланса интересов всего общества требует согласования стратегий всех коалиций  $c_n \in C$ . Этот баланс противоречит максимизации (8) и едва ли может быть реализованным без изменения существующих тенденций. Другими словами, стратегии устойчивого развития в ее текущем определении [United Nations, 2013] не существует. Рассматриваемый подход позволяет найти более реалистичные способы сбалансированного развития.

Будем называть ситуацией вектор стратегий всех коалиций  $c_n$  региона  $n$  и обозначать ее через  $sit_n(t)$ ,

$$\text{sit}_n(t) = (\text{scd}(c_{n,l,1}, t), \dots, \text{scd}(c_{n,L,K}, t)), \quad (9)$$

а через  $\text{Sit}_n(t)$  - множество ситуаций  $\text{sit}_n(t)$  в регионе  $n$ ,  $n=1, \dots, N$ ,  $l=1, \dots, L$ ,  $k=1, \dots, K$ .

После реализации ситуации коалиция  $c_{n,l,k}$  претендует на получение дохода, являющегося частью созданного всеми коалициями ресурса, величина которого, кроме ситуации, зависит от взноса коалиции  $\text{sc}(c_{n,l,k}, t)$ . Тем самым, необходимо проанализировать поведение каждой из коалиций  $c_{n,l,k} \in C_n$  и согласовать со стратегиями других коалиций, что вряд ли возможно, прежде всего, из-за значительного объема необходимой информации.

Но, для оценки состояния системы важнее сам результат определенной ситуации, где доход коалиции  $c_{n,l,k}$  зависит как от стратегий всех коалиций, так и от коалиций, которые изменяют только  $n, l, k$ -ресурс. Аналогично, согласование стратегий коалиций имеет двоякий характер – зависит, во-первых, от влияния участников только на  $n, l, k$ -ресурс, во-вторых, от результатов взаимодействия всех коалиций по всем ресурсам.

Текущие экономические тенденции соответствуют как раз первому из этих вариантов - согласование интересов определенной компании со всем обществом выполняется лишь за счет выполнения законодательных и других ограничений, т.е. коалиция не ищет совокупного баланса интересов, а использует стратегии, допустимые нормами общества. Цель нахождения стратегии сбалансированного развития указывает на приоритетность второго варианта развития - высшим есть баланс совокупных интересов, в пределах которого и должен искаться баланс интересов отдельной коалиции. Именно этот вариант используется далее.

Введем величину всех созданных и израсходованных коалициями ресурсов в ситуации  $\text{sit}(t_j)$ .-

$$\begin{aligned} Aw_n(t_j) &= \sum_{c \in C} \sum_{l=1}^L \sum_{k=1}^K \text{scd}(c_{n,l,k}, t_j) + Aw_n(t_{j-1}) = \text{scd}_n(t_j) + Aw_n(t_{j-1}), \\ Aw_n(t_0) &= \text{const} \end{aligned} \quad (10)$$

Будем определять выигрыш коалиции  $c_{n,l,k}$  (характеристической) функцией [Данилов, 2002] -

$$v(c_{n,l,k}, t_j) = \max_{S_{c_{n,l,k}}(t_j)} \min_{S_{C_n \setminus c_{n,l,k}}(t_j)} Aw_n(t_j), \quad (11)$$

т.е. выигрыш коалиции  $c_{n,l,k}$  является максимальным значением величины агрегата, которую  $c_{n,l,k}$  может получить исходя из множества ситуаций  $\text{Sit}_n(t_j)$ . Тотальная коалиция  $C_n$  (множество

игроков  $I_n$ ) региона  $n$  при этом получает величину

$$v(C_n, t_j) = \sum_{c_{n,l,k} \in C_n} v(c_{n,l,k}, t_j). \quad (12)$$

Можем определить (11) как характеристическую функцию кооперативной игры -

$$\Gamma_n(t_j) = \langle C_n, \text{Sit}(t_j), v(c_{n,l,k}, t_j), t_j \in [t_0, T] \rangle, \quad (13)$$

описывающей взаимодействие коалиций и игроков при создании и использовании ресурсов в регионе  $n$ .

Игру (13), включая соотношение (7) - (12), будем называть кооперативной ресурсной моделью развития региона.

Игру контролирует игрок высшего уровня - власть региона, целью которой должна быть максимизация (12), что отображает развитие региона, и, вместе с тем, сбалансированные доходы коалиций. Но, этот баланс может не соответствовать желаемому состоянию региона. В частности, если природные ресурсы находятся в состоянии упадка, дальнейшее развитие в соответствии с игрой (13) может поддерживать эту ситуацию. Для предотвращения этого необходимы дополнительные ограничения на дележ (11) и на использование ресурсов. Частично они задаются весовыми коэффициентами  $A_{w_{n,l,k}}(t)$ .

Положим, что для  $n, l, k$ -ресурсов региона, а именно для деструктивной составляющей  $sd_{n,l,k}$  стратегий по всем  $c_{n,l,k}$ , заданы величины  $L_{w_{n,l,k}}$ , для которых должно выполняться

$$A_{L_{n,l,k}} = sd_{n,l,k} - L_{w_{n,l,k}} \geq 0. \quad (14)$$

Величины  $L_{w_{n,l,k}}$  будем отождествлять с граничными значениями  $A_{w_{n,l,k}}$  и считать, что они определяют множества допустимых стратегий коалиций.

Прибавим к (13) условие (14) и перейдем к модели

$$BL_n(t_j) = \langle \Gamma_n(t_j), L_{w_{n,l,k}}, t_j \in [t_0, T] \rangle. \quad (15)$$

Условие (14) изменяет множество ситуаций и рассчитанный на нем дележ (11). Также, при его расчете не учитывается безвозвратная утрата ресурсов. Например, добыча полезных ископаемых, кроме экономической выгоды, создания рабочих мест и иной пользы, вредит окружающей среде, здоровью населения, которое вдобавок в будущем не сможет работать в этой сфере вследствие отсутствия этих ископаемых. Тем самым, коалиции, которые уничтожают ресурсы, должны компенсировать их потерю всей системе, поскольку это навсегда выводит систему из текущего состояния. Это также указывает на отсутствие стратегии устойчивого развития в ее определении [United Nations, 2013].

Другими словами, использование ресурсов, кроме получения максимальных (согласованных в соответствии с (11)) доходов должно реализовывать максимум (14) по каждому с n,l,k-ресурсов.

Будем считать, что n,l,k-ресурс является компенсационным, если для него не выполняется (14). Объем компенсации  $kw_{n,l,k}$  зависит от величины  $s_{cd_{n,l,k}}(t)$  в ситуации  $sit_n(t)$  и определяется всеми игроками или игроком высшего уровня, который контролирует игру (15). Нулевое значение  $kw_{n,l,k}$  указывает на то, что ресурс не является компенсационным

Пусть  $kw(i_{n,l,k})$  или  $kw(c_{n,l,k})$  - величина компенсации, которую должен уплатить игрок  $i_{n,l,k}$  или коалиция  $c_{n,l,k}$  другим участникам в случае использования компенсационного n,l,k-ресурса. В момент времени  $t_j$  коалиции должны уплатить величину

$$kw_n(t_j) = \sum_{l,k} kw_{n,l,k}(t_j)$$

Тогда, можно положить, что

$$Ak_{n,l,k}(t_j) = Aw_{n,l,k}(t_j) + kw_{n,l,k}(t_{j-1}) - kw_{n,l,k}(t_j),$$

включив в определение агрегата полученную в предыдущий момент и выплаченную в текущий момент времени компенсацию.

При этом коалиция  $c_{n,l,k}$  получает часть  $kw_{n,l,k}(c_{n,l,k}, t_{j-1})$  величины  $kw_{n,l,k}(t_{j-1})$ , объем которой зависит кроме  $kw_n(t_{j-1})$  от стратегии коалиции в сложившейся ситуации, т.е., объем компенсации должен определяться в соответствии с дележом (11). Такое определение создает предпосылки или для эффективного использования компенсационных ресурсов для создания перспективных для региона ресурсов, или для отказа от их использования.

Введение компенсаций позволяет перейти к рассмотрению агрегатов  $Ak_{n,l,k}$  по отдельным видам ресурсов k, а именно рассматривать вместо (6) только сумму по отдельному n,l,k-ресурсу, что можно перенести и на дележ (11), т.е. имеем

$$Ak_{n,l,k}(t_j) = \sum_{c \in C} s_{cd}(c_{n,l,k}, t_j) + Ak_{n,l,k}(t_{j-1}). \quad (16)$$

Таким образом, рассматривается главное условие учета баланса влияния всех коалиций на каждый из видов ресурсов. Сумма полученных при этом доходов и будет выражать совокупный сбалансированный доход коалиций.

Тогда, вместо (11) доход коалиции  $c_{n,l,k}$  от использования n,l,k-ресурса будем определять как

$$v(c_{n,l,k}, t_j) = \max_{S_{c_{n,l,k}}(t_j)} \min_{S_{c_{n,l,k}}(t_j)} Ak_{n,l,k}(t_j) = \max_{Sit_n(t_j)} \min_{Sit_n(t_j)} Ak_{n,l,k}(t_j),$$

а по всем видам и категориям ресурсов, которые относятся к этой коалиции, определим сумму:

$$v(c_n, t_j) = \sum_{l,k} v(c_{n,l,k}, t_j) . \quad (17)$$

Величину (17) будем отождествлять с оптимальным доходом коалиции, который она может получить от использования и создания всех ресурсов в регионе. Соответствующая ситуация и определяется как способ сбалансированного развития системы.

---

### Выводы и перспективы

В дальнейшем необходимо определить, что является сбалансированным состоянием ресурсов региона, поскольку анализ множеств ситуаций дает только сбалансированное по интересам всех коалиций изменение этого состояния. Для анализа этой задачи в пределах предложенного подхода была проведена экспериментальная оценка состояния регионов Украины по разным категориям ресурсов. Проведенное исследование показало перспективность предложенного подхода к поиску стратегий сбалансированного развития, позволило оценить потенциальные изменения направлений развития при введении компенсационных ресурсов, а также иначе взглянуть на методы индикативного анализа.

---

### Список литературы

- [United Nations, 2013] Global Sustainable Development Report – Executive Summary: Building the Common Future We Want. New York: United Nations Department of Economic and Social Affairs, Division for Sustainable Development. 2013 (online), <http://sustainabledevelopment.un.org/globalsdreport>.
- [Bossel, 1999] Bossel H. (1999) Indicators for Sustainable Development: Theory, Method, Applications. International Institute for Sustainable Development. IISD. [online] <https://www.iisd.org/pdf/balatonreport.pdf>.
- [Polumiienko, Rybakov, Trofymchuk, 2013] Polumiienko, S., Rybakov L., Trofymchuk O. (2013) The Game-Theoretical Approach to the Simulation of Sustainable Development Strategies // Journal of Earth Science and Engineering, No. 3, pp. 337 - 340.
- [Полумієнко, Рибаків, 2015] Полумієнко С.К., Рибаків Л.О. Теоретико-ігрова ресурсна модель збалансованого технологічного розвитку // Математичне моделювання в економіці, 2015, №1, с. 53 – 61.
- [Полумієнко, Горда, 2016] Полумієнко С.К., Горда. С.Є Індикативний аналіз процесів національного розвитку // Математичне моделювання в економіці, 2016, № 2 (6), с. 65-97.
- [Bajwa, 2008] Bajwa J S. Defining Elements of Comprehensive National Power, 2008 (online),

[http://www.claws.in/images/journals\\_doc/1302263399\\_JS Bajwa.pdf](http://www.claws.in/images/journals_doc/1302263399_JS Bajwa.pdf).

[Данилов, 2002] Данилов В.И. Лекции по теории игр. – М.: Российская экономическая школа, 2002. - 141 с.

---

### Информация об авторах

---



**Сергей Полумиенко** – Институт телекоммуникаций и глобального информационного пространства НАН Украины, заведующий отделом

*Major Fields of Scientific Research: game-theoretical modeling, sustainable development, information society development*



**Сергей Горда** – Институт телекоммуникаций и глобального информационного пространства НАН Украины, младший научный сотрудник

*Major Fields of Scientific Research: sustainable development*

### Main concepts of resource game-theoretical approach to the balanced development modeling

**Sergii Polumiienko, Sergii Gorda**

**Abstract:** *The paper proposes combining of resource and theoretical-game approaches to construct a balanced development model in the form of a cooperative game. The model reflects the interaction of participants in the initial system of the regional or national level and allows to determine the criteria for optimal development and to find an appropriate strategy that satisfies the diverse interests of the participants of the system.*

**Keywords:** *resource approach, cooperative model, balanced development*

## ФОРМИРОВАНИЕ ТЕСТОВ ДЛЯ ПРОВЕРКИ СПОСОБНОСТИ ДЕМАСКИРОВКИ ЗАИМСТВОВАНИЙ В ПРОГРАММАХ ВЫЯВЛЕНИЯ ПЛАГИАТА

Виктор Шинкаренко, Елена Куропятник

**Аннотация:** Исследования направлены на обеспечение своевременного реагирования на вновь открываемые возможности механического маскирования текстовых заимствований. Предлагается нестандартный подход к тестированию программного обеспечения, при котором выходные данные тестов формируется тестируемой программой. Выполнена формализация средств проверки способности демаскировки заимствований в программах выявления плагиата в виде конструкторов сценариев модификации текстов и процесса их применения. Основой методологии исследований является математический аппарат конструктивно-продукционного моделирования. основополагающий принцип этого подхода – все сущности и процессы реального и виртуального мира рассматриваются как конструкции и конструктивные процессы с соответствующей атрибутикой. Использование указанного подхода позволяет автоматизировать процессы подготовки текстов с маскировкой заимствований, а также дает возможность дальнейшего развития сценариев модификации.

**Ключевые слова:** конструктивно-продукционное моделирование, конструктор, заимствования текстов, маскировка заимствований, сценарий маскировки, тест.

**ITHEA Keywords:** F.4.2 Grammars and Other Rewriting Systems; I.2.7 Natural Language Processing; H.3.1 Content Analysis and Indexing; J.4 Social and Behavioral Sciences; D.2.5 Testing and Debugging.

---

### Введение

Развитие современных технологий накопления и обмена информацией в свободном доступе обострило проблему заимствования в академической среде. Это касается как научных публикаций, так и студенческих работ [Bakhtiyari, 2014, Букач, 2016 и др.], в том числе текстов компьютерных программ [Mahmoud, 2017].

Проблема заимствований заключается в использовании текстовой и другой информации без указания источника, а также обнародование ее под своим именем, не являясь автором. Данная проблема отождествляется с проблемой плагиата в широком понимании.

Для решения этой проблемы разработаны программные средства выявления недобросовестных заимствований [Михайловский, 2013, Vani, 2018 и др.], анализ которых выполнен в [Шинкаренко, 2017]. Не будем останавливаться на сложности решения этой задачи, связанной с такими факторами как объемы, значимость, выявление источников, переводом используемых фрагментов и др. [Meuschke, 2013].

В данной работе проблема заимствований рассматривается в контексте постоянного развития противоборствующих сторон: одна сторона (защищающаяся) пытается выявить заимствования, а другая (атакующая) – скрыть (замаскировать). Аналогичная проблема существует и в других сферах, в том числе – информационных технологий. Например, разработка и борьба с вирусами (естественными и компьютерными), взлом и защита информации.

Механическая маскировка заимствований заключается в использовании автоматизированных средств для внесения несемантических изменений в текст с целью усложнения выявления плагиата. К таким изменениям относятся: удаление или замена отдельных букв (в тексте изредка появляются орфографические ошибки), добавление пустых абзацев, вставка непечатных символов, вставка символов белого цвета, замена символов на такие же другого языка (с разным машинным кодом, но одинаковым отображением), изменение регистра некоторых символов и ряд других [Ересь, 2018, Шишкин, 2017].

Для своевременного реагирования на изменение методов атаки защищающейся стороне необходимо иметь настраиваемый инструментарий на адекватной теоретической основе. Такой основой в данной работе является конструктивно-продукционное моделирование, являющееся развитием теории формальных грамматик.

В работе показано применение конструкторов для имитации работы систем маскирования заимствований. Исходный и маскированный текст с заимствованиями представляют собой тест для проверки возможностей программ антиплагиата по демаскировке заимствований.

---

### **Связанные работы**

---

Исследованию и решению проблемы плагиата посвящено множество работ в сферах образования [Букач, 2016, Федоровская, 2016] и разработки программного обеспечения [Gulis, 2016]. Разработка программных систем антиплагиата является одним из основополагающих среди современных методов борьбы с плагиатом.

В зависимости от разных вариаций изменений входного текста, заимствования, выявленные в них, могут быть классифицированы согласно таксономии [Шинкаренко, 2017, Chowdhury, 2016].

Программы выявления плагиата включают различную предварительную обработку текста [Ceska, 2009, [Mahmoodi, 2014, Xu, 2012]. Такая обработка должна включать средства

---

---

демаскирования заимствований. Некоторые возможности по демаскировке приведены в [Ересь, 2018].

Представленные исследования выполнены на основании конструктивно-продукционного моделирования, основы которого заложены в [Shynkarenko, 2014], с приложением к задачам выявления заимствований [Шинкаренко, 2016].

---

### Цель и задача исследования

---

Цель работы можно рассматривать на нескольких уровнях:

1. Контроль добросовестного отношения к подготовке текстовых документов в академической среде средствами программ выявления плагиата;
2. Противодействие системам механического сокрытия заимствований в текстах. Своевременное реагирование на вновь открываемые возможности механического маскирования заимствований;
3. Разработка теоретических основ и соответствующего программного обеспечения для проверки способности демаскировки заимствований в программах выявления плагиата.

В соответствие с поставленными целями необходимо разработать конструктор сценариев модификации текстов и конструктор модификации текстов.

---

### Конструирование сценариев модификации текстов

---

Сценарии модификации – последовательности операций над текстом. Под операциями будем понимать действия, направленные на обработку текста: поиск и редактирование элементов без изменения семантики.

Для формализации сценариев используем конструктор [Shynkarenko, 2014]:

$$C = \langle M, \Sigma, \Lambda \rangle, \quad (1)$$

где  $M$  – неоднородный расширяемый носитель,  $\Sigma$  – сигнатура отношений и соответствующих операций: связывания, подстановки, вывода, над атрибутами,  $\Lambda$  – множество утверждений информационного обеспечения конструирования (ИОК). Данное множество включает: онтологию, цель, правила, ограничения, условия начала и завершения конструирования.

Выполним уточняющие преобразования конструктора для моделирования сценариев: специализацию, интерпретацию, конкретизацию, реализацию [Shynkarenko, 2014]. Данные преобразования приводят к расширению информационного обеспечения.

**Специализация конструктора** предполагает определение семантической составляющей носителя:

$$C = \langle M, \Sigma, \Lambda \rangle_{s \mapsto} C_{SC} = \langle M_{SC}, \Sigma_{SC}, \Lambda_{SC} \rangle, \quad (2)$$

где  $s \mapsto$  – операция специализации,  $M_{SC}$  – носитель, элементами которого являются терминалы и нетерминалы,  $\Sigma_{SC}$  – сигнатура отношений и операций для построения конструкций – алгоритмов и сценариев,  $\Lambda_{SC}$  – множество утверждений ИОК, которое будет рассмотрено далее.

Онтология конструктора  $C_{SC}$  представлена следующими утверждениями. Носитель  $M_{SC} = T \cup N$  является расширяемым. Терминалы ( $T$ ) включают конечное множество алгоритмов выполнения операций непосредственно над текстом  $\{B_i^0\}$  и дополнительных (вспомогательных)  $\{B_i^1\}$ , а также сконструированные алгоритмы и их последовательности – сценарии и необходимые для них данные. Базовые алгоритмы реализуют атомарные (условно неделимые) действия. Нетерминалы – вспомогательные элементы, составляющие множество абстрактных алгоритмов.

Текст, к которому применяются базовые алгоритмы, имеет ряд атрибутов:  $ch\_count \downarrow TXT$  – количество символов,  $w\_count \downarrow TXT$  – количество слов,  $p\_count \downarrow TXT$  – количество абзацев.

К базовым относятся алгоритмы операций над элементами текста:

- $B_1^0 \mid_{pos, el, kind, TXT}^{TXT'}$  – вставка элемента  $el$  типа  $kind = \{character, word, paragraph\}$  в позицию  $pos$  в тексте  $TXT$ , в результате выполнения которой  $TXT'$  включает  $el$  (здесь и далее  $TXT'$  – модификация  $TXT$ );
- $B_2^0 \mid_{pos, TXT}^{TXT'}$  – удаления элемента, находящегося на позиции  $pos$  в тексте;
- $B_3^0 \mid_{el_1, el_2, kind, TXT}^{TXT'}$  – замена элемента текста  $el_1$ , принадлежащего тексту  $TXT$ , на  $el_2$ ,  $kind$  – тип элемента ( $kind = \{character, word, paragraph\}$ );
- $B_4^0 \mid_{s\_pos, e\_pos, TXT}^{range}$  – присвоение  $range$  текстового фрагмента, который находится между указанными позициями  $s\_pos$  и  $e\_pos$ ;
- $B_5^0 \mid_{ft, range, TXT}^{TXT'}$  – изменения форматирования  $range$  текстового фрагмента, принадлежащего  $TXT$ , с использованием формата  $ft$ , который включает атрибуты шрифта, как в TrueType шрифтах Windows.
- $B_6^0 \mid_{el, TXT}^{res}$  – проверка наличия в тексте  $TXT$  таблиц и принадлежности элемента  $el$  одной из них, если принадлежит –  $res = true$ , иначе  $false$ ;

- $B_7^0 |_{el, TXT, start\_pos}^{pos}$  – поиск элемента  $el$  в тексте  $TXT$ , начиная с позиции  $start\_pos$ ,  $pos$  равен позиции  $el$  в тексте  $TXT$ , если  $el$  найден в  $TXT$ , иначе –  $pos = -1$ ;
- $B_8^0 |_{kind, num, TXT}^{el}$  – присвоение  $el$  элемента типа  $kind$  (символ, слово, абзац) по номеру  $num$  текста  $TXT$ ,  $el$  – идентифицированный элемент;
- $B_9^0 |_{el_1, el_2, TXT}^{TXT'}$  – вставка элемента  $el_1$  после  $el_2$  в  $TXT$ , результат –  $el_1 el_2$  содержатся в  $TXT'$ ;
- $B_{10}^0 |_{el_1, TXT}^{TXT'}$  – удаление  $el_1$  из  $TXT$ ,  $TXT'$  содержит все элементы  $TXT$ , кроме  $el_1$ .

Также к ним относится ряд вспомогательных алгоритмов:

- $B_1^1 |_q^{seq}$  – формирования последовательности  $seq$  из  $q$  пробелов;
- $B_2^1 |_{a,b}^c$  – генерации случайного числа  $c$  в диапазоне  $[a, b]$ ;
- $B_3^1 |_{c,D}^{val}$  – поиска символа  $c$  в словаре  $D$ ,  $val$  – значение символа по словарю, если символ не найден  $val = '1'$  (проверка наличия в словаре символьных «омографов»);
- $B_4^1 |_{a,b}^c$  – вычитания числа  $a$  из  $b$ ,  $c = a - b$  – результат;
- $B_5^1 |_{a,b}^a$  – присвоения  $a$  значения  $b$  ( $a := b$ ).

Форма  $I$  – совокупность элементов носителя  $M_{SC}$ , связанных отношениями  $\Sigma_{SC}$ .

Конструкция  $K$  – сентенциальная форма в текущий момент времени, содержащая только терминалы [Shynkarenko, 2014]. К конструкциям относятся алгоритмы и сценарии.

Сигнатура  $\Sigma_{SC}$  состоит из множества операций  $\Sigma_{SC} = \langle \Xi, \Theta, \Phi, \{\rightarrow\}, \Psi \rangle$ , где  $\Xi$  – отношения и одноименные операции, операции связывания и преобразования элементов носителя  $\{;, \cdot, \Pi\} \subset \Xi$ ,  $\Theta = \{\Rightarrow, \mid\Rightarrow, \|\Rightarrow\}$  – операции подстановки и частичного и полного вывода,  $\Phi$  – отношения и одноименные операции над атрибутами,  $\{\rightarrow\}$  – отношение подстановки.  $\Psi = \{\psi_i : \langle s_i, g_i \rangle\}$  – набор правил подстановки,  $s_i$  – последовательность отношений подстановки,  $g_i$  – набор операций над атрибутами. Если операции над атрибутами не выполняются, правило подстановки будет иметь вид  $\langle s_i, \varepsilon \rangle$ , где  $\varepsilon$  – пустой символ. В правилах вывода применяются отношения из  $\Xi$ , а при реализации конструктора – операции, соответствующие отношениям.

Операция  $\cdot(B_i, B_j)$  – конкатенация алгоритмов, предполагает последовательное выполнение  $B_j$  после  $B_i$ .

Операция  $:(b, A)$  – выполнение алгоритма  $A$  при условии, что выражение  $b$  – истинно ( $b = true$ ).

Операция  $\prod_{i=1}^n A$  – выполнение алгоритма  $A$   $n$  раз.

Отношение подстановки – двуместное отношение с атрибутами  $w_i I_i \rightarrow w_j I_j$ , где  $I_i, I_j$  –

сентенциальные формы [Shynkarenko, 2014]. Последовательность отношений подстановки  $s_n$

будем записывать как  $s_n = \langle I_i \rightarrow I_j, \dots, I_m \rightarrow I_k \rangle$ , где  $I_i, I_j, I_m, I_k$  – сентенциальные формы

Отношение подстановки может быть записано в сокращенной форме  $s_m = \langle I_i \rightarrow I_j \mid I_k \rangle$ , где

$I_i, I_j, I_k$  – формы; что эквивалентно  $s_m = \langle I_i \rightarrow I_j \rangle$ ,  $s_n = \langle I_i \rightarrow I_k \rangle$ .

Для заданной формы  $w_i I = w_0 \otimes (w_1 I_1, w_2 I_2, \dots, w_h I_h, \dots, w_k I_k)$  и доступного отношения подстановки

$w_p \rightarrow (w_h I_h, w_q I_q)$  такого, что  $w_h I_h$  – подформа  $w_i I$  ( $w_h I_h \prec w_i I$ ), результатом трехместной операции подстановки  $w_i I^* = w_p \Rightarrow (w_h I_h, w_q I_q, w_i I)$  будет форма

$w_i I^* = w_0 \otimes (w_1 I_1, w_2 I_2, \dots, w_q I_q, \dots, w_k I_k)$ , где  $\Rightarrow \in \Theta$ ,  $\otimes$  – любая операция связывания из  $\Xi$ .

Двухместная операция частичного вывода  $w_i I^* = w_p \mid \Rightarrow (\Psi, w_i I)$  ( $\mid \Rightarrow \in \Theta$ ) заключается в:

- выборе одного из доступных правил подстановки  $\psi_r : \langle s_r, g_r \rangle$  с отношениями подстановки  $s_r$ ;
- выполнении на его основе операций подстановки.

Двухместная операция полного вывода или просто вывода ( $\mid \Rightarrow (\Psi, w_i I)$ ,  $\mid \Rightarrow \in \Theta$ ) заключается в пошаговом преобразовании форм, начиная с начального нетерминала и заканчивая конструкцией, удовлетворяющей условию окончания вывода, что подразумевает циклическое выполнение операций частичного вывода.

Целью конструктора является конструирование более сложных алгоритмов на основе базовых, а также построение их последовательностей – сценариев.

Начальное условие для конструирования сценариев: вывод начинается с нетерминала  $SC$ .

Условие завершения конструирования: форма не содержит нетерминалов.

**Интерпретация операций конструктора** предполагает определение модели исполнителя, которую представим в виде базовой алгоритмической структуры (БАС):

$$\langle C_{SC}, C_{A,SC} = \langle M_{A,SC}, V_{A,SC}, \Sigma_{A,SC}, \Lambda_{A,SC} \rangle \rangle \mid \mapsto \langle C_{SC}, C_{SC} = \langle M_1, \Sigma_1, \Lambda_1 \rangle \rangle, \quad (3)$$

где  $\iota \mapsto$  – операция интерпретации,  $M_{A,SC} \supset M_{SC} \cup \Omega(C_{A,SC})$  – неоднородный носитель,  $\Omega(C_{SC})$  – множество сконструированных алгоритмов, которые удовлетворяют  $C_{SC}$ ;  $V_{A,SC} = \{A_i^0 |_{X_i}^{Y_i}\}$  – множество базовых алгоритмов [Shynkarenko, 2009],  $X_i, Y_i$  – входные и выходные данные алгоритма  $A_i^0 |_{X_i}^{Y_i}$ ,  $\Sigma_{A,SC}$  – сигнатура и  $\Lambda_{A,SC}$  – информационное обеспечение;  $\Lambda_{A,SC} = \{M_{A,SC} = \bigcup_{A_i^0 \in V_{A,SC}} (X(A_i^0) \cup Y(A_i^0)) \cup \Omega(C_{A,SC})\}$ ,  $\Omega(C_{A,SC})$  – множество алгоритмов реализуемых конструктором  $C_{A,SC}$ ;  $\Lambda_1 = \{(A_1^0 |_{A_i, A_j}^{A_i \cdot A_j} \lrcorner \text{"} : \text{"})$ ;  $(A_2^0 |_b^{A_i} \lrcorner \text{"} : \text{"})$ ;  $(A_3 |_{n, A}^{A \cdot A \dots} \lrcorner \text{"} \prod \text{"})$ ;  $(A_4 |_{I_h, I_q, f_i}^{f_i} \lrcorner \text{"} \Rightarrow \text{"})$ ;  $(A_5 |_{f_i, \Psi}^{f_j} \lrcorner \text{"} \Rightarrow \text{"})$ ;  $(A_6 |_{\sigma, \Psi}^{\bar{\Omega}} \lrcorner \text{"} \Rightarrow \text{"})$ ;  $\Lambda_1 \supset \Lambda_{SC}$ .

Конструктор  $C_{A,SC}$  включает алгоритмы выполнения операций:

- $A_1^0 |_{A_i, A_j}^{A_i \cdot A_j}$  – композиция алгоритмов,  $A_i \cdot A_j$  – последовательное выполнение алгоритма  $A_j$  после алгоритма  $A_i$ ;
- $A_2^0 |_b^{A_i}$  – условное выполнение: алгоритм  $A_i$  выполняется, если выражение  $b$  истинно;
- $A_3 |_{n, A}^{A \cdot A \dots}$  – выполнение алгоритма  $A$   $n$  раз;
- $A_4 |_{I_h, I_q, f_i}^{f_i}$  – подстановка,  $I_h, I_q, f_i$  – формы;
- $A_5 |_{f_i, \Psi}^{f_j}$ ,  $A_6 |_{\sigma, \Psi}^{\bar{\Omega}}$  – частичный и полный вывод, где  $f_i, f_j$  – формы,  $\sigma$  – аксиома,  $\bar{\Omega}$  – множество сформированных конструкций.

На основе интерпретации получаем конструктивную систему формирования сценариев.

**Конкретизация конструктора** определяет правила построения сценариев, на основе терминальных и нетерминальных элементов, в том числе обозначающие сконструированные алгоритмы.

Конкретизация конструктора сценариев:

$$C_{SC} \kappa \mapsto C_{SC} = \langle M_2, \Sigma_2, \Lambda_2 \rangle, \quad (4)$$

где  $\kappa \mapsto$  – операция конкретизации, которая заключается в определении правил и начальных условий конструирования алгоритмов и сценариев,  $M_2 \supset M_{SC} \cup T_1 \cup N_1$ ,  $T_1 = \{B_i^0, B_j^1, A_2^0\}$ ,  $N_1 = \{CA_\kappa, CR_1, CC_m, \alpha, SC\}$ ,  $\Lambda_2 \supset \Lambda_1$ .

Рассмотрим правила и начальные условия конструирования алгоритмов, которые далее будут использованы при построении сценариев.

*Начальные условия* конструирования алгоритмов:

- SC – нетерминал, с которого начинается вывод;
- в алгоритмах поиска стартовая позиция инициализируется нулем ( $start\_pos = 0$ );
- формат, указывающий на регистр ( $ft_1 = "case = lower"$  – формат нижнего регистра,  $ft_2 = "case = upper"$  – формат верхнего регистра,  $ft_3 = "case = next"$  – перейти к следующему регистру).

Обозначим сконструированные алгоритмы:  $CA_k$  – добавления;  $CR_i$  – удаления;  $CA_m$  – изменения.

Первое правило позволяет сконструировать алгоритм вставки последовательности  $qs$  пробелов после случайно выбранного элемента позиции  $qi$  в текст  $TXT$  :

$$s_1 = \left\langle CA_1 \mid_{qs, qi, TXT}^{TXT'} \rightarrow B_1^1 \mid_{qs}^{seq} \cdot \prod_{j=1}^{qi} (B_2^1 \mid_{a,b}^{num} \cdot B_8^0 \mid_{kind, num, TXT}^{el} \cdot B_9^0 \mid_{seq, el, TXT}^{TXT'}) \mid_{kind=word|paragraph} \right\rangle, \quad (5)$$

Правило  $s_2$  позволяет сконструировать алгоритм добавления в текст сторонних фрагментов (текстов)  $el$  на позицию, определяемую  $pos$  (начало, середина, конец) в текст  $TXT$

$$s_2 = \left\langle CA_2 \mid_{pos, el, TXT}^{TXT'} \rightarrow (A_0^2 \mid_{pos=0} : B_1^0 \mid_{0, el, TXT}^{TXT'}) \cdot (A_0^2 \mid_{pos=1} : (B_1^1 \mid_{a,b}^c \cdot B_1^0 \mid_{c, el, TXT}^{TXT'})) \cdot (A_0^2 \mid_{pos=2} : B_1^0 \mid_{ch\_count, TXT, el, kind, TXT}^{TXT'}) \mid_{kind=character, word, paragraph} \right\rangle, \quad (6)$$

Следующее правило позволяет сконструировать алгоритм добавления в текст  $qp$  пустых абзацев в случайные позиции в тексте

$$s_3 = \left\langle CA_3 \mid_{qp, TXT}^{TXT'} \rightarrow \prod_i^{qp} (B_1^1 \mid_{a,b}^c \cdot B_8^0 \mid_{kind, c, TXT}^{el} \cdot B_6^0 \mid_{el, TXT}^{res} \cdot (A_2^0 \mid_{res=false} : B_9^0 \mid_{el, el, TXT}^{TXT'})) \cdot (A_2^0 \mid_{res=false} : B_4^1 \mid_{i, 1}^i) \mid_{kind=paragraph} \right\rangle. \quad (7)$$

Правило  $s_4$  позволяет сконструировать алгоритм удаления из текста  $TXT$   $q$  абзацев начиная с заданного под номером  $sp$ .

$$s_4 = \left\langle CR_1 \mid_{sp, q, TXT}^{TXT'} \rightarrow \prod_{i=sp}^{sp+q} (A_2^0 \mid_{i < p\_count, TXT} : (B_8^0 \mid_{kind, i, TXT}^{el} \cdot B_{10}^0 \mid_{el, TXT}^{TXT'})) \mid_{kind=paragraph} \right\rangle. \quad (8)$$

Для конструирования алгоритма изменения раскладки  $type$  случайного символа в тексте может быть использовано такое правило:

$$s_5 = \left\langle CC_1 \mid_{type, TXT}^{TXT'} \rightarrow B_1^1 \mid_{a,b}^c \cdot B_8^0 \mid_{kind, c, TXT}^{el} \cdot \prod_{i=1}^k (A_2^0 \mid_{type=i} : B_3^1 \mid_{el, D_i}^{val}) \cdot (A_2^0 \mid_{val \neq '1'} : B_3^0 \mid_{el, val, kind, TXT}^{TXT'}) \mid_{kind=character} \right\rangle. \quad (9)$$

Правило  $s_6$  позволяет сконструировать алгоритм изменения раскладки всех вхождений символа  $c$ , содержащегося в тексте  $TXT$ ,  $l\_type$  – раскладка для модифицируемого символа.

$$s_6 = \left\langle CC_2 \mid_{l\_type, el, TXT}^{TXT'} \rightarrow \prod_{i=1}^k (A_2^0 \mid_{l\_type=i} : B_3^1 \mid_{el, D_i}^{val}) \cdot \alpha \right. \\ \left. \alpha \rightarrow B_7^0 \mid_{c, TXT, start\_pos}^{pos} \cdot (A_2^0 \mid_{pos \geq 0} : (B_3^0 \mid_{c, val, kind, TXT}^{TXT'} \cdot B_5^1 \mid_{start\_pos, pos}^{start\_pos} \cdot \alpha)) \mid_{kind=character} \right\rangle. \quad (10)$$

Приведенное ниже правило позволяет сконструировать алгоритм изменения регистра по признаку  $r$  (верхний / нижний) для всего текста  $TXT$

$$s_7 = \left\langle CC_3 \mid_{r, TXT}^{TXT'} \rightarrow B_4^0 \mid_{0, ch\_count, \neg TXT, TXT}^{range} \cdot (A_2^0 \mid_{r=true} : B_5^0 \mid_{ft_1, range, TXT}^{TXT'}) \cdot (A_2^0 \mid_{r=false} : B_5^0 \mid_{ft_2, range, TXT}^{TXT'}) \right\rangle. \quad (11)$$

Алгоритм изменения регистра символа текста  $TXT$  на позиции  $pos$  конструируется следующим образом

$$s_8 = \left\langle CC_4 \mid_{pos, TXT}^{TXT'} \rightarrow B_4^0 \mid_{pos, pos+1, TXT}^{range} \cdot B_5^0 \mid_{ft, range, TXT}^{TXT'} \right\rangle. \quad (12)$$

Правило  $s_9$  позволяет сконструировать алгоритм изменения регистра заданного символа  $c$ , принадлежащего  $TXT$

$$s_9 = \left\langle CC_5 \mid_{c, TXT}^{TXT'} \rightarrow B_7^0 \mid_{c, TXT, start\_pos}^{pos} \cdot CC_4 \mid_{pos, TXT}^{TXT'} \right\rangle. \quad (13)$$

Для конструирования алгоритма изменения позиций абзацев под номерами  $num_1$ ,  $num_2$  (меняются местами) применяется следующее правило:

$$s_{10} = \left\langle CC_6 \mid_{num_1, num_2, TXT}^{TXT'} \rightarrow B_9^0 \mid_{kind, num_1, TXT}^{el_1} \cdot B_6^0 \mid_{el_1, TXT}^{res_1} \cdot (A_2^0 \mid_{res_1=false} : (B_9^0 \mid_{kind, num_2, TXT}^{el_2} \cdot B_6^0 \mid_{el_2, TXT}^{res_2} \cdot (A_2^0 \mid_{res_2=false} : (B_3^0 \mid_{el_1, el_2, kind, TXT}^{TXT'})))) \mid_{kind=paragraph} \right\rangle. \quad (14)$$

Правило  $s_{11}$  позволяет сконструировать алгоритм изменения позиций слов под номерами  $w\_num_1$ ,  $w\_num_2$  (меняются местами)

$$s_{11} = \left\langle CC_7 \mid_{w\_num_1, w\_num_2, TXT}^{TXT'} \rightarrow A_2^0 \mid_{w\_num_1, w\_num_2 < w\_count, \neg TXT} : (B_9^0 \mid_{kind, w\_num_1, TXT}^{el_1} \cdot B_9^0 \mid_{kind, w\_num_2, TXT}^{el_2} \cdot B_3^0 \mid_{el_1, el_2, kind, TXT}^{TXT'}) \mid_{kind=word} \right\rangle \quad (15)$$

Следующее отношение подстановки позволяет сконструировать алгоритм изменения слова  $el_1$ , предполагающее его замену на  $el_2$ ,  $type$  – вид замены (первое вхождение/ все вхождения). При определенных значениях входных параметров данный алгоритм может быть использован для удвоения пробелов

$$s_{12} = \left\langle CC_8 \mid_{el_1, el_2, type, TXT}^{TXT'} \rightarrow (A_2^0 \mid_{type=2} : (B_7^1 \mid_{el, TXT, start\_pos}^{pos} \right. \quad (16)$$

$$\cdot (A_2^0 \mid_{pos \geq 0} \cdot (B_3^0 \mid_{el, el_2, TXT, kind}^{TXT'} \cdot B_5^1 \mid_{start\_pos, pos}^{start\_pos} \cdot CA_8 \mid_{TXT}^{TXT'})) \cdot \\ \cdot (A_2^1 \mid_{type=1} \cdot (B_7^1 \mid_{el, TXT, start\_pos}^{pos} \cdot (A_2^0 \mid_{pos \geq 0} \cdot (B_3^0 \mid_{el, el_2, kind, TXT}^{TXT'} \cdot B_5^1 \mid_{start\_pos, pos}^{start\_pos})))) \mid_{kind=word} \cdot$$

Сценарии могут быть сконструированы с использованием таких правил:

$$SC \mid_{param, TXT}^{TXT'} \rightarrow CA_1 \cdot SC \mid_{param, TXT}^{TXT'} \mid CA_2 \cdot SC \mid_{param, TXT}^{TXT'} \mid CA_3 \cdot SC \mid_{param, TXT}^{TXT'} \\ SC \mid_{param, TXT}^{TXT'} \rightarrow CR_1 \cdot SC \mid_{param, TXT}^{TXT'} \\ SC \mid_{param, TXT}^{TXT'} \rightarrow CC_1 \cdot SC \mid_{param, TXT}^{TXT'} \mid CC_2 \cdot SC \mid_{param, TXT}^{TXT'} \dots \mid CC_8 \cdot SC \mid_{param, TXT}^{TXT'} \\ SC \mid_{param, TXT}^{TXT'} \rightarrow CA_1 \mid CA_2 \mid CA_3 \mid CR_1 \mid CC_1 \mid CC_2 \dots \mid CC_8, \quad (17)$$

где  $param$  – множество входных значений алгоритмов  $\{CA_j\}, \{CR_k\}, \{CC_m\}$ . Для сокращения записи в (17) не приводятся входные и выходные данные алгоритмов, определённые в (5–16).

**Реализацией конструктора** является множество сценариев  $SC = \{SC_i\}$  для синтаксических и структурных изменений, которые сочетают алгоритмы операций обработки текстов с определёнными параметрами. Сценарий – упорядоченная последовательность сконструированных алгоритмов. Элементы последовательности операций могут быть различными, идентичными или быть сценариям с различными параметрами выполнения.

### Конструирование модификации текстов

Рассмотрим специализированный конструктор модификаций текстов для сокрытия заимствований:

$$C = \langle M, \Sigma, \Lambda \rangle_s \mapsto C_{AT} = \langle M_{AT}, \Sigma_{AT}, \Lambda_{AT} \rangle, \quad (18)$$

где  $M_{AT}, \Sigma_{AT}, \Lambda_{AT}$  – носитель, сигнатура и множество утверждений ИОК соответственно.

*Онтология конструктора*  $C_{AT}$  представлена следующими утверждениями. Элементами носителя являются терминалы и нетерминалы. Терминалы включают конечное множество сконструированных сценариев  $\Omega(C_{A, SC})$ , а также данные для их применения, тексты и их элементы: символы электронного представления текстов и языковые конструкции: слова, предложения, абзацы [20]. Нетерминалы – вспомогательные элементы. Сигнатура  $\Sigma_{AT}$  состоит из множества операций  $\Sigma_{AT} = \langle \Xi_{AT}, \Theta_{AT}, \Phi_{AT}, \{\rightarrow\}, \Psi_{AT} \rangle$ , где  $\Xi_{AT}$  – отношения и одноименные операции, операции связывания и преобразования элементов носителя  $\{;, \nabla\} \subset \Xi_{AT}$ .  $\Theta_{AT} = \{\Rightarrow, \mid\Rightarrow, \mid\Rightarrow\}$  – операции подстановки и вывода,  $\{\rightarrow\}$  – отношение подстановки,  $\Phi_{AT}$  –

отношения и одноименные операции над атрибутами,  $\Psi_{AT} = \{\psi_i : \langle s_i, g_i \rangle\}$  – аналогичны операциям и отношениям  $C_{SC}$ .

Операция  $\nabla_{(w_s, S, TXT)}$  предполагает применение сценария  $S$  к тексту  $TXT$ .

Остальные операции – такие же, как в конструкторе сценариев  $C_{SC}$ .

*Целью конструирования* является получение базы текстов, которая состоит из пар «оригинал-модифицированная копия», что является набором тестовых значений (данных) для систем выявления заимствований (антиплагиата).

*Ограничение конструктора модификации текста.* Количество и последовательность примененных сценариев ограничивается их взаимоисключающим действием (например, смены латинского алфавита на кириллический и обратно, смена регистра).

*Начальные условия модификации текстов:*

- $TXT$  – текст, содержащий более одного абзаца для возможности применения всех видов сценариев;
- $\sigma$  – нетерминал, с которого начинается вывод;
- $\{D_i\}, i = \overline{1, k}$  – словари двух и более алфавитов (например, латинского и кириллического), содержащие пары символов из разных алфавитов с одинаковым начертанием,  $k = \overline{2, N}$  – количество словарей в зависимости от количества разных алфавитов;
- набор входных значений сценария ( $param$ ), задающийся внешним исполнителем:  $qi$  – количество вставок,  $qs$  – длина последовательности,  $qp$  – количество пустых абзацев для вставки,  $q$  – количество абзацев для удаления,  $sp$  – номер абзаца, с которого начинается удаление,  $type$  – тип замены (однократная/все входжения),  $l\_type$  – раскладка для модифицируемого символа,  $r$  – признак установки нижнего регистра,  $pos$  – позиция символа в тексте для модификации,  $s$  – символ для модификации,  $num_1, num_2$  – номера абзацев для перестановки,  $w\_num_1, w\_num_2$  – номера слов для перестановки,  $el_1, el_2$  – слова для замены;

*Условие завершения конструирования:* форма не содержит нетерминалов.

**Интерпретация** предполагает определения модели исполнителя, которую представим в виде базовой алгоритмической структуры (БАС), рассмотренную в (2), дополнив алгоритмом выполнения операции применения сценария ( $A_8 |_{S, TXT}^{TXT} \leftarrow \nabla$ ).

На основе интерпретации получаем конструктивную систему сценариев.

**Конкретизация конструктора** модификации текстов:

$$C_{AT} R \mapsto C_{AT} = \langle M_4, \Sigma_4, \Lambda_4 \rangle, \quad (19)$$

где  $M_4 \supset M_{AT} \cup T_2 \cup N_2$ ,  $T_2 = \{SC_i, TXT\}$ ,  $N_2 = \{\sigma\}$ ,  $\Lambda_4 \supset \Lambda_3$ .

Правило, приведенное ниже, позволяет выполнить модификацию текста, применив к нему один или несколько сценариев.

$$s = \langle \sigma \Big|_{param, TXT}^{TXT'} \rightarrow \nabla(SC \Big|_{param, TXT}^{TXT'}, TXT) \cdot \sigma, \sigma \rightarrow \nabla(SC \Big|_{param, TXT}^{TXT'}, TXT) \rangle. \quad (20)$$

**Реализация конструктора**  $C_{AT}$  заключается в изменении элементов его носителя путем выполнения операции применения сценария с учетом утверждения ИОК:

$${}_A C_{AT} R \mapsto \Omega({}_A C_{AT}). \quad (21)$$

Таким образом, результатом реализации конструктора являются тексты, которые подвержены «механическим искажениям» – синтаксическим изменениям, которые не влияют или имеют незначительное влияние на смысл.

---

## Результаты

Разработан конструктор формирования сценариев модификации текстов, моделирующий бесконечное разнообразие последовательности приемов маскировки заимствований существующих и потенциальных программных средств по механическому сокрытию плагиата.

В паре с конструктором модификации текстов имеем теоретическую основу формирования базы тестов для проверки способности демаскировки заимствований в программах выявления плагиата.

Входом теста является текст с множеством заимствований и этот же текст с маскировкой, выполненной в результате реализации конструктора модификации текста. Выходом – результаты работы исследуемой программы выявления плагиата с одним и другим текстом. Тест считается успешно пройденным, если результаты совпадают.

Авторами разработана программа, реализующая представленные в данной работе конструкторы.

---

## Заключение

Применение конструктивного моделирования для разработки и совершенствования систем выявления плагиата позволяет на единой теоретической основе:

- формализовать сценарии модификации;

- создать платформу для моделирования новых изменений текста;
- автоматизировать построение сценариев и их применение для создания тестовой базы текстов для оценки возможностей демаскировки заимствований.

Разработанный формализм позволяет автоматизировать решение класса задач связанных с моделированием работы вредоносных программ маскировки заимствований.

---

### **Дальнейшая работа**

Авторы занимаются разработкой собственных средств выявления заимствований в текстах, в том числе в компьютерных программах и математических формулах. Представленная модель является частью этой работы.

---

### **Благодарности**

Статья публикуется с частичной поддержкой ITHEA ISS ([www.ithea.org](http://www.ithea.org)) и ADUIS ([www.aduis.com.ua](http://www.aduis.com.ua)).

*The paper is published with partial support by the ITHEA ISS ([www.ithea.org](http://www.ithea.org)) and the ADUIS ([www.aduis.com.ua](http://www.aduis.com.ua)).*

---

### **Литература**

- [Bakhtiyari, 2014] Bakhtiyari K., Salehi H., Embi M.A., Shakiba M., Zavvari A., Shahbazi-Moghadam M., Ebrahim N.A., Mohammadjafari M. Ethical and unethical methods of plagiarism prevention in academic writing. International Education Studies; Vol. 7. No. 7. Canadian Center of Science and Education, 2014. pp. 52-62. ISSN 1913-9020 E-ISSN 1913-9039. doi:10.5539/ies.v7n7p52
- [Ceska, 2009] Ceska Z., Fox C. The Influence of Text Pre-processing on Plagiarism Detection. In: Angelova, G and Bontcheva, K and Mitkov, R and Nicolov, N and Nikolov, N, (eds.) International Conference on Recent Advances in Natural Language Processing, Borovets, Bulgaria, 2009. Association for Computational Linguistics. pp. 55 - 59.
- [Gulis, 2016] Gulis I., Chuda D., Petrik J. Plagiarism Detection in Students' Assignments Written in Natural Language. Proceedings of the International Conference on e-Learning, 2016. pp.141 – 144, <http://elearning-conf.eu/docs/cp16/paper-21.pdf>
- [Mahmoodi, 2014] Mahmoodi M., Varnamkhasti M.M. Design a Persian Automated Plagiarism Detector (AMZPPD). International Journal of Engineering Trends and Technology (IJETT). Vol. 8 No. 8. 2014. pp. 465-467. ISSN 2231-5381. doi: 10.14445/22315381/IJETT-V8P280 <http://www.ijettjournal.org/volume-8/number-8/IJETT-V8P280.pdf>

- [Mahmoud, 2017] Mahmoud H.A. Detecting Disguised Plagiarism. arXiv preprint arXiv:1711.02149 <https://arxiv.org/ftp/arxiv/papers/1711/1711.02149.pdf>
- [Meuschke, 2013] Meuschke N., Gipp B. State-of-the-art in detecting academic plagiarism. International Journal for Educational Integrity Vol. 9 No. 1. 2013. pp. 50–71. ISSN 1833-2595
- [Chowdhury, 2016] Chowdhury H.A., Bhattacharyya D.K. Plagiarism: Taxonomy, Tools and Detection Techniques Knowledge. Library and Information Networking, NACLIN 2016, ISBN: 978-93-82735-08-3 <https://arxiv.org/abs/1801.06323>
- [Shynkarenko, 2014] Shynkarenko V.I., Ilman V.M. Constructive-Synthesizing Structures and Their Grammatical Interpretations. Part I. Generalized Formal Constructive-Synthesizing Structure. Cybernetics and Systems Analysis, Vol. 50, No 5. Springer, 2014. pp. 665 – 662. Part II. Refining Transformations. –Vol. 50, No 6, 2014. pp. 829 – 841. ISSN: 1060-0396 (Print) 1573-8337 (Online), doi: 10.1007/s10559-014-9655-z, <https://link.springer.com/article/10.1007/s10559-014-9655-z>, DOI 10.1007/s10559-014-9674-9, <https://link.springer.com/article/10.1007/s10559-014-9674-9>
- [Shynkarenko, 2009] Shynkarenko V.I., Ilman V.M., Skalozub V.V. Structural models of algorithms in problems of applied programming. I. Formal algorithmic structures. Cybernetics and Systems Analysis, Vol. 45, No 3. Springer, 2009. pp 329-339. ISSN: 1060-0396 (Print) 1573-8337 (Online), doi: org/10.1007/s10559-009-9118-0 <https://link.springer.com/article/10.1007/s10559-009-9118-0>
- [Vani, 2018] Vani K., Deepa G. Unmasking text plagiarism using syntactic-semantic based natural language processing techniques: Comparisons, analysis and challenges. Information Processing & Management, Vol. 54, No 3. Elsevier, 2018. pp. 408-432, ISSN 0306-4573, [doi:org/10.1016/j.ipm.2018.01.008](https://doi.org/10.1016/j.ipm.2018.01.008).  
<http://www.sciencedirect.com/science/article/pii/S0306457317300547>
- [Xu, 2012] Qin Xu, Tan Tang, Lan-su Nie Research of text plagiarism detection process. Proceeding of the international conference on information engineering and applications (IEA). Vol 4. Springer, 2012. pp. 401-410. ISSN 1876-1100 ISSN 1876-1119 (el) DOI 10.1007/978-1-4471-4853-1
- [Букач, 2016] Букач М.М., Жело Х.А. Академічна чесність як основа становлення майбутнього науковця. Наукові праці. Педагогіка, Т. 269, №. 257, 2016. с. 52 – 56. ISSN 2311-0287 (Print), <http://pednauki.chdu.edu.ua/article/view/89746/85127>
- [Ересь, 2018] Ересь А.В. Лучанинов Д.В. Анализ современных систем обхода заимствований текста. Постулат. №1. ISSN 2414-4487 <http://e-postulat.ru/index.php/Postulat/article/viewFile/1097/1124>
- [Михайловский, 2013] Михайловский Ю.Б. Система Anti-Plagiarism як інструмент запобігання плагіату в навчальній та науковій діяльності. Вісник Хмельницького національного університету.

Технічні науки, № 3, 2013. с. 162–168. ISSN 2307-5740, [http://journals.khnu.km.ua/vestnik/pdf/tech/2013\\_3/33myh.pdf](http://journals.khnu.km.ua/vestnik/pdf/tech/2013_3/33myh.pdf)

[Федоровская, 2016] Федоровская Н.А. Пути снижения показателей уровня плагиата в работах студентов творческих направлений в высших учебных заведениях. Международный журнал прикладных и фундаментальных исследований, №. 11-3, 2016. с. 533-536.

[Шинкаренко, 2017] Шинкаренко В.И., Куропятник Е.С. Проблемы выявления плагиата и анализ инструментального программного обеспечения для их решения. Наука и прогресс транспорта, № 1 (67), 2017. с. 131—142. ISSN 2307–3489 (Print), ISSN 2307–6666 (Online), doi: 10.15802/stp2017/94034, <http://stp.diit.edu.ua/article/view/94034/92573>

[Шинкаренко, 2016] Шинкаренко В.И., Куропятник Е.С. Конструктивно-продукционная модель графового представления текста. Проблемы программирования, № 2 – 3, 2016. с. 63 – 72. ISSN 1727- 4907, <http://dspace.nbu.gov.ua/bitstream/handle/123456789/126391/07-Shinkarenko.pdf?sequence=1>

[Шишкин, 2017] Шишкин Ю.Е. Исследование возможностей систем обнаружения заимствований в методологии больших данных. Фундаментальные основы инновационного развития науки и образования. Пенза: МЦНС «Наука и просвещение», 2017. с. 55-73. ISBN 978-5-907012-88-2

---

### Информация об авторах

---



**Виктор Шинкаренко** – д.т.н., профессор, зав. кафедрой «Компьютерные информационные технологии» Днепропетровского национального университета железнодорожного транспорта имени академика В. Лазаряна; ул. Лазаряна, 2, 49010, Днепр, Украина; e-mail: [shinkarenko\\_vi@ua.fm](mailto:shinkarenko_vi@ua.fm)

Основные области научных исследований: конструктивно-продукционное моделирование, качество программного обеспечения, искусственный интеллект.



**Елена Куропятник** – аспирант Днепропетровского национального университета железнодорожного транспорта имени академика В. Лазаряна; ул. Лазаряна, 2, 49010, Днепр, Украина; e-mail: [olena.kuropyatnyk@gmail.com](mailto:olena.kuropyatnyk@gmail.com)

Основные области научных исследований: методы и средства выявления заимствований, инженерия программного обеспечения

## **Creation tests for checking plagiarism detection programs' ability of unmasking borrowings**

**Viktor Shynkarenko, Olena Kutopiatnyk**

*Abstract: The research is aimed at ensuring timely response to the newly discovered opportunities for mechanical masking of text borrowings. A non-standard approach to software testing is proposed, in which the test output data is generated by the program under test. The formalization of means to verify the ability to unmask borrowing in plagiarism detection programs in the form of constructors of texts modification scenario and process of their application were developed. The basis of the methodology of research is the mathematical apparatus of constructive-synthesizing modeling. The fundamental principle of this approach is that all the entities and processes of the real and virtual world are considered as constructs and constructive processes with the corresponding attributes. The usage of this approach makes it possible to automate the processes of preparing texts containing disguises of borrowing, and also allows further development of scenarios for modification.*

**Keywords:** *constructive-synthesizing modeling, constructor, text borrowing, disguise of borrowing, masking scenario, test.*

## TABLE OF CONTENT

<i>Методологии систем и системного анализа как основа развития информатики</i>	
Наталья Д. Панкратова, Виолетта Н. Волкова .....	3
<i>О теоретических основах модели свр-анализа</i>	
В.П.Галушко, Т.Н. Болотова, В.И Макаров, В.Ю.Величко .....	18
<i>Идентификация смысловой близости фрагментов текстов наукометрических баз</i>	
Светлана Петрасова, Нина Хайрова, Валерия Киселева .....	32
<i>Анализ взаимозависимости временных рядов биткоина и активности сообществ в социальных сетях</i>	
Людмила Кириченко, Тамара Радивилова, Виталий Булах, Вадим Чакрян .....	43
<i>Формирование обучающих выборок для Синтеза Адаптивной логической сети типа «трапецеидальная матрица»</i>	
Владимир Опанасенко, Сергей Кривый, Станислав Завьялов .....	56
<i>Основные концепции ресурсного теоретико-игрового подхода к моделированию сбалансированного развития</i>	
Сергей Полумиенко, Сергей Горда .....	72
<i>Формирование тестов Для проверки способности демаскировки заимствований в программах выявления плагиата</i>	
Виктор Шинкаренко, Елена Куропятник .....	84
Table of content .....	100