

# Концепция стохастического метамоделирования интеллектуальных цифровых экосистем

О.Р. Попов

**Аннотация**— Предложена концепция стохастического метамоделирования интеллектуальных цифровых экосистем (ИЦЭ) – цифровых двойников, имитирующих организационные системы с выраженным характером передачи знаний. Реализуемая методология объединяет модели сетевого анализа (N подход) и вероятностные модели дистрибутивной семантики (S подход). К N подходу относят перколяционное моделирование (PM) и методы фрактального анализа (FA). К S подходу относится – вероятностное тематическое моделирование (PTM) и стохастическая модель прогнозирования трендов (SPM). Все перечисленные модели могут быть решены вычислительными методами, имитирующими параметры и переменные компьютерными алгоритмами.

В ходе последовательной реализации алгоритмов моделирования формируются два кортежа, составленные из наборов параметров данных, полученных при моделировании – сетевой  $M_n$  и семантической  $M_s$ . Представлена формальная мета-модель MM, выражающая авторскую методологию стохастического моделирования, составленная из самих моделей  $M_n$  и  $M_s$ , а также иерархических  $R_n$ , ассоциативных  $R_A$  и транзитивных  $R_T$  отношений между моделями. Для точного решения описательной задачи интеллектуального анализа возможных зависимостей между переменными в наборах данных, предложена концепция ДСМ-метода (правдоподобного вывода) автоматизированной поддержки научных исследований и методы теории нечетких множеств.

Рассмотрены общие результаты апробации предложенного подхода к исследованию ИЦЭ на основе обработки данных сетей цитирования научных публикаций и сетей научных терминов, составленных из выборок библиографических баз данных arXiv, Scopus и Google Scholar.

**Ключевые слова** — алгоритм, интеллектуальная цифровая экосистема, моделирование, параметр, семантика, сеть, стохастическая модель.

## I. ВВЕДЕНИЕ

Объектом исследования выступают цифровые экосистемы (ЦЭ) и результаты их интеллектуальной деятельности.

ЦЭ в современной экономике являются результатом объединения корпоративного сотрудничества и информационных технологий [1].

Характеристики интеллектуальной системы (ИС) – лучше всего понимаются в динамической интеграции информационных и системных свойств. Особенность самоорганизации ИС – это поддержание автономности сетью взаимодействий ее составных частей и диссипативность [2-4]. В случае интеллектуальных систем это выражается в приращении информации, знаний [5].

Таким образом, интеллектуальную цифровую экосистему (ИЦЭ), можно определить как автономную ЦЭ, нацеленную на производство новых знаний [6, 7]. В таком понимании – она может выступать как реальной организацией, так и виртуальной системой.

Примером ИЦЭ на макроуровне является система научного инжиниринга отрасли, на микро – научный отдел университета, библиографическая база данных (ББД).

Концепция «цифрового двойника» (Digital Twin, DT) возникла в бизнесе, при решении задач поиска эффективных инструментов для прогнозирования и управления технической стратегией [8]. Использование цифровых моделей сложных объектов (физических активов), обладающих зеркальной информационной взаимосвязью в реальном времени с активом, превращает активы в целостную управляемую кибер-физическую систему [9]. В условиях непрерывной цифровизации экономики данная идея представляет собой содержательное ядро глобального тренда в направлении создания более интеллектуальных и высокоавтоматизированных систем [10].

В качестве примера развития применения концепции «цифрового двойника» в промышленности приведем цифровой бетон [11]. Под цифровым бетоном понимаются все продукты, произведенные с использованием технологий изготовления с использованием цифровых моделей и машин с ЧПУ на основе цифровых данных. В последние годы, отвечая на более строгие требования с точки зрения классических показателей надежности, долговечности и прочности [12], конструкционный бетонный материал претерпел значительные изменения, поскольку его производство стало более высокотехнологичным.

Модели, используемые в интеллектуальном цифровом двойнике [10], обладают возможностью самостоятельного машинного обучения, применяемые к потоку обновляемых реальных данных для выявления закономерностей, которые затем используются в работе системы.

В меньшей степени феномены «интеллектуальной цифровой экосистемы» и «интеллектуального цифрового двойника» исследованы для применения в научно-образовательной среде.

В контексте постановки целей исследования ИЦЭ – это модель цифрового двойника, имитирующая организационные системы с выраженным характером передачи знаний.

В таких системах важен анализ таких показателей, как опубликованные научные статьи, материалы конференций, информационные сайты [13, 14]. Инструментом такого анализа является моделирование текстов и модели связей (сетей) цитирования публикаций.

Для моделирования таких сложных, нелинейных, высокоразмерных, насыщенных, динамических систем применяются, развиваемые в теории информации и прикладной математике, вероятностные подходы и методы сетевого анализа.

Однако, полноценной методологии, объединяющей такие подходы до сих не разработано и это является актуальной теоретической и прикладной задачей.

Статья получена 30 июля 2024.

О.Р. Попов – Южное отделение МОО «Академия информатизации образования» (e-mail: fu7urway@gmail.com)

## II. МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ

Для реализации теоретической цели исследования сформируем общую расширенную постановку методологии для построения интеллектуального цифрового двойника с помощью парадигмы стохастического моделирования.

Дистрибутивная гипотеза лингвистики (distributional hypothesis) основана на том, что, статистические зависимости употребления слов и терминов (термов), их частотность и порядок повторения могут быть использованы для нахождения заложенного в текст смысла.

Алгоритмы информационного поиска, которые используют данную гипотезу для выявления многих аспектов семантики, применяя статистические методы к большим корпусам текстов, исходят из положения, что каждый документ — это некоторый баланс тем. Назовем данный подход к моделированию — семантическим S.

Другая интерпретация интеллектуальных систем — графовая или сетевая N. Моделирование сети понятий позволяет сделать вывод о структурных особенностях и взаимосвязях ее компонентов, получить анализ групп тем путем соединения концепций.

Оба подхода имеют свои сильные и слабые стороны. Сосредоточение на распределениях и контекстах позволяет находить схожие понятия, определять значения слов или проверять логическую корректность. В свою очередь, динамика и нечеткость сети необходима, потому что семантика дистрибутивных моделей, как правило, фиксирована.

Проблема, которую мы решаем, заключается в объединении этих двух подходов для анализа более сложных явлений.

Кроме того, результатом моделирования с учетом динамики процесса, являются статистические характеристики переменных системы, накопленные (измеренные) в процессе работы модели.

В зависимости от того, каким образом генерируются события при моделировании эволюции системы во времени такого рода модели делятся на детерминированные и стохастические.

Классическим и широко используемым подходом к имитационному моделированию систем, изменяющихся во времени, является дискретно-событийное моделирование (discrete-event simulation) [15].

Наиболее обширная практика и наличие программно-инструментальных средств представлены в рамках детерминированных подходов к построению модели возникновения событий. Менее изучены стохастические модели (алгоритмы) генерирования изменения состояний.

В последнее время такого рода стохастическим интерпретатором поступившего набора данных из реального мира выступают большая языковая модель (LLM) или большая мультимодальная модель (LMM). Цифровой двойник выступает как в роли информационной системы (анализ данных реального мира), так и в роли системы моделирования (аугментации или расширения реальных данных) [16].

Базовая концепция многоуровневого метамоделирования Object Management Group (OMG) представляют собой иерархию или сеть метамоделей для интернета вещей (IoT) и DT, построенную путем повторного использования существующих моделей, которая может содержать несколько (перекрестных) моделей домена или настраиваемых приложений [17].

Предлагаемая методология стохастического моделирования для изучения ИЦЭ представляет собой сложную иерархическую взаимосвязь моделей.

Для решения базовой задачи моделирования построения структур цифровых экосистем необходимы графовые модели закономерностей формирования структуры и взаимосвязей компонентов сети. Это достигается применением методов теории перколяции и перколяционных геометрических фазовых переходов и фрактального

анализа спектров случайных процессов в информационном пространстве [18].

В свою очередь, для решения задачи моделирования выявления и прогнозирования трендов в ИЦЭ требуется строить генеративные тематические модели научного текста [19] и стохастические модели прогнозирования динамики событий [20].

Все перечисленные модели могут быть описаны в рамках комплекса теорий вероятности и математической статистики, а также графов и сетей и решены вычислительными методами стохастического моделирования, имитирующего параметры и переменные компьютерными алгоритмами.

Обозначим кортежем  $M_n$  объединение всех групп параметров, полученных в результате моделирования формирования структуры сети:

$$M_n = \langle N_k, N_l, N_p, N_f, N_t \rangle, \quad (1)$$

где  $N_k$  — набор параметров, характеризующих общее состояние структуры сети;  $N_l$  — набор параметров, характеризующих состояние связей в сети;  $N_p$  — критические показатели перколяции;  $N_f$  — показатели фрактальности (самоподобия);  $N_t$  — набор показателей динамики (временных рядов).

Обозначим кортежем  $M_s$  объединение всех групп параметров, полученных в результате моделирования выявления и прогнозирования трендов:

$$M_s = \langle S_p, S_m, S_v, S_r, S_t \rangle, \quad (2)$$

где  $S_p$  — набор гиперпараметров и регуляризаторов вероятностного тематического моделирования PTM;  $S_m$  — набор данных, характеризующих результаты PTM;  $S_v$  — набор метрик оценки качества PTM;  $S_r$  — метрики оценки релевантности (семантического сходства);  $S_t$  — параметры моделирования стохастической динамики информационных процессов.

Точная оценка параметров и показателей, входящих в наборы компонентов кортежей  $M_n$ ,  $M_s$ , критически важна для эффективности изучаемых ИЦЭ.

Предлагаемая методология измерения, обработки, оценки и выявления взаимодействия наборов данных заключается в определенной системе моделирования информационных процессов, основанная на теоретико-вероятностных подходах.

Каждая из стохастических моделей требует последовательного анализа, как собственных внутренних механизмов, так и механизмов их взаимодействия между собой.

На рис. 1 показана сложная иерархическая структура взаимосвязей стохастических моделей, представляющая собой методический макет исследования.

Возникновение перколяционных кластеров и фрактальных массивов в информационном пространстве взаимосвязей авторов и сетей цитирования является отображением информационных процессов в определенной тематической области, что сопровождается появлением новых научных тематик.

Комплекс сетевых, перколяционных и фрактальных метрик собирается в процессе перколяционного моделирования (PM) и фрактального анализа (FA).

Генеративная вероятностно тематическая модель probabilistic topic model (PTM) выявляет и фиксирует тематику политематических по сути научно-информационных потоков, представляя каждую тему дискретным распределением вероятностей терминов, а каждый документ — дискретным распределением вероятностей тем.

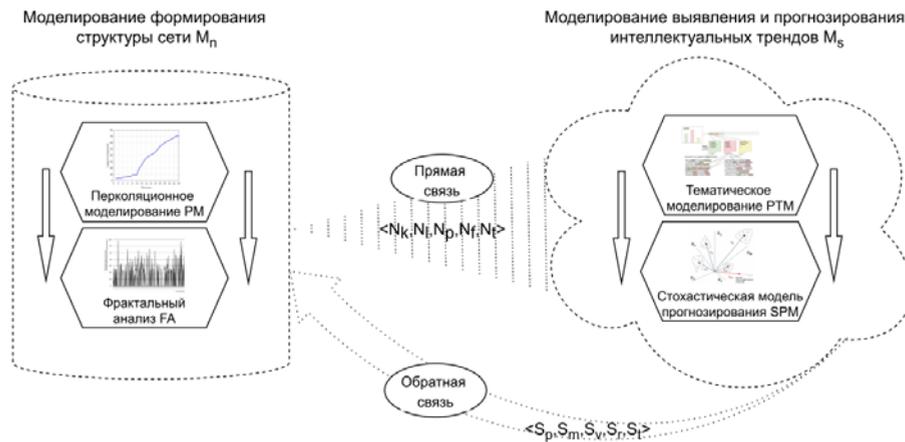


Рис. 1. Методический макет авторского стохастического метамоделирования ИЦЭ

Полученный в результате моделирования  $M_n$  комплекс сетевых параметров, критических показателей перколяции, показателей размерности фрактальных субструктур, а также параметров «степени» самоподобности влияет на точную настройку гиперпараметров РТМ. Темпоральная синхронизация интервалов РТМ и диапазона формирования перколяционного кластера также имеет значение при оптимизации моделей.

Тонкая настройка параметров стохастической модели прогнозирования (SPM) трендов реализуется за счет механизма, моделирующего ее поведение от условного времени вероятности перехода перколяционного порога. В рамках методологии стохастического моделирования данные механизмы взаимодействия выражаются прямой связью  $M_n \rightarrow M_s$  (см. рис. 1).

В свою очередь, семантический комплекс систематизированных концепций, терминов и тем, выявленных и упорядоченных в процессе РТМ, влияет на оптимизацию и детализацию в настройке параметров экспериментов РМ и FA, формирует новые области исследования. Оптимизация SPM работает в условиях двустороннего анализа взаимодействия с сетевым перколяционным моделированием РМ. Данные механизмы взаимодействия требуют постановки эксперимента с учетом обратной информационной связи  $M_s \rightarrow M_n$  (см. рис. 1).

Кроме того, информационное пространство высокой размерности и инфопотоки отличаются высокой плотностью обмена «больших данных», что приводит к нетривиальному характеру механизмов взаимодействия  $M_n \leftrightarrow M_s$ .

Таким образом, формальную Мета-модель ММ, выражающую авторскую методологию стохастического моделирования, можно представить следующим образом:

$$MM = \langle M_n, M_s, R_H, R_A, R_T \rangle, \quad (3)$$

где  $R$  – отношения между моделями  $M_n$  и  $M_s$ ;  $R_H$  – иерархические отношения;  $R_A$  – ассоциативные отношения;  $R_T$  – транзитивные отношения.

Для заданных множеств (кортежей) групп параметров  $M_n = \langle N_{ki}, N_i, N_p, N_f, N_t \rangle$  и  $M_s = \langle S_p, S_m, S_v, S_r, S_t \rangle$  построим функцию (классификатор)  $F$ :

$$F = M_n \times M_s \rightarrow \{\text{true}, \text{false}\}. \quad (4)$$

Функция  $F$  для пар параметров из разных кортежей  $(N_i, S_j)$  определяет истинностное значение true или false, причем true означает, что в параметре  $N_i$  кортежа  $M_n$  выражена какая-либо зависимость параметра  $S_j$  кортежа  $M_s$ ; иначе – false.

Отношения  $R$  между группами параметров моделей  $M_n$  и  $M_s$  представляют корреляции между номинальными категориальными переменными. Поэтому для описания используем бинарную шкалу, где значение колеблется от 0 до 1, где 0 указывает на отсутствие связи между переменными, а 1 указывает на наличие связи между переменными.

Для суммирования возможных зависимостей между всеми переменными в наборах данных моделирования и понимания взаимосвязей между ними построим таблицу сопряженности.

Для точного решения описательной задачи интеллектуального анализа наборов данных используем концепцию ДСМ-метода (правдоподобного вывода) автоматизированной поддержки научных исследований [21]. Целью ДСМ-метода является выявление причинно-следственных связей между множествами признаков (групп наборов параметров) и свойств (собственно параметров) [22].

На языке классической теории множеств, подмножество признаков (фрагмент)  $f = \{t_j\}$  будет являться возможной причиной интересующего свойства  $c_k$ , если данный фрагмент является максимальным сходством некоторого набора объектов, обладающих этим свойством (положительных примеров). Такие причины в ДСМ-методе представляются в форме гипотез  $h = \{f, c\}$ ,  $h \in H$ . На всем множестве  $H$  гипотезы могут быть как причинами наличия (положительные гипотезы), так и причинами отсутствия (отрицательные гипотезы) у объектов интересующих свойств.

Одна из процедур ДСМ-метода - вывод по аналогии, позволяет применять порожденные гипотезы для определения факта присутствия свойств у неопределенных объектов.

Для взвешивания гипотез в процессе классификации неопределенных объектов надо произвести оценку информативности, которую можно реализовать различными методами. В работе [22] показано, что наилучшим способом максимизации количества объектов, правильно распознаваемых гипотезой, и минимизации неправильно распознаваемых, является функция релевантной частоты RF (Relevance Frequency).

Данная функция относится к контролируемым методам взвешивания терминов, которые учитывают предварительную информацию о принадлежности документов к заранее определенным категориям (обучающим выборкам).

Основываясь на идее использования оценок выбора признаков, когда классификатор текста распределяет положительные и отрицательные тестовые документы, присваивая им разные баллы, получена формула [23]:

$$RF = \log \left( 2 + \frac{p}{\max(1, n)} \right), \quad (5)$$

где  $p$  – количество положительных примеров распределения документов в выборке,  $n$  – количество отрицательных примеров.

Таким образом, мы можем интерпретировать показатель RF, как основной показатель информативности при общей оценке веса зависимостей между параметрами  $N_i$  кортежа  $M_n$  и параметрами  $S_j$  кортежа  $M_s$ .

Данные оценки рассматриваются на бинарной шкале. Для обобщения оценки применения метамоделирования на  $n$ -арные шкалы и детализации с учетом иерархических  $R_H$ , ассоциативных  $R_A$  и транзитивных  $R_T$  отношений могут быть применены методы теории нечетких множеств.

Для формализации семантического пространства с возможностью анализа лингвистической переменной, выявления зависимостей тематически маркированной лексики и расчета весов термов применим семантический граф (СГ) с использованием алгоритма Гинзбурга [24]. Данному СГ соответствует система правил между двумя ключевыми словами (КС) СВЯЗЬ:

$$\{KC_j \xrightarrow{t_i} KC_k\}, \text{ где } t_i - \text{ сила связи.}$$

На основе этой концепции построены тройки элементов, используемых для реализации процедуры выявления причинно-следственных связей для проектируемой метамодел. К элементам такой концептуальной схемы отнесем: отношение, обозначающее связь между наборами параметров ( $R_H, R_A, R_T$ ), собственно два набора параметров из разных кортежей ( $N_i$  и  $S_j$ ) и возможные варианты представления любой тройки отношений в зависимости от уровня специфичности связи по отношению к данным наборам.

Весовые коэффициенты функций принадлежности термов лингвистической переменной определенной связи к одному из вариантов ее представления могут быть определены на основе анализа комплексных результатов стохастического метамоделирования (PM, FA, PTM, SPM).

### III. АПРОБАЦИЯ И РЕЗУЛЬТАТЫ

В исследовании есть два крупных направления моделирования, которые взаимодействуют между собой и в совокупности обеспечивают новый более глубокий взгляд на изучаемый объект.

*Моделирование формирования структуры сети  $M_n$ .*

В рамках разработанной методологии ход эксперимента задан параметрами кортежа  $M_n = \langle N_k, N_i, N_p, N_f, N_t \rangle$ , включающими группы:

$N_k$  – параметры, характеризующие общее состояние структуры сети: количество компонентов (узлов)  $i$ , наличие перколяционного кластера и его размер  $S$ , средняя дистанция  $l$ , диаметр  $d$ , коэффициент кластерности  $C_k$ ;

$N_i$  – параметры, характеризующие состояние связей в сети: общее число связей  $k$ , средняя степень вершин  $k_i$ , плотность  $D$ , центральность  $C$ , промежуточность  $B$ ;

$N_p$  – критические показатели перколяции  $\beta$ ,  $\nu$  и  $\gamma$ ;

$N_f$  – показатели самоподобия (фрактальные размерности)  $d_f$ ,  $d_{min}$  и  $d_i$ ;

$N_t$  – характеристики динамики (временных рядов): показатель Херста  $H$ , показатель отклонений  $\Delta L$ , коэффициент корреляции  $R(k)$ , индекс разброса дисперсии (фактор Фано)  $F(k)$ .

*Моделирование выявления и прогнозирования интеллектуальных трендов  $M_s$ .*

Для этого направления в ходе эксперимента необходимы параметры кортежа  $M_s = \langle S_p, S_m, S_v, S_r, S_t \rangle$ , включающие:

$S_p$  – гиперпараметры вероятностного тематического моделирования PTM:  $\beta_\omega$ ,  $\alpha$ , и регуляризаторы воздействия на матрицы  $\Phi$ ,  $\Theta$  и  $p_{idw}$ ;

$S_m$  – результаты PTM: множество выявленных тем  $T$  в коллекции, распределение вероятностей тем  $\theta_{id}$ , распределение вероятностей термов  $\varphi_{\omega i}$ ;

$S_v$  – метрики оценки качества PTM: перплексия (perplexity)  $P_m$ , когерентность (coherence)  $C_t$ , размер семантического ядра  $W_t$ , чистота  $pur_t$ , контрастность  $con_t$ , подъем  $\log Lift_t$ ;

$S_r$  – метрики оценки релевантности (меры сходства): косинусная  $cosine$ , ранговая корреляция Спирмена, KL-дивергенция;

$S_t$  – параметры моделирования стохастической динамики информационных процессов:  $x_\omega$ ,  $\varepsilon$  и  $\zeta$ .

Апробация предложенного подхода к исследованию ИЦЭ проводилась на основе обработки данных сетей цитирования научных публикаций и сетей научных терминов.

В ходе апробации методологии моделей направления  $M_n$  проведены обширные исследования динамики формирования научно-образовательных сетей с помощью моделирования методом Монте-Карло.

Получены многочисленные результаты по изучению параметров, характеризующих общее состояние структуры сети и состояние связей сети [25], а также критические показатели перколяции на примере исследования сетей терминов библиографической базы данных (ББД) Scopus [26], а также сетей цитат arXiv раздела hep-th по теоретической физике высоких энергий, включая их полное дерево цитирования [27].

В рамках ближайшей постановки находятся эксперименты с фрактальными показателями самоподобия и динамическими характеристиками (данными временных рядов, составленных из сетей цитирования научных публикаций). Оптимизирован объем выборки для тестирования индикаторов локального поведения временного ряда. Для валидации моделирования в качестве входных данных для временных рядов выбрана сеть аннотаций по теоретической физике высоких энергий (hep-th часть arXiv).

В ходе апробации моделей направления  $M_s$  и структурирования данных для PTM, при формировании глоссария прогностических терминов на базе сетевого сервиса «Википедия», применен комплексный алгоритмический подход, согласно которому интегрирован ряд условий, объединяющих возможности графового и семантического подходов [28].

В настоящее время на основе исследовательской базы данных, составленной из аннотаций взаимосвязанных статей раздела hep-th по теоретической физике высоких энергий ББД arXiv, проводятся исследования по оптимизации гиперпараметров и регуляризаторов тематических моделей на базе LDA и BigARTM, апробируются и калибруются основные метрики оценки качества PTM. Проводятся исследования по оценке релевантности (меры сходства) выявленных вероятностных распределений тем исходной концепции, а также эксперименты по разработке методики выявления скрытых концепций на основе совпадений в распределении вероятностей терминов.

Для решения прикладной цели исследования разрабатывается модифицированный алгоритм прогнозирования интеллектуальных трендов (SPM) на базе темпоральных вероятностных тематических распределений [29].

После сбора полного набора данных по завершенной схеме методического макета стохастического метамоделирования (PM, FA, PTM, SPM) будет реализована количественная валидация заявляемой методологии.

В зависимости от данного анализа на основе значимости каждой связи будет установлен интегрированный показатель информативности

метамодели.

#### IV. ЗАКЛЮЧЕНИЕ

Предложена концепция и структура стохастического метамоделирования интеллектуальных цифровых экосистем – цифровых двойников, имитирующих организационные системы с выраженным характером передачи знаний.

Теоретическое значение исследования заключается в разработке новых подходов к комплексному стохастическому моделированию эффективности деятельности динамических ИЦЭ с использованием оригинальных вычислительных методов и алгоритмов.

Основное прикладное значение работы заключается в том что, комбинация парадигмы стохастического моделирования и концепции цифрового двойника, формирует универсальную методику, позволяющую анализировать деятельность открытых динамических ИЦЭ любых территориальных уровней (как макро – отрасль, страна, так и микро – отдельная организация), любого информационного содержания (как технического, так и гуманитарного), любой организационной структуры (как реальной организации, так и виртуальной).

В рамках развития концепции интеллектуальных цифровых двойников представляется возможным применения данной методологии для системы мониторинга реальных физических систем, аналитики для извлечения знаний из данных, прогнозирования и оптимизации производственных процессов.

#### БИБЛИОГРАФИЯ

- [1] Темкин И. О., Клебанов Д. А., Дерябин С. А., Конов И. С. Построение интеллектуальной геоинформационной системы горного предприятия с использованием методов прогнозной аналитики // Горный информационно-аналитический бюллетень. – 2020. – № 3. – С. 114–125. DOI: 10.25018/0236-1493-2020-3-0-114-125.
- [2] Попов, О. Р. Способ поиска параметров порядка самоорганизующихся систем: информационный аспект / О. Р. Попов // Интеллектуальные ресурсы - региональному развитию. – 2020. – № 2. – С. 64-70. URL: <https://www.elibrary.ru/item.asp?id=43033193>.
- [3] Еремин А. Л. Ноогенез и теория интеллекта. Краснодар: СовКуб, 2005. — 356 с.
- [4] Хакен Г. Информация и самоорганизация. Макроскопический подход к сложным системам. – М.: Мир, 1991. – 240 с.
- [5] Попов, О. Р. Алгоритмы построения интеллектуальных систем обработки текстовой информации для задачи анализа мнений / О.Р. Попов, Е.В. Гребенюк // Интеллектуальные ресурсы - региональному развитию. – 2021. URL: <https://www.elibrary.ru/item.asp?id=46659138>.
- [6] Popov, O. R. Phenomenon of Information and Informational Ecology: Interaction and Definitions on the Language of Soft Computing / O. R. Popov, B. V. Martynov // 14th International Conference on Theory and Application of Fuzzy Systems and Soft Computing – ICAFS-2020, Budva, Montenegro, 27–28 августа 2020 года. – Budva, Montenegro: Springer International Publishing, 2021. – P. 694-701. URL: <https://elibrary.ru/item.asp?id=46509214>.
- [7] Research of tools for monitoring changes in natural and anthropogenic-transformed ecosystems / A. Belyaev, S. Kramarov, O. Mityasova [et al.] // IOP Conference Series: Earth and Environmental Science : Fundamental and Applied Scientific Research in the Development of Agriculture in the Far East (AFE 2021), Ussurijsk, 15 ноября 2021 года. – Ussurijsk: IOP Publishing Ltd, 2021.
- [8] About Some Issues of Developing Digital Twins for The Intelligent Process Control in Quarries / Deryabin, S.A., Temkin, I.O., Zykov, S.V. // Procedia Computer Science, 2020, 176, P. 3210–3216.
- [9] Прохоров А., Лысачев М. Научный редактор профессор Боровков А. Цифровой двойник. Анализ, тренды, мировой опыт. – М.: ООО «АльянсПринт», 2020. – 401 с.
- [10] Анализ тенденций развития цифровых двойников нового поколения / В. А. Сухомлин, Д. Е. Намиот, Д. А. Гапанович // International Journal of Open Information Technologies. – 2024. – Т. 12, № 7. – С. 119-130.
- [11] Цифровой бетон: открытый BIM, машиночитаемые стандарты, IoT, цифровые двойники, логистика 4.0, бережливое строительство и другие промышленные подходы на примерах транспортных инфраструктур / В. П. Куприяновский, О. Н. Покусаев, Д. Е. Намиот [и др.] // International Journal of Open Information Technologies. – 2021. – Т. 9, № 9. – С. 133-173.
- [12] Высокоэффективный способ приготовления парафиновой пасты – полифункциональной добавки для бетонов / Попов О.Р., Попов Р.И., Курочка П.Н., Промтов М.А. // Транспортное строительство. - 1994. - № 1. – С. 36-37.
- [13] Современные проблемы и перспективы интернационализации интеллектуальных ресурсов России: вызовы, стратегии, модели, интересы национального, регионального и отраслевого развития / А. А. Акишина, И. В. Антипина, А. И. Богуш [и др.]. – Москва : Издательский Центр РИОР, 2019. – 287 с. – ISBN 978-5-369-02021-0. – DOI 10.29039/02021-0. – EDN DHXUEA.
- [14] Перспективы и возможности формирования системы экспертно-аналитического сопровождения международной деятельности российских университетов : коллективная монография / Под ред. С.О. Крамарова, Н.В. Пелихова, В.И. Скоробогатовой. — Москва : РИОР, 2020. — 295 с. — (Международная деятельность). — DOI: <https://doi.org/10.29039/02044-9>.
- [15] Цифровые двойники и системы дискретно-событийного моделирования / Д. Е. Намиот, О. Н. Покусаев, В. П. Куприяновский, М. Г. Жабицкий // International Journal of Open Information Technologies. – 2021. – Т. 9, № 2. – С. 70-75.
- [16] Xia, Yuchen, et al. LLM experiments with simulation: Large Language Model Multi-Agent System for Process Simulation Parametrization in Digital Twins. arXiv preprint arXiv:2405.18092 (2024).
- [17] Jacoby M., Uslander T. Digital twin and internet of things — Current standards landscape // Applied Sciences. – 2020. – Т. 10. – С. 6519.
- [18] Ming Li, Run-Ran Liu, Linyuan Lu, Mao-Bin Hu, Shuqi Xu, Yi-Cheng Zhang. Percolation on complex networks: Theory and application. Physics Reports. 907: 1–68. DOI: 10.1016/j.physrep.2020.12.003.
- [19] Инкрементальное обучение тематических моделей для поиска трендовых тем в научных публикациях / Н. А. Герасименко, А. С. Чернявский, М. А. Никифорова [и др.] // Доклады Российской академии наук. Математика, информатика, процессы управления. – 2022. – Т. 508, № 1. – С. 106-108.
- [20] Жуков Д. О., Хватова Т. Ю., Зальцман А. Д. Моделирование стохастической динамики изменения состояний узлов и перколяционных переходов в социальных сетях с учетом самоорганизации и наличия памяти // Информатика и её применения. – 2021. – Т. 15. – № 1. – С. 102-110. – DOI: 10.14357/19922264210114.
- [21] Финн В. К. Эпистемологические основания ДСМ-метода. Ч. I // НТИ. Сер. 2. Информационные процессы и системы. 2013. № 9. – С. 1–29.
- [22] Котельников, Е.В. Выбор структуры данных для представления гипотез в варианте ДСМ-метода анализа текстов / Е.В. Котельников // Фундаментальные исследования. – 2016. – № 10(2). – С. 301–305.
- [23] Lan M., Tan C. L., Su J., Lu Y. Supervised and Traditional Term Weighting Methods for Automatic Text Categorization // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009. Vol. 31. No. 4. – P. 721–735.
- [24] Воронина, И. Е. Компьютерное моделирование лингвистических объектов : монография / И. Е. Воронина ; И. Е. Воронина. – Воронеж : Изд.-полиграфический центр Воронежского гос. ун-та, 2007. – 177 с.
- [25] Попов, О. Р. Исследование распространения информации в сетях, структурированных из набора прогностических терминов. / О.Р. Попов, С.О. Крамаров // Вестник кибернетики. – 2022. – № 1(45). – С. 38-45. – DOI 10.34822/1999-7604-2022-1-38-45.
- [26] Динамика формирования связей в сетях, структурированных на основе прогностических терминов / С. О. Крамаров, О. Р. Попов, И. Э. Джариев, Е. А. Петров // Russian Technological Journal. – 2023. – Т. 11, № 3. – С. 17-29. – DOI 10.32362/2500-316X-2023-11-3-17-29.
- [27] Перколяция и формирование связности в динамике сетей цитирования данных по физике высоких энергий / С. О. Крамаров, О. Р. Попов, И. Э. Джариев, Е. А. Петров // Russian Technological Journal, представлено к публикации.
- [28] Попов О. Р., Гросу А., Крамаров С. О. Комплексный сетевой алгоритм формирования глоссария контекстно-близких прогностических терминов // Современные информационные технологии и ИТ-образование. 2023. Т. 19, № 4. С. 910-926. doi: <https://doi.org/10.25559/SITITO.019.202304.910-926>.
- [29] Попов, О. Р. Адаптация мировых практик к проблеме долгосрочного технологического прогнозирования состояния самоорганизующихся интеллектуальных систем / О.Р. Попов // Интеллектуальные ресурсы - региональному развитию. – 2021. URL: <https://www.elibrary.ru/item.asp?id=46659136>.

# The concept of stochastic metamodeling of intelligent digital ecosystems

O.R. Popov

**Abstract** — The concept of stochastic metamodeling of intelligent digital ecosystems (IDE) is proposed – digital twins simulating organizational systems with a pronounced nature of knowledge transfer. The implemented methodology combines network analysis models (N approach) and probabilistic models of distributive semantics (S approach). The N approach includes percolation modeling (PM) and fractal analysis (FA) methods. The S approach includes probabilistic topic modeling (PTM) and stochastic trend forecasting model (SPM). All of the listed models can be solved by computational methods simulating parameters and variables by computer algorithms.

During the sequential implementation of modeling algorithms, two tuples are formed, composed of sets of data parameters obtained during modeling – network  $M_n$  and semantic  $M_s$ . A formal meta-model of MM is presented, expressing the author's methodology of stochastic modeling, composed of the models  $M_n$  and  $M_s$  themselves, as well as hierarchical  $R_H$ , associative  $R_A$  and transitive  $R_T$  relations between the models. For an accurate solution of the descriptive problem of intellectual analysis of possible dependencies between variables in data sets, the concept of the DSM-method (plausible inference) of automated support of scientific research and methods of fuzzy set theory are proposed.

The general results of testing the proposed approach to the study of the IDE based on processing data from citation networks of scientific publications and networks of scientific terms compiled from samples of bibliographic databases arXiv, Scopus and Google Scholar are considered.

**Keywords** — algorithm, intelligent digital ecosystem, modeling, parameter, semantics, network, stochastic model.

## REFERENCES

- [1] Temkin I. O., Klebanov D. A., Deryabin S. A., Konov I. S. Postroenie intellektual'noj geoinformacionnoj sistemy gornogo predpriyatiya s ispol'zovaniem metodov prognoznoj analitiki // Gornyj informacionno-analiticheskij byulleten'. – 2020. – #. 3. – S. 114–125. DOI: 10.25018/0236-1493-2020-3-0-114-125.
- [2] Popov, O. R. Sposob poiska parametrov poryadka samoorganizuyushchih sistem: informacionnyj aspekt / O. R. Popov // Intellektual'nye resursy - regional'nomu razvitiyu. – 2020. – #. 2. – S. 64-70. URL: <https://www.elibrary.ru/item.asp?id=43033193>.
- [3] Eremin A. L. Noogenez i teoriya intellekta. Krasnodar: SovKub, 2005. — 356 s.
- [4] Haken G. Informaciya i samoorganizaciya. Makroskopicheskij podhod k slozhnym sistemam. – M.: Mir, 1991. – 240 s.
- [5] Popov, O. R. Algoritmy postroeniya intellektual'nyh sistem obrabotki tekstovoj informacii dlya zadachi analiza mnenij / O.R. Popov, E.V. Grebenyuk // Intellektual'nye resursy - regional'nomu razvitiyu. – 2021. URL: <https://www.elibrary.ru/item.asp?id=46659138>.
- [6] Popov, O. R. Phenomenon of Information and Informational Ecology: Interaction and Definitions on the Language of Soft Computing / O. R. Popov, B. V. Martynov // 14th International Conference on Theory and Application of Fuzzy Systems and Soft Computing – ICAFS-2020, Budva, Montenegro, 27–28 avgusta 2020 goda. – Budva, Montenegro: Springer International Publishing, 2021. – P. 694-701. URL: <https://elibrary.ru/item.asp?id=46509214>.
- [7] Research of tools for monitoring changes in natural and anthropogenic-transformed ecosystems / A. Belyaev, S. Kramarov, O. Mityasova [et al.] // IOP Conference Series: Earth and Environmental Science : Fundamental and Applied Scientific Research in the Development of Agriculture in the Far East (AFE 2021), Ussurijsk, 15 noyabrya 2021 goda. – Ussurijsk: IOP Publishing Ltd, 2021.
- [8] About Some Issues of Developing Digital Twins for The Intelligent Process Control in Quarries / Deryabin, S.A., Temkin, I.O., Zykov, S.V. // Procedia Computer Science, 2020, 176, P. 3210–3216.
- [9] Prohorov A., Lysachev M. Nauchnyj redaktor professor Borovkov A. Cifrovoy dvojniki. Analiz, trendy, mirovoj opyt. – M.: OOO «Al'yansPrint», 2020. – 401 s.
- [10] Analiz tendencij razvitiya cifrovoy dvojniki novogo pokoleniya / V. A. Suhomlin, D. E. Namiot, D. A. Gapanovich // International Journal of Open Information Technologies. – 2024. – T. 12, # 7. – S. 119-130.
- [11] Cifrovoy beton: otkrytyj BIM, mashinochitaemye standarty, IoT, cifrovye dvojniki, logistika 4.0, berezhlivoe stroitel'stvo i drugie promyshlennye podhody na primerah transportnyh infrastruktur / V. P. Kupriyanovskij, O. N. Pokusaev, D. E. Namiot [i dr.] // International Journal of Open Information Technologies. – 2021. – T. 9, # 9. – S. 133-173.
- [12] Vysokoeffektivnyj sposob prigotovleniya parafinovoj pasty – polifunkcional'noj dobavki dlya betonov / Popov O.R., Popov R.I., Kurochka P.N., Promtov M.A. // Transportnoe stroitel'stvo. - 1994. - # 1. - S. 36-37.
- [13] Sovremennye problemy i perspektivy internacionalizacii intellektual'nyh resursov Rossii: vyzovy, strategii, modeli, interesy nacional'nogo, regional'nogo i otraslevogo razvitiya / A. A. Akishina, I. V. Antipina, A. I. Bogush [i dr.]. – Moskva : Izdatel'skij Centr RIOR, 2019. – 287 s. – ISBN 978-5-369-02021-0. – DOI 10.29039/02021-0. – EDN DHXUEA.
- [14] Perspektivy i vozmozhnosti formirovaniya sistemy ekspertno-analiticheskogo soprovozhdeniya mezhdunarodnoj deyatelnosti rossijskih universitetov : kollektivnaya monografiya / Pod red. S.O. Kramarova, N.V. Pelihova, V.I. Skorobogatovoj. — Moskva : RIOR, 2020. — 295 s. DOI: <https://doi.org/10.29039/02044-9>.
- [15] Cifrovye dvojniki i sistemy diskretno-sobyitnogo modelirovaniya / D. E. Namiot, O. N. Pokusaev, V. P. Kupriyanovskij, M. G. Zhabickij // International Journal of Open Information Technologies. – 2021. – T. 9, # 2. – C. 70-75.
- [16] Xia, Yuchen, et al. LLM experiments with simulation: Large Language Model Multi-Agent System for Process Simulation Parametrization in Digital Twins. arXiv preprint arXiv:2405.18092 (2024).
- [17] Jacoby M., Uslander T. Digital twin and internet of things — Current standards landscape // Applied Sciences. – 2020. – T. 10. – S. 6519.
- [18] Ming Li, Run-Ran Liu, Linyuan Lu, Mao-Bin Hu, Shuqi Xu, Yi-Cheng Zhang. Percolation on complex networks: Theory and application. Physics Reports. 907: 1–68. DOI: 10.1016/j.physrep.2020.12.003.
- [19] Inkremental'noe obuchenie tematiceskikh modelej dlya poiska trendovyh tem v nauchnyh publikacijah / N. A. Gerasimenko, A. S. Chernyavskij, M. A. Nikiforova [i dr.] // Doklady Rossijskoj akademii nauk. Matematika, informatika, processy upravleniya. – 2022. – T. 508, # 1. – S. 106-108.
- [20] Zhukov D. O., Hvatova T. Yu., Zal'cman A. D. Modelirovanie stohasticheskoj dinamiki izmeneniya sostoyanij uzlov i perkolyacionnyh perekhodov v social'nyh setyah s uchedom samoorganizacii i nalichiya pamyati // Informatika i eyo primeneniya. – 2021. – T. 15. – # 1. – S. 102-110. DOI: 10.14357/19922264210114.
- [21] Finn V. K. Epistemologicheskie osnovaniya DSM-metoda. Ch. I // NTI. Ser. 2. Informacionnye processy i sistemy. 2013. # 9. – S. 1–29.
- [22] Kotelnikov, E.V. Vybory struktury dannyh dlya predstavleniya gipotez v variante DSM-metoda analiza tekstov / E.V. Kotelnikov // Fundamental'nye issledovaniya. – 2016. – # 10(2). – S. 301–305.
- [23] Lan M., Tan C. L., Su J., Lu Y. Supervised and Traditional Term Weighting Methods for Automatic Text Categorization // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009. Vol. 31. No. 4. – P. 721–735.
- [24] Voronina, I. E. Komp'yuternoe modelirovanie lingvisticheskikh ob'ektov : monografiya / I. E. Voronina ; I. E. Voronina. – Voronezh : Izd.-poligraficheskij centr Voronezhskogo gos. un-ta, 2007. – 177 s.
- [25] Popov, O. R. Issledovanie rasprostraneniya informacii v setyah, strukturirovannyh iz nabora prognosticheskikh terminov. / O.R. Popov, S.O. Kramarov // Vestnik kibernetiki. – 2022. – # 1(45). – S. 38-45. – DOI 10.34822/1999-7604-2022-1-38-45.
- [26] Dinamika formirovaniya svyazey v setyah, strukturirovannyh na osnove prognosticheskikh terminov / S. O. Kramarov, O. R. Popov, I. E. Dzharijev, E. A. Petrov // Russian Technological Journal. – 2023. – T. 11, # 3. – S. 17-29. – DOI 10.32362/2500-316X-2023-11-3-17-29.
- [27] Perkolyaciya i formirovanie svyaznosti v dinamike setej citirovaniya dannyh po fizike vysokih energij / S. O. Kramarov, O. R. Popov, I. E. Dzharijev, E. A. Petrov // Russian Technological Journal, submitted for publication.
- [28] Popov O. R., Grosu A., Kramarov S. O. Kompleksnyj setevoj algoritm formirovaniya glossariya kontekstno-blizkih prognosticheskikh terminov // Sovremennye informacionnye

- tehnologii i IT-obrazovanie. 2023. T. 19, # 4. S. 910-926. doi: <https://doi.org/10.25559/SITITO.019.202304.910-926>.
- [29] Popov, O. R. Adaptaciya mirovyh praktik k probleme dolgosrochnogo tekhnologicheskogo prognozirovaniya sostoyaniya samoorganizuyushchih-sya intellektual'nyh sistem / O.R. Popov // Intellektual'nye resursy - regional'nomu razvitiyu. – 2021. URL: <https://www.elibrary.ru/item.asp?id=46659136>.