

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(ВИНИТИ РАН)

НАУЧНО · ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 1. ОРГАНИЗАЦИЯ И МЕТОДИКА
ИНФОРМАЦИОННОЙ РАБОТЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 11

Москва 2023

ОБЩИЙ РАЗДЕЛ

УДК [004+002]:001.5

И.М. Зацман

Научная парадигма информатики как третьей культуры*

Дано описание первых результатов создания научной парадигмы информатики, объединяющей широкий спектр информационных и компьютерных наук. Предметная область информатики рассматривается в рамках концепции полиадического компьютеринга Пола Розенблума. Основная цель работы – построение фрагментов верхних уровней двух классификаций сущностей предметной области информатики как третьей культуры (наряду с естественными и социально-гуманистическими науками, которые Ч. Сноу назвал двумя культурами). При этом используется ряд уже известных оснований для построения трех уровней классификации объектов предметной области (первая классификация) и двух уровней классификации трансформаций этих объектов (вторая классификация). Обе классификации планируется использовать для создания научной парадигмы информатики как третьей культуры. Предлагаемые фрагменты классификаций позиционируются как начальная стадия процесса формирования всей парадигмы. Рассматриваются контекст и предпосылки создания предлагаемой парадигмы и то, чем первые результаты ее формирования могут быть полезны уже сегодня.

* Исследование выполнено с использованием ЦКП «Информатика» ФИЦ ИУ РАН.

ВВЕДЕНИЕ

12 апреля 2022 г. профессор Университета Южной Калифорнии Пол Розенблум сделал итоговый обзор научных результатов, полученных им с середины-конца 1970-х гг. [1]. Одно из фундаментальных достижений профессора представляет собой пересмотр теоретических оснований компьютеринга, обусловленный предлагаемым им существенным расширением его предметной области (к информатике перейдем далее). В процессе расширения этой предметной области вся система научного знания была разделена им на четыре отрасли (*физические науки* в широком контексте, включая химию, геологию, астрономию и т.д., *науки о жизни, социально-гуманитарные науки и компьютеринг*), дано описание предметной области компьютеринга и применения его методов в трех других отраслях знания. Сфера применения методов, моделей и средств компьютеринга в этих трех отраслях знания, объединенная с предметной областью компьютеринга, была названа профессором как *polyadic computing* [2]. Отличие компьютеринга как отрасли знания от *polyadic computing* состоит в том, что сфера применения методов, моделей и средств компьютеринга в трех других отраслях знания не относится к предметной области самого компьютеринга как четвертой отрасли системы научного знания по П. Розенблуму.

Судьба инновационных идей П. Розенблума сложилась по-разному в США и в европейских странах. В первом случае на основе его идей им был создан новый курс обучения – «Введение в компьютеринг», из которого через несколько лет была исключена основная часть материала. Одной из причин исключения была негативная реакция слушателей этого курса [1]. В европейских странах идеи П. Розенблума были частично использованы при разработке стратегии компьютерного образования [3], которой посвящен следующий раздел настоящей статьи.

Существенное расширение П. Розенблумом предметной области компьютеринга в контексте крупного деления системы научного знания на четыре отраслиозвучны нашей позиции, но мы предлагаем всю систему знания разделить не на четыре отрасли, а на три культуры: точную и естественно-научную (science), социально-гуманитарную (humanities) и информационно-компьютерную (informatics). В контексте деления системы знания на три культуры в качестве перевода термина *polyadic computing* мы используем термин «информатика» как синоним внушительного спектра информационно-компьютерных наук с существенно расширенным представлением о предметной области информатики по сравнению с традиционными представлениями о ней. Предлагаемое деление во многом было подсказано известным сопоставлением первых двух культур [4, 5].

Исаия Берлин описал контраст между точными и естественными (далее кратко – естественные) и социально-гуманитарными науками так: «Конкретное и уникальное против повторяющегося и универсального, конкретное против абстрактного, ... качество против количества ... вот некоторые из аспектов контрастного сопоставления» [5, с. 109]¹. Согласно создаваемой парадигме в предметной области информатики присутствуют и конкретные сущности с качественными характеристиками, и абстрактные сущности с количественными. Подчеркнем, что методами, системами и средствами информатики первые могут преобразовываться во вторые и это преобразование может быть полностью или частично обратимым [6].

Важно отметить, что трансформации конкретных объектов в абстрактные не изучаются ни точными, ни естественными, ни социально-гуманитарными науками. В них не исследуются условия обратимости таких трансформаций. Однако именно методы, модели и средства информатики могут быть применены и для исследования, и для реализации подобных трансформаций [7, 8], что служит первым доводом в пользу предлагаемого позиционирования информатики как третьей культуры, дополняющей естественно-научную и социально-гуманитарную культуры в части описания трансформаций конкретных объектов в абстрактные, включая полностью и частично обратимые трансформации.

Второй довод – это более широкий спектр природы объектов предметной области информатики по сравнению и с точными, и с естественными, и с социально-гуманитарными науками (см. объекты информатики в пяти средах разной природы в разделе 2 настоящей статьи). Для такого разнообразия природы объектов ее предметной области нет своего «пространства» в традиционной дихотомии естественно-научной и социально-гуманитарной культур.

Предлагаемое позиционирование информатики как третьей культуры обуславливает объединение самого широкого спектра уже существующих информационно-компьютерных наук и смежных им дисциплин, в том числе, информационной науки [9–11], компьютерной науки [12, 13], теории генерации и управления знаниями [14, 15] и науки о данных [16]. Наша основная цель состоит в описании структуры научной парадигмы информатики и верхних уровней двух классификаций сущностей ее предметной области. В используемой нами структуре научной парадигмы эти две классификации принадлежат к одному из ее восьми компонентов согласно Абраму Соломонику [17–20].

¹ “The specific and unique versus the repetitive and the universal, the concrete versus the abstract, ... quality versus quantity ... these are some of the aspects of the contrast” [5, p. 109].

После описания ряда вопросов европейского компьютерного образования (раздел 1) мы рассматриваем контекст и предпосылки создания предлагаемой парадигмы, включая идеи П. Розенблума как одну из предпосылок (раздел 2), а затем даем описание трех верхних уровней классификации объектов исследования в информатике и двух верхних уровней классификации трансформаций этих объектов (раздел 3). Предлагаемая парадигма информатики – это только один из возможных вариантов теоретических оснований стратегии компьютерного образования. Отметим, что потребность в ее научной парадигме отчетливо проявилась в процессе подготовки стратегии европейского компьютерного образования [3].

1. О СТРАТЕГИИ ЕВРОПЕЙСКОГО КОМПЬЮТЕРНОГО ОБРАЗОВАНИЯ

Апрельский номер журнала «Communications of the ACM» за 2019 г. включает специальный раздел «Информатика в Европе», состоящий из статей, посвященных актуальным темам в сфере информатики как науки, вопросам ее преподавания и средствам разработки информационных технологий (ИТ) [21]. Раздел имеет 14 статей, в которых, в частности, описываются:

- факторы, определяющие направления перспективных европейских исследований и разработок [22];
- кадровый ландшафт европейского сектора ИТ [23];
- подход к позиционированию информатики в системе современного научного знания и построению стратегии европейского компьютерного образования [24].

По мнению авторов статьи [24], информатику следует рассматривать как дисциплину, не менее важную в системе образования, чем математика, естественные науки и иностранные языки. В этой статье анализируется «Первый план действий в области цифрового образования», одобренный Еврокомиссией (ЕК) в 2018 г. [25, 26]², а также Отчет с данными мониторинга преподавания информатики в европейском регионе, включая РФ и Ватикан, а также Израиль [28] (далее – Отчет).

В этом Отчете сформулирован ряд рекомендаций, ориентированных на улучшение ситуации с изучением информатики в Европе на всех уровнях образования. На основе материалов Отчета была подготовлена стратегия европейского компьютерного образования (далее – стратегия ЕКО или стратегия), одобренная Европейским комитетом по информатике (the Informatics Europe Board) и Европейским советом ACM (the ACM Europe Council). В стратегии и в Отчете информатика позиционируется как фундаментальная наука со своими собственными научными моделями, методами и подходами [3, 28].

По ряду аспектов позиции стратегии совпадают с инициативой Президента США от 2016 г. [29]. Но есть и одна принципиально иная позиция, сущес-

твенно отличающая стратегию от этой инициативы, – это диадический подход к построению стратегии ЕКО (two-tier strategy): согласно этой стратегии, с одной стороны, в курсах по информатике преподаются свойственные именно ей методы и модели, а с другой стороны, в рамках курсов по другим дисциплинам рассматривается *применение* методов, моделей и средств информатики.

1.1. Мониторинг состояния европейского компьютерного образования

Отчет [28] содержит результаты мониторинга состояния систем образования в области информатики³, а также результаты сопоставительного анализа этих систем для 55 административных единиц (AE) Европы⁴. В *Приложении С* этого Отчета приведены схемы и карты с результатами мониторинга и сопоставительного анализа, позволяющие сравнивать состояние систем образования в области информатики в разных AE между собой. Каждая карта или схема отражает, как правило, одну характеристику состояния сопоставляемых систем. В *Приложении D* Отчета приведено структурированное описание состояния системы образования в области информатики для каждой из 55 AE по комплексу характеристик⁵.

Поскольку авторы Отчета на основе данных мониторинга анализировали состояние преподавания информатики, они посчитали необходимым хоть как-то описать ее как науку. В 2017 г., когда был опубликован Отчет [28], отсутствовало конвенциональное определение предметной области информатики (сегодня ситуация не изменилась). Поэтому перед его авторами всталась весьма трудная задача, как охарактеризовать информатику. Их заслуживающий внимания подход включает две стадии весьма приблизительного описания этой предметной области. Сначала они сформулировали следующие положения:

- информатика кардинально изменилась за последние 50 лет и поэтому ранее используемые ее определения устарели (о чём еще в 2007 г., т. е. за 10 лет до публикации Отчета [28], писал Питер Деннинг [30]);
- современную информатику сложно уложить в «прокрустово ложе» естественных и/или технических наук;
- по сравнению с естественными науками информатику как научную дисциплину определить сложнее, так как у нее нет эмпирических основ как у них;

³ Подготовка к мониторингу началась в 2013 г. На сбор данных ушло два года. На основе результатов мониторинга в 2017 г. был опубликован Отчет [28].

⁴ Если в стране единая система образования в области информатики, то AE – это страна. Если в стране есть регионы с автономными системами образования, то AE – регион этой страны, например, Бавария. В Бельгии – 3 таких региона, в Соединенном королевстве – 4, в Испании – 15, в Германии – 16.

⁵ Структурированное описание состояния системы образования в области информатики в РФ приведено на с. 196 и 197 Отчета [28].

² Есть новый вариант плана действий Еврокомиссии в области цифрового (компьютерного) образования (2021–2027) [27].

- информатика – это нечто большее, чем использование только логических выводов (как в математике);
- информатика – это далеко не только сочетание инженерных принципов и технологий.

Перечисленные положения используются авторами Отчета на первой стадии описания информатики как образовательной дисциплины, которое естественно не претендует на ее строгое определение [28]. На второй стадии в Отчете перечисляются три публикации [31–33], которые, по мнению его авторов, содержат отдельные комплементарные характеристики информатики. По сути эти публикации можно рассматривать как три теоретических источника генезиса стратегии ЕКО.

1.2. Три теоретических источника стратегии ЕКО

На первых этапах становления информатики объектами ее исследований были в основном компьютеры и окружающие их явления. Например, в 1967 г. А. Ньюэлл, А. Дж. Перлес и Х. А. Саймон писали: «Везде, где есть феномены, может существовать наука для их описания и объяснения. ... Феномены порождают науки. Есть компьютеры. Следовательно, информатика [как компьютерная наука] предназначена для описания и изучения компьютеров. Явления, окружающие компьютеры, предстают как разнообразные, сложные и ценные»⁶ [34]. Такой подход к определению информатики как компьютерной науки долгие годы был доминирующим в системе компьютерного образования.

В 1986 г. в работе [31]⁷ был предложен концептуально иной поход к определению информатики именно как фундаментальной науки. Второй раздел «The definition of the term "informatics"» этой работы посвящен определению термина «информатика». Сначала, до определения этого термина, Кристен Нюгор пишет:

«Термин "computer science" следует заменить на "informatics". Несколько лет назад выбор между этими двумя терминами казался, скорее всего, несущественным. Обсуждения терминологии часто считают праздными, но иногда они могут отражать ключевые различия во мнениях или, по крайней мере, существенные акценты. Сегодня, к сожалению, используется термин "computer science". Этот термин имеет тенденцию поддерживать слишком узкое представление об информационных системах, которые в настоящее время интегрируют коллективы людей и разнообразные средства обработки информации, вза-

⁶ “Wherever there are phenomena, there can be a science to describe and explain those phenomena. ... Phenomena breed sciences. There are computers. Ergo, computer science is the study of computers. The phenomena surrounding computers are varied, complex, rich” [34, p. 1373].

⁷ Первый из трех источников стратегии ЕКО – это доклад Кристина Нюгера (Университет Осло), который был представлен им на Всемирном конгрессе IFIP (International Federation of Information Processing – Международная федерация по обработке информации) в Дублине в 1986 г.

имодействующие как посредством межличностных связей, так и посредством (все возрастающей доли) электронных каналов связи»⁸ [31]. Затем, после констатации важности четкого определения терминов «информатика» и «информация», он предлагает свою дефиницию первого термина: «Информатика – это наука, предметом которой являются информационные процессы и *связанные с ними феномены в артефактах, обществе и природе* (курсив мой – И. З.)»⁹ [31].

Далее К. Нюгор цитирует определение понятия «феномен» по словарю Webster 1960 г. («любой факт, обстоятельство или событие, которое сознание воспринимает сенсорно и может быть научно описано или оценено»¹⁰ [35]) и потом дает расширенное толкование этого понятия, которое кардинально отличается от его определения в словаре Webster 1960 г.:

«Важными примерами феноменов являются: живые организмы, неодушевленные объекты (включая артефакты, такие как, например, машины), события и процессы (например, выполнение компьютерных программ). Мы также можем говорить о *когнитивных феноменах, происходящих в сознании людей*, в отличие от явных [сенсорно воспринимаемых] феноменов, находящихся *вне сознания* (курсив мой – И. З.)»¹¹ [31].

Толкование К. Нюгера с делением феноменов, *связанных с информационными процессами*, на когнитивные, происходящие в сознании людей и формирующие концепты знания человека, и сенсорно воспринимаемые (например, тексты), существующие вне сознания, было в то время концептуально новым подходом к определению информатики как науки. Отметим, что одно из определений термина «феномен», расширенное по объему его значения, позже было включено и в современный онлайн-словарь Merriam-Webster¹² (в новом определении под номе-

⁸ “The term "computer science" should be replaced by "informatics". Some years ago the choice between the two terms seemed perhaps to be of little consequence. Discussions of terminology often are idle, but they may sometimes reflect basic differences of opinion or at least emphasis. Today it is unfortunate that the term "computer science" is used. The term tends to implant a too narrow way of thinking about information systems, that now are networks of people, information processing equipment and other machinery, interacting through direct inter-human and (an increasing proportion of) electronically supported communication links” [31, p. 4].

⁹ “Informatics is the science that has as its domain information processes and related phenomena in artifacts, society and nature” [31, p. 4].

¹⁰ “... any fact, circumstance, or experience that is apparent to the senses and that can be scientifically described or appraised” [35].

¹¹ “Important examples of phenomena are: living organisms, inanimate objects (including artifacts, like e.g. machines), events, and processes (e.g. computer program executions). We may also speak of cognitive phenomena, occurring within the minds of people, as opposed to manifest phenomena, occurring outside minds” [31, p. 4].

¹² В современном онлайн-словаре Merriam-Webster приводится следующая дефиниция под номером 2c для термина «феномен»: «Факт или событие, представляющие научный интерес, поддающееся научному описанию и объяснению – a fact or event of scientific interest susceptible to scientific descrip-

ром 2с удалена сенсорная воспринимаемость для феноменов, поддающихся научному описанию).

Можно предположить, что выбор авторами Отчета [28] работы [31], опубликованной более 35 лет назад, был обусловлен как раз существенным расширением предметной области информатики за счет включения широкого спектра феноменов, связанных с информационными процессами в технических, живых и социальных системах.

Из трех публикаций, выбранных авторами Отчета, второй источник – это книга Дэвида Харела (Научный институт Вейцмана), опубликованная в 1987 г., с изложением ряда важных и базовых тем информатики с алгоритмической точки зрения. В книге подчеркивается фундаментальный характер информатики и перечисляются три вида сложности проблем этой науки: вычислительная, системная и когнитивная [32].

Для определения предметной области информатики существенный интерес представляет сложность проблем третьего вида – когнитивная, которая порождается использованием в информационных технологиях сущностей, принадлежащих ментальной среде (концепты знания в сознании людей), или сущностей двойной природы, принадлежащих ее границам с другими средами предметной области информатики [37, 38]. Этот вид сложности связан с задачами, которые не удается решить с применением формальных алгоритмов, так как часто они не поддаются точной постановке, а сущностям, затрагиваемым в процессе их постановки и решения, трудно дать строго формальные определения. Примером такой задачи может служить поиск новых значений языковых единиц в текстах [39].

Д. Харел называет такие задачи «псевдоалгоритмическими» проблемами (pseudo-algorithmic problems), подчеркивая в своей книге, что он следует формальной трактовке термина «алгоритм». Главный вопрос при решении подобных проблем: «Каким образом представить сложное знание человека так, чтобы его представление поддавалось алгоритмической обработке, применению систем и средств информатики для их решения?» [32, с. 402].

Из трех публикаций, выбранных авторами Отчета [28], третий источник, комплементарный работам [31, 32], – это статья П. Деннинга и П. Розенблума [33], которая содержит краткое описание ранее вышеупомянутых идей Розенблума [1, 2]. В ней и был впервые предложен новый вариант позиционирования информатики в системе современного научного знания с делением ее на четыре отрасли.

Таким образом, эти три публикации неявно обосновывают необходимость существенного расширения предметной области информатики таким образом, чтобы она могла охватить неформальные задачи когнитивной сложности и широкий спектр феноменов, связанных с информационными процессами в технических, живых и социальных системах.

tion and explanation» [36]. Это определение соответствует толкованию К. Нюгара, но оно не содержит явного деления феноменов на когнитивные и сенсорно воспринимаемые.

2. КОНТЕКСТ СОЗДАНИЯ ПАРАДИГМЫ ИНФОРМАТИКИ И ЕЕ СТРУКТУРА

Рассмотренные работы П. Деннинга, П. Розенблума, К. Нюгара, Д. Харела, М. Касперсена и его соавторов, обсуждающих в том числе и вопросы позиционирования информатики в системе научного знания, были опубликованы в период с 1986 по 2019 г. Сегодня эти вопросы сохраняют свою актуальность и продолжается их активное обсуждение, в том числе с позиции поиска необходимой совокупности теоретических основ для разработки стратегии компьютерного образования. В статье [40] М. Тедре и Д. Паунен не только рассмотрели камни преткновения и проблемные места в поиске теоретических основ и в определении предметной области информатики в интересах компьютерного образования. Они также констатировали, что наибольшим препятствием выступает не что иное, как *отсутствие согласованной научной парадигмы информатики в трактовке Т. Куна* [41]. Необходимость создания именно куновской парадигмы в интересах компьютерного образования отмечалось и раньше [42]. Неослабевающий интерес к такой трактовке обусловлен тем, что с точки зрения Т. Куна научная парадигма любой науки, чтобы обеспечить эффективные научные исследования, должна позволять получить четкие ответы на следующие вопросы [41]¹³:

- Из каких основных сущностей состоит ее предметная область?
- Как они взаимодействуют друг с другом, а также с органами чувств и сознанием?
- Какие вопросы допустимо задать об основных ее сущностях и какие методы используются для поиска решений?

В настоящее время отсутствует конвенциональная парадигма информатики, которая отвечала бы на эти вопросы. В данном разделе нашей статьи рассматриваются контекст и предпосылки создания новой научной парадигмы информатики, фрагменты которой, предлагаемые нами, уже позволяют частично ответить на первые два вопроса.

2.1. Контекст создания научной парадигмы

Сложность построения и последующего сопоставления вариантов научной парадигмы информатики обусловлена отсутствием конвенциональных границ ее предметной области и консенсуса по смысловому содержанию ее базовых понятий [43–47]. Например, дискуссии о предмете информатики и о природе данных, информации и знания ведутся уже не один десяток лет. Согласно Ю.А. Шрейдеру, «споры о предмете [информатики] – это не столько споры об объективной истине, сколько стремление

¹³ «Effective research scarcely begins before a scientific community thinks it has acquired firm answers to the following: What are the fundamental entities of which the universe is composed? How do these interact with each other and with the senses? What questions may legitimately be asked about such entities and what techniques employed in seeking solutions?» [41, 42, p. 265].

отстоять свой взгляд на предмет, сделать его фактом общественного сознания сообщества исследователей. Это не значит, что представление о предмете произвольно – оно должно выражать плодотворный и эвристичный взгляд на реально осуществляющую деятельность, возможности ее развития и перспективы использования» [48]. Однако «плодотворные взгляды» могут различаться концептуально, и мы излагаем только один из возможных вариантов.

Для построения предлагаемого нами варианта научной парадигмы информатики выбрано десять публикаций как источников его зарождения [2, 3, 11, 28, 31, 33, 48–51], из которых взяты следующие ключевые положения:

- согласно Ю. А. Шрейдеру, определение базовых понятий информатики – это результат договоренности в сообществе исследователей, а не итоги изучения объективно существующих предметов [48], и поэтому в рамках предлагаемой парадигмы они и позиционируются как интерсубъективные сущности¹⁴;
- согласно Р. С. Гиляревскому, «если данные воспринимаются и интерпретируются человеком, то они становятся для него информацией» (курсив мой. – И. З.) [11, с. 10];
- в документе «Глубинное изменение – технологические переломные моменты и социальное воздействие» [49]¹⁵ предложен перечень новых информационных технологий, которые определяют кардинальный характер преобразования общества и экономики, получившего название «Четвертая промышленная революция» [50], и сказано, что прогнозируемый характер преобразования во многом будет обусловлен, в частности, теми ИТ, которые *охватывают сущности, принадлежащие средам разной природы*;
- в обзоре «Informatics Education in Europe» [28], опубликованном в 2017 г. и предваряющем разработку Европейской стратегии компьютерного образования «Informatics for All» [3], дана уже упоминавшаяся выше характеристика современной информатики: «Портрет информатики (также известной как “computing” или “computer science”) значительно изменился за последние 50 лет или около того. В то время как естественные науки определяются применительно к миру (world), в котором мы живем, информатику как

¹⁴ С точки зрения семиотического треугольника (*предмет – понятие – слово*) для объективных сущностей первичной вершиной треугольника служит *предмет*, в результате изучения которого появляются *понятие* и *слово*. Для интерсубъективных сущностей первичной вершиной служит *понятие*, варианты дефиниции которого необходимо обсудить в интересах достижения консенсуса. Если его удается достичь, то именно в процессе обсуждения появляются *предмет мысли, слово*, его обозначающее и выражающее *понятие*, которые вместе и образуют семиотический треугольник в результате генезиса интерсубъективной сущности.

¹⁵ В подготовке материалов для этого документа, который был разработан под эгидой Всемирного экономического форума (Давос, Швейцария), принимали участие около 800 экспертов и руководителей отрасли информационных и коммуникационных технологий.

научную дисциплину определить сложнее; у нее нет эмпирических основ, как у естественных наук; это нечто большее, чем мышление формальными символами, как в математике; и это далеко не просто компиляция инженерных принципов и технологий»¹⁶;

- в статье [51] И. Нонака разделяет явное (explicit) и неявное (tacit) знания и дает схему процессов их преобразования, которую планируется использовать при определении одного из компонентов парадигмы информатики, а именно – свойственных только ей базовых методов и моделей (этот компонент мы планируем описать в отдельной статье);

• в работах [2, 33] П. Деннинг и П. Розенблум предложили сгруппировать научные дисциплины в четырех отраслях знания и включить исследование информационных трансформаций в технических, живых и социальных системах в определяемую ими четвертую отрасль знания;

- в докладе К. Ньюгора [31] дано значительно расширенное определение предметной области информатики, которое уже цитировалось выше (см. сноску 9).

Это расширенное определение имплицитно вводит в предметную область информатики по К. Ньюгору разделение ее сущностей на категории разной природы: ментальной (например, результаты процессов генерации научного знания, происходящих в сознании исследователей) и сенсорно воспринимаемой, к которой относятся, в частности, знаковые формы представления результатов процессов генерации знания после его деления на концепты.

В развитие определения информатики по К. Ньюгору в работе [52] было дано описание основания для первой классификации объектов ее предметной области как феноменов разной природы и формирования следующих пяти ее сред, включающих объекты одной и той же природы:

1) ментальная среда – это совокупность когнитивных феноменов, используемых и/или формируемых в процессе познания, происходящего в сознании людей (феномены, формируемые в процессах познания с использованием знаковых систем как смысловые элементы знания, будем называть концептами);

2) сенсорно воспринимаемая среда, которую для краткости иногда будем называть информационной средой – это совокупность сенсорно воспринимаемых феноменов, находящихся вне сознания, но взаимодействующих с когнитивными феноменами;

3) цифровая среда – это совокупность компьютерных кодов;

4) нейросреда – это электрические потенциалы и магнитные поля, генерируемые мозгом, которые используются, например, в информационных технологиях

¹⁶ «The characterisation of Informatics (also known as computing or computer science) has developed significantly over the past 50 years or so. While natural sciences are defined with reference to the world in which we live, Informatics as a scientific discipline is harder to define; it does not have the empirical foundations of the natural sciences; it is more than symbolic reasoning as in mathematics; and it is far from just being a compilation of engineering principles and technology» [28, p. 8].

управления роботизированной рукой [37] и в других ИТ, применяющих интерфейсы «мозг – компьютер»;

5) ДНК-среда – это совокупность цепочек РНК и ДНК¹⁷.

Как было отмечено ранее, К. Нюгор имплицитно включил в предметную область информатики объекты ментальной среды. Мы предлагаем, с одной стороны, к объектам ментальной, информационной и цифровой среды добавить объекты нейро- и ДНК-среды, с другой – возможность включить в будущем объекты тех сред, которые не используются сейчас в системах и средствах информатики. Поэтому в предлагаемой нами научной парадигме информатики такая возможность включения устанавливается с самого начала.

В соответствии с перечисленными средами информатики первая классификация будет включать как минимум пять классов объектов ее предметной области, каждый из которых содержит объекты одной среды: ментальной, информационной, цифровой, нейро- или ДНК-среды. При этом с ростом разнообразия природы объектов информатики верхний уровень классификации может пополняться новыми классами, природа объектов которых отличается от природы сред, ранее включенных в первую классификацию [53].

Это может произойти, например, в том случае, когда при проектировании информационных технологий встретятся сущности, которые по своей природе не относятся ни к одной из ранее уже выделенных сред [54]. Таким образом, в нашем варианте парадигмы информатики классификацию объектов ее предметной области предлагается сделать открытой, что обусловлено возможным включением в будущем в нее объектов ранее не рассматривавшейся природы (открытость классификации).

2.2. Компоненты научной парадигмы информатики

Согласно А. Соломонику, научная парадигма «зрелой» науки состоит из следующих восьми компонентов, которые могут разрабатываться отдельно, но объединяются в единую и цельную конструкцию [17–20]: философские основы, предмет изучения, методы изучения, аксиоматика, классификации науки, система терминов, языки [знаковые системы] науки и методы верификации результатов. Сам термин «научная парадигма» трактуется им в соответствии с теорией Т. Куна, которая описывает процесс смены научных парадигм [41]. При этом А. Соломоник отмечает тот факт, что в книге Куна мы не находим ответа на вопрос: «Из чего должна состоять парадигма любой "зрелой" науки?» [17, с. 23].

Отметим, что в нашей статье основное внимание уделяется только одному компоненту парадигмы и

¹⁷ Например, модели трансляции естественных ДНК, созданные микробиологами, используются в информатике при разработке методов и средств записи для длительного хранения данных большого объема с применением синтезированных цепочек ДНК.

только двум классификациям¹⁸ (объектов предметной области информатики и трансформаций этих объектов). Благодаря первым результатам формирования этих двух классификаций, далее мы частично ответим на первые два из трех перечисленных выше вопросов Т. Куна. Частичность ответов обусловлена тем, что при описании фрагментов этих классификаций мы рассматриваем, в основном, только три среды предметной области информатики (ментальная, информационная и цифровая – рис. 1).

3. ВЕРХНИЕ УРОВНИ ДВУХ КЛАССИФИКАЦИЙ

Из трех основных подходов в информатике к описанию ее предметной области (объектный, трансформационный и синтетический) предлагаемое нами описание двух классификаций ближе всего к синтетическому. Примерами первого подхода, в рамках которого основное внимание уделяется объектам предметной области информатики (например, данные, информация и/или знание) и в меньшей степени отношениям между ними, могут служить работы [55–57]; примерами второго подхода, в рамках которого основное внимание уделяется информационным трансформациям и в меньшей степени трансформируемым объектам – работы [33, 58], примерами третьего, синтетического подхода, в котором уделяется внимание и объектам предметной области информатики, и отношениям между ними, могут служить работы [11, 48, 59–61].

Для построения двух классификаций наибольший интерес представляют трансформационный и синтетический подходы. В 2009 г. П. Деннинг и П. Розенблум сформулировали суть информатики как компьютеринга следующим образом: «... информатика – это не просто алгоритмы и структуры данных; это преобразования представлений» [33]¹⁹. Чуть позже, в контексте краткого описания парадигмы информатики как компьютеринга, П. Деннинг и П. Фриман изменили эту формулировку на такую: «Центральный объект внимания в информатике можно определить как информационные процессы – естественные или искусственные процессы, преобразующие информацию (курсив мой – И. З.)» [58]²⁰.

На синтетическом подходе остановимся чуть подробнее, так как именно его обобщение использовалось в процессе построения обеих классификаций. Ранее уже цитировалось положение Р.С. Гиляревского об отношении между данными и информацией (*после восприятия и интерпретации данные становятся информацией*) [11]. В работах [48, 59–61] основное внимание также уделяется не только таким объектам предметной области как информация и знание, но и отношениям между ними. Согласно Ю.А. Шрейдеру «Информационно-знаниевые процессы включают преобразование человеческих (в зна-

¹⁸ Табличная классификация интерфейсов информатики была описана в [54].

¹⁹ «... computing is not just algorithms and data structures; it is transformations of representations» [33, p. 28].

²⁰ «The central focus of the computing paradigm can be summarized as information processes – natural or constructed processes that transform information» [58, p. 29–30].

чительной мере, личностных) знаний, существующих «здесь» и «теперь», в социальную информацию, доступную «везде» и «всегда» и гарантирующую возможность извлечения из нее знаний» [48]. Отметим, что уровни социализации знания планируется определить на одном из следующих уровней предлагаемой классификации объектов (скорее всего, на четвертом), который в настоящей статье не рассматривается.

Главный вывод П. Ингверсена состоит в том, что понятие информации должно удовлетворять двум требованиям. С одной стороны, информация является результатом *преобразования в знаковую форму* когнитивных структур человека-генератора. С другой стороны, это нечто такое, при восприятии и осознании чего имеющиеся знания получателя информации подвергаются влиянию и трансформируются [59]. В описании предметной области информационной науки Дж. Фаррадей сопоставляет как центральные понятия этой предметной области знания и информацию, а также трансформацию информации в знания [60]. По мнению В. Хьорланда, в определении информационной науки в явном виде должна быть выражена соотнесенность информации со знанием [61].

Перечисленные положения из работ [11, 48, 59–61] и стали предпосылками формирования двух классификаций сущностей предметной области информатики: ее объектов исследования (см. рис. 1) и их трансформаций (рис. 2). В использовании этих положений можно увидеть наш подход к формированию классификаций: по максимуму использовать уже известные в информатике положения, которые не противоречат друг другу, как источники создания предлагаемых классификаций.

На сегодняшний день процесс построения предлагаемых классификаций еще далек от завершения. Однако первые итоги их формирования уже применяются:

- в определении отношений между компонентами модели (иерархии) DIKW (data-information-knowledge-wisdom – данные-информация-знание-мудрость) [62, 63];
- для развития модели DIKW в интересах проектирования информационных технологий извлечения нового знания из данных и/или текстовой информации [15];
- для формализованного описания статистической обработки текстовой информации в базах данных [6].

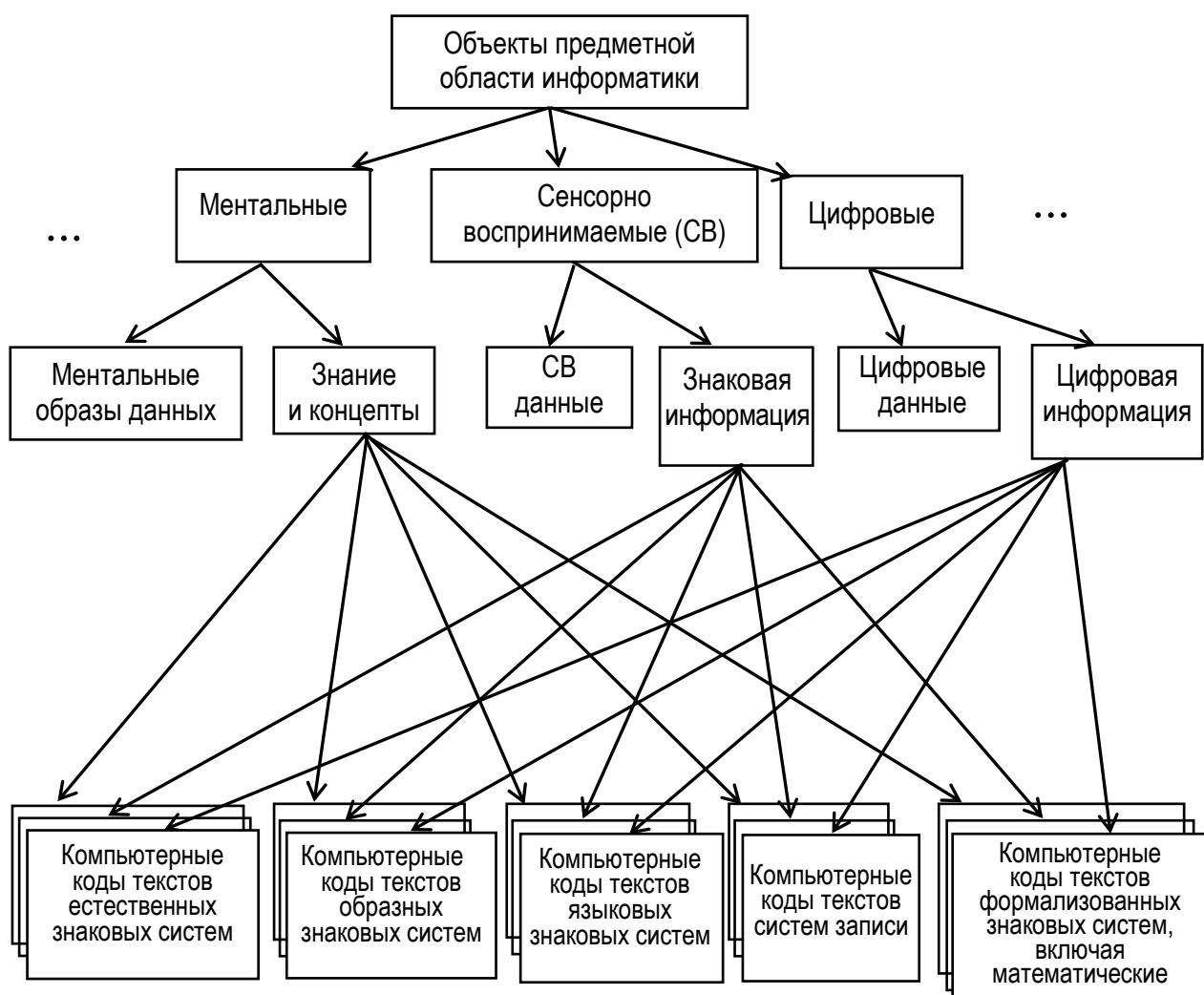


Рис. 1. Три верхних уровня классификации объектов предметной области информатики

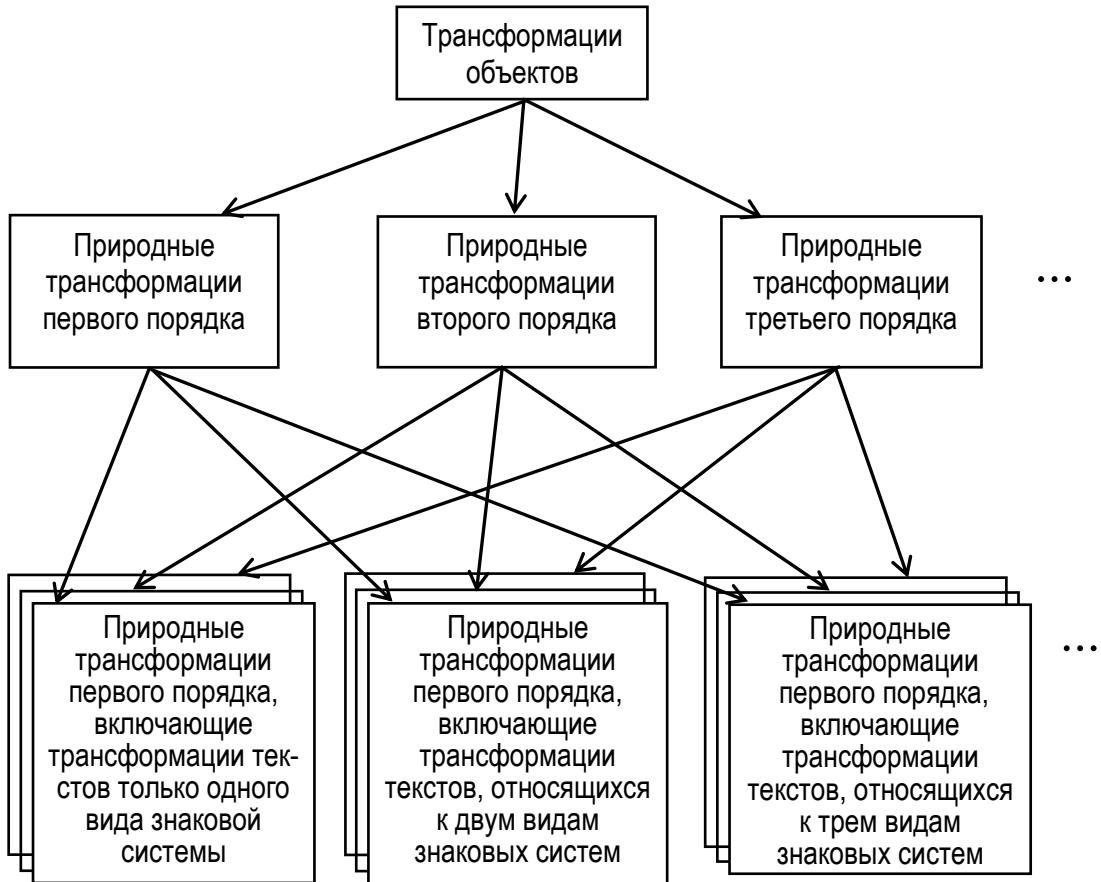


Рис. 2. Два уровня классификации трансформаций объектов предметной области в предлагаемой парадигме информатики как третьей культуры

Верхний уровень первой классификации (см. рис. 1) включает результаты спецификации объектов предметной области согласно их распределению по средам разной природы (объекты ментальной, сенсорно воспринимаемой, цифровой и других сред, обозначенных многоточием) [52], о которых мы писали выше. Здесь приведены и далее рассматриваются объекты только трех сред.

Верхний уровень второй классификации (см. рис. 2) согласно степени разнообразия природы объектов, вовлеченных в трансформации, включает классы трансформаций:

- первый класс – трансформации объектов только одной природы (далее – природные трансформации первого порядка), например, конкатенация трехзначных чисел N с $2N$ и $3N$, результат которой содержит 9 цифр кроме 0, может служить примером трансформации первого порядка [64];
- второй класс – трансформации объектов, относящихся к двум средам разной природы (далее – природные трансформации второго порядка), например, кодирование в компьютере текстов с помощью таблиц Unicode или представление ментальных понятий словами естественного языка;
- третий и последующие классы – трансформации объектов, относящихся к трем и более средам разной природы (далее – природные трансформации третьего и более высоких порядков), например, присво-

ение одного компьютерного кода понятию и слову как сочетанию букв, которое его выражает, что относится к третьему классу трансформаций объектов [38].

В каждом классе верхнего уровня первой классификации выделим два компонента, разместив их на следующем, втором, уровне. Согласно ранее предложенному определению сенсорно воспринимаемой (информационной) среды из работы [52] второй уровень содержит, как минимум, класс знаковой информации и класс сенсорно воспринимаемых данных. Необходимость выделения этих двух классов объектов проиллюстрируем на примере кардиограммы, сформированной в процессе цифровой регистрации электрических полей, образующихся при работе сердца. Кардиограмма на бумаге или на экране до ее интерпретации кардиологом – это данные сенсорно воспринимаемой среды. Информация из заключения кардиолога, созданная им как результат содержательного анализа данных кардиограммы, является сенсорно воспринимаемой знаковой информацией и тоже принадлежит к этой же среде. Таким образом, эта среда содержит, как минимум, два класса объектов: знаковую информацию и сенсорно воспринимаемые данные. При их компьютерном кодировании получаем соответственно цифровую информацию и цифровые данные, т. е. получаем два теоретически принципиально разных класса объектов цифровой природы.

Пример результата знаковой трансформации второго порядка

Оригинальное предложение	Перевод на английский язык
Мне эта подковочка дорога как память.	This little horseshoe is dear to me as a memento.
И позвольте вам за то, что вы ее сохранили, вручить двести рублей.	And, for having preserved it, allow me to give you two hundred roubles.
И он тотчас вынул из жилетного кармана деньги и вручил их Аннушке.	And he took the money from his waistcoat pocket at once and handed it to Annushka.

Примечание.

Для иллюстрации классов трансформаций использовались следующие предложения романа М.А. Булгакова «Мастер и Маргарита» (1929–1940) и их перевод Р. Певиа и Л. Волхонской (1997) на английский язык (перевод как пример знаковой трансформации первого порядка):

Мне эта подковочка дорога как память.

This little horseshoe is dear to me as a memento.

И позвольте вам за то, что вы ее сохранили, вручить двести рублей.

And, for having preserved it, allow me to give you two hundred roubles.

И он тотчас вынул из жилетного кармана деньги и вручил их Аннушке.

And he took the money from his waistcoat pocket at once and handed it to Annushka.

Процесс содержательного анализа и интерпретации данных кардиограммы, а также подготовки заключения состоит из нескольких этапов трансформаций объектов трех сред: ментальной, сенсорно воспринимаемой и цифровой. Сначала кардиолог сенсорно воспринимает данные кардиограммы, полученной в процессе цифрового мониторинга работы сердца (цифровая среда), декодирования и визуализации его результатов на бумаге или экране (сенсорно воспринимаемая среда). На следующем этапе появляются ментальные образы данных в сознании кардиолога как результат их сенсорного восприятия (первый класс ментальных объектов). Затем в процессе интерпретации генерируется знание кардиолога (второй класс ментальных объектов) о наличии или отсутствии нарушений в работе сердца пациента как результат креативного процесса, включающего понимание кардиологом ментальных образов данных.

Далее следует этап подготовки заключения на некотором естественном языке, включающий генерацию концептов как результат деления полученного знания и его выражения словами этого языка в сенсорно воспринимаемой среде. Если готовится двухязычное заключение, например, параллельно на русском и английском языках, то деление знания на концепты выполняется по-разному в системе каждого из двух языков. Другими словами, деление одного и того же знания кардиолога (о наличии или отсутствии нарушений в работе сердца пациента) на концепты зависит от используемого естественного языка как вербальной знаковой системы.

Таким образом, второй по порядку уровень первой классификации включает результаты следующей детализации объектов ментальной, сенсорно воспринимаемой и цифровой сред:

- знание и ментальные образы сенсорно воспринимаемых данных [15]²¹;

- знаковую информацию и сенсорно воспринимаемые данные (СВ-данные);

- цифровую информацию и цифровые данные.

Второй уровень второй классификации охватывает классы трансформаций объектов, которые сформированы согласно видам знаковых систем, определенных на основе их типологии по А. Соломонику [18: с. 131]²²: естественные знаковые системы, образные, естественно-языковые, вербально-несловесные системы записи²³, формализованные знаковые системы, включая математические. Этот уровень содержит следующие классы применительно *только к текстам, вовлеченным в процесс трансформации*:

- первый класс включает трансформации текстов только одного вида знаковой системы (далее – знаковые трансформации первого порядка), например, перевод текста с одного языка на другой²⁴ (см. примечание к табл. 1);

- второй класс включает трансформации текстов, относящихся к двум видам знаковых систем (далее – знаковые трансформации второго порядка), например, формирование в редакторе Word на основе параллельных текстов на русском и английском языках (естественно-языковые знаковые системы) таблицы, каждая строка которой содержит оригинальное предложение в первом столбце и его перевод во втором (вербально-несловесная система записи) (см. табл. 1);

²² В двух последующих монографиях А. Соломоник делит формализованные системы на знаковые формализованные системы первого и второго порядка [19, с. 76, 20, с. 64].

²³ Под системой записи понимается знаковая система, сочетающая вербальные знаки с несловесными (языки нотной записи, карт, таблиц и др.).

²⁴ Отметим, что перевод текста с одного языка на другой как знаковая трансформация первого порядка относится к классу природных трансформаций второго порядка верхнего уровня классификации трансформаций объектов, относящихся к двум средам разной природы: сенсорно воспринимаемой и цифровой сред для машинного перевода и ментальной и сенсорно воспринимаемой сред для человека-переводчика.

²¹ Деление на личностное, коллективное, организационное и конвенциональное знание [14] (конвенциональное у Ю. Шрейдера названо социальным [48]) в настоящей статье не рассматривается.

**Пример результата знаковой трансформации третьего порядка
(последний столбец содержит элементы арифметического вектора отношений длин)**

Оригинальное предложение	Длина оригинала	Перевод на английский язык	Длина перевода	Отношения длин
Мне эта подковочка дорога как память.	37	This little horseshoe is dear to me as a memento.	49	0,76
И позвольте вам за то, что вы ее сохранили, вручить двести рублей.	66	And, for having preserved it, allow me to give you two hundred roubles.	71	0,93
И он тотчас вынул из жилетного кармана деньги и вручил их Аннушке.	66	And he took the money from his waistcoat pocket at once and handed it to Annushka.	82	0,80

- третий класс включает трансформации объектов, относящихся к трем и более видам знаковых систем (далее – знаковые трансформации третьего и более высоких порядков), так, если к формированию таблицы в примере второго класса добавить запись результата вычисления отношения длины оригинального предложения к длине его перевода для каждой строки, то это будет представлять собой знаковую трансформацию²⁵ третьего порядка (естественно-языковые знаковые системы → вербально-несловесная система записи → математическая запись арифметического вектора отношений длин).

Следующий, третий уровень первой классификации в ее текущей версии ограничен только *знаковой, цифровой информацией и тем знанием ментальной среды, которое может быть представлено в знаковой форме*. Соответственно, спектр тех знаний *ментальной среды*, которые представлены как знаковая информация, распределяется на пять классов согласно тем видам знаковых систем, которые использовались для их представления в сенсорно воспринимаемой среде. Это позволяет в *цифровой среде* кодировать принадлежность знака к той или иной знаковой системе, отражая ее в компьютерном коде, а также значение (в статье синоним концепта) и форму этого знака.

ЗАКЛЮЧЕНИЕ

Перечисленные в настоящей статье классы верхних уровней двух классификаций, относящиеся только к одному компоненту предлагаемой научной парадигмы информатики, и визуализация этих классов, приведенная на рис. 1 и 2, представляют собой только отдельные фрагменты этого компонента. Наполнение каждого из уровней классификаций, естественно, также не претендует на полноту. Но и такое фрагментарное описание этого компонента парадигмы уже позволяет частично ответить на первые два из трех вопросов Т. Куна.

На первый вопрос (Из каких основных сущностей состоит предметная область информатики?) предлагается следующий ответ:

²⁵ Вычисление характеристик текстов будем также называть знаковыми трансформациями.

«Все основные сущности сначала делятся на два глобальных класса: *объекты* предметной области информатики и их *трансформации*. Природа объектов может быть ментальной, сенсорно воспринимаемой и цифровой, но она не ограничена только этими тремя средами. С появлением интерфейса «мозг-компьютер» были добавлены объекты нейросреды. Разработка методов записи и хранения данных с использованием синтезированных цепочек ДНК обусловила добавление объектов ДНК-среды, т. е. не исключено дальнейшее увеличение числа сред разной природы в предметной области информатики и соответствующее увеличение перечня ее основных сущностей-объектов».

Подобное увеличение *природы объектов* предметной области информатики не наблюдается ни в естественных, ни в социально-гуманитарных науках, что рассматривается как один из доводов к позиционированию информатики как третьей культуры.

Ответ на второй вопрос (Как сущности взаимодействуют друг с другом, а также с органами чувств и сознанием?) состоит из трех положений:

«1. С органами чувств человека взаимодействуют объекты сенсорно воспринимаемой среды предметной области информатики.

2. Виды взаимодействий объектов описаны в классификации их трансформаций, строящейся на верхних двух уровнях по следующим основаниям: порядок природной трансформации (от первого и выше) и порядок знаковой трансформации (от первого и выше).

3. Трансформации, приведенные в классификации, как сущности предметной области информатики, могут объединяться в рамках полисредовых технологий, охватывающих объекты разной природы» [49].

Более подробно взаимодействие объектов информатики друг с другом и технологическое сочетание их трансформаций описано в работах [15, 62, 63].

Таким образом, предлагаемое нами позиционирование информатики в системе научного знания как третьей культуры обусловлено:

- разнообразием природы объектов предметной области информатики, которое не вписывается в дихотомию естественно-научной и социально-гуманитарной культур;

- широким спектром трансформаций конкретных объектов информатики в абстрактные, которые не изучаются ни точными, ни естественными, ни социально-гуманитарными науками, в том числе, не изучается обратимость трансформаций конкретной знаковой информации в абстрактную²⁶ [6–8].

Такое позиционирование информатики можно было бы использовать и при пересмотре номенклатуры научных специальностей, добавив в качестве ее самого верхнего уровня деление всей системы научного знания на три культуры, и выделив в третьей культуре новое направление исследований «Общая информатика» для описания всех восьми компонентов научной парадигмы. Общая информатика могла бы стать концептуально единым теоретическим основанием всего спектра информационно-компьютерных образовательных дисциплин.

Деление на три культуры позволило бы информатике, объединяя информационно-компьютерные науки в рамках третьей культуры, способствовать фундаментальному переосмыслению и гармонизации взаимосвязей точных, естественных, социально-гуманитарных и информационно-компьютерных наук.

СПИСОК ЛИТЕРАТУРЫ

- Rosenbloom P. From Designing Minds to Mapping Disciplines. 12.04.2022. – URL: <https://www.youtube.com/watch?v=myC7oteFsWM>
- Rosenbloom P.S. On Computing: The Fourth Great Scientific Domain. – Cambridge: MIT Press, 2013. – 307 p.
- Caspersen M. E., Gal-Ezer J., Hall W., McGetrick A., Nardelli E., Schnabel B. Informatics for All: The Strategy. – New York, NY, USA: ACM, 2018. – 16 p. – URL: <https://www.acm.org/binaries/content/assets/public-policy/acm-europe-ie-i4all-strategy-2018.pdf>
- Сноу Ч.П. Две культуры. – Москва: Прогресс, 1973. – 144 с.
- Berlin I. Against the current: Essays in the history of ideas. – London: The Hogarth Press, 1979. – 394 p.
- Вакуленко В.В., Зацман И.М. Формализованное описание статистической обработки информации в базах данных // Информатика и её применения. – 2023. – Т. 17, № 3. – С. 93–99.
- Зацман И.М., Мамонова О.С., Щурова А.Ю. Обратимость и альтернативность генерализации моделей перевода коннекторов в параллельных текстах // Системы и средства информатики. – 2017. – Т. 27, № 2. – С. 125–142.
- Зацман И.М. Методология обратимой генерализации в контексте классификации информационных трансформаций // Системы и средства информатики. – 2018. – Т. 28, № 2. – С. 128–144.
- Михайлов А.И., Черный А.И., Гиляревский Р.С. Основы информатики. – Москва: Наука, 1968. – 756 с.
- Гиляревский Р.С. Основы информатики. – Москва: Изд-во «Экзамен», 2003. – 320 с.
- Информатика как наука об информации: Информационный, документальный, технологический, экономический, социальный и организационный аспекты / под ред. Р.С. Гиляревского. – Москва: ФАИР-ПРЕСС, 2006. – 592 с.
- Королев Л.Н., Миков А.И. Информатика. Введение в компьютерные науки. – Москва: Изд-во «Высшая школа», 2003. – 341 с.
- Computer Science Curricula 2023 (March 2023). – URL: <https://csed.acm.org/wp-content/uploads/2023/03/Version-Beta-v2.pdf>
- Nissen M. E. Harnessing knowledge dynamics: Principled organizational knowing & learning. – London: IRM Press, 2006. – 278 p.
- Zatsman I. Digital spiral model of knowledge creation and encoding its dynamics // 18th Forum (International) on Knowledge Asset Dynamics Proceedings. – Matera, Italy: Arts for Business Institute, 2023. – P. 581–596. – URL: https://www.researchgate.net/publication/371640416_Digital_Spiral_Model_of_Knowledge_Creation_and_Encoding_its_Dynamics
- Наука о данных. Материал из Википедии. – URL: https://ru.wikipedia.org/wiki/Наука_о_данных
- Соломоник А.Б. Парадигма семиотики. – Минск: MET, 2006. – 335 с.
- Соломоник А.Б. Философия знаковых систем и языков. – Москва: Изд-во ЛКИ, 2011. – 408 с.
- Соломоник А.Б. Опыт современной теории познания. – Санкт-Петербург: Алетейя, 2019. – 232 с.
- Solomonick A.B. The Modern Theory of Cognition. – Newcastle: Cambridge Scholars Publishing, 2021. – 242 с.
- Fatourou P., Hankin C. Welcome to the Europe Region Special Section // Communications of the ACM. – 2019. – Vol. 62, № 4. – P. 30.
- Gulliksen J. Incorporating Europe's Values in Future Research // Там же. – P. 40–41.
- Korrigane L. A Demographic Snapshot of the IT Workforce in Europe // Там же. – P. 32–33.
- Caspersen M.E., Gal-Ezer J., McGetrick A., Nardelli E. Informatics as a Fundamental Discipline for the 21st Century // Там же. – P. 58–63.
- Communication on the First Digital Education Action Plan. – The European Commission, 2018. – 18 p. – URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0022&from=EN>
- Communication on the First Digital Education Action Plan: Commission staff working document // The European Commission. – 2018. – 31 p. – URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018SC0012&from=EN>
- Digital Education Action Plan (2021–2027) // The European Commission, 2020. – URL: <https://education.ec.europa.eu/ru/focus-topics/digital-education/action-plan>

²⁶ Основание обратимость/необратимость многоступенчатых трансформаций конкретных объектов в абстрактные, рассмотренное в работах [6–8], планируется в будущем добавить в классификацию трансформаций объектов предметной области информатики.

28. The Committee on European Computing Education. Informatics Education in Europe: Are we all in the same boat? Technical Report. – New York, NY, USA: ACM, 2017. – 251 p. – URL: <https://portals.acm.org/hippo/ceccereport.pdf>
29. FACT SHEET: President Obama Announces Computer Science for All Initiative. – The White House, Office of the Press Secretary, 2016. – 12 p. – URL: <https://obamawhitehouse.archives.gov/the-press-office/2016/01/30/fact-sheet-president-obama-announces-computer-science-all-initiative-0>
30. Denning P. Computing is a natural science // Communications of the ACM. – 2007. – Vol. 50, № 7. – P. 13–18.
31. Nygaard K. Program Development as a Social Activity // Information Processing 86: Proceedings from the IFIP 10th World Computer Congress, Dublin, Ireland, September 1–5, 1986 / ed. H.-J. Kugler. – North Holland: Elsevier Science Publishers B.V., IFIP, 1986. – P. 189–198.
32. Harel D. Algorithmics – The Spirit of Computing. – Reading, MA, USA: Addison-Wesley, 1987. – 514 p.
33. Denning P., Rosenbloom P. Computing: The fourth great domain of science // Communications of the ACM. – 2009. – Vol. 52, № 9. – P. 27–29.
34. Newell A., Perlis A., Simon H. Computer science // Science. – 1967. – Vol. 157, № 3795. – P. 1373–1374.
35. Webster's New World Dictionary of the American Language. – New York, NY, USA: The World Publishing Company, 1960. – 1724 p.
36. Definition of phenomenon // Merriam-Webster's dictionary. – URL: <https://www.merriam-webster.com/dictionary/phenomenon>
37. Зацман И. М. Интерфейсы третьего порядка в информатике // Информатика и её применения. – 2019. – Т. 13, № 3. – С. 82–89.
38. Зацман И.М. Кодирование концептов в цифровой среде // Информатика и её применения. – 2019. – Т. 13, № 4. – С. 97–106.
39. Zatsman I. A model of goal-oriented knowledge discovery based on human-computer symbiosis // 16th Forum (International) on Knowledge Asset Dynamics Proceedings. – Matera, Italy: Arts for Business Institute, 2021. – P. 297–312.
40. Tedre M., Pajunen J. Grand theories or design guidelines? Perspectives on the role of theory in computing education research // ACM Transactions on Computing Education. – 2022. – Vol. 23, № 1. – P. 1–20. DOI: 10.1145/3487049.
41. Кун Т. Структура научных революций / пер. с англ. – Москва: Прогресс, 1977. – 302 с.; Kuhn T. The Structure of Scientific Revolutions. – Chicago: University of Chicago Press, 1962. – 264 p.
42. Fincher S., Tenenberg J. Using Theory to Inform Capacity-Building: Bootstrapping Communities of Practice in Computer Science Education Research // Journal of Engineering Education. – 2006. – Vol. 95, № 4. – P. 265–277. DOI: <https://doi.org/10.1002/j.2168-9830.2006.tb00902.x>
43. Rosenbloom P. The Computing Sciences and STEM Education // Proceedings of Ubiquity Symposium «The Science in Computer Science». – 2014. – Vol. 14. – URL: <http://ubiquity.acm.org/article.cfm?id=2590530>
44. Frailey D.J. Computation is process // E-print. Proceedings of Ubiquity Symposium «What is computation?». – 2010. – Vol. 10. – URL: <https://ubiquity.acm.org/article.cfm?id=1891341>
45. Conery J.S. Computation is symbol manipulation // E-print. Proceedings of Ubiquity Symposium «What is computation?». – 2010. – 7 p. – URL: <https://ubiquity.acm.org/article.cfm?id=1889839>
46. Denning P.J. Opening statement // E-print. Там же. – 11 p. – URL: <https://ubiquity.acm.org/article.cfm?id=1880067>
47. Bajcsy R. Computation and information // E-print. Там же. – 3 p. – URL: <https://ubiquity.acm.org/article.cfm?id=1899473>
48. Шрейдер Ю.А. Информация и знание // Системная концепция информационных процессов. – Москва: ВНИИСИ, 1988. – С. 47–52.
49. Deep Shift – Technology Tipping Points and Societal Impact. – Geneva: WE Forum, 2015. – URL: http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf
50. Шваб К. Четвертая промышленная революция. – Москва: Эксмо, 2018. – 288 с.; Schwab K. The Fourth Industrial Revolution. – Geneva: World Economic Forum, 2016. – 172 p.
51. Nonaka I. The knowledge-creating company // Harvard Business Review. – 1991. – Vol. 69, № 6. – P. 96–104.
52. Зацман И.М. Теоретические основания компьютерного образования: среди предметной области информатики как основание классификации ее объектов // Системы и средства информатики. – 2022. – Т. 32, № 4. – С. 77–89.
53. Зацман И.М. О научной парадигме информатики: верхний уровень классификации объектов ее предметной области // Информатика и её применения. – 2022. – Т. 16, № 4. – С. 108–114.
54. Зацман И.М. Таблица интерфейсов информатики как информационно-компьютерной науки // Научно-техническая информация. Сер. 1. – 2014. – № 11. – С. 1–15.
55. Ackoff R. From data to wisdom // Journal of Applied Systems Analysis. – 1989. – Vol. 16, № 1. – P. 3–9.
56. Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy // Journal of Information Science. – 2007. – Vol. 33, № 2. – P. 163–180.
57. Frické M. H. Data-Information-Knowledge-Wisdom (DIKW) Pyramid, Framework, Continuum // Encyclopedia of Big Data / eds. L. Schintler, C. McNeely. – Cham: Springer, 2018. DOI: 10.1007/978-3-319-32001-4_331-1.
58. Denning P., Freeman P. Computing's paradigm // Communications of the ACM. – 2009. – Vol. 52, № 12. – P. 28–30.
59. Ingwersen P. Information and information science // Encyclopaedia of Library and Information Sci-

- ence. – Vol. 56, sup. 19. – New York: Marcel Dekker Inc., 1992. – P. 137-174.
60. Farradane J. Knowledge, information, and information science // Journal of Information Science. – 1980. – Vol. 2, № 2. – P.75-80.
61. Hjorland B. Library and information science: practice, theory, and philosophical basis // Information Processing and Management. – 2000. – Vol. 36, № 3. – P. 501-531.
62. Зацман И.М. Данные, информация и знание в научной парадигме информатики // Информатика и её применения. – 2023. – Т. 17, № 1. – С. 116–125.
63. Зацман И.М. Трансформация иерархии Акоффа в научной парадигме информатики // Информатика и её применения. – 2023. – Т. 17, № 3. – С. 107-113.
64. Chen L., Zaharia M., Zou J. How is ChatGPT's behavior changing over time? E-print 2307.09009. 2023. – URL: <https://arxiv.org/abs/2307.09009>

Материал поступил в редакцию 25.08.23.

Сведения об авторе

ЗАЦМАН Игорь Моисеевич – доктор технических наук, зав. отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), Москва
e-mail: izatsman@yandex.ru

ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ

УДК [02+002.6]:004.8.032.26

О.Н. Шорин

Возможные направления применения искусственного интеллекта в библиотеках и информационных центрах

Показаны возможности использования нейронных сетей в библиотечной деятельности для автоматического аннотирования и поиска похожих документов, каталогизации изображений, анализа библиографических ссылок в научных статьях. В свою очередь, информация, хранящаяся в библиотеках, может быть полезна для обучения нейронных сетей.

Ключевые слова: библиотека, искусственный интеллект, нейронная сеть, обучение нейронной сети, аннотирование, поиск, генеративные трансформеры

DOI: 10.36535/0548-0019-2023-11-2

ВВЕДЕНИЕ

К середине XX в. возможности электронных вычислительных машин достигли такого уровня, что скорость выполняемых на них операций превзошла человеческие возможности. Поскольку лидерство в выполнении рутинных вычислительных операций окончательно перешло к компьютерам, то вопрос об осуществлении ими творческой деятельности оставался открытым. И если до этого момента только философы задавались вопросом о том, способны ли будут машины мыслить как человек, то в середине прошлого столетия к обсуждению этого вопроса активно подключились математики, программисты и инженеры. Появилось понятие «искусственный интеллект», которое имеет множество толкований, но в контексте настоящей статьи уместно использовать следующее: искусственный интеллект – свойство искусственных систем выполнять творческие функции, которые традиционно считаются прерогативой человека [1].

В 1950 г. Аллан Тьюринг предложил тест, с помощью которого можно определить, превосходит ли машина по разумности человека или нет [2]. Суть теста заключалась в следующем: человек взаимодействует с одним человеком и одним компьютером, при этом все участники не видят друг друга; человек задает вопросы и получает ответы от человека и от компьютера. По этой беседе человек должен определить с кем он разговаривает: с человеком или с компьютерной программой. Задача компьютера состоит в том, чтобы ввести человека в заблуждение, заставив

его сделать неверный выбор. Этот тест Тьюринга представляет собой значительное сужение творческой составляющей вычислительных систем, поэтому достаточно быстро появились инструменты, которые помогли компьютерам успешно его пройти. В частности, поскольку тест основан на серии вопросов и ответов, то появление языка программирования LISP (LISP Processing language), в основу которого положена удобная работа со списками, в том числе и строковыми, в совокупности с использованием нехитрых приемов построения беседы позволило компьютерной программе формально пройти тест Тьюринга, но очевидно, что это не имеет ничего общего с творческой составляющей мыслительного процесса.

Другой классической вехой в развитии искусственного интеллекта принято считать турнир по шахматам, который состоялся в мае 1997 г. между Гарри Каспаровым и компьютером Deep Blue фирмы IBM. В том турнире компьютер одержал победу над человеком, но, как и в случае с успешным прохождением теста Тьюринга, здесь не совсем уместно говорить об искусственном интеллекте. Дело в том, что компания IBM создала очень хороший алгоритм для решения одной конкретной задачи, и этот алгоритм был основан на переборе и сравнении возможных позиций в ходе партии с последующим отсечением тех направлений развития игры, которые ведут к ухудшению положения компьютера [3]. Затем в программу были встроены дополнительные оптимизации по использованию сценариев из других партий, которые применялись в начале и в заключении игры. Всё это

позволило компьютеру одержать верх над человеком, но, тем не менее, в основе алгоритма лежал банальный перебор комбинаций, что свидетельствует о потрясающей скорости работы вычислительной машины, но нисколько – о творческой составляющей компьютерной программы.

НЕЙРОННЫЕ СЕТИ

Спустя некоторое время после появления термина «искусственный интеллект» были разработаны различные подходы к созданию программно-аппаратных комплексов, которые позволили бы компьютерам достичь определенных успехов в творческой деятельности. Одним из таких подходов стала технология машинного обучения с использованием нейронных сетей, которая получила мощное развитие в последние несколько лет.

Нейронная сеть является попыткой смоделировать процессы, происходящие в мозгу человека. Она состоит из множества простых процессоров, каждый из которых способен получать сигнал, как-то на него воздействовать и отправлять этот сигнал дальше. Все процессоры объединены в сложную структуру, на вход которой можно подать данные: эти данные, пройдя через сеть процессоров, каким-то образом будут преобразованы и выведены в качестве результата [4].

Работа с нейронной сетью состоит из двух этапов: обучения и использования. В процессе обучения на вход нейронной сети подаются специально подобранные данные, а полученный результат оценивается, исходя из того, насколько он приближен к желаемому. В зависимости от этой оценки нейронная сеть изменяет коэффициенты воздействия отдельных процессоров своей сети, стараясь подобрать их таким образом, чтобы результат был максимально приближен к желаемому. С математической точки зрения, обучение – это многопараметрическая задача нелинейной оптимизации.

В процессе эксплуатации корректно обученная нейронная сеть сможет при получении сходных, но не эквивалентных, входных данных выдать результат, приближенный к тому, которому её обучили. Нейронные сети успешно используются в таких задачах, как распознавание образов. Например, нейронная сеть, обученная находить кошку на картинке, с большой долей вероятности сможет найти кошку, если ей на вход подать изображение, которое не участвовало в обучении сети.

И это может быть одним из таких направлений применения нейронных сетей в библиотеках, как автоматическое проставление тэгов в зависимости от того, что изображено на иллюстрациях произведений, хранящихся в электронной библиотеке, для облегчения последующего поиска по изображениям.

Самое сложное в работе нейронных сетей – это их обучение. В процессе обучения может возникнуть множество проблем. Часть из них связана с настройками и параметрами обучения. Например, сеть может быть обучена не оптимально. Это означает, что решением задачи нелинейной оптимизации является некий локальный экстремум, а не глобальный. Но при

обучении нейронной сети большинство проблем связано с данными, на которых её обучают.

Одна из проблем – это недостаточность данных для обучения сети. В этом случае сеть фактически «запоминает» правильные ответы, а не учится искать некую взаимосвязь в подаваемых на вход данных. Проблемой является и плохое качество данных, на которых происходит обучение. Эта проблема известна под названием «мусор на входе, мусор в результате» (Garbage In, Garbage Out – GIGO). Дело в том, что нейронная сеть, обученная на некачественных данных, не способна выдать что-то стоящее при получении на вход качественных данных, поскольку она с ними никогда не сталкивалась и не обучена тому, как правильно на них реагировать.

Таким образом, для правильного обучения нейронной сети требуется большое количество качественных данных. Это как раз то, в чем сильны библиотеки – они хранят огромное количество информации, которая прошла редакторскую правку, а научные публикации – ещё и рецензирование. Следовательно, еще одно возможное пересечение деятельности библиотек и развития нейронных сетей – это предоставление большого объема качественной информации, хранящейся в библиотеках, для обучения сетей [5].

Развитие нейронных сетей

Идеи об устройстве и функционировании нейронных сетей возникли в середине XX в., однако большое распространение они получили в начале XXI в. Скорее всего, это связано с развитием производства элементной базы для микрочипов. Технологии достигли такого уровня, что уже практически в каждом современном телефоне есть встроенный нейропроцессор, содержащий миллиарды узлов в своей нейронной сети. На развитие нейронных сетей сильное влияние оказalo появление языков программирования, направленных на оптимизацию человеческих усилий при их обучении.

Оба этих фактора привели к тому, что даже инженеры и программисты, которые создают нейронные сети, подбирают данные для обучения, пишут алгоритмы, проводят обучение и тестируют их, не могут предсказать, что выдаст сеть, если ей на вход подать те или иные данные. Дело в том, что в нейронных сетях между процессом создания и процессом использования появляется процесс обучения, результат которого просчитать человеку невозможно, поскольку он чисто физически не способен проанализировать работу миллиардов нейронов.

Возможно, из-за такой непредсказуемости конечного результата, а также из-за доступности и удобства использования нейронные сети и получили столь широкое распространение. Человеку даже из любопытства может быть интересно ввести какой-то запрос на телефоне или в специальной программе на компьютере и получить результат, который он не мог спрогнозировать, но который, как минимум, не противоречит его запросу. Этот «вау-эффект» подогревает интерес общества к внедрению технологии нейронных сетей в различные аспекты жизнедеятельности.

В 2017 г. в Российской национальной библиотеке (РНБ) был проведен эксперимент с использованием нейронной сети *Watson*, разработанной компанией IBM. На тот момент в РНБ возникла идея, что поиск научных публикаций должен отличаться от поиска художественных произведений. Зачастую поиск осуществляется по ключевым словам, которые пользователь вводит в качестве поискового запроса. Дополнительно он может ввести ограничения по ряду параметров: год издания, издательство, название произведений, фамилия автора и т.п. Но в любом случае поиск осуществляется по ключевым словам. Допустим, что научный сотрудник, являясь специалистом в какой-то определенной области и имея публикации по этой теме, заинтересован в поиске публикаций, которые каким-то образом пересекаются с полем его интересов. В этом случае вместо того, чтобы вычленять ключевые слова из своей публикации, вводить их в качестве поискового запроса, а потом пытаться проранжировать результаты поиска, этот сотрудник может подать в качестве данных для поиска какую-нибудь свою статью, нейронная сеть её проанализирует и подберет те публикации, которые схожи с исходной.

В 2017 г. нейронная сеть *Watson* проанализировала все авторефераты по истории, которые хранились в электронной библиотеке РНБ. В качестве поискового запроса *Watson*'у были предложены случайным образом выбранные из того же множества авторефераты по истории. Очевидно, что на первом месте в списке результатов появлялся тот же самый автореферат, который был подан в качестве запроса. В подавляющем большинстве случаев дальше шел список, который по мнению специалистов, был не очень сильно релевантен исходному автореферату, но достаточно редко нейронная сеть выдавала результат, который был не ожидаем для экспериментаторов.

В частности, можно привести пример, когда на вход был подан автореферат по экономике забайкальского края в 1921–1924 гг. Ожидаемо, что на первом месте оказался сам этот автореферат, на последующих местах – шли авторефераты, посвященные экономике забайкальского края, но в другие годы, например, 1918–1921 гг. или 1924–1927 гг. Но экспериментаторов поразило, что на следующих позициях поискового результата *Watson* выдал авторефераты, посвященные экономике соседних краёв того же или смежных периодов: например, экономика прибайкальского края в 1921–1924 гг.

К сожалению, такие поразительно неожиданные результаты выдавались *Watson*'ом лишь в менее 10% случаев, что было катастрофически недостаточным для внедрения этой технологии поиска в промышленную эксплуатацию в тот момент. С тех пор количество узлов, используемых в нейронных сетях, значительно выросло, а технологии обучения, тестирования и фильтрации заведомо неподходящих результатов вышли на качественно новый уровень, что позволяет вернуться к экспериментам по поиску литературы в библиотеках не только с использованием ключевых слов, но и по поиску схожих текстов.

Генеративные трансформеры

В середине второго десятилетия XXI в. исследователи стали направлять свои усилия на создание генеративных предварительно обученных трансформеров (Generative Pre-trained Transformer – GPT). GPT – это тип больших языковых моделей, которые обучаются на больших наборах текстовых данных и способны генерировать текст, схожий с человеческим. Пионером в этом направлении стала основанная в 2015 г. Илоном Маском и Сэмюэлем Олтменом компания OpenAI [6]. В конце 2022 г. этой компанией был запущен чат-бот с искусственным интеллектом ChatGPT, который вызвал всплеск интереса за счет больших возможностей по ведению диалога с пользователем, учету контекста и генерации текстов, которые практически невозможно отличить от человеческой речи.

ChatGPT основан на языковой модели GPT-3.5, которая является улучшенной версией модели GPT-3. В свою очередь, в модели GPT-3 для обучения использовались наборы данных на различных языках общим объемом 570 ГБ текстов. Разработчики GPT-3 обучили восемь различных моделей, архитектуры которых были схожи друг с другом, но различались по количеству параметров машинного обучения. У максимальной модели количество параметров машинного обучения составляет 175 миллиардов [7]. Это привело к тому, что GPT-3 умеет отвечать на вопросы по прочитанному тексту, сочинять стихотворения в заданном стиле, решать математические примеры, генерировать тексты на заданную тему.

В 2015–2016 гг. в РНБ было проведено масштабное исследование, посвященное тому как пользователи читают книги, размещенные в электронной библиотеке РНБ. Исследование проводилось с 10 марта 2015 г. по 31 августа 2016 г. За этот период в электронной библиотеке РНБ было открыто 1 932 603 книги и просмотрено 20 376 397 страниц. Таким образом, в среднем пользователи просматривают 10,54 страницы в одной книге. Можно сделать вывод о том, что основной сценарий работы пользователя выглядит следующим образом: поиск книг в электронном каталоге с использованием поискового запроса, выбор нескольких книг в результатах поиска, открытие выбранных книг, ознакомление с несколькими первыми страницами, на которых располагаются аннотация и оглавление, закрытие книги. Подобное поведение пользователя сигнализирует о том, что информации, представленной в результатах поиска, недостаточно для принятия решения о необходимости прочтения книги: человеку требуется открыть её и ознакомиться с несколькими первыми страницами. Тогда и возникла идея о том, что размещение небольшой аннотации на странице с результатами поиска помогло бы пользователю более осознанно принимать решение о необходимости прочтения книги.

В 2016 г. в РНБ был проведен эксперимент с использованием нейронной сети *Watson* компании IBM по автоматической генерации аннотаций на основе текста книги. Тогда этот эксперимент был признан провальным, поскольку сгенерированные тексты были бессвязными, содержали большое количество ошибок и вызывали раздражение у читателей.

Недавние исследования показывают, что ChatGPT превосходит людей в задаче аннотирования текстов. Например, в эксперименте, проведенном в марте-мае 2023 г., на 6 183 твитах и газетных статьях, ChatGPT превзошел нанятых сотрудников в задачах определения значимости текста, занятой политической позиции, предмета обсуждения и выявления структуры сообщения [8]. Вполне возможно, что имеет смысл вернуться к эксперименту по автоматическому аннотированию текстов, проведенному в РНБ в 2016 г., но уже с использованием современных моделей нейронных сетей.

Таким образом, еще одно из возможных направлений использования искусственного интеллекта в библиотеках – это автоматическое создание аннотаций для электронных книг с использованием генеративных предварительно обученных трансформеров.

Создание научных статей

В современных редакторах изображений зачастую встроена опция, позволяющая «улучшить» ту или иную фотографию с использованием искусственного интеллекта. Более того, в последних поколениях смартфона iPhone компании Apple подобный функционал встроен в стандартное приложение «Камера» и является неотключаемым. Возникает вопрос: каким образом программное обеспечение понимает, что представляет собой улучшение, а что нет? Ведь все люди разные, и то, что для одного пользователя является улучшением, например, уменьшение носа и увеличение губ, для другого абсолютно неприемлемо.

Скорее всего, функция «улучшения» настраивается индивидуальным образом и зависит от истории редактирования пользователем своих фотографий. Если человек на всех своих исходных фотографиях изменяет одну и ту же часть, то можно сделать вывод, что именно это преобразование и является «улучшением» в его понимании. Программное обеспечение может сохранять исходную версию изображения и версию, созданную пользователем, и, сравнивая эти версии, выделить некий шаблон изменений, который человек использует раз за разом. Возможно, применив этот шаблон к только что сделанной фотографии, получится изображение, которое с большей вероятностью понравится владельцу телефона.

Подобно тому, как пользователь телефона раз за разом редактирует свои фотографии, доводя их до совершенства, точно так же и научный сотрудник работает над своими публикациями, изменения структуру, расположение блоков текста, дополняя статью фактами и перефразируя различные стилистические обороты. Теоретически можно создать нейронную сеть и обучить её на массиве промежуточных вариантов публикаций автора для того, чтобы сеть смогла выделить характерные шаблоны изменений, присущие именно этому автору, и в дальнейшем была бы способна применять эти шаблоны для «улучшения» статей еще на этапе авторского редактирования.

Еще один рутинный процесс при создании научных статей – это работа со списками библиографических ссылок. Нейронные сети вполне способны по-

нять, в каком месте научной статьи автор ссылается на данные из другой статьи, корректно оформить цитату или ссылку на источник, а в библиографический список добавить упоминание исходной статьи в том стандарте, который принят правилами журнала или учреждения, в котором работает автор. Нейронные сети могут автоматически проверить все встречающиеся в статье библиографические ссылки на точность, корректность, соответствие источникам, а также на непротиворечивость приводимой информации с данными из статьи, на которую дана ссылка.

ЗАКЛЮЧЕНИЕ

За последние десять лет в развитии нейронных сетей произошел значительный прогресс. То, что они выдают на запросы пользователей, поражает воображение и, одновременно, обескураживает. Тексты, генерируемые нейронными сетями, невозможно отличить от тех, которые пишет человек. А связность текста, логичность повествования, соблюдение причинно-следственных связей в этих текстах позволяет им претендовать на звание научной статьи.

К сожалению, невозможно утверждать, что искусственный интеллект может выполнять творческие функции. На самом деле, он всего лишь научился работать с огромными массивами информации, вычленять из них какие-то данные, компоновать их между собой и предоставлять человеку в том виде, который ему привычен и удобен.

Используя принцип самоприменимости, среди исследователей стало модным спрашивать, например, у того же ChatGPT, что он думает о той или иной проблеме, связанной с искусственным интеллектом. В частности, на вопрос о том, способен ли искусственный интеллект написать научную статью, ChatGPT честно ответил, что он лишь способен сгенерировать большой объем текста на заданную тему, но это никаким образом не будет оригинальным исследованием, поскольку этот текст будет основан на существующих разработках, т. е. этот сгенерированный текст не будет содержать каких-либо новых идей, подходов и выводов [9]. Таким образом, пока рано говорить, что искусственный интеллект и нейронные сети, как один из видов реализации искусственного интеллекта, способны к действительно творческой деятельности в максимально широком значении этого понятия.

Тем не менее, на современном этапе развития нейронных сетей можно предположить, что они могут облегчить библиотекарям выполнение операций, которые напрямую не связаны с творческой составляющей, а именно:

- автоматическое проставление тэгов в зависимости от того, что изображено на иллюстрациях, хранящихся в электронной библиотеке произведений, для облегчения последующего поиска по изображениям;
- поиск по электронному каталогу не только с использованием ключевых слов, но и поиск схожих документов на основе полного текста;
- автоматическое аннотирование научных статей и книг, хранящихся в электронной библиотеке;

- использование шаблонов редактирования статей, характерных для того или иного научного сотрудника;
- анализ библиографических ссылок на точность, корректность, соответствие источникам и не-противоречивость.

В свою очередь, библиотеки могут выступать поставщиками больших массивов проверенных данных, которые необходимы нейронным сетям для обучения.

Таким образом, можно предположить, что в течение ближайших лет мы станем свидетелями взаимно выгодного сотрудничества библиотечного сообщества и программно-аппаратных комплексов, основанных на использовании систем искусственного интеллекта.

СПИСОК ЛИТЕРАТУРЫ

1. Аверкин А.Н., Гаазе-Рапопорт М.Г., Поступолов Д.А. Толковый словарь по искусственному интеллекту. – Москва: Радио и связь, 1992. – 256 с.
2. Turing A.M. Computing machinery and intelligence // Mind. – 1950. – Vol. LIX, Issue 236 (October). – P. 433–460. DOI: <https://doi.org/10.1093/mind/LIX.236.433>.
3. Campbell M., Hoane A.J.Jr., Feng-Hsiung Hsu. Deep Blue // Artificial Intelligence. – 2002. – Vol. 134, № 1. – P. 57-83. DOI:10.1016/S0004-3702(01)00129-1.
4. Orlov I.I., Larskikh E.L. Neural networks are a new milestone in the development of science and

- designe. Brief overview or the problem // 25 апреля 2021 года, 2021. – P. 77-81. DOI: 10.37539/SRP296.2021.13.83.005.
5. Каптерев А.И. Когнитивный менеджмент и искусственный интеллект в библиотеках: возможности и особенности // Научные и технические библиотеки. – 2023. – № 6. – С. 113-137. DOI: <https://doi.org/10.33186/1027-3689-2023-6-113-137>.
 6. Официальный сайт компании OpenAI. – URL: <https://openai.com> (дата обращения: 18.08.2023).
 7. Brown T. et al. Language models are few-shot learners // Advances in neural information processing systems. – 2020. – Vol. 33. – P. 1877-1901. DOI: <https://doi.org/10.48550/arXiv.2005.14165>.
 8. Gilardi F., Alizadeh M., Kubli M. Chatgpt outperforms crowd-workers for text-annotation tasks // arXiv preprint arXiv:2303.15056. – 2023. DOI: <https://doi.org/10.48550/arXiv.2303.15056>.
 9. Grimaldi G., Ehrler B. et al. Machines Are About to Change Scientific Publishing Forever // ACS Energy Letters. – 2023. – Vol. 8(1). – P. 878-880. DOI: 10.1021/acsenergylett.2c02828.

Материал поступил в редакцию 21.08.23.

Сведения об авторе

ШОРИН Олег Николаевич – кандидат технических наук, директор Библиотеки по естественным наукам Российской академии наук (БЕН РАН), Москва
e-mail: shorin@benran.ru

Особенности взаимодействия российских студентов с научными ресурсами: результаты опроса

Представлены результаты опроса студентов российских вузов различных степеней обучения (бакалавриат, магистратура, аспирантура) о том, как они решают задачи по поиску научной информации и получению доступа к ней. Опрос был проведен в апреле-мае 2023 г., в нем приняли участие 612 студентов. Анализ ответов показал, что заметен явный прогресс по мере продвижения студентов к более высоким ступеням обучения, однако в целом уровень научной грамотности даже в рамках рассматриваемых аспектов остается довольно низким, что может быть связано в первую очередь с отсутствием соответствующего системного обучения.

Ключевые слова: научная грамотность, научная коммуникация, новые медиа, студенты университета

DOI: 10.36535/0548-0019-2023-11-3

ВВЕДЕНИЕ

Огромные ресурсы и государство, и отдельные организации регулярно выделяют на информационное обеспечение развития научной сферы. Однако часто в стороне остается вопрос об обучении студентов научной грамотности – в том числе тому, где и как можно получить доступ к научной литературе, как ее отбирать и обрабатывать и т.д. Такое обучение, как правило, или отсутствует, или фактически происходит случайным образом уже в процессе самостоятельной научной работы [1, 2].

С первого курса студенты сталкиваются с задачами обработки определенного массива научных источников, но за неимением соответствующих знаний и навыков оказываются лишены инструментов для качественного решения проблемы их отбора и анализа. В результате внимание учащихся захватывают полулегальные ресурсы, содержащие агрегированную и часто не вполне корректную информацию. В дальнейшем эти студенты становятся аспирантами и продолжают обращаться к знакомым практикам, которые препятствуют как их собственному профессиональному росту, так и развитию научной сферы в целом. Так, проблемы плагиата и использования ненадежных источников частично связаны именно с незнанием, а не с преднамеренным нарушением учащимися правил научной этики. Представляется, что научная грамотность, а также культура научной коммуникации, в том числе работы с научной литературой, – это необходимое условие дальнейшего развития науки и общества в целом [3, 4].

В настоящем исследовании мы ставили задачи определить: 1) насколько российские студенты адаптированы к современной инфраструктуре научной

коммуникации и знакомы с ключевыми понятиями открытой науки; 2) как различается уровень соответствующих компетенций студентов на разных ступенях обучения. Решение этих задач позволяет оценить степень развития культуры научной коммуникации среди российских студентов и сформулировать рекомендации по исправлению сложившейся ситуации.

ОБЗОР ЗАРУБЕЖНОЙ ЛИТЕРАТУРЫ ПО ПРОБЛЕМЕ

Работа с научными источниками тесно связана с таким понятием, как «научная грамотность», хотя, конечно, его содержание гораздо шире, чем просто умение работать с научной информацией. С точки зрения современных исследователей, научная грамотность, как правило, включает в себя понимание базовых научных концепций, этики и ценностей науки, взаимосвязей между наукой, технологиями и обществом [5]. Однако полного согласия в использовании этого термина нет, еще в прошлом веке было опубликовано немало работ, авторы которых стремились как-то обобщить это понятие [6]. В результате объединения различных интерпретаций и получилось достаточно обширное определение, предполагающее знакомство человека с огромным количеством концепций в том числе из области общественных и социальных наук.

Некоторые исследователи отмечают, что существует несколько основных трактовок понятия «грамотность»: 1) как умения читать и писать и 2) как образованности. Если мы используем первую трактовку, то это понятие можно расширить до как «способность человека читать и понимать научные тексты и выражать мнение по научным вопросам» [7].

Первый шаг в освоении научной информации – ее поиск, получение доступа и возможности оценки разных источников на предмет их надежности. Для нас особый интерес представляют результаты изучения того, как через эти этапы проходят студенты, аспиранты и молодые исследователи в разных странах. На эту тему опубликовано огромное количество статей, в которых, как правило, описываются конкретные кейсы университетов, в лучшем случае – городов [8–14]. Специфика выбора тех или иных ресурсов различается в странах и дисциплинах, однако можно выделить несколько общих трендов и проблем, затронутых авторами: перенос внимания учащихся на электронные источники информации вместо печатных; постепенная переориентация на специализированные, в том числе подписные, ресурсы по мере продвижения студента к статусу исследователя и приобретения им соответствующих навыков; отсутствие специальной подготовки молодых исследователей для работы с научными базами и специализированными сервисами, в результате чего им часто приходится все осваивать самостоятельно. В некоторых случаях выводы специалистов особенно неутешительны: студенты имеют довольно слабое представление об инфраструктуре научной коммуникации даже после выпуска [15], предпочитают интернет-источники библиотечным и считают, что в интернете гораздо больше шансов найти релевантную информацию [16].

Основная причина отсутствия необходимых знаний и навыков у студентов в том, что они не получают соответствующего системного обучения. Так, например, авторы [17] считают, что программы обучения в основном направлены на передачу огромных объемов фактологического материала, и предполагается, что студенты каким-то «магическим образом» получат умения обрабатывать научную информацию за четыре года обучения. В результате нужные навыки приобретаются случайным образом, что зависит от того, насколько удачно для студента сложатся обстоятельства [2].

Результаты исследований, направленных на выявление влияния специального обучения на практики студентов и исследователей, демонстрируют эффективность такого подхода: в итоге зафиксировано значительное увеличение количества студентов, которые обращаются к специализированным базам данных вместо неопределенных источников из Интернета, используют более одной базы для поиска и большее количество критериев для оценки надежности информации из интернета [18].

Отсутствие соответствующих знаний и навыков для пользования базами данных в некоторых случаях сохраняется и после завершения обучения, и в начале профессиональной деятельности [14], что не может не сказываться на качестве научной работы.

Некоторые исследования, в рамках которых изучается использование научной информации, акцентируют внимание на роли библиотек [19] и новой функции библиотек как электронного портала к практически бесконечным объемам знаний [20].

Российские публикации, затрагивающие вопросы научной грамотности студентов, концептуально опираются на указанные выше труды и чаще всего со-

держат лишь рассуждения о возможностях или практике развития соответствующих навыков среди учащихся конкретных кафедр или факультетов. Результаты единственного схожего с нашим эмпирического исследования, которое удалось обнаружить, в значительной степени согласуются с полученными нами: хотя, по данным автора, все респонденты знакомы с такими базами, как Scopus и Web of Science, почти 84% участников опроса не проходили обучения их использованию. Отметим, что выборка данного исследования включает всего 31 человека, 24 из которых являются аспирантами или молодыми учеными [21].

МЕТОДИКА И РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Мы хотели выяснить, насколько результаты зарубежных исследователей соответствуют российской ситуации и насколько явно выделенные ими проблемы проявляются на примере студентов отечественных вузов. Чтобы получить представление о ситуации в стране, а не в нескольких конкретных учебных заведениях, мы провели опрос учащихся университетов во всех федеральных округах России, направив в вузы анкету с просьбой распространить ее среди студентов.

Нас интересовали лишь те аспекты научной грамотности, которые связаны со «способностью читать», т.е. знанием, где можно найти релевантные научные источники, как можно получить к ним доступ и оценить, насколько источник может быть достоин доверия. Можно посчитать, что мы несколько вышли за заявленные рамки, так как в анкете были и вопросы, помогающие прояснить, насколько студенты знакомы с некоторыми важными в контексте современной системы научной коммуникации тенденциями (открытый доступ и использование открытых лицензий). В данном случае мы использовали ответы как косвенную характеристику уровня знаний о том, где и как искать информацию и получать доступ к ней.

Нас интересовало несколько ключевых вопросов о взаимодействии студентов с научным контентом:

- 1) с какими типами источников и на каких языках они работают;
- 2) через какие платформы и сервисы они осуществляют поиск научной информации;
- 3) через какие платформы и сервисы они получают доступ к научной информации, как часто к необходимой информации не удается получить доступ, обращаются ли они к пиратским ресурсам;
- 4) знакомы ли им понятия «открытый доступ» и «открытые лицензии».

Опрос проводился в апреле-мае 2023 г., в нем приняли участие учащиеся бакалавриата, магистратуры и аспирантуры. Всего было получено 612 ответов.

Большая часть наших респондентов были студентами региональных университетов наиболее крупных и развитых городов региона (табл. 1).

Мы также выявили области науки, к которым относится направление подготовки обучающихся (табл. 2). В нашем опросе с некоторым перевесом преобладают представители социогуманитарной сферы, однако значительную долю составляют и студенты технических специальностей.

Таблица 1

Распределение респондентов по типу города обучения

Тип города обучения	Респонденты	
	количество	%
Областной / краевой / республиканский	363	59
Другое	180	29
Столица	69	11
Всего	612	100

Таблица 2

Распределение респондентов по областям знания

Область знания	Респонденты	
	количество	%
Гуманитарные науки	242	40
Технические науки	216	35
Социальные науки	78	13
Естественные науки	76	12
Всего	612	100

Таблица 3

Распределение респондентов по ступени обучения

Ступень обучения	Респонденты	
	количество	%
Бакалавриат	503	82
Аспирантура	63	10
Магистратура	46	8
Всего	612	100

Нам важно было понять, как меняются знания и навыки учащихся по мере их «взросления» – приобретают ли они необходимые для научной работы компетенции уже в бакалавриате или требуется переход на более высокие ступени обучения (табл. 3).

В опросе приняло участие не так много магистрантов и аспирантов, однако собранные данные позволяют сделать некоторые предположения и на их счет. Именно принадлежность к той или иной ступени обучения стала для нас одним из ключевых параметров сравнения того, как меняются подходы к работе с источниками научной информации по мере продвижения студента в обучении.

Безусловно, мы осознаем все ограничения настоящего исследования: оно основано на результатах опроса, а мы не вполне уверены в том, насколько на самом деле рефлексивно студенты относятся к своим практикам взаимодействия с научным контентом, возможны некоторые искажения (особенно на уровне бакалавриата), которые без привлечения качественных методов исследования не представляется возможным ни отследить, ни скорректировать.

В соответствии с поставленными нами задачами вопросы анкеты были разделены на четыре основных блока, каждый из которых связан с одним из аспектов взаимодействия учащихся с системой научной коммуникации.

**Источники научной информации:
типы и языки**

Один из ключевых показателей развитости навыков работы с научной информацией – это способность отбирать достаточно надежные источники, на которые можно опереться при работе над собственным исследованием. Рассматривая данные, собранные в рамках нашего опроса, мы можем наблюдать наглядное улучшение ситуации: по мере продвижения от бакалавриата к аспирантуре студенты все меньше внимания уделяют учебным работам, которые на первых курсах являются для них одним из основных источников информации, и все чаще обращаются к научным статьям, диссертациям и, что немаловажно, к монографиям, позволяющим более глубоко и подробно разобраться в изучаемом вопросе (табл. 4).

Помимо знакомства с научной литературой большое значение имеет способность студентов обращаться с источниками на иностранных языках, так как вне зависимости от тематики исследований наука – это международное явление, поэтому пренебрежение трудами зарубежных авторов практически неизбежно обедняет любую работу. И здесь мы тоже видим явную тенденцию к расширению спектра источников по мере «взросления» учащихся (табл. 5).

Таблица 4

**Типы источников научной информации, используемых учащимися разных ступеней обучения
(множественный выбор ответа)**

Типы источников	Студенты бакалавриата		Магистранты		Аспиранты	
	кол-во	%	кол-во	%	кол-во	%
Учебные работы	403	80	35	76	19	30
Статьи в научных журналах	394	78	39	85	61	97
Диссертации	156	31	24	52	55	87
Монографии	88	17	12	26	46	73
Другое	10	2	3	7	2	3
Методические и учебные материалы	3	1	0	0	1	2
Ничего	1	0	0	0	0	0
Всего	503	100	46	100	63	100

Таблица 5

**Языки источников научной информации, используемых учащимися разных ступеней обучения
(кроме русского, множественный выбор ответа)**

Язык публикации	Студенты бакалавриата		Магистранты		Аспиранты	
	кол-во	%	кол-во	%	кол-во	%
Английский	131	26	15	33	35	56
Немецкий	4	1	0	0	2	3
Другие	14	3	0	0	0	0
Любой	0	0	0	0	2	3

В процессе исследования мы установили, что источники на английском языке используют больше половины опрошенных аспирантов, а небольшая их доля, благодаря привлечению современных онлайн-переводчиков и искусственного интеллекта, взаимодействует с литературой на любом языке. Наиболее важна для нас здесь именно тенденция: если в бакалавриате и магистратуре к иностранным источникам обращается только около трети студентов, то в аспирантуре это уже почти две трети учащихся.

Поиск научной информации

Понять, что собой представляют используемые студентами источники научной информации, можно с помощью определения площадок, которые они выбирают для поиска литературы и доступа к ней. Нам было важно узнать, с какими сервисами студенты знакомы достаточно хорошо и какими они пользуются для поиска научной информации. На этот вопрос у нас не было готовых вариантов ответа. Полученные от респондентов ответы мы вручную кластеризовали следующим образом:

- общие поисковые сервисы – «Яндекс», Google, другие поисковые сервисы, Интернет;
- специализированные научные поисковые сервисы – Google Scholar, Scopus, Web of Science, Dimensions;
- электронные научные библиотеки – eLibrary.ru, «КиберЛенинка», сайт РГБ, Dissercat, библиотеки университетов;
- соцсети и блоги – различные каналы и паблики, в том числе научно-популярные;

- специализированные соцсети – Academia, ResearchGate;
- Электронные библиотечные системы – «Лань», «Юрайт», «Университетская библиотека онлайн»;
- интернет-ресурсы – «Студопедия», «Инфоурок», сайты с готовыми учебными работами;
- видеосервисы – YouTube;
- Научно-популярные СМИ – «Элементы», N+1, Naked science.

Среди сайтов научных / образовательных организаций респонденты называли сайт Российской академии наук, а также, по-видимому, сайты своих университетов. Из сайтов научных журналов и издательств чаще всего упоминались ScienceDirect и Springer Link.

К категории «другое» были отнесены ресурсы, не располагающие собственным научным контентом: базы «Гарант» и «Консультант Плюс», Росреестр, различные источники статистических данных.

Такой подход к работе с респондентами позволил избежать ситуации, при которой они выбирают в списке знакомое название, однако на самом деле мало знают сервис или отмечают категорию ресурсов, сущности и назначения которых на самом деле не понимают.

В результате мы выявили некоторую разницу в подходах к поиску информации среди студентов разных ступеней обучения (табл. 6). Если для студентов бакалавриата и магистрантов характерно в первую очередь обращение к универсальным поисковым сервисам, то аспиранты в основном рассчитывают на электронные научные библиотеки (в подавляющем большинстве это eLibrary и «КиберЛенинка»). Аспиранты относительно активно используют университетские электронные библиотечные системы (ЭБС).

Использование онлайн-инструментов и сервисов для поиска научной информации учащимися бакалавриата, магистратуры и аспирантуры (множественный выбор ответа)

Инструменты поиска	Студенты бакалавриата		Магистранты		Аспиранты	
	кол-во	%	кол-во	%	кол-во	%
Общие поисковые сервисы	195	39	20	43	12	19
Электронные научные библиотеки	121	24	16	35	43	68
Википедия	65	13	4	9	2	3
Другое	41	8	2	4	6	10
Интернет-ресурсы	35	7	5	11	2	3
ЭБС	24	5	1	2	7	11
Специализированные научные поисковые сервисы	14	3	3	7	15	24
Соцсети и блоги	8	2	2	4	0	0
Сайты научных / образовательных организаций	8	2	0	0	0	0
Сайты научных журналов и издательств	6	1	2	4	10	16
Видеосервисы	6	1	0	0	0	0
Научно-популярные СМИ	4	1	1	2	1	2
Специализированные соцсети	0	0	0	0	4	6
Всего студентов	503	100	46	100	63	100

Можно отметить, что интерес учащихся к «Википедии» как к «точке входа» в тему постепенно теряется, что, впрочем, может быть связано с тем, что на уровне аспирантуры молодой исследователь уже, как правило, хотя бы в некоторой степени, разбирается в изучаемой проблематике и соответствующей литературе, т. е. возможно, при переходе к новой теме, обращение к энциклопедии вновь становится актуальным даже на последней ступени обучения.

Еще одно заметное изменение в выборе сервисов – обращение к специализированным поисковым системам, что почти не практикуется до аспирантуры, хотя Google Scholar, например, не требует подписки или каких-то специальных навыков взаимодействия с ним. Дополнительно можно отметить гораздо более интенсивное обращение к сайтам научных журналов и издательств, а также к специализированным социальным сетям, о которых студенты бакалавриата и магистранты, похоже, не имеют представления.

Доступ к научной информации

Кластеризация сервисов для данного раздела осуществлялась по тому же принципу, что и для ресурсов поиска научной информации. Нам было важно разграничить эти этапы работы в том числе и для того, чтобы понять, насколько для самих студентов они различаются и осознаются как связанные, но все-таки отдельные (с учетом всех особенностей системы научной коммуникации) процессы.

В целом результаты опроса на тему инструментов доступа к научной информации (табл. 7) очень схожи с тем, что мы видели в отношении инструментов поиска научной информации. Это проявляется и в ди-

намике изменения предпочтений при продвижении учащегося от ступени к ступени.

Еще одно наблюдение, которое нам удалось сделать, состоит в том, что если студенты бакалавриата используют один-два типа источников информации, то в аспирантуре это уже три-четыре вида поисковых инструментов / ресурсов с научной литературой, что свидетельствует о развитии навыков работы с научной информацией и расширении спектра источников, с которыми студент способен работать.

При этом проблемы доступа к научной литературе чаще возникают именно у аспирантов, несмотря на то что они демонстрируют более полные знания о возможностях комплексного использования различных научных ресурсов (табл. 8).

Это объясняется прежде всего тем, что именно у аспирантов значительно возрастает потребность в доступе к зарубежной литературе, что сейчас обеспечить особенно сложно. Кроме того, аспиранты гораздо чаще для поиска используют специализированные базы данных и платформы крупнейших издательств, а это неизбежно приводит их к литературе, доступной только по подписке.

Отдельно мы постарались прояснить вопрос об использовании студентами пиратских ресурсов, и здесь их доля оказалась заметно выше, чем может показаться из табл. 8 (табл. 9).

Такая разница в показателях, объясняется, вероятнее всего, тем, что использование пиратских сервисов не является рутинной практикой у студентов, так как они предпочитают обращаться к статьям на русском языке, размещенным в российских же онлайн-библиотеках.

Таблица 7

Использование онлайн-инструментов и сервисов для доступа к научной информации учащимися бакалавриата, магистратуры и аспирантуры (множественный выбор ответа)

Инструменты поиска	Студенты бакалавриата		Магистранты		Аспиранты	
	кол-во	%	кол-во	%	кол-во	%
Электронные научные библиотеки	153	30	18	39	50	79
Общие поисковые сервисы	115	23	10	22	7	11
Википедия	49	10	2	4	1	2
Другое	44	9	2	4	4	6
ЭБС	41	8	1	2	6	10
Интернет-ресурсы	29	6	2	4	4	6
Соцсети и блоги	14	3	1	2	0	0
Специализированные научные поисковые сервисы	13	3	0	0	11	17
Сайты научных / образовательных организаций	13	3	2	4	4	6
Сайты научных журналов и издательств	5	1	1	2	6	10
Видеосервисы	3	1	0	0	0	0
Научно-популярные СМИ	2	0	0	0	1	2
Специализированные соцсети	1	0	0	0	5	8
Пиратские ресурсы	0	0	1	2	1	2
Всего	503	100	46	100	63	100

Таблица 8

Частота возникновения проблем с доступом к научным материалам у учащихся разных ступеней обучения

Частота возникновения проблем	Студенты бакалавриата		Магистранты		Аспиранты		Все	
	кол-во	%	кол-во	%	кол-во	%	кол-во	%
достаточно часто, но более половины контента доступно	225	45	19	41	32	51	276	45
довольно редко, в случае специфических источников	187	37	20	43	17	27	224	37
каждая вторая статья или чаще	48	10	6	13	11	17	65	11
никогда	43	9	1	2	3	5	47	8

Таблица 9

Использование учащимися разных ступеней обучения пиратских ресурсов для доступа к научной информации

Ответ	Студенты бакалавриата		Магистранты		Аспиранты		Все	
	кол-во	%	кол-во	%	кол-во	%	кол-во	%
да	90	18	9	20	12	19	111	18
нет	413	82	37	80	51	81	501	82

Таблица 10

Ответы учащихся разных ступеней обучения на вопрос о том, встречают ли они научную информацию в социальных сетях, %

Встречаете ли Вы научную информацию в социальных сетях?	Студенты бакалавриата	Магистранты	Аспиранты
да	57	50	38
нет	43	50	62

Социальные сети, в которых учащиеся встречают научную информацию

Соцсеть	Респонденты	
	кол-во	%
ВК	231	65
Telegram	69	19
Instagram*	14	4
YouTube	10	3
Tiktok	8	2
Другое	10	3

*Принадлежит компании Meta, которая признана экстремистской организацией и запрещена на территории России.

Оценка полезности информации из социальных сетей для целей науки и образования

Полезна ли информация из соцсетей для учебы / исследований?	Респонденты	
	количество	%
Да	233	66
В некоторой степени	33	9
Нет	22	6
Затруднились с ответом	66	19
Всего	354	100

Среди наиболее популярных видов пиратских ресурсов были названы торренты, Telegram-каналы и Sci-Hub. Это позволяет предположить, что «научную информацию» студенты трактовали довольно широко, так как торренты как минимум вряд ли могут быть первоочередным инструментом для доступа к результатам наиболее актуальных научных исследований.

Последний аспект в контексте получения доступа к научной информации – использование в этих целях социальных сетей (табл. 10). Здесь мы можем видеть тенденцию сокращения доли респондентов, обращавшихся к данному источнику информации, хотя ранее было установлено, что только в аспирантуре студенты начинают пользоваться специализированными социальными сетями, которые действительно могут быть очень полезны для получения доступа к научной литературе – особенно зарубежной.

Всего научную информацию в социальных сетях встречают 354 опрошенных. Вероятно, термин «научная информация» и здесь трактуется ими расширительно, так как лишь очень немногие привели в своих отвечах ссылки на ResearchGate и Academia.edu (табл. 11) – их было так мало, что в итоге они попали в категорию «другое».

Социальная сеть «ВКонтакте» за последние полтора десятилетия стала настоящим хабом контента по самым разным темам, и возможность поиска документа по бескрайним хранилищам очень способствует использованию этой сети студентами, поэтому результаты опроса не вызывают удивления. Telegram используется для запуска различного вида ботов, с помощью которых можно получить доступ к научным статьям. В то же время остается не вполне ясным, как этим целям могут служить такие социаль-

ные сети, как Instagram, YouTube и Tiktok. Это требует отдельного исследования.

Помимо вопроса о самом факте контакта с научной информацией в соцсетях мы также задавали вопрос о том, насколько полезной она оказывается (табл. 12).

Полученные нами результаты не позволяют делать выводы о характере, глубине, научной ценности и значимости тех материалов, которые студенты получили из социальных сетей, но зато они демонстрируют, насколько важна эта коммуникационная среда для молодого поколения, насколько высоко его представители ценят социальные сети, доверяют им и какое место отводят в своем медиапотреблении.

Осведомленность об открытом доступе

Последний из рассматриваемых нами аспектов связан с открытым доступом. Мы посчитали необходимым затронуть этот вопрос в связи с тем, что он, с одной стороны, позволяет прояснить, насколько студенты в принципе знакомы с особенностями функционирования системы научной коммуникации на данном этапе, а с другой стороны – показывает, как они понимают свои права и возможности в обращении с научным контентом.

Один из ключевых элементов для открытого доступа – открытые лицензии, без которых невозможно обеспечить должные правовые условия для организации постоянного открытого доступа к научному контенту. С нашей точки зрения, знакомство с этим правовым инструментом в значительной степени характеризует степень осведомленности о теме открытого доступа в целом. Как мы можем видеть, пока не так много студентов обладают соответствующими знаниями (табл. 13).

Таблица 13

Ответы респондентов на вопрос о том, известно ли им понятие «открытые лицензии»

Ответ	Студенты бакалавриата		Магистранты		Аспиранты		Все	
	кол-во	%	кол-во	%	кол-во	%	кол-во	%
нет	270	54	32	70	33	52	335	55
да	233	46	14	30	30	48	277	45

Таблица 14

Ответы респондентов на вопрос о том, известно ли им понятие «открытый доступ»

Ответ	Студенты бакалавриата		Магистранты		Аспиранты		Все	
	кол-во	%	кол-во	%	кол-во	%	кол-во	%
Нет	184	37	14	33	15	24	213	35
Да	319	63	32	67	48	76	399	65

Почти половина респондентов заявила о том, что им известно об открытых лицензиях, однако не вполне понятно, насколько этим утверждениям можно доверять, так как при ответе на аналогичный вопрос об открытом доступе (табл. 14) доля осведомлённых студентов оказалась еще выше, однако на практике лишь 23 человека (4%) затем смогли привести определение, в котором фигурировали формулировки из Будапештской инициативы «Открытый доступ» или упоминались открытые лицензии. В основном респонденты или затруднились с ответом (36%), или под открытым доступом понимали «бесплатно доступный в интернете» (58%). Никакой взаимосвязи точности трактовки термина со степенью обучения нами обнаружено не было.

Полученные нами результаты следует интерпретировать крайне осторожно – значительная доля тех, кто уверен в своей осведомленности, вероятно, на самом деле просто где-то слышал указанные термины, но с их точным значением не знаком.

ВЫВОДЫ И ДИСКУССИЯ

Все аспекты научной грамотности, которые мы рассматривали, являются начальными, не требует какой-то сложной подготовки, владения специальными программами или наличия дорогостоящих подписок на научные сервисы. Единственное, что необходимо для их использования (степень его эффективности мы сейчас оставляем в стороне, так как соответствующие навыки мы не оценивали), – это знакомство с основным набором инструментов и сервисов, с помощью которых можно искать научную информацию и получать к ней доступ.

На данный момент можно заключить, что на системном уровне даже такого поверхностного обучения не проводится. Безусловно, к аспирантуре многие обучающиеся уже лучше ориентируются в инфраструктуре научной коммуникации и, возможно, в каких-то современных инструментах работы с информацией, но, судя по результатам опроса, можно

предположить, что наша ситуация вполне соответствует результатам предшествующих исследований: навыки в основном приобретаются «волею случая» – наблюдением за действиями коллег, случайным обнаружением сервисов в Интернете, знакомством с информацией об их использовании в социальных сетях и т.д. Поэтому значительная часть студентов так и остается без самых необходимых знаний и навыков поиска и использования источников, что не может не сказываться на качестве учебных и научных работ.

Особые опасения вызывает повышенное внимание студентов к готовым учебным работам вместо научной литературы. Конечно, нет ничего плохого и в том, чтобы ориентироваться на труды коллег-студентов, если это работы из университетской библиотеки или профильного репозитория, но часто источниками этих учебных работ являются сомнительные сайты, которые невозможно корректно процитировать, так как их авторы часто не указываются, а тексты состоят из заимствований непонятного происхождения. Более того, в них возможны ошибки, постоянное воспроизведение которых в новых работах препятствует исправлению ситуации.

Еще одна заметная проблема – отсутствие практики использования учащимися литературы на иностранных языках, хотя сегодня для эффективной работы даже с научными текстами существует достаточно инструментов для перевода. Источником проблемы может быть или представление о том, что незнание языка – непреодолимый барьер, или же непонимание того, что для большинства исследований пренебрежение зарубежными источниками неизбежно оставляет лакуны в работе и негативно сказывается на их качестве. Оба этих вопроса решаются с помощью информирования и обучения, которым сегодня, очевидно, не уделяется достаточно внимания.

Результаты опроса об использовании учащимися конкретных инструментов и сервисов интересны в первую очередь отсутствием упоминания некоторых ресурсов. Например, ни разу в списках источников не были упомянуты репозитории открытого доступа,

сервисы типа Unpaywall, «Архив Интернета», «Открытая библиотека», а также пиратские сервисы, которые активно используются исследователями по всему миру. Если детально разобрать пункт «Электронные научные библиотеки», то выясняется, что в основном вся литература берется учащимися из двух источников – eLibrary и «КиберЛенинки», что гораздо лучше, чем сайты рефератов, но недостаточно для полноценных научных исследований. Вопрос использования пиратских ресурсов для доступа к научной информации очень сложен и неоднозначен, но знакомство с этой стороной научной коммуникации также необходимо для того, чтобы студенты осознавали все риски, связанные с такого рода сервисами.

Обращение к непрофильным социальным сетям в поиске научной информации несет самые разные риски. Для их нивелирования необходима разъяснительная работа, которая относится скорее к медиаграмотности в целом, чем отдельно к научной грамотности. Практики обмена информацией и обсуждения в том числе научных проблем в социальных сетях могут способствовать повышению качества исследований и развитию исследовательской культуры, однако для этого участники коммуникации должны иметь соответствующую подготовку. Также остается вопрос о том, как сервисы типа YouTube и Tiktok используются для передачи научной информации, но эта проблема требует отдельного комплексного исследования.

Открытый доступ и связанные с ним практики пока остаются за пределами знаний учащихся о современной научной коммуникации. В условиях осложненного доступа к научным ресурсам особенно важно уделять больше внимания этим вопросам и обучать студентов работе с соответствующими сервисами, а также разъяснить правовые условия использования контента, опубликованного под разными лицензиями, в том числе данных и сопутствующих исследовательских материалов, пригодных к повторному использованию.

СПИСОК ЛИТЕРАТУРЫ

1. Rothstein H.R. Accessing relevant literature // APA Handbook of Research Methods in Psychology. – 2012. – Vol. 1. – P. 133–144. DOI: 10.1037/13619-009.
2. Cirino L.A., Emberts Z., Joseph P.N., Allen P.E., Lopatto D., Miller C.W. Broadening the Voice of Science: Promoting Scientific Communication in the Undergraduate Classroom // Ecol Evol. – 2017. – Vol. 7. – P. 10124-10130. DOI: 10.1002/ece3.3501.
3. Brownell S. E., Price J. V., Steinman L. Science Communication to the General Public: Why We Need to Teach Undergraduate and Graduate Students this Skill as Part of Their Formal Scientific Training // Journal of Undergraduate Neuroscience Education (JUNE): a publication of Faculty for Undergraduate Neuroscience (FUN). – 2013. – Vol. 12, №1. – E6–E10. – URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3852879/>
4. Kuehne L.M., Twardochleb L.A., Fritschie K.J., Mims M.C., Lawrence D.J., Gibson P.P., Stewart-Koster B., Olden J.D. Practical Science Communication Strategies for Graduate Students // Conservation Biology. – 2014. – Vol. 28. – P. 1225-1235. DOI: 10.1111/cobi.12305.
5. Ozdem Yilmaz Y., Çavaş P., Çavas B., Çakiroglu J., Ertepınar H. An Investigation of Elementary Students' Scientific Literacy Levels // Journal of Baltic Science Education. – 2010. – Vol. 9. – P. 6-19. – URL: <http://oaji.net/articles/2014/987-1404740965.pdf>
6. Pella M.O., O'Hearn G.T., Gale C.W. Referents to Scientific Literacy // Journal of Research in Science Teaching. – 1966. – Vol. 4. – P. 199-208.
7. Miller J.D. Scientific Literacy: A Conceptual and Empirical Review // Daedalus. – 1983. – Vol. 112, № 2. – P. 29-48. – URL: <https://www.jstor.org/stable/20024852>
8. Sampath Kumar B.T., Kumar G.T. Perception and Usage of E-resources and the Internet by Indian Academics // The Electronic Library. – 2010. – Vol. 28, № 1. – P. 137-156. DOI: 10.1108/02640471011023432.
9. Naqvi T.H. Use of Electronic Databases by Post-graduate Students and Research Scholars at GBPAT Library, India // Library Philosophy and Practice (e-journal). – 2012. – 809. – URL: <https://digitalcommons.unl.edu/libphilprac/809>
10. Lee J.Y., Paik W., Joo S. Information Resource Selection of Undergraduate Students in Academic Search Tasks. Information Research // An International Electronic Journal. – 2012. – Vol. 17, № 1. – URL: <https://eric.ed.gov/?id=EJ971948>
11. Madge O.-L. Assessing the training needs of users in accessing electronic resources // Qualitative and Quantitative Methods in Libraries. – 2017. – Vol. 4, № 4. – P. 973-981. URL: <http://www.qqml.net/index.php/qqml/article/view/296>
12. Soni N., Gupta K., Shrivastava J. Awareness and Usage of Electronic Resources among LIS Scholars of Jiwaji University, Gwalior: A Survey // DESIDOC Journal of Library and Information Technology. – 2018. – Vol. 38. – P. 56-62. DOI: 10.14429/djlit.38.1.11524.
13. Idiegbeyan-ose J., Ifijeh G., Aregbesola A., Emmanuel O., Botu T. E Resources vs Prints Usages and Preferences by Undergraduates in a Private University Nigeria // DESIDOC Journal of Library and Information Technology. – 2015. – Vol. 39. – P. 125-130. DOI: 10.14429/djlit.39.2.13885.
14. Acheampong L., Mingle N., Smart P., Osei-Kofi P., Bekoe S. Investigating Awareness and usage of Electronic Resources by Research Scientists in Ghana // DESIDOC Journal of Library and Information Technology. – 2020. – Vol. 40. – P. 204-210. DOI: 10.14429/djlit.40.4.14906.
15. Ameen K. Graduate Researchers' Perceptions and Expectations: An Exploratory Study about Reference and Information Services // Reference Services Review. – 2019. – Vol. 48, № 2. – P. 227-242. DOI: 10.1108/RSR-02-2019-0009.

16. Fry L. Student Attitudes Towards Library Usage and Sources at a Turkish University // IFLA Journal. – 2016. – Vol. 42(2). – P. 126–133. DOI: 10.1177/0340035216645370.
17. Coil D., Wenderoth M.P., Cunningham M., Dirks C. Teaching the Process of Science: Faculty Perceptions and an Effective Methodology // CBE Life Sci Educ. – 2010. – Winter, 9(4). – P. 524-35. DOI: 10.1187/cbe.10-01-0005.
18. Shanahan C.M. Transforming Information Search and Evaluation Practices of Undergraduate Students // International Journal of Medical Informatics. – 2008. – Vol. 77, № 8. – P. 518-526. DOI: 10.1016/j.ijmedinf.2007.10.004.
19. Salubi O., Ondari-Okemwa E., Nekhwevha F. Utilisation of Library Information Resources among Generation Z Students: Facts and Fiction // Publications. – 2018. – Vol. 6, №2. – P. 16. DOI: 10.3390/publications6020016.
20. Byerley S.L., Beth Chambers M., Thohira M. Accessibility of web-based library databases: the vendors' perspectives in 2007 // Library Hi Tech. – 2007. – Vol. 25, № 4. – P. 509-527. DOI: 10.1108/07378830710840473.
21. Повинич А.Ю. Дефициты подготовки молодых исследователей к работе с зарубежными базами данных // Актуальные тенденции и инновации в развитии российской науки: сборник научных статей. Том Часть X. – Москва: Изд-во «Перо», 2021. – С. 56-61.

Материал поступил в редакцию 05.08.23.

Сведения об авторе

ТРИЩЕНКО Наталия Дмитриевна – кандидат филологических наук, старший научный сотрудник факультета журналистики Московского государственного университета имени М.В. Ломоносова; старший научный сотрудник ГПНТБ СО РАН
e-mail: trishchenko.nataliia@yandex.ru

ДОКУМЕНТАЛЬНЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ

УДК [001+002.6]:311.311:004.9

Н.Б. Баканова, Д.В. Волчков

ФУНКЦИОНАЛЬНАЯ ПОДДЕРЖКА ИНФОРМАЦИОННЫХ СЕРВИСОВ АНАЛИЗА ПУБЛИКАЦИОННОЙ АКТИВНОСТИ СОТРУДНИКОВ НАУЧНОЙ ОРГАНИЗАЦИИ*

Рассматриваются вопросы разработки программных сервисов, обеспечивающих автоматизированный анализ качественных показателей публикационной активности сотрудников научной организации. Предмет исследования – подходы и методы, позволяющие реализовать взаимодействие информационных систем организации с внешними библиометрическими базами данных. Извлеченная информация может использоваться как в процессах внутренней деятельности организации для накопления данных по её публикационной результативности, так и при формировании отчетов по выполняемым проектам для подготовки документов на семинары, конференции, конкурсы. Показаны возможности взаимодействия информационной системы организации с базами данных: «Российский индекс научного цитирования (РИНЦ)», «Web of Science (WoS)», «Scopus».

Ключевые слова: библиометрические базы данных, наукометрические показатели, публикационная результативность, информационное взаимодействие

DOI: 10.36535/0548-0019-2023-11-4

ВВЕДЕНИЕ

Современные процессы информатизации организационного управления, развитие средств телекоммуникации, создание новых возможностей автоматизированной обработки документальной информации определили появление новых задач и разработки новых функций в системах информационной поддержки управленческой деятельности.

Новые функции предназначены для реализации современных требований поддержки управленческой деятельности, включая: оперативную обработку больших объемов информации, многокритериальный анализ и оценку результатов работы, формирование различных видов отчетности, информационное взаимодействие с вышестоящими и подведомственными организациями, поддержку принятия управленческих решений в случае возникновения проблемных ситуаций. Для научной организации важнейшие из них: анализ и оценка результатов научной деятельности,

формирование планов развития научных направлений, необходимость участия в информационном взаимодействии для оперативного обмена научной информацией [1].

Анализ и оценка результатов научной деятельности базируются на современных исследованиях в области научного направления «Науковедение», составной частью которого является «наукометрия». Термин «наукометрия» изначально рассматривался в плане глобальных задач управления наукой, однако в процессе развития исследований появилась возможность его практического использования для оценки эффективности научных направлений, результативности научной деятельности, анализа перспективности конкретных исследований, проводимых организацией. В рамках развития практических приложений наукометрии стали создаваться базы данных, накапливающие реферативную информацию о публикациях и позволяющие на их основе вычислять научометрические показатели. Наиболее известные из них: Российский индекс научного цитирования (РИНЦ), Web of Science (WoS), Scopus [2, 3].

В настоящее время на основе научометрических показателей осуществляется анализ научно-технического уровня организаций, оценка выполнения науч-

* Статья подготовлена по Государственному заданию: FFMN-2022-0005 «Теоретические и прикладные проблемы информационных технологий, реалистичной компьютерной графики, визуальной аналитики и обработки многомерных данных».

но-исследовательских работ, принимаются решения о начале или продолжении финансирования исследований по конкретной тематике. Кроме того, научометрические показатели позволяют оценивать научный потенциал организаций, учитываются при анализе конкурсных заявок, присуждении научных премий, определении победителей грантов [1, 4].

На основе научометрических показателей регулярно формируются рейтинги периодических изданий, которые показывают важность издания для научного направления, позволяют оценивать и сравнивать ведущие мировые журналы, обеспечивают авторам возможность выбирать издания для публикации результатов своих исследований [4, 5].

Необходимость использования научометрических показателей при подготовке отчетов о научной деятельности потребовала разработки автоматизированных систем. Функции таких систем предусматривают мониторинг процессов научной деятельности, накопление данных о результатах научных работ, формирование научометрических показателей, определяющих оценки публикационной активности отдельных сотрудников, подразделений или отдельных направлений исследовательских работ организации.

Данные автоматизированного анализа научометрических показателей позволяют, на уровне управления наукой, оперативно расширять или ограничивать развитие научных направлений, планировать проведение актуальных исследований, поддерживать перспективные направления. В целом, можно сказать, что данные научометрического анализа дают возможность обоснованно планировать и корректировать развитие и финансирование научных проектов [3, 6].

Информационные системы организации могут существенно отличаться от полнофункциональных научометрических систем. В общем случае для решения задач организации достаточно разработки собственных функциональных модулей, которые могут выполнять требуемый для внешних отчетов анализ показателей и, в то же время, решать задачи, учитывающие специфику организации.

Настоящая статья посвящена проблеме расширения функциональных возможностей одного из направлений информационной поддержки управленческой деятельности научной организации (учет публикационной результативности), которое обеспечивает сбор и обработку данных о публикационной активности сотрудников, в соответствии с научными направлениями организации.

ПОСТАНОВКА ЗАДАЧИ

В представленном исследовании рассматривается задача расширения функционала системы «Учет публикационной результативности» (далее Система) путем включения режимов взаимодействия с библиометрическими базами данных и создания сервисов оперативного анализа научометрических показателей по результатам опубликованных научных работ сотрудников организации. Сервисы предполагается использовать для оценки выполнения научно-исследовательских работ, а также для принятия решений о начале или продолжении финансирования исследований по тематикам организации.

Принципиальный подход к решению задачи показан на рис. 1 и в табл. 1. Создаваемые программные сервисы должны обеспечивать:

- взаимодействие с внешними библиометрическими базами данных для сбора информации о публикациях сотрудников научной организации;
- накопление научометрических данных о публикациях сотрудников научной организации в базе организации;
- обработку данных для вычисления публикационной активности сотрудников по научным направлениям данной организации.

Для решения поставленной задачи необходимо разработать и подключить к Системе ряд функциональных модулей, которые смогут формировать научометрические показатели и выполнять задачи анализа данных, необходимые для оценки и корректировки направлений исследований организации.

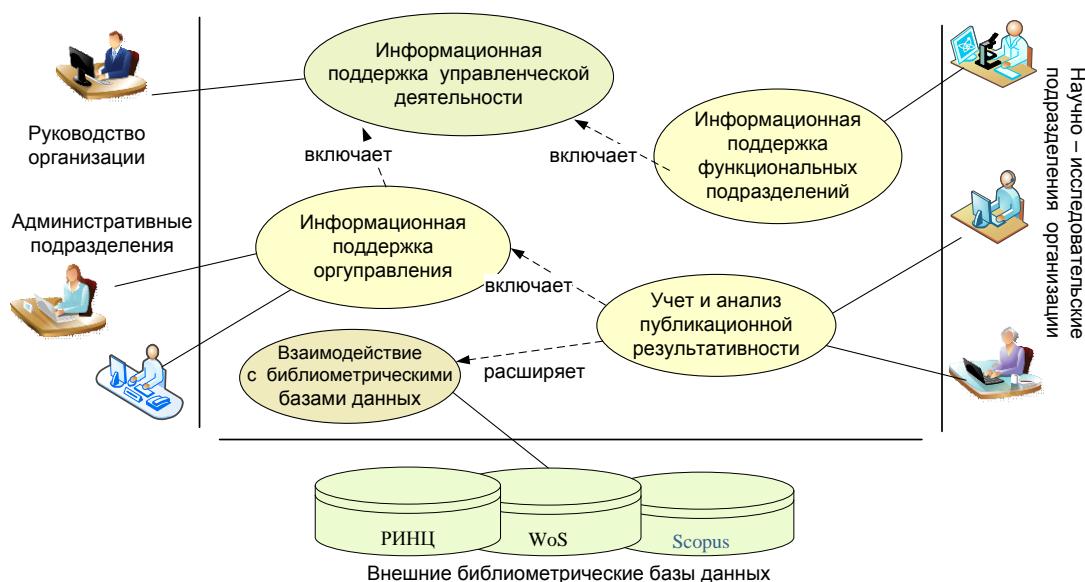


Рис. 1. Функциональная схема информационного комплекса организации

Перечень возможных форматов выгрузки библиометрических баз данных

Библиометрические базы данных	Форматы выгрузки
Web of Science (WoS) – политеатральная реферативно-библиографическая и научометрическая база данных, разработанная Институтом научной информации США (Institute of Scientific Information).	BibTeX, EndNote, CSV
Scopus – библиографическая и реферативная база данных и инструмент для отслеживания цитируемости статей, опубликованных в научных изданиях. База данных индексирует научные журналы, материалы конференций и серийные книжные издания, а также "профессиональные" журналы (Trade Journals).	BibTeX, RiS, CSV, HTML
Российский индекс научного цитирования (РИНЦ) – библиографическая база данных научных публикаций по российской научной периодике и индекс цитирования научных статей. Проект РИНЦ разрабатывается с 2005 г. компанией «Научная электронная библиотека» (eLibrary.ru).	XML, HTML

Иллюстрация включения в информационный комплекс научной организации разработанного сервиса приведена на рис. 1. В центре функциональной схемы представлены информационные системы, обеспечивающие основные функции поддержки управляемой деятельности научной организации, которые содержат информационную поддержку организационного управления, планирование деятельности и подготовку отчетов. На этом рисунке они подключены к основному программному комплексу пунктирными стрелками с надписью «включает».

Система «Учет и анализ публикационной результативности» входит как составная часть в комплекс систем поддержки организационного управления (пунктирная стрелка с надписью «включает»). Расширением этой системы является разработанный сервис, обеспечивающий взаимодействие информационной системы организации с внешними библиометрическими базами данных. Связь этого сервиса с системой показана пунктирной стрелкой с надписью «расширяет».

Важной задачей представленной разработки является реализация информационного взаимодействия информационной системы организации с внешними библиометрическими базами данных, которые содержат сведения об индексировании и статусе изданий, где размещены публикации авторов организации [7]. Чтобы достичь максимально возможного извлечения данных для наполнения информационной базы организации, требуется исследование существующих форматов выгрузки различных библиометрических баз данных (табл. 1).

При разработке алгоритмов анализа публикационной активности сотрудников, подразделений или отдельных направлений исследовательских работ организации требуется возможность оперировать следующими элементами записей о публикациях:

- название публикации,
- краткое и полное написание имен авторов,

- год публикации, номер издания, страницы, название издания и издательства, DOI публикации,
- идентификационный номер – ISSN (для газет, журналов и любых категорий периодических изданий на любом носителе),
- международный стандартный книжный номер – ISBN (выдаётся регистрационным агентством),
- язык и тип публикации,
- аннотация, ключевые слова,
- адреса авторов,
- данные о поддержке публикации (гранты, фонды, организации и др.),
- данные об аффилиации авторов.

Подходы к решению задачи

Для анализа форматов выгрузки, которые можно использовать при разработке программных средств конвертирования данных, были исследованы следующие ресурсы сети Интернет:

<https://ru.wikipedia.org/wiki/BibTeX>,
<https://ru.wikipedia.org/wiki/EndNote>,
<https://ru.wikipedia.org/wiki/CSV>,
<https://tallanto.com/ru/chto-takoe-csv-fayl-i-kak-ego-preobrazovat-v-excel>,
<https://www.azfiles.ru/extension/ris.html>,
<https://ru.wikipedia.org/wiki/HTML>.

Проведенный анализ (табл. 2) показал возможность выгрузки информации из библиометрических баз данных для решения поставленных задач без использования процедур ручного ввода информации.

Выгрузка данных из системы Web of Science осуществляется по поисковому запросу, в котором следует указать возможные варианты написаний названия организации. В запросе можно указать базы данных для выборки (Web of Science Core Collection, RSCI и др.), перечислить необходимые типы публикаций, выбрать период выгрузки. Результаты выполнения запроса можно выгрузить в любой из распространенных форматов, в том числе BibTeX, EndNote, CSV. На рис. 2 приведен фрагмент данных записи о публикации в формате EndNote.

Результаты анализа существующих форматов выгрузки данных

№ п/п	Название формата	Характеристика	Пример
1	BibTeX	Программное обеспечение для создания форматированных списков библиографии. Используется совместно с LaTeX. BibTeX использует bib-файлы специального текстового формата	@ARTICLE{tag, author = {Список авторов}, title = {Название публикации}, year = {Год}, journal = {Название издания} }
2	EndNote	Коммерческая система управления библиографической информацией, применяемая для управления ссылками и библиографией. В EndNote предусмотрена возможность выгрузки списка библиографии в текстовый файл. Каждая запись начинается с двухсимвольного кода PT, а заканчивается кодом ER	PT<Тип публикации> AU<Список авторов> TI<Название публикации> SO<Название издания> PY <Год> ER
3	CSV (Comma-SeparatedValues – значения, разделённые запятыми)	Текстовый формат, предназначенный для представления табличных данных. Каждая строка – это отдельная строка таблицы, столбцы которой разделены специальными символами – разделителями (например, точкой с запятой)	AU; TI; SO; PY Список авторов; Название публикации; Название издания; Год
4	RIS	Формат файла, разработанный Research Information Systems (RIS) для использования библиографических цитат. Хранится в текстовом формате ASCII. Информация отображается в строках, состоящих из двухсимвольного кода и соответствующего значения	TY -<Тип публикации> AU -<Список авторов> TI -<Название публикации> C3 -<Название издания> PY -<Год>ER -
5	HTML HyperText Markup Language	Язык гипертекстовой разметки – стандартизованный язык гипертекстовой разметки документов для просмотра веб-страниц в браузере	

PT J
 AU Markov, MB
 Uskov, RV
 Zhukovskiy, ME
 AF Markov, Mikhail B.
 Uskov, Roman, V
 Zhukovskiy, Mikhail E.
 TI MONTE CARLO MODELING OF THE PHOTON-ELECTRON CASCADE IN HETEROGENEOUS MATTER
 SO MATHEMATICA MONTISNIGRI
 LA English
 DT Article
 DE Photon-electron cascade processes; Hybrid parallelization; Monte Carlo weight algorithm
 AB Algorithms of statistical simulating the processes of the photon-electron cascade transport in objects using heterogeneous computers is worked out.
 C1 [Markov, Mikhail B.; Uskov, Roman, V; Zhukovskiy, Mikhail E.] RAS, KIAM, Miusskaya Sq 4, Moscow 125047, Russia.

Рис. 2. Фрагмент данных записи о публикации в формате EndNote

Авторы	Название	Год	Название исп.-Tom	Выпуск	Статья №	Страница	Страница DOI	ISSN	ISBN	CODEN	Тип документа	EID
Phillips H. Global data on earthwi	2021 Scientific Data	8	1	136			10.1038/s41597- 20524463			Data Paper	2-s2.0-85106756923	
Chetverus Computational multisc	2021 Journal of Cor	440		110352			10.1016/j.jcp.20	219991		JCTPA	Article	2-s2.0-85107297164
Ovchinnik Magnetic attitude conf	2021 Acta Astronau	186			203		210 10.1016/j.actaa	945765		AASTC	Article	2-s2.0-85107152678
Marenkov Assessment of laser in	2021 Nuclear Mater	28		101029			10.1016/j.nmre.2	23521791			Article	2-s2.0-8510690855
Shevchen Photometry of selecte	2021 Planetary and	202		105248			10.1016/j.pss.20	320633		PLSSA	Article	2-s2.0-85106345373
Zipunova Regularization and the	2021 Journal of Cor	53		101376			10.1016/j.jocs.21	18777503			Article	2-s2.0-85105079861
Khanina L Value of large Quercus	2021 Forest Ecolog	491		119172			10.1016/j.forecc	3781127		FECMD	Article	2-s2.0-85104992159
Micheli M Optical observations o	2021 Acta Astronau	184			251		10.1016/j.actaa	945765		AASTC	Article	2-s2.0-85104577247
Chetverus Numerical solution of	2021 Journal of Cor	390		113374			10.1016/j.cam.2	3770427			Article	2-s2.0-85099600430

Рис. 3. Фрагмент выгрузки списка публикаций в формате CSV

```

<item id="50382984"><linkurl>https://elibrary.ru/item.asp?id=50382984</linkurl>
<genre>статья в журнале</genre><type>научная статья</type>
<source><journal><title>Открытые системы. СУБД</title><issn>1028-7493</issn>
</journal></source>
<publisher>Издательство "Открытые системы"</publisher>
<issue><year>2023</year><number>1</number></issue>
<pages>46</pages><language>RU</language><cited>0</cited><titles>
<title lang="RU">PostgreSQL изнутри и снаружи</title></titles>
<doi>10.51793/OS.2023.93.28.001</doi><edn>VDOQIV</edn><authors><author num="1" lang="RU">
<lastame>Волков</lastame><initials>Дмитрий</initials><authorid>14747</authorid>
<affiliations><affiliation num="1" lang="RU"><orgname>ИПМ им. М. В. Келдыша РАН</orgname>
<orgid>375</orgid></affiliation></affiliations></author></authors><abstracts>
<abstract lang="RU">В помощь специалистам, хорошо знакомым с устройством реляционных СУБД.</abstract></abstracts><keywords><keyword>PostgreSQL</keyword>
<keyword>Академия ОС</keyword></keywords></item>

```

Рис. 4. Фрагмент данных записи о публикации в системе eLibrary

Каждая запись включает в себя следующие основные данные о публикации: краткое (AU) и полное (AF) написание имен авторов, название публикации (TI), название издания (SO) и серии (SE) (при наличии), ISSN (SN), ISBN (SB), язык публикации (LA), ее тип (DT), ключевые слова (DE), аннотацию (AB), адреса авторов (C1), название издательства (PU), год публикации (PY), том (VL), номер издания (IS), номера первой (BP) и последней (EP) страниц, DOI публикации (DI) и ее уникальный идентификатор WoS (UT), данные о поддержке публикации (гранты, фонды, организации и др.) (FU), а также данные о категориях WoS (WC) и областях науки (SC), к которым она относится. На основании данных поля C1 можно определить аффилиации авторов.

Выгрузка данных из системы Scopus – для этого необходимо выбрать название организации из списка. В разделе «Документы, все учреждение» на экране отобразится список всех публикаций организации, включая филиалы. Такой список можно модифицировать, применяя различные фильтры, в том числе по дате, типу публикации и др. Полученный список публикаций можно выгрузить в различные форматы, в том числе BibTeX, RIS, CSV, HTML, предварительно выбрав необходимые поля.

Набор полей в списке выгрузки из Scopus совпадает с набором полей в Web of Science. Вместо уникального идентификатора публикации WoS используется уникальный идентификатор Scopus (EID). На рис. 3 приведен фрагмент выгрузки списка публикаций в формате CSV.

Выгрузка данных из системы eLibrary – в этой системе реализована функция экспорта списка публикаций организации в формате XML, доступная в разделе «Инструменты» карточки организации. Возможен подбор публикаций по годам, типу, индексом издания (ВАК, РИНЦ, Web of Science, RSCI, Scopus), тематикам и др. На рис. 4 приведен фрагмент данных записи о публикации.

В качестве уникального идентификатора публикации в РИНЦ используются eLibrary ID и EDN. В остальном набор полей в структуре xml файла РИНЦ аналогичен Web of Science и Scopus, однако отсутствуют данные о поддержке публикации.

Другой способ получения данных из РИНЦ – это анализ списка публикаций организации в формате HTML, который открывается при выборе показателя «Число публикаций в РИНЦ» на странице карточки организации. Однако в HTML-коде далеко не все элементы данных о публикации расположены в отдельных тегах.

Сами теги не имеют идентификаторов, по которым можно однозначно определить, какого рода информация в них содержится. Для импорта данных необходимо разработать парсер кода HTML, который позволит по расположению элемента данных в структуре страницы определить его тип. Кроме того, на странице списка публикаций отсутствуют очень важные сведения о публикации, такие как DOI, ISSN/ISBN издания, аффилиации авторов, ключевые слова, тип публикации, области науки, информация о финансовой поддержке. Для получения этой информации необходимо переходить к карточке каждой отдельной публикации и анализировать ее исходный код.

К достоинствам системы eLibrary можно отнести наличие на странице организации большого количества различных отчетных показателей, таких как: КБПР с разбивкой по годам, индекс Хирша, данные о цитировании, публикации по областям знания, показатели по годам, показатели за 5 лет с разбивкой по категориям, включая количество патентных документов, монографий, число публикаций с участием зарубежных авторов, средневзвешенный импакт-фактор журналов, в которых были опубликованы статьи. В системе eLibrary можно получить статистические отчеты с распределением публикаций по тематике, типу, журналам, авторам, ключевым словам, числу цитирований и др.

ЗАКЛЮЧЕНИЕ

Подходы и методы информационного взаимодействия программных комплексов организационного управления с внешними библиометрическими базами данных, рассмотренные в настоящей статье, направлены на расширение функциональных возможностей информационных систем поддержки управленческой деятельности организаций.

Предлагаемые методы подключения к информационным системам учета публикационной активно-

сти режимов взаимодействия с библиометрическими базами данных и сервисы оперативного анализа наукометрических показателей значительно расширят функциональные возможности этих систем. Новый функционал позволит оперативно проводить оценку данных о выполнении научно-исследовательских работ, использовать автоматизированные режимы при анализе научно-технического уровня организации, обоснованно принимать решение о начале или продолжении финансирования исследований по конкретной тематике.

Создание удобных сервисов анализа данных публикационной активности сотрудников найдет применение во всех областях деятельности научной организации, включая проведение семинаров, подготовку и изучение перспективных направлений, составление научных отчетов об исследованиях, аттестацию сотрудников.

СПИСОК ЛИТЕРАТУРЫ

1. Арлазаров В.Л., Крапухина Н.В., Петровский А.Б., Фарсобина В.В. Модели материального стимулирования публикационной активности ученых // Искусственный интеллект и принятие решений. – 2018. – № 3. – С. 122-126. DOI:10.14357/20718594180322.
2. Цветкова В.А., Мохначева Ю.В. Библиометрические показатели, публикационная активность и публикации // Информация и инновации. – 2017. – Спец. вып. – С. 127–131.
3. Арлазаров В.Л., Брудно А.А., Кривцов В.Е., Попов А.П. Некоторые факторы модели управления наукой // Труды Института системного анализа Российской академии наук. – 2015. – Т. 65, № 2. – С. 27-32.
4. Мохначева Ю.В., Харыбина Т.Н. Научная продуктивность учреждений РАН и вузов: сравнительный библиометрический анализ // Вестник Российской академии наук. – 2011. – Т. 81, № 12. – С. 1065–1070.
5. Левин В.И. Новый подход к оценке качества научных исследований // Высшее образование в России. – 2017. – № 6. – С. 136–146.
6. Баканова Н.Б. Разработка информационных сервисов для реализации мониторинга публикационной активности сотрудников научной организации // Электросвязь. – 2022. – № 4. – С. 47-50. DOI: 10.34832/ELSV.2022.29.4.008.
7. Баканов А.С., Волчков Д.В., Баканова Н.Б. Разработка сервисов поддержки принятия управленических решений с использованием данных систем организационного управления // Информационные технологии и вычислительные системы. – 2020. – № 3. – С. 101-107. DOI 10.14357/207186322003010.

Материал поступил в редакцию 30.08.23.

Сведения об авторах

БАКАНОВА Нина Борисовна – доктор технических наук, доцент, ведущий научный сотрудник Института прикладной математики им. М.В. Келдыша Российской академии наук (ИПМ им. М.В. Келдыша РАН), Москва
e-mail: nina@keldysh.ru

ВОЛЧКОВ Дмитрий Владимирович – ведущий программист Института прикладной математики им. М.В. Келдыша Российской академии наук
e-mail: vol@keldysh.ru

УВАЖАЕМЫЕ КОЛЛЕГИ!

ВИНТИ РАН предлагает Вашему вниманию База данных (БД) ВИНТИ РАН в режиме online

База данных (БД) ВИНТИ РАН — Федеральная база отечественных и зарубежных публикаций по естественным, точным и техническим наукам. Генерируется с 1981 г., обновляется ежемесячно, пополнение составляет более 600 000 документов в год.

БД ВИНТИ РАН включает 26 тематических фрагментов, состоящих из более чем 190 разделов.

Документы БД содержат библиографию, ключевые слова, рубрики и реферат первоисточника.

На основе БД ВИНТИ пользователям доступны следующие продукты:

- online доступ к базе данных круглосуточно, без выходных;
- выполнение тематического поиска специалистом ВИНТИ по запросу заказчика;
- по заявкам предоставляются любые наборы тематических фрагментов БД ВИНТИ или их разделов на любых видах электронных носителей, или через FTP-сервер;
- для ознакомления с возможностями поиска имеется демо-версия базы данных bd.viniti.ru.

База данных ВИНТИ зарегистрирована Российским агентством по правовой охране программ для ЭВМ, баз данных и топологий интегральных микросхем (РосАПО) (Свидетельство № 960034 от 23.09.1996г.)

Подробную информацию Вы можете получить:

Адрес: 125190, Россия, Москва, ул. Усиевича, 20, ВИНТИ РАН

Телефон: 8 499-152-54-81

E-mail: feo@viniti.ru