

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 1. ОРГАНИЗАЦИЯ И МЕТОДИКА
ИНФОРМАЦИОННОЙ РАБОТЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 8

Москва 2023

ОБЩИЙ РАЗДЕЛ

УДК [002:502.22]:1

Ю.Н. Столяров

«Третий мир» и его «странные» законы. Документософское эссе

Обосновывается тезис, что концепция трёх миров (физического, ментального и мира объективированного знания), выдвинутая Карлом Раймундом Пóппером, имеет своим предметом документ как материализованную информацию. К. Б. Гельман-Виноградов предложил ввести понятия «документная среда ноосферы» и «документальная [документская] память ноосферы», подразумевая под первым – пространство, а под вторым – время, включающие всё множество существующих и вновь создаваемых документов.

А.В. Соколов помещает третий мир в ноосферу, именуя его Библиологосом. Будучи равновеликим наряду с двумя другими, мир документов обладает собственными понятиями, законами, особенностями, не сводимыми к законам как материального, так и идеального миров. Проводится мысль, что в силу всемирной цивилизационной важности проявлений третьего мира в наступивший век цифровизации именно от него зависит будущее человечества, современная наука должна выдвинуть разработку проблем документологии на передний план своего внимания.

Ключевые слова: философия, третий мир, ноосфера, документология, К.Р. Поппер, П. Отле, К.Б. Гельман-Виноградов, А.В. Соколов, Библиологос, Г.Н. Швецова-Водка

DOI: 10.36535/0548-0019-2023-08-1

Неологизм «Документософия» как философское осмысление сущности документа предложен Аркадием Васильевичем Соколовым (род. 1934 г., Ленинград) [1]. Он констатировал, что «в настоящее время мы располагаем обширными массивами эмпирического знания о всевозможных документах и активно развивающимися, хотя и размытыми документными теориями. При этом документософия никогда не рассматривалась как объект философской рефлексии. В книговедческой науке известна библиософия – философия книги..., а в документоведческих опусах философия документа – документософия – даже никогда не упоминалась». Между тем философское познание документа, по его убеждению, совершенно необходимо для решения стратегических вызовов завтрашнего дня. Принимая его замечание и на свой счёт, я уже сделал попытку несколько исправить ситуацию в статье [2] как продолжении стремления хотя бы в первом приближении с широких философских позиций по-размышлять над феноменом, именуемым Документом.

ПОСТАНОВКА ПРОБЛЕМЫ

С того момента, когда сэр Карл Раймунд Пóппер (*Karl Raimund Popper*; 28 июля 1902 – 17 сентября 1994) выдвинул концепцию трёх миров (физического, ментального и мира овеществлённого знания), мировая научная общественность восприняла её как данность, т. е. и не критиковала её, но и не развивала. В российской науке наиболее ярким и последовательным пропагандистом попперовских идей можно считать Валерия Павловича Леонова (род. 1942 г.), который подробно излагает эти идеи в своих книгах [3, с. 24-28; 4, с. 101-104]. Прежде чем сделать первый шаг на этом пути, придётся вкратце представить автора нового мировоззрения.

Наибольшую известность К. Попперу принесли труды по философии науки, а также по социальной и политической философии. Творчески переплавив в своём сознании взгляды преимущественно К. Маркса, Ф. Энгельса, К. Каутского, Э. Бернштейна, А. Эйнштейна, З. Фрейда и А. Адлера, он стал основоположником качественно нового взгляда на мир. Докторскую диссертацию он защитил по философии (на тему о методологии когнитивной психологии, 1928), но, будучи преподавателем, отлично владел и математикой, и естественными науками. С 1946 до середины 1970-х гг. был профессором логики и деканом факультета философии, логики и научного метода в Лондонской школе экономики и политических наук. За научные заслуги в 1964 г. посвящён в рыцари. В 1976 г. К. Поппер стал членом Лондонского королевского общества, в 1982 г. был награждён орденом Кавалеров Почёта и получил премию Алексиса Токвиля (1984) (из советских учёных ею был награждён всего один человек – выдающийся русский философ, писатель, социолог, публицист Александр Александрович Зиновьев (1922–2006)) и премию Киото (1992).

Идеи К. Поппера прозвучали сразу в нескольких областях: эпистемологии (как теоретика критического реализма), социологии (как теоретика открытого общества) и гносеологии. Нас в данном случае интересуют именно его поздние работы гносеологическо-

го характера – «Объективное знание: эволюционный подход» (*Objective Knowledge: An Evolutionary Approach*. 1972); со второго исправленного и дополненного издания в 1979 г. сделан русский перевод, впоследствии повторявшийся) и «Постскриптум к «Логике научного открытия»» (*Postscript to the Logic of Scientific Discovery*. – 1982. – Vol. 1). В них он выдвинул теорию существования трёх миров:

- 1) мира физических объектов и состояний;
- 2) мира психических и ментальных состояний сознания;
- 3) мира объективированного знания.

Содержание третьего мира составляют научные гипотезы, литературные произведения и другие, не зависящие от субъективного восприятия, объекты – по сути дела, все искусственно созданные объекты, т. е. вся сфера культуры, но документы среди них занимают первое и главенствующее место. Российский философ и культуролог А. В. Соколов помещает третий мир в ноосферу, именуя его Библиологосом и обнаруживая истоки его анализа в античной философии, у Анаксагора, жившего и творившего в V в. до н. э. [5, с. 42]. Библиологосом он предложил считать коллективный разум исторически устойчивой общности людей, владеющих книжной культурой, пишущих и читающих книги и непосредственно участвующих в производстве библиосферы. Классическая Библиосфера определена как суперсистема книжно-коммуникационных систем, обеспечивающая воспроизводство, сохранение и дальнейшее развитие национальной книжной культуры. Структуру библиосферы образуют социально-культурные институты (системы или отрасли книжного дела), а именно: издательское дело, полиграфическая промышленность, книжная торговля, библиотечное дело, библиографическое дело. А. В. Соколовым рассмотрена разработанная классиком философии XX в. Эрнстом Кассирером (*Ernst Cassirer*; 28 июля 1874, Бреславль, ныне Вроцлав – 13 апреля 1945, Принстон, Нью-Джерси, США) символическая концепция природы и сущности человека, которая обосновывает культурно-историческое единство «человек – книга». Согласно философии символических форм, мир книг, представляющий собой библиосферу, является для культурного человека столь же естественной и необходимой составляющей среды обитания, как пригодная для жизни экология [6]. Принимая во внимание знаковую природу культуры, Э. Кассирер обошёл вниманием единство знака и материальной составляющей произведений культуры и, таким образом, до понимания третьего мира во всей его полноте не дорос. А. В. Соколов убеждён, что перед библиотечной наукой, педагогикой и библиотечной профессией в целом стоит важная и ответственная задача — обеспечить использование многообразных ресурсов библиотек для гуманистических ноосферных преобразований современного общества [7].

Мир физических объектов взаимодействует с миром психических состояний, а тот порождает мир объективированного знания¹, который существует безот-

¹ Термин «знание» К. Поппер не дефинирует. Если считать знанием усвоенную и запомненную индивидом информа-

носителем к своим создателям. Таким образом, знание (т. е. информация), согласно К. Попперу, не зависит от познающего субъекта. Первый и второй миры связаны между собой через посредство третьего мира, который, с одной стороны, рукотворен, но, с другой стороны, сверхчеловечен, поскольку в нём сосредоточены «знания без субъекта знания» [8, с. 116-120], и они не просто хранятся в библиотечных фондах, но обладают способностью взаимодействовать между собой и порождать новое знание. Овеществлённое знание К. Поппер считал самым важным цивилизационным феноменом. Если все достижения человеческого разума по какой-либо причине исчезнут, замечал он, но библиотеки с их сокровищами материализованного человеческого разума сохранятся, то цивилизация сохранит единственный шанс возродиться.

Со ссылкой на академика Константина Владимировича Анохина (род. 1957), директора Института перспективных исследований мозга МГУ им. М.В. Ломоносова, заведующего лабораторией нейробиологии памяти НИИ нормальной физиологии им. П.К. Анохина В.П. Леонов полагает, что книга – это «полная отражённая система субъективного опыта, сформированного у организма в процессе эволюции, индивидуального развития и обучения». Книга в нас уже есть, она уже «заложена», достаточно её лишь перевести. Создание нового текста «есть не более чем предмет перевода» [9]. Идею о том, что и грамотность в человеческое сознание привнесена извне, В.П. Леонов проводит и в других своих работах, особенно таких, как «Книга как космический субъект: философско-культурологическое эссе» [10] и «Книга как врождённая программа человека» [11]. Это концептуально новый взгляд на происхождение письменности, и он коррелируется со взглядами К. Поппера. В своей новейшей монографии – «Нейрокогнитивное: опыт когнитивного исследования» (2023) – В.П. Леонов развивает идею книги как «внешнего когнитивного». Под когнитивом понимается «полная отражённая система субъективного опыта, сформированного в организме в процессе эволюции и человеческом Я» [12, с. 18-19]. Фрагмент субъективного опыта, перенесённый во внешнюю среду, есть внешний когнитив: книга, или шире – документ. Следует, по-видимому, полагать, что когнитивный элемент передаётся по наследству и вбирает весь когнитивный опыт человечества. Индивидууму остаётся только извлечь и осознать нужную информацию из глубин своего внутреннего мира. Чтение книги, т. е. внешнего когнитива, «представляет собой заполнение разрыва между внешним когнитивом и субъективным опытом читателя» [12, с. 21]. Как бы то ни было, но В.П. Леонов тоже признаёт существование мира, не вписывающе-

цию, то оно предстаёт как категория субъективного характера. Иными словами, знание присуще только субъекту, и выражение «мир объективированного знания» более корректно понимать как «мир объективированной информации», поскольку в отличие от знания информация может пониматься безотносительно к её носителю. К. Поппер не углублялся в вопрос о соотношении знания и информации, поэтому ему простительно рассматривать эти понятия как синонимы.

гося однозначно в дилемму «материя или сознание». Для исследования сущности и законов функционирования когнитива существующих наук недостаточно, требуется соединить нейронауки и гуманитарные знания.

Правильное отношение к законам третьего мира.

Для того, чтобы проникнуться «странными» законами третьего мира, надо прежде всего подойти к ним без предубеждения, но с пониманием их естественности и неизбежности. С пониманием, для начала, того, что они совершенно своеобразны, но вместе с тем вполне согласуются с понятиями и правилами как первого, так и второго миров. Более того, будучи их производным, они поддаются объяснениям на уровне логики, но логики, так сказать, неевклидовой.

Постижение законов третьего мира прежде всего требует отказаться от понятий здравого смысла, о котором Эйнштейн сказал, что он извинителен только для юнцов, и перейти на уровень мышления в границах формальной логики. Формальная логика хороша тем, что даёт непротиворечивые, но подчас совершенно неприемлемые с точки зрения здравого смысла ответы на поставленные вопросы. Так, с точки зрения житейской практики вопрос: «Могут ли быть слоны с крыльями?» будет рассмотрен как попросту нелепый и на него будет дан однозначно отрицательный ответ. Однако если его задать нормальному математику, то услышим: «Слоны не только могут быть с крыльями, но и сейчас буквально все слоны имеют множество крыльев. Это множество включает в себя ноль элементов». И с точки зрения формальной логики этот ответ следует признать единственно корректным. К тому же он вполне согласуется с житейской практикой, где и впрямь все слоны имеют нулевое количество крыльев.

СЛЕДСТВИЯ ИЗ ТЕОРИИ ТРЕТЬЕГО МИРА

Итак, в дополнение к известным двум привычным мирам – физическому (естественному, материальному) и ментальному (духовному, идеальному) К. Поппером предложено признать существование ещё одного, равноправного, равновеликого им мира – мира овеществлённого знания. Физически знания объективируются в виде документов. Как современник основоположника документологии Поля Отле (*Paul Marie Ghislain Otlet*; 23 августа 1868, Брюссель, Бельгия – 10 декабря 1944, там же) К. Поппер мог бы пользоваться термином и понятием «документ», но углубляться в эти тонкости предстоит его последователям. К. Поппер сделал более важное открытие: он обратил внимание на существование нового мира, параллельного двум известным ранее. Из нового взгляда следует, что в третьем мире существуют собственные законы и закономерности, не сводимые к давно известным философским атрибутам. В привычных понятиях первого и второго миров они кажутся странными, но в собственном мире естественны и логичны, иными быть не могут.

Новые утверждения и должны выглядеть странно, ни на что не похоже, в ином случае они представляли бы собой не более чем разновидность либо материального, либо идеального мира. Их надо выявлять и познавать, что называется, с нуля, как это имело ме-

сто в материальном и идеальном мирах. Но там такое изучение осуществляется веками и тысячелетиями, а идеологам попперовского мира предстоит навёрстывать достижения сложившегося философского взгляда на мир поистине семимильными шагами, да и они оказываются слишком медленными.

Документ как единство идеального и материального. Особенность документа состоит в том, что он представляет собой единство идеального и материального, т. е. единство противоположностей. В зависимости от того, что видится более важным в том или ином случае, на первый план выступает та или иная сторона данного единства. Это обстоятельство легко проиллюстрировать ссылками на такие безусловно авторитетные нормативно-правовые акты, как государственные стандарты. Когда стандартизация в нашей стране только вводилась (1970), документ определялся как “*средство* закрепления различным способом на специальном материале информации о фактах, событиях, явлениях объективной действительности и мыслительной деятельности человека” [13]. Вскоре (1983) было понято, что суть документа не сводится к средству, и акцент перенесли на *материальный носитель* (с содержащейся в/на нём информацией) [14]. К настоящему времени дефиниция переосмыслена ещё раз – настолько, что носитель и информация поменялись местами и на первое место выдвигается *информация*, зафиксированная на материальном носителе [15].

Но двуединство информации и носителя признавалось не всеми и не всегда. Отдельные исследователи придавали значение только одной из названных составляющих (на самом деле их значительно больше, что стало известно относительно недавно), и это приводило их к катастрофическим выводам. Так, Николай Александрович Рубакин (1862–1946) указывал на совершенно очевидный с его точки зрения обыденный факт: любая книга представляет собой не что иное как бумагу с нанесёнными на неё знаками, т. е. никаким содержанием не обладающую. В понятиях документологии он признавал материальную и знаковую составляющие, а в информационной (семантической) составляющей книге отказывал, перенося её в голову читателя. На этом методологическом убеждении базируется вся его библиопсихологическая теория, повсеместно вызывающая стойкое неприятие и соответственно отказ от её практического применения. Но среди множества его критиков не было ни одного, кто указал бы классику на глубинную ошибку его рассуждений, и подвижническая деятельность Н.А. Рубакина протекала в обстановке непонимания, неприятия вплоть до (в советской практике) отторжения и навешивания на него совершенно незаслуженных политических и идеологических ярлыков. В.П. Леонов опирается в своих построениях на авторитет Н.А. Рубакина, тоже видевшего в чтении чисто психологический феномен и полагавшего, что собственного содержания у книги нет, что у неё столько содержаний, сколько читателей, поскольку каждый вкладывает в читаемый текст содержание из своего жизненного опыта.

Однако основоположник кибернетики Норберт Винер (1894–1964), как известно, утверждал, что ин-

формация – это не материя и не энергия, что информация – это информация. Н. Винер прав и неправ одновременно. Прав в том, что информацию нельзя сводить ни к веществу, ни к энергии. Неправ, во-первых, в противопоставлении материи и энергии: энергия суть тоже форма существования материи; во-вторых, на уровне проявления информация нуждается в материальной форме, вне этой формы её обнаружить невозможно. Но, будучи облечена в материальную форму (вещественный или энергийный носитель и знаковое выражение), информация превращается в документ. И вот это качественное видоизменение становится камнем преткновения для многих информатиков. Например, тот факт, что информация *всегда* проявляется в том или ином материальном виде и знаковой форме, привёл наших современников А.В. Соколова и его приверженца Т.Ф. Берестову (1948–2022) к рассмотрению материально-знаковой формы как имманентно сопутствующей информации. Но, поскольку эта форма явно противостоит сущности информации, а вместе с ней и здравому смыслу, ими делается попытка снять противоречие объявлением существования неких «информационных очков», наличия у информации амбивалентной сущности [16].

Такие утверждения можно рассматривать как научный нонсенс. Но он имеет далеко идущие практические последствия: разработку так называемого информационного ресурсоведения [17–19], которое, в свою очередь, кладётся в основу государственного библиографического стандарта нового поколения. Теоретическая несостоятельность этого явления рассмотрена отдельно [20] и в данном случае опускается.

Где границы Документа с точки зрения его информационной и материальной составляющих? В качестве равноценных документов могут рассматриваться и по отдельности, и в совокупности как, например, CD-ROM (иной соответствующий гаджет), так и отображение его содержания на экране компьютера, так и самый этот компьютер и т. д. вплоть до всей структуры, а затем и инфраструктуры Интернета. В самом деле, взятый сам по себе диск с записанной на нём информацией можно уподобить книге как автономному и самодостаточному Документу. Однако для того, чтобы воспринять содержание этого диска, его нужно вставить в дисковод компьютера и обратиться к экрану, которые, таким образом, являются естественным продолжением CD-ROMа, или флеш-памяти. Таким образом, компьютер в целом правомерно рассматривать как полноценную многокомпонентную материальную оболочку для электронного документа. Поскольку её составной частью является источник электропитания, то и этот источник, вместе с системой передачи электричества, корректно рассматривать как часть Документа. К слову, широко известное слово «ноутбук» – не что иное как калька с английского «note book» – «записная книжка, тетрадь, блокнот». Иными словами, ноутбук весь, целиком, есть единый Документ. И в конечном счёте можно прийти к выводу, что всю систему Интернета вместе с её инфраструктурой правомерно считать одним гигантским Документом, наподобие распределённой мегаэнциклопедии, в котором сосредоточена

вся электронная информация. Создание именно такой всеобъемлющей книги предрекал в своё время Поль Отле, когда утверждал, что «тип документа будущего ... будет синтез наиболее эффективных приобретённых характеристик, присоединённых к тем окончательным, которые должны быть ему предназначены» [21, с. 303]. Библиология, отмечал он, «уже рассматривает все Книги как одну безостановочно увеличивающуюся Книгу» [21, с. 303].

Информация по тому или иному вопросу может быть сосредоточена в одном или множестве документов, но при этом она может быть рассредоточена в пространстве настолько причудливым образом, что вместо определения её истинного местонахождения в каждый данный момент времени легче признать, что она находится везде одновременно и при этом нигде конкретно, – вот и ещё одна странность, свойственная современному третьему миру. Облачные технологии постоянно расширяют возможность распыления информации, вместе с тем обеспечивая её мгновенный сбор в затребованном месте.

В абстрактном смысле о мире Документов можно сказать, что по содержанию и форме он безграничен как в сторону увеличения, так и в сторону уменьшения – вплоть до того, что он одновременно бесконечно содержателен (вбирает в себя содержание буквально всех документов), но одновременно и бессодержателен (поскольку по принятому в логике определению объём понятия находится в обратной зависимости от содержания). При этом названные характеристики постоянно прирастают, хотя, казалось бы, ни бесконечности, ни пустоте прирастать или уменьшаться некуда. Впрочем, и в первом мире Вселенная бесконечна, однако существует теория её непрерывного расширения.

С точки зрения содержания, мир документов (или – на усмотрение наблюдателя – всего лишь единственный документ, но вбирающий в себя содержание всех реальных и потенциальных документов) бесконечен потому, что в этих документах зафиксированы все знания (или вся информация; в данном случае это не суть важно) человечества по всем вопросам. Документный потенциал человечества равен его духовному потенциалу, выраженному на материальном носителе в знаковой форме. И по мере существования человечества этот документный потенциал непрерывно возрастает. Документировать (т. е. превращать в документ) можно абсолютно любую информацию – как по содержанию, так и по форме, и величине, и времени, к которому она относится. На материал записи, знаковую систему, средства и способы записи принципиальных ограничений не существует. Это означает, что любая информация может быть записана на каком угодно носителе и каким угодно способом. И коль скоро в третьем мире все ограничения сняты, допустимо представить, что информация о чём угодно, причём о его прошлом, настоящем и будущем, теоретически может быть как записана на чём угодно какого угодно размера, так и извлечена из чего угодно как сколь большого, так и сколь малого размера. Вопрос состоит только в обнаружении способов такой записи или извлечения информации. Пока человечество не то чтобы

стоит на пороге создания соответствующих технологий. Оно стоит всего лишь на пороге осознания, для начала теоретического, такой возможности и её признания. При всём разнообразии и непрерывном появлении всё новых носителей, знаковых систем, средств и способов записи информации человечество ещё не освоило и ничтожной доли своих возможностей такого рода, хотя уверенно и успешно в этом направлении движется.

Итак, мы вышли на первейшую странность: обнаруживается, что помимо двух известных измерений (материи и сознания) мир имеет и третье – документское измерение. И мир документов существует в одно и то же время и отдельно от двух других миров, и параллельно с ними. Это вторая странность. Мир документов существует одновременно как внутри (в виде усвоенной из них информации), так и вне нас – и это третья странность.

Мир документов правомерно рассматривать как конкретное воплощение ноосферы – скорее всего, далеко не единственное, но, по крайней мере, на сегодняшний день отчётливо видимое. На связь понятия «документ» с понятием «нооинформация» я обратил внимание ещё в 1981 г. [22, с. 69]. Десятилетием позже связь документа и ноосферы обнаружил известный специалист в области документоведения и архивоведения, а ранее документалист К.Б. Гельман-Виноградов (1925–2010). Ещё в 1992 г., т. е. более 30 лет тому назад, он опубликовал статью «Документальная память ноосферы как новый объект познания (к постановке проблемы)» [23]. По его словам, П. Тейяр де Шарден (1881–1955) заметил, что ещё древняя китайская цивилизация выражала себя непосредственно в письме. Устную и письменную передачу мысли французский философ считал искусственным дополнением к естественным формам наследственности индивида как члена общества. Зарождение ноосферы В. И. Вернадский тоже связал с возникновением письменности, т. е. отнёс за 5-7 тысяч лет до нашего времени. Со ссылкой на В.И. Вернадского К.Б. Гельман-Виноградов утверждал, что ноосфера выходит за рамки естественнонаучного истолкования. Это вполне согласуется с концепцией К. Поппера.

К схожим выводам пришёл А.В. Соколов, когда начал углублённо заниматься проблемой документосферы [24]. Десятью годами раньше этим вопросом озаботился Р.С. Мотульский [25].

Таким образом, человечество уже давно существует в ноосфере, и наивно полагать, будто она возникнет в неопределённом будущем и представит собой однозначно позитивное явление. Г.Н. Швецова-Водка (1943–2023) обоснованно сомневалась в однозначно положительной оценке большинства концепций ноосферы: «ведь мыслящая оболочка Земли формировалась из самых различных идей и стремлений. Одни из них – прогрессивные – способствуют человеколюбию, миролюбию, милосердию, бережному отношению к природе, накоплению естественнонаучных и социологических знаний, другие поощряют экстремизм, воинственность, нетерпимость, алчность, потребительское отношение к природе, суеверия. Среди результатов практической, якобы “разумной”, деятельности людей – эксплуатация человека человеком,

разрушение природной среды, войны, убийства» [26, с. 101]. Всё это откладывается в документном ресурсе. Поставленную Г. Н. Швецово-Водкой альтернативу «ноосфера – это либо реальность, с которой приходится жить человечеству, либо далёкий от действительности идеал, прекрасное будущее, утопия» [26, с. 101], есть основания решать в пользу реальности, сколь бы противоречивой и мало вдохновляющей она ни была.

Заметим также, что в отличие от бесконечно продолжительных по времени и пространству существования первого и второго миров история третьего мира ограничена и конечна. Он существует (пока что?) в границах Земли и ближнего космического пространства, по моменту возникновения совпадает со временем становления Homo legens (человека читающего) и находит реальное воплощение в материализации мысли. Можно согласиться с К.Б. Гельманом-Виноградовым, предложившим ввести понятия «документная среда ноосферы» и «документальная [документская] память ноосферы», подразумевая под первым – пространство, а под вторым – время, вбирающие в себя всё множество существующих и вновь создаваемых документов.

Думается, изложенного в настоящей статье для начала, для первого приближения к осознанию новой реальности достаточно, чтобы привыкнуть к этим странностям, смириться с их объективным существованием, признать, что человечество давно развивается в границах третьего мира, что третий мир – синоним ноосферы. Всемирная цивилизационная важность проявлений третьего мира в наступивший век цифровизации состоит в том, что будущее человечества зависит именно от проявлений третьего мира. Поэтому современная наука должна выдвинуть разработку проблем документологии на передний план своего внимания.

Для того, чтобы глубже погрузиться в этот странный мир, требуется дальнейшая конкретизация заявленных тезисов.

СПИСОК ЛИТЕРАТУРЫ

- Соколов А.В. Документ как предмет научного познания // Научные и технические библиотеки. – 2021. – № 8. – С. 32-33. DOI: <https://doi.org/10.33186/1027-3689-2021-8-13-38>.
- Столяров Ю.Н. Документософия как философское осмысление сущности документа // Научные и технические библиотеки. – 2022. – № 9. – С. 127-146. DOI: <https://doi.org/10.33186/1027-3689-2022-9-127-146>.
- Леонов В.П. Пространство библиотеки. Библиотечная симфония. – Москва : Наука, 2003. – 121 с.
- Леонов В.П. Besame mucho. Путешествие в мир книги, библиографии и библиофильства. – Москва : Наука, 2008. – 268 с.
- Соколов А.В. Библиотеки в умных городах: к столетию концепции ноосферы // Университетская книга. – 2020. – Апрель. – С. 42-45.
- Соколов А.В. Эссе о библиотечной постсовременности // Вестник культуры и искусств. – 2020. – № 1(61). – С. 7-18.
- Соколов А.В. Ноосферный человек в библиотеке будущего // Библиотековедение. – 2019. – Т. 68, № 1. – С. 7-17. DOI: <https://doi.org/10.25281/0869-608X-2019-68-1-7-17>.
- Поппер К.Р. Объективное знание : эволюционный подход / пер. с англ. Д. Г. Лахути. – Москва : УРСС, 2002. – 381 с.
- Леонов В.П. Новые горизонты науки о книге и библиографии // Материалы Всероссийской научно-практической конференции «Библиотека и культурное пространство региона (К 45-летию кафедры библиотечных и документально-информационных технологий)» (г. Пермь, 17–18 декабря 2020 г.) / отв. ред. Е.М. Вафина. – Пермь, 2021. – С. 64.
- Леонов В.П. Книга как космический субъект: философско-культурологическое эссе // Леонов В. П. Избранное. – Санкт-Петербург : БАН, 2021. – С. 9-33.
- Леонов В.П. Книга как врождённая программа человека // Там же. – Санкт-Петербург : БАН, 2021. – С. 34-43.
- Леонов В.П. Нейрокиббегедение: опыт когнитивного исследования. – Санкт-Петербург: БАН, 2023. – 79 с.
- ГОСТ 16487-70. Делопроезводство и архивное дело. Термины и определения. – https://studopedia.ru/16_27823_opredelenie-dokumenta-v-gosudarstvennom-standarte.html.
- ГОСТ 16487-83. Делопроезводство и архивное дело. Термины и определения (Термин 3 «документ» определен как «материальный объект с информацией, закреплённой созданным человеком способом для её передачи во времени и пространстве»). – URL: <https://files.stroyinf.ru/Data2/1/4294835/4294835961.pdf>
- ГОСТ Р 7.0.8-2013. Делопроезводство и архивное дело. Термины и определения. (Термин 7 «документ» определен как «зафиксированная на носителе информация с реквизитами, позволяющими её идентифицировать»). – URL: <https://legalacts.ru/doc/gost-r-708-2013-natsionalnyi-standart-rossiiskoi-federatsii/>
- Соколов А.В. «Информационные очки» как методологический инструмент // Энергия: экономика, техника, экология. – 2007. – № 6. – С. 74-79.
- Берестова Т.Ф. Информационное ресурсоедение как новое научное направление: постановка проблемы // Научно-техническая информация. – Сер. 1. – 2015. – № 7. – С. 1–9. – URL: <http://lamb.viniti.ru/sid2/sid2free?sid2=J13633527>
- Берестова Т.Ф. Адаптивно-эволюционная концепция информации и информационных ресурсов // Библиосфера. – 2018. – № 4. – С. 11-17.
- Берестова Т.Ф. Теоретическое информационное ресурсоедение. – Челябинск : Челябинский государственный ин-т культуры, 2019. – 338 с.

20. Столяров Ю.Н. Библиотечное фондоведение и библиографоведение: точки пересечения // Библиография и книговедение. – 2022. – № 2. – С. 111-125.
21. Отле П. Библиотека, библиография, документация : избранные труды пионера информатики. – Москва : ФАИР-ПРЕСС, Пашков дом, 2004. – 349 с.
22. Столяров Ю.Н. Библиотека: структурно-функциональный подход. – Москва : Книга, 1981. – 255 с.
23. Гельман-Виноградов К.Б. Документальная память ноосферы как новый объект познания (к постановке проблемы) // Международный форум по информации и документации. – 1992. – Т. 17, № 1. – С. 8-16.
24. Соколов А.В. Документный ресурс ноосферы: отклик на выход в свет учебного пособия Ю.Н. Столярова «Документный ресурс» (Москва : Либерия-Бибинформ, 2009) // Научные и технические библиотеки. – 2010. – № 1. – С. 119-129.
25. Мотульский Р.С. Механизм формирования и распространения нооинформации // Научные и технические библиотеки. – 2000. – № 10. – С. 4-15.
26. Швецова-Водка Г.Н. Ноосферная тематика в документологии // Научные и технические библиотеки. – 2018. – № 9. – С. 92-105.

Материал поступил в редакцию 20.05.23.

Сведения об авторе

СТОЛЯРОВ Юрий Николаевич – доктор педагогических наук, профессор, главный научный сотрудник Российской государственной библиотеки, Государственной публичной научно-технической библиотеки России и Научного и издательского центра «Наука» РАН, Москва.
e-mail: yn100@narod.ru.
ORCID.org 0000-0002-9597-4275

Экосистемы цифрового бизнеса: российский опыт

На примере российских бизнес-экосистем анализируются тенденции объединения различных цифровых сервисов в комплексные экосистемы. Обзор основных экосистем позволил сделать вывод, что использование таких систем в современных условиях – важный фактор успеха развития бизнеса.

Ключевые слова: цифровая экосистема, бизнес-экосистема, цифровизация, платформизация, мультисервисная подписка, стратегия, конкуренция

DOI: 10.36535/0548-0019-2023-08-2

ВВЕДЕНИЕ

В последние годы заметна тенденция к объединению различных цифровых сервисов и продуктов, созданных и/или предоставляемых под брендом компании, в единые бизнес-экосистемы. Этому способствовала пандемия коронавируса, которая активировала процессы цифровой трансформации и кардинально повлияла на изменение подходов к формированию корпоративных систем управления. Цифровизация позволяет трансформировать сферы хозяйственной деятельности за счет внедрения новых технологий или бизнес-моделей. Ввиду того, что требования потребителей постоянно возрастают и усиливается деловая конкуренция между компаниями, отдельно взятой организации становится всё сложнее внедрять различного рода инновации. По этим причинам многие российские и зарубежные компании вынуждены активно работать над созданием бизнес-экосистем.

Концепция бизнес-системы активно используется в управленческой деятельности. В настоящее время не выработано четкого определения термина «бизнес-экосистема» и, как правило, под ней понимают взаимосвязанную сеть фирм с центральной фирмой (лидером экосистемы) и сообществом участников [1]. Бизнес-экосистемы представляют собой набор организаций, состоящих из учреждений, клиентов, поставщиков, конкурентов, дистрибьюторов, которым удаётся сотрудничать во время поставки заказанного продукта и/или услуги. В этом случае организации формируют для себя новые задачи и создают инновации за счёт сотрудничества с дополнительными компаниями [2].

Цель настоящего исследования – сравнительный обзор российских бизнес-экосистем на основе данных об их российском рынке, научных трудов отече-

ственных и зарубежных исследователей. Мы использовали эмпирические и теоретические методы: анализ, обобщение, группировка, сравнение.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

В современном бизнесе всё больше компаний стремятся применять экосистемный подход в своей стратегии развития. Характер, масштабы и влияние цифровой трансформации выходят далеко за рамки одного хозяйствующего субъекта и разворачиваются в бизнес-экосистемах, состоящих из нескольких взаимозависимых фирм [3]. Это связано с тем, что совокупность различных сервисов и направлений внутри экосистемы при формировании ее модели становится более эффективной. Объединение сервисов и услуг в комплексные предложения позволяет сделать привлечение клиентов более выгодным: уменьшаются затраты и, следовательно, увеличивается прибыль компаний. Если фирма становится частью экосистемы, то она получает доступ к сервисам компаний-партнеров и, в этом случае, не вкладывает инвестиции в развитие непрофильных направлений. Всё это способствует оптимизации ресурсов и повышению конкурентоспособности компаний [4].

Цифровизация приводит к изменению традиционных бизнес-моделей и способствует переменам в организационных структурах управления. В связи с этим в научном сообществе цифровизация исследуется с разных точек зрения и особое внимание направлено на изучение цифровых экосистем [5-7], цифровых бизнес-экосистем [8, 9], а также цифровых платформ [10-12].

Важным фактором, способствующим ускоренному развитию экосистем, является глобальный тренд

на цифровизацию продуктов и услуг. В наше время сложно представить свою жизнь без Интернета и электронных устройств (смартфоны, планшеты, ноутбуки и т.д.), которые стали неотъемлемой частью профессиональной сферы и повседневной жизни современного общества. Портативные электронные устройства – основной способ поиска и изучения информации. Этот процесс поддерживается также развитием мобильного Интернета, который способствует сокращению цифрового неравенства между социально-экономическими группами и субъектами страны.

Согласно отчету Global Digital 2023 на начало 2023 г. пользуются мобильным Интернетом 5,44 млрд чел., что составляет 68% от общей численности населения мира. Количество уникальных пользователей мобильных телефонов за 2022 г. выросло почти на 3% (168 млн новых пользователей) [13]. В Российской Федерации на начало 2023 г. насчитывалось 127,6 млн интернет-пользователей, следовательно, проникновение Интернета составляет 88,2% [14]. Таким образом, повышение уровня мобильного Интернета содействует созданию и развитию цифровых экосистем, которые предлагают интегрированные и удобные сервисы для пользователей.

Сегодня не существует точного и универсального определения термина «экосистема». В исследовании Московской школы управления «Сколково» и компании «МТС» показано, что экосистема – это стратегия, при которой независимые в операционной деятельности, действующие на свой риск бизнес-единицы создают совместную ценность для потребителей, используя общие правила, инструменты и знания [15]. Министерство экономического развития Российской Федерации в концепции государственного регулирования цифровых платформ и экосистем определяет эти бизнес-единицы как клиентоцентричную бизнес-модель, объединяющую две и более группы продуктов, услуг, информации для удовлетворения конечных потребностей клиентов [16].

Таким образом, бизнес-экосистема – это набор сервисов, продуктов и услуг, объединенных вокруг одной компании и направленных на удовлетворение широкого спектра потребностей клиента.

Выделяют следующие стратегии развития бизнес-экосистем [16-18].

1. Развитие собственных сервисов. Так, на российском рынке первые экосистемы создавались путем организации собственных сервисов или интеграции с другими компаниями. По этому пути идут Яндекс, Сбер и МТС. Такой подход позволяет полностью контролировать сервисы в рамках экосистемы и хранить данные о её охвате.

2. Заключение партнерства. Компания заключает партнёрство с другими компаниями для построения экосистемы, что позволяет за более короткий срок охватить большее число потребностей пользователя и одновременно с этим сократить издержки на разработку собственных сервисов и продуктов или покупку других компаний. При данном направлении развития экосистемы применяется стратегия *win-win* или «эффективной взаимозависимости». Например, приложение Тинькофф-банка закрывает практически все основные потребности потребителей: онлайн-банкинг,

покупка билетов, онлайн-шопинг, заказ еды, запись к врачу и др. В результате доверие к бренду повышается, а упоминание о нём вызывает множество ассоциаций. Выгоду в этом случае получают как клиент (удобство и простота использования) и бизнес (постоянный поток клиентов), так и банк (высокий доход от комиссий и других выплат партнеров).

3. Гибридная. Экосистема компании сочетает собственные и партнерские сервисы. Эту стратегию использует компания VK (ВКонтакте), которая имеет платформу мини-приложений, однако все ключевые сервисы, такие как социальная сеть, Музыка, Видео и Клипы были либо разработаны самой компанией, либо приобретены или получены через совместное предприятие со Сбербанком. Несмотря на это, развитие финансовых сервисов осуществляется преимущественно через партнерства. Компания, применяя гибридную стратегию, стремится использовать партнерские отношения для развития финансовых сервисов, дополняя свою экосистему и предоставляя дополнительные удобства своим пользователям.

Можно выделить следующие основные признаки цифровой экосистемы:

1) ориентация компании на максимальное удовлетворение потребностей клиентов;

2) применение *data-driven* подхода, предполагающего использование информации о клиентах для оптимизации бизнес-рекламы и маркетинговых коммуникаций;

3) внедрение автоматических систем и технологий, позволяющих снижать стоимость процессов, повышать качество взаимодействия с клиентами и предлагать новые услуги или продукты, способствующие увеличению общей ценности продукта. Путем автоматизации компании могут оптимизировать свои операции, устранять ручные и трудоемкие процессы, а также повышать эффективность использования ресурсов. Кроме того, автоматизация способствует созданию более удобного и персонализированного клиентского опыта, улучшает скорость обслуживания и снижает возможность ошибок. В результате, компания может предлагать новые продукты и услуги, которые ранее были недоступны, и, тем самым, увеличивать свою ценность для клиентов;

4) адаптация и моментальное реагирование на рыночные изменения. Если экосистема не сможет следовать за динамикой рынка, то есть риск того, что пользователи начнут предпочитать более передовые платформы.

В России существует несколько цифровых экосистем, которые объединяют различные компании и сервисы, предлагая пользователям широкий спектр услуг и продуктов [19], среди них: СБЕР, Яндекс, МТС, Тинькофф, ВТБ, OZON, VK, МегФон, X5 Retail Group, Wildberries и Авито. В то же время по результатам исследования Московской школы управления «Сколково» и компании «МТС» в России назвать экосистемами можно только следующие компании: VK (до переименования – Mail.ru Group), МТС, Сбербанк и Яндекс [15, 16]. Этот вывод был сделан на основе количества сервисов, входящих в каждую экосистему, основных сегментов рынка, которые они закрывают, и количества пользователей [20]. В табл. 1 представлен

сравнительный анализ ведущих отечественных экосистем по различным сегментам.

Сильные национальные компании являются значимой особенностью российского рынка, отличающей его от многих других. В настоящее время локальный рынок подвержен воздействию различных факторов. С одной стороны, место иностранных компаний, которые покинули рынок, успешно занимают российские компании, и для большинства из них это является естественным расширением их основного бизнеса. С другой стороны, рост аудитории требует дополнительных инвестиций в инфраструктуру и ускоренного развития собственных сервисов, особенно связанного с расширением ассортимента предлагаемых продуктов и услуг [21].

Российские бизнес-экосистемы, в отличие от иностранных, чаще всего выбирают развитие по горизонтальному или комплексному пути. Именно поэтому они конкурируют во множестве различных отраслевых рынков. Сохраняется особенность собственных поисковых сервисов, социальных медиа и картографических сервисов, где присутствуют локальные компании в сфере электронной коммерции, успешно конкурирующие не только с зарубежными компаниями, но и между собой [20].

В отличие от стран Европейского союза, где регуляторы акцентируют свое внимание на антимонопольном регулировании американских интернет-гиган-

тов, в России, с присутствующей серьезной конкуренцией между цифровыми экосистемами, более актуальными становятся вопросы, связанные с защитой персональных данных и прав потребителей, а также проблемы дискриминации поставщиков. Это означает, что российский рынок сосредоточен на обеспечении конфиденциальности данных пользователей и защите их прав, а также на обеспечении равных условий для всех участников.

Российские цифровые экосистемы пока еще не полностью реализовали свой потенциал, в отличие от их зарубежных аналогов. Например, средняя чистая прибыль у российских экосистем в 60 раз ниже, чем у мировых лидеров в этой области [20]. Это указывает на то, что российским компаниям еще предстоит пройти долгий путь развития и повышения своей рентабельности.

Еще одна особенность российских экосистем заключается в том, что убыточные сервисы внутри экосистемы финансируются из основной деятельности компании. Это означает, что прибыль, получаемая от основного вида деятельности, используется для поддержания и развития сервисов, которые на данный период не приносят достаточного дохода или даже убыточны. Такой подход свидетельствует о стремлении компаний сохранить и расширить свою экосистему, несмотря на временные финансовые трудности.

Таблица 1

Конкуренция в различных сегментах ведущих российских бизнес-экосистем*

Сегменты	Экосистема			
	Сбербанка	МТС	Яндекс	VK
Цифровой контент	Okko, СберЗвук, Rambler Media	Kion, MTC Music, MTC Строки, WASD.tv	Кинопоиск, Яндекс.Музыка, поиск «Яндекса»	Mail.ru, VK, Одноклассники, Мой мир, VK Музыка, VK клипы
Финансы	Сбербанк, ЮMoney	МТС Банк, МТС Cashback	Яндекс.Инвестиции, Yandex.Pay	VK Pay
Телекоммуникации	СберМобайл	Сотовая и фиксированная связь, ТВ, интернет	Почта «Яндекса»	Почта «Mail.ru», ISQ, мессенджер
E-commerce	СберМегаМаркет	Интернет-магазин МТС	Яндекс.Маркет	VK Market
Образование	СберОбразование (Edutoria, СберКласс и Школа 21)	-	Яндекс.Практикум	SkillBox, GeekBrains, SkillFactory
Технологии	Sber Devices	CloudMTS, MTC Digital, VisionLabs	Yandex.Cloud, Yandex Self-Driving Group	Облако Mail.ru
Здоровье	СберЗдоровье, Сбер Аптека	SmartMed	Яндекс.Здоровье	Сервис «Все аптеки»
Еда	СберМаркет, Самокат	-	Яндекс.Еда, Яндекс.Лавка, Delivery Club	Самокат
Мобильность	Ситимобил, СитиДрайв	-	Яндекс.Такси, Яндекс.Драйв	Ситимобил, СитиДрайв

*Источник: составлено авторами

Мультисервисные подписки среди лидеров российских бизнес-экосистем*

Экосистема	Подписка	Количество пользователей		Доля «подписки» в пользовательской базе, %
		подписки, млн чел.	экосистемы, млн чел.	
Яндекс	Яндекс.Плюс	10	103	9,7
МТС	Premium	7,4	79,7	9,3
VK	VKcombo	4,5	90	5,0
СБЕР	СБЕРПрайм	4	103	3,9

* Источник: составлено и рассчитано авторами по данным [20, 22]

Экосистема, которая объединяет множество различных сервисов, невозможна без внедрения механизмов, способных их связывать. Один из таких механизмов – модель подписки. Мультисервисные подписки являются одним из основных инструментов удержания пользователей в цифровых экосистемах. Пользователь может получить доступ к нескольким популярным сервисам за умеренную плату и, таким образом, сэкономить на оплате каждого сервиса отдельно. В свою очередь, компания стимулирует свою лояльную аудиторию использовать и другие продукты, даже если они на данный момент не так популярны, как альтернативные предложения конкурентов. В табл. 2 приведено количество участников подписки на сервис и количество пользователей экосистемы.

Подписка удобна пользователям тем, что позволяет им удовлетворить несколько своих потребностей одновременно и, при этом, сэкономить денежные средства благодаря различным бонусам, скидкам, повышенному кэшбеку и другим привилегиям. Сама по себе подписка может быть убыточной, но ее долгосрочная ценность для экосистемы заключается в снижении затрат на привлечение аудитории к ее сервисам и увеличении LTV (прибыль компании, полученная от одного клиента за всё время сотрудничества с ним).

Российские цифровые экосистемы начали осваивать это направление совсем недавно: первой на российском рынке свою подписку запустила компания «Яндекс» в 2018 г., в 2019 г. за ней последовала Mail.ru Group (после переименования в 2021 – VK), в 2020 г. – Сбер, МТС. Остановимся на них подробнее [23]:

Яндекс Плюс. В базовой версии подписки предоставляется доступ к 24 сервисам, охватывающим различные направления, такие как медиа и развлечения, товары, питание, мобильность, офлайн-досуг, технологии, финансы и специальные предложения. Ключевые сервисы подписки – Яндекс.Музыка, КиноПоиск HD, Яндекс.Такси и Яндекс.Еда; преимущества – большое количество доступных сервисов и бесплатный доступ к популярным стриминговым ресурсам. Однако есть и некоторые недостатки, включая сложности с получением кэшбека, который зависит от нескольких условий, таких как размер покупки и участвующие товары.

СберПрайм. В базовой версии подписки предоставляется доступ к 13 сервисам, охватывающим различные направления, такие как товары, здоровье, пи-

тание, мобильность, медиа и развлечения, мобильная связь, технологии и финансы. Ключевые сервисы подписки – СберМаркет, Okko и СберЗвук; преимущества – выгодные условия использования продуктов этой экосистемы. Однако в базовой версии подписки количество бонусов для владельцев банковских карт ограничено.

VK Combo. В рамках подписки предоставляется доступ к 17 сервисам, охватывающим различные направления, включая медиа и развлечения, питание, мобильность, технологии и образование. Ключевые сервисы подписки – Музыка ВКонтакте и Delivery Club; преимущества – большое количество сервисов с доступом на выгодных условиях, низкая цена. Эта подписка единственная, в рамках которой можно воспользоваться услугами образовательных онлайн-порталов или игровых сервисов, а скидки на их использование достигают до 55%. Однако в подписку входит не самый популярный онлайн-кинотеатр.

Premium от МТС. В рамках подписки предоставляется доступ к 13 сервисам, охватывающим различные направления, включая мобильную связь, онлайн-развлечения, финансы, офлайн-досуг, технологии, здоровье и спецпредложения. Ключевой сервис подписки – мобильная связь МТС; преимущества – выгодные условия использования телеком-услуг. Однако предложения по остальным направлениям достаточно слабы.

Мультисервисные подписки, вместе с остальными типами подписок, позволяют компаниям получать регулярные и предсказуемые доходы, а также более точно прогнозировать свои будущие финансовые потоки. Однако они выполняют и еще одну важную функцию – способствуют кросс-продажам новых и малоизвестных сервисов внутри экосистемы. Такой подход позволяет создавать и укреплять привычку пользователей и перетягивать на развивающиеся сервисы ту часть аудитории, которая ранее использовала более популярные продукты конкурентов и вынуждена была приобретать отдельные подписки на каждый из них.

Появление цифровых экосистем, которые объединяют поставщиков и потребителей и позволяют им взаимодействовать, снижая издержки, – это естественный результат прогресса интернет-технологий и конкуренции на рынке. Благодаря сетевым эффектам экосистемы создают значительные преимущества для всех участников. Аналитики предсказывают значи-

тельный рост рынка экосистемных подписок с точки зрения как денежных показателей, так и количества подписчиков. Это свидетельствует о значимом вкладе экосистемных моделей в экономику и их стремительном распространении. Согласно прогнозам международной консалтинговой компании McKinsey, к 2025 г. цифровые экосистемы будут генерировать около 30% дохода организаций. Пользователи выбирают этот формат сервисов за удобство (доступ ко всем сервисам и продуктам единой экосистемы доступен по одному ID) и финансовую выгоду. Кроме того, экосистемная цифровая трансформация крупного бизнеса дает пользователю возможность получать более привлекательные условия, чем при обращении в разные компании [24].

ВЫВОДЫ

Современный бизнес очень быстро меняется, и компании, которые не следят за трендами и не адаптируют свои продукты и сервисы под изменяющиеся потребности рынка, могут быстро потерять свою конкурентоспособность. Объединение сервисов в экосистемы позволяет предлагать клиентам комплексные и удобные решения, объединяющие различные аспекты их жизни. Компании предоставляют пакетные предложения, включающие в себя широкий спектр сервисов, таких как телекоммуникационные услуги, финансовые сервисы, онлайн-покупки, развлекательный контент и многое другое. Такие экосистемы делают привлечение клиентов экономически эффективным, так как компании могут предлагать интегрированные решения, удовлетворяющие множество потребностей клиентов. Экосистема дает возможность сервисам «обмениваться» технологиями и решениями, а значит, быстрее их развивать. За счет взаимодействия сервисов становятся выше эффективность и удовлетворение запросов потребителей.

Российские цифровые экосистемы находятся на начальной стадии своего развития и сталкиваются с рядом вызовов, включая низкую рентабельность и финансирование убыточных сервисов. Однако, с учетом потенциала и роста интереса со стороны пользователей, отечественные компании имеют возможность улучшить свои экосистемы, привлечь больше клиентов и достичь конкурентоспособности на мировом уровне. Это требует стратегического подхода, инноваций и укрепления финансовой устойчивости компаний.

СПИСОК ЛИТЕРАТУРЫ

1. Liu G., Aroean L., Ko W.W. Service innovation in business ecosystem: The roles of shared goals, coopetition, and interfirm power // *International Journal of Production Economics*. – 2023. – Vol. 255. – P. 108709. DOI: <https://doi.org/10.1016/j.ijpe.2022.108709>.
2. Espina-Romero L.C., Guerrero-Alcedo J.M., Ossio C. 7 topics that business ecosystems navigate: Assessment of scientific activity and future research agenda // *Heliyon*. – 2023. – e16667. DOI: <https://doi.org/10.1016/j.heliyon.2023.e16667>.
3. Mann G., Karanasios S., Breidbach C. F. Orchestrating the digital transformation of a business ecosystem // *The Journal of Strategic Information Systems*. – 2022. – Vol. 31, Iss. 3. – P. 101733. DOI: <https://doi.org/10.1016/j.jsis.2022.101733>.
4. Suuronen S., Ukko J., Eskola R. et al. A systematic literature review for digital business ecosystems in the manufacturing industry: Prerequisites, challenges, and benefits // *CIRP Journal of Manufacturing Science and Technology*. – 2022. – Vol. 37. – P. 414-426. DOI: <https://doi.org/10.1016/j.cirpj.2022.02.016>.
5. Колмыкова Т.С., Ковалев П.П., Уколова Л.А. Эволюция цифровых экосистем в Финтехе // *Регион: системы, экономика, управление*. – 2021. – № 4. – С. 16-24. DOI: [10.22394/1997-4469-2021-55-4-16-24](https://doi.org/10.22394/1997-4469-2021-55-4-16-24).
6. Морозов М.М., Морозов М.А. Цифровые экосистемы как инструмент трансформации сервисной экономики // *Вестник Академии знаний*. – 2021. – № 45. – С. 221-227. DOI: [10.24412/2304-6139-2021-11361](https://doi.org/10.24412/2304-6139-2021-11361).
7. Романюк Е.В., Байракова И.В., Трусевич Е.В. Цифровые экосистемы на современном этапе развития // *Проблемы социально-экономического развития Сибири*. – 2021. – № 4. – С. 60-64. DOI: [10.18324/2224-1833-2021-4-60-64](https://doi.org/10.18324/2224-1833-2021-4-60-64).
8. Трейман М.Г., Ян Я. Бизнес-экосистемы в эпоху цифровых технологий // *Управленческий учет*. – 2022. – № 5-1. – С. 253-258. DOI: [10.25806/uu5-12022253-258](https://doi.org/10.25806/uu5-12022253-258).
9. Шилов С.А., Ганина С.А. Современные российские бизнес-экосистемы как результат развития цифровой экономики // *Вестник Российского нового университета. Серия: Человек и общество*. – 2022. – № 3. – С. 19-27. DOI: [10.18137/RNU.V9276.22.03.P.019](https://doi.org/10.18137/RNU.V9276.22.03.P.019).
10. Литвин И.Ю. Развитие цифровых платформ и организационные методы цифрового инжиниринга в российской промышленности // *Инновационное развитие экономики*. – 2022. – № 3-4. – С. 65-69. DOI: [10.51832/2223798420223-465](https://doi.org/10.51832/2223798420223-465).
11. Еферин Я.Ю., Россото К., Хохлов Ю.Е. Цифровые платформы в России: конкуренция между национальными и зарубежными многосторонними платформами стимулирует экономический рост и инновации // *Информационное общество*. – 2019. – № 1-2. – С. 16-34.
12. Лебедева Я.О., Макарова С.Н., Сергеев С.М. Разработка цифровых платформ функционирования логистических сетей // *Инновации*. – 2022. – № 2. – С. 49-53. DOI: [10.26310/2071-3010.2022.280.2.007](https://doi.org/10.26310/2071-3010.2022.280.2.007).
13. Digital 2023: Global overview report. – URL: <https://datareportal.com/reports/digital-2023-global-overview-report> (дата обращения: 25.05.2023).
14. Digital 2023: The Russian Federation. – URL: <https://datareportal.com/reports/digital-2023-russian-federation> (дата обращения: 25.05.2023).
15. Эксперты назвали компании с признаками экосистем. – URL: https://www.rbc.ru/technology_and_media/01/02/2022/61f3d76f9a794775ff544309 (дата обращения: 25.05.2023).

16. Концепция общего регулирования деятельности групп компаний, развивающих различные цифровые сервисы на базе одной «экосистемы». – URL: https://www.economy.gov.ru/material/file/cb29a7d08290120645a871be41599850/konceptiya_21052021.pdf (дата обращения: 25.05.2023).
17. Экосистемы в бизнесе // SberKnowledge. Новые тренды в менеджменте. – 2019. – № 19. – 38 с.
18. Кобылко А.А. Перспективы развития бизнес-экосистем: конкуренция, сотрудничество, специализация // Russian Journal of Economics and Law. – 2022. – Т. 16, № 4. – С. 728–744. DOI: <http://dx.doi.org/10.21202/2782-2923.2022.4.728-744>.
19. Абдрахманова Г.И., Гохберг Л.М., Демьянова А.В. и др. Платформенная экономика в России: потенциал развития: аналитический доклад / под ред. Л.М. Гохберга, Б.М. Глазкова, П.Б. Рудника, Г.И. Абдрахмановой. – Москва: ИСИЭЗ ВШЭ, 2023. – 72 с.
20. Цифровые экосистемы в России: эволюция, типология, подходы к регулированию. – URL: <https://www.iep.ru/ru/novosti/institut-gaydara-provel-issledovanie-tsifrovyykh-ekosistem.html> (дата обращения: 25.05.2023).
21. Маркова В.Д., Кузнецова С.А. Экосистемы как инновационный инструмент роста бизнеса // ЭКО. – 2021. – № 8. – С. 151–168.
22. Рейтинг сервисов подписок в России 2022. – URL: <https://frankrg.com/wp-content/uploads/2022/01/a3722ff8ba82.pdf> (дата обращения: 25.05.2023).
23. Российский рынок подписок экосистем // VC.ru. – URL: <https://vc.ru/u/163530-nikolay-sedashov/311381-rossiyskiy-rynok-podpisok-ekosistem-klyuchevye-igroki-osobennosti-i-perspektivy> (дата обращения: 25.05.2023).
24. Почему холдинги и корпорации превращаются в экосистемы // Ведомости. – URL: <https://www.vedomosti.ru/partner/articles/2021/12/21/901907-pochemu-holdingi> (дата обращения: 25.05.2023).

Материал поступил в редакцию 25.05.23.

Сведения об авторах

ГУСЕВА Мария Николаевна – доктор экономических наук, профессор кафедры управления проектом Государственного университета управления, г. Москва
e-mail: boxgusevoy@yandex.ru

БРИКОШИНА Ирина Станиславовна – кандидат экономических наук, заведующая кафедрой управления проектом Государственного университета управления, г. Москва
e-mail: is_brikoshina@guu.ru

ЯШАЛОВА Наталья Николаевна – доктор экономических наук, заведующая кафедрой экономики и управления Череповецкого государственного университета, г. Череповец
e-mail: natalij2005@mail.ru

Обзор методов и технологий обнаружения и борьбы с недостоверными новостями и искаженными данными в социальных сетях

Анализируются методы и технологии, применяемые для выявления и противодействия неправдивым новостям и искаженным данным в социальных сетях. Представлен обзор существующих методов обнаружения новостей, содержащих недостоверную информацию, а также ряд рекомендаций для пользователей социальных сетей по распознаванию неправдивых сведений. Разработан модуль, который успешно комбинирует два действия: обнаружение и предотвращение распространения недостоверной информации.

Ключевые слова: анализ данных, машинное обучение, нейронные сети, фейковые новости, социальные сети, достоверность, обнаружение, технологии

DOI: 10.36535/0548-0019-2023-08-3

ВВЕДЕНИЕ

Проблема выявления и противодействия неправдивым новостям и искаженным данным актуальна в современном мире. Социальные сети значительно влияют на повседневную жизнь людей и часто используются в качестве основного источника информации. Однако, имея потенциал для распространения полезной информации, социальные сети могут быть использованы для распространения неправдивых новостей и фальшивых сведений. В последнее время произошел резкий рост интереса к этой теме как исследователей, экспертов, так и общественности в целом. Это связано с увеличением количества искаженных сведений, которые быстро распространяются в социальных сетях, могут нанести значительный ущерб обществу и создавать неправильное представление о событиях, людях и мире в целом, вводить общество в заблуждение, влиять на принятие важных решений и приводить к негативным социальным, экономическим и политическим последствиям.

Цель настоящего исследования заключается в том, чтобы добиться доверия к информации, распространяемой в социальных сетях, и снизить риск возникновения негативных последствий путем разработки эффективного метода и технологии для борьбы с неправдивыми новостями и фальшивыми сведениями.

Для этого необходимо:

- проанализировать и классифицировать недостоверные новости и искаженные данные;
- изучить причины возникновения фейковых новостей и искаженных данных;
- исследовать влияние недостоверных новостей и искаженных данных на общественное мнение и процессы принятия решений;

- оценить риски, связанные с распространением недостоверных новостей и искаженных данных;
- разработать рекомендации для пользователей социальных сетей по распознаванию недостоверных новостей и искаженных данных и борьбе с ними;
- проанализировать существующие методы и технологии для обнаружения и борьбы с недостоверными новостями и искаженными данными;
- разработать новый метод и технологию для обнаружения неправдивых сведений.

ДЕЗИНФОРМАЦИЯ В СОЦИАЛЬНЫХ СЕТЯХ: АНАЛИЗ, ПРОТИВОДЕЙСТВИЕ И ИНТЕГРИРОВАННЫЙ МОДУЛЬ

Искаженные данные в социальных сетях могут быть классифицированы по нескольким категориям в зависимости от их природы и причин возникновения:

- 1) недостоверные новости — сознательно ложная информация, которая распространяется с целью манипулирования общественным мнением, и может иметь различные формы. Например, это фальшивое заявление о событии, недостоверная статистика или фотография, которая была изменена или подделана;
- 2) искажение фактов — неправильная интерпретация фактов, может быть нечаянной или сознательными и приводить к неправильному восприятию событий. Например, может служить созданию ложной причинно-следственной связи, т. е. созданию связи между двумя событиями, которых на самом деле не было;
- 3) манипуляция информацией — техника манипулирования сознанием. Например, создание впечатления, что определенные события или идеи более популярны, чем они есть на самом деле, или обесценивание мнения определенных групп;

4) выборочное представление данных – использование только тех данных, которые соответствуют определенным интересам или точкам зрения, и игнорирование других данных, которые могут им противоречить. Например, создание ложных статистических данных для подтверждения определенной точки зрения;

5) использование низкокачественных, недостаточно авторитетных и ненадежных источников информации, которые могут содержать неточности и ошибки, для подтверждения определенных утверждений;

6) изменение контекста – представление данных в неточном контексте, что может приводить к неправильному восприятию событий. Например, изменение хронологии, или порядка событий, или временных рамок, чтобы исказить смысл фактов;

7) негативный комментарий – толкование событий и отзывов, которые могут влиять на мнение пользователей. Например, оскорбительные или угрожающие комментарии для дискредитации определенных личностей, мнений или организаций;

8) использование ботов – компьютерных программ для автоматического распространения информации, которые могут вводить в заблуждение пользователей и создавать иллюзию популярности определенных идей или событий;

9) привлечение сторонних исполнителей – для создания текстовых материалов. Существующие площадки для заказа текстов или рерайтинга предоставляют возможность найти специалиста, который напишет

новости, статьи или отзывы на любую тематику. Однако часто такие наемные работники не обладают достаточными знаниями в выбранной тематике или же имеют ограничения на контекст. Например, требование писать только положительные комментарии.

Для большей наглядности категории распространения ложной информации мы классифицируем в виде схемы, приведенной на рис. 1.

Недостоверная информация в социальных сетях может возникать по разным причинам:

- извлечение выгоды – получение дохода от рекламы, привлечение трафика на веб-сайты или продвижение продукции, услуг или идей;
- манипуляция общественным мнением – с целью формирования определенной точки зрения или укрепления имеющихся стереотипов;
- распространение определенных политических идеологий;
- желание привлечь внимание к определенной теме, событию или личности, разрушить доверие к СМИ и укрепить альтернативные информационные каналы, спровоцировать реакцию пользователей социальных сетей, например, с целью вызвать их гнев или возмущение.

Важно проверять информацию на достоверность и быть осторожным при ее распространении, поскольку искаженная информация может привести к неправильному пониманию реальности и повлиять на принятие важных решений.

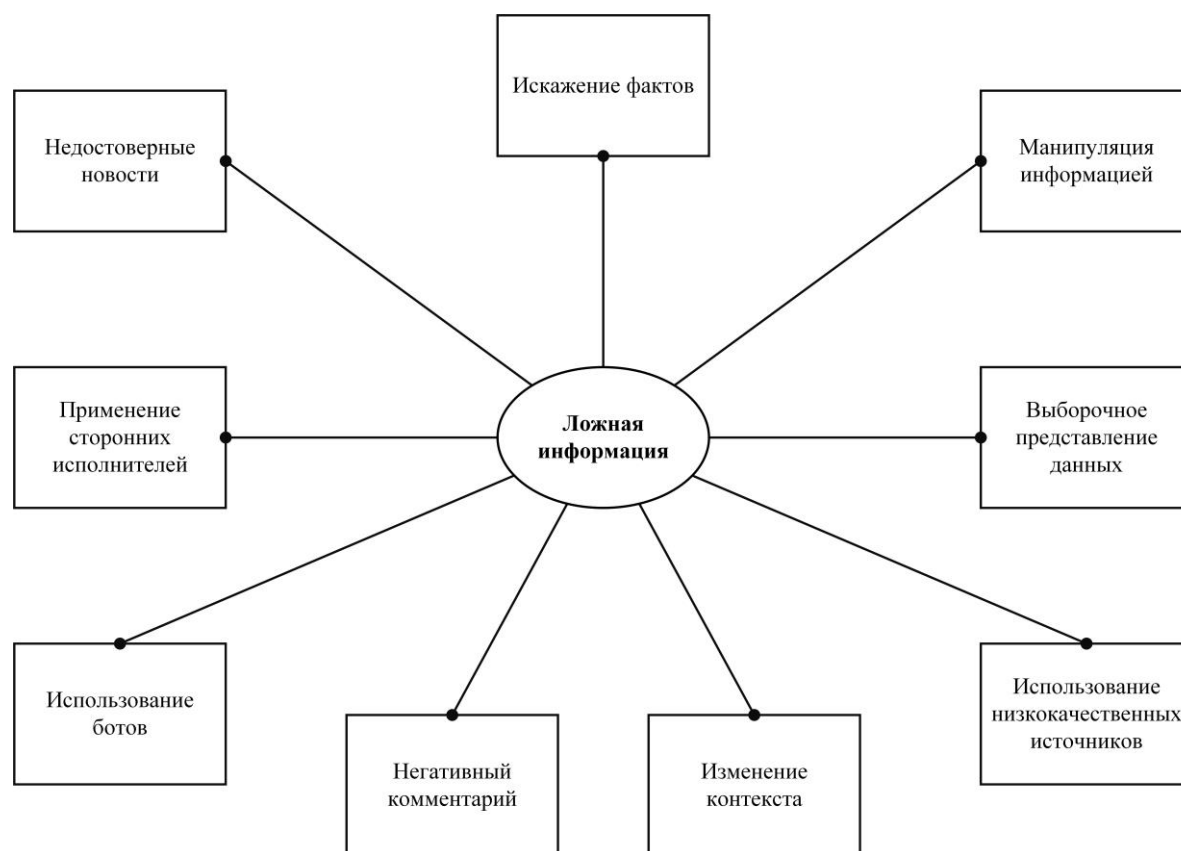


Рис. 1. Классификация ложной информации

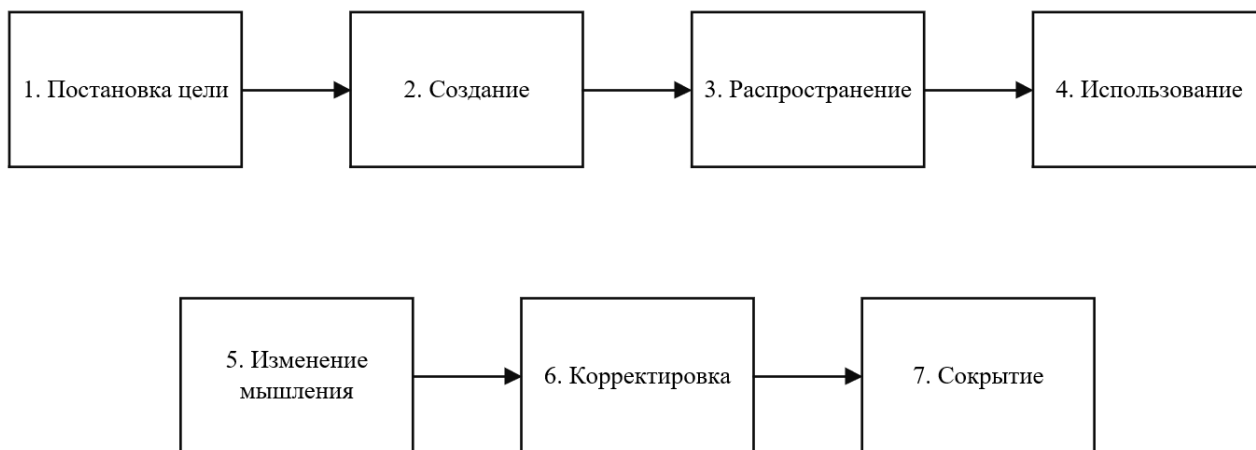


Рис. 2. Алгоритм создания ложной информации

Алгоритм создания ложной информации может использоваться в различных целях, включая манипуляцию общественным мнением, создание ложных обвинений, провоцирование конфликтов и т.д. Каждый из этапов такого алгоритма требует тщательной проработки и может быть выполнен в различных формах в зависимости от конкретной ситуации. Исходя из представленного на рис. 2 алгоритма создания недостоверных сведений, распишем каждый его шаг подробнее.

1. **Постановка цели.** Началом процесса является определение цели, которую хочет достичь агент дезинформации. Это может быть распространение фейковой информации, введение людей в заблуждение или дискредитация какой-то группы или индивидуума. Цель может быть как конкретной, так и широко сформулированной.

2. **Создание.** После постановки цели дезинформатор создает материал, который может включать в себя фальсифицированные фото- и видеоматериалы, ложные цитаты, измененные документы. Это может быть довольно сложным и трудоемким процессом, поскольку требуется выдержать баланс между достоверностью и заниженным качеством материала.

3. **Распространение.** После того, как материал был создан, дезинформатор начинает его распространение. Он может использовать социальные сети, фейковые аккаунты, ботов и другие методы, чтобы убедительно донести свою ложную информацию до целевой аудитории. Важно, чтобы материал попал в максимальное количество ресурсов.

4. **Использование.** Дезинформатор может использовать фальшивые материалы для достижения своих целей – попытаться заставить людей изменить взгляды на определенную тему, разжечь международные конфликты, повлиять на результаты выборов, а также чтобы дискредитировать оппонентов или создать более благоприятную среду для своих интересов.

5. **Изменение мышления.** Это один из ключевых элементов дезинформации. Дезинформатор может использовать ложную информацию для воздействия

на убеждения людей и заставить их изменить свое отношение к определенной теме или событию, применяя различные методы, такие как ложные доказательства, научная риторика и эмоциональные аргументы.

6. **Корректировка.** После распространения ложной информации, дезинформатор должен продолжать мониторить ситуацию и корректировать свои методы в зависимости от результатов. Например, если ложная информация не достигла нужного эффекта, то дезинформатор может изменить способ распространения, формулировки или даже саму информацию.

7. **Соккрытие.** Может потребоваться сокрытие происхождения ложной информации, чтобы не нести ответственность за ее создание и распространение. Соккрытие может включать анонимные источники, подделку данных об авторе или удаление следов происхождения информации.

Научиться распознавать недостоверные сведения становится крайне важным для тех, кто хочет быть информированным и принимать правильные решения. Для пользователей социальных сетей попытаемся сформулировать рекомендации, которые помогут разобраться в информации и отличить правду от лжи:

- проверяйте дату публикации, чтобы убедиться, что новость не устарела. Иногда недостоверные новости могут быть старыми, измененными или повторно опубликованными в надежде обмануть людей;
- контролируйте фотографии и видео с помощью поисковых систем, чтобы найти исходные изображения или видео и убедиться, что они не были изменены;
- не доверяйте источникам, которые просят о распространении новости, это может быть признаком фейковой информации;
- подписывайтесь на надежные источники, чтобы получать правдивые и достоверные новости, а не попадать на уловки недостоверных новостей и манипуляторов;
- не поддерживайте недостоверные новости – не комментируйте, не лайкайте и не распространяйте недостоверные новости и искаженные данные. Это

может поощрять других пользователей делать то же самое и приводить к еще большему распространению ложной информации;

- будьте внимательны к языку новостей – недостоверные источники могут использовать язык, который направлен на вызов определенных эмоций или настроений. Будьте внимательны и держите свой ум в здравом состоянии, чтобы избежать манипуляций;

- включайте критическое мышление – анализируйте информацию, которую вы получаете в социальных сетях. Не верьте всему, что вы читаете, и используйте логику и здравый смысл, чтобы отличать правду от лжи;

- не забывайте о личной ответственности – каждый пользователь социальных сетей несет личную ответственность за свои действия в сети. Помните, что вы можете влиять на то, как распространяется информация в социальных сетях, и что борьба с недостоверными данными начинается с каждого пользователя;

- используйте несколько источников информации, чтобы получить более полное представление о событии или новости. Это поможет вам убедиться, что сведения, которые вы получаете, достоверны;

- пользуйтесь социальными сетями с осторожностью – они могут быть потенциально опасными, если вы не знаете, как правильно взаимодействовать с ними. Будьте осторожны с тем, что вы делаете и распространяете в социальных сетях.

Следуя этим рекомендациям, вы сможете более осознанно использовать социальные сети и защитить себя от манипуляций и обмана, что будет способствовать созданию более информированного и ответственного интернет-сообщества.

Для обнаружения и борьбы с фейковыми новостями и искаженными данными в социальных сетях существует множество методов и технологий. Рассмотрим некоторые из них.

Методы обнаружения фейковых сообщений:

- машинное обучение использует алгоритмы для выявления особенностей ложной информации. Требуется большого объема данных, на которых алгоритмы машинного обучения будут обучаться;

- анализ семантики текста, с помощью алгоритмов обработки естественного языка. Это позволяет выявлять несоответствия в тексте, которые могут свидетельствовать о том, что информация является недостоверной;

- использование специальных баз данных, которые содержат информацию о недостоверных сведениях, для проверки информации и сравнения ее с уже известными случаями искаженных данных;

- визуальный анализ для обнаружения искаженных изображений, основанный на обнаружении изменений в цвете, форме и размере объектов на изображении.

Технологии борьбы с фейковыми сообщениями:

- факт-чекинг – это процесс проверки достоверности информации в новостях и сообщениях в социальных сетях. Факт-чекеры проверяют информацию на соответствие реальности и выносят вердикт о ее достоверности;

- блокчейн-технология может быть использована для создания децентрализованных и безопасных платформ для обмена информацией и проверки ее достоверности;

- обучение критическому мышлению может помочь пользователям социальных сетей улучшить свои навыки анализа информации и выявления искаженных данных. Эта технология позволяет пользователю более точно оценивать информацию и отличать достоверную информацию от неправдивой;

- коллаборативная фильтрация используется для анализа поведения пользователей в социальных сетях и соблюдения их интересов, показывая пользователям только ту информацию, которая для них наиболее релевантна, тем самым уменьшая количество искаженных данных;

- участие пользователей социальных сетей может помочь в борьбе с неправдивыми новостями и искаженными данными, сообщая о подозрительной информации и помогая другим пользователям их распознавать.

В табл. 1 и 2 представлены достоинства и недостатки методов обнаружения и технологий для предотвращения фейковых новостей и фальшивых данных.

Каждый метод обнаружения недостоверных данных имеет свои преимущества и недостатки, а их использование зависит от конкретных ситуаций. Например, методы машинного обучения и анализа семантики могут быть эффективными для обнаружения фейковых новостей и искаженных данных в текстовой форме, а методы визуального анализа лучше применять для обнаружения фейковых изображений. Объединение нескольких методов может быть более эффективным, чем один метод. Комбинация различных методов обеспечивает наиболее эффективный способ борьбы с фейковыми новостями и искаженными данными в социальных сетях.

Каждая технология борьбы с фейковыми новостями и искаженной информацией в социальных сетях имеет свои преимущества и недостатки, и выбор метода зависит от конкретной ситуации и доступных ресурсов. Факт-чекинг может быть эффективным для обнаружения искаженных данных, в то время как блокчейн-технология может обеспечить безопасность и защиту информации от фальсификации. Обучение критическому мышлению и коллаборативная фильтрация информации могут помочь пользователям распознавать фейковые новости и искаженные данные, но не являются эффективным способом борьбы с ними. Участие пользователей может повысить уровень осведомленности, но требует их активного участия и может быть недостаточно эффективным в обнаружении всех случаев фейковых новостей и искаженных данных.

В свете исследуемой проблемы, для обнаружения и борьбы с неправдивыми новостями и искаженными данными в социальных сетях представляется необходимым разработать новый метод и технологию, которые могут помочь как пользователям социальных сетей, так и редакторам контента, журналистам и другим специалистам, работающим с информацией в социальных сетях.

**Достоинства и недостатки методов обнаружения неправдивых новостей
и фальшивых данных**

Методы обнаружения	Достоинства	Недостатки
Машинное обучение	– Высокая точность обнаружения недостоверных сведений. – Автоматизированное обнаружение. – Возможность использования для обработки большого объема данных.	– Требуется большое количество данных для обучения. – Машинное обучение не всегда может адекватно обрабатывать новые типы искаженных данных.
Анализ семантики	– Способность выявлять несоответствия в тексте. – Возможность обработки естественного языка.	– Может быть недостаточно эффективным при обработке неформального языка и сленга. – Требуется определенных знаний в области лингвистики.
Использование баз данных	– Позволяет быстро идентифицировать уже известные случаи ложной информации. – Не требует сложной обработки.	– Не обнаруживает новые типы искаженных данных.
Визуальный анализ	– Способность обнаруживать искажения на изображениях. – Возможность использования для обработки большого объема изображений.	– Не может распознать искажения, которые произошли на основе других изображений. – Может быть неэффективным при обработке очень сложных искажений.

Таблица 2

**Достоинства и недостатки технологий борьбы с неправдивыми новостями
и фальшивыми данными**

Технологии борьбы	Достоинства	Недостатки
Факт-чекинг	– Позволяет быстро проверить достоверность информации. – Широко распространен и доступен в Интернете. – Участие экспертов может повысить качество проверки.	– Не может покрыть все случаи искаженных данных. – Занимает много времени и требует значительных усилий, чтобы провести достаточно проверок.
Блокчейн-технология	– Обеспечивает высокую степень безопасности и защиты от фальсификации информации. – Позволяет создавать децентрализованные платформы для проверки информации.	– Требуется значительных усилий и ресурсов для реализации. – Еще не имеет широкого распространения и может быть сложным для использования для большинства пользователей.
Обучение критическому мышлению	– Позволяет пользователям развивать навыки критического мышления, что может помочь им распознавать неправдивые новости и искаженные данные. – Может быть эффективным в долгосрочной перспективе.	– Не предоставляет непосредственного способа борьбы с неправдивыми новостями и искаженной информацией. – Требуется значительных усилий и времени для достижения результатов.
Коллаборативная фильтрация	– Позволяет пользователям настраивать свои предпочтения и получать рекомендации, которые соответствуют их интересам и предпочтениям. – Может помочь пользователям получать информацию, которая более подходит для их личных нужд.	– Не является эффективным способом борьбы с искаженной информацией. – Может привести к созданию «информационных пузырей», когда пользователи получают информацию, которая соответствует только их суждениям и мнениям.
Участие пользователей	– Пользователи социальных сетей могут помочь в борьбе с искаженными данными, сообщая о подозрительной информации и помогая другим пользователям распознавать фальшивые сведения. – Может повысить уровень осведомленности пользователей.	– Не обнаруживает все случаи искаженных данных. – Может вызвать фальшивые сигналы, если пользователи сообщают о подозрительной информации без должной проверки.

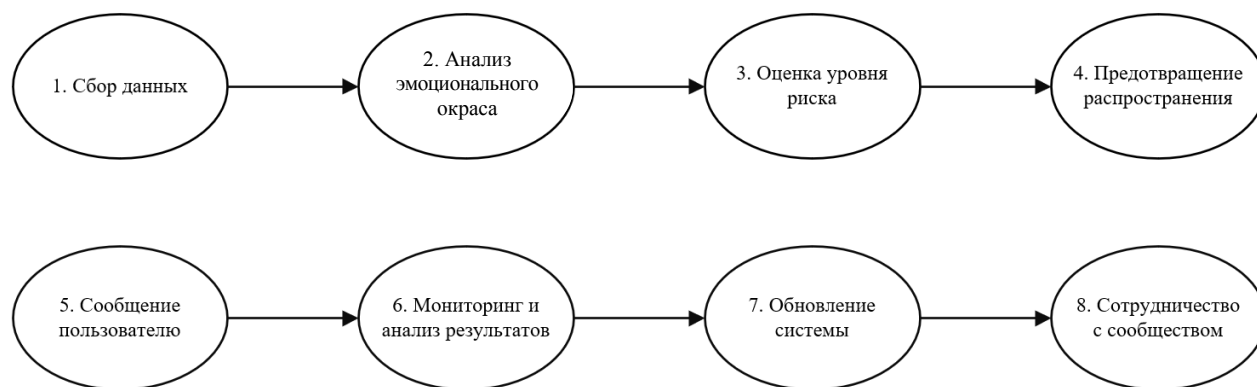


Рис. 3. Обобщенная структура модуля анализа эмоционального окраса текста

Мы предложили объединить метод обнаружения и технологию предотвращения поддельных новостей и неправдивых сведений в социальных сетях в единый модуль, который будет основан на анализе эмоциональной окраски текстовых сообщений и комментариев пользователей в социальных сетях. Предлагаемый модуль использует методы машинного обучения для автоматического определения эмоциональной окраски сообщений, а также алгоритмы обнаружения новостей и информации, содержащей неправдивые сведения, основанные на анализе структуры и контекста сообщений. Такой подход позволит значительно увеличить эффективность обнаружения и предотвращения недостоверных новостей и ложной информации в социальных сетях.

На рис. 3 представлена обобщенная структура разработанного модуля, а теперь для обнаружения и предотвращения поддельных и неправдивых сведений рассмотрим каждый этап подробнее.

Сбор данных: на этом этапе концентрируются данные из различных источников, таких как публичные посты и комментарии в социальных сетях, новостные статьи и блоги. Затем эти данные будут использоваться для анализа и обнаружения недостоверных сведений. Например, можно применить API (*Application Programming Interface*) социальной сети для получения публичных сообщений и комментариев с целью их дальнейшего анализа, а также использовать веб-скрейпинг для сбора данных с других веб-ресурсов.

Анализ эмоционального окраса: для определения можно использовать такие алгоритмы машинного обучения, как анализ сентимента. С этой целью необходимо составить обучающую выборку сообщений с разными эмоциональными окрасками и обучить модель, чтобы она могла автоматически определять эмоциональную окраску других сообщений.

Оценка риска: на основе анализа модуль может определить вероятность того, что сообщение с чрезмерно негативной эмоциональной окраской может содержать фейковую информацию или ложную новость.

Предотвращение распространения фейковой информации и ложных новостей: для этого можно использовать различные механизмы. Например, блокировать или скрывать сообщения, которые имеют высокий риск содержания фейковой информации, а

также предупреждать пользователей о возможном риске, если они все же решат воспользоваться таким сообщением.

Сообщение пользователю: о фейковой информации или ложной новости пользователя следует ставить в известность и рекомендовать ему проверить информацию перед ее распространением. Например, можно показать предупреждение, которое появляется перед тем, как пользователь сможет поделиться сообщением в социальной сети, и рекомендовать проверить факты источника, который опубликовал сообщение.

Мониторинг и анализ результатов: для дальнейшего улучшения работы системы необходим постоянный её мониторинг. Например, можно анализировать количество заблокированных или скрытых сообщений, или предупреждений, которые были показаны пользователям, и оценивать фейковую информацию.

Обновление системы: для учета новых видов фейковой информации и ложных новостей необходимо регулярно обновлять систему. Например, использовать машинное обучение для автоматического обучения системы на новых видах фейковой информации.

Сотрудничество с сообществом: важно, чтобы пользователи участвовали в борьбе с фейковой информацией и ложными новостями. Например, можно пригласить пользователей социальной сети сообщать о фейковых и ложных новостях, которые они заметили, а также проводить образовательные кампании, чтобы пользователи стали более внимательными к источникам информации и научились проверять ее достоверность.

На рис. 4 представлен алгоритм анализа эмоционального окраса текста, а теперь рассмотрим его подробнее.

Обучающая выборка. Она состоит из текстовых данных с соответствующими эмоциональными метками или классами (например, негативный, позитивный или нейтральный) и используется для обучения модели и настройки ее параметров.

Извлечение признаков. На этом этапе из текстовых данных извлекаются признаки, которые будут использоваться для классификации эмоциональной окраски. Это могут быть методы NLP, такие как мешок слов (*bag-of-words*) или TF-IDF (*term frequency-inverse document frequency*), для представления текста в виде числовых векторов.

Обучение модели. Используя извлеченные признаки и размеченную обучающую выборку, модель машинного обучения, например, классификатор (наивный байесовский классификатор, логистическая регрессия, метод опорных векторов или нейронная сеть), обучается находить связь между признаками и эмоциональной окраской текста.

Тестирование модели. После обучения модель тестируется на неразмеченных данных, которые не использовались в обучающей выборке. Это позволяет оценить точность и эффективность модели в предсказании эмоциональной окраски текста.

Классификация новых текстов. После успешного тестирования модель может быть использована для классификации новых текстовых данных. Она должна анализировать текст и предсказывать его эмоциональную окраску на основе обученных параметров и предварительно извлеченных признаков.

Предлагаемый нами модуль борьбы с недостоверной информацией и ложными новостями в социаль-

ных сетях выглядит довольно комплексным и включает в себя широкий спектр мероприятий, которые позволят снизить риск распространения ложной информации.

Сбор данных из различных источников, анализ эмоционального окраса, оценка риска, предотвращение распространения, сообщение пользователям, мониторинг и анализ результатов работы системы, обновление системы и сотрудничество с сообществом — все эти шаги позволяют не только выявлять фейковую информацию и ложные новости, но и уменьшать вероятность их дальнейшего распространения в социальных сетях.

Важно отметить, что этот модуль должен быть постоянно совершенствуемым и адаптируемым под изменяющиеся условия и требования. Следует активно сотрудничать с сообществом, вовлекать пользователей в процесс выявления и борьбы с недостоверной информацией и проводить образовательные мероприятия для повышения уровня осведомленности и критического мышления в сфере новостей и информации в целом.



Рис. 4. Структурная схема алгоритма анализа эмоционального окраса текста

ЗАКЛЮЧЕНИЕ

В настоящей статье были рассмотрены и проанализированы различные методы и технологии для обнаружения и борьбы с недостоверными новостями и данными в социальных сетях. Они включают в себя использование искусственного интеллекта для обнаружения неправдивых новостей и искаженной информации, анализ контента, проверку фактов, технологии блокировки фейковых новостей и анализ социальных сетей. Кроме того, технология обучения может помочь пользователям улучшить свою способность критически мыслить и отличать правдивую информацию от ложной.

Каждый из этих методов и технологий имеет свои преимущества и ограничения, и эффективность их применения может зависеть от конкретного контекста. Для обеспечения эффективной борьбы с фейковыми новостями и искаженной информацией в социальных сетях могут потребоваться несколько методов и технологий в сочетании.

Модуль, основанный на анализе эмоциональной окраски текстовых сообщений и комментариев пользователей в социальных сетях, предназначен для выявления и противодействия искаженным данным. Использование предлагаемого подхода позволит существенно повысить эффективность обнаружения и предотвращения распространения недостоверных новостей и ложной информации в социальных сетях.

Борьба с недостоверными данными в социальных сетях – сложная и многогранная задача, которая требует применения различных методов и технологий, а также участия пользователей, социальных сетей и общественности. Только совместными усилиями мы можем справиться с этой проблемой и обеспечить более надежную и достоверную информацию в социальных сетях.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Козловский Б. Максимальный репост: Как соцсети заставляют нас верить фейковым новостям. — Москва: Альпина Пабlisher, 2018. — 198 с. — URL: <https://e.lanbook.com/book/125817> (дата обращения: 13.02.2023). — Режим доступа: для авториз. пользователей.
2. FAKE NEWS. Актуальность, распознавание и построение ложных новостей // itnan.ru. — URL: <https://itnan.ru/post.php?c=1&p=692196> (дата обращения: 14.02.2023).
3. Дорофеева В.В. Фейковые новости в современном медиaprостранстве // Вопросы теории и практики журналистики. — 2019. — №4. — URL: <https://cyberleninka.ru/article/n/fejkovyie-novosti-v-sovremennom-mediaprostranstve> (дата обращения: 09.03.2023).
4. Шолле Ф. Глубокое обучение с R и Keras / пер. с англ. В.С. Яценкова. — Москва : ДМК Пресс, 2023. — 646 с. — URL: <https://e.lanbook.com/book/315488> (дата обращения: 15.03.2023). — Режим доступа: для авториз. пользователей.
5. Ингерсолл Г.С., Мортон Т.С., Фэррис Э.Л. Обработка неструктурированных текстов. Поиск, организация и манипулирование. — Москва : ДМК Пресс, 2015. — 414 с.
6. Бонцанини М. Анализ социальных медиа на Python. Извлекайте и анализируйте данные из всех уголков социальной паутины на Python / пер. с англ. А.В. Логунова. — Москва : ДМК Пресс, 2018. — 288 с.

Материал поступил в редакцию 15.05.23.

Сведения об авторах

КОЛОСОВ Максим Игоревич – студент 4-го курса кафедры информационных систем и сетей КФ МГТУ им. Н.Э. Баумана
e-mail: mgtumax@yandex.ru

Модель опережающего информационного обслуживания (на базе научно-технической библиотеки судостроительного предприятия)

Разработана модель опережающего информационного обслуживания специалистов в научно-технической библиотеке организации, которую предлагается встроить в систему электронного документооборота путем создания личного кабинета специалиста. Формирование рекомендаций для специалистов основано на коллаборативной фильтрации документов, адаптированной под особенности работы научно-технической библиотеки и ее аудитории.

Ключевые слова: опережающее информационное обслуживание, научно-техническая библиотека, рекомендательная система

DOI: 10.36535/0548-0019-2023-08-4

Для разработки модели опережающего информационного обслуживания мы использовали общие принципы организации работы судостроительного предприятия и особенности функционирования его научно-технической библиотеки. Документы в библиотеке представлены как на бумажных, так и на электронных носителях. Для сотрудников организации доступен электронный каталог и электронная библиотека на основе автоматизированной библиотечно-информационной системы (АБИС) «Руслан». Оперативная деловая коммуникация поддерживается системой электронного документооборота. Поступающие в фонд библиотеки документы обрабатываются с помощью внутреннего тезауруса, предметные рубрики и ключевые слова которого разработаны совместно со специалистами отраслевых подразделений предприятия. Тезаурус включает специфические термины судостроительной отрасли и способен наиболее точно отражать содержание документа, к тому же (что важно для обслуживания специалистов) он понятен самим специалистам [1].

Цель создания модели опережающего информационного обслуживания – упреждающее удовлетворение информационных потребностей пользователей, которые не были ими сформулированы в информационном запросе. В этом нам могут помочь сами специалисты, как писала В.А. Минкина: «специалист... знает о своем «незнании» больше, чем указывает в запросе» [2, с. 21]. Такое знание о «незнании» можно выявить, наблюдая за поведением специалистов: что читают, как оценивают прочитанное. Работники библиотеки должны взаимодействовать со специалистами – объективно оценивать содержание профессиональной литературы могут только люди, «погруженные в предмет». Поэтому разрабатываемая модель должна обеспечивать ряд возможностей: отслеживать и анализировать читательское поведение специалистов; распространять на весь коллектив ре-

комендации, которыми специалисты обмениваются в узком кругу (что особенно актуально на крупных предприятиях, где далеко не все сотрудники, работающие в одной сфере, знакомы друг с другом); передавать сформированные рекомендации специалистам на их рабочие места (что важно для территориально распределенных предприятий, каким является исследуемое нами). На рис. 1 в общем виде представлена модель опережающего информационного обслуживания специалистов.

Для реализации модели мы предлагаем создать Личный кабинет специалиста, который можно встроить в систему электронного документооборота, что не исключает необходимости предварительной регистрации. Пользователю нужно будет заполнить минимальные сведения о себе и предложить пароль. Перспективность интеграции Личного кабинета в систему электронного документооборота предприятия обусловлена тем, что уже сегодня через систему можно переходить в электронную библиотеку и ее каталог. Именно фонд библиотеки станет базой предоставления документов для их оценки, необходимой для дальнейшего формирования рекомендаций. В результате работы модели образуется список потенциально актуальных документов или «Перечень рекомендаций». Он будет строиться по принципу коллаборативной фильтрации. При таком подходе рекомендация формируется на основе поведения конкретного пользователя и/или группы пользователей с близкими интересами [3].

Допустим, у нас есть специалист А, который высоко оценил первую, вторую и третью книги, и есть специалист Б, который высоко оценил первую и вторую книги. Так как оценки специалистов А и Б совпадают в отношении первой и второй книг, то мы можем предположить, что специалист Б высоко оценит и третью книгу. На рис. 2 показан принцип работы коллаборативной фильтрации документов.

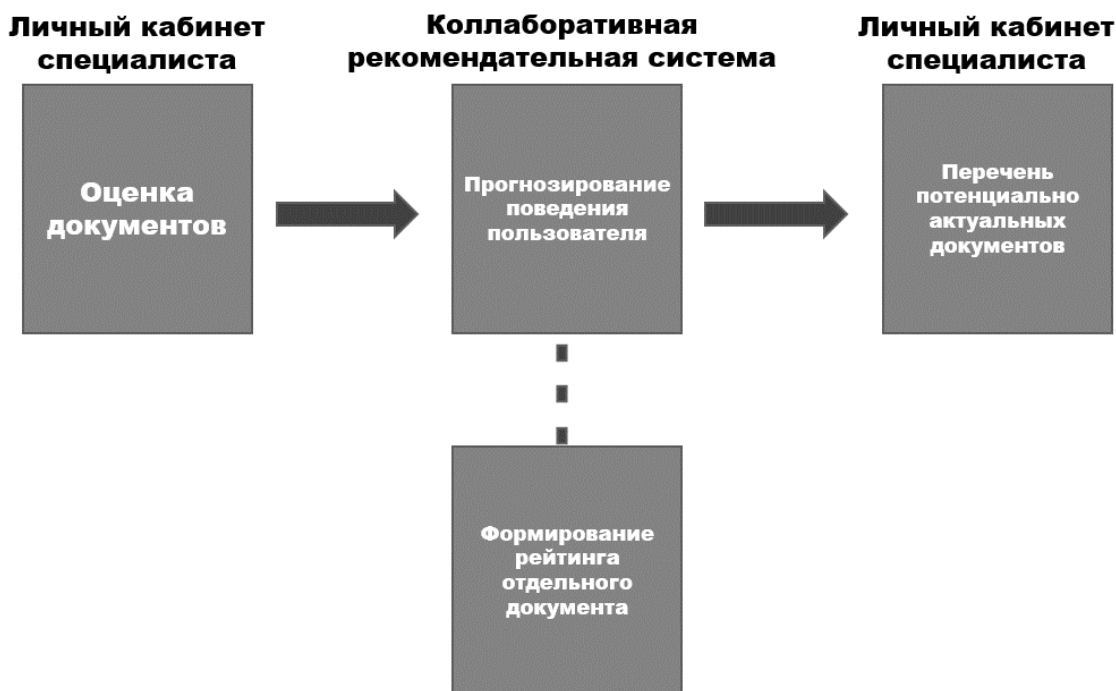


Рис. 1. Общий вид модели опережающего информационного обслуживания специалистов
(Пунктиром обозначены параллельно выполняемые действия)

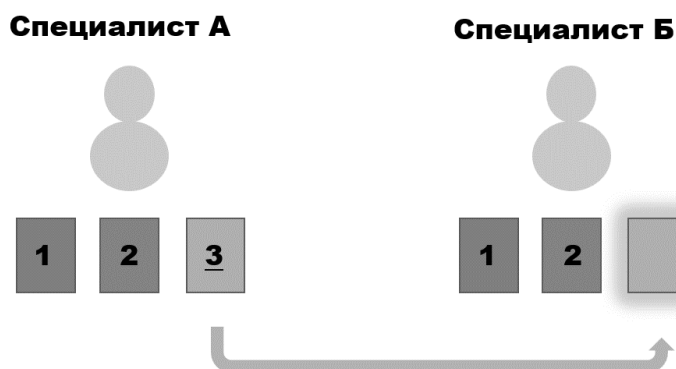


Рис. 2. Общий принцип работы коллаборативной фильтрации документов

Алгоритм подготовки рекомендаций основан на сравнении поведения пользователей, называется *User-based* и включает следующие шаги:

- 1) определение совпадения интересов всех пользователей с интересами того пользователя, для которого подбирается рекомендация;
- 2) отбор пользователей, интересы которых в большей степени совпадают с интересами пользователя, для которого подбирается документ;
- 3) прогнозирование оценки, которую поставит пользователь документу, на основе оценок других пользователей (их называют «соседями») [4, 5].

Оценки, которые пользователь выставляет документу, будут собираться и обрабатываться в Личном кабинете специалиста. Оценка имеет диапазон от 0 до 5, где «0» – документ не содержал информации для решения поставленной задачи, «5» – документ содержал информацию для решения поставленной задачи. Установленная система оценок является показателем для определения поведения (интересов) специалиста. Эта шкала оценок будет появляться не только в «Перечне рекомендаций» в Личном кабинете, но и в Электронном каталоге (рядом с библиографической записью), и в электронной библиотеке. Это

позволит пользователю оценивать документ сразу же после ознакомления с его полным текстом в электронном виде.

Для формирования рекомендации каждой оценке документа присваивается весовой коэффициент, который влияет на формирование общего рейтинга этого документа. Например, оценка руководителя подразделения будет иметь больший «вес», чем оценка инженера этого подразделения.

Рассматривая конечный результат, который выдаст система, мы можем столкнуться с большим количеством документов, с полными текстами которых пользователь не сможет ознакомиться. Эта проблема решается следующим образом: мы можем установить ограничение на рейтинг документов, которые попадают в список. Например, будут выдаваться только документы с рейтингом не ниже «5», т. е. документы с более низким рейтингом пользователь не увидит. Можно также установить ограничение на количество показываемых документов.

Однако у системы на основе коллаборативной фильтрации существует проблема «холодного старта», так как на этапе запуска системе может не хватить информации о документе, поведении пользователя для построения рекомендаций [6].

Чтобы новые поступления библиотеки попадали в «Перечень рекомендаций» пользователей, их следует оценивать, а чтобы получить эту оценку документ должен поступать в систему модели.

Проблему «холодного старта» пользователей можно решить, включив в функционал библиотекаря:

- рассылку информации о наличии новой системы с предложением регистрации через систему электронного документооборота;
- взаимодействие со специалистом при его визите в библиотеку: предлагать зарегистрироваться в системе, помогать в этом, и основываясь на данных из читательского билета о документах, которые специалист уже заказывал, рекомендовать ему оценить эти документы (при условии, что они загружены в систему). Если читатель заинтересовался в получении конечного перечня документов, но по субъективным причинам не хочет самостоятельно их оценивать, то можно предложить вариант поддержки профиля в этой системе (на предприятии подобная практика существует с наукометрическими базами данных). Библиотекарь будет иметь доступ к Личному кабинету читателя (с его согласия), сможет вносить оценки документов и скачивать «Перечень рекомендаций». Оценки можно собирать и иным способом, например, распечатать шкалу и предлагать читателям заполнять её при посещении библиотеки;
- простановку документу предполагаемой оценки с большим весовым коэффициентом и «распределение» этого документа в рекомендации конкретным специалистам на основе данных тезауруса;
- информирование о новой системе постоянных читателей, с которыми налажен личный контакт, а они будут информировать коллег.

Решить проблему «холодного старта» документов мы предлагаем следующим образом: загружать в систему новые документы уже получившими оценку. Это необходимо, чтобы избежать некорректного старта работы алгоритма. У каждого пользователя

свое понимание ценности документа, оценки будут разными. За первоначальное, эталонное мнение можно принять оценку, которую ставит информационный лидер в каждом направлении деятельности предприятия. Выявление информационных лидеров, взаимодействие с ними, получение оценок – всё это должен делать библиотекарь, знающий круг своих читателей, их поведение. С полученной оценкой документ загружается в систему и далее «обрастает» оценками пользователей и получает определенный рейтинг. Следует отметить, что работа с информационными лидерами должна носить постоянный характер (процесс повторяется с каждым новым документом).

Все документы, поступающие в фонд библиотеки, проходят процедуру аналитико-синтетической переработки. Таким образом они получают предметную рубрику и ключевые слова из внутреннего тезауруса предприятия. Для получения более точной рекомендации мы можем просить специалистов при регистрации в системе в Личном кабинете из предлагаемого списка (используя внутренний тезаурус) выбирать те ключевые слова и предметные рубрики, которые отражают их научные интересы.

Этим ключевым словам и предметным рубрикам присваивается коэффициент значимости (устанавливаемое число, влияющее на приоритет документа при формировании «Перечня рекомендаций»). Так, попадание в список рекомендаций и место документа в нем зависят от рейтинга и коэффициента значимости. Например, пользователь оценивает документы из разных сфер деятельности предприятия (возможно, он решит протестировать систему и в электронном каталоге начнет проставлять оценки всем документам, не привязывая их к тематике). Вероятность попадания таких документов в список будет менее, чем тех документов, которым присвоены ключевые слова и рубрики, указанные в качестве научных интересов в Личном кабинете специалиста (даже тогда, когда рейтинг «нетематического» документа выше рейтинга «тематического»).

Это можно использовать в продвижении собственных изданий предприятия к потребителю. Пусть коэффициент значимости отраслевых предметных рубрик и ключевых слов равен «1». Принимая в фонд библиотеки вышедшее в свет издание, мы, кроме отраслевых предметных рубрик и ключевых слов, присваиваем ему слово «новинка» с коэффициентом «2». Таким образом (после получения эталонной оценки от Информационного лидера) с учетом отраслевой принадлежности специалистам в первую очередь будет рекомендоваться «свежее» издание предприятия. Через определенный промежуток времени слово «новинка» убирается или выставляется коэффициент значимости «0», а документ продолжает рекомендоваться «на общих основаниях».

При регистрации пользователя в системе опережающего обслуживания мы будем просить его указывать следующие данные:

- занимаемая должность;
- ключевые слова и предметные рубрики, отражающие научные интересы (пользователь не самостоятельно формулирует слова, а выбирает из сформированного списка на основе тезауруса);
- пароль от Личного кабинета.

Таким образом, обладая минимальными данными о пользователе, мы сможем выполнять опережающее информационное обслуживание. Оценки документов будут анонимными. Автора оценок должен знать только библиотекарь и только в случае, когда он (с согласия специалиста) будет осуществлять поддержку профиля. Анонимность пользователя (мы не просим указывать ФИО) и анонимность оценок – это немаловажный фактор, так как из-за специфики работы судостроительного предприятия многие специалисты отрицательно относятся к распространению информации о себе (даже внутри предприятия, более того – внутри подразделения). Предлагаемая модель позво-

лит наладить не только опережающее обслуживание, но и обмен (пусть даже конфиденциальный) знаниями между специалистами. На рис. 3 представлен макет интерфейса Личного кабинета специалиста.

Для управления рекомендациями библиотекарю тоже необходим личный кабинет, но кабинет библиотекаря-администратора. Возможность контроля позволит избежать ситуации, когда положительные оценки документа формируются без опоры на «качество» научных знаний. С учетом изложенного мы получаем модель опережающего информационного обслуживания специалистов предприятия, представленную на рис. 4.



Рис. 3. Макет интерфейса Личного кабинета специалиста

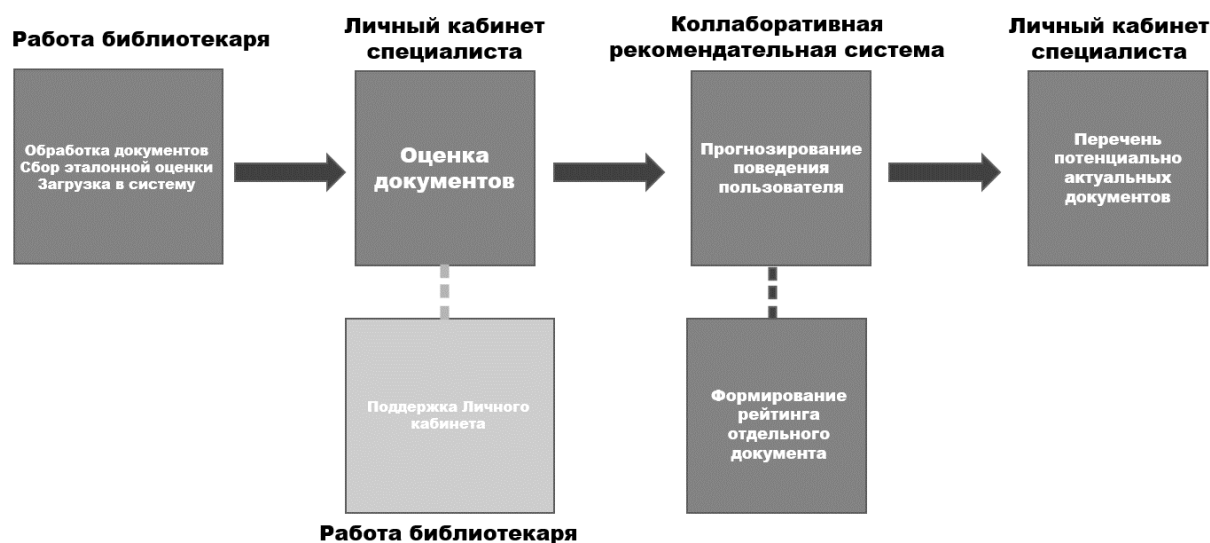


Рис. 4. Модель опережающего информационного обслуживания специалистов
(Темным цветом обозначены обязательные элементы модели, светлым – факультативные)

Улучшить эту модель может решение ряда некоторых административных проблем предприятия. В качестве перспектив можно предложить создание единой библиографической базы данных:

- технических документов: стандартов, регламентов;
- патентных документов;
- отчетов о НИОКР.

Созданная таким образом единая библиографическая база данных поможет упорядочить документы предприятия и включить их в систему обслуживания специалистов.

СПИСОК ЛИТЕРАТУРЫ

1. Крулев А.А. Системный рубрикатор и тезаурус ведомственной технической библиотеки: особенности, возможности и перспективы // Научно-техническая информация. Сер. 2. – 2016. – № 3. – С. 28-33.
2. Минкина В.А. Библиотечно-библиографическое обслуживание в непрерывном образовании специалистов : дис. ... д-ра пед. наук : 05.25.03. – Санкт-Петербург: СПбГИК, 1993. – 390 с. – Прил.: с. 1-190.
3. Карауш А.С. Рекомендательные системы в публичных библиотеках // Библиосфера. – 2009. – № 1. – С. 41-43.
4. Гомзин А.Г., Коршунов А.В. Системы рекомендаций: обзор современных подходов // Труды Института системного программирования РАН. – 2012. – Т. 22. – С. 401-417.
5. Полетаева Ю.С. Разработка сетевой рекомендательной системы для научно-технической библиотеки ИРНИТУ // Материалы IV Всероссийской научно-практической конференции с международным участием «Научные коммуникации. Профессиональная этика». – Омск : Омский государственный технический университет, 2019. – С. 120-127.
6. Волкова Л.Л., Токарева А.А., Ланко А.А. О разработке рекомендательной системы, предлагающей книги по предпочтениям пользователя // Новые информационные технологии в автоматизированных системах. – 2017. – № 20. – С. 239-244.

Материал поступил в редакцию 30.05.23.

Сведения об авторе

ПУНДА Анастасия Сергеевна – магистрант кафедры информационного менеджмента библиотечно-информационного факультета Санкт-Петербургского государственного института культуры
e-mail: punda.nastya@mail.ru

ДОКУМЕНТАЛЬНЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ

УДК 347.77.028:004.822:004.85(470+571)

А.В. Горбунов, Б.Л. Генин, Д.С. Золкин

Задача выявления элементов семантического кластера патентных документов для поиска уровня техники*

Представлена концепция, приведено определение и раскрыта методика формирования семантических кластеров патентных документов. Предложено рассматривать поиск уровня техники как задачу выявления элементов такого кластера.

Описаны созданные в Роспатенте и доступные заинтересованным лицам наборы данных, содержащие семантические кластеры российских патентных документов, – датасет для машинного обучения и тестовый датасет.

Ключевые слова: патентный поиск, семантический кластер, критерий качества поиска, машинное обучение, тестовая коллекция

DOI: 10.36535/0548-0019-2023-08-5

ВВЕДЕНИЕ

В настоящей статье мы рассмотрим задачу автоматического поиска уровня техники для целей экспертизы изобретений. Во многих современных системах патентного поиска имеется функциональность, направленная на решение подобных задач, и часто формулируемая как поиск документов, похожих на заявку на изобретение (далее по тексту – «похожих»), или в более общем случае – на документ – образец.

Решением этой задачи будем считать автоматический поиск документов, которые в традиционном режиме эксперты находят и включают в виде соответствующих ссылок в отчеты о поиске по заявке и в поле с кодом ИНИД (56) (список документов уровня техники) в документы о выдаче патента на изобретение [1, 2]. Юридически более строгое определение приведено в Приказе Минэкономразвития России: «Информационный поиск проводится для определения уровня техники, с учетом которого будет осуществляться проверка патентоспособности заявленного изобретения, по не отозванным или не признанным отозванными заявкам, прошедшим формальную экспертизу с положительным результатом, в случае удо-

влетворения ходатайства о проведении экспертизы заявки “по существу”»¹.

Для удобства изложения введем понятие «семантический кластер патентных документов», который должен включать патентные документы, отражающие уровень техники в конкретной предметной области. Мы предлагаем следующую формальную дефиницию такого кластера, опирающуюся на устойчивые особенности патентных документов и вытекающую из особенностей патентного права.

Семантический кластер патентных документов включает заявку на изобретение, описание изобретения к патенту (если он выдан по данной заявке), публикации, содержащие полнотекстовую информацию о различных этапах жизненного цикла изобретения, а также следующие патентные документы: семейство патентов-аналогов, в которое входит данная заявка на изобретение (в том числе все публикации, содержащие полнотекстовую информацию о различных этапах жизненного цикла изобретений); документы, на которые ссылаются эксперты в публикациях по результатам экспертизы отчетов о поиске по заявке и в поле с кодом ИНИД (56) в сведениях о вы-

* Работы проводились в рамках НИР «Исследование и развитие использования искусственного интеллекта для патентного поиска в многоязычных базах данных при экспертизе изобретений и полезных моделей (шифр НИР: 1-ИТ-2022).

¹ «Порядок проведения информационного поиска при проведении экспертизы по существу по заявке на выдачу патента на изобретение и представления отчета о нем», утвержден приказом Минэкономразвития России от 25 мая 2016 года № 316.

даче патента на изобретение; документы семейств патентов-аналогов, в которые входят документы, на которые ссылается экспертиза; документы, указанные заявителем в качестве прототипов в подаваемой заявке; документы семейств патентов-аналогов, в которые входят документы, на которые указывают ссылки, приведенные заявителем.

Таким образом, семантический кластер состоит из документов, которые, по мнению экспертов, должны быть учтены или уже учтены при экспертизе заявки на изобретение.

С учетом введенного определения, решением заявленной нами задачи будем считать автоматический поиск документов, которые входят в соответствующий семантический кластер.

КРИТЕРИЙ КАЧЕСТВА РЕШЕНИЯ ЗАДАЧИ АВТОМАТИЧЕСКОГО ПОИСКА УРОВНЯ ТЕХНИКИ

Патентный документ (за исключением самого документа-образца), найденный при поиске похожих на документ-образец, будем считать релевантным, если он входит в семантический кластер (за исключением самого документа-образца и документов, входящих в семейство патентов-аналогов документа-образца), в составе которого имеется этот документ-образец.

В зависимости от задачи пользователя, в интересах решения которой производится оценка, можно применять дополнительные критерии фильтрации результатов поиска. Например, при поиске предшествующего уровня техники в задачах экспертизы изобретений можно считать релевантными только патентные документы, опубликованные ранее даты приоритета документа-образца.

Качество системы патентного поиска «похожих» будем оценивать по двум метрикам степени достижения качества целевого решения [3].

Максимальное качество целевого решения (обозначим его $R(K)$) будет достигнуто, если все релевантные документы система выдаст в $top(K)$ документов поисковой выдачи. Тогда вероятность достижения цели можно оценить так:

$$R(K) = \frac{1}{K} \sum_{k=1}^K (\pi_k),$$

где:

K – количество документов в поисковой выдаче, отсортированной по релевантности, по которым оценивается качество системы;

(π_k) – бизнес-метрика, вычисляемая по формуле:

$$(\pi_k) = \begin{cases} 1, & \text{если все релевантные документы попали в } top(K) \text{ выдачи,} \\ 0, & \text{если хотя бы один релевантный документ не попал в } top(K) \text{ выдачи.} \end{cases}$$

Такая характеристика качества патентного поиска «похожих» представляет собой отношение количества поисков, в которых были найдены все релевантные документы в $top(K)$ результатов поиска, к общему количе-

ству поисков. В этом подходе учитывается специфика задачи поиска при экспертизе изобретений.

Другая оценка качества патентного поиска «похожих» определяется как отношение количества поисков, в которых был найден хотя бы один релевантный документ в $top(K)$ результатов поиска, к количеству поисков. В этом случае оценка вероятности нахождения хотя бы одного релевантного документа в каждом поиске будет определяться так:

$$P(K) = \frac{1}{K} \sum_{k=1}^K (\pi_k),$$

где:

K – количество документов в поисковой выдаче, отсортированной по релевантности, по которым оценивается качество системы;

(π_k) – бизнес-метрика, вычисляемая по формуле:

$$(\pi_k) = \begin{cases} 1, & \text{если хотя бы один релевантный документ попал в } top(K) \text{ выдачи,} \\ 0, & \text{если ни одного релевантного документа нет в } top(K) \text{ выдачи.} \end{cases}$$

КОЛЛЕКЦИИ НА ОСНОВЕ СЕМАНТИЧЕСКИХ КЛАСТЕРОВ

Для обеспечения машинного обучения и тестирования систем поиска различными сообществами в разное время готовились коллекции документов, специально подобранные, очищенные от «шумов» и снабженные тестовыми наборами, оцененными по релевантности командами «человеческих» экспертов. Это в первую очередь хорошо известные коллекции CLEF-IP (патенты США и Европы), а также NTCIR (Япония), TREC-CHEM (коллекция документов по химии), WIPO и EPO (Европейские документы) [4].

Коллекции патентных документов для машинного обучения и тестирования обычно представляют собой многоцелевые наборы данных для неопределенного круга задач тестирования и сравнительного анализа систем поиска и аналитической обработки [5].

Наряду с очевидной полезностью этих коллекций, необходимо отметить, что их использование для обучения и тестирования систем поиска имеет ряд сложностей: во-первых, не все коллекции создавались под задачу поиска уровня техники, во-вторых, недостаточное языковое разнообразие коллекций – основная масса документов представлена на английском языке, документы на русском языке отсутствуют полностью. Наконец, главное препятствие для развития таких коллекций – необходимость «вручную» проводить оценку релевантности отобранных документов и формировать из них тестовые наборы. Ввиду большой трудоемкости этой работы коллекции не могут обновляться часто, а это снижает их ценность с учетом быстрого изменения мировых трендов патентования.

В настоящей работе мы концентрируем внимание на отдельной специфической задаче патентного поиска для экспертизы изобретений. Это позволяет создать эффективную для решения данной задачи тестовую коллекцию, которая как и другие известные

коллекции патентных документов, фиксируется в некоторых временных и тематических рамках и обеспечивает стабильность, воспроизводимость и сравнимость результатов обработки при сравнении и анализе решения различных задач. Но главной особенностью и отличительной чертой предлагаемой нами коллекции является ее нацеленность на эффективное использование при машинном обучении системы автоматического поиска уровня техники при экспертизе изобретений.

Мы исходим из того, что машинное обучение на текстах семантических кластеров с возможным привлечением библиографических данных (например, классификационных) патентных документов [6], входящих в эти кластеры, может дать хорошие результаты при решении задачи, заявленной на семантических кластерах.

При этом структура набора данных на основе семантических кластеров, объединяющих тексты и библиографические данные патентных документов, входящих в эти кластеры, позволяет специалистам, организующим машинное обучение, абстрагироваться от специфических особенностей и деталей форматирования патентных документов и сконцентрировать свое внимание на задачах собственно машинного обучения. В первую очередь это связано с тем, что выявление лексического пространства семантического кластера обеспечивает простоту и удобство векторизации терминов и построения дистрибутивного тезауруса лексики патентных документов.

Вместе с тем, мы не накладываем никаких ограничений на использование любых патентных документов, а также тех или иных методов и средств искусственного интеллекта и параметров машинного обучения. Мы лишь, в соответствии с общей формулировкой задачи автоматического поиска уровня техники для целей экспертизы изобретений, будем оценивать результаты машинного обучения с использованием тестового массива, также основанного на семантических кластерах.

НЕКОТОРЫЕ ВАЖНЫЕ ДЕТАЛИ ФОРМИРОВАНИЯ СЕМАНТИЧЕСКИХ КЛАСТЕРОВ

Исследуемый массив содержит семантические кластеры патентных документов, основанные на официальных публикациях Роспатента о заявках на изобретения и о выдаче патентов на изобретения и полезные модели, опубликованных в 1994–2020 гг.

Для создания набора данных мы не использовали патентные документы, указанные заявителем в подаваемой заявке, а также патентные документы семейств патентов-аналогов, включающих документы, на которые указывают ссылки, приведенные заявителем. Это связано с отсутствием в официальных публикациях Роспатента соответствующих данных в структурированной форме. В соответствии с целями создания набора данных, направленными на оценку качества поиска и анализ систем поиска в русскоязычных массивах, при формировании кластеров мы использовали только ссылки экспертизы на опубликованные Роспатентом документы на русском языке.

В официальных публикациях Роспатента ссылки экспертизы в поле (56) на патентные документы могут быть плохо структурированы и иногда содержат формальные ошибки. Программа формирования кластеров проверяла ссылки, корректировала их структурирование и при необходимости устраняла формальные ошибки.

Каждый семантический кластер представляет собой одну запись – объединение соответствующих документов в формате XML на основе стандарта ВОИС ST 96 [7].

Коллекция состоит из двух частей: первая – набор данных (датасет) для машинного обучения и вторая – тестовый датасет.

Датасет для машинного обучения содержит семантические кластеры патентных документов, основанные на официальных публикациях Роспатента о заявках на изобретения и о выдаче патентов на изобретения и полезные модели, опубликованных в 1994–2020 гг.:

- общее количество записей в датасете 1 493 023;
- общий объем массива 416 875 млн байт;
- объем массива в архивах 31 283 млн байт.

Формирование семантических кластеров патентных документов, на основе официальных публикаций Роспатента приводит к тому, что кластеры с учетом более поздних публикаций могут включать в себя полностью или частично информацию кластеров, основанных на более ранних публикациях, выявленных по ссылкам экспертизы. Возникающее таким образом частичное дублирование информации мы сознательно не устраняли, так как считали необходимым сохранить информацию о хронологии развития групп ссылок экспертизы и семейств патентов-аналогов.

Характер расширения исходных документов при формировании семантических кластеров показан в табл. 1 на примере данных за апрель в 2017–2020 гг.

Тестовый датасет содержит семантические кластеры, основанные на официальных публикациях Роспатента о заявках на изобретения и о выдаче патентов на изобретения и полезные модели, опубликованных в 2021 г.:

- общее количество записей в датасете 46 665;
- общий объем массива 22 753 млн байт.

Такое количество данных в тестовом датасете превышает обычные потребности в объеме данных для тестирования, поэтому в наших экспериментах мы использовали лишь относительно небольшую часть тестового массива. При этом представительность массива, с точки зрения равномерности распределения данных по тематическим областям, обеспечивалась отбором семантических кластеров, созданных на основе документов, опубликованных за ограниченный временной период.

Наборы данных, состоящие из семантических кластеров, созданных на основе официальных публикаций Роспатента на русском языке, выставляются на поисковой платформе Роспатента и предоставляются бесплатно заинтересованным лицам по запросу. Каждый набор данных сопровождается необходимой документацией.

Характер расширения исходных документов при формировании семантических кластеров за апрель в 2017 –2020 гг.

	Годы			
	2020	2019	2018	2017
Количество публикаций	3993	5486	5745	6265
Найдено цитат:				
1	850	1156	1213	1133
2	576	972	992	918
3	405	688	756	662
4	212	390	395	249
5	98	136	119	72
6	25	34	29	20
7	7	10	1	0
8	2	7	4	3
9	1	2	0	0
10	0	10	1	0
Всего цитат	2176	3405	3520	3059
Аналогов цитат	1785	2730	2889	2610
Аналогов документов	3158	4668	5152	5017
Всего документов в кластерах за месяц	11112	16289	17306	16951

УТИЛИТА ОЦЕНКИ КАЧЕСТВА ПОИСКА

Мы уже отмечали, что наборы данных на основе семантических кластеров и, в частности, тестовый датасет, могут использоваться для оценки качества автоматического поиска уровня техники или патентного поиска «похожих» и сравнительного анализа систем патентного поиска.

Теперь кратко опишем унифицированный инструмент для вычисления таких оценок – утилиту оценки качества автоматического патентного поиска «похожих», состоящую из трех основных блоков.

Блок отбора документов для проведения поиска и вычисления оценки. Отбор производится из тестового массива документов, т. е. из массива, не использованного для машинного обучения, случайным (квазислучайным) образом. Поиск нужных документов выполняется с использованием для поиска API Платформы патентного поиска. Для этого нужно указать конкретную дату или диапазон дат и, опционально, один или несколько кодов вида документов, соответствующих зарегистрированным (выданным) патентам. Но для ведомств с большим количеством заявок и патентов при этом все равно находится слишком много документов. В связи с этим дополнительно имеется возможность из списка результатов поиска по описанному критерию отбирать каждый второй (или третий и т.д.) документ, причем эта кратность указывается параметром. На конечной стадии отбора проверяется наличие активных патентных ссылок, что позволяет вычислять критерий качества поиска, и хотя бы одного из текстовых полей описания изобретения *abstract, description, claims* (это позволяет проводить поиск похожих документов). На выходе блок выдает (и сохраняет) список идентификаторов отобранных документов и их количество.

Функции этого блока могут не выполняться, если планируется использовать ранее созданный и сохраненный список идентификаторов отобранных документов.

Блок поиска похожих на отобранные документы в оцениваемых системах и получения результатов. Поиск проводится в массивах мирового патентного фонда на Платформе патентного поиска Роспатента. Также обеспечивается возможность поиска «похожих» в некоторых других поисковых системах (например, Google.Patent, Yandex.Patents), предоставляющих соответствующую функциональность.

Для сравнительного анализа качества поиска «похожих» в оцениваемых системах блок обеспечивает их поиск на Платформе патентного поиска Роспатента с использованием предоставляемого Платформой API (базовый, опорный вариант).

Результатом работы блока является массив отсортированных по релевантности списков результатов поиска «похожих» по документам-образцам, отобранным блоком отбора. При этом в список результатов включаются первые 20 (или другое, указанное параметром K, количество, но не более 1000) отсортированных по релевантности результатов поиска (top 20).

Блок вычисления результатов оценки качества патентного поиска «похожих». Здесь вычисляются две описанные выше метрики качества патентного поиска «похожих» на основе определения релевантности найденных документов.

В результате обработки утилита выдает вычисленные значения качества автоматического патентного поиска «похожих» и вероятность успешного поиска всех релевантных документов, а также сохраняет в файлах результаты поисков, по которым вычислялись метрики, и выявленные признаки релевантности (хотя бы один или все релевантные).

НЕКОТОРЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ АВТОМАТИЧЕСКОГО ПОИСКА ПОХОЖИХ ДОКУМЕНТОВ С ИСПОЛЬЗОВАНИЕМ УТИЛИТЫ ОЦЕНКИ КАЧЕСТВА ПОИСКА

Первые результаты исследований на созданном по семантическим кластерам русскоязычном тестовом датасете приведены в табл. 2. Наряду с подтверждением ранее опубликованных результатов сравнения качества автоматического патентного поиска «похожих» с использованием искусственного интеллекта поисковой Платформы Роспатента с подобной функциональностью систем патентного поиска Google.Patent и Yandex.Patents, цели нашего исследования включали анализ стабильности результатов оценки при изменении объемов тестовой выборки и количества учитываемых при вычислении оценки документов поисковой

выдачи, а также изменения оценок для патентных документов первой и вторичной публикации. Была также сделана попытка проанализировать устойчивость оценки в зависимости от тематической области поиска.

В экспериментах использовался фрагмент описанного выше тестового датасета, включающий публикации за январь 2021 г. с кодами вида публикации C1 и C2 с отбором части публикаций (каждый второй или каждый четвертый и т.п.). Оценки вычислялись с использованием функциональности поиска «похожих» с использованием искусственного интеллекта поисковой Платформы Роспатента, а также, для сравнения систем патентного поиска Google.Patent и Yandex.Patents. В табл. 2 приведены результаты вычисления метрики $P(K)$ для K , равного 20 и 50, и результаты вычисления полноты $R(K)$ для K , равного 20 и 50.

Таблица 2

Результаты первых исследований с использованием утилиты оценки качества поиска на тестовом датасете семантических кластеров (январь 2021 г.)

Тестовый датасет			Платформа Роспатента		Google.Patent		Yandex.Patents	
код вида	часть	кол-во	P(20)/P(50)	R(20)/R(50)	P(20)	R(20)	P(20)	R(20)
C2	1/2	222	26,13	4,05	9,01	0	9,01	0,45
C1	1/4	338	50,0/56,80	10,65/15,09	0	0	19,23	4,14
C1	1/2	675	50,81	10,96				
C1	1	1338	51,27	13,00	4,93	0,82	23,47	3,21

Таблица 3

Распределение публикаций документов с кодами видов C1 и C2 по тематике подклассов МПК

№ п/п	Код вида C1 (документов в 2021 г. 17281) / %		Код вида C2 (документов в 2021 г. 6235) / %		Код вида C1 (всего документов 490449) / %		Код вида C2 (всего документов 273677) / %	
1	A61B	7	A61K	17	A23L	6	A61K	11
2	G01N	7	A61P	13	A61K	6	A61P	8
3	A61K	6	C12N	6	G01N	5	A61B	4
4	A61P	4	C07K	6	A61B	5	G01N	4
5	E21B	3	C07D	5	A61P	4	C07D	4
6	H04W	3	A61B	4	E21B	4	A23L	3
7	G06F	3	H04W	3	B01D	2	G06F	3
8	A61F	3	G01N	3	C04B	2	C12N	3
9	C12N	3	B01J	3	A61F	2	E21B	3
10	B01D	2	B01D	3	B01J	2	C07C	3

Таблица 4

Результаты оценки качества автоматического поиска похожих документов по тематическим областям*

G01N					A61K			
520					420			
K	P(K)	R(K)	P(K)	R(K)	P(K)	R(K)	P(K)	R(K)
20	54,81	19,81	51,92	19,23	53,25	16,57	47,86	16,19
50	59,81	25	59,62	23,08	59,17	20,12	53,33	19,05
100	62,88	28,65	62,5	26,92	62,72	24,26	57,62	23,1

* Представлены данные по оценкам $P(K)$ и $R(K)$ поиска похожих документов (в процентах) на тестовых выборках по тематическим областям.

Напомним, что документы с кодом вида С2 в публикациях Роспатента содержат сведения о выдаче патента после появления сведений о заявке на изобретение. Документы с кодом вида С1 в публикациях Роспатента – это первые публикации сведений о выдаче патента без предварительной публикации сведений о заявке на изобретение. В табл. 2 показано, что результаты автоматического поиска «похожих» существенно лучше для патентных документов первой публикации с кодом вида публикации С1. Причем это улучшение характерно как для поисковой платформы Роспатента, так и для Yandex.Patents.

Важно и то, что при увеличении количества поисков для вычисления оценок с 338 до 1338, как и при увеличении количества учитываемых при вычислении оценки документов поисковой выдачи с 20 до 50, оценки возрастали, но незначительно, что свидетельствует о стабильности получаемых оценок качества поиска.

В табл. 3 представлено распределение публикаций по тематике (по подклассам МПК отдельно по публикациям 2021 г. и по всем российским публикациям).

В табл. 4 показаны первые результаты оценки качества автоматического поиска «похожих» на поисковой платформе Роспатента по тематическим областям, относящимся к наиболее активным областям патентования в России в 2021 г.:

A61K – Лекарства и медикаменты для терапевтических, стоматологических или гигиенических целей;

G01N – Исследование или анализ материалов путем определения их химических или физических свойств.

Данные табл. 4 получены с использованием утилиты оценки качества автоматического поиска похожих документов на тестовом массиве семантических кластеров патентов, опубликованных в 2021 г.

ЗАКЛЮЧЕНИЕ

В процессе решения задачи автоматического поиска уровня техники для целей экспертизы предложено новое понятие «семантический кластер патентных документов»; описаны реализация наборов данных, содержащих такие кластеры патентных документов на русском языке, предназначенных для машинного обучения и тестирования систем патентного поиска, и утилита оценки качества автоматического патентного поиска похожих документов для выявления предшествующего уровня техники на основе тестового датасета.

Результаты изучения автоматического патентного поиска похожих документов с использованием этой утилиты на созданном по семантическим кластерам русскоязычном тестовом датасете, позволили создать наборы данных, содержащие семантические кластеры патентных документов на русском языке, предназначенные для машинного обучения и тестирования

систем патентного поиска и установить их на поисковой Платформе Роспатента в бесплатном доступе.

СПИСОК ЛИТЕРАТУРЫ

1. Горбунов А.В., Генин Б.Л., Золкин Д.С., Киселев С.Л. Опыт применения методов искусственного интеллекта для повышения качества поиска "сходных" документов в патентно-поисковой системе экспертизы изобретений // В сб. «Роль интеллектуальной собственности в прорывном научно-техническом развитии общества». – Москва: Федеральный институт промышленной собственности, Роспатент, 2018. – С. 17-19.
2. Genin B.L., Zolkin D.S. Similarity search in patents databases. The evaluations of the search quality // World Patent Information. – 2021. – Vol. 64, March. DOI: <https://doi.org/10.1016/j.wpi.2021.102022>.
3. Кураленок И.Е., Некрестьянов И.С. "Оценка систем текстового поиска", Программирование 28 (4): 226-242, Системный анализ, управление и обработка информации. Санкт-Петербург: Санкт-Петербургский гос. ун-т, 2004.
4. Patent retrieval: a literature review. Walid Shalaby. Wlodek Zadrozny Knowledge and Information Systems. DOI: <https://doi.org/10.1007/s10115-018-1322-7>.
5. A Horizontal Patent Test Collection, Mihai Lupu, Alexandros Bampoulidis, Luca Papariello, July 2019, Conference: the 42nd International ACM SIGIR Conference. DOI: 10.1145/3331184.3331346.
6. Горбунов А.В., Генин Б.Л., Золкин Д.С. Искусственный интеллект в работе патентных ведомств // Информационные ресурсы России. – 2021. – № 3. – С. 18-23.
7. WIPO Standard ST.96, Version 6.0 (approved by the Task Force, October 3, 2022), Standard ST.96 (Main Body): <https://www.wipo.int/export/sites/www/standards/en/pdf/03-96-01.pdf>, Annexes I to VII: <https://www.wipo.int/standards/en/st96/v6-0/>

Материал поступил в редакцию 26.05.23.

Сведения об авторах

ГОРБУНОВ Александр Владимирович – советник директора, Федеральный институт промышленной собственности
e-mail: gorbunov@rupto.ru

ГЕНИН Борис Лемелевич – кандидат технических наук, ведущий научный сотрудник, Федеральный институт промышленной собственности
e-mail: bguenine@rupto.ru

ЗОЛКИН Дмитрий Сергеевич – начальник отдела проектирования информационно-поисковых систем, Федеральный институт промышленной собственности
e-mail: db_dept@rupto.ru