

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 8

Москва 2022

ОБЩИЙ РАЗДЕЛ

УДК 001.89:338

П.А. Калачихин

Методология целеполагания научных исследований*

Рассмотрены ключевые особенности целеполагания в сферах экономики и управления развитием науки. Выделены уровни вертикали управления научной деятельностью. Сформулированы требования к научной стратегии. Показана важность роли прогнозирования в целеполагании науки. На основе анализа стратегических документов вынесены на обсуждение слабые места методов и показателей, используемых в целеполагании отечественной науки. Предложены либеральная и директивная схемы декомпозиции целей при переходе с одного уровня вертикали управления наукой на другой. Высказаны соображения по оптимизации выбора показателей для целеполагания научных исследований. Аргументирована взаимодополняемость целевых показателей и приоритетов в управлении наукой.

Ключевые слова: научно-технологическое развитие, стратегическое планирование, стратегия, управление наукой

DOI: 10.36535/0548-0027-2022-08-1

* Работа выполнена в рамках исследования по теме 0003-2022-0001 Госзадания ВИНТИ РАН и при поддержке Российского фонда фундаментальных исследований (проект РФФИ № 20-07-00014).

ВВЕДЕНИЕ

Обстановка в современном мире связана с экономическими трудностями, напряженными внешнеполитическими отношениями и глобальными формажорными обстоятельствами. Для того чтобы отечественная наука смогла дать достойный ответ многочисленным вызовам, необходимо повысить эффективность управления ею, в частности, обновив систему целеполагания, которая на сегодняшний день вызывает немало критики. Однако, как гласит крылатое выражение, «все новое – это хорошо забытое старое», и здесь уместны аллюзии с историческим прошлым, в котором советская наука решала важнейшие задачи народного хозяйства и обороны. Реагируя на современные кризисы и угрозы, следовало бы воспользоваться уже имеющимся опытом. Большим подспорьем в этом деле является то, что современные цифровые технологии позволяют ставить такие амбициозные цели, которые невозможно было достичь ранее.

Ориентир на оптимизацию управления наукой в сторону ужесточения процедур планирования и контроля за результатами исследований придает важность критериям отбора стратегических показателей, недостаток в точности оценок целевых значений которых мы предлагаем устранить более формальным подходом к расчетам.

Исследование показателей целеполагания, используемых в управлении научной деятельностью, требует анализа факторов, влияющих на планирование достижений. В настоящей статье мы попытаемся выполнить систематизацию таких элементов системы стратегического целеполагания, как область целеполагания и вовлеченные в целеполагание субъекты.

СТРАТЕГИЧЕСКОЕ ЦЕЛЕПОЛАГАНИЕ В ЭКОНОМИКЕ И НАУКЕ

Теория целеполагания (*goal setting*) соотносится с психологией и теорией мотивации, личной эффективностью и самоменеджментом. Успехи в данном направлении связаны прежде всего с публикациями Э.А. Локка (E.A. Locke) и Г.П. Лэйтэма (G.P. Latham). Для мотивации сотрудников частных организаций повсеместно используются ключевые показатели эффективности – *KPI (Key Performance Indicators)*. В научной сфере примером использования *KPI* является перевод научных сотрудников на эффективные контракты и внедрение показателей результативности научной деятельности. Показатели *KPI* выстраиваются в сбалансированную систему показателей оценки эффективности деятельности организаций – *BSC (Balanced Scorecard)*.

SWOT-анализ (*Strengths, Weaknesses, Opportunities, Threats* – сильные стороны, слабые стороны, возможности, угрозы), *PEST*-анализ (*Political, Economic, Social, Technological* – политические, экономические, социальные и технологические факторы) и его разновидности, а также *SNW*-анализ (*Strength, Neutral, Weakness* – сильные, нейтральные, слабые стороны) – это проверенные временем инструменты менеджмента, но применительно к науке они позволяют делать выводы лишь весьма общего характера. С точки зрения целеполагания, гораздо более конкретен *GAP*-анализ (*gap* – разрыв). Помимо этого, следует отметить ме-

тодологию *SMART (Specific, Measurable, Achievable, Relevant, Time-bound* – конкретный, измеримый, достижимый, значимый, ограниченный по времени), которая вводит ограничительные требования при определении новых целей и задач. В свою очередь, подходы *MBO (Management By Objectives* – управление по целям) и *OKR (Objectives and Key Results* – цели и ключевые результаты) дополняют методологию *SMART*.

Существуют и менее известные методы целеполагания. Так, постановка цели по методу *TASKED* подразумевает:

- указание объекта управления и общего направления его изменения, в том числе во времени (*Total*);
- согласование общей цели с подцелями управления соподчиненных субъектов управления (*Agreed*);
- декомпозицию общей цели по уровням управления (*Sectionalization*);
- описание управленческих воздействий на объект управления (*Know*);
- указание относительного изменения объекта управления в процентах (*Effect*);
- определение консолидированных расходов, направляемых на достижение результата (*Disbursement*) [1, с. 41].

Выбор методов для стратегического целеполагания на государственном уровне сильно зависит от расклада политических сил в отдельно взятой стране или регионе, который обуславливает концепцию внутреннего развития, в том числе в сфере науки. Помимо этого, оказывают влияние климатические, культурные и прочие местные особенности, относящиеся не только к странам, но и регионам, а также к интеграционным структурам.

Важным требованием к целевым показателям, используемым в стратегическом планировании науки на государственном уровне, является возможность их подсчета на основании информации, которой располагают статистические службы. Подобная информация может собираться по данным электронных платформ, а также отчетов, предоставляемых организациями в контролирующие органы.

Обозначим научных сотрудников, научно-исследовательские организации и государственный аппарат как, соответственно, *субъекты* низшего, среднего и высшего уровней (*levels*) управления наукой. В то время как целеполагание на высшем и среднем уровнях управления наукой более организовано, на низшем уровне этот процесс ограничен рамками стратегии в меньшей степени.

Стратегия – это не только обоснованные и сформулированные цели и средства их достижения, но еще и содержащий их документ. Если отвлечься от конкретных целей и способов их достижения, то нужно добиться следующих полезных свойств стратегии развития науки – она должна:

- вдохновлять исследователей, помогать им выверять актуальные направления исследований;
- быть достаточно гибкой, чтобы учитывать возможность изменения приоритетов, технологий, законодательства и объемов финансирования.

Стратегия предполагает выбор цели и средств для ее достижения. Выбор цели зависит от её ценности и

стоимости. Ценность цели обусловлена ожидаемым эффектом, который позволительно оценивать приблизительно, но зато можно точно подсчитать, сколько будет стоить достижение этой цели. Выбор средств достижения цели – это определение ресурсов для оплаты ее стоимости. Помимо этого, у цели есть ряд таких параметров, как как длительность периода достижения и целевое значение, заданное в числовом или текстовом виде.

Как правило, верную стратегию нужно выбирать из нескольких вариантов. Это могут быть: борьба за лидерство, восстановление утраченных позиций, перелом отрицательных тенденций, заимствование чужого опыта, борьба с кризисом, фокусирование на отдельных направлениях и, наконец, инертные стратегии, применяемые в отсутствие ресурсов и возможностей для развития науки. Стратегия может быть как «оборонительной», так и «наступательной».

Ситуация, когда государство и бизнес были нацелены на рост научных исследований и увеличение расходов на НИОКР, осталась в прошлом [2, с. 139]. В современном мире сложно представить «агрессивную» научную стратегию.

Цели и задачи, которые сегодня стоят перед исследователями, либо будут достигнуты в указанные сроки, либо нет. На пути к их достижению, как правило, возникают препятствия, а принятые на высшем уровне меры могут быть как успешными, так и неэффективными, поэтому научная стратегия должна учитывать различные сценарии развития событий.

Построение сценария возможно двумя способами – статистическим и вероятностным. В первом случае выявляются основные факторы, определяющие в целом характер реализации программы, строятся динамические ряды этих факторов, рассчитываются тренды и производится их экстраполяция. Во втором случае каждый из факторов представляется пессимистической, наиболее вероятной и оптимистической оценками, что увеличивает многообразие возможных состояний в будущем [3, с. 368-369].

Согласно Стратегии научно-технологического развития Российской Федерации¹, научно-технологическое развитие может осуществляться по двум сценариям:

- импорт технологий и фрагментарное развитие исследований и разработок, интегрированных в мировую науку, но занимающих в ней подчиненные позиции;
- лидерство по избранным направлениям научно-технологического развития в рамках как традиционных, так и новых рынков технологий, продуктов и услуг и построение целостной национальной инновационной системы.

При формировании долгосрочных прогнозов, корректировке сценариев и установлении целевых индикаторов следует учитывать факторы, предупреждающие о возникновении негативных тенденций, и события, имеющие низкую вероятность возникнове-

ния, но способные обернуться значительными последствиями [4, с. 6].

Следуя курсу на интеграцию с мировым научным сообществом, индикаторы стратегических документов были привязаны к международным наукометрическим базам данных *Web of Science* и *Scopus*. В ожидании развития внешних отношений с другими странами по неблагоприятному сценарию, на высшем уровне рассматривалась инициатива по пересмотру прежнего курса, предполагающая привязку индикаторов стратегических целей к крупнейшей отечественной наукометрической базе РИНЦ [5]. Отныне при построении суверенной наукометрической системы следует избегать цифровой зависимости (*digital dependence*) от зарубежных поставщиков данных.

Исправление дефектов системы целеполагания может предполагать не только регламентированную процедуру корректировки стратегии или параметров отдельных целей, но и радикальное изменение методов целеполагания.

ОБСУЖДЕНИЕ СИСТЕМЫ ЦЕЛЕПОЛАГАНИЯ ОТЕЧЕСТВЕННОЙ НАУКИ

В информационной среде есть недоверие к исходным количественным данным, на основании которых принимаются стратегические решения, или сомнения, что эти данные интерпретируются верно. Из официальных источников поступает информация о росте затрат на науку, некоторые же активные представители научного сообщества заявляют об обратном. Подобные противоречия можно объяснить устареванием информации, фальшивыми новостями (*fake news*), а также различиями в методиках подсчета объемов финансирования, например, поправками на уровень инфляции или пересчетом абсолютных величин в относительные. Так или иначе, информационный шум вносит ощутимые помехи в управление наукой.

Государственный аппарат весьма консервативен в методологических вопросах, к которым относятся выбор индикаторов для управления научной деятельностью и методик их подсчета. В значительной мере «узкие места» в системе целеполагания проведения фундаментальных исследований являются системными и унаследованы от проблем стратегического целеполагания государственной политики в целом.

Многие из действующих документов стратегического планирования устарели и утратили актуальность, некоторые – слабо увязаны между собой. Кроме того, отсутствует объективный подход к приоритизации целей, предусмотренных в различных документах стратегического планирования. При этом некоторые цели систематически не достигаются и переносятся на следующий перспективный период без тщательного анализа причин их недостижения и выработки системы мер по исправлению ошибок [6, с. 139].

Значительная часть мероприятий, хоть и не являющихся прорывными, но от этого не менее важных, и соответствующая им доля расходов не попали в документы программно-целевого планирования [7].

Вынуждены констатировать, что в стратегическом управлении отечественной наукой тиражируется единственное средство – выделить из бюджета на достижение цели некоторую часть ресурсов и назна-

¹ Указ Президента Российской Федерации от 1 декабря 2016 г. N 642 "О Стратегии научно-технологического развития Российской Федерации" – URL: <https://docs.cntd.ru/document/420384257?marker=6580ip..>

чить ответственных за ее исполнение. В качестве средств достижения стратегических целей в науке следовало бы использовать методы бюджетного и внебюджетного финансирования, льготного налогообложения и кредитования, страхования финансовых рисков, ускоренной амортизации основных фондов и нематериальных активов, лизинга научного оборудования и приборов [8, с. 67].

Из стратегических документов не вполне ясен порядок и очерченность достижения целей.

Сложившаяся государственная политика в области науки не ориентирована на конкретные научные достижения, целеполагание выстроено во многом формально, а ориентиры для выбора тематики научных исследований заданы чрезмерно широко и в основном базируются на предложениях научных организаций, у которых отсутствуют ресурсы для проведения крупных исследовательских проектов. Существующая система индикаторов не способна обеспечить востребованность научных достижений, поскольку оценивает результаты, которые поддаются имитации [9, с. 87].

Стратегические документы имеют силу прежде всего для чиновников и лиц, занимающих в науке руководящие должности. Научным работникам в большей степени нужна повестка, в соответствии с которой они могли бы выстраивать свои цели. Отсутствие заданной повестки невозможно компенсировать научными коммуникациями. В результате реформ ключевые решения в отечественной науке принимаются чиновниками, а сами ученые лишь выражают согласие или несогласие, вплоть до протеста, с решениями, принимаемыми на вышестоящих уровнях.

Отечественные ученые отмечают отсутствие ясного социального заказа для науки [10, с. 64–65].

Цели документов стратегического планирования часто не сопровождаются количественными параметрами оценки. Порядка 30% показателей стратегических документов носят качественный, трудно оцениваемый характер [11, с. 79].

Среди дефектов документов стратегического планирования следует выделить отсутствие согласованности и преемственности этих документов, а также недостижимость обозначаемых целевых ориентиров. Причиной появления таких дефектов может быть объективная ограниченность человеческих ресурсов [12, с. 51].

Целевые библиометрические индикаторы не всегда основаны на корректных и проверенных расчетах. Основными ошибками в планировании таких показателей следует считать недооценку темпов развития других государств и завышенную оценку собственных возможностей [13, с. 12].

Несмотря на то, что в целом методики оценки показателей, используемых в документах стратегического планирования науки, как правило прозрачны, методики вычисления целевых значений показателей не задокументированы.

Так, для национальной цели развития Российской Федерации "Цифровая трансформация"² подробно из-

ложены и раскрыты все методики, но непонятно, откуда берутся целевые показатели роста 125%, 200% и пр. к 2030 г. для показателя «Достижение цифровой зрелости» при том, что фактические значения показателей в ряде случаев уточняются.

Приложения с описанием методики расчета показателей есть далеко не во всех документах. Создается впечатление, что в государственной программе Российской Федерации "Научно-технологическое развитие Российской Федерации"³ механически был проставлен ряд значений от текущего значения к целевому. При этом часто бывает так, что в ближайшем отчетном году после даты утверждения документа показатель явно занижен.

Развитие науки предполагает ее рост, но многие целевые наукометрические показатели Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 годы)⁴ являются константными.

Стратегические цели управления наукой, цели научных организаций и научных работников недостаточно хорошо взаимосвязаны между собой. Передача целей между уровнями – это одно из узких мест системы целеполагания.

ФОРМИРОВАНИЕ ЦЕЛЕВЫХ ЗНАЧЕНИЙ НАУЧНЫХ ПОКАЗАТЕЛЕЙ

Цель можно видеть как желаемое состояние или предвосхищаемый результат управленческих действий, либо как систему параметров, индикаторов и ориентиров [14, с. 67]. Выбор между показателем и формулировкой в виде текста зависит от видения цели.

Согласно подходам *SMART* и *TASKED*, цель должна содержать показатель. При этом следует задать как сам количественный показатель, так и методику его оценки, а также целевое значение и сроки: «*в течение T лет показатель X должен достичь значения Y*». Если сроки не указаны явно, то они могут быть поняты из контекста стратегического документа.

Целевые значения показателей можно записывать точно или приближенно, в обычном, интервальном или нечетком (например, «достичь высокого значения») виде.

В документах стратегического планирования точные или приближенные целевые значения показателей задаются в виде временных рядов, которые способны увеличиваться:

- равномерно;
- с ускорением роста;
- с замедлением роста.

ции "Цифровая трансформация". – URL: <https://docs.cntd.ru/document/573320665>.

³ Государственная программа Российской Федерации "Научно-технологическое развитие Российской Федерации" (утв. постановлением Правительства Российской Федерации от 29 марта 2019 г. № 377) – URL: <http://static.government.ru/media/files/AAVpU2sDAvMQkIHV20ZJZc3MDqcTxt8x.pdf>.

⁴ Программа фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021 – 2030 годы) (утв. распоряжением Правительства Российской Федерации от 31 декабря 2020 г. № 3684-п). – URL: <http://www.ras.ru/fstorage/download.aspx?id=48c1fef5-a422-4bf7-8774-67e002578c0f>.

² Приказ Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации № 600 от 18.11.2020 "Об утверждении методик расчета целевых показателей национальной цели развития Российской Федера-

При расчете целевых показателей следует учитывать различные сценарии, используя ограничения на значения этих показателей.

Динамика ряда значений целевых показателей может различаться в зависимости от сценария, по которому пойдет развитие науки. Компактным способом объединения временных рядов показателей является запись целевого значения в минимаксной форме:

$$X_t \geq \min_s \{Y_{st}\} \wedge X_t \rightarrow \max_s \{Y_{st}\}, \quad (1)$$

где: X_t – фактическое значение показателя X в t -м периоде;

Y_{st} – целевое значение показателя X в t -м периоде согласно сценарию s .

В качестве предельно критического значения определяется число, превышение (при растущем показателе) или снижение (при падающем показателе) которого свидетельствует о вхождении системы в зону риска. Величину предельно критических значений необходимо устанавливать из результатов специальных исследований, экспертных оценок и математического моделирования [15, с. 20].

Формула (1) применяется в случае, если целевое значение показателя должно быть увеличено. Тогда $\min_s \{Y_{st}\}$ является критическим значением показателя X в периоде t , а $\max_s \{Y_{st}\}$ – его оптимальным значением.

Согласно Стратегии цифровой трансформации отрасли науки и высшего образования⁵ результат её реализации должен влиять на достижение национальных целей. Так, в карточке проектов Стратегии есть пункты «Связь с другими национальными целями» и «Связь с показателями национальных целей». Помимо этого, в каждом проекте есть раздел «Влияние на достижение национальных целей», который содержит колонку «Оценка влияние результата проекта на достижение показателя национальной цели (экспертная оценка)». Таким образом, цели развития верхнего уровня управления наукой увязаны между собой.

Наибольшую же сложность представляет взаимосвязь целевых показателей при понижении уровня управления наукой, где показатели вышестоящего уровня должны декомпозироваться в показатели нижестоящего уровня. В рамках трехуровневой модели управления наукой можно выделить два варианта декомпозиции целей: когда стратегическая цель декомпозируется в цель организации и когда цель организации декомпозируется в цели сотрудников.

Подобная декомпозиция может осуществляться в либеральной (мягкой) или директивной (жесткой) форме. Либеральная форма декомпозиции предполагает рекомендательный характер вышестоящих целей, при котором нижестоящие цели формулируются

по взаимной договоренности сторон. Директивная декомпозиция подразумевает прямое подчинение, когда нижестоящий уровень должен обеспечить выполнение поставленных сверху задач и отчитаться по ним.

Взаимодействие учредителей и научных организаций в рамках федеральных целевых программ укладывается в директивную схему декомпозиции целей. Государственное задание может принимать либеральный характер (если темы научных исследований формируются по принципу «снизу-вверх») или, наоборот, приобретать черты директивного инструмента [16, с. 22]. Грантовое финансирование научных проектов строится по либеральной схеме.

Согласно Стратегии научно-технологического развития Российской Федерации большие вызовы (*big challenges*) определяются как объективно требующие реакции со стороны государства совокупности проблем, угроз и возможностей, сложность и масштаб которых таковы, что они не могут быть решены, устранены или реализованы исключительно за счет увеличения ресурсов.

Целеполагание в чисто либеральной или исключительно директивной формах является крайностью, которой нужно избегать. Тем не менее, при наличии больших вызовов возможен значительный крен в сторону рассматриваемой нами далее схемы.

Целевое значение показателя на более высоком уровне должно быть достигнуто за счет суммарного вклада субъектов нижестоящего уровня. Однако такой вклад не может быть одинаковым, поскольку субъекты различаются как по специализации, так и по «мощности производства знаний» (*knowledge production capacity*).

Индикаторы, являющиеся абсолютными величинами, такие как показатели публикационной активности, декомпозируются элементарно.

Например, если поставлена стратегическая цель достичь необходимого количества публикаций, то каждая организация и каждый её сотрудник должны внести вклад в достижение общей цели:

$$Y = \sum Y_\alpha = \sum_\alpha \sum_\beta Y_{\alpha\beta}, \quad (2)$$

где: Y – запланированное число публикаций по всем организациям;

Y_α – количество публикаций, которые должна подготовить организация α ;

$Y_{\alpha\beta}$ – количество публикаций, которые должен подготовить сотрудник β организации α .

Однако на практике с показателями публикационной активности дела обстоят сложнее ввиду наличия связей между организациями и сотрудничеством авторов.

Примеры абсолютных показателей: прирост числа публикаций по результатам исследований и разработок в ведущих научных журналах, прирост числа патентных заявок, поданных по результатам исследований и разработок.

⁵ Стратегия цифровой трансформации отрасли науки и высшего образования (утв. Минобрнауки России). – URL: <https://minobrnauki.gov.ru/upload/iblock/e16/dv6edzmr0og5dm57dtm0wyllr6uwtujw.pdf?ysclid=l20eljqkxj>.

Предположим, что в стратегическом документе предлагается оптимизировать в сторону увеличения относительный показатель X , который рассчитывается так:

$$X = \frac{P}{Q}, \quad (3)$$

где X – это, к примеру, среднее количество публикаций, приходящееся на одного сотрудника,
 P – это суммарное количество публикаций;
 Q – число сотрудников.

Если требуется увеличить среднее количество X публикаций на одного сотрудника на Y процентов, то следует выполнить расчет, на сколько единиц необходимо увеличить суммарное количество публикаций P . Но тогда нужен прогноз ожидаемого количества сотрудников Q .

Примеры относительных показателей: доля статей в соавторстве с иностранными учеными в общем числе публикаций российских авторов, индексируемых в международных системах научного цитирования, доля ученых в общей численности российских и зарубежных ученых, работающих в российских организациях и имеющих статьи в научных изданиях первого и второго квартилей, индексируемых в международных базах данных, вклад в достижение места по удельному весу в общем числе статей в областях, определяемых приоритетами научно-технологического развития, в изданиях, индексируемых в международных базах данных.

С понижением уровня управления целевой показатель трансформируется в подчиненные показатели со структурными изменениями. При этом имеет значение не только размерность целевого показателя, но и его тип, который может быть наукометрическим, специальным, рейтинговым или многокомпонентным.

Общим трендом является то, что в стратегическом планировании науки сейчас преобладают целевые показатели, которые нельзя отнести к наукометрическим. Как мы полагаем, это связано с тем, что наука была перенацелена на создание инноваций, а фундаментальные исследования стали еще менее востребованными. Поскольку такие показатели, как «хиршеподобные» метрики, плохо декомпозируются и агрегируются, библиометрические индикаторы применяются в научном целеполагании еще реже, чем наукометрические. Вебометрики и альтметрики не получили широкого применения в целеполагании, поскольку легко поддаются «накрутке».

Особенность экспериментальных исследований состоит в применении специальных показателей результативности, подстроенных под конкретную область. Например, в Федеральной научно-технической программе развития синхротронных и нейтронных исследований и исследовательской инфраструктуры на 2019–2027 годы⁶ фигурирует количество создан-

ных специфических технологий, установок, метрологических методик. В отношении прикладных исследований акцент делается на патентную активность, а именно используются показатели количества полученных патентов или заявок на получение патентов.

Помимо этих показателей, к специальным относятся экономические показатели, такие как объем финансирования науки и количество занятых сотрудников в научной инфраструктуре.

Декомпозиция целевых значений рейтинговых показателей заключается в том, что главная стратегическая цель достижения желаемой позиции в рейтинговом списке влечет за собой подчиненные показатели. Если известно текущее место по заданным показателям, то возможно приблизительно оценить, насколько следует увеличить показатели, участвующие в формировании рейтинга, чтобы повысить позицию вверх на необходимое число пунктов. Такой принцип используется, когда ставится задача занять 5-е место в рейтинге мировых экономик, 10-е место в списке ведущих научных держав или обеспечить присутствие пяти университетов в списке 100 лучших университетов мира. Однако нужно учитывать, что рейтинги конкурентов также будут меняться, и дать объективный прогноз в этом случае довольно сложно.

Примеры рейтинговых показателей: место по удельному весу в общем числе статей в областях, определяемых приоритетами научно-технологического развития, в изданиях, индексируемых в международных базах данных, место по численности исследователей в эквиваленте полной занятости среди ведущих стран мира⁷.

Данные о значениях многокомпонентных показателей должны быть достаточно детализированы для диагностики отдельных компонентов, агрегированных при помощи целевой функции или более изобретательным способом. Например, это многокомпонентный показатель «достижение цифровой зрелости отраслей экономики и социальной сферы, процент», который рассчитывается помимо прочих показателей на основании индикаторов, связанных с научной и образовательной деятельностью.

Помимо количественных параметров и индикаторов, цель может быть задана как достижение в определенной научной области. Достижениями являются результаты интеллектуальной деятельности, а также успехи более общего характера, такие как ликвидация пробелов в знаниях. Оценивать степень выполнения сформулированных в текстовом виде целей надлежит поручать экспертам.

Целевые показатели задают объемы, в которых нужно производить научные знания. Целеполагание по приоритетам указывает, какие именно знания необходимо производить.

16 марта 2020 г. № 287). – URL: https://minobrnauki.gov.ru/common/upload/library/2020/08/fntp_mega.pdf?ysclid=I20fajorw.

⁷ Паспорт национального проекта «Наука» (утв. президиумом Совета при Президенте Российской Федерации по стратегическому развитию и национальным проектам (протокол от 24 декабря 2018 г. N 16)). – URL: <http://static.government.ru/media/files/vCAoi8zEXRVsuy2Yk7D8hvQbpbUSwO8y.pdf>.

⁶ Федеральная научно-техническая программа развития синхротронных и нейтронных исследований и исследовательской инфраструктуры на 2019 – 2027 годы (утв. постановлением Правительства Российской Федерации от

Планирование научно-исследовательских работ по распоряжениям руководства государства осуществляется в рамках директивной схемы. Двустороннее согласование формулировок тем государственных заданий соответствует либеральной схеме.

Список приоритетных направлений развития отечественной фундаментальной науки составляется на основании экспертных опросов и консультаций с учеными и специалистами. При этом в основе таких списков лежит анализ приоритетов наиболее технологически развитых стран мира с поправкой на отечественную специфику [17].

Согласно Стратегии научно-технологического развития Российской Федерации приоритетами следует считать те направления, которые позволяют получать научные и научно-технические результаты и создавать технологии, представляющие собой основу инновационного развития внутреннего рынка продуктов и услуг, устойчивого положения России на внешнем рынке.

Важно отметить, что приоритеты развития науки носят качественный характер и формулируются в текстовом виде, однако выявляются они не только экспертными, но и формальными методами.

Так, приоритетные направления исследований можно определять за счет выявления исследовательских фронтов методами:

- анализа публикационной активности (*analysis of publication activity*);
- прямого анализа цитирований (*direct citation analysis*);
- анализа со-цитирований (*co-citation analysis*);
- библиографического сочетания (*bibliographic coupling*);
- контент-анализа (*content analysis*);
- кластеризации ссылок (*citation topics*) [18, 19].

Таким образом приоритеты развития науки декомпозируются в темы государственных заданий и далее перетекают в заголовки публикаций.

Поскольку руководство зачастую способно только упрощенно формулировать, какие именно характеристики полезных свойств достижений предстоит получить, то окончательное определение научных достижений происходит на исполнительных уровнях.

ЗАКЛЮЧЕНИЕ

Можно выделить низший – индивидуальный, средний – организационный и высший – государственный уровни управления наукой, каждый из которых требует особого подхода к целеполаганию.

На выбор научной стратегии государством оказывают влияние внутренние и внешние политические факторы. Большинство современных научных стратегий являются оборонительными, различаясь лишь по мере жесткости антикризисных мер. Научная стратегия государства должна быть основана на прогнозах и учитывать различные сценарии развития событий.

По поводу сложившейся системы стратегического целеполагания в отечественной науке разные авторы высказывают критические замечания. Многие проблемы отечественной науки унаследованы от узких мест используемой общей методологии управления

государством. Наибольшие трудности возникают в процессе перехода от стратегических намерений к действиям.

При переходе между высшим и средним, средним и низшим уровнями управления можно выделить либеральный и директивный способы декомпозиции научных целей. В условиях больших вызовов директивная схема целеполагания наиболее предпочтительна.

Нижестоящие субъекты вносят неравнозначный вклад в поставленные цели. Для директивного целеполагания показатели нужно подбирать так, чтобы их проще было «разворачивать» в показатели нижестоящего уровня.

Целевые значения количественных показателей необходимо вычислять, а не «прикидывать». Для целеполагания научных достижений, сформулированных в неявном виде, следует использовать экспертные показатели и технологии.

В управлении наукой не приветствуется внедрение не прошедших апробацию экспериментальных методик. Тем не менее, следует уходить от «методологического минимализма», чтобы научные цели не казались такими «невзрачными».

Трехуровневую модель управления наукой можно расширить такими дополнительными уровнями, как: надгосударственные структуры, подразделения организаций и коллективы сотрудников.

Изложенные в настоящей статье соображения, на наш взгляд, могут быть полезными при формировании наборов индикаторов стратегий, программ и проектов по развитию науки.

СПИСОК ЛИТЕРАТУРЫ

1. Тхориков Б.А. Методология индикативного управления организациями социальной сферы: проблемы целеполагания, метод TASKED // Экономика. Информатика. – 2015. – Т. 36, № 19(216). – С. 39-42.
2. Артамонова Ю.Д., Демчук А.Л., Карнеев А.Н., Сафонова В.В. Современные стратегии развития науки и подготовки научных кадров: международный опыт // Высшее образование в России. – 2018. – № 4. – С. 135-148.
3. Глухов В.В., Коробко С.Б., Маринина Т.В. Экономика знаний: учеб. пособие. – Санкт-Петербург: Питер. – 2003. – 528 с.
4. Брутян М.М. Проблема прогнозирования и мониторинга целевых индикаторов создания научно-технического задела на примере авиастроения // Интернет-журнал «Науковедение». – 2017. – Т. 9, № 4.
5. Алиев Д.Ф. Современная наукометрия: вызовы и возможности. – 2022. – URL: https://rgsu.net/netcat_files/multiUpload/%20D.F%20SOVREMENNAYa%20NAUKOMETRIYa%20VYZOVY%20I%20VOZMOZHNOStI%20itogovyi%CC%86.pdf.
6. Наумов С.Н. Проблемы целеполагания в государственных программах // Редакционная коллекция. – 2019. – Т. 29. – С. 138-143.
7. Ярёмченко И.А. Национальные проекты, госпрограммы, национальные цели. Как они скоординированы // Метод. – 2019. – № 4. – С. 40-43.

8. Годосийчук А.В. Экономический механизм реализации государственной научно-технической политики // Научно-технические исследования. – 2021. – № 2021. – С. 66-90. DOI: 10.31249/scis/2021.00.04.
9. Шепелев Г.В. Об управлении российской наукой // Управление наукой: теория и практика. – 2020. – Т. 2, № 2. – С. 65-92. DOI: 10.19181/sntp.2020.2.2.3.
10. Юревич А.В. Стратегии развития российской науки // Социология науки и технологий. – 2010. – Т. 1, № 1. – С. 52-66.
11. Соколов И.А., Дрожжинов В.И., Райков А.Н., Куприяновский В.П., Намиот Д.Е., Сухомлин В.А. Искусственный интеллект как стратегический инструмент экономического развития страны и совершенствования ее государственного управления. Часть 2. Перспективы применения искусственного интеллекта в России для государственного управления // International Journal of Open Information Technologies. – 2017. – Т. 5, № 9. – С. 76-101.
12. Турганбаев А.О. Применение искусственного интеллекта в стратегическом планировании в государственном управлении // Государственная служба. – 2019. – Т. 21, № 5(121). – С. 49-54. DOI: 10.22394/2070-8378-2019-21-5-49-54.
13. Стародубов В.И., Кузнецов С.Л., Куракова Н.Г., Цветкова Л.А., Арефьев П.Г., Иванов А.В., Еремченко О.А. Публикационная активность российской медицинской науки в фокусе актуальной научной политики: оценка достижимости целевых показателей // Вестник Российской академии медицинских наук. – 2013. – Т. 68, № 3. – С. 8-14. DOI: 10.15690/vramn.v68i3.595.
14. Слинкова О.К., Скачков Р.А. Анализ научных подходов к определению сущности целеполагания и его роль в процессе управления // Известия Уральского государственного экономического университета. – 2015. – № 3(59). – С. 65-72.
15. Глазьев С.Ю. О создании систем стратегического планирования и управления научно-техническим развитием // Инновации. – 2020. – Т. 2, № 256. – С. 14. DOI: 10.26310/2071-3010.2020.256.2.002.
16. Гусев А., Ладный А., Белоклоков Е., Юревич М. О государственном задании в сфере науки // Общество и экономика. – 2017. – № 11. – С. 5-23.
17. Мухутдинова Т.З. Приоритетные направления развития науки, технологий и техники Российской Федерации и критические технологии федерального уровня: История разработки и динамика развития // Вестник Казанского технологического университета. – 2012. – Т. 15, № 14. – С. 247-254.
18. Мазов Н.А., Гуреев В.Н. Проблемные аспекты при выявлении приоритетных направлений научных исследований // Управление наукой: теория и практика. – 2020. – Т. 2, № 3. – С. 37-51. DOI: 10.19181/sntp.2020.2.3.2.
19. Green B. InCites citation topics. – 2021. – URL: <https://wok.mimas.ac.uk/support/documentation/presentations/citationtopics202101.pdf>.

Материал поступил в редакцию 03.06.22

Сведения об авторе

КАЛАЧИХИН Павел Андреевич – кандидат экономических наук, ведущий научный сотрудник ВИНТИ РАН, Москва
e-mail: pakalachikhin@viniti.ru, studypavel@mail.ru

Тренды публикационной активности российских организаций в 2000–2019 гг.

Неоднократно отмечалось, что количество публикаций российских исследователей быстро росло в 2012–2021 гг., однако этот рост был неравномерен как в количественном, так и в качественном отношении. Более того, в некоторых научных организациях публикационная активность не только не росла, но даже подала. Для полноценного понимания реального развития научного сектора необходим анализ динамики десятков показателей для сотен организаций. Однако как исследовать столь многочисленные ряды данных?

Для решения этой задачи используются метод классификации трендов временных рядов, основанный на комбинации из трех статистических критериев случайности или отсутствия тренда: критерий Кокса-Стюарта, критерий инверсий и критерий Манна-Кендалла. С его помощью для каждой из 701 ведущей российской научной организации и университета были классифицированы ряды наукометрических параметров, характеризующих различные аспекты публикационной активности, включая комплексный публикационный балл (аналог КБПР), средний уровень качества публикации и долю статей в сборниках научных трудов конференций. Показаны различия между трендами организаций разного типа и категорий результативности. На их основе сформулированы новые исследовательские вопросы, связанные с причинами возникновения тех или иных тенденций развития российской научной сферы.

Ключевые слова: наукометрия, публикационная активность, фракционный счет, Scopus, проект 5-100, критерии случайности или отсутствия тренда, КБПР, критерий Кокса-Стюарта, критерий инверсий, критерий Манна-Кендалла

DOI: 10.36535/0548-0027-2022-08-2

ВВЕДЕНИЕ

Публикационная активность российских исследователей за последние десять лет претерпела серьезные изменения. В стране была выстроена научная политика, ориентированная на повышение как количества, так и качества научных результатов. Динамика количества российских публикаций, индексируемых в международных научных базах данных *Web of Science* и *Scopus*, стала заметно опережать среднемировую, что привело к двукратному увеличению их доли в мировом публикационном потоке (*Web of Science*: с 1,5% в 2012 г. до 2,7 в 2020 г.; *Scopus*: с 1,6% до 3,6% соответственно). Ряд исследователей посвятили свои работы изучению российского «публикационного взрыва». Анализу и оценке подверглись достижения российской науки с точки зрения публикационной активности [1], результатов Проекта 5-100 [2–6], коллабораций [2, 7, 8], миграции ученых [9, 10, 11]. Поднимались также вопросы о метриках оценки результа-

тивности научных организаций [12–14] и проблемах в академической экспертизе [15].

Обзор литературы по исследуемой нами теме показал, что аномальный рост публикационной активности научных организаций неравномерен как в количественном, так и в качественном отношении. Когда одни научные организации выходят вперед за счёт внешнего стимулирования и собственного кадрового потенциала, не акцентируя внимание на качестве публикационного потока, другие, при общем увеличении числа публикаций, пытаются этот поток сохранить и демонстрируют более консервативные модели развития. Более того, могут существовать группы организаций, в которых, несмотря на общий рост публикационной активности, она падает по субъективным или объективным причинам. Эти особенности следует учитывать при изменении научной политики, процедур оценки и мер стимулирования. Выявление организаций с типичными и нетипичными трендами развития является актуальной задачей.

В наукометрии такие работы встречаются нечасто из-за того, что многие специалисты предпочитают сводить исследования к рассмотрению небольшого количества числовых рядов, поскольку их легче анализировать и визуализировать. При изучении большого числа объектов (например, университеты страны) обычно рассматриваются их совокупные показатели, но не тренды развития каждой организации. Отчасти по этой причине наблюдается дефицит сложившихся практик алгоритмической обработки временных рядов наукометрических данных. Поэтому основная цель настоящего исследования заключается в заполнении этого пробела и апробации метода определения трендов публикационной активности на примере массива наукометрических показателей российских организаций.

В настоящей статье анализируются временные ряды наукометрических показателей организаций, разделенных статистическими методами по их типу (рост/нет тренда/падение). Классификация временных рядов строилась нами на основе результатов проверки статистических гипотез о случайности или отсутствии тренда по нескольким критериям с помощью метода, изложенного в [16].

Для изучения было отобрано несколько наукометрических показателей, характеризующих особенности публикационной активности организаций. Помимо числа публикаций, полученных целочисленным счетом, был использован метод фракционного счёта, суть которого состоит в определении веса публикаций, относящихся к конкретной организации, пропорционально числу авторов и их аффилиаций, указанных в публикациях [12, 17]. Для изучения качественной характеристики публикационного потока использовался показатель, близкий к Комплексному баллу публикационной результативности [13]. Ещё одним важным показателем является доля публикаций, подготовленных с внешним сотрудничеством¹, поскольку это значительно повышает шансы на их цитируемость [18, 19], а также на размещение таких статей в более рейтинговых журналах [20, 21].

Многие из указанных исследований выявили резкий рост публикационной активности вузовского сегмента, что объясняется в том числе и «конференционным взрывом»: беспрецедентным для российской науки увеличением числа публикаций в сборниках материалов конференций. «К 2020 г. этот тренд принял угрожающие масштабы и требует анализа и скорейшей реакции во избежание ряда серьёзных негативных последствий» [22]. Разделяя эти опасения, мы включили два соответствующих показателя в наше исследование.

Отметим, что анализ трендов публикационной активности не является самодостаточным инструментом, поскольку для обоснования большинства гипотез необходимо применять иные аналитические методы, детализирующие причины возникновения,

¹ Под внешним сотрудничеством, здесь и далее, понимается участие в публикации авторов из разных организаций.

изменения или сохранения этих трендов. В настоящем исследовании мы сознательно ограничиваемся применением одного методического инструмента, и поэтому вынуждены воздержаться от обсуждения исследовательских вопросов, связанных с причинами возникновения тех или иных тенденций. Вместо этого, полученные результаты позволяют нам сформулировать вопросы (в тексте они будут выделены *курсивом*), а ответы на них могут быть получены в последующих работах.

ДАнные И МЕТОды ИССЛЕДОВАНИЯ

Данные. Информация о публикационной активности организаций за период с 2000 по 2019 гг. была получена нами из базы данных *Scopus* с использованием *Scopus Search API*. В ней аффилиациям авторов присваиваются уникальные идентификаторы, однако, в связи различными написаниями названий организаций и их перевода или транслитерации на английский язык количество таких «уникальных» организаций за указанный период превышает 59 тысяч. У многих организаций формируются дублирующие профили, в связи с этим для корректного анализа требуется очистка и дедупликация полученных баз данных [23]. Проблема усугубляется активными процессами трансформации академических институтов и организаций высшего образования, многочисленными их объединениями и ликвидациями. Процесс дедупликации был основан на сопоставлении аффилиаций из *Scopus* с идентификационным номером налогоплательщика (ИНН) организаций, что происходило преимущественно в ручном режиме. В связи с этим «длинный хвост» аффилиаций с незначительным количеством публикаций был отброшен. Не были включены в анализ и «зонтичные» организации – Российская академия наук, ее региональные отделения, а также федеральные министерства и агентства, появление которых связано преимущественно с некорректным или неполным определением аффилиационных данных. Чтобы исключить учреждения, которые были ликвидированы или присоединены к другим, мы проанализировали данные организаций, имеющих хотя бы одну публикацию в 2017–2019 гг. Из рассмотрения были исключены организации, имеющие невысокую публикационную активность (менее 100 публикаций за 20 лет), и академические институты, получившие 3-ю категорию в рамках национальной оценки публикационной результативности.

Окончательный набор данных содержал информацию о публикационной активности 701 организации науки и высшего образования, среди которых оказалось 289 академических институтов, ранее подведомственных РАН, РАМН или РАСХН), 296 высших учебных заведений и 116 других научных организаций, в том числе имеющих отношение к Национальному исследовательскому центру «Курчатовский институт», и подведомственных департаменту здравоохранения Москвы, Роскосмосу, Роспотребнадзору и ряду других министерств и ведомств. В табл. 1 представлены 6 групп организаций.

Группы организаций, включенных в исследование

Код	Количество организаций	Описание
АИ-1	155	Академические организации, получившие 1-ю категорию результативности по итогам оценки в 2016-2018 гг.
АИ-2	134	Академические организации, получившие 2-ю категорию результативности по итогам оценки в 2016-2018 гг.
ВУЗ-1	23	Вузы, участвовавшие в «Проекте 5-100», а также МГУ и СПбГУ
ВУЗ-2	92	Вузы, включенные в программу «Приоритет 2030», не входящие в группу ВУЗ-1
ВУЗ-3	181	Остальные вузы
ДР	116	Прочие организации

Таблица 2

Используемые метрики публикационной активности и их обозначения

Тип	Метрика	Назначение
1	AC	Количество публикаций целым счетом
	FC	Количество публикаций фракционным счетом
	FS	Комплексный публикационный балл
2	$CC = 1 - FC / AC$	Коэффициент коллабораций
	$CQ = FS / FC$	Коэффициент «качества»
3	CFC	Количество публикаций в сборниках трудов конференций фракционным счетом
	$CSH = CFC/FC$	Доля публикаций в сборниках трудов конференций

Таблица 3

Весовые коэффициенты публикаций в соответствии с квартилем журнала в JCR

Тип журнала	Без квартиля	Q4	Q3	Q2	Q1
Весовой коэффициент	1	2,5	5	10	20

Для анализа применялись три типа специальных метрик публикационной активности (табл. 2). Метрики первого типа оценивали общую публикационную активность организаций, базовыми здесь стали: общее число публикаций (AC) организации за год и количество публикаций фракционным счетом (FC). Метрика FC представляет собой сумму всех долей публикаций определенной организации за год; метрика FS – аналог комплексного балла публикационной результативности (КБПР) в версии для всех направлений, кроме гуманитарных и общественных наук. Для вычисления FS источники (журналы) из БД *Scopus* сравнивались с данными *Journal Citation Report* по ISSN, и им был сопоставлен наилучший квартиль по этому рейтингу. Необходимо отметить, что расчетные FS не соответствуют в полной мере КБПР, так как в расчете не учтены публикации, не проиндексированные в БД *Scopus*, в том числе часть журналов, входящих в *Emerging Sources Citation Index*, около половины журналов из *Russian Science Citation Index* и большая часть журналов из списка ВАК. Эта метрика вычисляется аналогично фракционному количеству

публикаций с добавлением каждой публикации весового коэффициента за квартиль журнала (табл. 3).

Метрика второго типа – коэффициент коллабораций (CC) демонстрирует среднюю фракционную долю участия других организаций в публикационной активности рассматриваемой организации за конкретный год. Например, у некоторой организации *A* за определенный год было 5 публикаций, из которых только одна подготовлена во внешней коллаборации с организацией *B* и имеет фракционную долю 0,5. Получается, если $AC = 5$, $FC = 4,5$, $FC / AC = 0,9$ и $CC = 0,1$, то это означает, что 1/10 часть публикаций организации *A* досталась другим организациям (в данном примере организации *B*). Эта метрика принимает значения от 0 до 1 (за исключением самого значения 1), где 0 – означает отсутствие внешней коллаборации у организации, а близкие к 1 значения – что у организации крайне низкая фракционная доля, что возможно только при публикации статей исключительно в больших коллаборациях. Коэффициент «качества» (CQ) представляет собой средний балл за публикацию рассматриваемой организации и принимает значения от 1 до 20 (характеризуется шкалой из табл. 3).

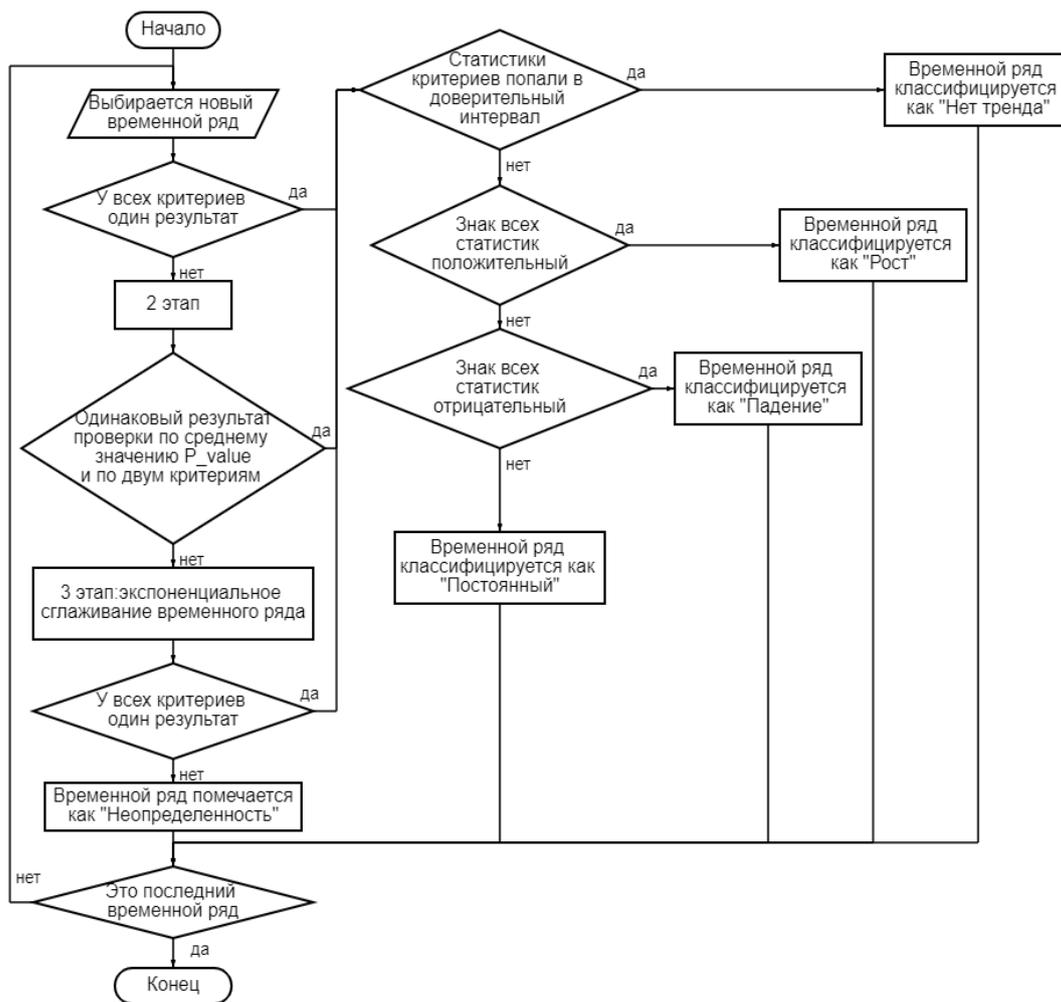


Рис. 1. Упрощенная блок-схема алгоритма классификации совокупной проверки [16].

Метрики третьего типа предназначены для оценки публикационной активности организаций в сборниках материалов конференций. К этому типу метрик относятся: количество публикаций в сборниках материалов конференций фракционным счетом (CFC) и доля таких публикаций от их общего числа (CSH).

Методы. Для каждой организации и для каждой метрики за период с 2000 по 2019 гг. нами были построены временные ряды разной длины, так как отсчет для каждой организации начинается не строго с 2000 г., а с года первой публикации любого типа за рассматриваемый период. Сделано это для того, чтобы в растущей подгруппе не оказались организации, начавшие публиковать свои разработки с опозданием (по сравнению с нулем стабильное число публикаций может быть опознано как рост).

Каждый временной ряд мы проверяли на наличие или отсутствие тренда публикационной активности в математическом ожидании. Под трендом понимается основная тенденция изменения значений ряда, направление преимущественного движения исследуемой переменной (в данном случае, одной из метрик).

В нашей работе временные ряды научных организаций классифицируются совокупной проверкой по

нескольким критериям случайности или отсутствия тренда, которые в англоязычной литературе именуется как *Trend Test*. Мы использовали три вида изменений: критерий Кокса-Стюарта [24], критерий инверсий [25, 26] и критерий Манна-Кендалла [27, 28], которые являются наиболее мощными при проверке относительно конкурирующих гипотез – наличие тренда в математическом ожидании [26, 29, 30]. Каждый отдельный критерий проверяет гипотезу при величине ошибки первого рода (уровне значимости) $\alpha = 0.1$.

Цель совокупной проверки – однозначно классифицировать временной ряд. В зависимости от поведения статистик критериев временные ряды классифицируются как возрастающие, убывающие, безтрендовые и неопределенные. Совокупная проверка состоит из трех последовательных этапов: на первом классифицируется большая часть временных рядов, классификация происходит, если все критерии приходят к одному выводу (вероятно, есть тренд или, вероятно, тренда нет); на втором и третьем этапах последовательно классифицируются оставшиеся неклассифицированными на предыдущем этапе временные ряды. На втором этапе упрощаются условия классификации, он ориентирован на классификацию

временных рядов, которые не были классифицированы одним из критериев из-за погрешностей. На третьем этапе методом экспоненциального сглаживания устраняется дискретность данных и повторяется проверка, как и на первом этапе. Временные ряды, которые не были классифицированы ни на одном из этапов, помечаются как «неопределенность». Классификация таких рядов потенциально будет нести в себе большое число ошибок из-за недостатка данных. На рис. 1 представлен общий алгоритм классификации совокупной проверкой (рис. 1.).

ОБЩИЕ РЕЗУЛЬТАТЫ КЛАССИФИЦИРОВАНИЯ

Результаты применения алгоритма для временных рядов каждой организаций по типу полученной классификации и по каждой метрике представлены в табл. 4. На рис. 2 показано соотношение классифицируемых рядов для каждой из рассмотренных нами метрик.

Почти у 90% организаций наблюдается рост числа публикаций целочисленным счётом, а у 77% – фракционным счетом. Таким образом, несмотря на общий ярко выраженный рост публикационной активности, до 5% организаций имеют убывающие тренды, а ещё почти в 20% случаев заметных изменений не фиксируется. В связи с этим следует сформулировать вопрос:

Чем характерны такие организации и с чем связано отсутствие роста публикационной активности в каждой из них: некачественное управление, отсутствие необходимых ресурсов, достижение собственного «потолка» возможностей или внешние факторы?

Коэффициент коллаборации у 60% организаций также растет, что согласуется с мировыми тенденциями, когда все чаще исследовательские задачи требуют объединения усилий двух и более коллективов авторов из разных организаций. В связи с этим возникает вопрос:

По каким причинам у оставшихся 40% организаций ожидается расширение коллабораций не происходит?

Количество организаций с растущим и убывающим показателем CQ примерно одинаково и составляет около 20%, при этом более чем у половины из них он существенно не меняется. Споры о падении качества публикаций при резком росте их количества идут в России на протяжении последних 10 лет. Распределение показателя CQ демонстрирует стакан, который наполовину пуст, или наполовину полон. Нам кажется более корректным говорить не о падении, а о *размывании* качества публикаций. Оно происходит на фоне увеличения публикационной активности, когда к стабильному потоку «сильных» статей, добавляется увеличивающийся поток публикаций более низкого качества.

Таблица 4

Распределение организаций по классификациям разных метрик

Тип метрик	1			2		3	
	AC	FC	FS	CC	CQ	CFC	CSH
Возрастающие	618	539	545	421	146	343	217
Нет тренда	51	102	94	227	368	233	346
Убывающие	17	32	30	17	149	34	35
Неопределенность	15	28	32	36	38	91	103



Рис. 2. Соотношение временных рядов для каждой рассматриваемой метрики, %

Является ли этот феномен следствием административных решений и принятых в организациях политик поощрения публикационной активности? Какие обстоятельства позволяют организациям поднимать уровень или сохранять высокое качество своих научных работ?

Показатели абсолютного и относительного количества статей в сборниках материалов конференций интересны для изучения феномена «российского конференционного взрыва» [22]. Около трети организаций увеличили долю таких публикаций, что обычно происходит за счет статей в научных журналах – более значимых видов публикаций.

Может ли со временем большое количество публикаций в сборниках материалов конференций перерасти в качественные статьи в высокорейтинговых журналах, и если да, то при каких условиях?

С другой стороны, у более чем трети организаций число публикаций в трудах конференций не изменилось или даже снизилось.

Какие организации увеличили свои публикационные показатели, не меняя практики участия в конференциях?

Отметим, что в третьем типе метрик находится наибольшее количество неопределенных трендов (14%), которые попадали в эти ряды из-за очень маленькой размерности, сильных флуктуаций, а также из-за значительных периодов роста и падения.

СРАВНЕНИЕ ТРЕНДОВ ПУБЛИКАЦИОННОЙ АКТИВНОСТИ ПО ГРУППАМ ОРГАНИЗАЦИЙ

Проанализируем группы организаций, имеющие одинаковые тренды.

Различия между классификациями по разным метрикам могут проиллюстрировать существующие стратегии в развитии публикационной активности.

В табл. 5 представлены соотношения организаций разных групп, попавших в подмножества, образуемые различными классификациями трендов по трем метрикам (АС, FC, FS) общей публикационной активности. Такая группировка метрик интересна ещё и потому, что сочетание роста публикационной активности по АС и отсутствия тренда или падения по FC указывает на увеличение степени коллаборации, а несоответствие трендов по FC и FS может быть свидетельством преобладания качества публикаций над количеством или наоборот.

Большее количество организаций (502 из 701), в том числе 263 вуза и 199 академических институтов) находится в подмножестве № 1 (см. табл. 5) одновременного роста всех метрик. Доля вузов (большинство из которых подведомственны Минобрнауки) в этом подмножестве составляет более 90%, а академических институтов (до 2014 г. подведомственных государственным академиям наук, с 2014 по 2018 гг. – ФАНО России) – 67-70%. Лишь 34% организаций группы ДР (относящихся к иным министерствам и ведомствам) вошли в это подмножество. Отсюда можно сделать вывод, что рост публикационной активности организации в значительной степени зависит от её ведомственной принадлежности, в рамках которой определяются степень важности показателей и меры по стимулированию их достижения.

Подмножества № 2 и № 3 (см. табл. 5) являются наиболее показательными примерами разных стратегий организаций в условиях вероятного ограничения ресурсов. Наше предположение состоит в том, что эти организации нашли возможности для роста публикационной активности, однако их не хватило для того, чтобы объемы и качество научной продукции монотонно росли.

Таблица 5

Распределение организаций по классификациям метрик АС, FC и FS («↑» – растущий тренд, «↓» – падающий тренд, «0» – тренд отсутствует)

№	АС	FC	FS	Все, %	АИ-1, %	АИ-2, %	ВУЗ-1, %	ВУЗ-2, %	ВУЗ-3, %	ДР, %
1	↑	↑	↑	71,6	69,7	67,9	100,0	93,5	85,1	34,5
2	↑	↑	0	3,7	1,9	5,2	–	–	3,9	7,8
3	↑	0	↑	4,4	8,4	9,7	–	–	–	4,3
4	↑	0	0	2,4	2,6	0,7	–	1,1	2,2	6,0
5	↑	0	↓	0,3	0,6	–	–	–	–	0,9
6	↑	↓	↑	0,1	0,6	–	–	–	–	–
7	↑	↓	↓	0,3	–	–	–	–	–	1,7
8	0	0	↑	0,6	0,6	2,2	–	–	–	–
9	0	0	0	3,6	4,5	3,0	–	–	2,2	8,6
10	0	0	↓	0,6	0,6	–	–	–	–	2,6
11	0	↓	0	1,0	–	1,5	–	–	0,6	3,4
12	0	↓	↓	0,6	0,6	–	–	–	0,6	1,7
13	↓	↓	0	0,1	–	–	–	–	–	0,9
14	↓	↓	↓	2,1	–	0,7	–	–	–	12,1
-	Остальные			8,6	9,9	9,1	–	5,4	5,5	15,5
	Всего			701	155	134	23	92	181	116

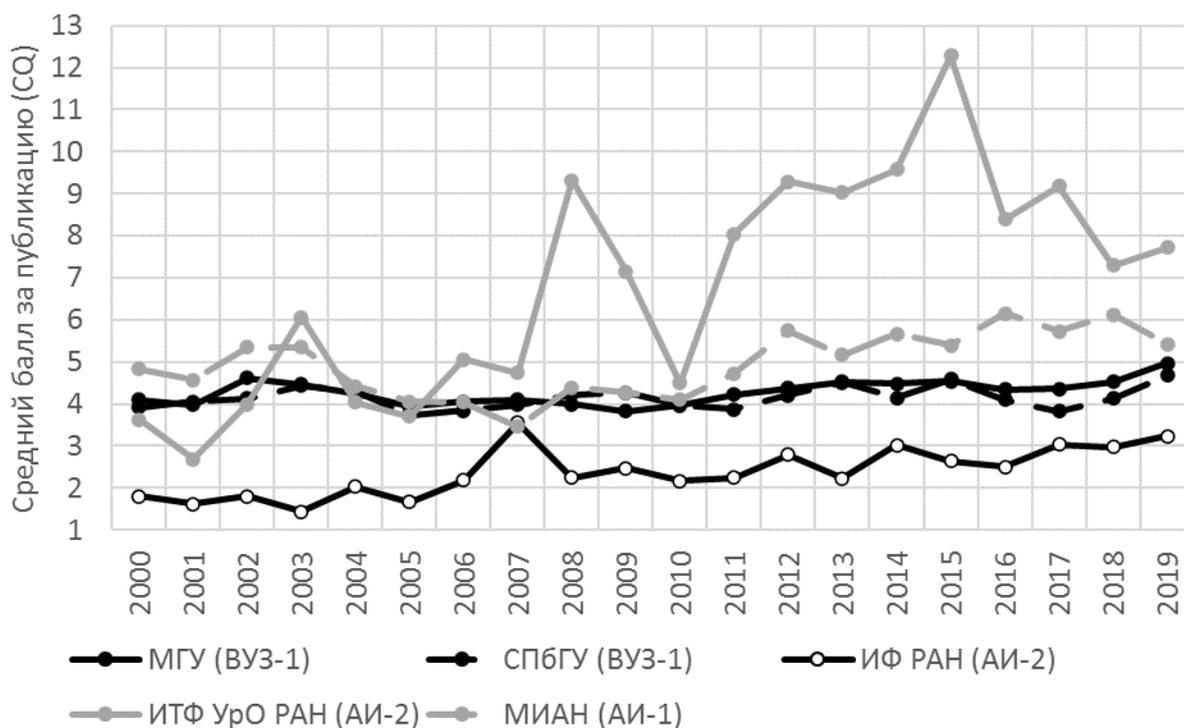


Рис. 3. Средний балл за публикацию у некоторых организаций по годам

В первом случае рост числа публикаций по АС и FC и его отсутствие по FS свидетельствуют о заметном падении среднего комплексного балла публикационной активности (КБПР) на одну публикацию. Такие организации стали публиковать больше статей, но в низкорейтинговых журналах и в сборниках материалов конференций. Второй случай является признаком того, что данные организации стали публиковать свои работы преимущественно в журналах более высокого квартиля – фракционное число публикаций (FC) осталось неизменным, а КБПР (FS) растет.

Подмножества 4-7 (см. табл. 5) характеризуются значительно возросшей долей коллабораций, среди которых можно найти организации, часто публикующиеся в мегаколлаборациях. Самой крупной организацией в подмножестве № 4 является НИЦ Курчатовский институт, а в подмножестве № 7 – подведомственный ему Институт физики высоких энергий и Институт теоретической и экспериментальной физики имени А.И. Алиханова. В подмножество № 5 попало только два института: Петербургский институт ядерной физики Курчатовского института и Институт ядерной физики им. Г.И. Будкера СО РАН. Обе эти организации занимаются физикой высоких энергий и активно участвуют в мегаколлаборациях, организованных на базе Большого адронного коллайдера (БАК).

Поскольку участники мегаколлабораций больше всех страдают от фракционного подсчета публикационной результативности, следует ли для них предусмотреть отдельные процедуры оценки или смягчить влияние существующих?

Подмножество № 8 (см. табл. 5) содержит только четыре академических института и характеризуется

ростом среднего балла за публикацию аналогично подмножеству № 3. Отличие заключается только в отсутствии роста числа публикаций целым счетом. Самым крупным в этом подмножестве является Институт физиологии им. И.П. Павлова РАН, а самый значительный рост среднего балла зафиксирован у Института теплофизики УрО РАН (средний балл за публикацию у этих организаций представлен на рисунке 3).

В подмножестве полной стабильности № 9 (см. табл. 5) большинство составляют академические институты. По числу публикаций наиболее крупными представителями являются – Институт спектроскопии РАН, Институт физики металлов УрО РАН, ФНИЦ эпидемиологии и микробиологии им. Гамалеи, Институт высшей нервной деятельности и нейрофизиологии РАН.

В подмножество № 10 (см. табл. 5) попали Физико-энергетический институт им. А. И. Лейпунского, Институт «Гиредмет» и Институт физики твердого тела РАН. У организаций подмножества № 11 средний балл за публикацию растет, компенсируя падение FC. Среди таких организаций: Институт теоретической и экспериментальной биофизики РАН, Институт высокомолекулярных соединений РАН, Санкт-Петербургский государственный технологический институт. В подмножество № 12, демонстрирующее падение количества публикаций фракционным счетом и КБПР при сохраняющемся общем уровне публикационной активности, попали Институт теоретической физики РАН и Московский независимый университет. Единственной организацией в подмножестве № 13 является АО Радиевый институт имени В.Г. Хлопина. Подмножество № 14 состоит из организаций группы «другие», за ис-

ключением Института физических проблем им. П.Л. Капицы РАН из «АИ-2».

Отметим, что распределение в группах АИ-1 и АИ-2 различается очень незначительно, т. е. научные организации первой и второй категории демонстрируют очень близкие тренды (возможно, в одной из групп эти тренды выражены более ярко, однако такой анализ необходимо проводить другими инструментами).

Отдельный интерес представляют научные организации, которые относятся к лидерам и обладают нехарактерными трендами. Так, из 155 институтов 1-й категории, лишь 24 не имеют роста публикационной активности по показателю FS (КБПР). Удивительно, что 15 из них имеют физический профиль, а 12 – нерастущий тренд по показателю FC (число публикаций фракционным счётом): АФИ, ИЛФ СО РАН, ИОФ РАН, ИСАН, ИТПЭ РАН, ИТФ РАН, ИФВД РАН, ИФМ УрО РАН, ИФП СО РАН, ИФТТ РАН, ИЯФ СО РАН, ФИЦ ХФ РАН.

Следующие три института обладают растущим трендом по FC, но без роста КБПР: САО РАН, ФИАН, ИПА РАН. Следует отметить, что 1-я категория была присвоена и ряду других организаций, выполняющих исследования в области физических наук, у которых был зафиксирован растущий тренд публикационной активности по показателю КБПР.

Физика – одна из основных специализаций в нашей стране, в перечисленных организациях работают признанные в мире эксперты. Так, наблюдается противоречие: экспертное сообщество считает лидерами те организации, у которых не наблюдается увеличение КБПР, однако такой рост был заложен Минобрнауки в методику оценки эффективности на уровне 10% ежегодно [13].

Следует ли утверждать, что отсутствие роста формальных показателей необязательно свидетельствует о стагнации организации, так как оно может быть объяснено стабильно высоким количеством публикаций на протяжении всего периода наблюдений, достижением потолка ресурсных ограничений или расширением участия в междисциплинарных исследованиях и мегаколлаборациях, которые не дают заметного прироста количества публикаций фракционным счётом?

СРАВНЕНИЕ ТРЕНДОВ ПУБЛИКАЦИОННОЙ АКТИВНОСТИ ПО ПОКАЗАТЕЛЯМ КОЭФФИЦИЕНТОВ КОЛЛАБОРАЦИЙ И «КАЧЕСТВА»

В табл. 6 представлены результаты классификации показателей публикационной активности, характеризующих «качество» и уровень коллабораций.

У организаций подмножества №1 (см. табл. 6) доля коллабораций увеличивалась одновременно с ростом среднего «качества» публикаций (метрика CQ). 70% этого подмножества составляют академические институты, причем академических институтов 1-й категории заметно больше (49 из 103). Наиболее крупной организацией в данном подмножестве является МГУ им. М.В. Ломоносова. Наибольший рост среднего балла на публикацию был обнаружен у Института физиологии растений им. К.А. Тимирязева РАН и ФНЦ физико-химической медицины ФМБА. Доля коллаборации выросла сильнее всего (в 9 раз) у НИИ вакцин и сывороток им. И.И. Мечникова (коэффициент «качества» вырос в 4 раза).

Поскольку в этом подмножестве наиболее выделены академические институты и университеты 1-й группы, следует ли считать комбинацию этих двух метрик (коэффициенты коллабораций и «качества») хорошим предиктором организаций, претендующих на научное лидерство?

Подмножество № 2 составляют организации, которые сохранили «качество» публикаций и увеличили степень сотрудничества с другими организациями. В большинстве случаев у этих организаций число публикаций также имеет восходящий тренд. Это самая большая группа – количество академических институтов и вузов в ней примерно одинаковое. СПбГУ наиболее крупная организация, попавшая в данное подмножество.

В подмножестве № 3 (см. табл. 6) находятся организации, у которых рост сотрудничества сочетается со снижением показателя CQ. Эта ситуация требует особого внимания, поскольку обычно формирование коллабораций приводит к улучшению качества исследований за счет расширения компетенций и возможностей проектной группы. Среди организаций в этом подмножестве 68% являются вузами 2-й и 3-й групп, а также 5 университетов из 1-й группы: БФУ, УрФУ, ЮУрГУ, МИФИ, СПбПУ им. Петра Великого.

Таблица 6

Классификация организаций по коэффициентам коллабораций и «качества» («↑» – растущий тренд, «↓» – падающий тренд, «0» – тренд отсутствует)

№	СС	CQ	Все, %	АИ-1, %	АИ-2, %	ВУЗ-1, %	ВУЗ-2, %	ВУЗ-3, %	ДР, %
1	↑	↑	14,7	31,6	18,7	34,8	5,4	3,3	8,6
2	↑	0	30,8	31,0	27,6	21,7	32,6	30,4	35,3
3	↑	↓	11,6	7,1	7,5	21,7	25,0	12,2	8,6
4	0	↑	5,3	5,2	9,7	–	2,2	3,9	6,0
5	0	0	17,1	11,0	23,9	–	12,0	20,4	19,8
6	0	↓	7,8	1,9	2,2	8,7	12,0	13,8	9,5
7	↓	↑	0,6	0,6	1,5	–	–	–	0,9
8	↓	0	0,9	–	1,5	–	1,1	0,6	1,7
9	↓	↓	1,0	0,6	–	–	1,1	2,2	0,9
-	Остальные		10,2	11,0	7,4	13,1	8,6	13,2	8,7
Всего			701	155	134	23	92	181	116

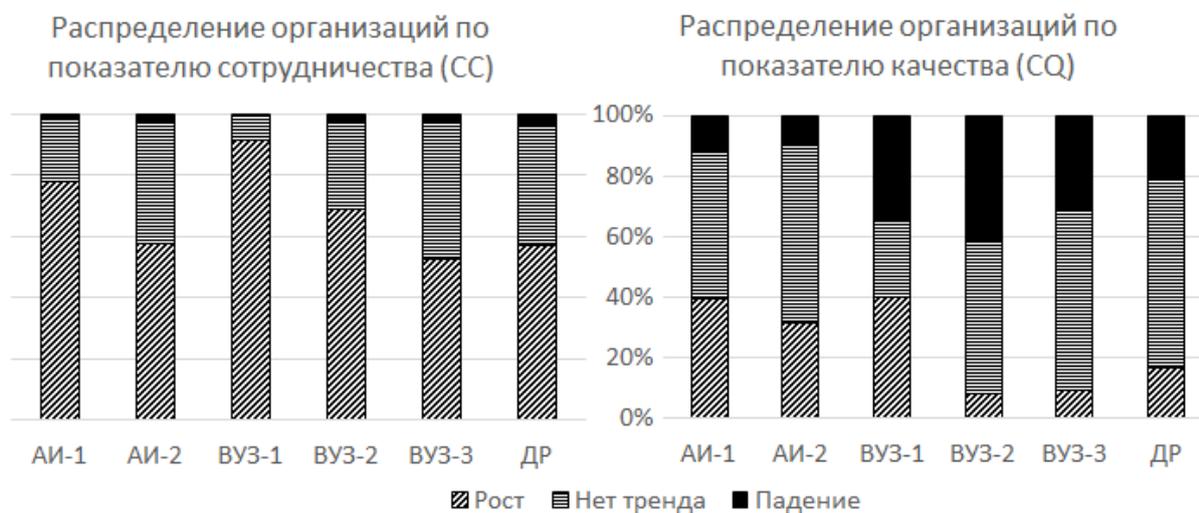


Рис. 4. Различия между группами организаций по классификациям метрик CC и CQ

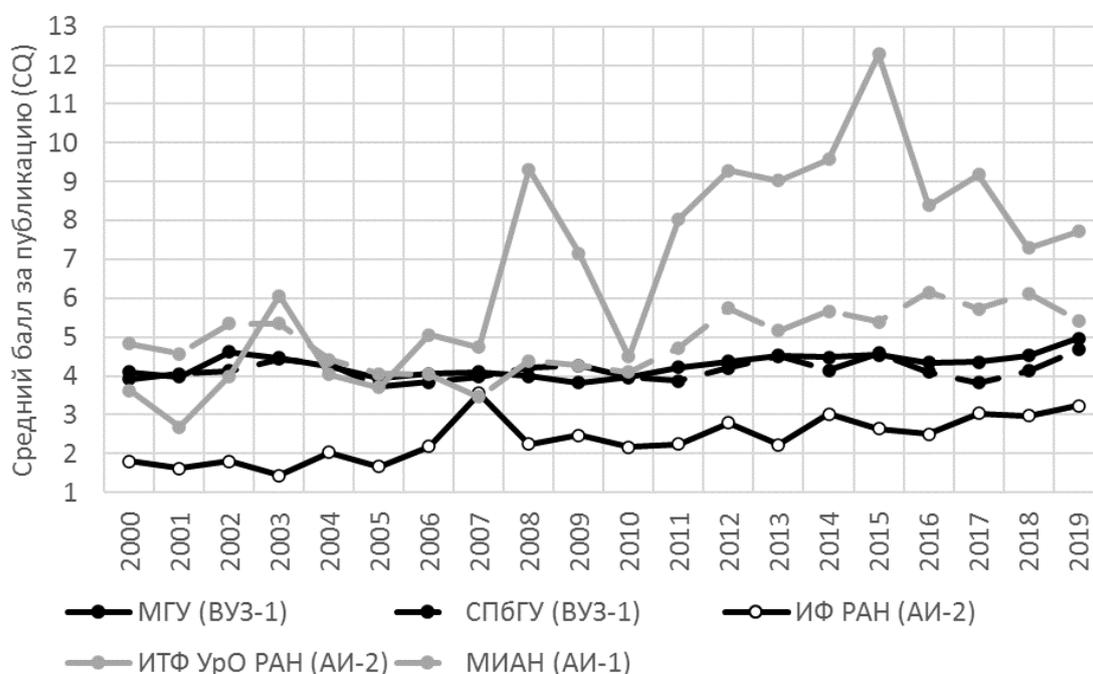


Рис. 5. Динамика изменения среднего коэффициента качества CQ по группам организаций

Есть ли прямая связь между развитием коллабораций и развитием качества, или это эффект от влияния разных факторов?

Подмножества № 4-6 (см. табл. 6) характеризуются разным поведением показателя CQ при стабильном уровне сотрудничества. Аналогично подмножествам № 1-3: там, где качество улучшается, преобладают академические институты, а где качество ухудшается – больше вузов. В подмножестве № 4 наиболее высокий рост среднего балла за публикацию демонстрируют Центр теоретических проблем физико-химической фармакологии РАН (значения показателя изменились с 1 до 6,5) и Институт теплофизики Уро РАН (с 3 до 8,1) – рост от средней публикации в журналах Q4 до Q2. В подмножество №6

попали два университета из первой группы организаций – это КФУ и ЛЭТИ. Часть авторов КФУ были замечены в публикациях в хищнических журналах [31, 32], у ЛЭТИ значительно выросла доля публикаций в сборниках материалов конференций.

Подмножества № 7-9 (см. табл. 6) малочисленны и его составляют организации, у которых снизился уровень сотрудничества. При этом рост «качества» (метрика CQ) характерен для академических институтов, наиболее крупными из которых являются Физико-технологический институт РАН и Институт проблем переработки углеводов СО РАН.

Таким образом, в большинстве организаций уровень сотрудничества возрос, чаще это происходило в группах AI-1 и VUZ-1 (рис. 4). Показатель «каче-

ства» CQ (средний балл на статью) находится в некотором балансе, его рост чаще наблюдается в академических институтах, а его падение – в вузах. Падение среднего балла на статью наблюдается у 94 вузов (32% от всех вузов) и 30 (10%) академических институтов. Рост же наблюдается только у 30 (10%) вузов и 98 (34%) академических институтов. Чуть более половины организаций осталось без изменений.

Для оценки степени роста/падения «качества» публикаций на рис. 5 показана динамика изменения средних показателей метрики CQ по группам организаций. Как можно заметить, средние показатели у академических институтов 1-й категории самые высокие. Следующими идут вузы 1-й группы, у которых падение коэффициента качества с 2012 г. сменилось ростом в 2017 г. Самые низкие показатели наблюдаются у вузов 2-й и 3-й групп, причем с заметным падением с 2012 г. у 2-й группы и с 2013 г. у 3-й группы. Синхронное падение показателей отмечалось во всех группах в 2016 г., что связано с резким увеличением числа российских публикаций в сборниках трудов конференций.

Значительный рост публикационной активности и увеличение уровня сотрудничества наблюдаются как у университетов, так и у академических институтов, однако при этом между ними фиксируются и заметные различия. У вузов число публикаций выросло в 3,6 раза (фракционным счетом) с 2012 г., при этом у большей их доли отмечается падение среднего балла за публикацию.

Можно ли утверждать, что, резкий рост публикационной активности вузов обеспечен в значительной степени публикациями в низкорейтинговых журналах и в сборниках материалов конференций, что приводит к размыванию качества?

Публикационная активность академических институтов фракционным счетом выросла лишь на 80%, однако относительно высокий средний балл за публикацию сохранился на протяжении всего исследуемого периода. Стоит отметить, что публикационная активность академических институтов 1-й категории заметно выше, чем институтов 2-й категории,

средний балл за публикацию у них выше и чаще был замечен рост этого показателя.

Свидетельствует ли сочетание повышения среднего качества публикаций и роста коэффициента коллабораций об активном включении организации в международное разделение научного труда?

ВЛИЯНИЕ КОНФЕРЕНЦИЙ НА ПУБЛИКАЦИОННУЮ АКТИВНОСТЬ

Теперь рассмотрим сочетание классификаций числа публикаций и их доле в сборниках материалов конференций. В табл. 7 представлены результаты классификаций по двум метрикам – числу публикаций фракционным счетом (FC) и их доле в сборниках материалов конференций (CSH) (рис. 6).

Рассмотрим сначала все подмножества, в которых наблюдается явный рост числа публикаций фракционным счетом по различным классификациям и по их доле в сборниках материалов конференций.

В подмножестве № 1 (см. табл. 7) вузы занимают 66%. Это подмножество второе по величине и состоит из организаций, у которых наблюдается ускоренный рост публикаций в сборниках материалов конференций относительно остальных видов на фоне их общего роста. Подмножество №2 содержит наибольшее число организаций, в нем представлены в равной мере академические институты и вузы. Это подмножество характеризуется пропорциональным ростом всех публикаций и статей из сборников материалов конференций. Подмножество №3 немногочисленно и состоит преимущественно из академических институтов 1-й категории, относящихся к биологическим или химическим наукам, у которых, вероятно, возросла публикационная активность в журналах. Во всех остальных случаях, где рост доли конференционных материалов не наблюдается, академических институтов больше, чем вузов. Отметим, что в 2019 г. доля публикаций в сборниках материалов конференций среди всех рассматриваемых организаций составляет 32,4 %, у академических институтов – 21,7 %, у университетов – 38,9 %. Например, в группе ВУЗ-2 средний показатель метрики CSH увеличился с 20,4% в 2012 г. до 47,4% в 2019 г.

Таблица 7

Классификация организаций по числу публикаций и их доле в сборниках материалов конференций («↑» – растущий тренд, «↓» – падающий тренд, «0» – тренд отсутствует)

№	FC	CSH	Все, %	АИ-1, %	АИ-2, %	ВУЗ-1, %	ВУЗ-2, %	ВУЗ-3, %	ДР, %
1	↑	↑	27,5	16,1	23,9	52,2	50,0	38,1	7,8
2	↑	0	34,4	38,1	37,3	30,4	31,5	34,2	29,3
3	↑	↓	2,6	7,1	2,2	–	1,1	0,6	1,7
4	0	↑	2,0	1,9	3,0	–	1,1	1,7	2,6
5	0	0	10,3	15,5	11,9	–	1,1	3,3	21,6
6	0	↓	1,1	3,2	0,7	–	–	–	1,7
7	↓	↑	0,4	–	–	–	–	0,6	1,7
8	↓	0	2,9	1,3	2,2	–	–	–	12,9
9	↓	↓	1,1	–	–	–	–	0,6	6,0
10	Остальные		17,7	16,8	18,8	17,4	15,2	20,9	14,7
Всего			701	155	134	23	92	181	116

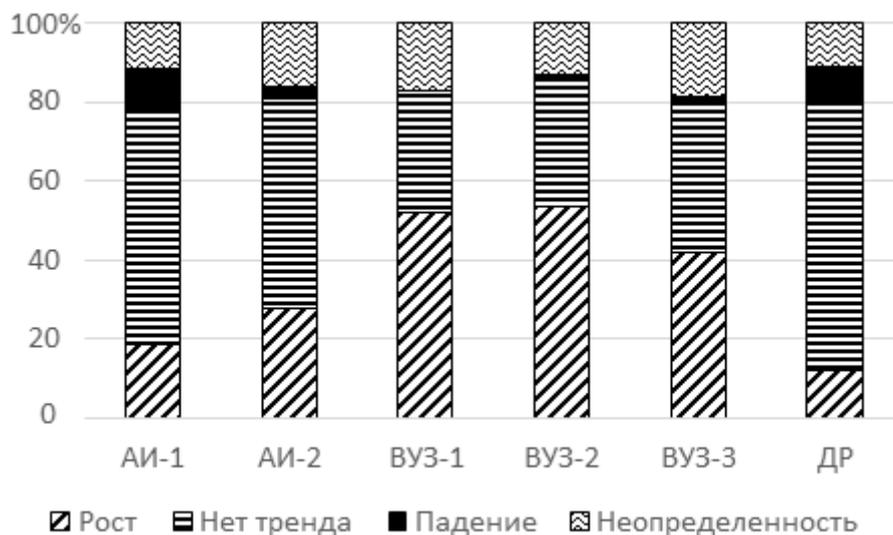


Рис. 6. Классификация групп организаций по метрике CSH (доля конференций, %)

С чем связан аномальный рост публикаций в трудах конференций, издаваемых российскими вузами?

Часть ответов на этот вопрос дана в статье И.А. Стерлигова [22], однако этот феномен «российского конференционного взрыва» следует изучить более детально.

ЗАКЛЮЧЕНИЕ

В 2010–2020 гг. наука в России стала одной из реальных целей национального развития: на 13,4% в постоянных ценах выросли внутренние затраты на исследования и разработки [33], был принят и реализуется ряд программ и проектов поддержки научных исследований. Хорошо заметен эффект, выраженный в значительном росте количества публикаций российских исследователей, индексируемых в международных базах данных, растет уровень национальных и международных коллабораций. Масштаб этого роста позволяет ожидать, что он обеспечен если не всеми, то хотя бы ведущими российскими вузами и научными организациями, этот рост закладывается и в планируемые показатели. Однако анализ показал, что ряд научно-исследовательских организаций, в том числе признанных лидеров, не демонстрирует рост публикационной активности, а некоторые показывают даже снижение. Имеются организации, показывающие, вопреки общему тренду, снижение уровня коллабораций. Эти феномены требуют отдельного изучения для понимания причин и, возможно, принятия управленческих или политических решений. Мы полагаем, что некоторые организации, являющиеся признанными научными лидерами, могут обеспечивать стабильно высокие показатели публикационной активности в связи с достижением потолка ресурсных ограничений, а в некоторых случаях даже снижать их из-за расширения своего участия в междисциплинарных исследованиях и особенно мегаколлаборациях.

В то же время настораживает низкая цитируемость работ российских ученых [33], почти в два раза отличающаяся от среднемировых показателей. В значительной степени это может объясняться высокой

долей публикаций в низкорейтинговых журналах и сборниках материалов конференций. Анализ показал, что сохранение «качества» публикаций с точки зрения рейтингов журналов в основном достигается в академических институтах, в то время как в вузах наблюдается его снижение, или, как мы предполагаем, «размывание» за счет опережающего роста числа низкорейтинговых публикаций. Особенно настораживает аномально высокая доля публикаций в сборниках материалов конференций основной массы вузов. Необходим анализ природы этого явления, в том числе с точки зрения последующего публикационного следа авторов – могут ли эти исследователи в последующем дорасти до качественных статей в высокорейтинговых журналах и если да, то при каких условиях?

Необходимо объяснить также различное поведение разных групп организаций с растущим уровнем коллабораций – у некоторых организаций при этом растет средний уровень качества публикаций, а у каких-то, наоборот, падает. Традиционно коллаборации (особенно международные) рассматриваются как драйвер для роста качества исследований за счет увеличения междисциплинарности и интенсивного обмена опытом. Возможно, снижение качества и рост коллабораций у второй группы организаций являются следствием независимых процессов, но если это не так, то причины такой связи важны для формулирования научной политики и принятия решений.

В настоящей статье апробирован метод определения трендов публикационной активности для временных рядов наукометрических показателей на примере российских университетов и научных организаций. Полученные результаты демонстрируют, что метод является вполне работоспособным и может применяться на относительно коротких рядах данных для выявления групп объектов со сходными тенденциями. Проведен также сравнительный анализ результатов публикационной активности среди наиболее значимых организаций РФ, занимающихся научными исследованиями. Изучение их временных рядов методами проверки статистических гипотез от-

сутствия тренда привело к формулированию ряда вопросов, возможности поиска ответов на которые лежат за пределами предложенного метода и могут быть предметом отдельных исследований.

Необходимо ещё раз подчеркнуть, что инструмент изучения трендов публикационной активности целесообразно использовать вместе с другими аналитическими инструментами и социологическими исследованиями, которые позволят более точно выявлять и обосновывать причины их возникновения.

СПИСОК ЛИТЕРАТУРЫ

1. Moed H.F., Markusova V., Akoev M. Trends in Russian research output indexed in Scopus and Web of Science // *Scientometrics*. – 2018. – Vol. 116, № 2. – С. 1153–1180.
2. Matveeva N., Sterligov I., Yudkevich M. The effect of Russian University Excellence Initiative on publications and collaboration patterns // *Journal of Informetrics*. – 2021. – Vol. 15, № 1. – С. 101110.
3. Shibanova E., Platonova D., Lisyutkin M. The 5–100 Project University Development Trends and Patterns // *Russian Education and Society*. – 2018. – Vol. 60. – P.682–715. DOI: 10.1016/j.joi.2020.101110
4. Guskov A. E., Kosyakov D. V., Selivanova I. V. Boosting research productivity in top Russian universities: the circumstances of breakthrough // *Scientometrics*. – 2018. – Vol. 117, № 2. – P. 1053–1080.
5. Turko T. et al. Influence of the program “5–top 100” on the publication activity of Russian universities // *Scientometrics*. – 2016. – Vol. 109, № 2. – P. 769–782.
6. Гуськов А.Е., Косяков Д.В., Селиванова И.В. Стратегии повышения публикационной активности университетов – участников Проекта 5–100 // *Научные и технические библиотеки*. – 2017. – № 12. – С. 5–18.
7. Matveeva N., Ferligoj A. Scientific collaboration in Russian universities before and after the excellence initiative Project 5-100 // *Scientometrics*. – 2020. – Vol. 124, №3. – P. 2383–2407. DOI: 10.1007/s11192-020-03602-6
8. Mazov N.A., Gureev V.N. Bibliometric analysis of the flow of publications by Novosibirsk State University in collaboration with the RAS Siberian Branch // *Herald of the Russian Academy of Sciences*. – 2017. – Vol. 87, № 5. – P. 445–453. DOI: 10.1134/S1019331617050057
9. Gureev V.N., Guskov A.E., Mazov N.A. Russian Scientists in Global Scientific Migration Processes // *Herald of the Russian Academy of Sciences*. – 2021. – Vol. 91, № 4. – P. 428–437. DOI: 10.1134/s1019331621040055
10. Гуськов А.Е., Селиванова И.В., Косяков Д.В. Миграция российских исследователей: анализ на основе наукометрического подхода // *Библиосфера*. – 2021. – № 1. – С. 3–15.
11. Subbotin A., Aref S. Brain drain and brain gain in Russia: Analyzing international migration of researchers by discipline using Scopus bibliometric data 1996–2020 // *Scientometrics*. – 2021. – Vol. 126, № 9. – P. 7875–7900.
12. Гуськов А.Е., Косяков Д.В. Национальный фракционный счёт и оценка научной результативности организаций // *Научные и технические библиотеки*. – 2020. – Т. 9, № 1. – С.15–42. DOI: 10.33186/1027-3689-2020-9-15-42
13. Методика подсчета комплексного балла публикационной активности. – URL: https://minobrnauki.gov.ru/upload/Methodika_novaya.pdf (дата обращения: 07.06.2022).
14. Гуськов А.Е., Косяков Д.В., Селиванова И.В. Методика оценки результативности научных организаций // *Вестник Российской академии наук*. – 2018. – Т. 88, № 5. – С. 430–443. DOI: 10.7868/S0869587318050092.
15. Guba K., Tsivinskaya A. Evaluating the Evaluators in Russia: When Academic Citizenship Fails // *Europe-Asia Studies*. – 2021. – Vol. 73, № 6. – P. 1010–1036. DOI: 10.1080/09668136.2021.1932759
16. Блинов П.Ю. Методология классификации временных рядов по наличию тренда в наукометрии // *Вычислительные технологии*. – 2022. – Т. 27, № 1. – С. 52–69
17. Waltman L., van Eck N.J. Field-normalized citation impact indicators and the choice of an appropriate counting method // *Journal of Informetrics*. – 2015. – Vol. 9, № 4. – P. 872–894.
18. Thelwall M., Maflahi N. Academic Collaboration Rates and Citation Associations Vary Substantially Between Countries and Fields // *Journal of the Association for Information Science and Technology*. – 2020. – Vol. 71, № 8. – P. 968–978.
19. Мжельский А.А., Москалёва О.В. Научная кооперация как фактор, влияющий на цитируемость статьи. Аналитический обзор // *Управление наукой: теория и практика*. – 2020. – Т. 2, № 3. – С. 138–164.
20. Macháček V., Srholec M. Globalization of Science. Evidence from Authors in Academic Journals by Country of Origin. – URL: <https://idea.cerge-ei.cz/studies/2019-05-globalization-of-science-evidence-from-authors-in-academic-journals-by-country-of-origin> (дата обращения: 07.06.2022).
21. Багирова А.В., Косяков Д.В., Гуськов А.Е. 50 самых высокоцитируемых обзоров 2013–2017 гг. // *Научно-техническая информация. Сер. 1*. – 2021. – № 7. – С. 22–36. DOI: 10.36535/0548-0019-2021-07-4; Bagirova A.V., Kosyakov D.V., Guskov A.E. The 50 Most Highly Cited Reviews of 2013-2017 // *Scientific and Technical Information Processing*. – 2021. – Vol. 49, № 3. – P. 168-184.
22. Стерлигов И.А. Российский конференционный взрыв: масштабы, причины, дальнейшие действия // *Управление наукой: теория и практика*. – 2021. – №2. – С. 222-251. DOI: 10.19181/sntp.2021.3.2.10.
23. Селиванова И.В., Косяков Д.В., Гуськов А.Е. Влияние ошибок в базе данных *Scopus* на оценку результативности научных исследований // *Научно-техническая информация. Сер. 1* – 2019. – № 9. – С. 25-32; Selivanova I.V., Kosyakov D.V., Guskov A.E. The Impact of Er-

- rors in the Scopus Database on the Research Assessment // Scientific and Technical Information Processing. – 2019. – Vol. 46, № 3. – P. 204–212.
24. Cox D.R., Stuart A. Quick sign tests for trend in location and dispersion // *Biometrika*. – 1955. – Vol. 42. – P.80–95.
25. Химмельблау Д. Анализ процессов статистическими методами. – Москва: Мир, 1973.
26. Веретельникова И.В. Исследование и применение критериев проверки гипотез об отсутствии тренда и критериев однородности: дис. ... канд. техн. наук. – Новосибирск: НГТУ, 2019. – 203 с.
27. Mann H.B. Non-parametric tests against trend // *Econometrica*. – 1945. – Vol. 13. – P. 245-259.
28. Kendall M.G. Rank Correlation Methods. – London: Charles Griffin, 1975.
29. Лемешко Б.Ю., Комиссарова А.С., Щеглов А.Е. Свойства и мощность некоторых критериев случайности и отсутствия тренда // *Научный вестник НГТУ*. – 2012. – Т. 46, № 1. – С. 53-66.
30. Yue S., Pilon P. A comparison of the power of the test, Mann-Kendall and bootstrap tests for trend detection // *Hydrological Sciences Journal*. – 2004. – Vol. 49. – P. 21–37.
31. Грубникова Е.И. Проект 5–100: взгляд через призму теории институциональной коррупции // *Мир России*. – 2020. – Т. 29, № 2. – С. 72–91. DOI: 10.17323/1811-038X-2020-29-2-72-91
32. Комиссия РАН по противодействию фальсификации научных исследований. При участии А.А. Абалкиной, А.С. Касьяна, Л.Г. Мелиховой. Иностранцы хищные журналы в Scopus и WoS: переводной плагиат и российские недобросовестные авторы. – URL: <https://kpfra.ru/wp-content/uploads/plagiarism-by-translation-2.pdf> (дата обращения: 07.06.2022).
33. Индикаторы науки: 2021: статистический сборник / Л.М. Гохберг, К.А. Дитковский, Е.И. Евневич и др.; Нац. исслед. ун-т «Высшая школа экономики». – Москва: НИУ ВШЭ, 2021. – 352 с.

Материал поступил в редакцию 14.06.22.

Сведения об авторах

БЛИНОВ Павел Юрьевич – кандидат технических наук, старший научный сотрудник лаборатории наукометрии и научных коммуникаций, Российский научно-исследовательский институт экономики, политики и права в научно-технической сфере, Москва
e-mail: p.blinov@iep.ru

КОСЯКОВ Денис Викторович – заместитель заведующего лаборатории наукометрии и научных коммуникаций, Российский научно-исследовательский институт экономики, политики и права в научно-технической сфере
e-mail: kosyakov@sciencepulse.ru

МАЛЫШЕВА Александра Валерьевна – младший научный сотрудник лаборатории наукометрии и научных коммуникаций, Российский научно-исследовательский институт экономики, политики и права в научно-технической сфере
e-mail: bag_bala@mail.ru

ГУСЬКОВ Андрей Евгеньевич – кандидат технических наук, заведующий лабораторией наукометрии и научных коммуникаций, Российский научно-исследовательский институт экономики, политики и права в научно-технической сфере
e-mail: guskov.andrey@gmail.com

Д.В. Виноградов

Несостоятельность оценки числа классов эквивалентности состояний в обучении по малым выборкам

В настоящей заметке мы докажем, что предложенный ранее коллегами оригинальный способ обнаружения классов эквивалентности состояний вероятностных процессов с быстрым перемешиванием по матрице переходов оказался плохим, он неправильно оценивает число классов (оценка несостоятельна).

Ключевые слова: орграф, полный подграф, класс эквивалентности, состоятельность оценки

DOI: 10.36535/0548-0027-2022-08-3

ОДИН ПОДХОД К ОБНАРУЖЕНИЮ КЛАССОВ ЭКВИВАЛЕНТНОСТИ СОСТОЯНИЙ

В статье Grusho A.A., Grusho N.A., Zabezhailo M.I., Timonina E.E., Kulchenkov V.V. «Probability model of concepts recovery in small sample learning»¹ был предложен оригинальный способ обнаружения классов эквивалентности состояний вероятностных процессов с быстрым перемешиванием по матрице переходов. С целью выявления изменения состояния процесса (например, функционирования линии связи) авторы предложили рассматривать бинарную матрицу возможных переходов из одного состояния в другое, предположив, что эти состояния разбиты на несколько классов эквивалентности M_1, \dots, M_k . Число таких состояний обозначалось через n , которое в дальнейшем устремлялось к бесконечности. Никаких ограничений на число кластеров k не накладывалось, кроме естественных $n \geq k > 1$.

В наших дальнейших выводах мы условимся использовать буквы i, j для обозначения отдельных состояний. Множество всех состояний обозначим через V , оно соответствует вершинам ориентированного графа (орграфа).

Опишем используемую в обсуждаемой статье вероятностную модель: имеется некоторая вероятность α сохранения класса состояний и серия испытаний Бернулли с параметром успеха α для заполнения матрицы переходов A . Эта серия будет иметь длину $T = \frac{(n-tnn)^2}{\alpha}$. Следует отметить свойство $T \rightarrow \infty$, которое плохо согласуется с декларацией «обучение в малых выборках» из названия обсуждаемой работы.

В дальнейшем мы будем обозначать элементы матрицы переходов через $a_{i,j}$, которые первоначально инициализированы нулями ($a_{i,j} = 0$).

Если в серии Бернулли появляется «успех», то сначала равновероятно выбирается некоторый класс M , а затем независимо и равновероятно выбираются $i \in M, j \in M$ и устанавливается элемент $a_{i,j} = 1$.

Если в серии Бернулли появляется «неудача», то сначала равновероятно выбираются различные кластеры $M \neq M'$, а затем независимо и равновероятно выбираются $i \in M, j \in M'$ и устанавливается элемент $a_{i,j} = 1$.

Эта модель может считаться адекватной при учете изменения состояний, если предполагать, что случайный процесс обладает свойством быстрой перемешиваемости, а изменения извлекаются через достаточно большие промежутки времени. В противном случае, очевидно, имеется коррелированность состояний, но для такой общей модели многие техники исходной статьи не работают.

Затем в обсуждаемой статье излагается некоторый алгоритм нахождения классов эквивалентности с помощью нахождения полных подграфов в получившемся орграфе (V, A) , остальные состояния присоединяются к классу с максимальным числом ребер.

Следует заметить, что с вычислительной точки зрения предложенный алгоритм не вполне удовлетворителен, так как поиск клика в графе является $\#P$ -полной задачей, а следовательно, потребуется экспоненциальный алгоритм. К сожалению, число состояний должно быть большим (в обсуждаемой статье предполагается, что $n \rightarrow \infty$), а это вызывает сомнения в практичности предложенных конструкций.

Но, самое главное, неверна ключевая предлагаемая авторами исходной статьи теорема, которая утверждает (мы ослабим ее утверждение, отбросив соответствие между классами эквивалентности и полными подграфами) состоятельность оценки числа кластеров.

¹ Grusho A.A., Grusho N.A., Zabezhailo M.I., Timonina E.E., Kulchenkov V.V. Probability model of concepts recovery in small sample learning // Communications of the ECMS. – 2020. – Vol. 34, Issue 1. – P. 393-397. – URL: <https://www.scs-europe.net/dlib/dl-index.htm>.

Нетеорема. При $n \rightarrow \infty$, $\alpha = 1 - \frac{C}{\ln n}$ и $T = \frac{(n \cdot \ln n)^2}{\alpha}$ число полных подграфов в орграфе (V, A) будет с вероятностью 1 сходиться к числу k классов эквивалентности.

Почему возникли сомнения в истинности ключевой теоремы исходной статьи? Дело в том, что число k классов является гиперпараметром, который не должен оцениваться по выборке, а это не так в описанном коллегами подходе. В таких случаях обычно нарушается состоятельность оценки, что мы и докажем в следующем разделе.

НЕСОСТОЯТЕЛЬНОСТЬ ПРЕДЛОЖЕННОГО ПОДХОДА

Опровергать теорему из исходной статьи мы будем для $C = 1$ (т. е. $\alpha = 1 - \frac{1}{\ln n}$) и $k = n$. Мы докажем, что с вероятностью $\geq \frac{1}{2}$ число полных подграфов будет меньше n .

При $k = n$ если в серии Бернулли появляется «успех», то устанавливается элемент $a_{i,j} = 1$ с вероятностью $\frac{\alpha}{n}$; если же в серии Бернулли появляется «неудача», то устанавливается элемент $a_{i,j} = 1$ при $i \neq j$ с вероятностью $\frac{1-\alpha}{\binom{n}{2}} = \frac{2}{n(n-1)\ln n}$.

Докажем некоторые вспомогательные леммы.

Лемма 1. После $T = \frac{(n \cdot \ln n)^2}{\alpha}$ шагов

$$P[\exists i(a_{i,i} = 0)] \leq n \cdot e^{-n \cdot \ln^2 n} \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Доказательство.

$$\text{Оценим } P[a_{i,i} = 0] = \left(1 - \frac{\alpha}{n}\right)^T \leq e^{-\frac{\alpha \cdot n^2 \cdot \ln^2 n}{\alpha}}, \text{ и}$$

применим неравенство Буля

$$P[\exists i(a_{i,i} = 0)] \leq \sum P[a_{i,i} = 0].$$

Лемма 2. После $T = \frac{(n \cdot \ln n)^2}{\alpha}$ шагов

$P[a_{i,j} = a_{j,i} = 1]$ не стремится к 0 при $n \rightarrow \infty$ для любых $i \neq j$.

Доказательство. Рассмотрим фиксированную пару различных состояний $i \neq j$ и пару соответствующих ячеек $a_{i,j}$ и $a_{j,i}$. Будем наблюдать сумму геометрических случайных величин $X + Y$. Случайная величина X соответствует времени ожидания первого попадания в любую из наших ячеек. Вероятность успеха для нее будет $p = \frac{4}{n(n-1)\ln n}$, как обычно, а вероятность неудачи $1 - p$ мы обозначим через q . Тогда $P[X = k] = p \cdot q^{k-1}$.

Случайная величина Y – время ожидания первого попадания в дополнительную ячейку матрицы – геометрическая случайная величина с вероятностью успеха $\frac{p}{2}$, т. е. $\frac{2}{n(n-1)\ln n}$. Тогда для нее имеем

$$P[Y = t - k] = \frac{p}{2} \cdot \left(1 - \frac{p}{2}\right)^{t-k-1} \geq \frac{p}{2} \cdot q^{t-k-1}.$$

Оценим

$$\begin{aligned} P[X + Y = t] &= \\ &= \sum P[X = k] \cdot P[Y = t - k] \geq \frac{(t-1) \cdot p^2}{2} \cdot q^{t-2}. \end{aligned}$$

Нас интересует событие $X + Y \leq T$, когда до момента T в обеих ячейках появятся 1. Имеем

$$\begin{aligned} P[X + Y \leq T] &= \\ &= \sum P[X + Y = t] \geq \frac{p^2}{2} \cdot \sum (t-1) \cdot q^{t-2} = \\ &= \frac{p^2}{2} \cdot [\sum x^t]'_{x=q} = \\ &= \frac{p^2}{2} \cdot [(x^T - 1) \cdot (x-1)^{-1}]'_{x=q} = \\ &= \frac{p^2}{2} \cdot \left[\frac{T \cdot x^{T-1} \cdot (x-1) - (x^T - 1)}{(1-x)^2} \right]_{x=q} = \\ &= \frac{1}{2} \cdot [1 - q^{T-1} \cdot (q + p \cdot T)] \geq \\ &\geq \frac{1}{2} \cdot \left[1 - \left(1 - \frac{4}{n(n-1)\ln n}\right)^{\frac{n^2 \ln^2 n}{\alpha} - 1} \cdot \left(1 - \frac{4}{n(n-1)\ln n} + \frac{4n^2 \ln^2 n}{n(n-1)\ln n}\right) \right]. \end{aligned}$$

Очевидно, что $\left(1 - \frac{4}{n(n-1)\ln n}\right)^{\frac{n^2 \ln^2 n}{\alpha} - 1}$ при $n \rightarrow \infty$

имеет асимптотику $n^{-\frac{4}{\alpha}}$, а это задавит асимптотику $4 \ln n$ второго сомножителя в скобках.

Теперь сформулируем наш результат.

Теорема 1. При $n \rightarrow \infty$, $\alpha = 1 - \frac{1}{\ln n}$ и $T = \frac{(n \cdot \ln n)^2}{\alpha}$

число полных подграфов в орграфе (V, A) не будет с вероятностью 1 сходиться к числу $k = n$ классов эквивалентности.

Доказательство. Если $\exists i(a_{i,i} = 0)$ после T шагов, то $k = n$ невозможно. Впрочем, по Лемме 1 вероятность этого экспоненциально мала.

Поэтому допустим противное: $\forall i(a_{i,i} = 1)$ после T шагов. Чтобы число кластеров оказалось меньше $k < n$, нужно для некоторой пары состояний $i \neq j$ одновременно наблюдать события $a_{i,j} = a_{j,i} = 1$. Тот факт, что это произойдет с отделенной от 0 вероятностью, следует из Леммы 2.

ЗАКЛЮЧЕНИЕ

В настоящей работе доказана несостоятельность оригинального метода обнаружения классов эквивалентности состояний вероятностных процессов с быстрым перемешиванием по матрице переходов, предложенного в статье А.А. Грушо и др., опубликованной в 2020 г. в журнале «Communications of the ECMS»

Можно ли исправить ситуацию изменением параметров? Думается, что это невозможно. Хотя наш контр-пример с $k = n$ можно устранить (гиперпараметр k зафиксирован), но тогда несостоятельность возникнет при другом числе классов эквивалентности.

Материал поступил в редакцию 26.05.22.

Сведения об авторе

ВИНОГРАДОВ Дмитрий Вячеславович – доктор физико-математических наук, ведущий научный сотрудник ФИЦ ИУ РАН, Москва
e-mail: KRRGuest@yandex.ru

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 811.161.1'37'322.2

К.К. Боярский, Е.А. Каневский, Е.Н. Клименко

Устойчивые словосочетания с двумя семантическими значениями

Рассматриваются особенности обработки неоднословных оборотов (фразем), способных выполнять функцию предлогов, при автоматическом анализе русскоязычных текстов. Сложность анализа таких оборотов заключается в достаточно высокой омонимичности, что снижает точность автоматического разбора предложений и, тем более, определение семантики.

Приведены результаты исследований омонимии 20 фразем, могущих играть роль предлогов или быть простыми сочетаниями слов. Разработаны правила, позволяющие снимать частеречную и синтаксическую омонимию при автоматическом анализе текста: сначала в правом и левом контекстах находятся управляющие слова, затем проводится семантико-синтаксический (с использованием семантического классификатора) анализ этих управляющих слов, на основании чего принимается решение, является ли словосочетание составным предлогом. Отмечено, что рассматриваемые фраземы кроме прямого лексического значения имеют и переносное, которое и соответствует предложному обороту.

Ключевые слова: автоматический анализ текста, классы, лексемы, словосочетания, омонимия, предложные обороты, фраземы

DOI: 10.36535/0548-0027-2022-08-4

ВВЕДЕНИЕ

В современном информационном мире важнейшее значение приобретает задача извлечения информации из текстовых источников, причем характер этой информации может быть самый различный: выделение ключевых слов, тематическая рубрикация, определение отношения автора к описываемым событиям и т. д. Разработаны разнообразные методы, в большинстве своем статистические, позволяющие решать эти задачи [1].

Однако, если объем анализируемых текстов ограничен, а их лексический состав довольно близок (например, научные работы определенной предметной области), то точность статистических методов может оказаться недостаточной. Становится необходимым проводить анализ текстов более детально, например, строить дерево подчинения и анализировать семантику как отдельных слов, так и связей между ними [2]. Для русскоязычных текстов проблема усугубляется высокой степенью омонимичности лексических единиц, связанной, в том числе, с их морфологической изменчивостью. Несколько контактно стоящих слов нередко образуют устойчивый

оборот – фразему. При этом лексическое значение фраземы часто не может быть выведено из лексических значений составляющих её слов, так что она фактически является идиомой [3]. Автоматическое выявление фразем, определение их синтаксической роли и семантики – это одна из важных задач компьютерного анализа текстов.

ПОСТАНОВКА ЗАДАЧИ

Фраземы в тексте могут изменяться или быть неизменяемыми. Большая часть неизменяемых фразем представляет собой обороты, выполняющие функции:

- наречий – в конце концов, время от времени, в свое время;
- предлогов – в зависимости от, в соответствии с, несмотря на;
- вводных оборотов – другими словами, к слову сказать;
- союзов – а также, в том числе и, вместо того чтобы, если бы;
- частиц – все же, как бы, как раз, к тому же, едва не, вроде бы;
- предикативных оборотов – не дай бог, лыка не вяжет;

- прилагательных – в миниатюре, милостью божьей, с бородой;
- междометий – бог в помощь, ёлки зеленые, скажи на милость.

Однако даже неизменяемые фраземы могут обладать омонимией: в зависимости от контекста представлять собой разные части речи или вообще быть свободными словосочетаниями, сохраняющими семантику составляющих слов. Например, некоторые случаи, когда сочетание местоимения с глаголом образует наречный или местоименный оборот, рассмотрены в [4, 5].

Наиболее полные списки оборотов приведены в Национальном корпусе русского языка (НКРЯ) [6] – более 2800 словосочетаний. В нашем исследовании использовались также словари С.А. Кузнецова [7] – более 1400 словосочетаний и Р.П. Рогожниковой [8] – 1500 сочетаний, эквивалентных слову. Следует сразу заметить, что все три источника отмечают неоднозначность оборотов, однако только Р.П. Рогожникова достаточно внимательно разбирает случаи омонимии рассматриваемых ею словосочетаний. С точки зрения филолога все нормально – зафиксирован случай употребления какого-либо словосочетания в той или иной роли, указана его семантика и этого вполне достаточно. Но для лингвиста задача не решена – для разработки правил снятия такой омонимии при автоматическом анализе текста необходимо иметь алгоритмы. А для этого надо переложить семантику на морфологию и/или на собственно слова, что само по себе достаточно сложная задача.

Один из важных синтаксических элементов, определяющих смысл предложения, – это предложные группы. Семантика предложных групп, в которых в качестве предлога выступают неоднословные обороты, обсуждалась, например, в [9]. Как показывает даже предварительный анализ, большинство из таких фразем не обладает омонимией и всегда является предложными оборотами. Фраземы, могущие быть предложными оборотами, можно разделить на три группы [10].

Словосочетания 1-й группы практически не имеют омонимов¹ и представляют собой предложные обороты. Они могут состоять из предлога и существительного (*в виде, в ходе, во время*) или оканчиваться предлогом (*бок о бок с, в ответ на, вместе с*).

Словосочетания 2-й группы могут выполнять функции предлогов или наречий. Для снятия омонимии достаточно использовать простейшее правило: если после такого словосочетания стоит слово в требуемом падеже, то это предложный оборот. В большинстве случаев это родительный падеж (*в конце, во главе, по праву*). В противном случае это наречие.

Словосочетания 3-й группы могут играть роль предлогов или быть простыми сочетаниями слов (кроме того иногда ещё играть роль других частей речи). В большинстве случаев такие предлоги требуют после себя родительного падежа. Многие из них при отсутствии слова в соответствующем падеже выступают простыми сочетаниями слов. Например: *без*

ведома, в пределах, с целью. Однако в ряде случаев словосочетание не является предлогом и при наличии справа от фраземы слова с требуемым падежом – нужны еще и дополнительные условия. В связи с этим при автоматическом анализе таких фразем необходимо выполнять достаточно сложные правила, рассмотрению которых и посвящена настоящая статья.

МЕТОДИКА ИССЛЕДОВАНИЯ

Для отработки правил и анализа текста мы используем парсер SemSin [11], к достоинствам которого относятся наличие словаря, классификатора и широкого набора команд [12, 13], позволяющего достаточно легко разрабатывать самые сложные правила. Если фразема является оборотом, то входящие в неё слова объединяются в один токен (единицу разбора). Это накладывает определенные условия на разрабатываемые правила, поскольку они должны быть выполнены еще до образования именных и предложных групп (до формирования каких-либо связей между токенами).

Анализ проводился на материале предложений из НКРЯ [6], выбранных по опции Excel. Если исследуемое словосочетание встречалось слишком редко, то использовался также и наш массив текстов (объемом около 50 млн слов), составленный из ряда повестей, новостных и спортивных статей. Отобранный материал подвергался дополнительной фильтрации, чтобы исключить случаи разрыва словосочетания знаками препинания.

Отобранные предложения разделялись на несколько частей с помощью разрабатываемых правил, а полученные наборы предложений исследовались экспертным образом. Как будет показано далее, это позволило достаточно надежно определять частоты встречаемости предложных оборотов и их омонимов.

Рассмотрим подробнее правила снятия омонимии у ряда фразем², имеющих, по меньшей мере, два четких лексических значения. Все отобранные фраземы характеризуются тем, что в случае, когда они имеют переносное значение, они играют роль предлога.

ФРАЗЕМЫ И ПРАВИЛА СНЯТИЯ ИХ ОМОНИМИИ

1. БЕЗ ПОМОЩИ – фразема содержит два лексических значения: или действие выполняется без помощи человека (это будет свободным сочетанием двух слов), или для чего-либо не используются материальные средства, денежные пособия, техника, продукты и т.п. (и это будет предложным оборотом).

Для анализа фраземы использовалось двухэтапное правило.

Первый этап – поиск управляющего слова, которым может быть одно из трех слов, стоящее справа от фраземы³, если оно является существительным в

¹ Русский язык характерен большим количеством исключений. Здесь и далее мы считаем допустимым, если количество ошибок не превосходит 4-5%.

² Из исследованных нами 55 фразем с двумя вариантами омонимии было отобрано 19, которые обладают наиболее ярко выраженными значениями.

³ Между управляющим словом и фраземой может стоять прилагательное, причастие или числительное в родительном падеже.

родительном падеже и выполняет одно из трех условий – слово однозначно, справа от него стоит знак пунктуации или стоит слово, не имеющее родительного падежа (82%):

● *Но без помощи эскимосов попытка окончилась провалом*⁴.

● *Столь масштабные дела невозможно расследовать без помощи военного ведомства.*

● *Правда, все это он делал без помощи своей прекрасной яхты.*

Если управляющее слово не найдено, то это – **свободное словосочетание** (18%):

● *В конце концов дю Россэ решил обойтись без помощи и ехать в одиночку.*

● *Мусульманин брат мусульманину, он не предаст его и не оставляет без помощи.*

Второй этап – анализ управляющего слова.

Свободное словосочетание (60%). Управляющее слово является одушевленным существительным и относится к классу «Человек» (лоцман, Станиславский, человек):

● *Без помощи Фридриха фон Герольта эта аудиенция едва ли состоялась бы.*

Предложный оборот (40%):

● *Теперь мы уже не могли высадиться на берег без помощи ветра.*

2. В ГЛАЗАХ – фраза содержит два лексических значения: в одном случае имеются в виду глаза человека (это будет свободным сочетанием двух слов), или общественное (много реже, частное) мнение (и это будет предложным оборотом).

Для анализа этой фраземы используется многозвенное правило. Заметим, что такая структура правила обеспечивает минимальную погрешность.

Свободное словосочетание. Контактно справа стоит одно из местоимений *его, её, их, моих* или *твоих* (10%):

● *Потом она оторвалась, взглянула на него, и в глазах её остановился ужас.*

Предложный оборот. Контактно справа стоит слово в родительном падеже, причем или справа стоят слова *всех, окружающих* или в окрестности фраземы⁵ имеется словоформа одной из лемм **АВТОРИТЕТ, ВЛАСТЬ, ДИСКРЕДИТИРОВАНИЕ, ДИСКРЕДИТИРОВАТЬ, ДОСТОИНСТВО, ЗАСЛУГА, ОДИОЗНЫЙ, ПЛОХОЙ, ПРЕСТИЖ, ПРЕСТУПЛЕНИЕ, СМЕШНОЙ, УВАЖЕНИЕ** или **УЯЗВИМЫЙ** (6%):

● *Если вы опуститесь до их уровня, то потеряете свой авторитет в глазах русских.*

Свободное словосочетание. В окрестности фраземы имеется слово из класса «Психика Душа Чувство» (апатия, грусть, ощущение) или словоформа одной из лемм **ВИДЕТЬ, ГОРЕТЬ, ЗАГОРАТЬСЯ, ЗАГОРЕТЬСЯ, МЕЛЬКНУТЬ, МЕРКНУТЬ, НАСМЕШКА,**

⁴Здесь и далее жирным шрифтом выделено собственно исследуемое словосочетание, управляющее слово, по которому предложение отнесено к данной группе, подчеркнуто

⁵Здесь и далее под окрестностью фраземы понимается ближайший контекст фраземы влево на 5 токенов (но не далее двоеточия и точки с запятой) и вправо на 7 токенов (но не далее знака препинания, не считая кавычек). При поиске текст самой фраземы не учитывается.

ПОСМОТРЕТЬ, ПОТЕМНЕТЬ, ПРОМЕРЦАТЬ, ПРОЧЕСТЬ, СЛЕЗА, УВИДЕТЬ или **ЧИТАТЬ** (32%):

● *Ей показалось, что в глазах милиционера мелькнуло сочувственное, жалостливое выражение.*

Предложный оборот. Контактно справа стоит слово в родительном падеже (22%):

● *Психологический статус подростков в глазах следователей резко расходился с юридическим.*

Свободное словосочетание (30%):

● *Но я что-то особого счастья у него в глазах не заметила.*

3. В ДЕЛЕ – фраза содержит два лексических значения: или имеется в виду деятельность (работа), а также персональное дело (это будет свободным сочетанием двух слов), или указывается на действие, выраженное отглагольным существительным (и это будет предложным оборотом).

Свободное словосочетание.

Контактно справа от фраземы стоит слово в родительном падеже, которое начинается с прописной буквы (6%):

● *Наиболее ярко этот спор был виден в деле Светланы Бахминой.*

● *В деле Возрождения вскрылись новые обстоятельства.*

Контактно справа от фраземы стоит одушевленное существительное в родительном падеже, которое относится к классу «Человек» (*капитан, мельник, осужденный*) и не является словом *созидания* (2%):

● *!воскликнул я, — в деле архитектора-вандала нет ни одного смягчающего обстоятельства.*

Предложный оборот.

Контактно справа от фраземы стоит существительное в родительном падеже (36%):

● *Опять повторяю, что неправдоподобность в деле чувства есть вернейший признак истины.*

Контактно справа от фраземы стоят прилагательное или причастие в родительном падеже и за ним существительное также в родительном падеже, которое оканчивается на “-ии”, “-ия” (3%):

● *Это был триумф, признание приоритета нашей страны в деле мирного использования энергии атома.*

Иначе это **свободное словосочетание** (53%).

● *Природную ренту сегодня определяют люди, которые "в деле" с олигархами.*

● *В деле Гуммеля было всего четыре странички*⁶.

4. В ЛИЦЕ – фраза содержит два лексических значения: в одном случае имеется в виду лицо человека (это будет свободным сочетанием двух слов), или самого человека, представляющего кого-либо (и это будет предложным оборотом).

Свободное словосочетание.

Контактно справа от фраземы стоит одно из слов *его, её, их* в родительном падеже, а правее стоит слово, не имеющее родительного падежа (4%):

● *В лице её не было ни кровинки.*

⁶ Отметим, что в эту же группу попадают случаи, когда слово, стоящее после фраземы, отсутствует в словаре, и падеж его неизвестен.

Контактно справа от фраземы стоит слово в родительном падеже, а в окрестности фраземы находится одна из словоформ лемм БЛЕСНУТЬ, БОЛЕЗНЕННЫЙ, ДРОГНУТЬ, КОЛЕБАНИЕ, ИРОНИЯ, КОЛЕБАНИЕ, МЕЛЬКНУТЬ, ОТТАЛКИВАТЬ, ОТЧАЯНИЕ, ПОКОЙ, ПРОМЕЛЬКНУТЬ, ПРОЧЕСТЬ, ПРОЧИТАТЬ, СТАТЬ, ХВАТАТЬ, ЧЕРТА, ЭКЗАЛЬТИРОВАННЫЙ или ЯСНОСТЬ (5%):

● *В лице сиреневого джентльмена чего-то не хватает, по-моему.*

Предложный оборот. Контактно справа от фраземы стоит слово в родительном падеже (57%):

● *Ходатайства того, в лице которого говорит римская власть?*

Свободное словосочетание (34%):

● *Напрасно ты так испугался и изменился в лице.*

5. В ОБЛАСТИ – фразема содержит два лексических значения: или действие происходит в ограниченной части поверхности/пространства (это будет свободным сочетанием двух слов), или имеется в виду отдельная отрасль знания/деятельности, сфера каких-либо представлений, занятий и т.п. (и это будет предложным оборотом).

Для анализа фраземы используется двухэтапное правило.

Первый этап – поиск управляющего слова, которым может быть одно из трех слов, стоящих справа от фраземы, если оно является существительным в родительном падеже при условии, что контактно справа от фраземы тоже слово в родительном падеже (прилагательное):

● *К примеру, она сильнее в области топливно-энергетических ресурсов.*

● *Одной из постоянных тем повестки дня ЭКОСОС является укрепление координации в области чрезвычайной гуманитарной помощи ООН.*

Если контактно справа от фраземы стоит слово, не имеющее родительного падежа, то это – **свободное словосочетание** (9%):

● *У нас в области жители посёлков на трассе никогда не любили город.*

Если же контактно справа от фраземы стоит слово в родительном падеже, но ни одно из стоящих справа трех слов не является существительным, то это – **предложный оборот** (3%):

● *О том же противоречии в области нравственных и политических воззрений мы уже говорили выше.*

Второй этап – анализ управляющего слова.

Свободное словосочетание (2%). Если выбранное слово относится к классу «Части_живого» (голова, лоб, пупок), то это простое сочетание двух слов:

● *Нога заживает полностью через 3, 4, 5 месяцев, а грубый рубец в области колена остаётся на всю жизнь.*

Предложный оборот (86%):

● *К примеру, она сильнее в области топливно-энергетических ресурсов.*

6. В ОБЛАСТЬ – фразема содержит два лексических значения: или действие происходит в ограниченной части поверхности/пространства (это будет свободным сочетанием двух слов), или имеется в виду отдельная отрасль знания/деятельности, сфера каких-либо представлений, занятий и т.п. (и это будет предложным оборотом).

Для анализа фраземы, как и в предыдущем случае, используется двухэтапное правило.

Первый этап – поиск управляющего слова, которым может быть одно из трех слов, стоящих справа от фраземы, если оно является существительным в родительном падеже при условии, что контактно справа от фраземы стоит слово в родительном падеже, которое начинается со строчной буквы:

● *Ну, а как быть, если эта отсталость отходит в область прошлого?*

● *Мы вступаем в область тропического леса, который опоясал подножие горы.*

● *Мне пришлось ударить его коленом в область его вождедеющей части тела.*

Если контактно справа от фраземы нет слова в родительном падеже, то это – **свободное словосочетание** (27%):

● *Совсем уж кстати: родители по каким-то делам отправляются в область.*

Если же контактно справа от фраземы стоит слово, начинающееся со строчной буквы, в родительном падеже, но ни одно из стоящих справа трех слов не является существительным, то это – **предложный оборот** (3%):

● *Переводя выше сказанное в область законодательной и общей метрологии, мы сталкиваемся с понятиями "калибровка" и "поверка".*

Второй этап – анализ управляющего слова.

Свободное словосочетание. Выбранное слово относится к классу «Части_живого» (голова, лоб, пупок), «Личность Профессия Военные» (десантник, партизан, рыцарь), «Температура» (жар, температура, тепло), «Погода» (антициклон, бриз, холод), «Природные зоны» (всполье, залежи, лесопосадка) или является словом давления (28%):

● *Противник наносит удар стволом в область шеи.*

● *Напоследок они ввели в область миротворцев.*

● *Совершил ряд путешествий на Новосибирские острова и в область реки Яны. Из области повышенного давления воздушные массы устремляются в область пониженного давления.*

Предложный оборот (42%):

● *Ну, сделал небольшое отступление в область истории...*

7. В ПЛАНЕ – фразема содержит два лексических значения: в одном случае это изображение чего-либо на плоскости или система мероприятий для выполнения работ, операций и т.п. (это будет свободным сочетанием двух слов), в другом случае это та или иная сфера проявления или путь осуществления чего-либо (и это будет предложным оборотом).

Свободное словосочетание. Контактно справа или через один токен стоит словоформа одной из

лемм ВЫПУСК, ДОМ, ИЗДАТЕЛЬСТВО, КОМПАНИЯ, МЕРОПРИЯТИЕ, ПОРТ, ПОСТРОЙКА, ПРАВИТЕЛЬСТВО, РАБОТА, ФОРМА, ЦЕХ (3%):

● ...и прочих необходимо надобных строений, которые означены в плане Севастопольского порта.

● В плане правительства до 2018 года для них предусмотрели налоговые льготы и привилегии при участии в гостендерах.

Предложный оборот.

Контактно справа от фраземы стоит союз, частица или наречие, а правее – слово в родительном падеже (2%):

● Я не в смысле поиска житейского партнера, а в плане вообще перспектив?

Контактно справа от фраземы в родительном падеже стоит существительное, местоимение, прилагательное или причастие (80%):

● А в плане управления рисками стоит обратиться к программам личного страхования.

Свободное словосочетание (15%):

● Греки быстро нашли указанное в плане место.

8. В РАЙОНЕ – фразема содержит два лексических значения: или это местность, а также единица административно-территориального деления (это будет свободным сочетанием двух слов), или ее смысл соответствует предлогу около (и это будет предложным оборотом).

Свободное словосочетание. Контактно справа стоит прилагательное в родительном падеже, которое относится к классу «Время» (нынешний, зимний, современный) (3%):

● Василий Розанов, живший в Костроме у Боровкова пруда (в районе нынешнего цирка), делился таким впечатлением детства.

Предложный оборот.

Контактно справа стоит слово в родительном падеже, которое относится к одному из классов «Время» (лето, утро, час), «Количество» (19, 16-х, сорок) или «Живой / Молодой-Старый» (возраст, молодой, старый) (12%):

● Я забежал в нее давно, когда она в феврале находилась в районе 40 долларов.

Свободное словосочетание (85%):

● Туда бы еще Шамрая из ГУЖА, и будет праздник в районе.

9. В РОЛИ – фразема содержит два лексических значения: или это художественный образ на сцене, в фильме и т. д. (это будет свободным сочетанием двух слов), или имеется в виду проявление в качестве кого-либо, чего-либо (и это будет предложным оборотом).

Свободное словосочетание. Справа хотя бы одно из двух слов начинается с прописной буквы (20%):

● ...и никому не приходило в голову попробовать его в роли какого-нибудь Лира или хотя бы Мальволио. Предложный оборот Контактно справа от фраземы стоит слово в родительном падеже или кавычки (74%):

● В роли авалиста может выступать банк.

Свободное словосочетание (6%):

● Вы же сами подобную ситуацию на себя в роли примеряли.

10. В РУСЛЕ – фразема содержит два лексических значения: или это углубление в почве для водного потока (это будет свободным сочетанием двух слов), или имеется в виду путь движения, развития чего-либо (и это будет предложным оборотом).

Свободное словосочетание.

Контактно справа стоит существительное из класса «Природные зоны» (канал, река, Днепр) (4%):

● Между Петербургским и Васильевским островами в русле Средней Невы на плане обозначено несколько небольших безымянных островков.

Или в окрестности фраземы находится словоформа одной из лемм РЕКА, РУЧЕЙ, КАМЕНЬ, ВОДА, ПРОСОХНУТЬ (3%):

● В русле её много камней, образующих пороги с валами и прижимами, многие из которых достаточно сложны.

● Небольшая по размерам трубка обнажалась в русле пересыхающего летом ручья.

Предложный оборот (93%):

● В русле подобных размышлений лежит и другая немецкая мифологема того времени.

11. В РЯДУ – фразема содержит два лексических значения: или это относится к однородным предметам, расположенным друг за другом (это будет свободным сочетанием двух слов), или имеется в виду совокупность предметов, среди которых выделяется кто-либо или что-либо (и это будет предложным оборотом).

Предложный оборот – одно из двух слов непосредственно справа от фраземы имеет родительный падеж (79%):

● Какое из животных стоит первым в ряду одомашненных видов?

● С общего согласия процедура была названа "умывание по Браунингу" и заняла достойное место в ряду себе подобных, о которых я знаю.

Свободное словосочетание (21%):

● Проект стоит только шестым в ряду.

12. В СОСТАВЕ – фразема содержит два лексических значения [6]: или это совокупность людей, предметов, образующих какое-либо целое (это будет свободным сочетанием двух слов), или это количество, численность (и это будет предложным оборотом).

Предложный оборот.

Контактно справа стоит числительное в родительном падеже (двух, шести, 129) (4%):

● В секретарскую спокойной деловой походкой входила милиция в составе двух человек.

Контактно справа от фраземы не стоит знак препинания, и в окрестности влево от фраземы находится словоформа одной из лемм АРМИЯ, БАТАЛЬОН, БРИГАДА, ДИВИЗИЯ, ПОЛК, ФРОНТ, ЭСКАДРА, ГРУППА, ДЕПУТАЦИЯ, КОЛЛЕГИЯ, КОМАНДА,

КОМИССИЯ, КОНСОРЦИУМ, ПОСОЛЬСТВО, ПРЕЗИДИУМ, ЭКИПАЖ, ЭКСПЕДИЦИЯ, или ТРИО (6%)⁷:

● *Первый раз встречали его у стен монастыря, под колокольный звон, депутацией в составе архимандрита Антония, купца первой гильдии Филиппова, городского головы Пирожникова и городского Невецкого.*

Свободное словосочетание (90%):

● *А георгиевский крест он получил ещё за бои с немцами в составе русской армии.*

13. В СТЕНАХ – фраза содержит два лексических значения: или имеется в виду вертикальная часть какого-либо здания (это будет свободным сочетанием двух слов), или высказываемое относится к самому зданию (и это будет предложным оборотом).

Свободное словосочетание.

В окрестности фраземы имеется слово из классов «Фрагмент_постройки» (*арка, бельведер, крыша*) или «Отверстие» (*бойница, люк, трещина*) (16%):

● *В стенах рыжели ободренные железные двери с глазками.*

Контактно справа стоит слово не в родительном падеже, которое не является ни наречием, ни частицей (4%):

● *Но нам нравилось думать, что руины умывали с зацементированными в стенах зеркалами чудом сохранились.*

Контактно справа от фраземы стоит знак препинания или в ее окрестности находится словоформа одной из лемм ПЕЩЕРА, ПРОЛОМ, ПРОХОД, БУРИТЬ, ЗАМУРОВАТЬ (6%):

● *Спасибо, только там есть ещё всякие (забыл, как называется) виноградовавилки, вырезанные в стенах пещер...*

● *Я родила ребенка и семь месяцев сидела в стенах.*

Предложный оборот (74%):

● *Все события происходят в течение учебного года в стенах школьного здания.*

14. ВО ВЛАСТИ – фраза содержит два лексических значения: или пребывание в составе должностных лиц, начальства, администрации (это будет свободным сочетанием двух слов), или нахождение под могуществом, господством, силой кого-нибудь, чего-нибудь (и это будет предложным оборотом).

Свободное словосочетание. Контактно справа от фраземы стоит слово в родительном падеже, а непосредственно перед фраземой находится словоформа одной из лемм ДОЛЖНОСТЬ, ЗАКРЕПЛЕНИЕ, КОММУНИСТ, НЕЖЕЛАТЕЛЬНЫЙ, ПОЯВЛЕНИЕ, ПРЕБЫВАНИЕ, ПРЕДСТАВИТЕЛЬСТВО, УЧАСТИЕ, ВИДЕТЬ, ИСКАТЬ, СМЕНИТЬ (3%):

● *Все должно пройти честно и открыто, а предложенные механизмы исключают появление во власти случайных людей.*

Предложный оборот.

Контактно справа от фраземы стоит наречие, кавычки или частицы *ещё, лишь, не*, а вслед за ними прилагательное в родительном падеже (2%):

● *И все они были во власти не менее сильного наркотика⁸.*

Контактно справа стоит слово в родительном падеже, но не слова *вакуума, изменения, лицемерия, области, человек* (50%):

● *Но здесь мы находимся во власти законодателя.*

Свободное словосочетание (46%):

● *А это партнерство будет нуждаться в своём представительстве во власти.*

15. ИЗ ОБЛАСТИ – фраза содержит два лексических значения: или действие происходит в ограниченной части поверхности/пространства (это будет свободным сочетанием двух слов), или имеется в виду отдельная отрасль знания/деятельности, сфера каких-либо представлений, занятий и т. п. (и это будет предложным оборотом).

Для анализа фраземы используется двухэтапное правило.

Первый этап – поиск управляющего слова, которым может быть одно из трех слов, стоящих справа от фраземы⁹, если оно является существительным при условии, что контактно справа от фраземы стоит слово в родительном падеже, которое начинается со строчной буквы (69%):

● *Но это уже из области мечтаний.*

● *Это заявление может вызвать только смех, потому что оно из области чистой фантастики!*

Если контактно справа от фраземы стоит слово в родительном падеже, начинающееся со строчной буквы, но ни одно из стоящих справа трех слов не является существительным, то это – **предложный оборот** (4%):

● *Да и не могло выйти, равно как и сейчас из области невероятного – нападение на Иран.*

Иначе – это **свободное словосочетание** (26%):

● *Он считал, что язык хайда развился на основе языка иммигрантов из области Южных морей.*

Второй этап – анализ управляющего слова:

Свободное словосочетание.

Выбранное слово относится к классу «Части_живого» (*голова, лоб, пупок*), «Личность_Профессия_Военные» (*десантник, партизан, рыцарь*), «Температура» (*жар, температура, тепло*), «Погода» (*антициклон, бриз, холод*), или является словом давления (3%):

● *Средоточие твоего мужского начала – это что-то из области пуза?*

● *Показано, что динамика истечения определяется соотношением импульсов истекающего газа и вытесняемого из области течения фонового газа.*

● *Из области повышенного давления воздушные массы устремляются в область пониженного давления.*

⁸ Здесь словосочетание *не менее* выполняет функцию наречия.

⁹ Между этим словом и фраземой может стоять одно или два слова в родительном падеже.

⁷ Приведены наиболее частотные леммы.

Предложный оборот (67%):

- *Иллюстрацией из области биологии может служить вопрос о происхождении сна.*

16. НА ВОЛНЕ – в одном случае имеются в виду волновые или электромагнитные волны (это будет свободным сочетанием двух слов), в другом случае – повторяющиеся явления в жизни общества (и это будет предложным оборотом).

Свободное словосочетание.

Справа от фраземы в кавычках стоит слово, начинающееся с прописной буквы (2%):

- *Кое-где горели костры, работали транзисторные приемники на волне "Маяка".*

Контактно справа от фраземы стоит числительное или в ее окрестности находится словоформа одной из лемм ПРИЕМ, РАДИО, РАДИОЛЮБИТЕЛЬ, РАДИОСТАНЦИЯ, ЛЕЖАТЬ, ОТХОДИТЬ, ПАРОХОД или ПЛАВАНИЕ (6%):

- *Слышу хорошо Москву на волне 32,8.*
- *Все полярные радиостанции продолжают непрерывно слушать на волне самолета.*

Предложный оборот.

Контактно справа от фраземы отсутствует знак препинания или стоят кавычки, причем хотя бы одно из двух слов справа имеет родительный падеж (77%):

- *На волне Пушкинского юбилея 1937 года появилось стихотворное обращение к поэту.*
- *А вслед за такими признаниями на волне "загниженного бытовизма" прорвались и настоящие стихи.*

Свободное словосочетание (15%):

- *Здесь только голову приподняла, покачиваясь на волне.*

17. НА ГЛАЗАХ У – фразема содержит два лексических значения: в одном случае имеются в виду глаза человека (это будет свободным сочетанием трех слов), в другом – общественное (много реже, частное) мнение (и это будет предложным оборотом).

Свободное словосочетание.

В окрестности фраземы имеется словоформа одной из лемм ВЛАГА или СЛЕЗА (5%):

- *Слёзы на глазах у людей.*

Предложный оборот (95%):

- *Экономика Китая развивается не по дням, а по часам на глазах у всего мира.*

18. НА РУБЕЖЕ – фразема содержит два лексических значения: это или линия раздела между смежными территориями, позиция вдоль линии фронта (это будет свободным сочетанием двух слов), или последняя грань, предел (и это будет предложным оборотом).

Предложный оборот.

Контактно справа стоит слово из класса «Количество» (количественные или порядковые числительные) или цифры (29%):

- *На рубеже шестидесятых – семидесятых годов один за другим выплыли в свет два очень смешных романа.*

Контактно справа стоит слово в родительном падеже и одно из слов, расположенных правее фраземы, относится к классу «Время» (неолит, столетие, эра) (46%):

- *Случилось это якобы на рубеже веков в Англии.*

- *Он жил на рубеже средней и верхней юры, около 160 миллионов лет назад.*

Контактно справа стоит слово в родительном падеже, и одно из слов правее фраземы является словоформой от лемм ГОСУДАРСТВО, ДЕЙСТВИЕ, ИЗМЕНЕНИЕ, ИСКУССТВО, КРИЗИС, КУЛЬТУРА, ПРОСВЕЩЕНИЕ, РЕВОЛЮЦИЯ, СТАНОВЛЕНИЕ или ЦИВИЛИЗАЦИЯ (2%):

- *Надлом произошел на рубеже становления индустриальной цивилизации.*

- *Дело в том, что в 2002 году ЗАО "СМАРТС" стоит на рубеже своих главных изменений в области управления.*

Свободное словосочетание (23%):

- *Приближаясь к роце, стоящей на рубеже отцовского владения, Лиза пошла тише.*

- *Но ветер до трёх метров в секунду особо на рубеже не мешал.*

19. ПРИ ПОМОЩИ, С ПОМОЩЬЮ – каждая из этих фразем содержит два лексических значения: или помощь оказывается человеком сознательно (это будет свободным сочетанием двух слов), или для чего-либо используются материальные средства, денежные пособия, техника, продукты и т.п. (и это будет предложным оборотом).

Для анализа фраземы используется двухэтапное правило.

Первый этап – поиск управляющего слова, которым может быть одно из трех слов, стоящих справа от фраземы¹⁰, если оно является существительным в родительном падеже и выполняется одно из трех условий: слово однозначно, справа стоит знак пунктуации или слово, не имеющее родительного падежа:

- *Ноги вымыть с помощью щётки и обрезать когти...*

- *Оценку состояния водных экосистем можно проводить при помощи биологических тестов.*

- *Или с помощью специальных съедобных украшений "выведите" на них имена приглашённых.*

Если такое слово не найдено, то это – предложный оборот (20%):

- *Но стоит ли при помощи чисто административных ножиц решать столь сложные и деликатные вопросы?*

- *Механизмы реализации информационных атак включают в себя методы и средства, с помощью которых нарушители проводят атаку.*

Второй этап – анализ выбранного ранее слова.

¹⁰ Между этим словом и фраземой может стоять прилагательное, причастие или числительное в родительном падеже.

Предлоги, требующие после себя родительного падежа

Оборот	Связь с хозяином	НКРЯ	Выбрано	Частота предлога, %	Точность, %
БЕЗ ПОМОЩИ	Как	1789	433	40	97
В ГЛАЗАХ	вПред	168770	1787	28	96
В ДЕЛЕ	вПред	7448	513	38	97
В ЛИЦЕ	вПред	9203	381	57	97
В ОБЛАСТИ	вПред	18166	405	88	97
В ОБЛАСТЬ	вВин	2302	421	45	97
В ПЛАНЕ	Как	2286	464	83	96
В РАЙОНЕ	Где, Когда	11056	458	12	98
В РОЛИ	Как	5096	333	80	96
В РУСЛЕ	Как	578	286	93	99
В РЯДУ	Где	1523	318	79	97
В СОСТАВЕ	Какой	7087	327	9	99
В СТЕНАХ	Где	1905	304	74	98
ВО ВЛАСТИ	вПред	1995	512	51	96
ИЗ ОБЛАСТИ	Изо	1316	310	71	96
НА ВОЛНЕ	Как	606	452	77	99
НА ГЛАЗАХ У	Как	2093	349	95	99
НА РУБЕЖЕ	Когда	1820	545	77	98
ПРИ ПОМОЩИ	Как	10346	379	94	96
С ПОМОЩЬЮ	Как	22032	443	90	98

Свободное сочетание.

Выбранное существительное является одушевленным и относится к классу «Человек» (лоцман, Станиславский, человек) (7%):

• Но кардинал Ришелье подкупает Васко де Гама и при помощи деду Гамильтон добивается посылки вас в Америку.

• С помощью знакомых из ОКР мы пытались отыскать их в архивах.

Предложный оборот (73%):

• Кого власти подавляют при помощи судебной системы?

• Триггер DFF оказался реализованным с помощью процедуры DFFPC.

ЗАКЛЮЧЕНИЕ

С использованием парсера SemSin нами выполнен анализ омонимии устойчивых словосочетаний (фразем), могущих играть роль предлогов или быть простыми сочетаниями слов, а также разработаны правила, позволяющие с высокой точностью снимать частеречную и синтаксическую омонимию. В то же время снятие семантической омонимии представляет собой значительно более сложную задачу, требующую дополнительного изучения.

Рассматривая правила для снятия омонимии, следует отметить, что они опираются как на морфологию и классы классификатора, так и, в ряде случаев, на простое перечисление слов или лемм. Кроме слов, расположенных непосредственно справа от фразем, для поиска используются также их окрестности.

Общие результаты анализа собраны в таблицу, в которой приведены данные о количестве вхождений фраземы в НКРЯ и количестве её вхождений в файл,

выбранный для анализа. Частота предлога (количество случаев, когда фраза является предложным оборотом) лежит в диапазоне от 12 до 95%. Как видно из таблицы, предлагаемые правила снятия омонимии практически обеспечивают точность результатов порядка 95%.

Следует отметить, что все разобранные в настоящей статье фраземы обладают некоторой особенностью – каждая из них может иметь прямое семантическое значение и переносное. Причем как раз переносное значение и соответствует предложному обороту. Это свойство может помочь решить проблему снятия омонимии в конкретном случае.

Мы не утверждаем, что абсолютно все фраземы такого рода (имеющие по два варианта омонимии: простое сочетание слов и предложный оборот) будут иметь указанное свойство. Однако ясно, что это свойство присуще не только описанным нами фраземам. Например, им обладают фраземы *в сфере*, *в счёт* и *с целью*, которые очень редко, но могут иметь и прямое значение, хотя практически являются предложными оборотами с частотой не менее 97%. Что же касается фразем, которые имеют 3 вида омонимии – свободное сочетание слов, предложный оборот и наречный или предикативный оборот – эти фраземы еще подлежат дальнейшему исследованию.

СПИСОК ЛИТЕРАТУРЫ

1. Daud Ali, Li Juanzi, Zhou Lizhu, Muhammad Faqir. Knowledge discovery through directed probabilistic topic models: a survey // In Proceedings of Frontiers of Computer Science in China. – 2010. –

- Vol. 4, – P. 280-301. DOI: <https://doi.org/10.1007/s11704-009-0062-y>.
2. Avdeeva N., Artemova G., Boyarsky K., Gusarova N., Dobrenko N., Kanevsky E. Subtopic Segmentation of Scientific Texts: ParametrOptimisation. Communications in Computer and Information Science Knowledge Engineering and the Semantic Web // 6th International Conference, KESW 2015. – Moscow: Proceedings, 2015. – Vol. 518, – P. 3–15.
 3. Zakharov V., Golovina A., Azarova I. Statistical analysis of Russian multiword prepositions // NORDSCI 2020 Conference Proceedings (12–14 October 2020). – Sofia: Saima Consult Ltd, Bulgaria. – Vol. 3, Book 1. – P. 191-200.
 4. Каневский Е.А., Клименко Е.Н., Силина Е.Ф. Анализ словосочетаний, эквивалентных слову (часть I) // Экономико-математические исследования: математические модели и информационные технологии. – Санкт-Петербург: Нестор-История, 2016. – С. 69–88.
 5. Боярский К.К., Каневский Е.А. Семантика устойчивых словосочетаний с глаголами // Научно-техническая информация. Сер. 2. – 2019 – № 11. – С. 23–31.
 6. Национальный корпус русского языка. – URL: <http://www.ruscorpora.ru/> (дата обращения: 22.02.2021).
 7. Кузнецов С.А. Большой толковый словарь русского языка. – Санкт-Петербург: Норинт, 1998.
 8. Рогожникова Р.П. Толковый словарь сочетаний, эквивалентных слову. – Москва: ООО «Издательство Астрель», 2003.
 9. Zakharov V., Golovina A., Alexeeva E., Gudkov V. Russian Secondary Prepositions: Methodology of Analysis // CEUR Workshop Proceedings. Computational Models in Language and Speech – 2020. – Vol. 2780. – P. 187-201.
 10. Боярский К.К., Каневский Е.А., Клименко Е.Н. Устойчивые словосочетания в роли предлогов. Компьютерная лингвистика и вычислительные онтологии. Вып. 5 // Сборник научных трудов XXIV Международной объединенной научной конференции «Интернет и современное общество», IMS-2021 (Санкт-Петербург, 24–26 июня 2021 г.). – Санкт-Петербург: Университет ИТМО, 2021. – С. 17–28.
 11. Боярский К.К., Каневский Е.А. Семантико-синтаксический парсер SEMSIN // Научно-технический вестник информационных технологий, механики и оптики. – 2015. – Т. 15, №5. – С. 869–876.
 12. Боярский К.К., Каневский Е.А., Стафеев С.К. Использование словарной информации при анализе текста // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – №3(79). – С. 87–91.
 13. Боярский К.К., Каневский Е.А. Язык правил для построения синтаксического дерева // Материалы XIV Всероссийской объединенной конференции «Интернет и современное общество». Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики. – Санкт-Петербург, 2011. – С. 233–237.

Материал поступил в редакцию 20.05.22.

Сведения об авторах

БОЯРСКИЙ Кирилл Кириллович – кандидат физико-математических наук, доцент, старший преподаватель Университета ИТМО; старший научный сотрудник Института проблем региональной экономики РАН, Санкт-Петербург
e-mail: boyarin9@yandex.ru

КАНЕВСКИЙ Евгений Александрович – кандидат технических наук, старший научный сотрудник, ведущий научный сотрудник Института проблем региональной экономики РАН, Санкт-Петербург
e-mail: eak300@mail.ru

КЛИМЕНКО Евгения Николаевна – научный сотрудник Института проблем региональной экономики РАН
alexklim2000@yandex.ru



Федеральное государственное бюджетное учреждение науки
Всероссийский институт научной и технической информации
Российской академии наук
(ВИНИТИ РАН)

ИНФОРМАЦИОННОЕ ПИСЬМО

о проведении 10-й научной конференции с международным участием

«НТИ-2022. Научная информация в современном мире:

глобальные вызовы и национальные приоритеты»,

посвященной 70-летию ВИНИТИ РАН

(25 – 26 октября 2022 г.)

Всероссийский институт научной и технической информации Российской академии наук (ВИНИТИ РАН), один из крупнейших мировых информационных центров, приглашает принять участие в работе **10-й научной конференции с международным участием «НТИ-2022. Научная информация в современном мире: глобальные вызовы и национальные приоритеты».**

Основная цель конференции – обсуждение актуальных проблем информационного обеспечения науки, аналитической и инновационной деятельности, разработки и внедрения методов интеллектуальной обработки информации, а также обмен опытом между профильными научными, производственными и учебными организациями.

Для участия в конференции приглашаются ученые и специалисты РАН, вузовской и отраслевой науки в области информационных технологий, работники информационных центров и библиотек, служб распространения информационных продуктов и услуг, представители издательств.

На «НТИ-2022» предполагается обсуждение следующих вопросов:

- информационное обеспечение научных исследований в условиях санкционного давления;
- роль реферативных служб в современном мире;
- информационно-библиотечная деятельность и обслуживание;
- развитие классификационных систем;
- проблемы науковедения и наукометрии;
- интеллектуальные технологии обработки информации;
- международное сотрудничество в сфере НТИ;
- методы и технологии популяризации научных знаний.

На конференции планируется проведение пленарного заседания, круглых столов и тематических секций, в том числе отдельной секции молодых ученых, интересующихся проблемами научно-информационного обеспечения научных исследований.

Рабочие языки конференции: русский и английский.

Формат участия: очно, заочно.

Участие в конференции бесплатное.

По итогам конференции будет подготовлен сборник докладов (РИНЦ), лучшие из которых, отобранные экспертами Программного комитета, предполагается опубликовать в журналах «Научно-техническая информация. Серия 1. Организация и методика информационной работы» (РИНЦ, переводная версия издания включена в международные индексы цитирования Scopus, Web of Science) и «Научно-техническая информация. Серия 2. Информационные процессы и системы» (входит в Russian Science Citation Index – RSCI, переводная версия издания включена в международный индекс цитирования Web of Science).

Требования по оформлению представлены на странице конференции <http://www.viniti.ru/news/349-10-ya-nauchnaya-konferentsiya-posvyashchennaya-70-letiyu-viniti-ran>

Прием докладов – до 20 сентября 2022 г.

Публикация программы на сайте конференции – после 18 октября 2022 г.

Конференция будет проходить по адресу: г. Москва, ул. Усиевича, д. 20

Дополнительная информация:

E-mail: viniti@viniti.ru

Контактное лицо:

Секретарь конференции

Камнева Ирина Евгеньевна

Член оргкомитета конференции

Самсонова Анастасия Станиславовна

Контактные телефоны:

+7 (499) 155-45-12, +7 (499) 155-43-33

УВАЖАЕМЫЕ КОЛЛЕГИ!

ВИНИТИ РАН предлагает Вашему вниманию Реферативный Журнал в электронной форме

РЖ в электронной форме (ЭлРЖ) выпускается по всем разделам естественных, технических и точных наук.

Каждый номер ЭлРЖ является полным аналогом печатного номера РЖ по составу описаний документов, их оформлению и расположению. Он сопровождается оглавлением, указателями.

ЭлРЖ представляет собой информационную систему, снабженную поисковым аппаратом и позволяющую пользователю на персональном компьютере:

- читать номер РЖ, последовательно листая рефераты;
- просматривать рефераты отдельных разделов по оглавлению;
- обращаться к рефератам по указателям авторов, источников, ключевых слов;
- проводить поиск документов по словам и словосочетаниям;
- выводить текст описаний документов во внешний файл.

ЭлРЖ могут быть:

- записаны на DVD-ROM;
- передаваться через FTP-сервер (клиенту предоставляется логин и пароль с доступом к FTP-серверу ВИНТИ, с которого он скачивает заказанные журналы).

Электронные реферативные журналы можно заказать за текущий год с любого номера, а также за предыдущие годы.

Подробную информацию Вы можете получить:

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

Телефон: 8 499-152-62-11

E-mail: feo@viniti.ru

УВАЖАЕМЫЕ КОЛЛЕГИ!

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

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

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

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

На основе БД ВИНИТИ пользователям доступны следующие продукты:

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

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

Подробную информацию Вы можете получить:

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

Телефон: 8 499-152-54-81

E-mail: feo@viniti.ru