

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

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

Издается с 1961 г.

№ 4

Москва 2018

ОБЩИЙ РАЗДЕЛ

УДК 001.4:001.895:002

Л.К. Бобров, И.П. Медянкина

О влиянии понятийного аппарата на постановку задач информационной поддержки инновационной деятельности

Приводятся различные трактовки понятий «инновационная экономика», «инновационная деятельность» и «инновации». Следствием неустоявшейся терминологии является разнообразие смысловой нагрузки термина «инновационная инфраструктура». Вытекающее из этого многообразие подходов к решению задач информационной поддержки инноваций иллюстрируется на примере анализа нормативно-правовых и программных документов восьми регионов Сибирского федерального округа. Как альтернатива созданию трудно реализуемой единой федеральной точки ориентирования в информационном пространстве предлагается построение сети региональных систем, обеспечивающих навигацию и доступ к соответствующим ресурсам информационного обеспечения инновационной деятельности. В привязке к внешним информационным ресурсам приводится модель жизненного цикла инноваций, охватывающая все этапы от генерации идей до вывода инновационного продукта с рынка.

Ключевые слова: инновации, инновационная экономика, инновационная деятельность, терминология, руководящие документы, инфраструктура инноваций, информационные ресурсы, навигация, информационные системы

ВВЕДЕНИЕ

Становление инновационной экономики предполагает активную роль регионов в реализации программ инновационного развития. В качестве региональных драйверов инноваций выступают, прежде всего, технополисы, зоны развития новых и высоких технологий, научно-технологические парки, инновационно-технологические центры и центры коммерциализации технологий, бизнес-инкубаторы и др. Разработка механизмов, обеспечивающих функционирование единой региональной информационно-аналитической рабочей среды, является приоритетным направлением для задач информационной поддержки инноваций. Острота вопроса наиболее сильно ощущается в инновационных проектах мультидисциплинарного характера, когда тесная кооперация информационных технологий и предметных областей инновационной деятельности способствует более качественному выполнению работ и осознанию важности формирования эффективной среды информационной поддержки всех этапов жизненного цикла инноваций.

В формировании базовой инфраструктуры инноваций внешние информационные ресурсы и технологии играют определяющую роль, поскольку их грамотное использование позволяет инноваторам сконцентрироваться на решении профессиональных задач, полноценно опираясь на уже существующие в мире наработки научно-технического характера. Отсутствие же единой региональной информационно-аналитической рабочей среды приводит к тому, что большие объемы формально публичной информации труднодоступны или недоступны внешним потребителям, что в существенной мере сдерживает развитие инновационной экономики (в частности, экономики Новосибирской области как региона с высоким научно-техническим потенциалом). Однако создание такой среды сопряжено с необходимостью решения ряда задач, включающих элементы методологического характера. В том числе это касается понятийного аппарата инноватики и его опосредованного влияния на постановку задач информационной поддержки инновационной деятельности на уровне отдельных регионов страны.

ПОНЯТИЙНЫЙ АППАРАТ

Памятуя о мудром высказывании Рене Декарта о том, что люди избавились бы от половины своих неприятностей, если бы смогли договориться о значении слов, имеет смысл прежде всего рассмотреть существующие трактовки понятий «инновационная экономика», «инновационная деятельность» и «инновации». Тем более, что эти понятия трактуются далеко неоднозначно.

Например, под «инновационной экономикой» понимается «экономика, основанная на существенном и массовом использовании научных знаний в производстве товаров и услуг. Вес научных знаний в производстве таких потребительных стоимостей, особенно с помощью высоких технологий, может составлять значительную часть (до 10–15%) от общей стоимости произведенного продукта» [1].

В толковом словаре «Инновационная деятельность» [2] термин «инновационная экономика» отсутствует, но есть следующее синонимичное определение: «Экономика знаний – экономика, основанная на производстве, обновлении, циркуляции, распределении и применении знаний. Это экономика, которая использует знания для обеспечения своего роста и конкурентоспособности... Продукты экономики знаний – научная и разнообразная высокотехнологичная продукция, высококвалифицированные услуги, образование».

Согласно Википедии «инновационная экономика (экономика знаний, интеллектуальная экономика) – тип экономики, основанной на потоке инноваций, на постоянном технологическом совершенствовании, на производстве и экспорте высокотехнологичной продукции с очень высокой добавочной стоимостью и самих технологий...» [3].

Несколько сложнее дело обстоит с представлением о том, что такое «инновационная деятельность». Согласно Федеральному закону «О науке и государственной научно-технической политике» [4] инновационная деятельность есть «...деятельность (включая научную, технологическую, организационную, финансовую и коммерческую деятельность), направленная на реализацию инновационных проектов, а также на создание инновационной инфраструктуры и обеспечение ее деятельности...».

Словарь бизнес-терминов [5] гласит, что инновационная деятельность – это «... деятельность, связанная с применением, внедрением новых идей (научных исследований и научно-технических разработок), усовершенствованных изделий, технологических процессов...».

В толковом словаре "Инновационная деятельность" [2] последняя определяется «...в узком смысле... как ...деятельность, связанная с использованием результатов фундаментальных научных исследований для разработки новой продукции или технологии...». Другие примеры определений, свидетельствующие о многообразии трактовок этого понятия, можно увидеть в [6–12].

Еще сложнее с трактовкой понятия «инновация»: поиск значений этого термина по словарям и в интернете дает десятки различных дефиниций, а в работе [13] отмечается, что в литературе насчитываются сотни определений, и авторы приводят 16 авторских трактовок этого понятия, взятых из монографий и учебных пособий. Заметим, что такое разнообразие мнений существует на фоне почти повсеместного употребления данного термина в различных выступлениях, публикациях, планах и пр. Вкрадывается подозрение, что каждый из выступающих (авторов, составителей, и др.) вкладывает в этот термин свой смысл.

Данную ситуацию подтверждают результаты всероссийских опросов [14], проведенных ВЦИОМ в 2007-2011 гг.: 53% опрошенных затруднились ответить на вопрос «Что такое, по вашему мнению, «инновация»? Попробуйте определить это понятие», 27% – ответили, что это абстрактные нововведения, 15% – внедрение современных технологий, 3% – использование достижений науки и техники (научно-технический прогресс).

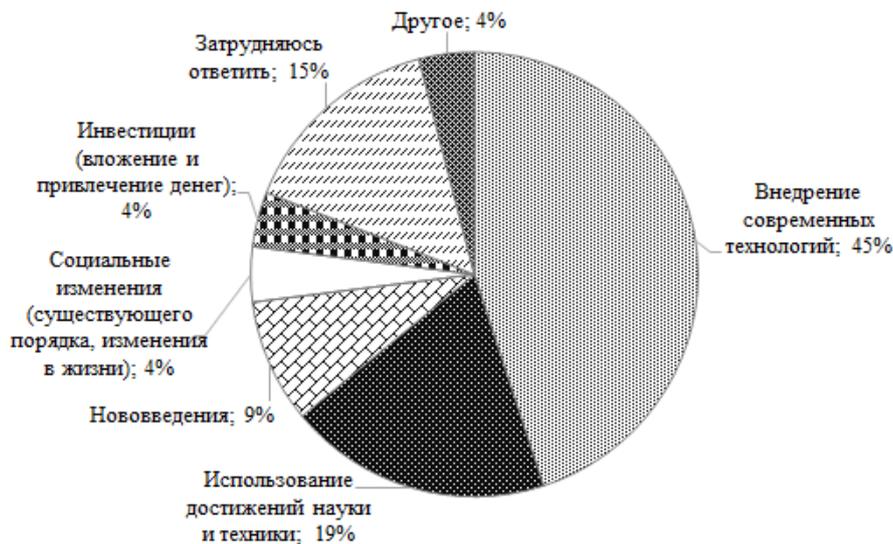


Рис. 1. Результаты ответа на вопрос «Что такое, по вашему мнению, «инновация»?»

Более поздние всероссийские опросы ВЦИОМ 2010-2016 гг. [15] показали несколько иную картину – затруднились с ответом только 15% опрошенных (рис. 1), однако разнообразие мнений существенно не сократилось.

Согласно различным словарям, авторским определениям, а также законодательным и нормативным документам, при всем разнообразии формулировок, инновация трактуется в одних случаях как процесс, в других – как результат [16–23].

Существует и третий вариант, когда понятие «инновация» рассматривается одновременно и как процесс, и как результат. Аргументированное обоснование такой точки зрения представлено в работе [24], а наиболее ярким примером может служить определение, сформулированное в проекте документа «Стратегия инновационного развития Российской Федерации на период до 2020 года» [25] (см. выделенное жирным шрифтом) и снабженное там примечательной ссылкой (см. курсив):

«Инновации – введенный в употребление новый или значительно улучшенный продукт (товар, услуга) или процесс, новый метод продаж или новый организационный метод в деловой практике, организации рабочих мест или во внешних связях». Термин может быть использован как для описания самого процесса создания новых продуктов, процессов, рынков (в этом случае синонимом является термин «инновационная деятельность»), так и для описания результата этого процесса (синонимы – «продукт (результат) инновации», «инновационная продукция (услуга, бизнес-модель, технология и т.д.)). Близким по значению является термин «нововведение».

В межгосударственном стандарте ГОСТ 31279-2004 «Инновационная деятельность. Термины и определения» [26] представлено не только общее определение термина «инновация» – «Новые или усовершенствованные технологии, виды продукции или услуг, а также организационно-технические решения производственного, административного, коммерческо-

го или иного характера, способствующие продвижению технологий, товарной продукции и услуг на рынок», но и рассматриваются продукт-инновация: «Инновация, связанная с разработкой и внедрением новой или усовершенствованной продукции (изделий) или уже реализованных в производственной практике других предприятий и распространяемых через технологический обмен (беспатентные лицензии, ноу-хау, консультации)» и процесс-инновация: «Инновация, связанная с разработкой и внедрением новых или значительно улучшенных производственных процессов, предполагающих применение нового производственного оборудования, новых методов организации производственного процесса или их совокупности».

Дискуссии терминологического плана и попытки найти единое, научно обоснованное и всеобъемлющее толкование термина «инновация» для однозначного понимания практических задач отражены в многочисленных публикациях [11, 24, 27, 28]. Более того, в работе [29] несовершенство понятийного аппарата рассматривается как один из факторов риска реализации стратегии «Инновационная Россия-2020»¹.

Неоднозначность толкования понятия «инновация» порождает и другие терминологические проблемы, оказывающие непосредственное влияние на практическую реализацию задач инновационного развития. Так, авторы работы [30], по результатам анализа внушительного перечня нормативных правовых актов, обращают внимание на неустоявшуюся

¹ Авторы данной работы ссылаются на то, что «...в сентябре 2011 г. президиумом Правительства РФ был одобрен проект Стратегии инновационного развития Российской Федерации на период до 2020 г...» и критически оценивают понятийный аппарат, приведенный в *Приложении 2* к данному документу. Однако в окончательную опубликованную редакцию Стратегии, утвержденную Распоряжением Правительства РФ от 8 декабря 2011 г. № 2227-р (<http://www.garant.ru/products/ipo/prime/doc/70006124/#ixzz4V4knaTwY>) *Приложение 2*. «Основные термины, используемые в Стратегии» не вошло.

терминологию и разнообразие смысловой нагрузки применительно к понятию «инновационная инфраструктура», что, как они справедливо отмечают, оказывает влияние на формирование правильного представления о составляющих национальной инновационной системы и адекватность оценки их функционирования. При этом состав элементов инфраструктуры варьируется в различных нормативных правовых актах, в то время как «...для достижения большей эффективности деятельности инновационной инфраструктуры России, понимания уровня ее развития в инновационной системе в целом, необходимо детальное понимание того, что она из себя представляет, включая ее региональный разрез...».

Помимо изложенного следует отметить, что существующие нормативно-правовые и программные документы далеко не всегда четко и однозначно характеризуют информационную составляющую как важную неотъемлемую компоненту инфраструктуры поддержки инноваций.

ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ КАК КОМПОНЕНТА ИНФРАСТРУКТУРЫ ПОДДЕРЖКИ ИННОВАЦИЙ (на примере Сибирского федерального округа)

Многие документы международного и федерального уровня [10, 31–33] указывают на то, что информационное обеспечение инноваций является важной инфраструктурной компонентой. В связи с этим представляет интерес краткое рассмотрение информационной инфраструктуры инновационной сферы через призму нормативно-правовых и программных документов субъектов РФ, входящих в состав Сибирского федерального округа (таблица). В качестве таковых выбраны регионы, входящие в первые пять десятков по рейтингу инновационного развития субъектов РФ [34].

Приведенные в таблице выдержки из утвержденных концепций, стратегий и программ развития регионов отражают соответствующие взгляды на роль, место и формы информационного обеспечения в решении задач развития инновационной экономики. В одних случаях внимание акцентируется на информационной поддержке процессов продвижения инновационной продукции, в других – на формировании и использовании официальной статистической информации, а иногда – на использовании ИТ для решения имиджевых задач, или – на информационном обеспечении процессов принятия решений региональными властями, и т.п. В некоторых документах институты информационной поддержки инноваций определены напрямую, или – опосредованно, или – неким образом подразумеваются либо отсутствуют вообще.

В целом же ни один из приведенных в таблице документов (равно как и другие, не приведенные здесь ввиду ограниченности объема публикации [35–37]), не дает четкого представления о видении системы информационной поддержки инноваций на всем протяжении их жизненного цикла: от возникновения идеи до вывода инновационного продукта с рынка. С известной долей вероятности можно предположить, что одной из причин этого является понимание инновации именно как внедренного (или нуждающегося в этом) и уже объективно существующего новшества. Пример другого рода: приведенное в Википедии и отражаемое в ряде работ [38, 39] утверждение о том, что в России ещё не сложилась информационная инфраструктура, которая обеспечила бы информационное сопровождение инновационного цикла от возникновения идеи до её внедрения в жизнь, – иллюстрирует сосредоточение на информационной поддержке исключительно процессов разработки инноваций до момента их внедрения (вывода на рынок), не принимая во внимание последующие этапы жизненного цикла.

Отражение задач информационной поддержки инноваций в нормативно-правовых и программных документах регионов Сибирского федерального округа

Регион	Документ	Точка зрения на информационную поддержку инноваций
Омская область (50 место в рейтинге)	Концепция развития инновационной инфраструктуры на территории Омской области до 2015 г. – URL: http://www.projects.innovbusiness.ru/content/document_r_74DF3429-B093-4303-AE2B-80744874CD32.html (дата обращения 11.11.17).	«...Мероприятия по информационной поддержке инноваций направлены на сбор, анализ и распространение: – сведений о разработчиках инновационных продуктов и технологий; – сведений о потребителях инноваций; – сведений о кадровых ресурсах Омской области в сфере инноваций; – информации о технологической базе инноваций, отраслевых инновационных центрах; – информации об имеющихся производственных мощностях и их технико-технологическом уровне; – сведений об источниках и механизмах финансирования инноваций; – сведений о страховых инструментах снижения инновационных рисков».

Регион	Документ	Точка зрения на информационную поддержку инноваций
Бурятия (49 место в рейтинге)	Концепция промышленной политики Республики Бурятия на период 2013 – 2017 гг. до 2025 г. – URL: http://docs.cntd.ru/document/473813294 (дата обращения 11.11.17).	<p>В разделе «Информационная поддержка промышленной политики» констатируется, что «... Одной из причин сформировавшейся в настоящее время диспропорции по инвестициям частного бизнеса в сферу промышленного производства республики, является отсутствие... достоверной экономической информации о стратегической прибыльности определенного бизнеса, т.е. о спросе и издержках в нем...»</p> <p>Для устранения информационных диспропорций ключевыми задачами государственных органов являются: ...представление достоверной и полной информации об экономической привлекательности бизнеса в различных отраслях...</p> <p>Информационные системы будут создаваться в целях обеспечения субъектов промышленной деятельности... экономической, правовой, статистической, производственно-технологической информацией, информацией в области маркетинга, необходимой для развития субъектов промышленной деятельности...».</p>
Кемеровская область (40 место в рейтинге)	Государственная программа Кемеровской области "Экономическое развитие и инновационная экономика Кузбасса" на 2014 - 2019 гг. – URL: http://docs.cntd.ru/document/412805057 (дата обращения 11.11.17).	<p>Задачи Государственной программы включают «...Развитие инновационной инфраструктуры, обеспечивающей внедрение инноваций...».</p> <p>Подпрограмма "Стимулирование инноваций" предусматривает выполнение ряда мероприятий, включая "Информационное обеспечение инновационной деятельности", в рамках которого «...предполагается размещение информационных материалов в средствах массовой информации, в том числе за рубежом, организация выпуска справочной и методической литературы по вопросам развития инновационной деятельности в Кемеровской области. Подготовленные репортажи будут размещаться на информационных лентах, официальных сайтах информагентств...»</p> <p>Другими формами реализации станут:</p> <ul style="list-style-type: none"> – публикация интервью с руководителями исполнительных органов государственной власти Кемеровской области; – организация и проведение пресс-конференций на тему развития инновационного сектора области...».
Иркутская область (31 место в рейтинге)	Государственная программа Иркутской области "Экономическое развитие и инновационная экономика" на 2015–2020 гг.– URL: http://docs.cntd.ru/document/424069250 (дата обращения 11.11.17).	<p>Инфраструктура инновационной деятельности Иркутской области включает 15 элементов регионального значения. Это технопарки, бизнес-инкубаторы, центры трансфера технологий, центры коллективного пользования научным оборудованием и приборами, консалтинговые организации, а также центры переподготовки и повышения квалификации кадров инновационной сферы.</p> <p>Подпрограмма "Поддержка и развитие малого и среднего предпринимательства в Иркутской области на 2015–2020 годы" включает задачу содействия деятельности организаций, образующих инфраструктуру поддержки малого и среднего предпринимательства и развитие научной, научно-технической и инновационной деятельности в Иркутской области.</p>

Регион	Документ	Точка зрения на информационную поддержку инноваций
Алтайский край (27 место в рейтинге)	Государственная программа Алтайского края "Экономическое развитие и инновационная экономика" на 2015 - 2020 гг. – URL: http://docs.cntd.ru/document/423852817 (дата обращения 11.11.17).	<p>Для достижения цели подпрограммы "Стимулирование инноваций" будут реализованы мероприятия, способствующие стимулированию спроса на инновационную продукцию; формированию конкурентоспособного сектора научных исследований и разработок; созданию и развитию объектов инновационной инфраструктуры; обеспечению эффективного взаимодействия субъектов инновационной деятельности.</p> <p>Подпрограмма "Совершенствование системы формирования и использования официальной статистической информации в рамках регионального плана статистических работ" нацелена на</p> <p>«...создание условий для обеспечения Губернатора Алтайского края, Председателя Правительства Алтайского края и органов исполнительной власти Алтайского края актуальной и достоверной статистической информацией...».</p>
Новосибирская область (11 место в рейтинге)	Стратегия социально-экономического развития Новосибирской области на период до 2025 г. – URL: https://www.nso.ru/sites/test.new.nso.ru/wodby_files/files/migrate/activity/socio-economic_policy/strat_plan/documents/1654.pdf (дата обращения 11.11.17).	<p>В разделе «Основные проблемы развития инновационной системы Новосибирской области отмечается, что «...разрывы в инновационном цикле возникают при переходе от фундаментальных и прикладных исследований к коммерческим технологиям из-за... неразвитости инновационной сети информационной инфраструктуры... Преодоление выделенных структурных диспропорций и дисбалансов между спросом и предложением в значительной степени зависит от развития институтов инновационной инфраструктуры: правовой, финансовой, организационной, информационной и др.»</p> <p>Раздел «Целевые задачи по развитию региональной инновационной системы» включает «...создание доступных информационных источников по проблемам функционирования экономики знаний, организацию интерфейса между производителями и потребителями знаний...».</p> <p>Раздел «Механизмы реализации инновационной стратегии» предусматривает «...участие в международной информационной среде для поддержки конкурентоспособности инновационной деятельности».</p>
Красноярский край (10 место в рейтинге)	Стратегия инновационного развития Красноярского края на период до 2020 г. – URL: www.krskstate.ru/dat/bin/art_attach/1609_strategy2020.doc (дата обращения 11.11.17).	<p>«Основные направления реализации Стратегии в области формирования системы информационного обеспечения инновационной деятельности:</p> <ol style="list-style-type: none"> 1) создание информационного портала в области развития и поддержки инновационной деятельности на территории края, обеспечивающего публичность оказания государственной поддержки инновационной деятельности...; 2) формирование баз данных технологических и организационных инноваций в соответствующих сферах, создание необходимых интернет-ресурсов для их публикации и обсуждения;

Регион	Документ	Точка зрения на информационную поддержку инноваций
		3) активная информационная поддержка развития инновационной деятельности на территории Красноярского края, направленная на повышение престижа и привлекательности инновационной деятельности в Красноярском крае; 4) формирование обновляемого информационного ресурса о научно-технической и инновационной продукции, производимой в Красноярском крае ...».
Томская область (9 место в рейтинге)	Закон об инновационной деятельности в Томской области. – URL: https://duma.tomsk.ru/upload/site/2017/03/21 (дата обращения 11.11.17).	К организационно-информационным мерам поддержки инновационной деятельности относятся: «...1) экспертное и консультативное обеспечение инновационной деятельности; 2) организационная поддержка участия субъектов инновационной деятельности при проведении выставок, конференций, форумов, ярмарок и иных информационно-рекламных мероприятий; ... 8) учет, хранение и актуализация информации о субъектах и объектах инновационной деятельности (в том числе ведение реестров субъектов и объектов инновационной деятельности); 9) проведение исследований и анализа состояния и развития инновационной деятельности в Томской области».

ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ НА ПРОТЯЖЕНИИ ЖИЗНЕННОГО ЦИКЛА ИННОВАЦИЙ

Приведенный краткий терминологический обзор и рассмотрение нормативно-правовых и программных документов свидетельствуют об отсутствии однозначного ответа на вопрос о том, что включает задача информационной поддержки инноваций и какого рода информационные ресурсы необходимы для ее успешного решения.

На этом основании информационное обеспечение на протяжении жизненного цикла инноваций предлагается осуществлять на основе процессного подхода к определению термина «инновация». Не рискуя вступать в терминологические дискуссии и давать авторскую трактовку данного термина, в качестве исходного прием развернутое определение, предлагаемое философским словарем [40].

«ИННОВАЦИЯ (нововведение) – процесс создания, освоения и практической реализации научно-технических достижений. Процесс инновации обычно включает следующие фазы: фундаментальные исследования, приводящие к научным открытиям; прикладные научные исследования и инженерное творчество, приводящие к созданию изобретений в виде устройств, способов или веществ; разработка и испытание экспериментальных образцов новых изделий, технологий, материалов и т.п.; проектирование новой техники и её промышленное освоение; реализация первых промышленных образцов и при достижении коммерческого успеха расширение

производства; фаза диффузии (проникновения) инновации в другие области науки, техники, производства, духовной сферы и т.д... Инновация имеет свой «жизненный цикл», соответствующий указанным фазам. «Жизненный цикл» инновации можно моделировать, проектировать, прогнозировать, что позволяет осуществлять управление процессом инновации и его ускорением. Инновация включает как одну из основных частей маркетинг и маркетинговые исследования».

Исходя из приведенного определения и детализируя содержание, направленность и приоритеты работ на отдельных этапах, приведем модель жизненного цикла инноваций в соответствии с внешними информационными ресурсами, необходимыми для успешного выполнения работ (рис. 2). Показанный на рисунке далеко не полный перечень ресурсов наглядно свидетельствует о том, что информационная поддержка жизненного цикла инноваций должна базироваться на комплексном использовании документальных и фактографических источников (на что справедливо обращали внимание авторы работы [41] и ряда других работ), отражающих следующие основные тематические и видовые аспекты:

- информацию о рынках;
- сведения о программах и направлениях исследований по практическому использованию результатов фундаментальных НИР;
- маркетинговую информацию;
- информацию о потенциальных инвесторах и конкурсах грантов;

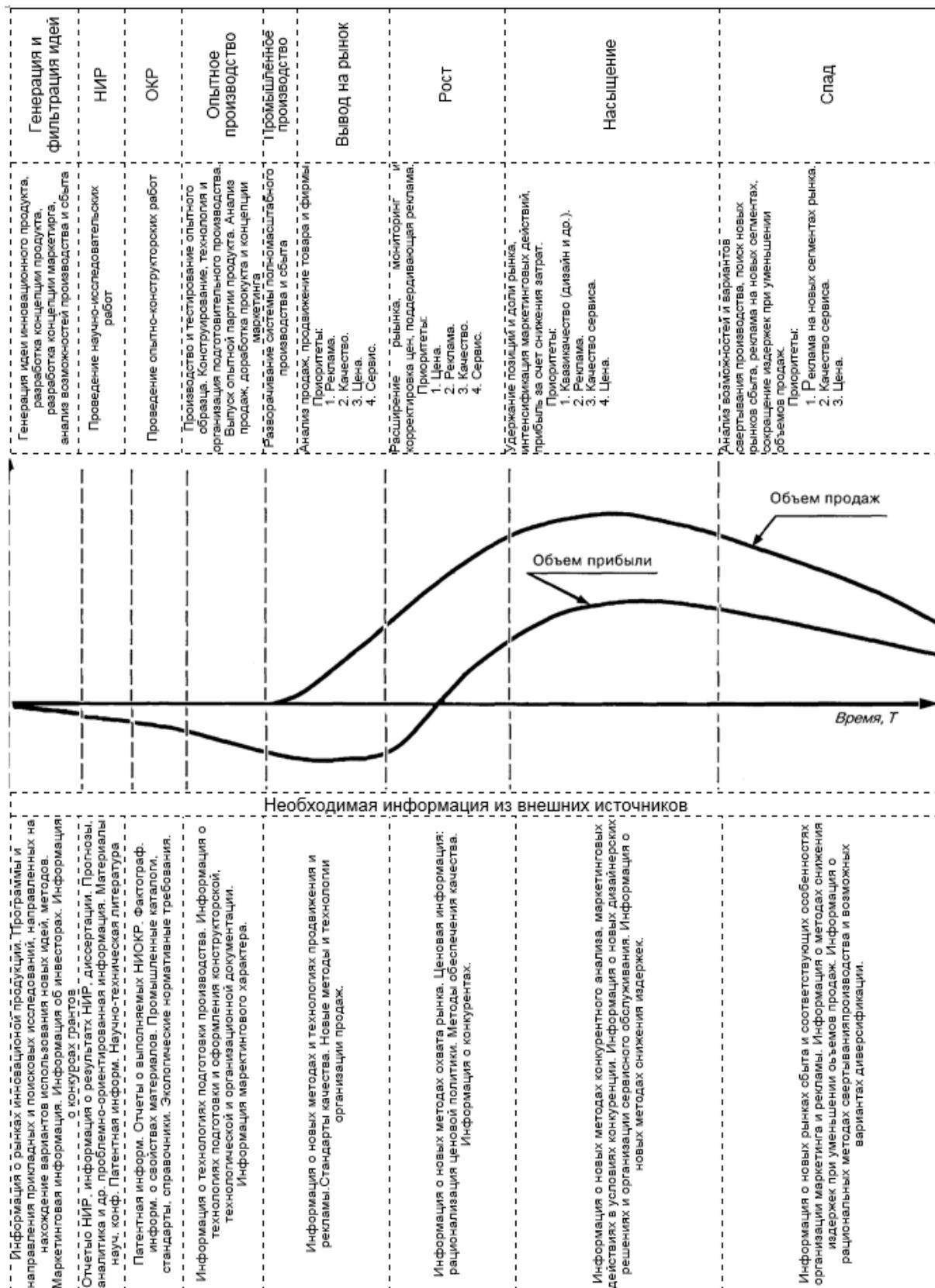


Рис. 2. Модель жизненного цикла инноваций в привязке к внешним информационным ресурсам

- научно-техническую и патентную информацию, отчеты о НИР и диссертации;
- материалы научных конференций;
- прогнозно-аналитическую информацию;
- конъюнктурно-коммерческую информацию;
- правовые и нормативные акты, стандарты и справочники;
- экономическую информацию (включая менеджмент), и др.

Перечисленные сведения можно почерпнуть из многочисленных баз и банков данных, доступных через сеть Интернет. При этом далеко не все ресурсы платные. Существует значительный пул общедоступных ресурсов, которые могут быть с успехом использованы для информационного обеспечения. К числу платных ресурсов относятся, прежде всего, базы бизнес-информации, зарубежные полнотекстовые и фактографические базы данных, а также крупные базы реферативной информации.

Однако эти информационные ресурсы разрознены, навигация по ним затруднена, и, как и 10 лет назад [41], отсутствует «единая точка входа», обеспечивающая получение комплексной информации для поддержки жизненного цикла инноваций. Есть основания полагать, что создание такой единой точки на федеральном уровне не представляется сегодня реальным, в силу чего более приемлем вариант построения сети региональных информационных систем, обеспечивающих навигацию и доступ к соответствующим ресурсам.

Таким образом, можно сделать вывод, что построение сети региональных информационных систем потребует некоторого пересмотра точек зрения на состав и наполнение информационной инфраструктуры как компонент общей инфраструктуры поддержки инноваций. И это неизбежно произойдет в процессе перехода на технологии цифровой экономики. При этом, с учетом наработок сети Федерального института промышленной собственности [42], прошлого опыта эксплуатации АСНТИ СО РАН, АС НАУКА [43] и др., а также решений по управлению информационными ресурсами и бизнес-информацией [44–47] задача создания региональных систем информационной поддержки инноваций не видится нереальной.

Одним из пилотных узлов такой сети вполне может быть Новосибирск, обладающий серьезным инновационным потенциалом и набором информационных ресурсов, достаточным для начала работ по созданию информационной среды поддержки инноваций.

ЗАКЛЮЧЕНИЕ

Современная инновационная деятельность протекает в условиях устойчивого развития и наращивания цифровых библиографических, фактографических и полнотекстовых коллекций, в создании которых принимают участие органы научно-технической информации, федеральные и отраслевые библиотеки, университеты, предприятия, ассоциации и консорциумы инновационной сферы, академические и отраслевые научные институты, научно-профессиональные общества и союзы, интеграторы бизнес-информации и др.

Констатируя устойчивую положительную тенденцию в развитии инновационной сферы, нельзя не принимать во внимание, что представители малого и

среднего инновационного бизнеса зачастую решают задачи создания и реализации своей продукции без учета уже имеющихся и опубликованных в открытой печати наработок, что приводит к нерациональной трате интеллектуальных усилий, а также ограниченных материальных и финансовых ресурсов. Кроме того, многочисленные факты свидетельствуют о неудовлетворительной информированности персонала инновационных предприятий в части возможного использования мировых информационных ресурсов как основы для принятия решений на различных этапах жизненного цикла инновационной продукции.

Представляется, что совокупная инновационная продуктивность региона зависит от качества информационной поддержки инноваций на каждом из этапов их жизненного цикла. В свою очередь, повышение качества информационной поддержки требует, как привлечения широкого спектра внешних информационных ресурсов, так и создания единой технологической среды навигации, хранения, поиска, обработки и предоставления информации на основе распределенных информационно-вычислительных сетей.

На основании изложенного можно сделать следующие выводы:

1) неоднозначность толкования понятия «инновация» порождает разнообразие смысловой нагрузки термина «инновационная инфраструктура (или инфраструктура инноваций)». Из этого следует отсутствие определенности в отношении состава и наполнения информационной инфраструктуры поддержки инновационной деятельности;

2) нормативно-правовые и программные документы регионов Сибирского федерального округа свидетельствуют о многообразии точек зрения на задачи информационной поддержки инноваций и методы их решения;

3) одной из важнейших причин неразвитости информационной инфраструктуры как компонента общей инфраструктуры поддержки инноваций можно назвать видение инновации исключительно как внедренного (или нуждающегося в этом) и уже объективно существующего новшества;

4) информационная поддержка инноваций на протяжении их жизненного цикла от момента генерации идей до вывода инновационного продукта с рынка предполагает активное использование широкого спектра внешних, по отношению к инновационной организации, документальных и фактографических информационных ресурсов;

5) разработка механизмов, обеспечивающих функционирование информационно-аналитической рабочей среды, включая навигацию и доступ к информационным ресурсам, является приоритетным направлением развития инфраструктуры поддержки инновационной деятельности;

6) в связи с объективными трудностями создания единой федеральной точки навигации и доступа к информационным ресурсам поддержки инновационной деятельности предлагается разработка распределенной сети информационного обеспечения инноваций, узлы которой образуют соответствующие региональные системы;

7) создание региональных информационных систем поддержки инноваций требует не только коррек-

тировки нормативно-правовых и программных документов соответствующих регионов, но и вызывает необходимость реинжиниринга существующих локальных систем, консолидации имеющихся информационных ресурсов, расширения информационной базы и вложения соответствующих материальных и финансовых ресурсов.

СПИСОК ЛИТЕРАТУРЫ

1. Лебедев С.А. Философия науки: словарь основных терминов. – М.: Акад. Проект, 2004; Киров : ОАО Дом печати – Вятка. – 316 с. – (Gaudeamus) (Учеб. пособие для вузов). – URL: http://philosophy_of_science.academic.ru, свободный (дата обращения 11.11.17).
2. Толковый словарь "Инновационная деятельность" : термины инновац. менеджмента и смеж. обл.: (от А до Я) / отв. ред. В.И. Суслов. 2-е изд., доп. – Новосибирск : Сиб. науч. изд-во, 2008. – 223 с.
3. Экономика знаний // Википедия. – URL: <https://ru.wikipedia.org/wiki>, свободный (дата обращения 11.05.17).
4. Федеральный закон от 23.08.1996 N 127-ФЗ (ред. от 23.05.2016) "О науке и государственной научно-технической политике" (с изм. и доп., вступ. в силу с 01.01.2017) // Собр. законодательства РФ. – 1996. – № 35. – Ст. 4137.
5. Словарь бизнес-терминов // АКАДЕМИК: Словари и энциклопедии на Академике. – URL: <http://dic.academic.ru/dic.nsf/business/3554> (дата обращения 11.05.17).
6. Словарь-справочник терминов нормативно-технической документации // АКАДЕМИК: Словари и энциклопедии на Академике. – URL: http://normative_reference_dictionary.academic.ru/ (дата обращения 11.05.17).
7. Большой юридический словарь / авт. и сост. А.Б. Борисов. – М.: Книжный мир, 2010. – 847 с.
8. Инновационная деятельность // Война и мир в терминах и определениях: военно-технический словарь / под общ. ред. Д.О. Рогозина; сост.: Д.О. Рогозин и др. – М.: Вече, 2016. – 270 с. – URL: http://war_peace_terms.academic.ru/367, свободный (дата обращения 11.11.17).
9. О Концепции инновационной политики Российской Федерации на 1998–2000 годы : постановление Правительства РФ от 24 июля 1998 г. № 832 // Собр. законодательства Рос. Федерации. – 1998. – № 32. – Ст. 3886.
10. Руководство Осло: рекомендации по сбору и анализу данных по инновациям : совместная публикация ОЭСР и Евростата: пер. на русс. яз. 3-е изд. – 2010. – URL: http://mgimo.ru/upload/docs_6/ruk.oslo.pdf, свободный (дата обращения 11.05.17).
11. Флёрова А.Н. Понятие инновации в законодательстве Российской Федерации // Рос. внешне-экономический вестник – 2006. – № 9. – С. 64-67.
12. О Стратегии инновационного развития РФ на период до 2020 г.: распоряжение Правительства Рос. Федерации от 08 дек. 2011 г. № 2227-р // Собр. законодательства РФ. – 2012. – № 1. – Ст. 216.
13. Агарков С.А., Кузнецова Е.С., Грязнова М.О. Инновационный менеджмент и государственная инновационная политика: учеб. пособие. – М.: Акад. естествознания, 2011. – 143 с.
14. Инновации и общество в России. Москва // ВЦИОМ, 2011. – URL: https://wciom.ru/fileadmin/file/reports_conferences/2011/2011-09-23-innovacii.pdf (дата обращения 29.05.17).
15. Инновации в российской экономике: от идей – к практике. Всероссийские опросы ВЦИОМ 2010–2016 гг. Санкт-Петербург // ВЦИОМ, 2016. – URL: https://wciom.ru/fileadmin/file/reports_conferences/2016/2016-06-17-inovacii.pdf (дата обращения 29.05.17).
16. Уткин Э.А., Морозова Г.И., Морозова Н.И. Инновационный менеджмент. – М. : АКАЛИС, 1996. – 207, [1] с. (с. 10)
17. Кулагин А.С. Немного о термине «инновация» // Инновации. – 2004. – № 2. – С. 48.
18. Энциклопедия социологии // Библиотека учебной и научной литературы. – URL: <http://sbiblio.com/biblio/content.aspx?dictid=97&wordid=812868> (дата обращения 11.11.17).
19. Некрасова Н.А., Некрасов С.И. Философия науки и техники: тематический словарь-справочник: учеб. пособие для студентов всех специальностей. – М.: МИИТ, 2009. – 424 с.
20. Санто Б. Инновация как средство экономического развития. – М.: Прогресс, 1990. – 295 с. (с. 24)
21. Морозов Ю.П., Гаврилов А.И., Городков А.Г. Инновационный менеджмент: учебное пособие для вузов. – М.: ЮНИТИ-ДАНА, 2003. – 471 с. (с. 17)
22. Твисс Б. Управление научно-техническими нововведениями. – М.: Экономика, 1989. – 456 с.
23. Мазуренко С. Инновации – это симбиоз государственной политики и рыночных отношений // Венчурная Россия. – URL: http://www.allventure.ru/articles/15/print_version.html (дата обращения 11.11.17).
24. Азгальдов Г.Г., Костин А.В. К вопросу о термине «инновация» // Библиотека оценщика «LABRATE.RU». – URL: http://www.labrate.ru/articles/azgaldov-kostin_doklad_2009-2_about-innovation.htm (дата обращения 26.05.17).
25. Стратегия инновационного развития Российской Федерации на период до 2020 года (проект) // Российская кластерная обсерватория. – URL: http://cluster.hse.ru/cluster-policy/low_base.php (дата обращения 23.05.17).
26. Межгосударственный ГОСТ (Межгосударственный стандарт). ГОСТ 31279-2004. Инновационная деятельность. Термины и определения = Innovative activities. Terms and definitions: межгосударственный стандарт. Изд. офиц. – Минск : Изд-во стандартов, 2005. – III. – 12 с.
27. Винокуров В.И. Основные термины и определения в сфере инноваций // Инновации. – 2005. – №4 (81). – С. 6-22.

28. Монастырный Е.А. Термины и определения в инновационной сфере // *Инновации*. – 2008. – № 2 (112). – С. 28-31.
29. Голиченко О.Г., Самоволева С.А. Риски реализации стратегии инновационного развития Российской Федерации на период до 2020 г. «Инновационная Россия-2020» // *Инновации* – 2012. – №4 (162). – С. 71-80.
30. Бахтурин Г.И., Турко Т.И., Плиева З.Р., Гудкова А.А., Одинцова Н.Н., Ольшевский Д.В. Мониторинг как инструмент оценки состояния национальной инновационной инфраструктуры // *Инноватика и экспертиза: научные труды*. – 2016. – № 3 (18). – С. 8-23. – URL: http://inno-exp.ru/archive/18/innov_2016-3_8-23.pdf (дата обращения 23.05.17).
31. Концепция научно-информационного обеспечения программ и проектов государств-участников СНГ в инновационной сфере. – Доступ из справ.-правовой системы «Гарант». – URL: <http://base.garant.ru/2568289/#friends> (дата обращения 23.05.17).
32. Стратегия развития информационного общества в Российской Федерации от 7 февраля 2008 г. № Пр-212 // *Российская газета* – 2008. – 16 февр. – URL: <https://rg.ru/2008/02/16/informacia-strategia-dok.html> (дата обращения 23.10.17).
33. Об утверждении государственной программы Российской Федерации "Информационное общество (2011 – 2020 годы) : постановление Правительства Рос. Федерации от 15 апреля 2014 г. № 313 (ред. от 26.10.2017) // *Собр. законодательства РФ*. – 2014. – № 18, ч. 2. - Ст. 2159.
34. Абдрахманова Г.И., Бахтин П.Д., Гохберг Л.М. и др. Рейтинг инновационного развития субъектов Российской Федерации. Вып. 5 / под ред. Л.М. Гохберга. – М.: НИУ ВШЭ, 2017. – 260 с. – URL: <https://issek.hse.ru/data/2017/06/09/1170533818/RIR2017.pdf>, (дата обращения 23.10.17).
35. Стратегия социально-экономического развития Алтайского края на период до 2025 года: одобрена постановлением Администрации Алтайского края от 28.12.2007 № 622 // *Центр поддержки предпринимательства: Алтайский край*. – URL: <http://www.altaicpp.ru/strat/strategija.html> (дата обращения 23.10.17).
36. Об утверждении Концепции социально-экономического развития Иркутской области на период до 2020 года: Распоряжение Губернатора Иркутской области от 04.06.2010 N 34-р // *ИРКУТ: Информационный портал Иркутской области*. – URL: <http://irkut.info/2010/06/koncersija36885.htm> (дата обращения 23.10.17).
37. Об утверждении Стратегии социально-экономического развития Кемеровской области до 2025 года: принят Советом народных депутатов Кемеровской области 9 июля 2008 года № 74-ОЗ (ред. от 02.11.2012) // Консорциум «Кодекс». Электронный фонд правовой и нормативно-технической документации. – URL: <http://docs.cntd.ru/document/990308346> (дата обращения 23.10.17).
38. Горбунова Ю.И., Горбунова О.Н. Информационная инфраструктура: современная сущность, подотрасли ее составляющие // *Социально-экономические явления и процессы*. – 2014. – № 2 (60). – С. 14-21.
39. Большеева С.А., Лопалева Н.С. Научные журналы в информационно-коммуникационной инфраструктуре реформы высшего образования в РФ // *Известия Уральского федерального университета. Сер. 1: Проблемы образования, науки и культуры*. – 2017. – Т. 162, № 23 (2). – С. 20-29.
40. *Философия науки и техники: тематический словарь*. – URL: http://science_philosophy.academic.ru/ (дата обращения 25.10.17).
41. Арский Ю.М., Цветкова В.А., Яшукова С.П. Информационные ресурсы для поддержки инноваций // *Информационные ресурсы России*. – 2007. – №6. – С. 8.
42. Основная деятельность // *ФИПС – Федеральный институт промышленной собственности*. – URL: http://www1.fips.ru/wps/wcm/connect/content_ru/about/osn_deya/ (дата обращения 23.10.17).
43. Елепов Б.С., Баженов С.Р., Бобров Л.К., Каленов Н.Е. Проектирование и эксплуатация региональных АСНТИ. – Новосибирск, Наука: Сибирское отделение, 1991. – 173 с.
44. Бобров Л.К., Крылов А.С., Медянкина И.П., Осипов А.Л., Пестунов А.И., Терещенко С.Н. Управление бизнес-информацией: технологии обработки и анализа. – Новосибирск: НГУЭУ, 2015. – 283 с.
45. Трусев А.В. Принципы разработки и организации функционирования информационных систем и процессов поддержки инновационного развития региона: автореф. дис. ... д-ра техн. наук. – М., 2017. – 38 с.
46. Бобров Л.К., Медянкина И.П., Осипов А.Л., Пашков П.М., Родионова З.В. О компетенциях менеджера бизнес-информации // *Научно-техническая информация. Сер.1*. – 2016. – № 5. – С. 5-14.
47. Бобров Л.К., Гиляревский Р.С., Родионов И.И., Цветкова В.А., Шрайберг Я.Л. Информационный менеджмент. – Новосибирск: Новосиб. гос. ун-т экономики и упр., 2009. – 314 с.

Материал поступил в редакцию 27.11.17.

Сведения об авторах

БОБРОВ Леонид Куприянович – доктор технических наук, профессор кафедры бизнес-информатики Новосибирского государственного университета экономики и управления «НИНХ», e-mail: l.k.bobrov@nsuem.ru

МЕДЯНКИНА Ирина Петровна – кандидат технических наук, доцент кафедры бизнес-информатики Новосибирского государственного университета экономики и управления «НИНХ» e-mail: i.p.medyankina@edu.nsuem.ru

К оценке связи библиометрических и альтметрических показателей: на материале российских публикаций SCI-E (2015 г.)*

На массиве данных о российских публикациях за 2015 г., зарегистрированных в БД SCI-E, осуществлена оценка характера корреляции между традиционными библиометрическими и альтметрическими показателями. С этой целью разработана специальная методика, основывающаяся на использовании коэффициента ранговой корреляции Кенделла. В случае публикаций оценка выполнена для корреляции между цитированием публикации (значение поля TC в БД SCI-E) и ее использованием (поля U1 и U2). В случае журналов-источников указанных публикаций корреляция определялась между значением Cited Half-life и значением показателей использования журнала, которые определялись как среднее значение U1 (U2), взятое по всем публикациям журнала. Установлено, что характер распределения журналов по величине Cited Half-life журналов-источников российских публикаций практически совпадает с характером распределения журналов-источников публикаций всего мирового потока: российское распределение несколько смещено в сторону больших значений Cited Half-life (медиана мирового распределения – 6,1, российского – 6,2). Это смещение резко возрастает при учете числа опубликованных в журнале российских и мировых публикаций соответственно (медиана мирового распределения – 5,8, российского – 6,8).

Ключевые слова: библиометрия, альтметрика, цитируемость, показатели использования, ранговая корреляция, корреляционная зависимость, российские публикации, Web of Science

ВВЕДЕНИЕ

Растущая роль социальных сетей привела к появлению новых показателей исследовательской деятельности – альтернативной метрике (альтметрике). Альтметрика стала популярной темой в течение последних пяти лет как показатель оценки социальных сетей. Поиск, выполненный нами в БД Web of Science Core Collection (17.07.2017), показал устойчивый и резкий рост числа публикаций по этой тематике: от двух статей в 2012 г. до 96 статей в 2016 г. Анализ 253 статей по альтметрике (WoS и Scopus за период 2005-2015 гг.), выполненный в работе [1], позволил выявить группу наиболее продуктивных авторов, организаций и коллабораций между ними. Причем, даже в этой «высокопродуктивной группе» представители США и европейских стран выделялись особо высокой продуктивностью.

Анализ литературных источников показал следующее. В работе [2], выполненной в Великобритании, были проанализированы около 333 тыс. публикаций по 45 разделам медицины из БД Scopus. Во всех указанных разделах медицины существует тесная корреляция между показателями цитируемости и количеством пользователей БД «Mendeley» в 2009 г. При этом оказалось, что степень корреляции несколько уменьшилась, когда были исключены из рассмотрения данные, соответствующие пользователям – студентам. По мнению авторов работы [2], индикаторы альтметрики могут быть использованы в качестве дополнительного показателя оценки публикации.

В работе [3] приведены результаты социологического опроса участников международной конференции «Показатели науки и техники 2012» (S&TI-2012). Опрос был выполнен под руководством проф. Хаустейн С. (Haustein S.) Согласно этим результатам социальные сети оказали заметное влияние на профессиональную деятельность почти половины респондентов. При этом отношение респондентов к использованию различных социальных платформ оказалось различным. Так социальной сетью LinkedIn пользовались 68% респондентов, тогда как платформы Mendeley, Academia.edu и

* Работа выполнена частично в рамках темы 0003-2014-0007 и частично при поддержке Российского фонда фундаментальных исследований (проекты РФФИ: 17-02-00078 и 17-02-00157)

ResearchGate использовали только порядка 20%. Авторы работы [3], отмечают также что по мнению подавляющего числа респондентов (72%) количество выгрузок статьи является наиболее ценным альтметрическим показателем.

Детальный анализ различных показателей альтметрики и их связи с метриками цитирования на материале массива, содержащего свыше 750 тыс. публикаций, приводится в работе [4] (эти показатели были предоставлены платформой Altmetrics.com). Результаты анализа показали, что использование публикации (количество ее выгрузок и т.п.) охватывает относительно небольшое число исследовательских статей: 15-24% от рассмотренного массива. Авторы работы [4] пришли к выводу, что показатели альтметрики наиболее целесообразно использовать при анализе следующих областей науки: медицина, науки о живой природе, социальные и гуманитарные науки.

Серьезный анализ «ажитожа», возникшего в связи с показателями альтметрики, опубликован в работе [5], где детально обсуждаются данные, предоставляемые платформой Altmetrics.com. Авторы этой работы указывают на недостатки так называемого «взвешенного показателя альтметрики», предоставляемого этой платформой. При этом авторы предупреждают сообщество ученых о серьезной опасности злоупотребления данными альтметрики.

С 2014 г. информационная платформа Web of Science Core Collection (WoS CC), наряду с показателями цитирования статей, предоставляет также альтернативные показатели – показатели использования статьи, которые в WoS включают два типа метрики: «Количество использований статьи за 180 дней» (Usage Counts Last 180 Days – U1) и «Количество использований статьи с 2013 года» (Usage Counts Since 2013 с 01.02.2013 – U2). В WoS в качестве использования публикации признается каждое из следующих действий пользователя: выгрузка (импорт) полного текста публикации; сохранение библиографического описания публикации в библиографическом справочном менеджере (например, EndNote) или в каком-либо другом выгружаемом формате. WoS фиксирует каждое такое действие, прибавляя единицу («1») к текущему значению показателей использования U1 и U2¹. Метрика использования статьи доступна для всех баз данных, размещенных на информационной платформе WoS, за исключением патентной БД Derwent Innovation Index. В 2016 и 2017 гг. были опубликованы работы [6, 7], посвященные использованию показателей альтметрики по данным WoS. В статье [6] в качестве источника анализа были выбраны пять журналов по предметной категории WoS «Информаци-

тика и библиотечные науки» («Information Science and Library Science») за период с 1945 по 1999 гг. и с 2013 по 2015 гг. Одним из результатов этого анализа является выявление факта предпочтения использования публикаций, опубликованных недавно. Однако высокоцитируемые статьи, опубликованные ранее, достаточно длительное время продолжают также активно использоваться.

Представительное эмпирическое исследование на материале публикаций из Бельгии, Израиля, Ирана и мира в целом (WoS CC, 2013 г.) было опубликовано в середине 2017 г. в журнале *Scientometrics* [7]. Это исследование посвящено изучению взаимосвязей между индикаторами использования статей (U2), показателями международного научного сотрудничества и цитируемостью. Выгрузка этих публикаций произведена в 2016 г., следовательно, окно цитирования и окно использования составляли три года. Для анализа распределения публикаций по цитируемости статей и их использования, авторы работы [7] применили ранее разработанный ими метод исследования моделей цитируемости, который они назвали «Characteristic Scale Scores». Из этого анализа следует, что между показателями цитируемости и количеством использований (U2) имеется значительная корреляция, особенно в области общественных наук. Однако авторы не обнаружили связи между цитируемостью публикаций с большим количеством авторов и количеством использований этих публикаций. По мнению авторов работы [7] стабильность распределений, полученных в процессе использования этого метода, свидетельствует о возможности его применения для характеристики показателей цитируемости статей и их длительного использования (U2). Наблюдалось явное различие в моделях цитирования и использования статей, но в некоторых областях знания проявилось неожиданное сходство.

Следующее исследование этих же авторов, принятое к печати [11], было предоставлено одному из авторов настоящей статьи на 13-й «Международной конференции по вебометрике, информетрии и наукометрии» и 18-м совещании COLLNET, состоявшихся 09.07–11.07 2017 г. в Университете Кента (Англия). Разработанные в предшествующих работах методы авторы работы [10] применили, используя мировой массив публикаций за 2013 г. Кроме трех основных показателей, включающих цитируемость, количество использований статьи (U2) и доли международного сотрудничества, в работе [10] вводится показатель «индекс использования журнала» (Journal Usage Index). Этот же инструментарий в работе [10] применен при анализе массивов публикаций из 15 стран по пяти крупным областям исследований и 12 подобластям (по классификации Leuven-Budapest). По их мнению, Journal Usage Index может быть применен как дополнительный показатель к цитируемости журнала. В указанной работе отмечается также различие между общественными науками и науками о жизни в характере связей между цитируемостью журнала и его использованием.

В настоящей работе поставлена задача определить характер связи, существующей между традиционными библиометрическими показателями – с одной стороны, и показателями альтметрики – с другой.

¹ Следует сказать, что еще в 2003 г. в ИС «Указатель РФФИ» [8, 9] был разработан и внедрен комплекс показателей, смысл и способ формирования которых во многом были подобны нынешним показателям использования в WoS. Увеличение показателя использования (в «Указателе РФФИ» этот показатель назывался «on-line цитированием») публикации происходило в случае, когда результат поиска по некоторому запросу пользователя системы, содержал эту публикацию. При этом фиксировался еще ряд параметров: запрос (набор ключевых слов), дата запроса, имя пользователя. Причем, все эти данные оказывались «привязанными» к публикации, «прцитированной on-line».

В качестве представителей альтметрики рассматриваются показатели использования, введенные в WoS с 2013 г. и рассмотренные выше: U1 и U2, в качестве представителей традиционных библиометрических показателей – значение TC (Times Cited – количество цитирований публикации) и Cited Half-life – CdHL (полупериод жизни статей из данного журнала). «Физический смысл» этого показателя состоит в том, что он обозначает медианный возраст (в годах) тех публикаций из данного журнала, которые были процитированы в заданном году (в год опубликования очередного выпуска JCR).

ИСХОДНЫЕ ДАННЫЕ АНАЛИЗА

Исходные данные – массив публикаций российских авторов, вышедших в 2015 г. и зарегистрированных в Science Citation Index Expanded (SCI-E). Общее число публикаций в этом массиве – 37279 документов. Были также сформированы «производные»

массивы, т.е. подмассивы, выделенные из исходного по следующим классификационным основаниям: по принадлежности публикаций заданной категории WoS (набору категорий WoS); по типу рассматриваемых элементов массива (публикация, журнал).

Из WoS массив был импортирован 15 марта 2017 г. Таким образом, значения TC, U1 и U2 были однозначно зафиксированы во времени, что очень важно в случае TC и U2. В случае U1 это не столь важно, так как значения U1 всегда соответствуют неизменному по величине временному отрезку (180 дней). Список массивов публикаций и соответствующих им массивов журналов представлен в табл.1, а наборы категорий WoS (категории WC), с помощью которых были сформированы эти массивы, – в ПРИЛОЖЕНИИ. Причем, в состав этого набора категорий могут входить категории из других наборов. Следовательно, одна и та же публикация может быть представлена в нескольких массивах.

Таблица 1

Общие характеристики исследуемых массивов публикаций (российские публикации за 2015 г., SCI-E)

Массивы публикаций	Число публикаций	Число наименований журналов ²	Доля публикаций с TC>0 (%)	Доля публикаций с U1>0 (%)	Доля публикаций с U2>0 (%)
1	2	3	4	5	6
Все публикации (исходный массив)	37279	3606	53,0	62,5	85,2
Публикации из зарубежных журналов	19645	3215 ³	63,9	64,7	86,8
Публикации, соответствующие хотя бы одной из 8-и категорий WoS по биологии (включающих слово «Biology»)	3059	503	53,4	66,9	90,3
Публикации, соответствующие категории WoS «Биохимия и молекулярная биология»	1448	178	52,8	62,0	88,7
Публикации, соответствующие хотя бы одной из 17-и категорий WoS по технологиям и инженерному делу (включающих слово «Engineering»)	3560	397	48,1	70,1	92,3
Публикации, соответствующие хотя бы одной из 3-х категорий WoS по математике (включающих слово «Mathematics»)	2264	154	42,0	29,5	63,1
Массив российских публикаций, соответствующих хотя бы одной из 9-и категорий WoS по медицине (включающих слово «Medicine»)	1084	123	38,0	52,3	76,8
Массив российских публикаций, соответствующих хотя бы одной из 2-х категорий WoS по микробиологии (включающих слово «Microbiology»)	696	132	57,0	69,3	93,7
Массив российских публикаций, соответствующих категории WoS «Физика конденсированного состояния»	1772	55	65,8	79,0	97,9

² В значения графы 3 помимо собственно журналов и сериальных изданий включены также и другие источники публикаций (исключение составляет значение в строке 2: см. следующую сноску). Следовательно, попытки получить число российских журналов путем вычитания из значения в строке 1 значения в строке 2 будут некорректны (действительное число российских журналов в анализируемом массиве составляет 146 наименований).

³ Здесь указано именно число журналов и сериальных изданий, т.е. тех источников, которые содержатся в JCR: данные о стране-издательстве источников других видов установить не удалось.

Если в качестве основных элементов выступают публикации (корреляция на уровне публикаций), мы будем использовать непосредственные значения ТС, U1 и U2. В тех случаях, когда в качестве основных элементов выступают журналы (корреляция на уровне журналов), значения U1 и U2 журнала будут представлены средним значением соответствующего показателя для данного журнала. Таким образом, в качестве первой характеристики журнала будут применены значения использования U1 и U2. В качестве второй характеристики журнала – показатель Cited Half-life (CdHL).

МЕТОДИКА ИССЛЕДОВАНИЯ

Определение зависимостей между альтметрическими и традиционными библиометрическими показателями будет выполнено на двух уровнях – публикаций и журналов. Если взяты публикации, мы использовали коэффициент ранговой корреляции Кенделла τ , а если журналы – коэффициент ранговой корреляции τ и графические представления соответствующих зависимостей.

Уровень публикаций: U1 (U2) ↔ ТС

Идея определения ранговой корреляции опирается на следующее положение: чем в большей степени порядок следования элементов данного множества,

упорядоченных (проранжированных) по некоторому одному признаку (показателю), совпадает с порядком следования этих элементов по другому признаку, чем больше таких совпадений, тем выше положительная корреляция между этими признаками.

Упорядочим (проранжируем) все 37279 публикаций по убыванию (невозрастанию) значения ТС. Затем аналогичным образом проранжируем эти же публикации по значениям U1 (U2).⁴ Отсортируем публикации по возрастанию рангов по первому признаку (ТС). В итоге получим таблицу, фрагмент которой приведен в табл. 2 (буквенные обозначения реальных публикаций из рассматриваемого массива содержатся в графе 1). Поскольку в табл. 2 публикации уже расположены по возрастанию рангов ТС (графа 4), то нет необходимости анализировать порядок следования по этому показателю. Следовательно, необходимо анализировать порядок следования публикаций только по U1, вернее, по рангам, соответствующим значениям U1 (графа 5).

Рассмотрим, пары публикаций в табл. 2 (графа 1). Будем двигаться сверху вниз табл. 2. Т.е. будем рассматривать пары вида: «a-b», «a-c», «a-d», ..., «a-j»; «b-c», «b-d», ..., «b-j»; ...; «h-i», «h-j»; «i-j». При этом пары в «обратном» порядке не рассматриваются, т.е. не рассматриваются пары вида: «b-a», «c-b» и т.п.

Таблица 2

Исходные данные для определения корреляции между ТС и U1*

Публикация	Значение ТС	Значение U1	Ранг публикации по ТС	Ранг публикации по U1	Число случаев, совпадения порядка следования для пар, с участием данной публикации	Число случаев, несовпадения порядка следования для пар, с участием данной публикации
1	2	3	4	5	6	7
a	812	216	1	1	9	0
b	726	42	2	28	7	1
c	535	26	3	62	6	1
d	462	106	4	5	6	0
e	382	7	5	1741	0	5
f	286	9	6	927	0	4
g	282	18	7	140	2	1
h	245	11	8	597	0	2
i	241	15	9	263	0	1
j	211	31	10	47	0	0
Итого					$P=30$	$Q=15$

* фрагмент реальной таблицы – приведены первые 10 (по величине ТС) публикаций (обозначены латинскими буквами)

⁴ В тех случаях, когда для нескольких публикаций значения данного показателя совпадают, публикациям присваивается один и тот же ранг, который вычисляется как среднее значение мест (рангов), занимаемых этими публикациями.

Публикацию, которая в заданной паре в табл. 2 по порядку следования стоит на первом месте, назовем первой публикацией в паре (ППВП), а публикацию, которая находится ниже первой – назовем второй публикацией в паре (ВПВП). Подсчитаем, количество тех пар из числа образованных соответствующей ППВП, в которых численное значение ранга ВПВП по U1 больше, чем численное значение ранга ППВП по этому же признаку. Это тот случай, когда порядки следования по ТС и по U1 совпадают. Затем для данной ППВП подсчитаем количество пар публикаций, у которых эти порядки не совпадают, т.е. когда ранг (его численное значение) ВПВП по U1 меньше, чем численное значение ранга ВПВП по этому же признаку. Очевидно, что для всех девяти пар, куда входит публикация «а» (пары «а-б», «а-с», «а-д», ..., «а-ж») порядок следования по обоим признакам совпадает: нет ни одной публикации, у которой ранг (его численное значение) по показателю U1, меньше, либо равен рангу по этому показателю публикации «а». Для пар, куда входит публикация «б», такое совпадение нарушено в случае пары «б-д»: ранг по U1 публикации «б» больше такого ранга публикации «д» (28>5), тогда как по рангу ТС порядок этих рангов противоположный (2<4), т.е. порядок следования по обоим признакам для пар с участием публикации «б» совпадает в 7 случаях, и в одном случае – не совпадает. Для пар с участием публикации «с» случаев совпадения 6, а случаев несовпадения – 1, и т.д.

Запишем полученное число совпадений для каждой заданной («первой» – ППВП) публикации пары в графу 6, а несовпадений – в графу 7. Сумму значений в графе 6 (число совпадений) обозначим как P , сумму значений в графе 7 – как Q .

Расчет коэффициента τ для данных из табл. 2 выполним по формуле (1).

$$\tau = \frac{P - Q}{\frac{1}{2}N(N - 1)}, \quad (1)$$

где: P – число случаев совпадения порядка следования по обоим признакам, взятое по всем парам публикаций (в табл. 2 $P=30$ – последняя строка, графа 6); Q – число случаев несовпадения порядка следования по обоим признакам (в табл. 2 $Q=15$ – последняя строка, графа 7); N – общее количество единиц (элементов) данного массива (публикаций, журналов; в табл. 2, N – число публикаций, $N=10$).

Для данных табл. 2 формула (1) будет выглядеть следующим образом:

$$\tau = \frac{30 - 15}{\frac{1}{2}10(10 - 1)} = 0,333. \quad (2)$$

Рассмотрим две взаимоисключающие гипотезы. *Первая*, так называемая *нулевая гипотеза* (H_0), согласно которой *корреляция* между рассматриваемыми признаками *совершенно отсутствует*. *Вторая*, напротив, утверждает, что такая *корреляция существует*. Задача состоит в том, чтобы доказать, что нулевая гипотеза неверна и ее следует отвергнуть. При

этом существует некоторая вероятность того, что отвергая нулевую гипотезу, мы совершаем ошибку. Эту вероятность обозначим символом α . Задавая, например, $\alpha=0,05$, мы тем самым требуем, чтобы вероятность ошибочного отказа от нулевой гипотезы была не больше 0,05. Или, другими словами, задавая $\alpha=0,05$, мы требуем, чтобы справедливость гипотезы о существовании корреляции была достаточно высокой (не менее 0,95). Чем меньше значение α , тем выше наше требование к вероятности того, что гипотеза о наличии корреляции верна: при $\alpha=0,001$ эта вероятность очень высокая – 0,999.

Обозначим через T_{kr} некое минимально-допустимое (критическое) значение τ , при котором еще можно говорить, что гипотеза о наличии корреляции верна. Расчет значения T_{kr} производится по формуле (3). Если величина τ при заданном значении α окажется меньше чем величина T_{kr} , ($\tau < T_{kr}$), то нулевую гипотезу следует признать справедливой. Если же τ при заданном значении α окажется больше чем величина T_{kr} , ($\tau > T_{kr}$), то нулевую гипотезу следует отвергнуть, тем самым признав справедливость гипотезы, утверждающей, что корреляция существует. Обычно этим оценкой наличия/отсутствия корреляции ограничивается. Нам представляется, что такая оценка в некоторых случаях может оказаться недостаточной, в частности, тогда, когда разность между τ и T_{kr} невелика и измеряется величиной порядка 0,001-0,010.

Чтобы преодолеть эту недостаточность, введем коэффициент $k = \tau / T_{kr}$.

Очевидно, что чем больше значение k , тем более тесная корреляция существует между рассматриваемыми признаками (как минимум, k должно быть равно или больше 1).

$$T_{kr} = z_{kr} \sqrt{\frac{2(2N + 5)}{9N(N - 1)}}, \quad (3)$$

где: N – как и в формуле (1) – число элементов в массиве (в табл. 2 число публикаций $N=10$); z_{kr} – при заданном значении α определяется по таблице функции Лапласа (при $\alpha=0,05$ $z_{kr}=1,96$).

Выполнив вычисления по формуле (3), получим $T_{kr} = 0,067$ и $k = 0,333 / 0,067 = 4,97$.

Ужесточим требование к вероятности того, что гипотеза о существовании корреляции была справедлива. Для этого зададим $\alpha=0,001$, тогда в нашем случае, согласно таблице функции Лапласа, $z_{kr}=3,3$ и $T_{kr}=0,108$ и $k=3,08$.

Уровень журналов: U1 (U2) ↔ CdHL

При определении зависимостей между традиционными библиометрическими показателями и показателями альтметрики на уровне журналов применялся как коэффициент ранговой корреляции Кенделла, так и графическое представление этих зависимостей.

Распределение журналов и соответствующих им публикаций по величине CdHL. Прежде чем перейти к определению зависимостей между использованием (U_1 и U_2) и полупериодом цитируемых публикаций (CdHL), целесообразно проанализировать распределения журналов по величине этого полупериода, а также рассмотреть соответствующие распределения статей, опубликованных в журналах с данным CdHL. Для этого сформируем группы журналов по величине CdHL. Затем для получения указанных выше распределений расположим журналы и соответствующие им публикации по этим группам. При этом мы попытались сопоставить распределения, включающие только журналы-источники российских публикаций и сами эти публикации (российские распределения), с аналогичными распределениями, охватывающими все мировые журналы и их публикации (мировые распределения). Для российских распределений данные были получены непосредственно из анализируемого массива публикаций за 2015 г. и с использованием данных JCR о величине CdHL журналов, в которых эти публикации вышли. Для получения мировых распределений мы использовали только JCR-SE: данные о величине CdHL журнала и число «Citable Items», т.е. число тех публикаций из данного журнала, которые могут быть процитированы. Как правило, это статьи и обзоры, которые в сумме составляют 75-90% всех публикаций журнала. Эти

данные необходимо представить на соответствующих графиках. Причем и для журналов и для соответствующих им публикаций на оси абсцисс будем откладывать значения CdHL, а на оси ординат – значение доли (от общего числа рассматриваемых журналов), которую занимает та или иная группа журналов (рис. 1). В случае публикаций на оси ординат расположим доли публикаций, принадлежащих каждой группе журналов – от общего числа публикаций в соответствующем массиве (рис. 2).

Следует сказать, что из 8591 журнала, представленного в JCR-SE за 2015 г., у 1683 журналов (19,6%) значения Cited Half-life превышают 10 лет. Причем, в JCR указывается только то, что этот показатель превышает 10 лет, но не уточняется сколько именно лет он составляет. Такие журналы были исключены из рассмотрения, так как информация о том, что $CdHL > 10$ фактически не является численной и ее не удастся использовать ни при расчете коэффициента корреляции, ни при графическом представлении соответствующих зависимостей. Кроме того, по 299 журналам (3,5%) в JCR данные отсутствуют (в поле «Citable Items» JCR-SE «указано «Not Available»). Таким образом, общее число мировых журналов, которое может быть включено в распределение по величине Cited Half-life, составляет 6609 наименований (76,9% от 8591), в которых опубликовано более миллиона статей и обзоров (1102824 – 84% от 1312319).

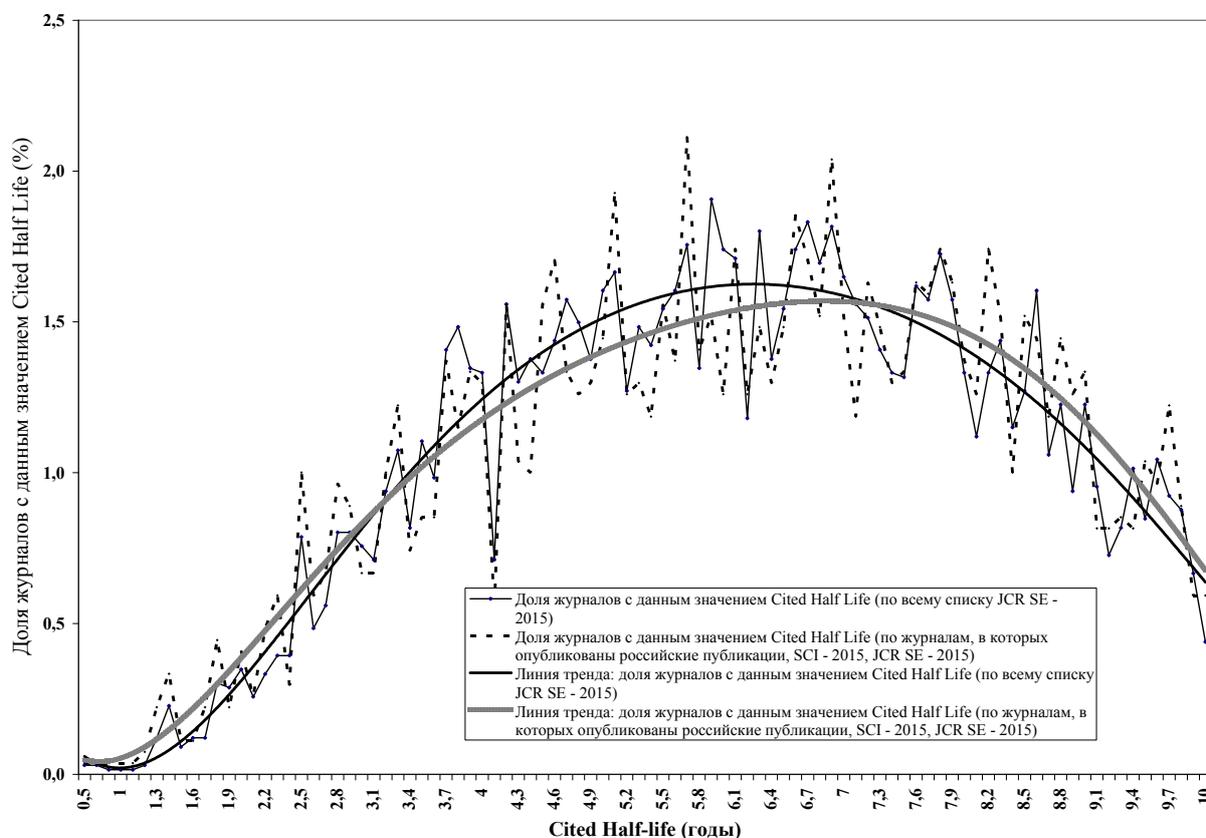


Рис. 1. Распределение журналов по величине Cited Half-life.

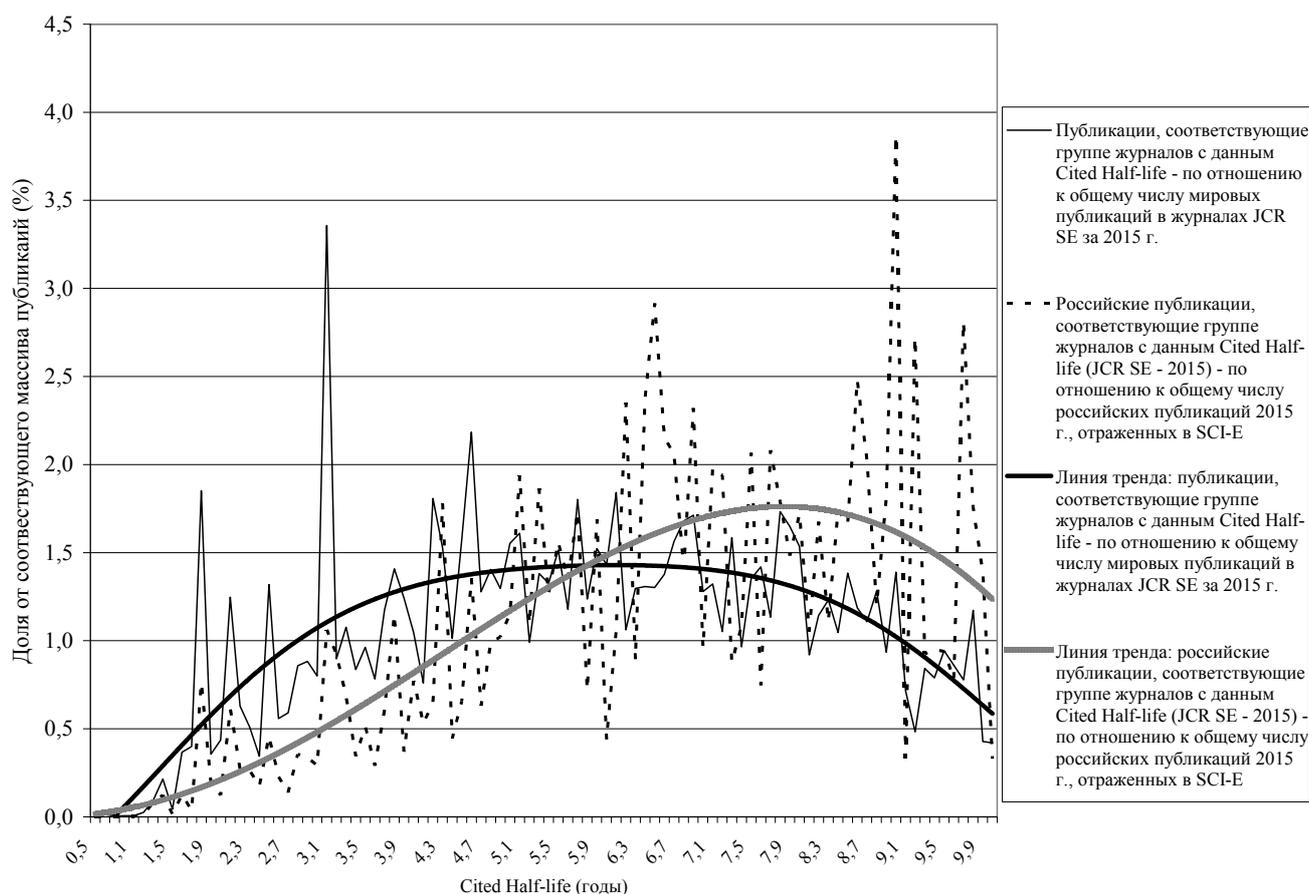


Рис. 2. Распределение публикаций по величине Cited Half-life журналов, в которых они были опубликованы

Если перейти к журналам, в которых опубликованы работы российских авторов, то из 3606 источников придется исключить 682 журнала, у которых $CdHL > 10$. Также необходимо исключить еще 45 журналов, у которых данные о значении $CdHL$ отсутствуют. Кроме того, из анализируемого массива пришлось исключить те источники (181 наименование), которые хотя и присутствовали в исходном массиве, однако, отсутствовали в JCR-SE. В результате общее число журналов, которые мы будем рассматривать при анализе российского массива, составляет 2698 наименований: 74,8% от общего числа журнальных и не журнальных источников российских публикаций и 78,8% от всех журналов – источников российских публикаций, представленных в JCR-SE. В этих журналах была опубликована 35801 статья российских авторов, что составляет 96% от исходного массива российских публикаций за 2015 г.

Вычисление коэффициентов корреляции на уровне журналов. Общая схема расчета коэффициента корреляции в случае журналов аналогична схеме расчета этого коэффициента для случая публикаций, описанной выше. При этом, в качестве библиометрического показателя был выбран полупериод цитируемых данным журналом публикаций ($CdHL$), а в качестве показателей альтметрики – показатели использования ($U1$ и $U2$). Значения $CdHL$ были получе-

ны непосредственно из JCR-SE за 2015 г., а показатели использования данного журнала рассчитывались как среднее число использований ($U1$, $U2$) в расчете на одну публикацию из этого журнала.

Определение корреляционных зависимостей. Характер исходных данных, которые мы предполагаем использовать для определения на уровне журналов взаимосвязи между альтметрическими и библиометрическими показателями, дает основания полагать, что применение коэффициента ранговой корреляции Кенделла для этих целей может оказаться недостаточным. Действительно, результат вычисления коэффициента корреляции – это всегда некоторое число, которое на графике будет выглядеть как точка. В случае определения характера зависимостей на уровне публикаций такая точечная характеристика вполне достаточна и оправдана: публикация рассматривается здесь как некоторая статистическая единица, принципиально не отличающаяся по своим свойствам от других публикаций – таких же статистических единиц. С переходом на уровень журналов, мы имеем дело с такими информационными объектами, рассмотрение которых в качестве безликих статистических единиц может существенно исказить картину: журналы обладают значительно более богатым набором атрибутов, чем публикации, и характеризуются значительно большей «индивидуальностью», чем публикации.

Итак, для определения характера зависимости между альтметрическими – с одной стороны, и традиционными библиометрическими показателями – с другой, мы сопоставим значения использования (U1 и U2), и значения CdHL, соответственно. При этом на оси абсцисс будем откладывать значения CdHL, характеризующие соответствующие группы журналов, а на оси ординат средневзвешенные значения U1 (U2) этих групп. Совершенно очевидно, что эти зависимости не могут быть жесткими, не являются функциональными зависимостями. Однако столь же очевидно, что характер связи между рассматриваемыми показателями необходимо представить в виде корреляционных зависимостей.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Уровень публикаций

Как следует из табл. 3, на уровне публикаций вычисление коэффициента ранговой корреляции τ было выполнено для двух пар показателей: U1 и TC; U2 и TC для каждого из девяти рассматриваемых массивов публикаций. Следовательно, общее число выполненных оценок наличия/отсутствия корреляции на уровне публикаций составляет – 18. Анализ полученных данных позволяет утверждать следующее.

Существует довольно тесная корреляция между цитируемостью и показателями использования для

всех без исключения массивов публикаций. Действительно, минимальное значение коэффициента k в случае показателя U1 составляет 4,1, а в случае показателя U2 – 5,9 (оба эти значения характеризуют массив публикаций по микробиологии – 8-я строка табл. 3), а максимальное – 30,9 и 50,2 соответственно (массив из всех публикаций – 1-я строка указанной таблицы). Отметим, что исключение публикаций из российских журналов не ведет к увеличению тесноты корреляции. Напротив, значение k оказывается даже ниже, чем для полного массива публикаций. Полученные данные обнаруживают также следующую тенденцию: с увеличением числа элементов (публикаций) увеличивается, как правило, и значение коэффициента k . Еще одно важное наблюдение: более длительному периоду, который охватывает показатель использования, соответствует более тесная корреляция между этим показателем и традиционными библиометрическими показателями, достаточно сравнить значения k в графах 4 и 7 в табл. 3.

Уровень журналов

Распределение журналов и соответствующих им публикаций по CdHL

Как следует из рис. 1, характер распределения по величине CdHL журналов, в которых опубликованы работы российских авторов, по своей форме мало чем отличается от аналогичного распределения всех мировых журналов, представленных в JCR-SE.

Таблица 3

Уровень публикаций: корреляция между показателями использованием и цитируемостью

Массивы публикаций	Корреляция между U1 и TC			Корреляция между U2 и TC		
	τ	T_{kr}	k	τ	T_{kr}	k
1	2	3	4	5	6	7
Все публикации (исходный массив)	0,354	0,011	30,9	0,552	0,011	50,2
Публикации из зарубежных журналов	0,345	0,016	21,6	0,516	0,016	32,2
Публикации, соответствующие одной из 8-и категорий WoS по биологии	0,345	0,040	8,7	0,549	0,040	13,8
Публикации, соответствующие категории WoS «Биохимия и молекулярная биология»	0,386	0,058	6,7	0,607	0,058	10,5
Публикации, соответствующие одной из 17-и категорий WoS по технологиям и инженерному делу	0,371	0,037	10,0	0,599	0,037	16,3
Публикации, соответствующие одной из 3-х категорий WoS по математике	0,200	0,045	4,4	0,421	0,046	9,1
Массив российских публикаций, соответствующих одной из 9-и категорий WoS по медицине	0,422	0,067	6,3	0,645	0,067	9,6
Массив российских публикаций, соответствующих одной из 2-х категорий WoS по микробиологии	0,339	0,084	4,1	0,490	0,084	5,9
Массив российских публикаций, соответствующих категории WoS «Физика конденсированного состояния»	0,359	0,052	6,9	0,513	0,052	9,8

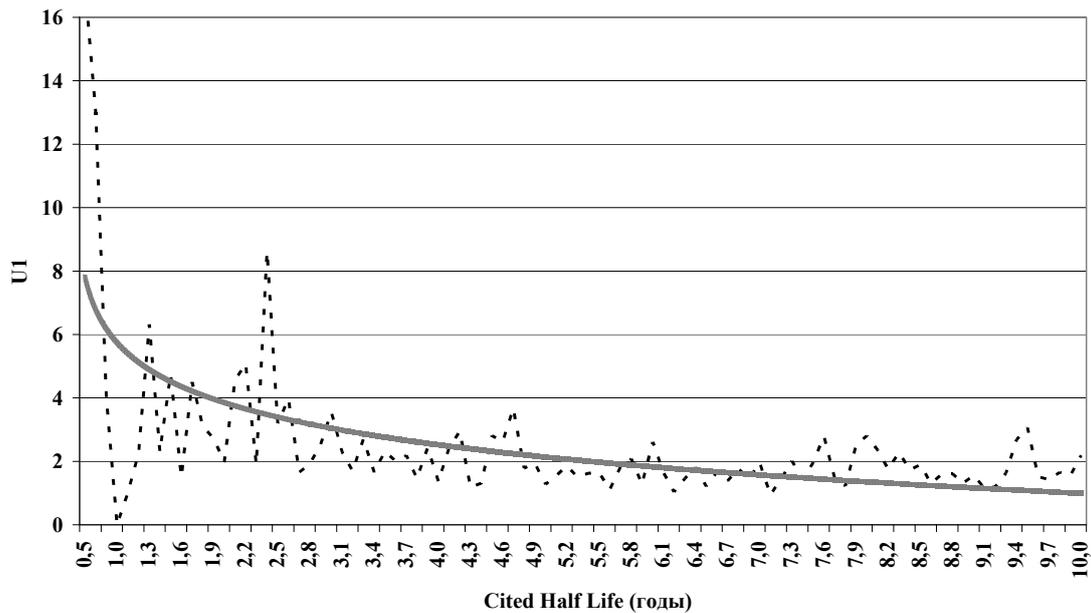


Рис. 3. Зависимость средних значений U1 (в расчете на 1 статью) от значений Cited_Half-life в журналах JCR-SE – по массиву российских публикаций 2015 г., представленных в SCI-E: пунктирная кривая – значения U1, сплошная кривая – линия тренда

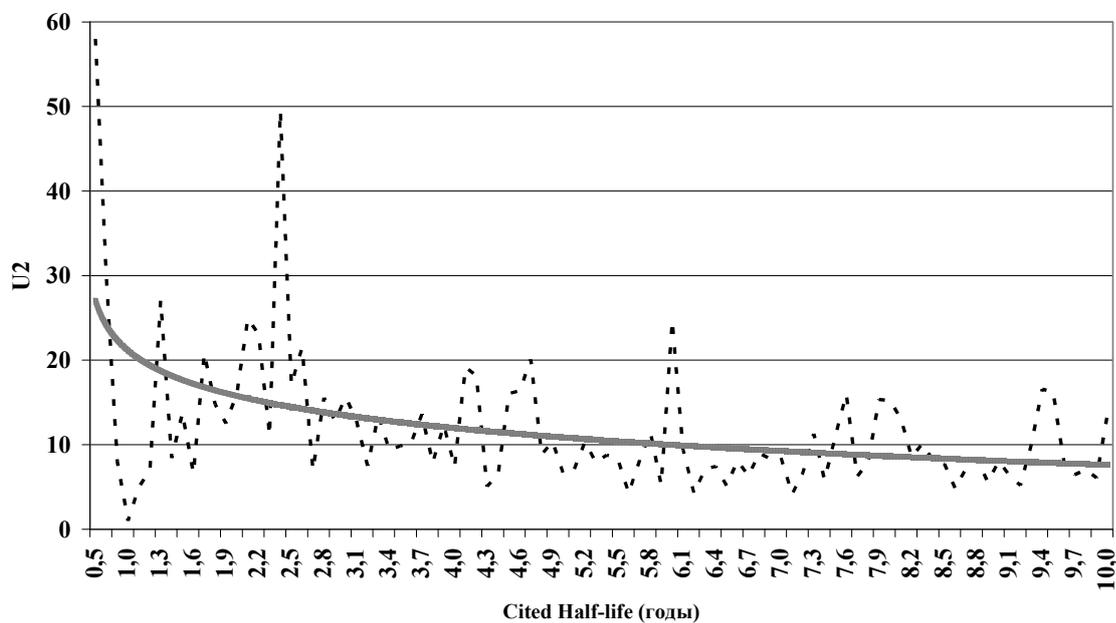


Рис. 4. Зависимость средних значений U2 (в расчете на 1 статью) от значений Cited_Half-life в журналах JCR-SE - по массиву российских публикаций 2015 г., представленных в SCI-E: пунктирная кривая – реальные расчетные значения U1, сплошная кривая – линия тренда

В обоих случаях кривые этих распределений по форме несколько напоминают «колокол» нормального распределения: по мере увеличения значений Cited Half-life доля журналов неуклонно растет, а затем, достигнув некоторого максимального значения, – так же неуклонно снижается⁵. Для мирового массива журналов медиана и мода распределения достаточно близки: медианное значение – 6,15 года, мода – примерно 6,25. Для российских распределений эти статистические характеристики отстоят друг от друга заметно дальше: медиана – 6,8, мода – 7,2. Судя по значению медианы, а также визуально, российское распределение журналов сдвинуто в сторону журналов с большими значениями CdHL. Следует помнить, что из-за отсутствия в JCR дифференциации для значений Cited Half-life, превышающих 10 лет, приведена на рис. 1 отсутствует «хвост» распределений. Причем доля журналов с такими значениями Cited Half-life значительна: для мирового распределения она составляет 15,3% от всех журналов, у которых есть данные о Cited Half-life, а в случае аналогичного российского – 20,2%.

Кривая мирового распределения доли публикаций, содержащихся в журналах с данными значениями Cited Half-life (рис. 2.) как и распределения самих журналов (рис. 1) напоминает «колокол» нормального распределения, а медиана и мода практически совпадают – обе характеристики находятся в пределах 5,8 года. Аналогичное российское распределение резко отличается от мирового. Оно еще больше сдвинуто в сторону больших значений Cited Half-life: медиана российского распределения больше медианы мирового распределения на целый год: 6,8 и 5,8 – соответственно. Значения медианы и моды российского распределения достаточно далеки друг от друга: 6,8 и 8,0 – соответственно, само распределение на графике выглядит также достаточно асимметрично. Как и в случае распределения журналов, в распределениях мировых и российских публикаций «обрублен хвост» распределения: отсутствующие на графике публикации (они соответствуют журналам, у которых $CdHL > 10$).

Коэффициенты корреляции на уровне журналов

Предварительный анализ данных показал, что с увеличением значений Cited Half-life значения $U1$ и $U2$ в целом уменьшаются. Поэтому, в отличие от корреляции на уровне публикаций, для определения корреляции на уровне журналов, минимальный, т.е. самый высокий, ранг присваивался группе журналов с максимальным значением Cited Half-life. Таким образом, в качестве нулевой гипотезы необходимо предположить, что увеличение значений Cited Half-life не ведет к уменьшению значений $U1$ и $U2$. Полученное значение τ для всего массива российских публикаций для значений $U1$ оказалось больше (хотя и не на много), чем соответствующее значение T_{kr} при 0,031 и 0,025 – соответственно. Следовательно, нулевую гипотезу можно отвергнуть и признать, что

существует обратная, причем достаточно слабая ($k = 0,031 / 0,025 = 1,24$) корреляция между значениями $U1$ и Cited Half-life. В случае $U2$ мы прибегли к аналогичному приему. Результаты расчета: $\tau = 0,027$. Вычислим значение T_{kr} , задав $\alpha = 0,05$, получим $T_{kr} = 0,025$. Следовательно $k = 1,08$. Отсюда, необходимо признать, что существует обратная, причем еще более слабая корреляция между значениями $U2$ и Cited Half-life. Столь малые значения τ и k подтверждают наше предположение о том, что в данном случае использование коэффициента корреляции на уровне журналов оказывается малоэффективным и необходимо переходить от коэффициента корреляции к корреляционным зависимостям.

Определение корреляционных зависимостей

На рис. 3 представлен график корреляционной зависимости значений $U1$ от значений Cited Half-life. Из рис. 3 следует, что с увеличением значений Cited Half-life значения $U1$ уменьшаются. Линия тренда на этом графике построена в Excel и представляет собой логарифмическую кривую. На рис. 4 представлен аналогичный график для показателя $U2$. Из рис. 4 следует, что с увеличением значений Cited Half-life значения $U2$ уменьшаются. Линия тренда на этом графике также построена в Excel и, как и в случае с $U1$, представляет собой логарифмическую кривую.

ЗАКЛЮЧЕНИЕ

Предложена методика оценки связи между традиционными библиометрическими и альтметрическими показателями, основывается на определении коэффициента ранговой корреляции Кенделла и определении корреляционных зависимостей. Предложен прием (коэффициент k), позволяющий более точно оценить степень корреляции. На уровне публикаций методика применена для оценки корреляции между цитированием (TC) и использованием ($U1$ и $U2$) массива публикаций российских авторов за 2015 г., а также еще 8 тематических подмассивов, сформированных из исходного массива. Во всех этих случаях оказалось, что существует тесная корреляция между указанными показателями. На уровне журналов методика использована для вычисления коэффициента корреляции между значениями Cited Half-life и показателями использования журналов, в которых вышли российские публикации за 2015 г. В этом случае обнаружена довольно слабая отрицательная корреляция между Cited Half-life и $U1$ ($U2$). Кроме того на уровне журналов построены корреляционные зависимости между этими показателями. Анализ показал, что связь между Cited Half-life и $U1$ ($U2$) описывается обратной логарифмической зависимостью.

Осуществлено сопоставление российских и мировых распределений журналов и соответствующих им публикаций по величине Cited Half-life. Это сопоставление показало, что российское распределение журналов практически совпадает с аналогичным мировым, однако несколько смещено в сторону журналов с большими значениями Cited Half-life (медиана российского распределения – 6,2, мирового – 6,1). Это смещение резко возрастает при переходе на уровень публикаций (медиана российского распределения – 6,8, мирового – 5,8).

⁵ Сплошные кривые на рис. 1 – это линию тренда, которые построены в Excel и представляют собой многочлены шестой степени, в котором в качестве независимой переменной выступает Cited Half-life (ось X), а в качестве зависимой – доля журналов с данным Cited Half-life от соответствующего массива.

СПИСОК ЛИТЕРАТУРЫ

- Gonzalez-Valiente CL, Pacheco-Mendoza J, Arencibia-Jorge R. A Review of Altmetrics as an Emerging Discipline for Research Evaluation // Learned Publishing. – 2016. – Vol. 29, № 4. – P. 229-238. DOI: 10.1002/leap.1043
- Thelwall M., Wilson P. Mendeley Readership Altmetrics for Medical Articles: An Analysis of 45 Fields // Journal of the Association for Information Science and Technology. – 2016. – Vol. 67, № 8. – P. 1962-1972. DOI: 10.1002/asi.23501.
- Haustein S., Peters I., Bar-Ilan J., Priem J., Shema H., Terliesner J. Coverage and Adoption of Altmetrics Sources in the Bibliometric Community // Scientometrics. – 2014. – Vol. 101, № 2. – P. 1145-1163. DOI: 10.1007/s11192-013-1221-3
- Costas R., Zahedi Z., Wouters P. Do "Altmetrics" Correlate with Citations? Extensive Comparison of Altmetrics Indicators with Citations from a Multidisciplinary Perspective // Journal of the Association for Information Science and Technology. – 2015. – Vol. 66, № 10. – P. 2003-2019. DOI: 10.1002/asi.23309
- Gumpenberger C., Glänzel W., Gorraiz J. The Ecstasy and the Agony of the Altmetrics Score // Scientometrics. – 2016. – Vol. 108, № 2. – P. 977-982. DOI: 10.1007/s11192-016-1991-5
- Wang X, Zhichao F, Sun X. Usage Patterns of Scholarly Articles on Web of Science // Scientometrics. – 2016. – Vol. 109, № 2. – P. 917-926. DOI: 10.1007/s11192-016-2093-2100
- Chi P., Glänzel W. An empirical investigation of the associations among usage, scientific collaboration and citation impact // Scientometrics. – 2017. – Vol. 112, № 7. – P. 403-412. DOI: 10.1007/s11192-017-2356-4 T7
- Арапов М.В., Либкинд А.Н., Либкинд И.А., Минин В.А. Информационная система «Указатель РФФИ» – средство создания сетевой структуры российской науки // Труды конференции САИТ–2005 «Системный анализ и информационные технологии». – М.: URSS, 2005. - Т. 2. - С. 9-14.
- Либкинд А.Н., Минин В.А., Либкинд И.А., Маркусова В.А., Янц М. Указатель РФФИ как инструмент информационного поиска и анализа тенденций в фундаментальной науке России // Материалы Международной конференции НТИ-2007 «Информационное общество, интеллектуальная обработка информации, информационные технологии» (Москва, 24-26 октября 2007 г.). – М.: ВИНТИ, 2007. – С. 180-183.
- Glänzel W. Characteristic scores and scales. A bibliometric analysis of subject characteristics based on long-term citation observation // Journal of Informetrics. – 2007. – № 1. – P. 92-102.
- Chi P., Glänzel W. Impact and usage indicators for the assessment of research in scientific disciplines and journals // Scientometrics. – 2017. – in press.

ПРИЛОЖЕНИЕ

Категории WoS, с помощью которых формировались массивы публикаций

Тематика	Категории WoS
Биология	Biology; Reproductive Biology; Cell Biology; Biochemistry & Molecular Biology; Developmental Biology; Evolutionary Biology; Mathematical & Computational Biology; Marine & Freshwater Biology
Биохимия и молекулярная биология	Biochemistry & Molecular Biology
Технологии и инженерное дело	Agricultural Engineering; Cell & Tissue Engineering; Computer Science, Software Engineering; Engineering, Electrical & Electronic; Engineering, Industrial; Engineering, Chemical; Engineering, Biomedical; Engineering, Environmental; Engineering, Civil; Engineering, Petroleum; Metallurgy & Metallurgical Engineering; Engineering, Aerospace; Engineering, Mechanical; Engineering, Geological; Engineering, Manufacturing; Engineering, Marine; Engineering, Multidisciplinary; Engineering, Ocean
Математика	Mathematics; Mathematics, Interdisciplinary Applications; Mathematics, Applied
Медицина	Medicine, General & Internal; Medicine, Research & Experimental; Nuclear Medicine & Medical Imaging; Critical Care Medicine; Oral Surgery & Medicine; Emergency Medicine; Medicine, Legal; Tropical Medicine; Integrative & Complementary Medicine
Микробиология	Microbiology; Biotechnology & Applied Microbiology
Физика конденсированного состояния	Physics, Condensed Matter

Сведения об авторах

ЛИБКИНД Александр Наумович – кандидат технических наук, ведущий научный сотрудник Всероссийского института научной и технической информации Российской академии наук (ВИНИТИ РАН), Москва
e-mail: anliberty@mail.ru

МАРКУSOVA Валентина Александровна – доктор педагогических наук, зав. Отделением ВИНИТИ РАН
e-mail: valentina.markusova@gmail.com

БОГОРОВ Валентин Григорьевич – руководитель образовательных программ
Компании Clarivate Analytics,
e-mail: valentin.bogorov@clarivateanalytics.com

Н.С. Редькина

Сайты библиотек сквозь призму веб-аналитики

Рассмотрены возможности инструментов веб-аналитики для детального исследования основных направлений работы библиотек в веб-среде, изучения поведения пользователей и анализа востребованности библиотечно-информационных продуктов и услуг, предлагаемых на сайтах. Представлены параметры оценки аудитории библиотечных веб-ресурсов (посещаемость, источники трафика, география пользователей и др.) с помощью плагинов, лог-анализаторов, счетчиков, рейтингов и сервисов (Google Analytics, Яндекс.Метрика, Hotlog). Проведен анализ сайтов ряда библиотек мира маркетинговыми инструментами (Hubspot Website Grader, Nibbler, CheckTrust, Google's page speed insights, Hemingway).

Ключевые слова: библиотеки, сайты, эффективность, оценка, веб-аналитика, Google Analytics, Яндекс.Метрика, маркетинговые веб-инструменты

ВВЕДЕНИЕ

Веб-технологии стали неотъемлемой частью взаимодействия библиотеки и пользователя, так как предоставляют широкие возможности для адаптации библиотечно-информационных ресурсов и услуг к потребностям современных пользователей, способствуют формированию благоприятного имиджа и положительной репутации библиотек в виртуальном мире, содействуют привлечению большего количества читателей за счет повышения социальной активности в веб-среде, применения новых способов и форм работы. Библиотеки создают актуальные экстерриториальные, мультимедийные, интерактивные и таргетированные ресурсы, развивают современные сервисы интерактивной коммуникации. Благодаря этому становятся возможными эффективные коммуникации с целевыми аудиториями, организация профессионального общения и взаимодействия с пользователями, внедрение новых форм библиотечно-информационного обслуживания, реализация обучающей деятельности на различных веб-платформах, реклама библиотечных культурно-массовых и иных мероприятий с использованием расширенного спектра информационных каналов и др.

Результаты исследования, проведенного в 2016 г., подтверждают активность библиотек в веб-среде. В результате анализа собранных данных было определено, что сайт имеют 92% библиотек (от 336 респондентов – представителей библиотек разных организационно-правовых форм). Большая часть сайтов (42%) расположена на платных хостингах (хостинг – предоставление сторонней организацией дискового пространства для размещения сайта), 21% – на бесплатных, 19% – представлены на сайтах организаций, в которые входят библиотеки [1]. Вместе с тем, как и любые другие библиотечно-информационные продукты, веб-сайты библиотек нуждаются в управлении, базирующимся на анализе различных данных,

получаемых по результатам экспертной оценки, опросов, маркетинговых исследований, показателей счетчиков и лог-анализаторов, инструментов веб-аналитики, позволяющем изучать востребованность ресурсов, проводить оценку их информативности и полезности для определенных категорий пользователей.

ОЦЕНКА БИБЛИОТЕЧНЫХ САЙТОВ ИНСТРУМЕНТАМИ ВЕБ-АНАЛИТИКИ

Наибольшую актуальность в оценке веб-ресурсов получают различные веб-аналитические системы измерения, сбора, анализа, представления и интерпретации информации о посетителях ресурсов, предоставляющие информацию в автоматизированном режиме, удобных форматах выгрузки данных по заданным показателям. Среди инструментов веб-аналитики интерес для библиотек могут представлять следующие: плагины (например, плагин статистики *WordPress Stats*), лог-анализаторы (*Analog*, *Webalizer*, *AWStats* и др.), счетчики (*OpenStat* (ранее *Spylog*), *24log.ru*), рейтинги (*LiveInternet*, Рамблер.Топ100, *top.mail.ru*, *Ruward*, *Directrix*) и сервисы (*Google Analytics*, Яндекс.Метрика, *Hotlog*) [2].

Р. O'Brien и др. в работе [3] справедливо отмечают, что лог-аналитика предоставляет очень подробную информацию о коммуникационной активности, но поскольку программное обеспечение управляется локально, оно способно налагать определенное административное бремя. Анализ лог-журналов может оказаться затруднительным, если он объединяет данные из более чем одного физического веб-сервера. Когда несколько серверов составляют веб-сайт организации, требуется ручная компиляция отчетов. С другой стороны, явное преимущество анализа журналов заключается в том, что пользовательская и институциональная информация не предоставляется третьим лицам.

Современные сервисы веб-аналитики являются хорошей альтернативой журналам. Они предлагают простоту использования, интеллектуальную презентацию представления данных в режиме реального времени, доступность и др. [4]. Но они также имеют определенные ограничения [5, 6]. Однако исследования показывают, что большинство академических библиотек по-прежнему их используют [1].

Обобщая возможности современных инструментов веб-аналитики, выделим параметры, позволяющие провести оценку аудитории веб-ресурса и проанализировать посещаемость сайта по различным данным:

- общая статистика, включающая количество посетителей на сайте (визитов), число просмотренных ими страниц (просмотры), количество новых/уникальных посетителей, география пользователей, учет демографических характеристик и др.;

- источники трафика пользователей (поисковый, прямой и т.д.), позволяющие проанализировать, откуда пришли пользователи, получить статистику по поисковым запросам (ключевым фразам), системам и т.д.;

- детальный анализ посещаемости: наиболее востребованные разделы и ресурсы сайта (страницы, на которые чаще всего заходят пользователи и дольше всего задерживаются), статистика просмотра страницы в отдельности или веб-ресурса в целом и каждой метрики в отдельности (например, средняя глубина просмотра сайта, конверсия, показатель отказов и др.), маршрут посещения (по каким путям идет просмотр различных страниц на веб-ресурсе, точки входы и выхода пользователей) и др.

Кроме того, инструменты веб-аналитики позволяют оценить технические параметры и такие проблемы, как например, некорректные ссылки ресурса (исходящие ссылки, ведущие на несуществующий домен в Интернете), ошибка 404 (переход с сайта на страницу, которая по каким-то причинам не находится в индексе поисковой системы), перегрузка страниц сайта, а также индексация поисковыми роботами. Метрики веб-аналитики помогают определить, насколько легко ориентируются пользователи на сайте и как быстро находят нужную информацию.

Инструменты веб-аналитики, хотя и не могут прогнозировать потребности пользователей, однако описывать их поведение. Эти данные неопенимы при принятии решений в отношении развития библиотечных веб-сайтов [7]. Отслеживание соответствующих показателей с помощью хорошо продуманной аналитической программы не только способствует принятию проектных решений, но и демонстрирует степень, в которой эти решения преуспели, и тем самым создают основу следующей итерации в процессе проектирования сайтов [8].

Положительный опыт использования данных веб-аналитики описан в профессиональной печати. В *Boise State University Albertsons Library* (США) анализ данных *Google Analytics* привел к созданию веб-сайта с показателем отказов на 10 или меньше, к уменьшению количества кликов, необходимых пользователям для доступа к содержимому библиотеки [9]. Сайты двух библиотек (*Rutgers-Newark Law Library* и *New Jersey Digital Legal Library*) также были переработаны и стали больше соответствовать ин-

формационным потребностям посетителей и библиотекарей [10]. Пример и рекомендации по визуализации статистики, интерпретации данных и настройке отчета на основе реального опыта *Texas Tech University Libraries* (<http://apps.lib.ttu.edu/>), приведен в статье *L. Yang и J. M. Perrin* [6].

Как показывает практика, на библиотечных сайтах чаще всего устанавливается не один счетчик, а два и даже более. Это обусловлено, прежде всего, многообразием инструментов, представленных в сети, различием метрик и способов визуализации. Недостатком установки двух систем (счетчика и анализатора логов) является то, что эти программные средства имеют разные алгоритмы сбора данных. Поэтому их результаты могут значительно отличаться. К примеру, показатель количества визитов за один месяц (июнь 2017 г.) сайта ГПНТБ СО РАН в системах *Google Analytics* и *AWStats* составляет 61 095 и 87 691, уникальных посетителей – 46 487 и 53 540 соответственно. Средняя длительность просмотра страницы также различается: *Google Analytics* – 2,3 минуты, *AWStats* – 2 минуты. В связи с этим для отчетности, особенно в рамках деятельности библиотек одной организационно-правовой формы или ведомственной принадлежности, рекомендуется использовать один счетчик, а для получения дополнительной аналитики применять другие системы.

Функционал инструментов веб-аналитики также может различаться. Бесплатные счетчики (серверы статистики) обычно содержат лишь базовые возможности, коммерческие обладают расширенной функциональностью. Для выбора инструментов веб-аналитики рекомендуется изучить различные варианты счетчиков и сервисов, метрик, используемых счетчиками, формируемые ими отчеты. Практические советы по работе с наиболее интересными системами веб-аналитики даны в вышедших в последние несколько лет изданиях А. Кошика, М. Хасслера и др. [11–13]. Кроме того, в литературе можно найти сравнение функциональных возможностей, достоинств и недостатков, показателей различных инструментов, которые наиболее широко применяются в России [14–19]. Проблемы анализа и оценки эффективности библиотечных сайтов рассматриваются в публикациях С.К. Канна, О.В. Кулевой, Е.В. Ковязиной и др. [20–26].

Ряд авторов отмечают, что Интернет не должен быть единственным источником информации для оценки веб-сайта. Рекомендуется использовать данные кликов в сочетании с другими методами тестирования юзабилити сайтов [27].

МАРКЕТИНГОВЫЕ ВЕБ-ИНСТРУМЕНТЫ АНАЛИЗА САЙТОВ

Эффективный сайт библиотеки – это востребованный, посещаемый ресурс, который по своему значению отвечает запросам пользователей. Важнейшими задачами являются не только создание библиотечно-информационного продукта по актуальной и интересной тематике, но и правильное его представление на сайте, разработка структуры сайта и составление семантического ядра. Под семантическим ядром подразумевается набор слов, словоформ

и словосочетаний, употребляемых определенной целевой аудиторией для поиска информации. Эти элементы должны очень точно описывать генерируемые библиотекой продукты и услуги. На основе корректно составленного семантического ядра создается структура сайта и определяется дальнейшая стратегия его продвижения, развития, совершенствования и оптимизации под меняющийся рынок информационных ресурсов.

Решающее значение для увеличения трафика, повышения коэффициента конверсии, приумножения количества потенциальных пользователей и повышения интереса к ресурсам приобретает оптимизация производительности веб-сайта. В этом вопросе могут помочь различные маркетинговые инструменты анализа сайтов (*Quicksprout, Hubspot Website Grader, MOZ Open Site Explorer* и др.). Нами были изучены более 30 различных веб-инструментов. Рассмотрим некоторые из них.

Hubspot Website Grader (<https://website.grader.com/>) используется для изучения видимости сайта поисковыми ботами, анализа заголовков и метаописания, оценки его «мобильности». Этот маркетинговый инструмент дает общую информацию и оценивает веб-сайт в масштабе от 1 до 100, предлагая советы по оптимизации. Среди основных рекомендаций выделим следующие: заголовки страниц следует делать не длиннее 70 символов, в них нельзя повторять ключевые слова; метаданные не могут содержать более 155 символов и должны относиться к странице; обязательно наличие карты сайта, позволяющей пользователям легко осуществлять навигацию.

Анализ сайта ГПНТБ СО РАН этим инструментом выявил некоторые проблемы, в частности то, что заголовочные метки отличаются от заголовков основной страницы. При этом к самим заголовкам страниц у *Website Grader* вопросов не возникло. Более того, отмечено, что названия страниц «великолепные».

Исследование сайтов за один месяц (июнь 2017 г.) нескольких библиотек мира с помощью *Hubspot Website Grader* (табл. 1) позволило выстроить их рейтинг по общему баллу и определить эффективность сайтов с точки зрения представления в сети, адаптации к мобильным версиям, *SEO* и безопасности (баллы в отношении к максимальному количеству баллов).

Nibbler (<http://nibbler.silktide.com/>) – маркетинговый инструмент, использующий более 20 тестов для оценки сайта. Данные разбиты на четыре основные области:

- 1) доступность – насколько сайт доступен для мобильных пользователей и пользователей с ограниченными возможностями;
- 2) удовлетворенность пользователей;
- 3) маркетинг – маркетинг и популярность сайта;
- 4) технологии – насколько сайт хорошо построен и спроектирован.

Программа *Nibbler* предлагает оценку от 1 до 10 для каждой из этих областей. По сравнению с *Hubspot Website Grader* он дает более подробную информацию о том, что необходимо для работы (от заголовков страниц до качества кода), сортирует предметы в соответствии с приоритетом – от самого высокого до самого низкого.

Таблица 1

Примеры анализа сайтов библиотек мира инструментом *Hubspot Website Grader*

Название библиотеки / адрес в сети	Средний балл	Представление	Адаптация под мобильные устройства	SEO	Безопасность
Библиотека Конгресса США (www.loc.gov)	81	21/30	30/30	20/30	10/10
ГПНТБ СО РАН (www.spsl.nsc.ru)	69	14/30	30/30	25/30	0/10
Немецкая национальная библиотека (http://www.dnb.de)	52	22/30	15/30	15/30	0/10
РГБ (rsl.ru)	37	12/30	0/30	15/30	10/10
Национальная библиотека Франции (http://www.bnf.fr)	36	26/30	0/30	10/30	10/10
ГПНТБ России (gpntb.ru)	32	22/30	0/30	10/30	0/10

Для сайта ГПНТБ СО РАН *Nibbler* определил рейтинг популярности (48 618 место в мире, июль 2017 г.), положительные моменты (например, сайт правильно обрабатывает недостающие страницы, отправляет 404 код статуса HTTP, использует кодировку GZIP, что уменьшает время загрузки веб-страницы) и следующие приоритеты для улучшения:

- перечень ошибок, препятствующих выполнению рекомендаций W3C (Консорциум *World Wide Web*, также известный как W3C): доступность сайта для лиц с ограниченными возможностями; доступность из разных веб-браузеров и с различных устройств; минимальное кодирование для обеспечения быстрой скорости загрузки и удобной навигации и др.;

- сокращение длинного якорного текста, т. е. текста ссылки, который виден пользователю (в 16 ссылках слишком длинный текст, ссылки не описывают ресурс, в 51 ссылке нет текста, описывающего их значение);

- отсутствие мета-описания на некоторых страницах сайта (только 25% страниц имеют мета-тег описания), позволяющего контролировать то, как будет выглядеть результат в поисковой системе;

- 4 страницы сайта используют презентационные HTML-элементы и / или атрибуты. Следует избегать использования презентационных HTML, таких как `` и `<p bgcolor="#FF0000">`;

- некоторые изображения (26%) не имеют определенных размеров с использованием атрибутов `width` и `height` в HTML. Если это исправить, то браузер сможет быстрее упорядочить страницу и др.

Woorank (<https://www.woorank.com/>), как и *Nibbler*, позволяет получить много информации о сайте. Он обеспечивает общую оценку от 1 до 100 и содержит подробную информацию о *SEO*, «мобильности», юзабилити, технологиях. Инструмент предоставляет контрольный список ключевых слов, заголовков, тегов для оптимизации сайта, а также показывает сложность контекста. По анализу трафика посещений также определяется рейтинг в стране и мире. Согласно *Woorank* сайт ГПНТБ СО РАН в рейтинге посещаемости (июль 2017 г.) в России находится на 3 908 месте, в мире – 48 688, что сопоставимо с данными *Nibbler*.

CheckTrust (<https://checktrust.ru/>) – сервис проверки качества сайтов. Он определяет качественные ссылочные доноры по показателям траста (от англ. –

доверие). Механизм работы данного инструмента основывается на теории, согласно которой существует некий пул сайтов, признанных авторитетными. Эти сайты ссылаются на другие сайты в сети, передавая им часть своего траста, те сайты, в свою очередь, ссылаются еще на другие и т.д. В итоге можно выстроить виртуальную цепочку ссылок, в начале которой находятся самые авторитетные сайты, а в конце сайты, которые доверия не вызывают. Условно наибольший авторитет обозначен как Траст=100, а наименьший авторитет – Траст=0, т.е. чем больше кликов по ссылкам отделяет интересующий нас сайт от самого авторитетного, тем меньше его траст. Примеры оценки качества сайтов библиотек инструментом *CheckTrust* приведены в табл. 2. По данным *CheckTrust* получить ссылку с сайтов анализируемых библиотек крайне полезно.

Google's page speed insights (<https://developers.google.com/speed/pagespeed/insights>) – инструмент, который дает возможность проверять скорость загрузки сайта и получать рекомендации по оптимизации страниц. Анализ страниц сайтов нескольких библиотек мира показал, что есть определенные проблемы, требующие решения (табл. 3). Уровень оптимизации страниц для компьютеров и мобильных устройств варьируется от плохого (страница не оптимизирована, будет загружаться долго), до среднего (страница может медленно загружаться, поскольку при ее создании не применены подходящие приемы оптимизации). Во всех случаях приведены конкретные рекомендации для улучшения работы сайта.

В частности, для веб-страницы ГПНТБ СО РАН предложения связаны: с удалением кода *JavaScript* и *CSS*, блокирующих отображение верхней части страницы (30 блокирующих скриптов на странице, 18 ресурсов *CSS*, замедляющих отображение контента); сокращением объема HTML-кода в верхней части страницы для того, чтобы она быстрее открывалась в браузере (после получения всех данных HTML с сервера удалось отобразить только 67 % от общего объема содержания из верхней части страницы); оптимизацией конкретных изображений для уменьшения их размера на 10,1 КБ (19 %) и др. Выполнение этих рекомендаций позволит оптимизировать сайт и улучшить работу с ним пользователей.

Таблица 2

Примеры оценки качества сайтов библиотек инструментом *CheckTrust*

Оценка донора	ТРАСТ	Примеры проверки библиотек России
Отличный донор	от 51 до 100	РНБ (100 траст / 100; 3 спам / 100) РГБ (100 траст / 100; 6.2 спам / 100) ГПНТБ СО РАН (84 траст / 100; 6.8 спам / 100)
Хороший донор	от 31 до 50	БАН РАН (37 траст / 100; 0.8 спам / 100)
Нежелательный донор	от 0 до 30	–

**Примеры оценки качества сайтов библиотек инструментом
Google's page speed insights**

Название библиотеки / адрес в сети	Для мобильных устройств	Для компьютеров	Уровень оптимизации для разных устройств
Библиотека Конгресса США (www.loc.gov)	49 / 100	58 / 100	Плохой
ГПНТБ СО РАН (www.spsl.nsc.ru)	68 / 100	84 / 100	Нужна доработка
Немецкая национальная библиотека (http://www.dnb.de)	37 / 100	45 / 100	Плохой
РГБ (rsl.ru)	47 / 100	57 / 100	Плохой
Национальная библиотека Франции (http://www.bnf.fr)	56 / 100	67 / 100	Плохой / Нужна доработка
ГПНТБ России (gpntb.ru)	64 / 100	72 / 100	Плохой / Нужна доработка

Отметим, что хороший уровень оптимизации для компьютеров обнаружен у сайта Мировой цифровой библиотеки (86 / 100). Инструментом *Google's page speed insights* установлено, что при создании этой страницы (<http://wdl.org/>) были учтены рекомендации, поэтому она является удобной для пользователей.

Помимо получения информации об эффективности сайта существуют инструменты, позволяющие оценить контент и корректировать его при необходимости. Провести анализ контента страниц сайта, упростить текст, расставить акценты, избавиться от ненужных наречий и, в итоге, облегчить его чтение помогает онлайн редактор *Hemingway* (<http://www.hemingwayapp.com/>), названный в честь известного писателя, который славился лаконичностью и простотой стиля письма. Оценить сложность веб-страницы можно, скопировав и вставив ее в редактор. После непродолжительной проверки сложные конструкции будут маркированы цветом, в подсказках отразятся рекомендации по изменениям и правке. Когда эти рекомендации будут выполнены, *Hemingway* поменяет присвоенный уровень со «сложного» на «легкий».

В результате анализа англоязычной страницы сайта по истории ГПНТБ СО РАН (<http://www.spsl.nsc.ru/en/o-biblioteki/biblioteka-segodnya/>) и аналогичной страницы Библиотеки Конгресса (LC) США (<https://www.loc.gov/about/history-of-the-library>) были получены следующие результаты: 3 фразы в тексте страницы ГПНТБ СО РАН и 2 LC имеют более простые альтернативы; 6 из 46 (2 из 21 LC) предложений трудно читать, 20 из 46 (15 из 21 LC) предложений очень трудно прочитать. Как показало наше исследование, трудности текста есть и в англоязычных сайтах, что может быть обусловлено алгоритмом, заложенным в редакторе и направленным на более упрощенный вариант подготовки текста в соответствии с требованиями поисковых систем и трендами в SEO-оптимизации (любой

текст на сайте должен быть максимально лаконичным и информативным). Результаты подобной оценки позволяют существенно доработать текст.

ВЫВОДЫ

Современные инструменты веб-аналитики являются эффективными для оценки деятельности библиотеки. Они позволяют объективно и по разным параметрам анализировать востребованность ресурсов и услуг, изучать поведение пользователей и их предпочтения, выявлять и устранять технические ошибки и, в результате, как оптимизировать веб-ресурсы, так и наметить стратегию развития библиотеки в веб-среде. Они не дадут ответов на все вопросы, но могут помочь библиотеке быть более продуктивной и эффективной в ее работе.

СПИСОК ЛИТЕРАТУРЫ

1. Редькина Н.С., Васильева Н.В., Коломенская А.С. Использование облачных сервисов в библиотеках России: результаты исследования // Научно-техническая информация. Сер. 1. – 2017. – № 6. – С. 21-30.
2. Редькина Н.С. Направления развития инструментов веб-аналитики // Научно-техническая информация. Сер. 2. – 2017. – № 5. – С. 5-10; Redkina N.S. The Development Tendencies of Web Analytics // Automatic Documentation and Mathematical Linguistics. – 2017. – Vol. 51, № 3. – P. 112-116.
3. O'Brien P., Arlitsch K., Sterman L., Mixer J. Undercounting File Downloads from Institutional Repositories // J. of Lib. Adm. – 2016. – Vol. 56, № 7. – P. 854-874. DOI: 10.1080/01930826.2016.1216224
4. Clark D.J., Nicholas D., Jamali H.R. Evaluating information seeking and use in the changing virtual world: the emerging role of Google

- Analytics // Learned publishing. – 2016. – Vol. 27, № 3. – P. 185-194.
5. Fagan J.C. The Suitability of Web Analytics Key Performance Indicators in the Academic Library Environment // J. Acad. Librarianship. – 2014. – P. 25-34. DOI: 10.1016/j.acalib.2013.06.005
 6. Yang L., Perrin J.M. How to Craft a Web Analytics Report for a Library Web Site // J. Web Librarianship. – 2014. – Vol. 8, № 4. – P. 404-417. DOI: 10.1080/19322909.2014.944296
 7. Barba I, Cassidy R, De Leon E, Williams B.J. Web Analytics Reveal User Behavior: TTU Libraries' Experience with Google Analytics // J. Web Librarianship. – 2013. – Vol. 7, № 4. – P. 389-400. DOI: 10.1080/19322909.2013.828991
 8. Loftus W. Demonstrating success: Web analytics and continuous improvement // J. Web Librarianship. – 2012. – Vol. 6, № 1. – P. 45-55. DOI: 10.1080/19322909.2012.651416
 9. Vecchione A., Brown D., Allen E. Tracking User Behavior with Google Analytics Events on an Academic Library Web Site // J. Web Librarianship. – 2016. – Vol. 10, № 3. – P. 161-175.
 10. Fang W. Using google analytics for improving library website content and design: A case study // Library Philosophy and Practice. – 2007. – URL: <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1121&context=libphilprac>
 11. Кошик А. Веб-аналитика 2.0 на практике : тонкости и лучшие методики / пер. с англ. – М.: Диалектика : Вильямс, 2016. – 526 с.
 12. Хасслер М. Веб-аналитика / пер. с нем. – М.: Эксмо, 2010. – 431 с.
 13. Яковлев А., Довжиков А. Веб-аналитика : основы, секреты, трюки. – СПб: БХВ-Петербург, 2010. – 266 с.
 14. Васьковский Е.Ю., Брумштейн Ю.М. Системный анализ функциональных возможностей счетчиков посещаемости сайтов // Прикаспийский журнал: управление и высокие технологии. – 2015. – № 3. – С. 96-113.
 15. Задорожный И.А., Медведева Е.С. Сравнительный анализ Яндекс метрики и Google Analytics как инструментов оценки эффективности диджитал-маркетинговых коммуникаций // Научные исследования: от теории к практике. – 2015. – Т. 2. – № 4 (5). – С. 161-164.
 16. Олевинский М.А. Веб аналитика. Сравнение систем веб аналитики // In Situ. – 2015. – № 4. – С. 46-48.
 17. Радзевич А.В. Веб-аналитика для бизнеса: как сделать правильные выводы об эффективности работы сайта // Интернет-маркетинг. – 2012. – № 4. – С. 218-225.
 18. Скородумов П.В., Холодов А.Ю. Анализ подходов и инструментальных средств анализа статистики посещения веб-сайта научной организации // Вопросы территориального развития. – 2015. – № 9 (29). – С. 6.
 19. Скородумов П.В., Холодов А.Ю. Анализ популярности веб-сайта научной организации с помощью различных систем сбора статистических данных // Вопросы территориального развития. – 2016. – № 1 (31). – С. 7.
 20. Канн С.К. Основы разработки системы показателей деятельности библиотечного Веб-сайта // Библиосфера. – 2012. – № 5. – С. 20-23.
 21. Канн С.К. Оценка деятельности библиотечного сайта с помощью средств веб-аналитики // Библиотеки национальных академий наук: проблемы функционирования, тенденции развития : науч.-практ. и теоретич. сб. – Киев, 2017. – Вып. 14. – С. 130-139. – URL: http://nbuv.gov.ua/UJRN/bnan_2017_14_16
 22. Канн С.К. Ключевые показатели эффективности библиотечного сайта // Труды ГПНТБ СО РАН. – Вып. 8 : Новые направления деятельности традиционных библиотек в электронной среде : материалы межрегион. науч.-практ. конф. (22-26 сент. 2014 г., г. Красноярск). – Новосибирск, 2015. С. 170-176.
 23. Ковязина Е.В. Оценка работы библиотеки с помощью веб-метрик // Вестн. БАЕ. – 2012. – № 4. – С. 47-51.
 24. Кулева О.В. Разработка системы оценки качества сайтов библиотек // Библиосфера. – 2009. – № 2. – С. 58 – 63.
 25. Turner S. J. Website Statistics 2.0: Using Google analytics to measure library website effectiveness // Techn. Serv. Quart. – 2010. – Vol. 27, № 3. – P. 261-278.
 26. Chandler A., Wallace M. Using piwik instead of google analytics at the Cornell university library // Serials Librarian. – 2016. – Vol. 71, № 3-4. – P. 173-179. DOI: 10.1080/0361526X.2016.1245645
 27. Farney T. A. Click Analytics: Visualizing Website Use Data // Inf. Tech. and Libr. – 2011. – Vol. 30, № 3. – P. 141-148.

Материал поступил в редакцию 15.12.17.

Сведения об авторе

РЕДЬКИНА Наталья Степановна – доктор педагогических наук, заместитель директора по научной работе, Государственная публичная научно-техническая библиотека Сибирского отделения Российской академии наук (ГПНТБ СО РАН), профессор Новосибирского государственного педагогического университета (г. Новосибирск)
e-mail: redkina@spsl.nsc.ru; to@spsl.nsc.ru

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 004.75.056

А.Ю. Щербаков

О разработке средств для формирования корпоративного распределенного реестра (блокчейн)

Обсуждается проблема разработки средств для формирования атомов корпоративного распределенного реестра (блокчейна), обеспечивающего защищенное распределенное хранение данных в связанной цепочке, рассматривается пример применения предлагаемых средств для реализации условного налогового платежа через смарт-контракт.

Ключевые слова: блокчейн, распределенное хранение данных, информационная безопасность, электронное голосование, шифрование, контроль целостности, смарт-контракт, средства разработки для блокчейн

ВВЕДЕНИЕ

Блокчейн как системная целостность должен состоять из отдельных элементов – звеньев, каждое из которых в свою очередь делится на элементарные компоненты (в [1] они названы **атомами блокчейна**). Атомы как элементы, ассоциированные или встроенные в компьютерную систему, могут быть пассивными и активными.

СТРУКТУРА БЛОКЧЕЙНА

Атомы блокчейна характеризуются идентификатором (именем) и имеют заданные свойства, необходимые как для формирования звеньев и всей цепочки, так и для выполнения свойств компьютерной (информационной) системы, которая использует блокчейн, например, свойства конфиденциальности пользовательских данных, помещаемых в распределенном реестре. Необходимо отметить, что существующие в настоящее время системы блокчейна не обеспечивают свойства конфиденциальности информации.

Выделим следующие типы атомов блокчейна.

1. Атом-граница (начало или конец) блокчейна, содержит ссылки на окончание предыдущего звена или начало следующего.

2. Атом-структура – определяет и описывает вектор идентификаторов атомов, имеющихся в звене и логично следует за атомом-граница (начало).

3. Атом-данные – содержит описание данных, интерпретируемых как пассивный компонент звена. Разновидностью атомов-данных являются:

- атом открытые данные (атом ОД) – характеризуется длиной и содержит данные, не подвергающиеся никакой обработке;

- атом целостные данные (атом ЦД) – характеризуется длиной данных зафиксированной целостности и идентификатором или ссылкой на процедуру контроля целостности данных;

- атом подписанные данные (атом ПД) – характеризуется длиной подписанных данных – идентификатором или ссылкой на процедуру проверки подписи под данными, а также ссылкой на ключ проверки подписи. Процедура проверки подписи может быть заданной внешне относительно блокчейна или быть атомом-субъектом;

- атом зашифрованные (закрытые) данные (атом ЗД) – характеризуется длиной зашифрованных данных – идентификатором или ссылкой на процедуру зашифрования/расшифрования данных, а также на ключ зашифрования или расшифрования;

- атом-подпись – содержит подпись атома или нескольких атомов – «подписанные данные» с заданным идентификатором (идентификаторами);

- атом-хеш – содержит эталон контроля целостности для атома или нескольких атомов «целостные данные» с заданным идентификатором (идентификаторами).

4. Атомы-субъекты – содержит описание данных, интерпретируемых как активный компонента звена.

Атом-субъект может быть:

- атомом-сценарием – содержащим интерпретируемый код для выполнения операций над атомами-данным;

- атомом-эксекутором – содержащим скомпилированный код для реального процессора, гипервизора компьютерной системы или атома-машины, в которой обрабатывается блокчейн;

- атомом-машиной – содержащим среду для выполнения атома-сценария или атома-экзакутора.

Далее рассмотрим элементарные модули, необходимые для формирования соответствующих атомов. Важно заметить, что в нашем случае речь идет о корпоративном распределенном реестре, поскольку в блокчейнах общего назначения атомы формируются иным образом. Однако предлагаемая технология позволяет защищать данные и в блокчейнах общего назначения, при этом информация предварительно формируется при помощи рассматриваемых здесь модулей, а затем уже помещается в блокчейн.

ЭЛЕМЕНТАРНЫЕ МОДУЛИ ДЛЯ ФОРМИРОВАНИЯ АТОМОВ КОРПОРАТИВНОГО БЛОКЧЕЙНА

Модуль первичного формирования исходного случайного числа и персонального идентификатора для пользователя

InitUser

Формат использования:

InitUser FileUserID UserPIN <RandomString>,

где:

FileUserID – имя файла с закрытым на пароле персональным идентификатором пользователя (может быть связано с именем пользователя),

UserPIN – пароль (пин-код или метод его ввода, например, чтения из USB-токена) для закрытия персонального идентификатора пользователя,

<RandomString> – необязательный параметр для улучшения работы датчика случайных чисел («разгонная строка»).

Модуль создает файл *random.bin* для дальнейшего использования датчика случайных чисел и файл с закрытым на пароле персональным идентификатором пользователя (фактически – защищенный контейнер для хранения и передачи персонального идентификатора пользователя).

Модуль возвращает типизированные ошибки, необходимые для интеграции вызовов модулей в смарт-контракт.

Возвращаемые ошибки:

- 1 – ошибка при тестировании модулей защиты,
- 2 – ошибка формата вызова,
- 3 – файл идентификатора уже существует,
- 4 – ошибка формирования случайного числа,
- 5 – ошибка обновления случайного числа,
- 6 – ошибка записи файла пользователя,
- 7 – ошибка контрольного чтения файла пользователя.

Модули формирования атомов блокчейна

GenAtomX

Модуль формирует атом-закрытые данные из открытых данных с использованием персонального идентификатора пользователя – атом типа X.

Формат использования:

GenAtomX FileUserID UserPIN file_or_string AtomFile,

где:

FileUserID – имя файла с закрытым на пароле персональным идентификатором пользователя,

UserPIN – пароль для закрытия персонального идентификатора пользователя,

file_or_string – информация для помещения в атом-3Д (закрытые данные), закрываемая на персональном идентификаторе пользователя,

AtomFile – файл, в который записывается атом.

Модуль использует файл с закрытым на пароле персональным идентификатором пользователя, открывает его и строит атом блокчейна согласно протоколу.

Возвращаемые ошибки:

- 1 – ошибка при тестировании модулей защиты,
- 2 – ошибка формата вызова,
- 3 – файл идентификатора пользователя не существует,
- 4 – ошибка формирования случайного числа,
- 5 – атом-файл уже существует,
- 6 – неверный пин-код,
- 7 – ошибка записи атома,
- 8 – ошибка контрольного чтения атома,
- 9 – ошибка обновления случайного числа.

GenAtomY

Модуль формирует атом-закрытые данные из открытых данных с использованием случайных данных для достижения заданного времени (PoW) – атом типа Y.

Формат использования:

GenAtomY file_or_string AtomFile Power,

где:

file_or_string – информация для помещения в атом-3Д (закрытые данные), закрытая на случайном числе заданной длины,

AtomFile – файл, в который записывается атом,

Power – двузначное число, обеспечивающее PoW (для тестирования рекомендуется значение 18-22).

Возвращаемые ошибки:

- 1 – ошибка при тестировании модулей защиты,
- 2 – ошибка формата вызова,
- 4 – ошибка формирования случайного числа,
- 5 – ошибка обновления случайного числа,
- 6 – атом-файл уже существует,
- 7 – ошибка записи атома,
- 9 – ошибка обновления случайного числа.

GenAtomH

Модуль формирует атом-хеш (атом целостные данные) из открытых данных с использованием персонального идентификатора пользователя – атом типа H.

Формат использования:

GenAtomX FileUserID UserPIN file_or_string AtomFile,

где:

FileUserID – имя файла с закрытым на пароле персональным идентификатором пользователя,

UserPIN – пароль для закрытия персонального идентификатора пользователя,

file_or_string – информация для помещения в атом-хеш, функция от открытых данных, вычисленная при помощи персонального идентификатора пользователя,

AtomFile – файл, в который записывается атом.

Возвращаемые ошибки:

- 1 – ошибка при тестировании модулей защиты,
- 2 – ошибка формата вызова,
- 3 – файл идентификатора пользователя не существует,

- 4 – ошибка формирования случайного числа,
- 5 – атом-файл уже существует,
- 6 – неверный пин-код,
- 7 – ошибка записи атома,
- 8 – ошибка контрольного чтения атома,
- 9 – ошибка обновления случайного числа.

Модули извлечения данных из звеньев блокчейна

ExcX

Модуль извлечения данных из атома типа X.

Формат использования:

ExcX FileUserID UserPIN AtomFile Result,

где:

FileUserID – имя файла с закрытым на пароле персональным идентификатором пользователя,

UserPIN – пароль для закрытия персонального идентификатора пользователя,

AtomFile – файл, в который помещен атом ЗД,

Result – файл с восстановленными данными.

Модуль имеет назначение тестирования пользователем своих действий и аудита пользователем транзакций.

Возвращаемые ошибки:

- 1 – ошибка при тестировании модулей защиты,
- 2 – ошибка формата вызова,
- 3 – файл идентификатора пользователя не существует,
- 4 – ошибка формирования случайного числа,
- 5 – атом-файл не существует,
- 6 – неверный пин-код,
- 7 – ошибка записи атома,
- 8 – ошибка контрольного чтения атома.

ExcY

Модуль извлечения данных из атома типа Y.

Формат использования:

ExcY AtomFile Result,

где:

AtomFile – файл, в который помещен атом ЗД,

Result – файл с восстановленными данными.

Модуль имеет назначение – выделение закрытых данных с заданной трудоемкостью PoW.

Возвращаемые ошибки:

- 1 – ошибка при тестировании модулей защиты,
- 2 – ошибка формата вызова,
- 3 – файл идентификатора пользователя не существует,
- 4 – ошибка формирования случайного числа,
- 5 – атом-файл не существует,
- 7 – ошибка записи результата.

CAtomH

Модуль проверяет целостность данных при помощи заранее вычисленного атом-хеш с использованием персонального идентификатора пользователя.

Формат использования:

GenAtomX FileUserID UserPIN file_or_string AtomFile,

где:

FileUserID – имя файла с закрытым на пароле персональным идентификатором пользователя,

UserPIN – пароль для закрытия персонального идентификатора пользователя,

file_or_string – информация для контроля целостности,

AtomFile – файл, в который был записан атом, содержащий контрольную информацию от данных.

Возвращаемые ошибки:

- 1 – ошибка при тестировании модулей защиты,
- 2 – ошибка формата вызова,
- 3 – файл идентификатора пользователя не существует,
- 4 – ошибка формирования случайного числа,
- 5 – атом-файл уже существует,
- 6 – неверный пин-код,
- 7 – ошибка записи атома,
- 8 – ошибка контрольного чтения атома,
- 9 – ошибка обновления случайного числа.

Модуль вычисления Power

Cpower

Формат использования

Cpower Time

Модуль возвращает число бит, которое необходимо для обеспечения трудоемкости в Time минут.

Все модули генерируют одноименные названию модуля текстовые журналы.

ПОНЯТИЕ СМАРТ-КОНТРАКТА

Первые идеи смарт-контрактов были предложены в 1996 г. Ником Сабо [2]. Практические реализации стали возможными, благодаря появлению в 2008 г. технологии блокчейн. Некоторые принципы смарт-контрактов были заложены в протоколе первой блокчейн-валюты (криптовалюты) *Bitcoin*, однако они не были реализованы в клиентском программном обеспечении, не обладали полнотой по Тьюрингу из соображений безопасности и не использовались на практике. С развитием технологии блокчейн, стали высказываться идеи, что поверх протокола биткойна могут быть созданы различные протоколы более высокого уровня, включая полноценные смарт-контракты, по аналогии с тем как поверх TCP/IP существует множество протоколов прикладного уровня.

Смарт-контракты впервые стали применяться на практике в проекте *Ethereum* (рус. Эфириум). Идея создания проекта появилась в 2013 г. В тот момент основатель журнала «*Bitcoin Magazine*» В. Бутерин пришёл к выводу, что биткойн плохо подходит в качестве базового протокола, поскольку изначально не был спроектирован под данную задачу, и написал в одной из своих статей об идее создания такого протокола с нуля.

Будем полагать, что смарт-контракт – исполняемый код, у которого зафиксирована целостность, помещенный в атом-эксектор и оперирующий с атомами и звеньями блокчейна. При этом результат работы смарт-контракта всегда помещается в новый атом или звено блокчейна. В противном случае идеология неизменно-сти звеньев блокчейна будет нарушена.

ПРИМЕР ФОРМИРОВАНИЯ И РАБОТЫ СМАРТ-КОНТРАКТА, ПОСТРОЕННОГО ИЗ ЭЛЕМЕНТАРНЫХ МОДУЛЕЙ

Полагаем, что пользователь nlp100 («налогоплательщик-100») имеет кошелек с криптовалютой (для оплаты налогов) spall00, на котором имеется 500 крипторублей.

Для тестового создания кошелька используем смарт-контракт:

```
genatomx nlp100_1 privet1 500 cna1100_u  
genatomx nlp100_3 privet3 500 cna1100_b.
```

В реальной платежной системе для создания кошелька будет необходимо выполнить согласованные с уполномоченным банком или оператором криптовалюты действия по идентификации пользователя, созданию кошелька и зачислению криптовалюты на него.

Пользователь *fns* (налоговая служба) выставил «налогоплательщику-100» налоговый платеж в 50 криптовалют в адрес банка *bank*, который оформил в виде смартконтракта *fns100_b.bat*.

Первично пользователь заводит свои контейнеры закрытого ключа по работе с кошельком *nlp100_1* для получения смартконтрактов и извещений от налоговой службы *nlp100_2* и для обмена с банком *nlp100_3* при помощи модуля *inituser*.

Смартконтракт представляет собой последовательный вызов методов создания атомов блокчейна *genatomx*, которые списывают с кошелька налогоплательщика-100 50 криптовалют и зачисляют их на счет банка, формируя атомы *a1* и *a2*:

```
genatomx nlp100_1 %1 -50 a1  
genatomx nlp100_3 %2 +50 a2.
```

Необходимо заметить, что для работы с кошельком налогоплательщика требуется предъявление пароля или приватного ключа, что указано аргументом *%1* (пароль или ключ вводится пользователем), соответственно, второй пароль или приватный ключ нужен для зачисления средств банку.

Налоговая служба заводит контейнер для обмена с «налогоплательщиком-100» – *fns100*, банк заводит контейнер для обмена с «налогоплательщиком-100» – *bank100*, банк для направления квитанций налоговой создает контейнер *bfns*.

Для создания контейнеров используется следующий смартконтракт:

```
inituser nlp100_1 privet1  
inituser nlp100_2 privet2  
inituser nlp100_3 privet3  
inituser fns100 privet2  
inituser bank100 privet3  
inituser bfns privet4.
```

Пароли или методы ввода приватного ключа используются для доступа пользователя к своему кошельку – *privet1* – приватный ключ, неизвестный никому, кроме пользователя, *privet2* – для обмена с налоговой службой, *privet3* – для обмена пользователя с банком и *privet4* – для обмена банка с налоговой службой.

Налоговая служба формирует атом *a0*, в котором находится зашифрованный текст смартконтракта *fns100_b.bat*:

```
genatomx nlp100_2 privet2 fns100_b.bat a0.
```

Налогоплательщик экстрагирует из атома смартконтракт:

```
excx nlp100_2 privet2 a0 user100.bat
```

и выполняет контракт *user100.bat*, вводя вместо *%1* и *%2* пароли или приватные ключи для доступа к своему кошельку и для передачи транзакции в банк.

Контракт создает звенья блокчейна *a1* и *a2*, в которых находятся измененное состояние кошелька налогоплательщика (уменьшение на 50 криптовалют) и распоряжение по зачислению необходимой суммы на кошелек банка.

Банк экстрагирует из атома распоряжение налогоплательщика по счету

```
excx nlp100_3 privet3 a2 b1.
```

Банк зачисляет деньги на свой кошелек (криптосчет) и высылает подтверждение налогоплательщику путем формирования очередного атома

```
genatomx nlp100_3 privet3 zachisleno=50 a3.
```

На стороне налогоплательщика автоматически списывается с кошелька зачисленная в банк сумма:

```
genatomx nlp100_1 privet1 450 cna1100_u  
genatomx nlp100_3 privet3 450 cna1100_b.
```

Новая копия кошелька создается со значением $450=500-50$ по уведомлению банка.

Банк имеет возможность проверить в *cna1100_b*, что сумма в кошельке изменилась:

```
excx nlp100_3 privet3 cna1100_b cna1100_t2.
```

На стороне клиента (налогоплательщика) операции с кошельком выполняются только при балансе средств – когда в кошельке клиента и банка одинаковые суммы.

Баланс проверяется при помощи следующего смартконтракта:

```
excx nlp100_1 privet1 cna1100_u cna1100_t1  
excx nlp100_3 privet3 cna1100_b cna1100_t2.
```

Последняя операция – формирование атома подтверждения *a4*:

```
genatomx bfns privet4 prinjato=nlp100=50 a4.
```

Налоговая служба экстрагирует атом *a4*

```
excx bfns privet4 a4 b3
```

и получает строку подтверждения оплаты налога.

Для реального использования необходимо ввести также подтверждение банковского платежа в казначейство.

Для демонстрации системы голосования можно выполнить следующий смартконтракт:

```
genatomx nlp100_1 privet1 Ivan_Ivanov g1  
genatomx nlp100_2 privet2 Petr_Petrov g2  
genatomy Ivan_Ivanov g3 19  
genatomy Petr_Petrov g4 20  
excx nlp100_1 privet1 g1 golos11  
excx nlp100_2 privet2 g2 golos21  
excy g3 golos12  
excy g4 golos22.
```

Для удобства используются уже созданные контейнеры.

Сначала формируются голоса за *Ivan_Ivanov* и *Petr_Petrov*, которые сформированы и доступны только проголосовавшим пользователям, а затем – доступны всем, но с раскрытием в заданное время, причем для *Petr_Petrov* время раскрытия в среднем в два раза больше. Затем голоса раскрываются в соответствующие файлы и сравниваются.

Журнал функции *excy* показывает примерное различие времен экстрагирования голосов – приблизительно 12 секунд и 31 секунда при одинаковом проценте перебора (около 30):

```
00:39:01 05.02.2018:Start excy
```

```
00:39:01 05.02.2018:Success Protect Function
```

00:39:01 05.02.2018:Ok Test Random
00:39:13 05.02.2018:RealCicle = 157705.000000
[Full cicles = 524288.000000] 30.079842 percent of
full cicles
Full time = 11.875000 sec
00:39:13 05.02.2018:Ok AtomFileY
00:39:13 05.02.2018:Start excy
00:39:13 05.02.2018:Success Protect Function
00:39:13 05.02.2018:Ok Test Random
00:39:44 05.02.2018:RealCicle = 410876.000000
[Full cicles = 1048576.000000] 39.184189 percent of
full cicles
Full time = 30.861000 sec
00:39:44 05.02.2018:Ok AtomFileY.

ВЫВОДЫ

Предлагаемая технология позволяет формировать практически весь необходимый спектр атомов блокчейна и выполнять операции с ними, используя технологию смарт-контрактов. Криптографические операции с атомами целесообразно реализовывать на основе отечественных криптоалгоритмов для обеспечения свойств доверенности и возможности последующей сертификации и аттестации разработанных решений у государственных регуляторов.

СПИСОК ЛИТЕРАТУРЫ

1. Биктимиров М.Р., Домашев А.В., Черкашин П.А., Щербаков А.Ю. Блокчейн: универсальная структура и требования // Научно-техническая информация. Сер. 2. – 2017. – № 11. – С. 1-4; Biktimirov M.R., Domashev A.V., Cherkashin P.A., Shcherbakov A.Yu. Blockchain Technology: Universal Structure and Requirements // Automatic Documentation and Mathematical Linguistics. – 2017. – Vol. 51, № 6. – p. 235 -238.
2. Ник Сабо. Умные контракты (Четвертая революция стоимости) // Компьютерра. – 1998. – № 38. – С. 12-19.

Материал поступил в редакцию 19.02.18

Сведения об авторе

ЩЕРБАКОВ Андрей Юрьевич – доктор технических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН, профессор МИЭМ НИУ ВШЭ, член Ученого совета ВИНТИ РАН, ведущий специалист по информационной безопасности Технопарк «Техпромбизнес», консультант по криптографии фирмы «Лаборатория Касперского», Москва
e-mail: x509@ras.ru

СПРАВОЧНО-ИНФОРМАЦИОННЫЙ РАЗДЕЛ

УДК 165.3:167:004.4

А.Л. Дзюбенко, Ю.Б. Чечиков

О связи объектно-ориентированного программирования с реальной жизнью

Рассматривается связь объектно-ориентированного программирования с реальным миром, философия объектно-ориентированного программирования, его методология в процессе отображения объектов реального мира на вычислительные средства.

Ключевые слова: объектно-ориентированное программирование, программирование, программное обеспечение, виды программирования, парадигма, методология, информационные процессы, информационные системы

Программирование – это в первую очередь процесс создания компьютерных программ, т.е. термин «программирование» означает процесс и искусство создания компьютерных программ с помощью специальных языков программирования.

В общем смысле слова, программирование есть формализация предопределенного состояния, реализуемого средствами математики или естественных наук.

В узком смысле слова, программирование рассматривается как кодирование алгоритмов на заданном языке программирования [1].

В более широком смысле программирование – процесс создания программ, т.е. разработка программного обеспечения.

Программирование включает:

- анализ поставленной задачи;
- проектирование – разработку комплекса алгоритмов;
- кодирование и компиляцию – написание исходного текста программы и преобразование его в исполнимый код с помощью компилятора;
- тестирование и отладку – выявление и устранение ошибок в программах;
- испытания и сдачу программ;
- процесс сопровождения.

На базе приведённых определений получаем, что программирование в общем понимании – это написание программ, где программа или приложение представляет собой просто набор команд, инструкций, выполняемых сценариев, которые исполняются машиной.

Перечислим несколько общепринятых видов программирования (рисунок):

Процедурное (императивное) программирование, состоящее из последовательности операторов, за-

дающих процедуру решения задачи, и являющееся отражением архитектуры традиционных ЭВМ программирования. Используется для реализации небольших подзадач, где очень важна скорость исполнения на современных компьютерах.

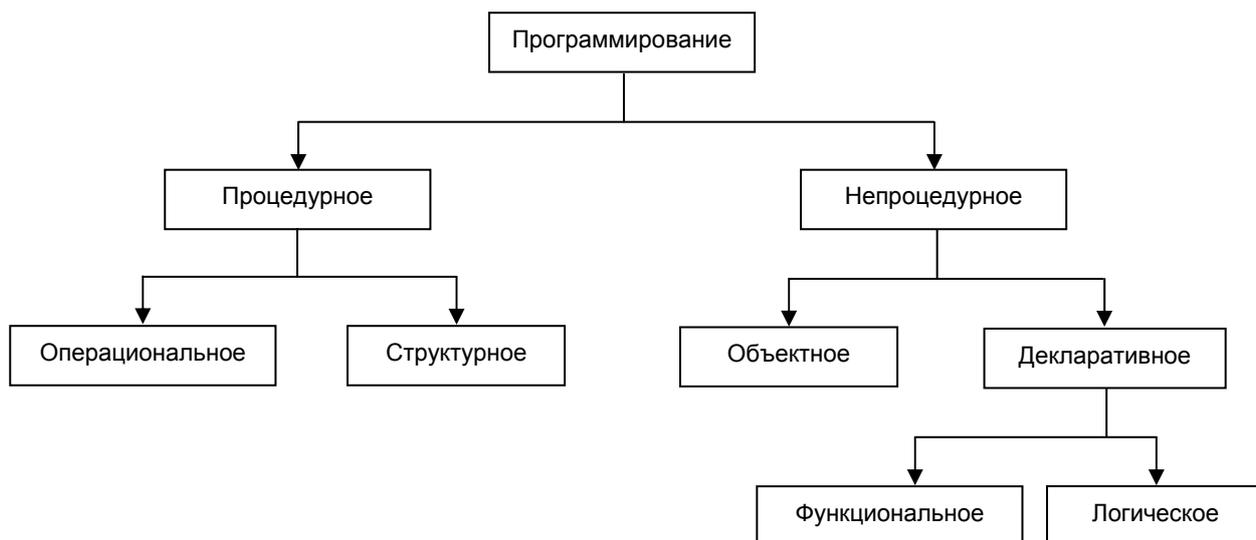
В основе структурного программирования лежит представление программы в виде иерархической структуры блоков, построенных из трёх типов базовых конструкций: последовательного выполнения, ветвления и цикла с предусловием.

В декларативном программировании особое внимание уделяется тому, что нужно сделать, а не тому, как это нужно сделать (в императивных языках). Здесь главным является точная формулировка задачи, а выбор и применение необходимого алгоритма для ее решения – проблема исполняющей системы, но не программиста.

Например, веб-страницы на языке HTML – декларативны, так как они описывают, что должна содержать страница, а не как отображать страницу на экране. Этот подход отличается от языков императивного программирования, требующих от программиста указывать алгоритм для исполнения.

Логическое программирование основано на математической логике. Как известно, самым распространённым языком логического программирования является ПРОЛОГ (Prolog).

Программа на языке ПРОЛОГ содержит две составные части: факты и правила. Факты представляют собой данные, с которыми оперирует программа, а совокупность фактов составляет базу данных ПРОЛОГа, которая фактически является реляционной базой данных. Основная операция, выполняемая над данными, – это операция сопоставления, называемая также операцией унификации или согласования.



Виды программирования

Объектно-ориентированное программирование – представление программы в виде совокупности объектов, каждый из которых является реализацией определенного класса, а классы образуют иерархию, основанную на принципах наследования.

Естественно, независимо от видов программирования, их реализация происходит через вычислительные средства.

Анализ пути развития основных языков программирования позволяет выделить следующие постоянно присутствующие, сменяющие друг друга тенденции [2]:

- смещение акцентов от частного (программирование деталей), к общему (программирование более крупных компонент);
- развитие и совершенствование инструментария программиста (языков программирования высокого уровня и рабочей среды);
- возрастание сложности программных и информационных систем.

Именно расширение области применения информационных технологий и вычислительной техники способствует развитию методов и инструментов построения программных систем.

Итак, на сегодняшний день программирование можно рассматривать и как науку, и как искусство, и как работу.

В качестве одного из его обобщенных определений можно предложить следующее: *программирование – это процесс отображения элементов предметной области на вычислительные средства.*

Если рассматривать программирование именно с позиций отображения, то его точность и полнота в значительной степени определяется инструментальными средствами разработчика и теми идеями, которые в них заложены.

Виды программирования соответствуют смене технологии программирования, что, практически, сродни смене парадигмы в науке.

Напомним, что *парадигма* – это исходная концептуальная схема, модель постановки проблем и их решения, методы исследования, господствующие в течение определенного исторического периода в научном сообществе. Самые большие революционные изменения в технологиях программирования произошли при переходе от парадигмы процедурно-ориентированного программирования (известной в истории как «разделяй и властвуй») к парадигме объектно-ориентированного программирования («все мое ношу с собой») [3].

Технология объектно-ориентированного программирования (ООП) позволяет проще формировать модели объектов реального мира. Сложная проблема раскладывается на составные части, и каждая составляющая становится самостоятельным *объектом*. С точки зрения системологии объект определяется структурой и поведением, т.е. содержит *свои* ресурсы и механизмы работы *именно* с этими ресурсами и все это имеет логически законченный вид.

Вспомним, что в программировании под объектом понимается объединение данных и процедур, их обрабатывающих, в одно целое.

Для дальнейшего рассмотрения выделим следующие три основных свойства объектно-ориентированного программирования:

1. Инкапсуляция (защита данных).
2. Наследование.
3. Полиморфизм.

Инкапсуляция обеспечивает высокую живучесть объекта и его специализацию. В программировании инкапсуляция – это механизм, который объединяет данные и код, манипулирующий с этими данными, а также защищает и то и другое от внешнего вмешательства. Целью инкапсуляции является сокрытие данных от несанкционированного доступа и изменения, а работа с данными ведется только через специально разработанные для этой цели процедуры.

Если рассмотреть это определение применительно к предметным областям, касающихся организаций, то в большинстве случаев в организациях, в первую очередь, подлежит сокрытию информация (ресурс), а не технология работы с ней.

Огромные средства тратятся на сохранение коммерческой тайны, а других решений, кроме как спрятать информацию и разрешить доступ к ней ограниченному кругу лиц из числа сотрудников компании, пока не придумано. А это и есть, по сути, инкапсуляция.

Как пример, каждое государство имеет свои ресурсы и процедуры для работы с этими ресурсами в виде министерств. Если стороннее государство захочет напрямую воспользоваться ресурсами другого государства, то это ему не удастся. В этом случае ему придется контактировать со специализированным министерством, которое имеет доступ к соответствующему ресурсу.

Если рассмотреть такие ресурсы как нефть и газ, то доступ к ним иностранных компаний осуществляется через специализированный механизм – «Газпром», являющийся интерфейсом между внешним миром и сокрытыми ресурсами государства.

Любой объект имеет определенную дисциплину доступа к своим ресурсам. Точно также в объектно-ориентированном программировании есть три уровня доступа – *private*, *protected* и *public*.

Так, у каждого человека есть информация личного характера, которую он ни при каких обстоятельствах никому не расскажет – это соответствует уровню *private*. У нас в стране даже принят закон о защите личной информации.

Второй уровень закрытости информации можно наблюдать в семейных преданиях, где доступ к информации имеет определенный круг лиц – члены семьи, и за пределы этого круга информация не уходит. Такой уровень называется *protected* – защищенный, он обеспечивает передачу информации по родственной иерархии, но защищает от доступа к информации извне. Если рассматривать этот уровень с точки зрения предприятия, то, например, информация, хранящаяся на головном предприятии, также доступна и дочерним отделениям, а со стороны доступ к такой информации полностью закрыт.

Третий уровень закрытости информации, как таковой, информацию не скрывает – это уровень *public*. Наиболее очевидный пример – это пресса, публичные издания. Все, что написано в газетах, доступно любому человеку без каких-либо ограничений. К ресурсам с модификатором доступа *public* можно отнести также магазины, библиотеки, школы.

Аналогично, объекты в объектно-ориентированном программировании, по умолчанию, имеют установленный уровень закрытости, который при необходимости можно изменять. Возвращаясь от абстракций к реальности, например, предприятие по выпуску гражданской продукции соответствует уровню *public*. Но, если какой-то цех перепрофилируется на производство спецпродукции, то его переводят на закрытый пропускной режим и уровень закрытости становится *private*.

Также возможно движение и в обратную сторону, когда, например, закрытые оборонные заводы, в рамках конверсии, начинают производить гражданскую продукцию и с уровня закрытости – *private* они переходят на уровень открытости – *public*.

Возникает вопрос: «А как из внешнего мира получить доступ к закрытым ресурсам?». Здесь надо иметь в виду, что объект, кроме ресурсов, имеет еще и механизмы для работы именно с этими ресурсами. Эти механизмы делают открытыми – *public*.

Например, коммерческая информация о деятельности предприятия – *private*, а структура предприятия (его отделы) – *public* – через эти отделы можно получить информацию о предприятии. На проходные большие предприятия вывешивают список отделов и их телефоны. Если не будет процедур, имеющих доступ к сокрытым данным, то эти данные останутся невостребованными (никто не сможет с ними работать).

Бухгалтерская информация закрыта и лица, не являющиеся сотрудниками бухгалтерии, не имеют к ней доступа.

Если нужно получить какую-то справку, например, о зарплате за прошлый год, то уволенные сотрудники не могут взять зарплатные ведомости и сделать себе справку – они не будут допущены к данным. Для решения этой задачи используется открытая функция – «Бухгалтер», которая из внешнего мира получает запрос, обслуживая его, обращается, как инкапсулированный механизм, к закрытым бухгалтерским документам и результат передает во внешний мир. Даже на самых закрытых предприятиях существуют подразделения, отвечающие за связь с внешним миром, например, отделы кадров.

Рассмотрим такой ресурс на предприятии как «Директор». Если доступ к ресурсу будет открыт – *public*, то все сотрудники предприятия по любым вопросам будут обращаться к «Директору», и система управления предприятием будет парализована. Если ресурс полностью закрыт – *private*, то к нему нельзя будет обратиться. Как решается такая проблема?

Классическое решение следующее. «Директор» имеет уровень закрытости – *private* и просто так к нему не войти. Нужно ждать, когда «Директор» освободится и ослабит уровень закрытости до *protected* и подчиненные смогут к нему зайти и подписать свои бумаги. Такое решение неэффективно. В соответствии с принципом инкапсуляции относительно объекта «Приемная Директора» для доступа к закрытому ресурсу «Директор» добавляется открытая функция «Секретарь». «Секретарь» контактирует с внешним миром, собирает документы для подписи, а потом, как член объекта, может обратиться к закрытому для других ресурсов «Директору».

Кроме того, в особые дни, например, раз в две недели, «Директор» становится открытым для всех своих сотрудников и принимает их по личным вопросам напрямую без «Секретаря».

В общем случае механизмы (процедуры) для работы с закрытыми данными являются открытыми. Но возникают ситуации, когда инкапсулированная процедура должна быть закрыта – *private*.

Например, служба внутренней безопасности предприятия занимается поведением сотрудников своего предприятия, и делать эту службу открытой для внешнего мира не имеет смысла.

Все здесь перечисленное имеет отношение к расширению первого принципа объектно-ориентированного программирования – инкапсуляции.

Вторым принципом объектно-ориентированного программирования является наследование, позволяющее одним объектам приобретать атрибуты и поведение других объектов, а также резко снизить затраты на получение результата. В программировании наследование помогает сделать разработку более экономной и обозримой, так как объекты пользуются одними и теми же атрибутами, и формами поведения без дублирования реализующих их программных кодов. Если рассматривать наследование в более широком смысле, то, например, можно долго и тяжело трудиться, чтобы накопить определенную сумму, а можно поступить проще – взять деньги у родителей, то есть воспользоваться сделанными ранее наработками.

Рассмотрим, как пример, классификацию животных в зоологии. Предком всех млекопитающих было крысоподобное животное, потомки которого, накапливая свои характерные признаки, дали все видовое разнообразие животного мира. От дикого предка пошла целая иерархия потомков, причем, каждый из них наследовал доступ к характеристикам и свойствам своих предков, но, при этом, имел что-то свое, чем он отличался от предков.

Так, из утверждения о том, что тигр относится к семейству кошачьих, вытекает сразу же несколько положений. Из этого утверждения следует, что у тигров хорошо развиты слух и обоняние, поскольку таковы характеристики кошачьих. Так как кошачьи входят в отряд хищных, это утверждение говорит еще о том, что тигры питаются мясом. Поскольку хищные являются позвоночными, мы узнаем и то, что у тигров есть позвоночник.

Вернемся к объектно-ориентированному программированию – процессу разработки иерархии объектов. При наследовании один объект может сохранять основные свойства другого объекта (предка) и добавлять к ним черты, характерные только для него. Целью наследования является использование существующих наработок. Наследование – процесс внесения уточнений.

И, наконец, третий принцип объектно-ориентированного программирования – полиморфизм (в переводе с греческого – «многообразие форм»).

Здесь, как пример, рассмотрим рулевое колесо автомобиля. Оно работает одинаково, независимо от того, используется ли при этом электропривод, механический привод или стандартное ручное управление. Главным здесь является то, что рулевое колесо везде одно и то же, а механизмы, выполняющие поворот, в каждом случае свои. Одно и то же имя используется для решения схожих, но технически разных задач.

Присвоение действию одного имени, которое совместно используется объектами всей иерархии, причем каждый объект иерархии применяет это действие своим собственным подходящим для него образом,

называется полиморфизмом. Целью полиморфизма, применительно к ООП, является наличие одного имени для задания близких по сути, но разных по реализации действий.

В общем случае концепцией полиморфизма является идея "один интерфейс и множество методов", т.е. создается общий интерфейс для группы близких по смыслу действий.

Например, пульта дистанционного управления телевизором внешне выглядят почти одинаково, но внутренняя реализация у каждой фирмы своя; анатомически мы все устроены одинаково, но каждый орган (процедура) у каждого из нас работает по-своему.

Размышляя о философии объектно-ориентированного программирования, которая органично проявилась из наблюдений и анализа реальности, нельзя не отметить две фундаментальные разновидности функций, инкапсулированных в объект – статические и виртуальные.

Статические функции характеризуются тем, что их действия описываются заранее до момента, когда в них возникает необходимость.

Например, в спектакле слова и действия каждого актера заранее прописаны в сценарии и отклоняться от сценария нельзя.

Врачи перед операцией собирают как можно больше информации о больном, составляют план операции, готовят необходимые инструменты и медикаменты и очень хорошо, если все прошло именно так, как заранее планировалось.

Проектная документация при строительстве объекта – вся смета составлена заранее.

Есть и другой подход к решению проблемы, и с ним мы часто сталкиваемся в ситуации – «когда я упрусь в проблему, тогда я ее и решу». У этого подхода есть объективная основа.

Часто возникают ситуации, когда понятно, что данная проблема обязательно будет решена, но как именно она решится – остается неизвестным до прояснения ситуации на месте. Такие функции называются *виртуальными*, т. е. якобы существующими.

Например, пожарные едут тушить пожар, но как именно они будут это делать, заранее никто сказать не может. Решение будет принято на месте после оценки ситуации.

Врач, который принимает больных, окажет помощь каждому пациенту, но что именно делать врач определит после беседы с конкретным пациентом.

Разница между вызовом статических и виртуальных методов – это разница между решением, принятым сейчас, и решением, принятие которого откладывается до прояснения ситуации в последний момент.

Хотелось бы еще отметить, что в объектно-ориентированном программировании, фактически, речь идет о прогрессирующей организации информации соответственно первичным семантическим критериям [4]: «важное/неважное», «ключевое/подробности», «родительское/дочернее», «единое/множественное».

Прогрессирование, в частности, на последнем этапе дает возможность перехода на следующий уровень детализации, что замыкает общий процесс.

Таким образом обычный человеческий язык, в целом, отражает идеологию ООП, начиная с инкапсуляции представления о предмете в виде его имени и заканчивая полиморфизмом использования слова в переносном смысле, что, в итоге, развивает выраженные представления через имя предмета до полноценного понятия-класса.

Итак, как мы видим, методология ООП позволяет довольно точно отображать объекты реального мира в различных предметных областях и системах – автоматических (чисто технических) и автоматизированных (человеко-машинных), и позволяет также отображать объекты реального мира, непосредственно связанные с *информационными процессами и информационными системами*.

В наиболее общем виде информационный процесс определяется как совокупность последовательных действий (операций, процедур, механизмов), производимых над информацией для получения какого-либо результата (достижения цели). Следует обратить внимание на то, что информационный процесс всегда протекает в какой-либо информационной системе – биологической, социальной, технической, социотехнической.

Методология объектно-ориентированного программирования при отображении объектов реального мира позволяет с высокой точностью показывать многие реальные процессы и системы.

СПИСОК ЛИТЕРАТУРЫ

1. Вайсфельд М. Объектно-ориентированное мышление. – СПб: Питер, 2014. – 304 с. – (Серия «Библиотека программиста»).
2. Прата С. Язык программирования С++. Лекции и упражнения, 6-е изд.: пер. с англ. – М.: ООО «И.Д. Вильямс», 2012. – 1248 с.
3. Павловская Т.А. С/С++. Процедурное и объектно-ориентированное программирование: учебник для вузов. Стандарт 3-го поколения. – СПб: Питер, 2015. – 496 с. – (Серия «Учебники для вузов»).
4. Мейер Б. Почувствуй класс: пер. с англ. / под ред. В.А. Биллига. – М.: Национальный Открытый Университет «ИНТУИТ»: БИНОМ. Лаборатория знаний, 2011. – 775 с.

Материал поступил в редакцию 12.02.18.

Сведения об авторах

ДЗЮБЕНКО Алла Леонидовна – кандидат технических наук, доцент, Финансовый университет при правительстве РФ, Москва
e-mail: al_dz@list.ru

ЧЕЧИКОВ Юрий Борисович – кандидат физико-математических наук, доцент, Финансовый университет при правительстве РФ; доцент, Московский авиационный институт
e-mail: yourych@mail.ru

ВНИМАНИЮ ПОДПИСЧИКОВ!

С 2018 года возобновляется издание информационного бюллетеня «Иностранная печать об экономическом, научно-техническом и военном потенциале государств-участников СНГ и технических средствах его выявления» серии «Экономический и научно-технический потенциал» (56741) взамен информационного бюллетеня «Экономика и управление»

Периодичность выхода – 12 номеров в год. Объем 48 уч.-изд. л. в год.

В бюллетене освещаются материалы иностранной печати по широкому спектру вопросов, касающихся сфер экономического и научно-технического развития России и стран СНГ: общие вопросы, финансы, промышленность, рынки, сельское хозяйство, космос, транспорт и связь, природные ресурсы, трудовые ресурсы, внешние торгово-экономические и научные связи

Оформить подписку на информационный бюллетень, начиная с любого номера, можно в ВИНТИ РАН по адресу: 125190, Россия, Москва, ул. Усиевича, 20,

Телефоны: (499) 151-78-61; (499) 155-42-85

Факс: (499) 943-00-60;

E-mail: contact@viniti.ru; sales@viniti.ru