

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

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

Издается с 1961 г.

№ 11

Москва 2017

ОБЩИЙ РАЗДЕЛ

УДК 004.27 : 004.6.056

М.Р. Биктимиров, А.В. Домашев, П.А. Черкашин, А.Ю. Щербаков

Блокчейн: универсальная структура и требования*

Обсуждается проблема разработки проекта архитектуры и математической модели распределенного хранения данных в связанной цепочке блоков (блокчейн), применение блокчейна для реализации различных актуальных информационных технологий, а также требования к нему, вытекающие из его системно-аналитических свойств.

Ключевые слова: блокчейн, распределенное хранение данных, информационная безопасность, электронное голосование, аукцион, шифрование, электронная подпись

ВВЕДЕНИЕ

В настоящее время вокруг весьма перспективной и актуальной модели распределенного хранения данных в связанной цепочке (блокчейна) [1] ведутся широкие научные и практические дискуссии, вклю-

чающие, к сожалению, множество недопониманий и спекуляций. Цель настоящей работы – разъяснить основные свойства этой технологии для обеспечения взвешенности решений о ее выборе и использовании как в корпоративных, так и в государственных информационных системах.

Как следует из названия технологии *blockchain* – «цепь» или «цепочка» блоков, блокчейн в первую очередь должен обеспечивать свойства цепи – неразрывность и прочность, которые являются парафразом свойства целостности.

Неразрывность определяется как свойство следования блоков (звеньев цепи) одного за другим в за-

* Публикация подготовлена в рамках исследований рабочей группы «Проблемы кибербезопасности» ФАНО России, а также в рамках работ по программе фундаментальных исследований Отделения математических наук РАН «Алгебраические и комбинаторные методы математической кибернетики и информационные системы нового поколения» и работ, поддержанных РФФИ, грант № 15-07-08522.

данной в процессе создания блокчейна последовательности, а прочность – как невозможность замены или удаления звена из цепочки.

Если рассматривать блокчейн как системную целостность, то он должен состоять из отдельных элементов – звеньев, каждое из которых в свою очередь делится на элементарные компоненты (назовем их атомами блокчейна). Атомы как элементы, ассоциированные или встроенные в компьютерную систему, могут быть пассивными и активными.

СТРУКТУРА БЛОКЧЕЙНА КАК СИСТЕМНОЙ ЦЕЛОСТНОСТИ

Атомы характеризуются идентификаторами (именами) и имеют некоторые свойства.

Выделим следующие типы атомов блокчейна:

1) атом-граница (начало или конец) блокчейна содержит ссылки на окончание предыдущего звена или на начало следующего;

2) атом-структура – определяет и описывает вектор идентификаторов атомов, имеющихся в звене, и логично следует за атомом-границей-началом;

3) атом-данные – содержит описание данных, интерпретируемых как пассивный компонент звена, и имеет разновидности:

- атом-открытые данные – характеризуется длиной и содержит данные, не подвергающиеся никакой обработке,

- атом-целостные данные – характеризуется длиной данных фиксированной целостности и идентификатором или ссылкой на процедуру контроля целостности данных,

- атом-подписанные данные – 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 данные – характеризуется длиной зашифрованных данных, идентификатором или ссылкой на процедуру шифрова-

ния/расшифрования данных, а также на ключ шифрования 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) атом-субъекты – содержит описание данных, интерпретируемых как активный компонент звена.

Атом-субъект может быть:

- атомом-сценарием, содержащим интерпретируемый код для выполнения операций над атомами-данными,

- 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-эксекютора.

Примеры применения блокчейна для актуальных технологий

Приведем пример звена блокчейна для регистрационных процедур (включая голосование).

Предположим, что происходит голосование N участников ($N=1, 2, \dots, i, \dots, N$), голосование может быть сложным, его результатом может быть не только бинарное «да» или «нет», но и произвольная строка, например, ФИО кандидата (строго говоря, «да» и «нет», равно как и «воздержался» тоже являются текстовыми строками).

Итак, U_i – участник голосования, G_i – его голос (в общем случае текстовая строка), X_i – секретный идентификатор, выработанный персонально U_i с использованием датчика случайных чисел и никому не известный, $[X_i]$ – секретный идентификатор, усеченный таким образом, чтобы обеспечить заданную трудоемкость перебора (например, не менее суток).

Участник голосования U_i формирует звено блокчейна Z_i (табл.1).

Таблица 1

Идентификатор	Название атома	Содержание	Примечание
1	Атом-граница-начало	Ссылка на звено Z_{i-1}	
2	Атом-структура		
3	Атом-открытые данные	U_i	
4	Атом-зашифрованные данные	$E(G_i, X_i)$	Зашифрованный голос участника, раскрыть который может только участник
5	Атом- зашифрованные данные	$E(G_i, [X_i])$	Зашифрованный голос участника, раскрыть который может устроитель голосования путем перебора заданной трудоемкости
6	Атом-хеш	$H([X_i])$	Информация для правильного выбора перебираемого ключа
7	Атом-подпись	Электронная подпись устроителя голосования под полями 1-6 и 8	Информация для фиксации неизменности звена
8	Атом-граница-окончание	Ссылка на звено Z_{i+1}	

При поступлении голосов устроитель голосования формирует и подписывает своей электронной подписью звено Z_i блокчейна, а затем при подведении итогов выполняет последовательный перебор ключей $[X_i]$, руководствуясь содержанием полей (атомов) 4 и 5, фиксируя голос участника как результат расшифрования атома 4.

Этим достигается невозможность подтасовки голосования, давления на голосующего в реальном времени, поскольку до подведения итогов его голос неизвестен, а также неизменность последовательности отдачи голосов и достоверное архивное хранение протоколов голосования в виде звеньев блокчейна.

В качестве развития описанной процедуры возможно вместо усечения долговременного идентификатора X_i использовать независимое от него случайное число R_i , такое что $|R_i| = |[X_i]|$.

Криптографический смысл такой замены – снижение нагрузки на X_i в случае, если голосование проводится не один раз для данного участника. Тогда определение $[X_i]$, которое происходит для восстановления голоса, не повлияет на уменьшение множества перебора целого X_i .

Для включения в функциональные возможности универсального блокчейна возможности выполнять работу заданной трудоемкости и получать за нее вознаграждение, дополним атомы-данные атомом-шаблоном или атомом-«будущие данные». Этот атом может быть как субъектом, так и объектом.

Атом-шаблон содержит указание на место в звене, которое предлагается заполнить участнику системы, чтобы получить заданное вознаграждение.

Рассмотрим регистрационную процедуру, когда U_i – участник системы (владелец объекта) регистрирует объект M_i – в общем случае бинарные данные (электронный документ, графический файл), X_i – секретный идентификатор, выработанный персонально U_i с использованием датчика случайных чисел и никому не известный, $[R_i]$ – секретный идентификатор, созданный таким образом, чтобы обеспечить заданную трудоемкость перебора. Далее рассмотрим ситуацию, когда другой участник может раскрыть – подтвердить регистрацию и физическое существование объекта M_i и получить за это вознаграждение.

Участник регистрационной процедуры U_i формирует следующее звено блокчейна Z_i (табл. 2).

Таблица 2

Идентификатор	Название атома	Содержание	Примечание
1	Атом-граница-начало	Ссылка на звено Z_{i-1}	
2	Атом-структура		
3	Атом-открытые данные	U_i	
4	Атом-зашифрованные данные	$E(M_i, X_i)$	Зашифрованный объект регистрации, раскрыть который может только владелец
5	Атом- зашифрованные данные	$E(M_i, [R_i])$	Зашифрованный объект регистрации, раскрыть который может другой участник перебора заданной трудоемкости
6	Атом-хеш	$H([R_i])$	Информация для правильного выбора перебираемого ключа
7	Атом-хеш	$H(M_i)$	Информация для окончательной проверки расшифрования объекта
8	Атом-шаблон	Пространство для внесения M_i (в том случае, если планируется раскрыть документ) или указание на функциональность атома, если раскрытие не планируется	
9	Атом-шаблон	Данные или инструкции (объект или субъект), которые выполняются при раскрытии данных, в том числе процедура получения вознаграждения	
10	Атом-подпись	Электронная подпись владельца под полями 1-7, 9 и 12	Информация для фиксации неизменности звена, пока данные не раскрыты
11	Атом-подпись	Электронная подпись владельца под полями 1-9 и 12	Информация для фиксации неизменности звена, когда данные раскрыты и помещены в шаблон (владелец может это сделать, поскольку обладает объектом M_i)
12	Атом-граница-окончание	Ссылка на звено Z_{i+1}	

Такая технология может применяться в случае, если требуется сохранение конфиденциальности данных в течение заданного срока, а затем их публикация для общего доступа.

Требования к блокчейну

Можно выделить следующие группы требований к блокчейну:

1) структурные, касающиеся наличия в звеньях блокчейна тех или иных типов данных (атомов) для обеспечения работы заданных технологий. Дополнительно они могут включать требование глобальной связанности блокчейна – наличие звена (или нескольких звеньев), описывающего структуру блокчейна в целом или его подмножества;

2) организационные, связанные с национальным или международным криптографическим регулированием, предполагающие применение национальных, рекомендованных или сертифицированных криптографических средств для формирования и обработки атомов блокчейна. Кроме того, данная группа может включать требования, связанные с национальными или ведомственными нормативами в областях применения – налоговая сфера, избирательные технологии, корпоративный документооборот и т.д.;

3) технологические, связанные с надежностью хранения звеньев блокчейна (например, с использованием технологии [2]), что должно обеспечивать заданные регуляторами соответствующих отраслей, где используется блокчейн, параметры надежности хранения и доступности этих звеньев. Кроме того, технологические требования должны описывать требования к производительности операций со звеньями и предельные объемы их накопления и хранения;

4) доверительные, имеющие четкую структуру блокчейна, регламентированные технологии работы со звеньями, а также интерфейс для выполнения операций над звеньями. Для обеспечения высокого доверия все прикладные интерфейсы должны быть доступны в исходных кодах. Кроме того, технология может быть формально верифицирована с помощью математических моделей.

ВЫВОДЫ

Рассмотренные свойства технологии блокчейна, типовые структуры данных и требования могут стать основой для формирования требований и решений корпоративного блокчейна, а также быть использованы для обеспечения взвешенности решений о выборе и использовании этой технологии как в корпоративных, так и в государственных информационных системах.

СПИСОК ЛИТЕРАТУРЫ

1. Централизованные криптовалюты. – URL: geektimes.ru/company/waves/blog/289379/
2. Зайцев А.В., Гостев С.С., Черкашин П.А., Щербаков А.Ю. О технологии распределенного хранения конфиденциальной информации в центрах обработки данных общего назначения // Научно-техническая информация. Сер. 2. – 2017. – № 5. – С.11-13.

Материал поступил в редакцию 21.09.17

Сведения об авторах

БИКТИМИРОВ Марат Рамилевич – кандидат технических наук, ВРИО директора ВИНТИ РАН, Москва
e-mail: marat@ras.ru

ДОМАШЕВ Алексей Владимирович – начальник отдела ФГУП НТЦ «Атлас», Москва
e-mail: domix@stcnet.ru

ЧЕРКАШИН Павел Александрович – заместитель генерального директора по информационным технологиям, «Концерн ГРАНИТ», Москва
e-mail: cherkashin@granit-concern.ru

ЩЕРБАКОВ Андрей Юрьевич – доктор технических наук, профессор НИУ ВШЭ, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН, Москва
e-mail: x509@ras.ru

Построение системы атрибутивного поиска в системах управления учетными или идентификационными доменами при внедрении интеграционных решений

Рассматривается задача проектирования подсистемы атрибутивного поиска при интеграции IDM-системы с современными сложными системами (например, ERP-системами), имеющими в своей основе гранулированный доступ. При внедрении интеграционных решений с большим количеством ролей и пользователей контекстный поиск, предустановленный в IDM-системах, показывает свою неэффективность. Предложено и реализовано решение этой задачи путем изменения подхода к организации поиска ролей и их описания, а также представлен подход к определению оптимального количества атрибутов, необходимых для эффективного поиска.

Ключевые слова: IDM-система, роль, ролевая модель доступа, атрибутивный поиск, поиск

ВВЕДЕНИЕ

Рост количества информационных систем, вовлеченных в бизнес-процессы современных корпораций любой из отраслей бизнеса, делает все более актуальным вопрос контроля и управления доступом к информационным ресурсам компании.

Потребность в решениях такого рода обусловило появление специализированных систем класса IDM¹ [1], которые позволяют автоматизировать процессы управления учетными записями и идентификационными данными пользователей, а также обеспечивают автоматизацию процессов внутреннего согласования при предоставлении доступа к информационным ресурсам компании.

Внедрение IDM-решений обеспечивает эффективную реализацию многих бизнес-задач и позволяет в значительной степени оптимизировать управление ИТ-инфраструктурой компании. Однако это очень сложный процесс, требующий особого внимания. В процессе внедрения типового IDM-решения и его адаптации под нужды конкретной компании возникают различные проблемы.

Примером такой проблемы может быть ситуация, возникающая при интеграции IDM-системы с современными сложными системами (например, ERP-системами), имеющими в своей основе гранулированный доступ. Особенностью такой формы доступа является то, что он выдается не одной ролью, а набором фильтров [2]. Если каждый набор таких фильтров выделять в отдельную роль – а так и происходит при интеграции целевых систем с IDM-системой – это приводит к тому, что количество ролей, доступных для запроса, может неконтролируемо увеличиваться. При этом эксплуатация такой системы становится крайне сложной.

Одним из известных решений проблемы большого количества загружаемых в систему ролей является выделение 100–200 основных ролей, а оставшийся сравнительно небольшой процент заявок предполагается выполнять вручную. Однако не для каждой системы удастся выделить сравнительно небольшое количество основных ролей для удовлетворения потребностей пользователей. Часто для доступа необходимо видеть сразу все роли. С таким количеством ролей, контекстный поиск, предустановленный в IDM-системах, показывает свою неэффективность, что обуславливает необходимость разработки решения по организации эффективной работы в IDM-системе при интеграции с целевыми системами, имеющими большое количество ролей.

¹ IDM (Identiti Management System) – системы управления учетными или идентификационными доменами.

АНАЛИЗ IDM-СИСТЕМ

Вопросы информационной безопасности играют сегодня огромную роль в бизнесе, где информация стала и продуктом, и сырьем. По данным аналитического центра компании InfoWatch [3] по сравнению с прошлым годом число утечек информации в мире выросло на 22%, в России – на 73%. В 55% виновниками утечек информации были настоящие или бывшие сотрудники – 54% и 1% соответственно.

Управление доступом к информационным ресурсам является ключевой функцией обеспечения информационной безопасности. Эта задача в том или ином виде решается в каждой информационной системе, а также на уровне всей ИТ-инфраструктуры компании или организации в целом. Управление доступом на современном уровне предполагает обеспечение процессного подхода с четко детерминированной ролевой моделью.

Один из самых острых вопросов при наличии большой ИТ-инфраструктуры – управление доступом сотрудников (а также сторонних пользователей) информационными ресурсами компании. В ситуации, когда таких ресурсов много, расположены они в разных информационных системах, имеют различных владельцев, ручное управление доступом представляется очень трудоемким процессом. Это отнимает много времени как у административного персонала (кадровые службы, ИТ-службы, ИБ-службы и т. д.), так и у конечных пользователей информационных систем, которые критически долго ожидают доступ для выполнения рабочих задач [4].

Для решения этих задач применяются специализированные системы класса IDM. Аббревиатуру IDM (Identity Management) можно расшифровать как «система управления учетными или идентификационными данными». Первоначально системы этого класса именовались именно так и подразумевалось управление учетными записями. С развитием функционала IDM-систем стали появляться новые классы решения, включающие в себя более широкий функционал. К таким решениям относятся IAM-решения (Identity and Access Management), IAG-решения (Identity and Access Governance), IGA-решения (Identity Governance and Administration) [5].

Основным функционалом IDM-систем является централизованное управление учетными записями, правами на доступ к информационным ресурсам, паролями и другими атрибутами в различных информационных системах, что позволяет автоматизировать процессы управления правами доступа, снизить нагрузку на ИТ-подразделения и обеспечить более высокий уровень информационной безопасности.

Решения данного класса востребованы в основном крупными компаниями. В настоящее время на рынке широко представлены как российские, так и зарубежные IDM-системы. Российские IDM-системы представляют такие компании, как «Аванпост», «ТрастВерс» и Solar Security. Также достаточно молодым представителем систем этого класса является IDM-система компания IIDM, построенная на базе открытых платформ (IC: Предприятие, OpenIDM, OpenICF). Зарубежные IDM-системы, которые получили признание в нашей стране и имеют своих клиентов,

представлены компаниями IBM и Oracle. Также в России с недавнего времени появилась компания SailPoint, которая хорошо известна на Западе и является лидером магического квадранта Gartner.

Современные системы управления доступом обеспечивают функционирование, как минимум, следующих процессов жизненного цикла:

- управление учетными записями (создание / изменение / удаление / блокирование / разблокирование учетных записей, управление паролями, управление учетными данными);
- управление информационными ресурсами (декларирование / изменение / удаление / назначение прав по умолчанию);
- управление правами и полномочиями (назначение / изменение / отзыв / блокирование / разблокирование / аудит(сверка));
- формирование отчетности (контроль всей совокупности полномочий, расследование исторических операций и пр.).

Все IDM-системы осуществляют управление доступом на основе ролевой модели – Role Based Access Control (RBAC).

Проблемы при использовании IDM-систем в интеграционных решениях

Одной из основных работ в проектах внедрения IDM является интеграция с управляемыми системами. Прежде всего, на стороне систем должны быть внешние интерфейсы управления доступом (создание, блокирование учетной записи, назначение/отзыв прав и т.д.).

Если система самописная, есть большая вероятность, что таких интерфейсов у нее нет, и их придется разрабатывать самостоятельно. Для этого нужно заранее уточнить, справятся ли с задачей собственные специалисты или потребуются привлечение сторонних экспертов. Кроме интерфейсов, в информационных системах могут быть еще две проблемы: гранулированный доступ и не стандартизированные названия ролей.

Гранулированный доступ является особенностью больших и сложных систем, таких как ERP, он выдается не одной ролью, а набором фильтров. Если каждый фильтр выделять в отдельную роль, это приведет к тому, что количество ролей, доступных для запроса, разрастется до 50 тыс. Понятно, что пользоваться такой системой невозможно.

Одним из решений проблемы является выделение 100–200 типовых ролей, обычно это покрывает 99% всех потребностей, а оставшийся процент заявок выполняется вручную, либо при запросе нестандартного доступа сначала формируется новая роль, а затем уже проходит сам запрос.

Необходимо разработать решение для организации удобной работы пользователей в Системе Управления заявками. Одним из очевидных и эффективных решений проблемы является изменение подхода к поиску ролей, попадающих в результаты поиска, который должен позволить значительно сократить их количество. В современном мире известны различные типы поиска, один из которых – поиск с использованием выбираемых атрибутов.

Модель атрибутивного поиска для IDM-систем

Для определения набора атрибутов, позволяющих эффективно осуществлять поиск ролей в IDM-системе, предлагается реализовать атрибутивный поиск.

В рамках предлагаемого подхода реализовано построение модели атрибутивного поиска, состоящей из трех этапов.

1. Определение возможных описателей (значений атрибутов поиска).
2. Группировка описателей и задание имен возможным атрибутам.
3. Выбор оптимального набора атрибутов для поиска.

Введем следующие определения.

Определение 1. Будем называть ролью пользователя корпоративной информационной системы (КИС (Role)) – именованный объект в КИС, характеризующийся набором характеристик, который позволяет идентифицировать действия пользователя в системе.

Пусть в системе имеются R-ролей, тогда множество ролей можно представить:

$$Role = \{Role_k\}, \text{ где } k=1, \dots, R \quad (1)$$

Например, роли в КИС «Х» можно представить в виде:

$Role_1$ = КИС "Х" \Отв за согласование заявки на закупку. Деп. 1.

$Role_2$ = КИС "Х" \Отв за согласование заказа на поставку. Компания. БУХ. Рук. ЦО ур. 2.

$Role_3$ = КИС "Х" \Спец по потребности. Филиал 16. ИТ.

$Role_4$ = КИС "Х" \Спец по закупкам. Филиал 20. ВХР.

Заметим, что каждая роль характеризуется набором атрибутов.

Определение 2. Под атрибутом роли будем понимать свойство/признак, характеризующее роль. Таким образом, роль может характеризоваться одним атрибутом или группой атрибутов.

Обозначим $Attr = \{a_j\}$, множество атрибутов, где j – идентификатор/номер атрибута $j=1, \dots, K$

Например,

$Role_3 = \{a_1, a_2, a_3\}$,

где a_1 = «Название»,

a_2 = «Филиал»,

a_3 = «Группа закупок»,

Замечание 1. Если порядок следования атрибутов в описании роли – фиксированный и не меняется от роли к роли, то атрибут однозначно определяется своим идентификатором, т.е. атрибут «Филиал» всегда будет иметь идентификатор «2».

Определение 3. Назовем множество значений, которые может принимать атрибут a_j доменом атрибута роли.

Обозначим D_j – домен j-го атрибута. D_j – множество всех возможных значений, которые может принимать атрибут a_j . Тогда, $D_j = \{d_i^j\}$, где d_i^j – i-ое значение j-го атрибута, $i=1, \dots, N$.

Например,

для атрибута a_2 = «Филиал» домен

$D_2 = \{\text{Филиал 01, Филиал 02, ..., Филиал 16, ..., Филиал 20}\}$,

$d_j^i = \text{«Филиал j»}$, где $j=1, \dots, 20$

Замечание 2. Декартово произведение $D_1 \times D_2 \times \dots \times D_K$ – определит множество всевозможных значений атрибутов ролей в системе.

Определение 4. Назовем коэффициентом покрытия атрибутом S множества ролей $\{Role_k\}$ – P_s , долю ролей, которые описываются заданным значением атрибута S, в общем количестве ролей. Значение P_s определяется по формуле:

$$P_s = \frac{\sum_{i=1}^N P_{is}}{R} * 100\% \quad (2)$$

где:

S – идентификатор атрибута;

P_{is} – количество ролей, описываемых i-значением атрибута S;

R – общее количество ролей в исследуемой системе.

Рассмотрим пример определения коэффициента покрытия для атрибута «Филиал» ролей КИС «Х».

Пусть $R=3200$ – количество ролей в КИС «Х», идентификатор атрибута «Филиал» – 2, т. е. $S=2$, общее количество ролей, в которые входят значения атрибута «Филиал» из домена D_2 равно 1920, тогда

$$P_2 = \frac{\sum_{i=1}^{20} P_{i2}}{3200} * 100\% = 60\% .$$

Заметим, что формула (2) справедлива только в случае, если выполняется условие Замечания 1. Кроме того, если необходимо определить коэффициент покрытия множества ролей группой атрибутов, то формула (2) станет более громоздкой. Это происходит в случае семантической независимости атрибутов ролей.

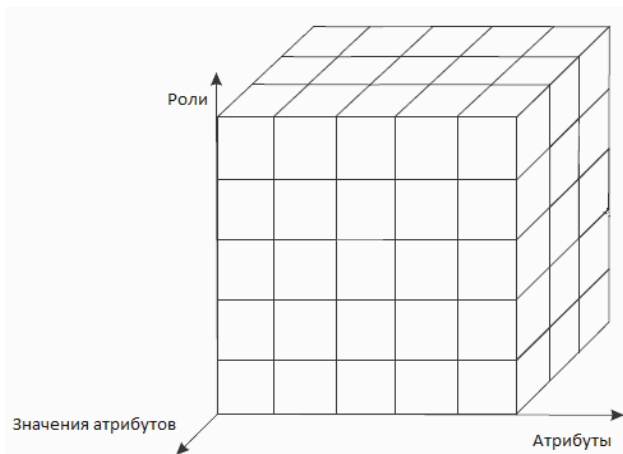
Далее список групп сортируется в порядке убывания значения P_s . В соответствии с сортировкой группе атрибутов присваивается ранг от 1 до N (количество выделенных групп).

Целью определения минимального набора атрибутов является выявление оптимального сочетания атрибутов, при котором достигается целевой показатель эффективности поиска. Этот показатель устанавливается заинтересованными сторонами, исходя из требований к выводу количеству ролей в результатах поиска.

В процессе определения набора атрибутов и их значений, описанных на предыдущем шаге, возможно определение значительного количества описательных групп атрибутов. Большая часть выявленных атрибутов может не оказывать значительного влияния на качество поиска и станет избыточной для описания ролей.

Поиск с использованием атрибутов представляет собой пересечение множеств ролей, описанных выбранными значениями атрибутов.

Описание каждой роли можно представить в виде трехмерной таблицы (рисунок).



Трехмерное представление описания ролей

Описанием роли являются данные, находящиеся в горизонтальном сечении куба.

Роли, описанные определенным значением атрибута, составляют кортеж

$$A(i,j) = \{R_{1i}, R_{2i}, \dots, R_{Ni}\}, \quad (3)$$

где i – идентификатор значения атрибута; j – идентификатор атрибута;

R_{Ni} – роль, описанная i -значением атрибута.

Пример:

$A(1,1) = \{\text{Бухгалтер. Филиал 01.ИТ. Бухгалтер. Филиал 01.ПФМ}, \text{Куратор. Филиал 01}, \dots\}$

Результатом поиска с использованием атрибутов является пересечение множеств ролей, описанных значениями атрибутов, определяется:

$$\bigcap_{T=1}^{N_0} A_T(i,j) = \underbrace{A_1(i,j) \cap A_2(i,j) \cap \dots \cap A_T(i,j)}_{N_0}, \quad (4)$$

где $A(i,j)$ – множество ролей, описанных i -м значением j -го атрибута;

T – идентификатор количества пересечений;

k – введенный идентификатор каждого пересечения;

N_0 – количество возможных пересечений.

Количество ролей, описанных определенным атрибутом, определяется:

$$K = M(A(i,j)), \quad (5)$$

где M – мощность множества.

Соответственно количество ролей, попадающих в результаты поиска с использованием определенных атрибутов, определяются (6):

$$\varphi_f = M_f \left(\bigcap_{T=1}^{N_0} A_T(i,j) \right), \quad (6)$$

где f – идентификатор пересечения.

Достижение целевого показателя определяется из расчета среднего значения результатов поиска в соответствии с (7), (8) с определенными атрибутами и их количеством в сравнении с принятым.

$$F_{\text{сред}} = \frac{\sum_f \varphi_f}{c}, \quad (7)$$

где f – идентификатор количества множеств,

C – общее количество осуществлённых пересечений.

$$F_{\text{сред}} \leq \text{целевой показатель} \quad (8)$$

Для определения достижения целевого показателя используется метод перебора всех возможных значений.

В цикле происходит проверка достижения целевого показателя при осуществлении пересечения множества со следующими по значению ранга множествами. Если целевой показатель достигнут – атрибут признается эффективным, пересечения с множествами, соответствующими группе атрибутов следующего ранга, не производятся.

Если значение целевого показателя не достигнуто – берется следующий по рангу атрибут и проверяется поиск с использованием всех возможных сочетаний значений атрибутов.

На каждом этапе добавления в группы множеств в процессе поиска оптимальной группы атрибутов производится дополнительная проверка. Если прирост снижения среднего значения результатов поиска менее 15%, то применение последующих групп атрибутов считается малоэффективным.

$$F_{\text{сред}}(v) - F_{\text{среднее}}(v+1) < F * 15\%, \quad (9)$$

где $F_{\text{сред}}$ – среднее значение результата поиска,
 v – порядковый номер цикла вычисления;

ЗАКЛЮЧЕНИЕ

Стандартный функционал IDM-систем предполагает использование только одного типа поиска – полнотекстового, когда пользователь вводит словосочетание, слово или часть слова и результатом поиска становится значение, куда данная «строка» входит полностью. В системах, с количеством ролей до 200, данный метод не вызывает трудностей. Однако сложные системы, которые используются в крупных компаниях, чаще всего содержат тысячи ролей, необходимых пользователям. Поиск ролей пользователями становится не эффективным и неудобным. Для поиска каждой роли пользователю приходится затрачивать значительное количество времени. Так как IDM-системы не предполагают поиск с использованием атрибутов в своем стандартном функционале, то на проекте внедрения IDM-системы «КУБ» в крупной энергетической компании была использована предложенная модель атрибутивного поиска.

Внедряемая и используемая IDM-система интегрирована с несколькими десятками систем компании заказчика. Для организации работы с более чем 10 целевыми системами, пользователям требуется

доступ к нескольким тысячам ролей. Для проверки адекватности модели была выбрана одна из систем заказчика, где содержится более 3200 ролей, которые должны быть доступны для заказа пользователями и руководителями в Системе Управления Заявками. Целевой показатель поиска, обеспечивающий удобный вывод для пользователей, определен из возможности вывода результатов поиска ролей на одной или двух страницах в web-интерфейсе Системы управления заявками. Количество выводимых ролей на одной странице вывода результатов поиска, установленное настройками системы, – 10 штук. Целевой показатель определен – 20 ролей.

В проекте были реализованы доработки программного обеспечения в соответствии с разработанной моделью. Внедрение доработок показало работоспособность и эффективность предложенной модели.

СПИСОК ЛИТЕРАТУРЫ

1. Хонин А. Обзор IdM/IAM-систем на мировом и российском рынке. – URL: https://www.anti-malware.ru/reviews/IdM_IAM_russian_market_2015# (дата обращения: 09.05.2017).
2. Кудрин П. Внедрение системы IDM за 10 шагов. – URL: http://www.jetinfo.ru/jetinfo_arhiv/entsiklopediya-idm/10-shagov-dlya-effektivnogo-vnedreniya-idm-sistem/2014 (дата обращения: 17.04.2017).
3. Компания Infowatch. Глобальное исследование утечек конфиденциальной информации в 2014 году. – URL: <https://www.infowatch.ru/report2014> - 26 февраля 2015г. (дата обращения: 02.04.2017).
4. Бондарь Д. Роль IDM в современной информационной инфраструктуре. – URL: <http://www.itsec.ru/articles2/Oborandteh/rol-idm-v-sovremennoi-informacionnoi-infrastryktyre> (дата обращения: 12.03.2017).
5. Шестова А.Д. Подсистема атрибутивного поиска для IDM-системы // 72-е Дни науки студентов МИСиС: международные, межвузовские и институтские научно-технические конференции. – М.: НИТУ «МИСиС», 2017. – С. 517-518.

Материал поступил в редакцию 19.09.17.

Сведения об авторах

БЕЛОВ Александр Владимирович – руководитель департамента прикладной математики МИЭМ ВШЭ – Национальный исследовательский университет Высшей школы экономики, Москва
e-mail: AVBelov@hse.ru

НЕЖУРИНА Марина Игоревна – кандидат технических наук, доцент, директор Института информационных бизнес систем Национального исследовательского технологического университета «МИСиС» (ИИБС НИТУ «МИСиС»), Москва
e-mail: min@misis.ru

ШЕСТОВА Анастасия Дмитриевна – магистр ИИБС НИТУ «МИСиС», выпускница 2017 г.
e-mail: nastica87@gmail.com

Современные наукометрические методы определения научных лидеров: новые математические модели*

Предлагается возможный подход к многоуровневой оценке научных организаций с использованием количественных данных, получаемых в ходе мониторинга, и учитывающий разнообразие видов результатов исследований и разработок.

Ключевые слова: научно-техническая деятельность, исследования и разработки, источники данных, научные организации, результативность организации, референтские группы, оценка организаций

ВВЕДЕНИЕ

Развитие сферы науки, признанной во всем мире в качестве одного из основных источников новых знаний и технических решений, необходимых для развития экономики, требует все более масштабных вложений со стороны государства и бизнеса. По оценкам ЮНЕСКО, в 2014 г. валовые расходы на исследования и разработки (ИР) в мире превысили 1,5 трлн долл. (UNESCO, 2016). Согласно данным отечественной статистики, в России в 2016 г. объем внутренних затрат на ИР достиг 943,8 млрд руб. (37,3 млрд долл. США в расчете по паритету покупательной способности), что ставит ее на десятую строчку в общемировом рейтинге [1] по данному показателю. По величине бюджетных ассигнований на гражданскую науку наша страна занимает 6-е место в мире, по численности занятых в сфере исследований и разработок – 4-е.

Рост инвестиций в развитие науки и новых технологий, наблюдающийся во многих странах, очевидно, требует и повышения их отдачи. Одним из решений стало внедрение в практику государственного управления регулярного мониторинга и оценки деятельности организаций, выполняющих исследования и разработки. Системы оценивания включают комплекс исследовательских и аналитических мероприятий, механизмов и процедур, нацеленных на анализ ре-

зультативности научной деятельности, выявление существующих в научных организациях проблем и ограничений, выработку рекомендаций по их преодолению. В зависимости от целей и задач политики, конкретные инструменты и подходы к измерению и оценке приобретают разные масштабы и формы.

Так, в Великобритании оценка результативности и качества научных исследований касается университетов и опирается на три группы параметров, включающих соответственно научные результаты, их влияние на развитие экономики и общества и качество научной среды¹. В Германии основным объектом оценки являются научные общества, которые, наряду с аудитом входящих в их состав институтов, проводят внешнюю оценку, привлекая отечественных и зарубежных экспертов. При этом направления оценки касаются как делового и административно-финансового состояния организаций, так и результативности научно-технической деятельности (см., например, [2]).

Франция реализует систему двухуровневой оценки научных организаций, входящих в состав Национального центра научных исследований (Centre national de la recherche scientifique – CNRS) [3]. Это предполагает проведение анализа финансовой отчетности, обязательной для государственных организаций страны, а также академическую оценку, охватывающую следующие уровни (CNRS, 2015): тематические направления исследований, программ и проектов, институтов и лабораторий, отдельных ученых.

Одна из наиболее сложных – система оценки результативности научной деятельности в Японии. Она предполагает проведение внутренней оценки и аудита деятельности организаций, выполняющих ИР, и

* Статья подготовлена в результате проведения исследования по теме: «Формирование «сети знаний» по обмену лучшими практиками управления научными исследованиями и научно-техническими результатами на основе межведомственной системы мониторинга результативности научных и образовательных организаций» (уникальный идентификатор проекта: RFMEFI60216X0017), реализованного при поддержке Минобрнауки России в 2016-2017 гг.

¹ Подробнее: официальный вебсайт Research Evaluation Framework – www.ref.ak.uk, раздел Panel Criteria and Working

двух независимых внешних оценок: всех действующих в стране университетов и научных организаций, а также совокупности организаций, входящих в состав национальных университетских корпораций [4]. При этом проводится анализ количественных показателей деятельности организаций (число научных статей и монографий, поданных патентных заявок и полученных патентов на изобретения, грантов, совместных проектов и др.) и профессиональная экспертиза качества полученных результатов (их научная ценность, социальный, экономический и культурный эффект).

Свои оценки реализуют США, Канада, Италия, Швеция и другие страны. Во многих случаях внедрение таких систем причислено к институтам развития, способным внести серьезный вклад в рост и структурную диверсификацию экономики. При этом в связи с абсолютным увеличением государственных расходов, повышением требований к результативности исследований и разработок и ослаблением контроля над текущим использованием бюджетных средств, инструменты оценивания постоянно совершенствуются, на законодательном уровне вводятся стандарты качества и нормативы управления, отвечающие новым условиям.

Начало активных дискуссий о необходимости оценки научной деятельности и перехода на новую модель финансирования науки на основе показателей ее результативности, а также первые конкретные шаги по внедрению этой практики в российском научном поле относятся к середине 2000-х гг. В 2006-2007 гг. под руководством Минэкономразвития России федеральные органы исполнительной власти провели оценку подведомственных им государственных научных центров.

Важной вехой в формировании подхода к оценке научной деятельности в России стали разработка методологии расчета коэффициента персональной результативности научной деятельности (ПРНД)² и проведение первого этапа мониторинга на ее основе (2006-2007 гг.)³. Далее в течение 2009-2011 гг. были приняты ведомственные документы, утверждающие положения о комиссиях по оценке результативности деятельности научных организаций, методики оценки, административных регламентов и др. В этот и последующий период отдельные ведомства и государственные корпорации провели практическую оценку результативности деятельности в подведомственных научно-технических комплексах. Однако ее итоги не вызвали заметных институциональных изменений.

² Утверждена Совместным Приказом Минобрнауки России, Минздравсоцразвития России и Российской академии наук № 273/745/68 от 3 ноября 2006 г. «Об утверждении порядка и условий применения стимулирующих выплат, обеспечивающих повышение результативности деятельности научных работников и руководителей научных учреждений и научных работников научных центров Российской академии наук».

³ Данная работа в определенном смысле задавала основы системы национальной оценки, практическая реализация которой стартовала с подписанием Постановления Правительства Российской Федерации от 08.04.2009 г. №312 «Об оценке результативности деятельности научных организаций...»

Хотя результативность научной деятельности за эти годы была в среднем невысокой, а в некотором случае показывала негативную динамику, ведомства, как правило, не слишком критично оценивали деятельность подведомственных им организаций, не стремились улучшать структуру научной сети.

В 2013 г. было решено перейти к очередному этапу реализации мониторинга и оценивания (Постановление Правительства России от 1 ноября 2013 г. № 979). В частности, новым элементом стало формирование системы мониторинга (ежегодно) и оценки (один раз в пять лет) результативности деятельности научных организаций, выполняющих научные исследования и разработки гражданского назначения. Основными инструментами процедуры оценки было объединение научных организаций в сопоставимые референтные группы и установление для них минимальных значений показателей результативности. Структура референтных групп, состав основных и дополнительных показателей результативности деятельности в них был утвержден протоколом заседания межведомственной комиссии (МВК) по оценке результативности от 14.01.2016 г. № ДЛ-2/14пр, согласно которому оценка должна проводиться по 39 исследовательским направлениям и трем профилям научной деятельности, отражающим ориентацию на различные типы результатов. Предложенный подход, уже показавший в себя в действии, тем не менее поставил ряд вопросов, связанных с возможностью расчета отдельных показателей, наполнения референтных групп достаточным числом организаций, учетом результативности деятельности междисциплинарных организаций и др. Цель настоящей работы – представление альтернативной экспериментальной комбинаторной модели распределения организаций по референтным группам, учитывая профиль и направление их деятельности, а также введение алгоритма определения категорий результативности на основании количественных наукометрических показателей.

МЕТОДЫ АНАЛИЗА И ИСТОЧНИКИ ДАННЫХ

Информационной основой для проведения анализа являются сведения, содержащиеся в федеральной системе мониторинга результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы, по состоянию на 05 сентября 2017 г. На указанную дату в системе были представлены сведения о 1795 организациях, из которых 1502 имеют заполненные профили. Для обеспечения относительной однородности в изучаемую совокупность вошли только научные организации, имеющие ненулевые затраты на исследования и разработки, и ненулевую численность исследователей. Размер финальной выборки составил 1018 единиц.

Последующий алгоритм формирования референтных групп и определения категорий результативности состоял из следующих этапов.

1. *Объединение научных организаций в группы* по 12 основным направлениям исследований в соответствии с Единым научным рубрикатором, разработанным Комиссией общественного контроля в сфере

науки и гармонизированным на верхнем уровне с классификатором областей науки организации экономического развития и науки (ОЭСР). Объединение организаций велось экспертно на основании сведений о направлениях исследований, представленных научными организациями. Сокращение числа исследовательских направлений, с одной стороны, обусловлено примерами лучшей зарубежной практики (примером в данном случае выступал опыт Италии, описанный в работах [5, 6]), с другой – необходимостью повышения компактности и операциональности модели.

2. *Определение профилей деятельности организаций* (далее – профиль) в каждом из направлений, в зависимости от их ориентации на определенный вид результата, для выявления которых используются следующие показатели.

А. Число научных публикаций, индексируемых в Web of Science, в расчете на 100 исследователей, (ед).

Этот показатель отличается от предложенного в материалах Минобрнауки России, так как обеспечивает большую строгость и верифицируемость данных. Кроме того, он уравнивает шансы быть представленным в системе для всех организаций за счет того, что дальнейшие расчеты соотносятся с областью науки (направлением исследований) организации.

Б. Количество созданных результатов интеллектуальной деятельности, имеющих государственную регистрацию и (или) правовую охрану в Российской Федерации или за ее пределами, а также количество выпущенной конструкторской и технологической документации, в расчете на 100 исследователей, (ед).

В. Объем выполненных работ, оказанных услуг (исследования и разработки, научно-технические услуги, доходы от использования результатов интеллектуальной деятельности) к общей численности работников, выполнявших исследования и разработки (тыс. руб.).

Для снижения эффекта масштаба каждый из показателей (А-В), отражающий определенный тип результатов, нормировался делением на количество сотрудников, преимущественно участвующих в его получении. В то время как исследователи, как правило, являются основными «производителями» научных публикаций и разработчиками технологий, финансовые результаты требуют привлечения различных категорий научного персонала.

Организация относилась к тому или иному профилю, если значение показателя (А – В) выражено, т. е. не равно нулю и не ниже соответствующего медианного значения по организациям, также относящимся к заданному направлению исследований. Подобный подход позволил идентифицировать четыре профиля:

- I. «Генерация знаний» (выражен показатель А);
- II. «Разработка технологий» (выражен показатель Б);
- III. «Научно-технические услуги» (выражен показатель В);
- IV. «Особый» (не выражен ни один из показателей А-В).

Пересечение профиля исследования, взятого для каждого из 12 исследовательских направлений, формирует референтную группу. При этом допускается, что одна организация может иметь от одного до трех

профилей деятельности, т.е. присутствовать в нескольких референтных группах, оставаясь при этом в рамках одной области исследования.

3. *Минимальные значения показателей результативности* в референтных группах рассчитываются как медианные значения соответствующих усредненных за пять лет показателей результативности (усредненным является показатель, определяющийся отношением соответствующих суммарных за пять лет числителей и знаменателей) для научных организаций данной референтной группы. В представленных экспериментальных расчетах использовались данные за один (2016) год.

4. *Категория результативности организации в референтной группе* устанавливается по следующему правилу:

- организация в референтной группе относится к 1-й категории результативности, если значение выраженного показателя для ее профиля не равно нулю и не ниже медианного значения соответствующего показателя по соответствующему направлению науки, увеличенного на 25%;
- организация в референтной группе относится к 3-й категории результативности, если значение выраженного показателя для ее профиля не выше медианного значения соответствующего показателя по соответствующему направлению науки, уменьшенного на 25%;
- организация в референтной группе относится к 2-й категории результативности, если она не относится ни к 1-й, ни к 3-й категории результативности в данной референтной группе.

Подобный подход позволяет присваивать организации различные категории результативности по всем ее профилям в рамках одного направления. В результате можно судить, по какому именно профилю деятельности (генерация знаний, разработка технологий или научно-технические услуги) организация лидирует, где она демонстрирует удовлетворительную научную результативность, а где утратила научную деятельность в качестве основной.

Руководствуясь описанными выше правилами были проведены тестовые расчеты. Результаты представлены в табл. 1 и 2.

РЕЗУЛЬТАТЫ ТЕСТОВЫХ РАСЧЁТОВ

Достигнутая выборка позволила идентифицировать четыре профиля исследований в 12 областях науки, что обеспечивает достаточную наполненность референтных групп (исключение составляет группа прочих естественных и точных наук (табл. 1). Каждый профиль исследования характеризуется уникальной конфигурацией ресурсов и результатов, для которых используются различные критерии оценки и пороговые значения.

Ряд профилей оказывается практически вырожденным. Например, для гуманитарных наук практически отсутствуют организации, имеющие охраняемые результаты интеллектуальной деятельности. Для технических, медицинских и сельскохозяйственных наук, напротив, эта группа довольно хорошо выражена. Подобные диспропорции можно увидеть и в других областях.

Результаты моделирования состава референтных групп

Область науки	Категория	Всего	в том числе по профилям			
			I	II	III	IV
Медицинские науки и общественное здравоохранение	Всего	158	79	79	79	20
	1-я категория		28	17	17	
	2-я категория		41	51	58	
	3-я категория		10	11	4	
Сельскохозяйственные науки	Всего	232	97	116	116	44
	1-я категория		30	25	29	
	2-я категория		54	77	85	
	3-я категория		13	14	2	
Социальные науки	Всего	80	40	30	40	13
	1-я категория		11	7	8	
	2-я категория		26	17	30	
	3-я категория		3	6	2	
Техника и технологии	Всего	202	101	101	101	25
	1-я категория		31	22	25	
	2-я категория		58	66	72	
	3-я категория		12	13	4	
Гуманитарные науки	Всего	38	19	16	19	7
	1-я категория		7	1	5	
	2-я категория		11	13	14	
	3-я категория		1	2	0	
Биологические науки	Всего	184	92	92	92	26
	1-я категория		31	20	21	
	2-я категория		53	64	65	
	3-я категория		8	8	6	
Компьютерные и информационные науки	Всего	61	31	31	31	9
	1-я категория		13	7	7	
	2-я категория		10	17	21	
	3-я категория		8	7	3	
Математика	Всего	25	13	13	13	4
	1-я категория		3	4	2	
	2-я категория		9	9	10	
	3-я категория		1	0	1	
Науки о Земле и смежные экологические науки	Всего	117	59	59	59	16
	1-я категория		17	15	12	
	2-я категория		38	35	45	
	3-я категория		4	9	2	
Прочие естественные и точные науки	Всего	16	8	8	8	3
	1-я категория		2	4	2	
	2-я категория		4	1	5	
	3-я категория		2	3	1	
Физика и астрономия	Всего	58	29	29	29	6
	1-я категория		9	9	8	
	2-я категория		17	17	20	
	3-я категория		3	3	1	
Химические науки	Всего	37	19	19	19	3
	1-я категория		5	5	4	
	2-я категория		12	10	15	
	3-я категория		2	4	0	

Таблица 2

Медианные значения показателей результативности в референтных группах

Область науки	В целом			в том числе по профилям											
	А	Б	В	А				Б				В			
				I	II	III	IV	I	II	III	IV	I	II	III	IV
0	9,09	6,55	585,64												
1	21,54	6,62	743,32	41,79	21,85	21,43	3,78	6,67	17,43	6,12	2,54	740,01	740,80	1 173,13	426,76
2	0,00	8,09	488,30	5,45	0,00	1,37	0,00	8,11	20,80	8,92	0,00	550,47	512,26	676,34	336,40
3	4,58	0,00	763,62	11,42	4,82	4,77	0,00	0,00	9,39	0,00	0,00	805,91	747,10	1 149,30	537,70
4	2,57	12,11	922,52	24,71	2,76	1,57	0,00	12,24	33,33	16,87	1,66	904,18	1 078,62	1 764,26	338,52
5	8,74	0,00	655,48	18,12	8,78	13,92	0,00	0,00	10,49	1,00	0,00	760,11	768,15	872,03	460,75
6	14,93	4,07	730,12	44,29	15,45	26,54	1,58	4,00	11,82	3,56	1,67	855,05	704,69	1 109,35	453,10
7	7,43	4,08	1 018,97	59,50	34,69	9,51	0,00	8,55	13,46	4,00	0,00	1 018,97	952,70	1 517,46	415,28
8	63,01	9,68	918,10	98,73	75,71	72,73	6,48	10,39	13,46	9,94	3,36	1 147,40	1 034,03	1 321,03	476,48
9	18,93	3,38	832,39	43,33	15,00	22,86	5,52	2,51	9,94	4,17	0,00	903,77	1 034,69	1 371,91	545,43
10	5,13	7,37	1 257,30	13,43	9,18	11,76	0,00	33,41	41,49	6,65	2,27	1 402,10	1 257,30	2 401,11	924,11
11	41,85	6,73	1 057,55	80,04	26,98	30,72	5,00	4,35	19,29	10,78	2,85	1 034,69	1 221,67	1 946,73	580,32
12	42,59	6,25	994,18	94,39	34,86	65,24	13,51	5,78	15,32	5,78	4,05	1 079,85	943,45	1 222,34	674,64

Условные обозначения областей науки:

- 0 – По всем организациям
 1 – Медицинские науки и общественное здравоохранение
 2 – Сельскохозяйственные науки
 3 – Социальные науки
 4 – Техника и технологии
 5 – Гуманитарные науки

- 6 – Биологические науки
 7 – Компьютерные и информационные науки
 8 – Математика
 9 – Науки о Земле и смежные экологические науки
 10 – Прочие естественные и точные науки
 11 – Физика и астрономия
 12 – Химические науки

Как видно из табл. 2, медианные значения показателей ожидаемо выше для организаций тех профилей, в которых они (показатели) являются определяющими. Кроме того, как и в случае с наполненностью референтных групп, по областям науки наблюдаются различия. Так, наиболее высокий уровень публикационной активности в расчете на 100 исследователей (показатель А, определяющий профиль «генерация знаний») зафиксирован в математике, наименьший – в сельскохозяйственных науках. Медианные значения показателя Б, характеризующего, прежде всего, результативность прикладной науки, наиболее высок для организаций технического профиля, а также точных наук (высокое значение показателя в последнем случае также может быть обусловлено малой наполненностью этого направления). Наконец, показатель В, определяющий общий объем выполненных исследований и разработок и научно-технических услуг, наиболее высок в области физики, а также в других естественных и точных науках.

ВЫВОДЫ

Предлагаемый подход для формирования референтных групп и общей количественной оценки результативности научной и научно-технической деятельности может быть легко воспроизведен для разных выборок при условии обеспечения их однородности по типу организаций, а также возможности четкого определения направлений исследований (на среднем и высоком уровне агрегирования). Кроме того, модель позволяет учитывать различные виды результатов и выделять несколько типов лидерства в зависимости от профилей деятельности организаций, создавая соответственно альтернативные рейтинги.

Несмотря на заявленные преимущества, предложенный подход имеет несколько ограничений, которые необходимо принимать во внимание при дальнейшем анализе. Так, результаты описанного распределения во многом определяются пороговыми значениями, установленными для каждой референтной группы. Изменение этих значений может привести к значительным изменениям всей картины, поскольку пороговые значения зависят от поведения всех научных организаций в выборке в рассматриваемый момент времени. Соответственно, любое изменение в составе выборки может повлиять на оценку отдельно взятой организации. Такие изменения могут быть как продуктивными (например, увеличение показателей эффективности некоторых организаций с течением времени может объективно ухудшить ситуацию для других), либо контрпродуктивным (например, путем введения в выборку необычных объектов, для которых предлагаемые показатели не будут создавать сильные выбросы или, напротив, не иметь выраженных значений).

Предложенный подход не позволяет проводить дальнейшую дифференциацию результатов научной и научно-технической деятельности по областям науки и оценку результативности деятельности мультидисциплинарных организаций. Это остается скорее системным ограничением, поэтому дальнейший подход может включать разработку методов верифика-

ции наукометрических данных и применение дробного счета для получения более точных оценок.

Наконец, использование нормированных показателей создает ситуацию, в которой крупные научно-исследовательские институты могут уступать средним и малым ввиду значительных масштабов деятельности. В результате их вклад оказывается недооценен. Решением может стать введение дополнительных оценок по размеру организаций (крупные, средние, малые).

В заключение следует отметить, что предложенная комбинаторная модель в определенном смысле является декларативной и должна использоваться только в качестве способа стратификации и предварительной оценки результативности научной и научно-технической деятельности организаций. Необходимым представляется проведение экспертизы, основанной на мнении независимых специалистов и учитывающих качественные характеристики (например, ценность) отдельных видов результатов. Дальнейшие исследования будут направлены на преодоление описанных ограничений и поиск оптимизированных решений.

СПИСОК ЛИТЕРАТУРЫ

1. Наука. Технологии. Инновации: 2017. Краткий статистический сборник. – М.: НИУ ВШЭ, 2017.
2. Fraunhofer Gesellschaft. Annual report 2014: the creative power of light. – München: Fraunhofer-Gesellschaft, 2015.
3. CNRS. Rapport d'activité du CNRS 2014. – Paris: Centre National de la Recherche Scientifique, 2015.
4. Murata N. Evaluation System for National University Corporations in Japan: Its Research Aspect. – Washington, DC: National Science Foundation, 2009.
5. Abramo G., D'Angelo C.A. The VQR, Italy's second national research assessment: Methodological failures and ranking distortions // Journal of the Association for Information Science and Technology. – 2015. – Vol. 66(11). – P. 2202-2214.
6. Ancaiani A., Anfossi A. F., Barbara A., Benedetto S., Blasi B., Carlett, V., Constantini M. Evaluating scientific research in Italy: The 2004–10 research evaluation exercise // Research Evaluation. – 2015. – Vol. 24(3). – P. 242-255.

Материал поступил в редакцию 21.09.17.

Сведения об авторах

КУЗЬМИН Глеб Николаевич – ведущий эксперт отдела проектирования информационных систем Института статистических исследований и экономики знаний НИУ ВШЭ, Москва
e-mail: gkuzmin@hse.ru

ФУРСОВ Константин Сергеевич – кандидат социологических наук, заведующий отделом исследований результативности научно-технической деятельности Института статистических исследований и экономики знаний НИУ ВШЭ, Москва
e-mail: ksfursov@hse.ru

Анализ вебометрических показателей основных сайтов, агрегирующих политематическую научную информацию

Обоснована актуальность решения задач оценки доступности, объемов и востребованности научно-технической информации (НТИ), размещенной на сайтах организаций-агрегаторов. Показана целесообразность разделения НТИ на две группы: научную информацию и инженерно-техническую. Приведены принципы категоризации сайтов-агрегаторов научной информации, предложено их разделение на 33 категории. Рассмотрены возможности и недостатки используемых в России методик оценки вебометрических показателей сайтов, содержащих научную информацию. Описана авторская методика оценки объемов и востребованности информации на научных сайтах-агрегаторах.

Апробация предложенной методики проведена на сайтах нескольких ведущих зарубежных и российских организаций-агрегаторов научной информации политематического характера. Охарактеризовано место этих сайтов в научно-информационном пространстве мира и России; состав и источники научной информации, агрегируемой на этих сайтах; условия доступа к ней; использование на сайтах систем расчета наукометрических показателей. Для всех рассмотренных сайтов приведены и прокомментированы их вебометрические показатели.

Ключевые слова: научная информация, сайты-агрегаторы, классификация агрегаторов, информационно-коммуникационные технологии, вебометрические показатели сайтов, web-аналитика, объемы информации, показатели посещаемости, востребованность информации, поведение пользователей

ВВЕДЕНИЕ

Повышение эффективности создания и использования научно-технической информации (НТИ), в том числе за счет ее агрегации на интернет-сайтах и обеспечения к ней удобного доступа, усиление интеграции России в международное научно-информационное пространство – это необходимые условия обеспечения успешного социально-экономического развития России. Выполнение последнего условия, в свою очередь, имеет два основных направления: 1) повышение известности, доступности и востребованности НТИ российского происхождения в зарубежном информационном пространстве; 2) улучшение доступности и эффективности использования иностранной НТИ российскими исследователями – в том числе за счет знакомства их с номенклатурой и возможностями зарубежных баз данных (БД) и иных хранилищ НТИ.

Совершенствование создания и использования НТИ в России важно для решения следующих конкретных задач: обеспечение интеллектуального развития страны, регионов, организаций, отдельных исследователей; поддержка эффективности функционирования/развития социально-экономических и социотехнических систем различных иерархических уровней; улучшение «информационной инфраструктуры общества»; совершенствование ин-

формационного обеспечения научной и инженерно-технической деятельности за счет поддержки доступа к оперативной и ретроспективной информации; поддержка научных коммуникаций, включая международные; формирование научных школ, новых научных направлений исследований и разработок [1-5]. При этом в условиях быстрого роста объемов информации, размещенной в Интернете (в том числе в «электронных научных библиотеках»), важны вопросы ее достоверности и объективности [6].

Доступность НТИ для пользователей имеет несколько аспектов: принципиальная возможность доступа к необходимой научной информации (НИ) из конкретной страны или организации; доступность для пользователей единиц хранения информации на бесплатной основе; наличие «барьеров» доступа из-за высоких уровней оплаты доступа к НИ по отношению к личным доходам исследователей или финансовым возможностям их организаций – это относится и к «академическим пользователям»; удобство использования сайтов – прежде всего их информационно-поисковых систем (ИПС); продолжительность открытия страниц сайтов; длительность обработки запросов в ИПС сайтов [7–13].

В СССР ключевую роль в процессах накопления и, частично, распространения научной информации в

стране играли библиотеки различных типов [1, 14], включая научные, инженерно-технические, ведомственные, учебные и пр.; территориальные и ведомственные центры НТИ (ЦНТИ); организации – депозитарии НТИ, включая ведомственные; организации-агрегаторы сведений об НТИ советского и зарубежного происхождения – прежде всего ВИНТИ РАН [15].

Для обеспечения доступа к информации использовались различные технологии: (1) временная выдача пользователям из фондов библиотек и ЦНТИ в читальных залах, а в некоторых случаях – и на дом книг, журналов, копий описаний изобретений и авторских свидетельств, информационных листов, иных документов и пр.; (2) копирование библиотечными и ЦНТИ имеющихся материалов (целиком или фрагментарно) на бумагу, фотопленку, микрофиши – как в присутствии пользователей, так и с пересылкой по почте копий, которые были заказаны дистанционно. В некоторых случаях такие копии использовались и для формирования фондов библиотек; (3) получение необходимых исследователям специальных изданий (или их копий) через межбиблиотечный абонемент – для временного использования; (4) оперативная рассылка пользователям копий оглавлений новых номеров периодических изданий, включая малодоступные советские и зарубежные научные журналы; (5) централизованное планирование/управление выпуском научной литературы издательствами; (6) ведение в библиотеках и ЦНТИ различных видов библиографических указателей по материалам, имеющимся в их фондах; указателей (ключей) к каталогам; летописей журнальных статей; перечней новых поступлений в фонды книг, журналов и пр. Это давало заинтересованным лицам удобные возможности выявления нужных им материалов в обширных фондах этих организаций; (7) периодические выставки новой литературы, поступившей в фонды отдельных библиотек; всесоюзные выставки новой литературы, выпущенной издательствами; тематические выставки при проведении научных конгрессов и т.п.; (8) ключевую роль в агрегации и обеспечении известности информационных материалов (особенно зарубежных) играл реферативный журнал (РЖ) ВИНТИ, который состоял из различных тематических серий [15]. По крайней мере, часть серий была доступна во многих крупных библиотеках и иных организациях. На основе публикуемых в РЖ кратких рефератов (иногда – авторских рефератов или просто названий работ) исследователи могли заказать копии нужной им информации – за счет как собственных средств, так и средств организаций, в которых они работали; (9) ряд научных журналов (прежде всего академических) практиковал рассылку авторам отписок (копий) опубликованных ими статей.

С развитием информационно-телекоммуникационных технологий (ИТКТ) ситуация принципиально изменилась. (А) Стала актуальной проблема оперативного мониторинга и анализа мирового потока научно-технической информации [16] по мере ее появления. (Б) Расширились функции современной системы НТИ в России [17], в том числе в отношении поддержки инновационных процессов, обучения специалистов и пр. (В) Основную роль в накоплении

и обеспечении доступности НТИ для потенциальных пользователей в России и за рубежом начали играть интернет-сайты [2, 8, 18, 19] и ИПС этих сайтов, поисковые системы Интернета. (Г) Доступность и востребованность научной информации увеличилась за счет широкого распространения и использования возможностей мобильных устройств, включая планшеты, смартфоны, ноутбуки. (Д) Широкое развитие получили автоматизированные средства наукометрического анализа научной информации, в том числе показателей отдельных исследователей, а также журналов, организаций, стран [20]. (Е) Резко увеличилось внимание к интеграции России в международное информационное пространство. Это особенно касается публикаций российских ученых в тех научных изданиях, материалы которых индексируются в международных системах учета цитирований и оценки наукометрических показателей, признаваемых Высшей аттестационной комиссией (ВАК) России [20]. Термин «признаваемые» здесь и далее означает, что публикации в таких изданиях специально учитываются ВАК при защитах диссертаций и пр. При этом исходя из решения ВАК от 15.06.2017 №1-пл/1 следует считать, что статус таких публикаций более высокий, чем работ, опубликованных в российских журналах, включенных в списки изданий, рекомендованных ВАК для публикации результатов кандидатских и докторских диссертаций.

В связи с этим необходимо отметить следующее. (1) Значительная часть НТИ, особенно информационно-рекламного характера, размещается на многих сайтах вообще без какого-либо контроля качества и достоверности. Иногда такая информация носит тенденциозный характер, направлена только на продвижение выполненных разработок, товаров, информационных и иных видов услуг. (2) По запросам пользователей поисковые системы Интернета часто выдают слишком много ссылок, причем сортировка их по релевантности нередко неудовлетворительная. Переформулирование поисковых запросов для улучшения качества выдач этих систем далеко не всегда позволяет получать удовлетворительные результаты. Возможности улучшения качества обработки поисковых запросов на практике достаточно ограничены – даже при использовании средств расширенного поиска, средств интеллектуализации этих систем и пр. (3) На сайтах-агрегаторах НТИ [1, 17, 19] размещаемая информация предварительно контролируется и индексируется. Это позволяет пользователям обеспечивать селективный отбор материалов на основе использования внутренних ИПС сайтов, которые, обычно, имеют возможности как простого поиска, так и расширенного. Однако даже такие поисковые системы часто дают много нерелевантных ссылок даже при вполне корректных запросах. Тем не менее, исследователи осуществляют поиск научной информации в основном именно на сайтах-агрегаторах. Поэтому весьма актуальны оценки удобства использования таких сайтов [13], показателей их востребованности у пользователей – они могут быть даны на основе различных средств web-аналитики (включая счетчики посещаемости сайтов), а также методов альтметрики. Вебометрические показатели (ВМП)

сайтов, включая их динамику, принципиально важны для анализа и управления потоками НТИ, определения эффективности работы научных организаций и вузов, оценки и управления полезностью их сайтов, построения (калибровки) математических моделей потоков НТИ и пр. [21–29]. Есть и публикации, специально посвященные исследованиям ВМП российских сайтов научных организаций [28, 30].

ОБЩИЕ ВОПРОСЫ ДОСТУПНОСТИ И КАЧЕСТВА НАУЧНОЙ ИНФОРМАЦИИ

Большинство существующих научных публикаций и нормативных документов посвящено научно-технической информации. Однако для анализа веб-метрических показателей сайтов-агрегаторов целесообразно разделить НТИ на собственно научную и инженерно-техническую (включая производственную) информацию. Четкой границы между этими двумя группами часто нет, поэтому они могут размещаться как на общих, так и на разных интернет-ресурсах. Нами рассматриваются только сайты-агрегаторы научной информации, хотя на части этих сайтов есть и инженерно-техническая информация.

В состав научной информации входят различные виды (категории) документов, причем они агрегируются на сайтах в разной степени и из разных источников. Подчеркнем, что НИ представляет собой не только опубликованные научные статьи и монографии, но и другие виды документов, включая сведения о предстоящих научных мероприятиях, персональную информацию об ученых и их достижениях, сведения о научных школах [4] и т.д. Важную роль в доступности научной информации играют сервисные средства Интернета. Они могут быть интегрированы с БД сайтов-агрегаторов или располагаться на отдельных интернет-сайтах (т.е. использовать внешние БД).

Источниками агрегируемой НИ являются, прежде всего, публикации в различных периодических научных изданиях, материалы конференций, сборники трудов, научные монографии (включая коллективные), препринты. При этом предполагается, что в отношении этих материалов имеется некоторый редакционный контроль качества и достоверности информации. Защита информационного пространства от недоброкачественной (низкоуровневой) научной информации осуществляется за счет таких мер, как достаточно жесткий отбор, который проходят издания, включаемые в ведущие международные системы учета цитирований и оценки наукометрических показателей; ограниченность количества источников, из которых отбираются работы в *Scholar Google (SG)*, иные специализированные ИПС по научной информации; предъявление определенных требований к изданиям, включаемым в списки ВАК России. Отметим также решение Российского индекса научного цитирования (РИНЦ) от 29.04.2017 об исключении 344 изданий из числа индексируемых этим ресурсом.

В типичных случаях организации-агрегаторы получают научные издания в электронной форме, уже снабженные необходимой гипертекстовой разметкой. Это позволяет минимизировать трудозатраты этих организаций по включению новых материалов в их БД.

На различных интернет-сайтах могут находиться как копии одних и тех же материалов, так и ссылки на места хранения в Интернете их оригиналов.

Первый вариант приводит к информационной избыточности. Однако при этом повышается надежность хранения материалов, увеличивается вероятность их обнаружения поисковыми системами Интернета и внутренними ИПС сайтов-агрегаторов. В частности, материалы, опубликованные в научных журналах многих российских вузов, размещаются как на сайтах этих организаций, так и на сайте *elibrary.ru*, сайте КиберЛенинки и пр.

Второй вариант размещения НИ исключает дублирование материалов и искажения библиографических сведений при их копировании. Однако он более уязвим в отношении угроз информационной безопасности, особенно при отсутствии зеркал для серверов, на которых находятся хранилища оригинальных материалов.

Средствами улучшения потенциальной доступности информации, в общем случае, может быть деятельность по переводу научной информации на другие языки и целенаправленная популяризация (информационное продвижение) результатов исследований [31–33]. В СССР для повышения доступности иноязычных документов, наиболее важные из них переводились на русский язык, а для информирования об этих переводах выпускались тематические библиографические указатели, которые рассылались по библиотекам и ЦНТИ.

Сейчас профессионально выполненные русскоязычные переводы размещаются на интернет-сайтах наряду с соответствующими документами на языках оригиналов. Однако специальные указатели переводных материалов в Интернете, как правило, не размещаются. Возможности электронных переводчиков для получения качественных переводов достаточно ограничены из-за многозначности используемых в научных документах терминов и сложности конструкций фраз.

Перечислим основные причины, по которым на интернет-сайтах относительно мало русскоязычных переводов иноязычных научных документов: невысокая востребованность таких переводов, особенно по узкоспециальной тематике; относительно высокий уровень владения английским языком лиц, профессионально занимающихся научной деятельностью, знание ими специальных терминов в своих предметных областях; возможность применения электронных переводчиков (включая интернет-переводчики) и электронных словарей, представляющих образцы выражений, используемых в различных предметных областях, источниках; частичная компенсация недостатков стилистики электронных переводов за счет профессиональных знаний специалистов, которые занимаются научной деятельностью в соответствующих предметных областях.

Отметим некоторые факторы, важные для последующего изложения.

Существующие подходы к оценке востребованности научной информации в электронной форме ориентированы на размещение ее на сайтах-агрегаторах.

Фактическая востребованность русскоязычных научных материалов выше, чем количество ссылок в научных публикациях, учитываемых РИНЦ и другими системами учета цитирований и оценки.

Расширенные оценки востребованности отдельных русскоязычных научных документов потенциально могут даваться с помощью альтметрики РИНЦ (включая количества просмотров материалов, их скачиваний, включений в подборки). Однако эти параметры для расчетов показателей в РИНЦ не используются по следующим причинам: комплексные методики вычисления востребованности научной информации пока не разработаны; альтметрика дает больше возможностей для злоупотреблений, чем подсчет количества ссылок в научных статьях.

Для сайтов показатели альтметрики имеют иное содержание, чем для статей. Использование таких показателей достаточно активно обсуждается в научных публикациях [24–26], в том числе в отношении показателя *трафик сайта* как характеристики его востребованности в Интернете [23]. Однако трафик конкретного сайта (включая его динамику во времени), могут определить лишь системные администраторы сайтов, установив специальные программы на серверах [22] или подключив внешние средства вебометрики [22, 28, 30]. В Интернете данные о трафике отдельных сайтов, как правило, не публикуются. Поэтому сведения о динамике изменения трафика сайтов (суточного, недельного, месячного) могут использоваться для поддержки принятия решений лишь менеджментом самих сайтов.

Основная масса научных материалов в мире сейчас публикуется на английском языке. Быстро растет число англоязычных публикаций российских авторов в зарубежных изданиях. Немало таких статей и в российских журналах. Много работ появляется и на других языках, включая китайский, японский, корейский, французский, испанский. Их доступность для русскоязычных пользователей обеспечивается в значительной степени опубликованием англоязычных аннотаций этих статей, а иногда – использованием электронных переводчиков/словарей. Небольшую часть научных статей на белорусском и (с некоторыми оговорками) украинском языках, русскоязычные пользователи могут читать без специального перевода.

Редакциями российских научных журналов статьи на английском языке обычно приветствуются. Это увеличивает вероятность включения журналов в международные системы учета цитирований и оценки наукометрических показателей, признаваемые ВАК России. Для большинства российских авторов трудоемкость подготовки статей на английском языке существенно выше, чем на русском. Доступность таких англоязычных статей для русскоязычных пользователей ниже, чем русскоязычных, что обычно приводит к уменьшению количества ссылок на такие публикации в русскоязычной литературе.

Доступность русскоязычных публикаций для англоязычных ученых обеспечивается размещением англоязычных аннотаций или версий статей на интернет-сайтах, деятельностью фирмы «МАИК/Наука Интерпериодика» по переводу на английский язык ряда российских научных журналов. Последний ва-

риант (русскоязычная + англоязычная публикации) наиболее интересен для авторов – одновременно обеспечивается доступность статей как для русскоязычных, так и для иноязычных исследователей. Это потенциально приводит к увеличению количества ссылок на эти статьи [34], улучшению их показателей альтметрики, расширению личных контактов исследователей.

СТАТУС РОССИЙСКИХ НАУЧНЫХ ЖУРНАЛОВ, СИСТЕМ УЧЕТА ЦИТИРОВАНИЙ И ОЦЕНКИ НАУКОМЕТРИЧЕСКИХ ПОКАЗАТЕЛЕЙ С ПОЗИЦИЙ ВАК ПРИ МИНОБРНАУКИ РОССИИ

Высшей аттестационной комиссией (ВАК) используются две категории рецензируемых научных журналов, рекомендуемых для публикации результатов кандидатских и докторских диссертационных исследований: издания, входящие в международные системы учета цитирований и оценки наукометрических показателей, признаваемые ВАК (российские журналы представлены в списке С1); рецензируемые российские научные издания, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата и доктора наук (список С2).

На 1 июля 2017 г. ВАК признавала следующие международные системы учета цитирований и оценки наукометрических показателей: *Web of Science, Scopus, Springer, ZbMath, MathSciNet, GeoRef, PubMed, Agris, Astrophysics Data System (ADS)*¹. Первые три ресурса являются универсальными (политематическими), а остальные целесообразно считать специализированными. Для попадания в списки этих ресурсов издатели научных журналов должны подавать заявления, проходить специальные (и довольно жесткие) процедуры отбора в отношении соблюдения правил публикационной этики, обеспечения качества и актуальности публикуемых работ, соответствия тематики статей их названиям, аннотациям и декларированным направлениям деятельности изданий. На конец июня 2017 г. в списке С1 были указаны 1003, а в списке С2 – 2143 российских издания.

Категоризация сайтов-агрегаторов научной информации возможна на основе различных подходов [16]. Для разделения сайтов-агрегаторов на категории по их характеристикам нами используется следующая терминология:

T1 – системы учета цитирований и оценки наукометрических показателей;

T2 – библиографические базы данных (содержат только библиографические сведения);

T3 – реферативные БД – содержат рефераты (аннотации) материалов из научных журналов и иных изданий;

T4 – полнотекстовые БД научных материалов;

T5 – поисковые системы по научным статьям и другим материалам;

¹ По решению ВАК от 15.06.2017 №1-пл/1 с 01.01.2018 г. перестанут «признаваться» системы Agris и ADS.

Т6 – поисковые, информационные и информационно-аналитические системы по научным журналам и сайтам;

Т7 – сервисные средства для обеспечения работы с персональной информацией об ученых. Ряд рассматриваемых далее интернет-ресурсов обладает сразу несколькими функциональными возможностями из числа указанных выше.

В том или ином объеме научная информация представлена на интернет-сайтах различных агрегаторов. Однако известность этих сайтов и их востребованность значительно различаются. В этом отношении важна и конкуренция между сайтами, особенно если доступ к одним и тем же материалам можно получить на разных ресурсах. В частности, в России на сайтах многих вузов и НИИ размещены в открытом доступе не только тексты научных статей, опубликованных в изданиях этих организаций, но и материалы, связанные с защитой диссертаций. Однако обычно на таких сайтах информация (особенно по диссертациям) оказывается доступной для пользователей только, если они четко знают, *где и что* искать.

Мы выделим 33 категории сайтов-агрегаторов НИ, исходя из назначения и функциональности сайтов; характера содержащейся на них информации; *необходимости и достаточности* всех категорий; удобства проведения анализа (сравнения) зарубежных и российских ресурсов.

К1. Сайты международных политематических (универсальных) систем учета цитирований и оценки наукометрических показателей, признаваемые ВАК России.

К2. Сайты международных политематических (универсальных) хранилищ научной информации, не признаваемых ВАК России, в том числе и ресурсы, содержащие некоторые оценки наукометрических показателей.

К3. Основные российские интернет-ресурсы научной информации универсального характера.

К4. Международные тематически специализированные системы учета цитирований и оценки наукометрических показателей, признаваемые ВАК России.

К5. Международные тематически специализированные хранилища научной информации, не признаваемые ВАК.

К6. Сайты некоторых зарубежных стран, специально предназначенные для информационной поддержки национальных публикаций.

К7. Интернет-ресурсы с каталогами диссертаций и авторефератов, их текстами.

К8. Информационные и информационно-аналитические системы по научным журналам.

К9. Зарубежные сайты отдельных журналов или их групп, имеющих общего издателя.

К10. Российские сайты отдельных журналов или их групп, имеющих общего издателя.

К11. Сайты для присвоения DOI статьям и ORCID авторам.

К12. Тематически специализированные российские ресурсы (хранилища) научной информации, не отнесенные к другим категориям.

К13. Сайты международных и зарубежных научных и научно-технических организаций, их объединений, научных обществ.

К14. Сайты российских научных и научно-технических организаций, их объединений, научных обществ.

К15. Сайты зарубежных вузов широкого профиля деятельности.

К16. Сайты российских вузов широкого профиля деятельности.

К17. Сайты зарубежных вузов, специализированных по профилю деятельности.

К18. Сайты российских вузов, специализированных по профилю деятельности.

К19. Универсальные и специализированные поисковые системы по научной информации.

(а) Только структурированные «коллекции» гиперссылок на интернет-ресурсы, иногда в сочетании с некоторой информацией о содержании этих ресурсов.

(б) Системы, имеющие собственные базы данных (индексные каталоги) – в том числе в отношении мест расположения БД научных материалов на внешних ресурсах, объемов и моментов актуализации таких БД.

(в) Системы, которые по запросам пользователей автоматически собирают и агрегируют информацию из всех доступных интернет-источников.

К20. Аннотированные каталоги информационных ресурсов, в том числе по средствам сервисного характера.

К21. Системы для выявления неоригинального текста (плагиата, случаев некорректного цитирования) в научных статьях и иных материалах, находящихся на интернет-сайтах в электронном виде.

К22. Сайты традиционных библиотек зарубежных стран и России.

К23. Сайты издательств научной литературы.

К24. Электронно-библиотечные системы издательств.

К25. Сайты электронных библиотек, не отнесенные к вышеприведенным категориям.

К26. Специализированные сайты со сводной информацией по предстоящим или прошедшим научным конференциям, семинарам, другим видам научных мероприятий.

К27. Интернет-ресурсы с информацией о научных организациях, их рейтингах.

К28. Ресурсы с персональной информацией об отдельных ученых, их коллективах.

К29. Сайты профессиональных социальных сетей и другие ресурсы, предназначенные для обеспечения научных коммуникаций.

К30. Сайты с нормативными документами в отношении научной деятельности, требований к научной публикационной этике.

К31. Сайты российских и зарубежных фондов, предоставляющих гранты на проведение научных исследований, организацию научных мероприятий и пр.

К32. Сайты, связанные с обеспечением процессов подготовки научных кадров, занятия вакантных должностей научных работников, прохождения стажировок.

К33. Сайты, содержащие сводную статистическую, обзорно-аналитическую и т.п. информацию о

результатах научной деятельности, о прогнозах такой деятельности.

Поскольку приведенная классификация относится только к НИ, в ней не отражены группы сайтов-агрегаторов с патентной информацией, со сведениями о программном обеспечении, БД и пр. Такая информация по российским разработкам есть, в частности, на сайте Федерального института промышленной собственности (ФИПС) России.

Вопросы поддержки информационных взаимосвязей между перечисленными категориями интернет-ресурсов (в том числе за счет использования гиперссылок) весьма актуальны. Однако эта тематика нуждается в отдельном рассмотрении.

Для организаций-агрегаторов НТИ большое значение имеет выбор источников агрегируемой информации с учетом ее достоверности, актуальности, качества, полноты охвата тематики; принципов и практических технологий агрегирования информации из различных источников [9, 33], а также исключения неоправданного дублирования материалов на одних и тех же ресурсах; тематической классификации и назначения адекватных дескрипторов для хранимых материалов; обеспечения информационной безопасности для отдельных информационных материалов и сайтов-агрегаторов в целом.

В настоящей статье из-за ограничений по объему анализируются вебметрические показатели только сайтов, относящихся к категориям К1–К3.

МЕТОДИКА ИССЛЕДОВАНИЙ

Сайты-агрегаторы для рассматриваемых трех категорий (К1–К3) выбирались исходя из их известности в международном и российском информационном пространстве. Для поиска адресов интернет-сайтов были использованы поисковые системы (машины), расположенные на сайтах <http://www.google.ru>, <http://www.yandex.ru>, а для информации по организациям-агрегаторам – на <https://ru.wikipedia.org>.

Сведения о направлениях деятельности ресурсов и их объемах брались как с сайтов самих ресурсов, так и из других источников. Отметим, что по одним и тем же сайтам-агрегаторам данные об объемах ресурсов в разных источниках иногда существенно различаются, причем не только из-за того, что они относятся к разным датам по времени.

Потребность в объективной оценке вебметрических показателей сайтов обуславливает развитие средств веб-аналитики [21], включая разработку и использование вебметрических индексов сайтов. Фактически они представляют собой интегральные показатели, позволяющие обобщить (свернуть) информацию по совокупности частных ВМП сайтов. Могут применяться такие варианты оценки вебметрических показателей.

(v1) По данным, передаваемым системными администраторами сайтов или уполномоченными представителями организаций в некоторую организацию, концентрирующую такие сведения. Эти данные получают за счет использования размещенных на сайтах программ (включая счетчики посещаемости [22]) и/или «подключения инструментов анализа веб-трафика: *Google Analytics*; Яндекс.Метрика».

В этих случаях удается оценивать/сравнивать посещаемости разных групп страниц сайтов или даже отдельных страниц. Подчеркнем, что такой подход возможен лишь в весьма ограниченном количестве случаев, так как большинство организаций не заинтересованы в передаче данных о ВМП сайтов, их сравнении с другими организациями, публикации таких сведений в открытых источниках.

В редких случаях на стартовых страницах сайтов-агрегаторов указываются текущие количества посетителей или накопленные количества входов посетителей (просмотров) за определенный период времени. Значительный интерес представлял бы отдельный учет посетителей таких категорий: из *своих* организаций по локальной вычислительной сети (это количество системным администраторам оценить достаточно легко); пользователей извне организаций, но из *своих* населенных пунктов; из других населенных пунктов России; зарубежных пользователей (как максимум – с разбивкой по странам).

(v2) Сведения, публикуемые на специальных вебметрических сайтах [30] по совокупностям специально изучаемых (мониторируемых) сайтов организаций.

(v3) Информация о сайтах, получаемая из основных поисковых систем Интернета. Такой подход использован, в частности, в [28,30] при сравнительном анализе сайтов академических организаций России. Результаты мониторинга этих ресурсов размещаются на сайте <http://webometrix.ru>. Однако авторами [28] признается, что значения, выдаваемые разными поисковыми системами Интернета, существенно отличаются. Отметим, что и процессы совершенствования этих систем также могут приводить к изменениям ВМП сайтов-агрегаторов. Как следствие данные, полученные за разные периоды, могут быть не полностью сопоставимы.

Для получения рейтингов сайтов в [28] рассчитывался показатель по алгоритму, представленному на <http://w.ict.nsc.ru/ranking/>. В нем используются такие частные вебметрические показатели. а) Среднее арифметическое количества внешних ссылок на сайт по трем системам (*Yandex*, *Google*, *Bing*). При этом относительные популярности указанных поисковых систем не учитывались. б) Среднее арифметическое количества страниц на сайте (определенное по тем же трем системам). в) Среднее арифметическое от количества *PDF*, *DOC*, *PPT* файлов (в сумме) – усреднение только по системам *Google* и *Yandex*. г) Индексы цитирования, полученные из систем *Индекс цитирования Яндекса* и *Scholar Google*. Отметим, что целесообразность использования *Scholar Google* для таких целей в литературе продолжает обсуждаться [35]. Затем по всем четырем параметрам (а...г) проводилось ранжирование организаций. В конечном счете для каждой из рассматриваемых организаций ранги для всех четырех рейтингов складывались. Суммирование рангов (не первичных данных) позволяет работать с разнородными показателями. Однако принятие по умолчанию одинаковой «важности» для всех четырех параметров с нашей точки зрения логически уязвимо.

В связи с этим укажем, что в [28] при построении рейтингов сайтов учитывалось пять показателей.

1. Размер сайта – оценивался по количеству его страниц.

2. Видимость сайта – оценивалась по количеству входящих ссылок, видимых (обнаруживаемых) на сайте поисковыми системами Интернета.

3. Альтметрика – в [28] она понимается как видимость сайта в сетях профессионального общения и сотрудничества ученых.

4. Количество полнотекстовых файлов, обнаруживаемых поисковыми системами (машинами) Интернета.

5. Научность сайта – в [28] оценивается как количество ссылок на сайт и материалы на нем, обнаруживаемых *Scholar Google*.

Представляется, что если говорить о качестве научных материалов, то целесообразно было бы дополнительно принимать во внимание и количество цитирований по каждому материалу, отражаемому на *Scholar Google*. Для вычисления вебометрического индекса конкретного сайта перечисленные показатели в [28] учитываются с весовыми коэффициентами соответственно: 0.1, 0.25, 0.25, 0.1, 0.30. Эта система весовых коэффициентов взята, видимо, с www.webometrics.info и соответствует методике Кибернетической лаборатории Центра научной информации и документации Национального исследовательского совета Испании. Недостатками этой методики, с нашей точки зрения, являются следующие: по пунктам 1 и 2 ([28, с. 4]) отсутствие учёта для используемых систем *Yandex*, *Rambler* их относительной популярности в России, по пункту 3 [28] – отсутствие учета популярных в нашей стране социальных сетей «В контакте» (и, с оговорками, «Одноклассники») – ими российские ученые также пользуются.

Для России считается целесообразной разработка национальных вебометрических индексов [36]. На практике такие индексы начали создаваться преимущественно для сайтов научных организаций России, причем независимо различными центрами [28, 30, 37].

(v4) Информация о показателях сайтов, получаемая при запросах из *Scholar Goggle* [28, 30, 37]. Считается, что эти данные показывают качество выставленных в общий доступ материалов, так как *Scholar Goggle* индексирует лишь отдельные объекты. В основном это архивы научных журналов, материалы научных конференций, научные труды сотрудников. Однако *Scholar Goggle* вообще не может индексировать содержимое некоторых сайтов, в частности таких, для доступа к ресурсам которых необходим ввод логина-пароля.

(v5) Данные, собираемые специальными программными средствами (включая общедоступные и платные), которые в дистанционном режиме позволяют анализировать показатели нужных сайтов.

(v6) Данные о гиперссылках на сайтах, собираемые специально разработанными роботами-краулерами [37], возможно, тематически специализированными [38].

В настоящей статье авторами применена методика оценки вебометрических показателей сайтов, исполь-

зующая подходы на основе вариантов типов «v5» и «v4». Она в основном аналогична той, которая описана авторами ранее в [39]. Ниже эта методика частично воспроизводится – для обеспечения понимания читателями информации в приводимых ниже таблицах (в них применяются общепринятые англоязычные термины и их переводы). Оценка ВМП сайтов-агрегаторов выполнялась нами по пяти направлениям.

1) Доступность сайта (*Ping*) оценивалась по среднему времени (*Average time, AT*) открытия стартовой страницы в секундах. Оценки были получены с применением сервиса для проверки скорости загрузки страниц сайтов с разных точек мира – <http://sitespeed.me>.

2) Параметры наполнения сайта (*Content*): (а) Веб-размер сайта – определялся с помощью его адресной системы на основе количества адресов статических или динамически генерируемых файлов (*Count, URLs*). Авторы применяли такие параметры этой подгруппы: количество страниц сайта (*text/html*); количество графических файлов (*image*), размещенных в файловой системе сайта и используемых на нем; количество размещенных на сайте файлов (*application*). Эти обозначения соответствуют спецификации *Internet Media Types* – стандартным идентификаторам типа файла (*Content-Type*).

Значения указанных параметров были получены из стандартных отчетов, сгенерированных программой *Xenu's Link Sleuth* (<http://xenus-link-sleuth.en.softonic.com>). Она проверяет веб-сайты на наличие гиперссылок по *URL*, находящихся в следующих объектах: в тегах *<a>*; в тегах изображений; в плагинах; в местных картах изображений; в таблицах стилей; в скриптах и *Java*-апплетах. Программа *Xenu's Link Sleuth* проходит по ссылкам на интернет-страницы и проверяет также все ссылки, которые есть на этих страницах. Такой алгоритм потенциально позволяет проверить весь сайт, но он может работать чересчур долго на сайтах большого размера.

(б) Объем сайта в мегабайтах (*Size, Mb*) определялся также с помощью программы *Xenu's Link Sleuth*.

3) Заметность (*Visibility*) сайта оценивалась по количеству ссылок на него, которые размещены на других сайтах, т.е. по количеству входящих ссылок. Для определения заметности использовался сервис, размещенный на <https://majestic.com>.

4) Для оценок удобства работы с информацией на сайте авторами настоящей статьи использовались такие параметры: (а) Количество внутренних ссылок (KBC) на сайте (φ) оценивалось по

$$\varphi = \sum_{j=1}^m LIn_j, \quad (1)$$

где m – количество внутренних *URL* на сайте, LIn_j – количество ссылок на j -й внутренний *URL* на сайте. При этом учитывались только *URL* файлов типа *text/html*, т.е. с указанием на страницы.

(б) Количество исходящих ссылок (КИС) с сайта учитывалось по двум показателям: (б1) Количество уникальных ссылок на интернет-страницы (количество исходящих URL на сайте) – обозначим его как « Ω ». (б2) Общее количество исходящих ссылок (ссылок на исходящие URL на сайте) – обозначим его как « Ψ ». Оно оценивалось по формуле

$$\Psi = \sum_{i=1}^n LIn_i \quad (2)$$

где n – количество исходящих URL на сайте; LIn_i – количество ссылок на сайте на i -й исходящий URL.

Значения КВС и КИС оценивались на основе стандартных отчетов *Xenu's Link Sleuth*.

Помимо указанных выше абсолютных показателей для ссылок, авторами рассчитывались и относительные – с нормировкой на количества страниц сайтов. Это позволяет сравнивать анализируемые характеристики для сайтов разного размера (объема).

5) Востребованность сайтов оценивалась по трем показателям: (а) количество уникальных посетителей за месяц; (б) количество просмотров за месяц; (в) средняя длительность просмотра (*Average Visit Duration, AVD*), т.е. средняя продолжительность работы посетителя с сайтом за один сеанс доступа к нему. При этом для показателей «1» и «2» месяцем считался период в 30 дней, предшествующих дате получения оценок.

Показатели «а» и «б» (для п. 5) оценивались с помощью сервиса, размещенного на <http://pr-cy.ru/>. Этот сервис автоматически проверяет открытые счетчики статистики и на их основе получает основные параметры посещаемости. Сначала ищется открытый счетчик Яндекс.Метрика. Если на нем нет сведений о сайте, то проверяется ресурс *LiveInternet*. Если в этих источниках ничего не было найдено, то система ресурса http://pr-cy.ru делает прогноз посещаемости сайта на основе рейтинговой системы. Ее работа основана на использовании соотношения посещаемости конкретных сайтов (общего количества их просмотров и частот посещений) и прочих интернет-сайтов. Для таких оценок используется ресурс *Alexa Rank* на <http://www.alexa.com>. Чем ниже показатель *Alexa Rank*, тем более популярен интернет-ресурс у пользователей. В сервисе, размещенном на сайте <http://pr-cy.ru>, реализован алгоритм, который позволяет конвертировать прогнозный показатель *Alexa Rank* в примерные прогнозные данные по посещаемости оцениваемого сайта.

Показатель средней длительности просмотра страниц сайтов (подпункт «в») определялся с помощью сервиса, находящегося на сайте <https://www.similarweb.com>.

Суммарная продолжительность работы пользователей с сайтом-агрегатором может быть оценена как

произведение «средней длительности просмотра» на «суммарное количество просмотров». Однако к этому показателю следует относиться с некоторой осторожностью, так как пользователи могут «держаться» открытой страницы сайта-агрегатора в свернутом окне и пользоваться ей только время от времени. Более глубокий анализ (он доступен только системным администраторам сайтов) может включать в себя определение средних показателей активности пользователей за один просмотр сайта: промежуток времени между выполнением любых действий, включая прокрутку страниц с результатами выдач ИПС сайтов и текстов открытых файлов, включение пользователями материалов в личные подборки и пр.; количество выполненных запросов; количество открытых и/или скачанных файлов.

Кроме того, оценивалось количество объектов сайтов, индексируемых *Scholar Google*.

В связи со значительной разнородностью рассматриваемых в статье сайтов-агрегаторов авторы не предлагают для них обобщенных (интегральных) критериев оценки.

Анализ данных по сайтам-агрегаторам осложняется тем, что некоторые ресурсы не имеют отдельных сайтов, а размещены на страницах сайтов других организаций или их подразделений. При таких условиях для внешних пользователей с использованием общераспространенных программ обычно нельзя получить показатели для групп страниц, соответствующих сайтам-агрегаторам. Однако показатели таких групп (и даже отдельных страниц) могут получать системные администраторы сайтов-агрегаторов за счет применения счетчиков для отдельных страниц [22] сайтов, а также подключения к сайтам внешних средств веб-аналитики [21, 22].

В приводимых далее таблицах вся информация относится к периоду с 1.07.2017 по 25.07.2017 (получить данные за более короткий промежуток не удалось из-за высокой трудоемкости работ). Поскольку со временем данные будут меняться, анализ динамики ВМП сайтов может быть отдельным направлением исследований. В табл. 1 приведено сравнение ВМП на основе различных подходов.

На основании табл. 1 можно сделать следующие выводы: 1) результаты, получаемые с использованием различных подходов, существенно отличаются; 2) разные поисковые системы (машины) Интернета дают значительно различающиеся результаты для одних и тех же сайтов; 3) использование *Scholar Google* для оценки количества ссылок по ряду сайтов оказывается непродуктивным.

Кроме того, отметим, что по сравнению с существующими подходами, описанными в начале данного раздела, предлагаемая нами методика имеет большую информативность, особенно в отношении оценки востребованности сайтов.

Сравнение вебметрических показателей, полученных разными методами

Ресурс	По методике авторов		На основе подходов из [28]			
	KBC*	KC*	Yandex KC	Google KBC	Google KC	Scholar Google KC
www.scopus.com	7433278	1	1	1050000	1	0
www.webofknowledge.com	1632266	1	1	2740000	2610	0
www.springer.com	16152325	>1000000**	209000	29800000	1100000	89
www.springeropen.com	281455	56945	1000	309000	20000	18
doaj.org	3581433	59	23000	189000	514000	0
indexcopernicus.com	986019	138	0	81200	4	22
www.proquest.com	3542085	5962	2000	3720000	5260	0
www.researchbib.com	104156	15	19	86700	187	0
www.science.gov	171925	176	246000	8670000	695000	1
www.sciencedirect.com	25891818	79	384000	4740000	14700000	9550000
www.hub.sciverse.com	7336	2	1	49600	1	0
www.scienceresearch.com	8324	34	0	3430000	26	0
cogprints.org	69007	198	8000	119000	109000	1750
sci-hub.io	96094	3	2	1320000	2	0
libgen.io	323244	189	2	3190000	3670	6
elibrary.ru	191615	4	3000000	767000	4440000	1670000
cyberleninka.ru	1188351	1673138	2000000	175000	1300000	1270000
www2.viniti.ru	55214	294	383	23100	322	1
sci-lib.com	751643	348907	11000	2220000	585000	0
gbu.bookchamber.ru	1082	29	37	113	46	0
sciencepublic.ru	479	60	5	76900	746	50
arbicon.ru	210371	607	3000	6050	18300	156
archive.neicon.ru	218698	62779	0	100	64100	110

* KBC – количество входящих ссылок на сайт; KC – количество страниц на сайте (для предлагаемой методики KC соответствует сумме показателей *Text/Html* и *Application* в таблицах с четными номерами);

** – из-за большого объема обрабатываемых данных более точную величину получить не удалось.

ЗАРУБЕЖНЫЕ ПОЛИТЕМАТИЧЕСКИЕ СИСТЕМЫ УЧЕТА ЦИТИРОВАНИЙ И ОЦЕНКИ НАУКОМЕТРИЧЕСКИХ ПОКАЗАТЕЛЕЙ, ПРИЗНАВАЕМЫЕ ВАК

В этом и последующих разделах сначала дается содержательная характеристика ресурсов, входящих в соответствующую категорию, а затем приводятся таблицы с ВМП для этих ресурсов.

Scopus – ресурс размещен на сайте www.scopus.com. Представляет собой реферативную и наукометрическую БД. Индексируется более 18500 наименований научно-технических и медицинских журналов, издаваемых примерно 5 тыс. организаций.

Для доступа к основной массе размещенной на ресурсе научной информации заинтересованные организации должны заключать специальные договоры и оплачивать владельцу ресурса определенные суммы. Поэтому в полном объеме в России информация на ресурсе доступна, в основном, лишь со служебных компьютеров крупных вузов, НИИ, ведущих библиотек. Приводимые далее показатели относятся к ресурсу на www.scopus.com в целом.

Web of Science (WoS) – ресурс размещается по адресу www.webofknowledge.com. Основные части: *Science Citation Index Expanded (SCI-EXPANDED)* – информация по естественным и некоторым прикладным наукам; *Social Sciences Citation Index (SSCI)* – по

социологическим дисциплинам; *Arts&Humanities Citation Index (A&HCI)* – по искусству и гуманитарным наукам. Информация в указанных частях приводится начиная с 1991 г. Эти ресурсы не содержат полных текстов статей, но включают библиографические списки к работам.

На сайте ВАК в разделе «Информация об изданиях», входящих в список «С1», приведены сокращенные обозначения для частей *WoS*. Их расшифровку мы даем на основе сведений из различных источников: (а) *SCI-Expanded (Science Citation Index Expanded)*; (б) *ESCI-Emerging Science Citation Index*, является частью *Web of Science Core Collection*. В эту базу, согласно сведениям на сайте <http://www.openscience.in.ua/ru/esci.html>, «могут приниматься журналы регионального уровня, которые освещают локальные вопросы»; (в) *BIOSIS* соответствует *BIOSIS citation index*; (г) *ZR* – соответствует *Zoological Record*. Приведенные выше аббревиатуры SSCI и A&HCI в списке «С1» ВАК не упоминаются вообще.

В списке «С1» ВАК аббревиатура *RSCI (Russian Science Citation Index)*, соответствующая ядру научных журналов РИНЦ, в явном виде не упоминается. Однако информация по этой части журналов отражена, например, на https://ru.wikipedia.org/wiki/Web_of_Science. Отметим, что помимо *RSCI* на *WoS* располагается и информация аналогичного статусно-

го уровня об иных группах журналов, например, *KCI-Korean Journal Database*.

Укажем также сайт с web-аналитикой для *WoS* (*Clarivate Analytics*): <https://clarivate.com/products/web-of-science/>. Есть и русскоязычная версия этого сайта (<http://wokinfo.com/russian/>).

Springer – сайт www.springer.com. Политематическая БД одного из крупнейших в мире медиахолдингов. На сайте декларируется учет статей из 2.9 тыс. журналов и около 280 тыс. книг, находящихся в продаже. Имеется собственная ИПС, использование которой не требует регистрации пользователей. Рубрикация научных журналов по отраслям достаточно удобная.

Отдельно отметим сайт **SpringerOpen**, где сконцентрированы статьи из ограниченного количест-

ва журналов, находящихся в открытом доступе (www.springeropen.com). На этом сайте также есть удобный рубрикатор по отраслям наук.

В табл. 2 и табл. 3 приводятся вебметрические показатели для сайтов-агрегаторов категории «К1».

По табл. 2 можно сделать вывод, что с применением описанных средств оценить объемы большинства рассматриваемых сайтов-агрегаторов фактически не удастся. Причины: наличие парольных доступов к сайтам; чрезмерно высокая длительность работы программных средств для сайтов большого объема. При этом учтем, что БД с НИ, используемые этими и другими рассматриваемыми далее Интернет-ресурсами, могут располагаться и вне самих сайтов-агрегаторов.

Таблица 2

Показатели для сайтов агрегаторов категории «К1» (первая часть)

№	Название ресурса	АТ, сек.	Count, URLs			Size, MB	SG
			Text/html	Image	Application		
1	<i>Scopus</i>	1,17	1	0	0	0	0
2	<i>Web of Science</i>	0,83	1	0	0	0	0
3	<i>Clarivate Analytics</i> https://clarivate.com/products/web-of-science/	5,83	8	96	8	1,03	0
	<i>Clarivate Analytics</i> http://wokinfo.com/russian/	0,37	660	187	135	250	0
4	<i>Springer</i>	0,33	>1000000*	**	**	**	89
5	<i>SpringerOpen</i>	0,44	7983	310	48962	70766	18

SG – количество объектов (материалов) индексируемых в *Scholar Google*;

* Из-за большого объема обрабатываемых данных более точную величину получить не удалось;

** Из-за большого объема обрабатываемых данных значение величины получить не удалось.

Таблица 3

Показатели для сайтов-агрегаторов категории «К1» (вторая часть)

Название ресурса		Абсолютное кол-во ссылок				Кол-во уни- кальных посетителей за месяц	Кол-во просмотров за месяц	AVD, чч:мм:сс
		Вх	Вн	Исх				
				Ω	Ψ			
1	Scopus	7433278	0	0	0	346906	1387620	00:05:23
2	Web of Science	1632266	0	0	0	497556	1990230	00:07:25
3	Clarivate Analytics https://clarivate.com/products/ web-of-science/	65590	40	254	1702	***	***	***
	Clarivate Analytics http://wokinfo.com/russian/	180098	6680	614	6165	***	***	***
4	Springer	1615232 5	**	**	**	1918445	7673790	00:02:37
5	SpringerOpen	281455	180455	49665	196623	126233	504930	00:01:40

** Из-за большого объема обрабатываемых данных значение величины получить не удалось.

*** Величины могут быть измерены только для сайта в целом.

По табл. 3 можно сделать следующие выводы:

1) очень высокие вебометрические показатели для многих приведенных сайтов определяются тем, что это сайты «мирового» уровня, обслуживающие пользователей, владеющих английским языком; 2) некоторые ВМП рассматриваемых сайтов предложенная нами методика оценить не в состоянии (из-за слишком большой продолжительности получения ВМП); 3) значения *AVD* для сайтов *Scopus* и *Web of Science* значительно превышают аналогичные показатели для *Springer* и *SpringerOpen*. В то же время количество уникальных посетителей для *Springer* значительно больше, чем для *Scopus* и *Web of Science*, вместе взятых. Такая ситуация может быть связана не только с различиями между объемами информации, находящейся в платном и бесплатном доступе на сравниваемых ресурсах, но и с их разной известностью у пользователей; 4) к сожалению, получить с помощью общедоступных программ статистику для этих ресурсов по пользователям только из России не представляется возможным.

ЗАРУБЕЖНЫЕ ПОЛИТЕМАТИЧЕСКИЕ ХРАНИЛИЩА НАУЧНОЙ ИНФОРМАЦИИ, НЕ ПРИЗНАВАЕМЫЕ ВАК РОССИИ

Представим лишь те ресурсы, которые можно считать основными, их отбор носит в определенной степени субъективный характер.

DOAJ (*Directory of open access journals* – сайт <https://doaj.org/>). Для включения материалов, публикуемых в журналах (включая и российские), в этот ресурс редакции изданий должны предпринимать определенные дополнительные действия и вносить определенную плату. Поэтому некоторые российские журналы, в том числе и входящие в списки ВАК, не стремятся к отражению на этом ресурсе публикуемых в них статей.

В последнее время услуга получения кода *DOI* для статей в отдельных журналах начала предлагаться авторам работ за отдельную плату. В то же время в зарубежной практике размещение научных публикаций в *DOAJ* (с присвоением им кода *DOI*) широко распространено. Считается, что потенциально это может значительно увеличивать количество ссылок на них и, как следствие, улучшать наукометрические показатели для журналов и авторов. Подчеркнем, что включение публикуемых работ в *DOAJ* возможно только для журналов с бесплатным (открытым) доступом к публикуемым материалам.

Wilson Company (сайт <http://www.hwwilson.com/>). По сведениям на сайте http://www.nlr.ru/res/inv/ic_www/cat_show.php?p=2&rid=77 *Wilson Company* предлагает 39 полнотекстовых и 25 библиографических баз данных (в том числе 14 реферативных). Поэтому данный ресурс рассматривается нами в рамках категории «К2». Однако при попытке открытия сайта <http://www.hwwilson.com/> пользователь автоматически перенаправляется на <https://www.ebsco.com/products/research-databases/h-w-wilson-databases>.

INSPEC (<http://www.theiet.org/resources/inspec>). Содержит рефераты публикаций из более чем 3,5 тыс. журналов. Включает более 8 млн. записей;

охватывает 2 тыс. трудов конференций, большое количество книг, диссертаций и отчетов.

Index Copernicus (сайт <http://indexcopernicus.com>). Польская наукометрическая база данных, рассматриваемая как международная научная платформа, предназначенная для поддержки научного сотрудничества. Допускает инициативное размещение научных материалов отдельными авторами и журналами.

ProQuest (<http://www.proquest.com/>). Ресурс считается ориентированным на поддержку деятельности библиотек, издателей научной литературы, а также распространителей этих изданий. В БД этого ресурса содержатся научные статьи, электронные книги, а также более 3 млн. диссертаций. Интерфейс ресурса может переключаться на ряд языков, но русского среди них нет.

ResearchBib (<https://www.researchbib.com/>). Имеет подзаголовок *Academic Resource Index*. Содержит архивы статей на нескольких языках. Индексирует научные журналы, материалы научных конференций, сведения о вакансиях для научных работников, о возможностях стажировок.

Science.gov (<https://www.science.gov/>). Сайт США с «правительственной» научной информацией и результатами исследований, выполненных на средства федерального бюджета. Работает с 60 базами данных и 2,2 тыс. сайтов. Обеспечивает доступ к более чем 200 млн. страниц. Может рассматриваться как надстроечный по отношению к другим ресурсам, которые отражены на нем.

Elsevier (<http://www.sciencedirect.com/>). Политематические ресурсы *Science Direct*.

Elsevier (<http://www.hub.sciverse.com/>). Политематические ресурсы *Science Hub Diverse*.

Pecypc <http://www.scienceresearch.com>. Может рассматриваться и как поисковая система по научным публикациям *Science Research*.

LAR (<http://www.thelar.org/2015-01-30-20-08-08/detailed-information-about-lar-journals>). Политематическая БД по научным журналам и статьям в них, включая находящиеся в открытом доступе. В табл. 4 приводятся данные для ресурса в целом, т.е. для <http://www.thelar.org>.

Cogprints (<http://cogprints.org>). Ресурс с материалами открытого доступа.

SCI-HUB (<https://sci-hub.io>). По сведениям, приведенным на самом ресурсе, он обеспечивает доступ к 62 млн научных статей. Создатель ресурса принципиально отрицает право интеллектуальной собственности и копирайта для научно-образовательной информации, ссылаясь на экономическую недоступность научных статей, находящихся в платном доступе, по крайней мере, для студентов. Особенность внутренней ИПС ресурса – можно вводить *URL* статьи, *PMI/DOI* или строку для поиска. На сайте предлагается также скачать и установить специальное расширение для *Google Chrome*. Ресурс работает с пиратской библиотекой *Library Genesis*.

Library Genesis (<http://libgen.io>). На сайте размещена поисковая система и онлайн-хранилище, предоставляющее бесплатный доступ к различным материалам, причем не только научного характера.

Мы не рассматриваем в рамках категории «К2» *Scholar Google*, так как считаем ее специализированной поисковой системой по научным материалам, доступным в Интернете.

Все показатели для сайтов категории «К2» сгруппированы в табл. 4 и 5.

Весьма «скромные» оценки объемов большинства ресурсов, сведения о которых включены в табл. 4, могут говорить о том, что БД, с которыми они работают, располагаются на каких-то «внешних» источниках. Специально отметим, что SG обнаруживает большое количество объектов только в одном случае (сайт www.sciencedirect.com).

Таблица 4

Показатели для сайтов-агрегаторов категории «К2» (первая часть)

№	Название ресурса	АТ, сек.	Count, URLs			Size, Mb	SG
			Text/html	Image	Application		
1	<i>Directory of Open Access Journals</i>	0,93	47	55	12	5,5	0
2	<i>Wilson Company</i>	5,6	2	8	0	0,039	0
3	<i>INSPEC</i>	0,4	3	12	0	0,189	0
4	<i>Index Copernicus</i>	0,88	136	72	2	21,1	22
5	<i>ProQuest</i>	0,7	5962	2	0	226	0
6	<i>ResearchBib</i>	2,4	15	6	0	0,2	0
7	<i>Science.gov</i>	1,11	118	210	58	69	1
8	<i>www.sciencedirect.com</i>	0,7	79	0	0	2	9550000
9	<i>www.hub.sciverse.com</i>	0,54	2	1	0	0,005	0
10	<i>www.scienceresearch.com</i>	0,33	33	89	1	5,2	0
11	<i>LAR</i>	3,78	116	29	0	4,3	0
12	<i>Cogprints</i>	0,38	195	21	3	12,5	1750
13	<i>SCI-HUB</i>	0,19	2	17	1	0,55	0
14	<i>Library Genesis</i>	0,67	160	9	29	195	6

Таблица 5

Показатели для сайтов агрегаторов категории «К2» (вторая часть)

№	Название ресурса	Абсолютное к-во ссылок				Количество уникальных посетителей за месяц	Количество просмотров за месяц	AVD, чч:мм:сс
		Вх	Вн	Исх				
				Ω	Ψ			
1	Directory of Open Access Journals	3581433	554	471	763	79272	317100	00:04:06
2	Wilson Company	8474	1	111	111	****	****	****
3	INSPEC	21753	2	9	9	****	****	****
4	Index Copernicus	986019	2575	22	350	40499	162000	00:05:12
5	ProQuest	3542085	129795	11329	80253	543553	2174220	00:04:01
6	ResearchBib	104156	175	65	205	26750	107010	00:02:53
7	Science.gov	171925	571	1237	1483	63343	253380	00:00:48
8	www.sciencedirect.com	25891818	130	203	776	2347970	9391890	00:03:45
9	www.hub.sciverse.com	7336	1	4	4	8574	34290	00:00:10
10	www.scienceresearch.com	8324	288	132	195	9763	39060	00:05:49
11	LAR	303	3923	42	315	****	****	****
12	Cogprints	69007	1785	39	544	20388	81540	00:00:54
12	SCI-HUB	96094	2	6	6	307902	1231620	00:05:50
13	Library Genesis	323244	1818	1768	4452	721243	2974800	00:07:37

**** У используемых авторами программных средств для получения ВМП нет доступа к статистике по сайту.

Из результатов, приведенных в табл. 5, можно сделать следующие выводы: 1) *AVD* для рассматриваемых ресурсов значительно различаются; 2) для ресурсов открытого доступа *AVD* меньше, чем аналогичные показатели для *Springer*; того же порядка, что и для *Scopus*, *WoS*; 3) наибольшие *AVD* имеют место для «пиратских ресурсов»; 4) самое высокое «количество просмотров» за месяц среди сайтов категории «К2» имеет ресурс www.sciencedirect.com (даже больше, чем у *Springer*); 5) однако и для «пиратских ресурсов» количества просмотров очень велики и значительно превышают, например, то, что имеет место для *Directory of Open Access Journals (DOAJ)*.

ОСНОВНЫЕ РОССИЙСКИЕ РЕСУРСЫ ПОЛИТЕМАТИЧЕСКОЙ НАУЧНОЙ ИНФОРМАЦИИ

Подчеркнем, что многие рассматриваемые в этом разделе ресурсы категории «К3» могут быть отнесены и к другим категориям. Однако в силу особой роли или функций эти ресурсы рассматриваются именно в данном разделе.

1. Общий сайт российской НЭБ [40] и РИНЦ находится по адресу elibrary.ru. Ресурс является полифункциональным и включает: БД научных материалов; внутреннюю ИПС, персональные сведения по исследователям; результаты расчетов наукометрических показателей. Для значительной части изданий, размещенных в НЭБ, обеспечивается возможность доступа к подробной информации по опубликованным материалам – либо на самом сайте, либо с переходом с него по гиперссылкам на внешние интернет-ресурсы. Для получения доступа к полным текстам опубликованных статей и решения некоторых иных задач, необходима регистрация пользователей на сайте ресурса. Достоинством сайта являются подробные наукометрические показатели, рассчитываемые РИНЦ для отдельных журналов; для их групп, издаваемых определенными организациями; для отдельных авторов. Приводятся и некоторые показатели альтметрики для отдельных публикаций, но не журналов. На стартовой странице сайта отражается количество пользователей, работающих на нем в текущий момент (обычно от нескольких тысяч до первых нескольких десятков тысяч человек). В последнее время РИНЦ активно занимается улучшением качества контента [41], исключением из числа индексируемых изданий тех, которые не рецензируют поступающие материалы, публикуют работы низкого качества. Для русскоязычных исследователей этот сайт сейчас играет важнейшую роль, особенно при решении задач прикладной наукометрии [42].

2. Электронная библиотека КиберЛенинка (www.cyberleninka.ru). Обеспечивает бесплатный доступ к опубликованным в России и странах ближнего зарубежья статьям, включая журналы из списков ВАК «С1» и «С2». Имеет развитую ИПС. Материалы частично дублируются с elibrary.ru. КиберЛенинка по информации на ее сайте – третья в мире электронная библиотека по степени «видимости» материалов в Google Scholar (http://repositories.webometrics.info/en/top_portals?sort=asc&order=scholar).

3. Ресурс ВИНИТИ, содержащий каталоги НТИ, располагается по адресу www2.viniti.ru. На ресурсе может быть организован поиск различных видов информации. Базы данных, в основном, реферативные. К части материалов доступ возможен только по логину/пароллю, предоставляемому в читальных залах ВИНИТИ или при заключения внешних организациями договоров о платном дистанционном доступе. Это один из немногих сайтов, который по результатам поиска выдает времена обработки поисковых запросов.

4. На сайте ГПНТБ России располагается ряд каталогов, в том числе «электронный каталог» по адресу <http://library2.gpntb.ru/>.

5. Большая научная библиотека (БНБ) размещена на сайте <http://sci-lib.com/> и содержит в основном информацию по естественным наукам и технике. Особенность данного ресурса – наличие форума, предназначенного для обсуждения различных научных проблем.

6. Ресурс свободного доступа «Электронные летописи Российской книжной палаты» (gbu.bookchamber.ru). На нем отражаются не только книги, но и журнальные/газетные статьи, авторефераты диссертаций, рецензии, нотные и картографические издания.

7. Сайт <http://sciencepublic.ru/> создан специалистами Международной научно-исследовательской федерации «Общественная наука» и содержит два основных ресурса. (А) *Science Public Index* (<http://direct.sciencepublic.ru/>) – позиционируется как наукометрическая база. (Б) *Science Public Direct* (<http://direct.sciencepublic.ru/>) – позиционируется как открытая база научных материалов. Некоторыми российскими платными изданиями декларируется не только размещение материалов на *Science Public Direct*, но и присвоение *IDSP* номеров для всех материалов, помещаемых в «*SP library*». На id.sciencepublic.ru последняя характеризуется как международная научная электронная библиотека. В связи с этим отметим, что *SP library* опознается поисковыми системами Интернета и как сайт Сингапурской политехнической библиотеки (<http://library.sp.edu.sg/>).

8. Межрегиональная аналитическая роспись статей МАРС (<http://arbicon.ru>) – проект Ассоциации региональных библиотечных консорциумов (АРБИКОН) – межведомственной межрегиональной библиотечной сети, объединяющей 220 библиотек-участниц различных систем и ведомств. База данных включает 2,1 млн статей из более чем 2 тыс. российских журналов, начиная с 2001 г.

9. НЭИКОН-архив (<http://archive.neicon.ru>) – обеспечивает доступ к архивам научных журналов некоторых западных (зарубежных) издательств, причем за достаточно давние годы. Оценить степень дублирования материалов по отношению к elibrary.ru достаточно сложно.

Вебометрические показатели для сайтов категории «К3» сведены в табл. 8 и 9.

Результаты табл. 8 показывают: 1) объемы информации на первых двух сайтах достаточно высоки и сопоставимы с ресурсами, рассмотренными в предыдущих разделах (категорий «К1» и «К2»); 2) нам не удалось с использованием примененных методов

оценить, какая часть размещенных на сайтах материалов русскоязычная, а какая – англоязычная. При этом англоязычными могут быть не только материалы зарубежного происхождения, но и статьи российских исследователей, а также переводы русскоязычных статей; 3) для сайта ВИНТИ анализ объемов информации на основе используемых программных средств фактически оказывается невозможным. Обращает на себя внимание также весьма высокая оценка длительности открытия стартовой страницы этого сайта. Возможно, для пользователей только из России это время является значительно меньшим.

Из приведенных в табл. 9 данных можно сделать следующие выводы: 1) у русскоязычных пользователей наиболее востребованными ресурсами являются первые два сайта, отраженные в табл. 8 и 9; 2) по количеству уникальных посетителей за месяц сайт КиберЛенинки значительно опережает сайт НЭБ/РИНЦ (elibrary.ru); 3) однако для обоих сайтов речь идет о сотнях тысяч посетителей за месяц, а не о миллионах, как для некоторых рассмотренных зарубежных сайтов. Такая ситуация отражает многократно меньшее количество русскоязычных исследователей в мире по сравнению с англоязычными. Однако, если нормировать количества посетителей сайтов на оценки количеств англоязычных и русскоязычных исследователей, то по крайней мере первые два российских сайта категории «КЗ» (из табл. 8 и 9) будут иметь даже некоторое преимущество по сравнению с функционально аналогичными им зарубежными сайтами-агрегаторами; 4) для сайтов НЭБ/РИНЦ и КиберЛенинка количество просмотров значительно больше, чем количество уникальных посетителей. Поэтому (а также с учетом *AVD*) можно сделать вывод, что многие исследователи пользуются этими ресурсами на регулярной основе; 5) для сайта НЭБ/РИНЦ (elibrary.ru) количество уникальных посетителей за месяц (266829) составило немногим менее половины от количества авторов публикаций, проиндексированных РИНЦ (578495 чел.). Какова доля авторов работ среди уникальных посетителей сайта? Определить это с помощью программных

средств, доступных для внешних пользователей сайта, нельзя. Однако с учетом того, что в количестве авторов РИНЦ учтены и уже умершие исследователи, можно считать, что востребованность этого сайта у живых авторов достаточно высока; 6) вопрос о том, влияет ли сезонность отпусков преподавателей вузов и научных работников России на показатели посещаемости сайтов НЭБ/РИНЦ, КиберЛенинки и др. нуждается в дополнительном изучении; 7) показатель *AVD* сайта elibrary.ru более чем в два раза превышает *AVD* для сайтов КиберЛенинки и ВИНТИ (для последних *AVD* примерно одинаковы). Это может свидетельствовать о большем объеме работы, выполняемом пользователями на elibrary.ru за один вход на сайт (просмотр), чем на КиберЛенинке и сайте ВИНТИ; 8) показатель *AVD* сайта elibrary.ru несколько выше чем для *Scopus* и лишь немногим уступает аналогичному показателю для *WoS*. Это можно интерпретировать таким образом, что технологии работы с научной информацией на сопоставляемых ресурсах примерно аналогичны для зарубежных и российских пользователей; 9) малое количество внутренних и исходящих ссылок для сайта elibrary.ru можно считать следствием непрозрачности этой информации для использованных авторами настоящей статьи программных средств; 10) по количеству входящих ссылок КиберЛенинка опережает elibrary.ru почти на порядок; 11) для сайта ВИНТИ количество входящих ссылок достаточно скромное – с учетом того, какую роль призван играть этот ресурс в российской науке; 12) для БНБ большинство ВМП весьма высоки, однако *AVD* составляет всего 42 сек.; 13) относительно высокая востребованность ресурса МАРС может определяться регулярными служебными потребностями работников многочисленных библиотек России; 14) обращают на себя внимание также достаточно высокие значения большинства показателей для НЭИКОН-архива. Однако *AVD* этого ресурса низкое – 58 сек. Таким образом, по соотношению рассматриваемых показателей этот ресурс близок к БНБ.

Таблица 8

Показатели для сайтов-агрегаторов категории «КЗ» (первая часть)

№	Название ресурса	AT, сек.	Count, URLs			Size, MB	SG
			Text/html	Image	Application		
1	Сайт elibrary.ru	1,28	4	9	0	0,106	1670000
2	КиберЛенинка	0,72	947921	2091	725217	7260891	1270000
3	ВИНТИ	14,94	107	107	187	252	1
4	Сайт http://library2.gpntb.ru	0,66	2156	10	0	73	0
5	Большая научная библиотека	0,37	348907	71	0	12492	0
6	Электронные летописи РКП	0,51	27	156	2	3,7	0
7	Сайт sciencepublic.ru	2,43	57	21	3	7,6	50
8	МАРС	2,54	548	439	59	237	156
9	НЭИКОН архив	0,86	62779	16	0	2811	110

Показатели для сайтов агрегаторов категории «КЗ» (вторая часть)

№	Название ресурса	Абсолютное к-во ссылок				Кол-во уникальных посетителей за месяц	Кол-во просмотров за месяц	AVD, чч:мм:сс
		Вх	Вн	Исх				
				Ω	Ψ			
1	Caïm elibrary.ru	191615	4	0	0	266829	1067310	00:06:23
2	КиберЛенинка	1188351	3073859	4646	4742015	366829	1467330	00:02:31
3	ВИНИТИ	55214	3941	203	684	8736	34950	00:2:46
4	http://library2.gpntb.ru/	16	9580	135	1571	21691	86760	00:01:56
5	Большая научная библиотека	751643	71509137	73916	6442776	4906	13048	00:00:42
6	Электронные летописи РКП	1082	43	4	8	13111	52440	00:02:23
7	sciencepublic.ru	479	666	37	292	3708	14820	00:02:54
8	МАРС	210371	4493	851	2607	9051	36210	00:03:05
9	НЭИКОН-архив	218698	1404408	14	391933	11315	21605	00:00:58

ВЫВОДЫ

1. Показано, что в условиях развития информационно-телекоммуникационных технологий сайты-агрегаторы научной информации играют ключевую роль в обеспечении ее доступности для заинтересованных лиц и организаций.

2. Предложена оригинальная классификационная схема для сайтов-агрегаторов.

3. Показаны недостатки существующих подходов к оценкам показателей сайтов-агрегаторов, а также построения интегральных (обобщенных) оценок по совокупностям этих показателей.

4. Представлена методика дистанционной оценки частных вебметрических показателей сайтов, использующая бесплатные программы и общедоступные сервисы Интернета.

5. Оценены вебметрические показатели трех категорий сайтов-агрегаторов: зарубежных политематических систем учета цитирований и оценки наукометрических показателей, признаваемых ВАК; зарубежных реферативных и полнотекстовых баз научных материалов, не признаваемых ВАК; основных российских политематических сайтов-агрегаторов научной информации.

5. Сопоставление полученных вебметрических показателей позволило сделать выводы в отношении популярности зарубежных и российских сайтов; востребованности размещенных на них материалов, а также особенностей работы пользователей с ресурсами научной информации на рассмотренных сайтах-агрегаторах.

СПИСОК ЛИТЕРАТУРЫ

1. Брумштейн Ю.М. Анализ влияния информационно-коммуникационных технологий на структуру создаваемой в России научно-технической информации // Научно-техническая информация. Сер. 1. – 2016. – № 12. – С. 7-17; Brumshtein Yu.M. An Analysis of the Impact of Information and Communication Technology on the

Structure of Scientific and Technical Information in Russia // Scientific and Technical Information Processing. – 2016. – Vol. 43, № 4. – P. 257-267.

2. Лопатина Н.В. Информационная инфраструктура общества: проблемы изучения и управления // Научно-техническая информация. Сер. 1. – 2016. – № 5 – С.1-4; Lopatina N.V. The Information Infrastructure of Society: Problems of Research and Management // Scientific and Technical Information Processing. – 2016. – Vol. 43, № 2. – P. 95-98.
3. Сюнтюренько О.В. Информационное обеспечение: факторы развития, управление, эффективность // Научно-техническая информация. Сер. 2. – 2016. – №6. – С. 7-15; Syuntyurenko O.V. The Evolutionary Factors, Management, and Efficiency of Information Support // Automatic Documentation and Mathematical Linguistics. – 2016. – Vol. 50, № 3. – P. 117-125.
4. Ильясов Б.Г., Герасимова И.Б., Карамзина А.Г., Суяргулова Ю.Р. Модель накопления знаний в научной школе в результате самоорганизации процесса обмена информацией // Труды СПИИРАН. – 2016. – № 4 (47). – С. 144-162.
5. Зацман И.М. Электронные библиотеки научных документов в Интернет: структуризация, формальное описание и поиск невербальной информации // Научно-техническая информация. Сер. 2. – 1998. – № 11. – С. 12-18; Zatsman I.M. Internet electronic scientific document libraries: structure, formal descriptin, and nonverbal information retrieval // Automatic Documentation and Mathematical Linguistics. – 1998. – Vol. 32, № 6. – P. 14-21.
6. Зелепухина В.А. Проблема достоверности и объективности информации внутри научного интернет-сообщества, построенного на принципах WEB 2.0 // Прикаспийский журнал: управление и высокие технологии. – 2013. – № 4. – С. 157-165.
7. Общедоступные ресурсы: государственные инициативы и практические решения // Университетская книга. – 2012. – № 4 (май). – С. 20-27.

8. Гуров А.Н., Гончарова Ю.Г., Бубякин Г.Б. Открытый доступ к научным знаниям: состояние, проблемы, перспективы развития // Научно-техническая информация. Сер. 1. – 2016. – № 4. – С. 10-16; Gurov A.N., Goncharova Yu.G., Bubyakin G.B. Open Access to Scientific Knowledge: Its State, Problems, and Prospects of Development // Scientific and Technical Information Processing. – 2016. – Vol. 43, № 2. – P. 88-94.
9. Прокудин Д.Е., Чугунов А.В. Сетевые электронные репозитории, системы агрегации и сохранения в концепции свободного доступа к научной информации / В сборнике: Интегрированные цифровые ресурсы: организационно-технологические и научно-методические основы развития. Сборник научных трудов. Сер. «Электронная библиотека» Президентская библиотека имени Б.Н. Ельцина; науч. ред. Е.Д. Жабко. – СПб, 2015. – С. 162-183.
10. Björk, Bo-Christer Open Access to Scientific Publications – An Analysis of the Barriers to Change? // Information Research. – 2004. – Vol.9, № 2. – P. 170. – URL: <http://hdl.handle.net/10227/647> (Date of access 04.07.2017).
11. Ларук О., Гаранович М.В. Оценка доступа к научной информации для академических пользователей в Интернете // Глобальный научный потенциал. – 2014. – № 9 (42). – С. 135-138.
12. Прохорова А.М. Понятие юзабилити сайта: показатели и стандарты проектирования // Новый университет. Серия: Экономика и право. – 2016. – № 9 (67). – С. 87-90.
13. Лайкова А.А. Юзабилити сайта: принципы и методы оценки // Актуальные проблемы авиации и космонавтики. – 2016. – Т. 2, № 12. – С. 745-747
14. Егоров В.С. Особенности доступа к научной литературе в электронном обществе: история вопроса // Научно-техническая информация. Сер. 1. – 2012. – № 11. – С. 18-23.
15. Арский Ю.М., Быков В.А. Деятельность ВИНТИ РАН – базовой организации государств – участников СНГ по межгосударственному обмену научно-технической информацией // Научно-техническая информация. Сер. 1. – 2013. – № 5. – С. 23-30; Arskii Yu.M., Bykov V.A. The Activities of VINITI RAS as a Key Organization of the Commonwealth of Independent States for the Exchange of Scientific and Technological Information // Scientific and Technical Information Processing. – 2013. – Vol. 40, № 2. – P. 101-108.
16. Калинин Ю.П., Хорошилов А.А., Хорошилов А.А. Принципы создания системы мониторинга и анализа мирового потока научно-технической информации // Системы и средства информатики. – 2016. – Т. 26, № 1. – С. 139-165.
17. Мельникова Е.В. Расширение функций современной системы НТИ России в контексте ее совершенствования и развития инновационной направленности // Научно-техническая информация. Сер. 1. – 2014. – № 2. – С. 9-13; Mel'nikova E.V. Enhanced Functions of the Modern STI System in Russia in the Context of its Improvement and of the Development of the Innovative Orientation // Scientific and Technical Information Processing. – 2014. – Vol. 41, № 1. – P. 29-32.
18. Брумштейн Ю.М., Васьковский Е.Ю., Куаншкалиев Т.Х. Поиск информации в Интернете: анализ влияющих факторов и моделей поведения пользователей // Известия Волгоградского государственного технического университета. – 2017. – № 1 (196). – С. 50-55.
19. Биктимиров М.Р., Глебский В.Л., Долгов Б.В., Поликарпов С.А. Использование информационных технологий и инфраструктур для агрегации научной информации. Опыт Канады, Нидерландов, США // Моделирование и анализ информационных систем. – 2015. – Т. 22, № 1. – С. 114-126.
20. Лягушкина Е.А., Богатов В.В. Анализ результативности фундаментальных научных исследований: публикационный аспект // Интернет-журнал «НАУКОВЕДЕНИЕ». – 2017. – Том 9, № 2. – URL: <http://naukovedenie.ru/PDF/10EVN217.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.
21. Редькина Н.С. Направления развития инструментов веб-аналитики // Научно-техническая информация. Сер. 2. – 2017. – № 5. – С. 5-10.
22. Васьковский Е.Ю., Брумштейн Ю.М. Системный анализ функциональных возможностей счетчиков посещаемости сайтов // Прикаспийский журнал: управление и высокие технологии. – 2015. – №3. – С. 45-58
23. Гуськов А.Е., Быховцев Е.С., Косяков Д.В. Альтернативная вебометрика: исследование веб-трафика сайтов научных организаций // Научно-техническая информация. Сер. 1. – 2015. – № 12. – С. 12-28; Guskov A.E., Bykhovtsev E.S., Kosyakov D.V. Alternative Webometrics: Study of the Traffic of the Websites of Scientific Organizations // Scientific and Technical Information Processing. – 2015. – Vol. 42, № 4. – P. 274-289.
24. Thelwall Mike. Data Science Altmetrics // Journal of Data and Information Science. – 2016. – Vol. 1, № 2. – P. 7-12 (DOI: 10.20309/jdis.201610)
25. Torres-Salinas Daniel, Cabezas-Clavijo Alvaro, Jimenez-Contreras Evaristo Altmetrics: New Indicators for Scientific Communication in Web 2.0 /arXiv:1306.6595v1 [cs.DL] DOI: 10.3916/C41-2013-05; 27 Jun 2013 18:20:32 GMT (515kb)
26. Zahedi Zohreh , Costas Rodrigo , Wouters Paul. How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications /arXiv:1404.1301v1 [cs.DL], DOI: 10.1007/s11192-014-1264-0; Fri, 21 Feb 2014 10:01:48 GMT (368kb)
27. Голицына О.Л., Куприянов В.М., Максимов Н.В. Информационные и технологические решения в задачах управления знаниями // Научно-техническая информация. Сер. 1. – 2015. – № 8. – С. 1-12; Golitsina O.L., Kupriyanov V.M., Maksimov N.V. Information and Technological Solutions Applied for Knowledge-Management Tasks // Scientific and Technical Information Processing. – 2015. – Vol. 42, № 3. – P. 150-161.

28. Третьякова О.В., Кабакова Е.А. Возможности использования вебметрического анализа в оценке сайта научного института // Вопросы территориального развития. – 2014. – № 2 (12). – С. 5.
29. Шокин Ю.И., Веснин А.Ю., Добрынин А.А., Клименко О.А., Рычкова Е.В. Математическое моделирование информационных процессов в веб-пространстве // Вычислительные и информационные технологии в науке, технике и образовании (СITech-2015). – 2015. – С. 300-315.
30. Косяков Д.В., Гуськов А.Е., Быховцев Е.С. Академические институты России в зеркале вебметрики // Вестник российской академии наук. – 2016. – Т. 86, № 11. – С. 1015–1025
31. Иванова В.И. Перевод научных публикаций как средство популяризации результатов исследований в глобальном мире // Русский язык и культура в зеркале перевода. – 2015. – № 1. – С. 207-215.
32. Семенюк Э.П. Глобализация информационного пространства и человечество // Научно-техническая информация. Сер. 1. – 2015. – № 1. – С. 1-13; Semeniyuk E.P. Globalization of the Information Area and its Significance for Mankind // Scientific and Technical Information Processing. – 2015. – Vol. 42, № 1. – P. 1-12.
33. Сковородина И.С., Ефремов П.В. Технологии создания, агрегации и использования научного и образовательного контента // Научная периодика: проблемы и решения. – 2013. – № 6 (18). – С. 31-34.
34. Аксентьева М.С., Кириллова О.В., Москалева О.В. К вопросу цитирования в Web of Science и Scopus статей из российских журналов, имеющих переводные версии // Научная периодика: проблемы и решения. – 2013. – № 4 (16). – С. 4-18.
35. Halevia Gali, <http://www.sciencedirect.com/science/article/pii/S1751157717300676> - cor0005, Moedb Henk, Bar-Ilan Judit. Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the Literature // Journal of Informetrics. – 2017 August. – Vol. 11, Issue 3. – P. 823–834. – URL: <https://doi.org/10.1016/j.joi.2017.06.005>
36. Антопольский А.Б. О целесообразности российского национального вебметрического индекса // Научно-техническая информация. Сер. 1. – 2014. – № 2. – С. 14-18; Antopolskii A.B. On the Feasibility of the Russian National Webometric Index // Scientific and Technical Information Processing. – 2014. – Vol. 41, № 1. – P. 33-37.
37. Печников А.А., Луговая Н.Б., Чуйко Ю.В., Косинец И.Э. Разработка инструментов для вебметрических исследований гиперссылок научных сайтов //Вычислительные технологии. – 2009. – Т. 14, № 5. – С. 66-78.
38. Блеканов И.С., Сергеев С.Л., Мартыненко И.А. Построение тематико-ориентированных веб-краулеров с использованием обобщенного ядра // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. – 2012. – Т. 5, № 157. – С. 9-15.
39. Брумштейн Ю.М., Васьковский Е.Ю., Горбачева А.Н. Интернет-сайты органов исполнительной власти в прикаспийских регионах России: анализ номенклатуры, функциональности и востребованности ресурсов // Прикаспийский журнал: управление и высокие технологии. – 2016. – № 4 (36). – С. 52-68.
40. Осипова В.А. НЭБ: История, устройство, новый этап развития // Университетская книга. – 2013. – № 10. – С. 64-66.
41. Еременко Г.О. Elibrary.ru: курс на повышение качества контента // Университетская книга. – 2016. – № 3. – С. 62-68.
42. Юрков А.В. Прикладная наукометрия: elibrary.ru vs Академия Google // Вестник Российского университета дружбы народов. Серия: Информатизация образования. – 2015. – № 2. – С. 21-27.

Материал поступил в редакцию 02.08.2017.

Сведения об авторах

БРУМШТЕЙН Юрий Моисеевич – кандидат технических наук, доцент Астраханского государственного университета, доцент
e-mail: brum2003@mail.ru
ORCID <http://orcid.org/0000-0002-0016-7295>

ВАСЬКОВСКИЙ Евгений Юрьевич – программист 1-й категории Управления информационных технологий и телекоммуникаций Астраханского государственного университета
e-mail: vaskovskiy_evgeniy@mail.ru
ORCID <http://orcid.org/0000-0002-4937-3305>

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 81'322'367.623

Г.И. Кустова

Об одной конструкции прилагательных*

Рассматривается проблема статуса предложной группы в предложном падеже, которая обычно считается валентностью прилагательного. Показано, что прототипическое прилагательное, будучи обозначением признака (свойства) предмета или лица, является одноместным вневременным предикатом, т.е. имеет только одну валентность – носителя признака. В случаях типа Х аккуратен в посещении занятий мы имеем дело с конструкцией, полученной в результате синтаксического преобразования других конструкций: Х аккуратно посещает занятия; аккуратное посещение Х-ом занятий, т.е. прилагательное в конструкции Х аккуратен в посещении занятий – это характеристика ситуации (посещение, посещать), метонимически отнесенная к лицу в результате транспозиции наречия.

Ключевые слова: прилагательное, валентность, конструкция, предикат, синтаксическое преобразование

Глагол как прототипический предикат обозначает актуальную – притом динамическую – ситуацию, прилагательное как непрототипический предикат обозначает вневременное свойство (признак) предмета, и в этом качестве имеет одну валентность – носителя признака. Все остальные зависимые прилагательного – отклонения от его исходного статуса вневременного признака, связанные с тем, что прилагательные (в разной степени, в зависимости от семантики) могут «продвигаться» в зону прототипических предикатов – глаголов.

Прилагательное как предикатное слово может иметь два типа валентностей, которые условно можно назвать глагольными и адъективными.

Глагольными мы будем, вслед за [1] (ср. раздел «Глагольные валентности прилагательных»), называть синтаксически обязательные валентности, выражаемые падежными и предложно-падежными зависимыми.

Глагольные валентности прилагательных часто (хотя и не всегда) выражаются теми же формами, что и у соотносительных глаголов, но самое существенное их свойство – обязательность. Можно выделить три основных группы прилагательных с обязательными валентностями.

Первая группа – прилагательные с модальной и оценочной семантикой, которые максимально близки к глаголам: *должен + инф.*, *согласен (согласился) на Вин / инф.*, *готов к Дат / на Вин / инф.*; *склонен к Дат*; *годен на Вин / для Род*.

Вторая группа – прилагательные, выражающие двухместные отношения. Их валентности тоже всегда так или иначе выражаются: *Х похож на Y = X и Y похожи* (субъект во мн.ч.); *Х равен (равняется) Y-у*; ср. также: *тождественный, сходный, симметричный, аналогичный*. Эти прилагательные соответствуют двухместным стативным глаголам отношения (ср. *походить на Вин*; *равняться Дат*).

Наконец, третья группа прилагательных – это обычные прилагательные, которые имеют конструктивно обусловленное значение (по В.В. Виноградову) и, по определению, употребляются с зависимыми: *силен в математике, здоров поспать* – поскольку это значение реализуется только в соответствующей конструкции. Сюда же, по-видимому, относятся случаи типа *богат витаминами*. Эти прилагательные можно присоединить к первой группе, ср. *разбирается в математике, любит поспать (не дурак поспать), содержит витамины*.

В настоящей работе такие прилагательные анализироваться не будут.

Мы будем рассматривать «обычные» прилагательные и их специфичные, условно говоря, «адъективные» валентности, а также другие зависимые, которые являются, как мы покажем далее, результатом синтаксических преобразований. Здесь отчетливо

* Исследование выполнено в рамках проекта РНФ № 16-18-02003, осуществляемого в МПГУ. Языковые примеры извлечены из Национального корпуса русского языка (www.ruscorgo.ru), далее – Корпус.

противопоставляются главная валентность прилагательных – носитель признака, выражаемый определяемым существительным, – и другие валентности, стандартно выражаемые зависимыми формами.

Валентность носителя признака иногда называют «контрвалентностью» [2, с. 155], она обычно не рассматривается, так как с синтаксической точки зрения она невыразима. Точнее, специфическая валентность на определяемое существительное выражается нестандартным для валентности способом – существительное не подчиняется прилагательному синтаксически, а, наоборот, подчиняет его (о нестандартных способах выражения валентностей см. [3, 4]). Таким образом, эта валентность необычна, так как выражается не подчиненным, а подчиняющим элементом – стандартная же валентность выражается дополнением. Но с семантической точки зрения именно «контрвалентность» – главная валентность прилагательного как предиката (предикатного слова), без реализации которой прилагательное не может быть употреблено. И то, что прилагательное подчинено определяемому существительному, в конечном счете, не так уж парадоксально: глагол подчиняется первому актанту-подлежащему; прилагательное в предикативном употреблении тоже подчиняется главному актанту-подлежащему. А в атрибутивной функции, подчиняясь существительному, прилагательное ведет себя как атрибутивная репрезентация глагола – причастие.

Все остальные зависимые прилагательного похожи на глагольные в том отношении, что выражаются подчиненными падежными формами или предложными группами.

Анализ материала показывает, что основных типов валентностей у прилагательного, как и у глагола, несколько, и они стандартно выражаются, в основном, одними и теми же падежными или (чаще) предложно-падежными формами. К основным распространителям прилагательных можно отнести: *для Род, в Пр, к Дат, с Твор* и др., ср. *желательное для нас решение комиссии, необходимые для работы инструменты, строг в оценках; добр к детям; вежлив с клиентами*, см. [5], ср. также [6].

В работе [6] все такие распространители рассматриваются как управление, т.е. как валентности прилагательных. Мы попытаемся показать, что эти случаи неодинаковы.

Специфика прилагательного как вневременного, неактуального признака состоит, в частности, в том, что его невыраженные валентности в общем случае имеют дефолтную интерпретацию (по умолчанию). Так, валентность субъекта (в широком смысле – экспериенцера, субъекта оценки, стандарта, см. [5, 6]), если она не выражена, интерпретируется как ‘для любого X-а’: *интересный фильм* – ‘для всех, кто будет смотреть’. Выраженная валентность функционирует как ограничитель: *интересный для детей фильм* – не для всех.

ПРИМЕЧАНИЕ. Другой вариант – референтная интерпретация валентности: она не выражена, так как вычисляется на базе материала предложения или известна из текста:

На важные встречи он обычно надевает галстук – ‘важные для него встречи’.

В некоторых случаях, если обозначение субъекта-экспериенцера отсутствует, предложение не становится неправильным, просто уменьшается количество информации: *Язык для формулирования поискового запроса должен быть простым и удобным для пользователя* [«Информационные технологии», 2004]. При отсутствии указания на субъект – *Язык для формулирования поискового запроса должен быть простым и удобным* – действует дефолтная интерпретация: ‘для релевантной в данной ситуации группы субъектов’.

Аналогично ведет себя распространитель со значением аспекта (по Дат): *Ньютону хоть яблоко на голову упало — предмет эстетичный по форме и аппетитный по содержанию* [«Техника - молодежи», 1977] (предложение без форм по Дат: *Ньютону хоть яблоко на голову упало — предмет эстетичный и аппетитный* – остается правильным), а также некоторые другие распространители, ср. к Дат: *А вода считается более агрессивной к человеку средой, чем даже космос* [«Сочи», 2002.08.22] – *Вода считается более агрессивной средой, чем космос* = ‘вода отрицательно воздействует на какие-то объекты’.

Забегая вперед, заметим, что в данном случае речь идет об одном и том же – вневременном – признаке, и наличие или отсутствие валентностей на это не влияет.

Однако бывают ситуации, когда отсутствие ограничителей может привести к искажению смысла или даже фактов: *Odette была совершенно безжалостна в своей оценке французской юстиции* [Г. А. Газданов. История одного путешествия (1938)] – предложение *Одетта была безжалостна* имеет совершенно другой смысл и не соответствует исходному: это утверждение о том, что Одетта – безжалостный человек (это черта характера, свойство личности), Одетта не проявляет жалости к другим людям. В исходном же предложении речь шла только о безжалостных оценках, хотя предикат *безжалостна* относился к лицу.

Далее мы будем рассматривать семантические и синтаксические особенности конструкции *А в Пр*, где *А* – прилагательное (нередко выступающее в краткой форме), *в Пр* – зависимая от него форма предложного падежа с предлогом *в*. Но сначала дадим краткий обзор конструкций *А в Пр*, чтобы выделить наш материал, поскольку форма *в Пр* скрывает за собой несколько разных значений и синтаксических статусов:

в Пр – глагольная валентность прилагательного:

уверен в чем, сведущ в чем, бессилён в чем;

в Пр – обозначение сферы локализации признака:

Вернер фон Браун, может быть, был гениален в своей области [Евгений Евтушенко. Ягодные места (1982)]; *Даже компетентные в военном деле современники затруднялись разумно объяснить такое поведение русского военачальника* [«Знание-сила», 2012];

в Пр – обстоятельство (потенциальное придаточное):

Я держу в руках этот, по-настоящему драгоценный в ситуации уже начавшегося солнечного затмения, «оптический инструмент» [«Наука и жизнь», 2009] – здесь *в Пр* обозначает обстоятельство (в данном случае – обстоятельство причины, основания), которое может быть развернуто в придаточ-

ное: Этот «оптический инструмент» драгоценен, так как уже началось солнечное затмение;

в Пр – результат преобразования некоторой исходной конструкции:

Х бескомпромиссен в оценках ← оценки Х-а бескомпромиссны, Х бесчеловечен в своих требованиях ← требования Х-а бесчеловечны.

Нас будет интересовать последний случай – преобразование исходной конструкции. Наша задача – рассмотреть характер этого преобразования и определить статус зависимой от прилагательного предложной группы **в Пр**.

Вообще, конструкцию вида *Х благоразумен в ответах* можно интерпретировать разными способами.

Первый – это два разных значения (если угодно – две ипостаси) признака *благоразумный*: одно является характеристикой личности (*благоразумный человек*), в этом случае у него нет валентностей, кроме носителя признака; второе – характеристика поведения человека в какой-то ситуации, это значение имеет валентность на ситуацию *в Пр*. Показательно, однако, что если семантические различия типа *зеленый лист* (цвет) и *зеленый виноград* («неспелый») или *твердая поверхность* и *твердое решение* в словарях фиксируются, то различия в парах *Х благоразумный человек* и *Х благоразумен в ответах* или *Х вежливый* и *Х был вежлив с клиентами* в словарях никогда не отражаются (и не считаются многозначностью): различие между *благоразумный Х* и *Х благоразумен в Пр*, скорее, аспектуальное – свойство и актуальное проявление, поведение, установка субъекта. При этом важно отметить, что *благоразумен в Пр* или *вежлив с Твор* совершенно не обязательно являются актуальными проявлениями постоянного свойства, ср. *Клиентам нравится этот вежливый юноша* vs. *Он сегодня в кои-то веки был вежлив с клиентами, обычно хамит*. А в некоторых случаях признак, который приписывается человеку в рамках конструкции **А в Пр**, как самостоятельная характеристика к человеку вообще не применим, ср.: *Второй, помолуже, плавный в движениях и во взглядах, как-то незаметно и почти неслышно устроился рядом с Нонной* [Н. Леонов, А. Макеев. Эхо дефолта (2000-2004)] = *плавные движения*, ≠ **плавный человек*; *Красивее они стали, на мой взгляд, размашистее в движении* [Татьяна Тарасова, Виталий Мелик-Карамов. Красавица и чудовище (1984-2001)] = *размашистые движения*; ≠ **размашистый человек*.

Второй – конструкция **А в Пр** – результат преобразования некоторой исходной конструкции. Суть этого преобразования заключается в том, что признак действия или результата деятельности человека метонимически смещается на самого человека:

Осторожны, двойственны в оценках бригадира и директор базы, и начальник управления – директор и начальник не являются осторожными людьми и тем более двойственными людьми. Осторожными и двойственными являются их оценки;

Х благоразумен в ответах – признак приписывается лицу, но мы не можем утверждать, что Х – благоразумный человек. Такой импликации нет. И это легко показать: *Х, обычно легкомысленный, был на удивление благоразумен в ответах*.

Другими словами, прилагательное в конструкции **А в Пр** – это не постоянная характеристика человека (свойство, черта характера), а актуальная характеристика человека в ситуации, обозначенной группой **в Пр**, т.е. характеристика действия, деятельности человека. Разумеется, не всякая характеристика действия (и не всякого действия) может метонимически смещаться на субъекта этого действия. Например, физические действия, особенно единичные, «сопротивляются» такому переносу: *Она тщательно промыла овощи теплой водой* – ? *Она тщательно в промывке / в мытье овощей*, ср., однако: *Русские иконописцы, не столько тщательно в отделке, всегда прилагают старание о рисунке и значении изображения* [В. И. Григорович. Очерк путешествия по Европейской Турции (1848)] ← *тщательно отделяют*; *Крекиш Петр — человек любопытный и тщательный в собирании российских древностей и редкостей* [Н. И. Новиков. Опыт исторического словаря о российских писателях (1772)] ← *тщательно собирает*. *Тщательный в Пр* всегда означает *тщательно делает что-то*, поскольку автономная характеристика *Х – тщательный человек* едва ли возможна.

Таким образом, в рассматриваемом нами случае **А в Пр** – это характеристика ситуации с участием Х-а и характеристика деятельности Х-а в рамках этой ситуации.

Соответственно, у конструкции **А в Пр**, в общем случае, может быть два исходных коррелята, соответствующих двум разным способам представления ситуации, а именно – (а) ситуация может быть обозначена существительным (часто, но не всегда, отглагольным), тогда ее характеризует прилагательное; (б) ситуация может быть обозначена глаголом, тогда вместо прилагательного выступает однокоренное наречие (есть и другие варианты, см. ниже):

Х благоразумен в ответах – *Х отвечал благоразумно*; *ответы Х-а благоразумны*;

Она плавна в движениях – *она плавно двигается*; *ее движения плавны*;

Они были невняты в изложении – *они невнятно изложили*; *их изложение невнятно*;

Х аккуратен в посещении занятий – *Х аккуратно посещает занятия*; *аккуратное посещение занятий*;

Блэр стал авторитарным во многих своих действиях – *Блэр действует авторитарно*; *действия Блэра авторитарны*.

Если существительное отглагольное (*действие, изложение*), обычно возможны два варианта исходной синтаксической конструкции: *авторитарные действия* – *авторитарно действовал*; *невнятно изложение* – *невнятно изложил*.

Однако нередко случаи неполной реализации данной схемы либо реализации в других конструкциях.

Иногда возможен только глагол с наречием:

бережливый в собственных расходах – *бережливо расходует [средства]*; **бережливые расходы*, ср., однако: *бережливое расходование*.

Иногда возможно только существительное:

Родители безутешны в своем горе – *безутешное горе*, выражение *безутешно горюют* в Корпусе не встретилось; впрочем, в Корпусе встречаются другие

сочетания, из которых наиболее частотны: *безутешно плакать, рыдать, скорбеть*;

деспотичны в своих требованиях – требования деспотичны, ?деспотично требуют;

амбициозный в своих планах – его планы амбициозны, ?амбициозно планирует;

*изменчивая в своих чувствах героиня – чувства изменчивы, *изменчиво чувствует, ср., однако: чувства меняются*;

*назойлив в своих подозрениях – подозрения назойливы, *назойливо подозревает*.

В приведенных примерах глагол есть, но наречно-глагольная конструкция неупотребительна. Бывают случаи, когда глагола, по тем или иным причинам, нет. Так, если существительное не отглагольное (а, например, отадективное), то глагольная конструкция, естественно, невозможна: *Сюжеты не безграничны в своем многообразии – многообразие сюжетов не безгранично*. То же в случае, когда существительное обозначает не конкретную ситуацию, а сферу деятельности: *Х авторитарен во внутренней политике = политика Х-а авторитарна*. Конкретный глагол здесь невозможен, но в конечном счете характеристика, выраженная прилагательным, относится к каким-то конкретным ситуациям: авторитарны какие-то действия, решения, относящиеся к области внутренней политики.

Иногда прилагательное преобразуется не в наречие при глаголе, а в другой глагол:

беззаботные в определении точного времени люди средневековья – люди средневековья не заботились об определении точного времени;

бесстрашны в разговорах – не боятся разговаривать [на определенные темы] (ср., однако, с неодушевленным субъектом: Бесстрашное в выводах об суждение – выводы бесстрашны);

Критики осторожны и ленивы в своих высказываниях – ленятся делать высказывания (что-то высказывать).

В случае модальной семантики всегда вместо наречия выступает модальный глагол:

*Абоненты беззащитны в борьбе в этом sms-мусором – *беззащитно борются → не могут бороться*;

беспомощен в борьбе с болезнями – не может бороться;

бессильны в наведении порядка – не могут навести порядок;

бессильны в прогнозах – не могут дать (правильный) прогноз.

Наконец, вместо наречия может употребляться предикатив:

ленивый в учении – ему лень учиться, ≠ лениво учится (ср., однако: ленивый в движениях – преобразуется по обычной схеме: лениво двигается, ленивые движения);

лёгкие в исполнении способы приготовления еды – легко исполнить;

Вещества просты в изготовлении, безопасны в применении – их просто изготовить, безопасно применять.

Впрочем, несмотря на такое разнообразие в конкретных случаях, логически здесь нужно усматри-

вать две исходные конструкции, связанные между собой, ср. *ответы Х-а дипломатичны vs. Х отвечает дипломатично*, которые в результате преобразования дают одну и ту же производную конструкцию: *Х дипломатичен в ответах*. В обоих случаях признак ситуации оказывается смещенным к субъекту.

Каждой исходной конструкции – назовем их субстантивной и глагольно-наречной – соответствует свой тип преобразования.

Если взять субстантивную конструкцию: *движения Х-а степенны – Х, степенный в движениях; поступки Х-а благоразумны – Х благоразумен в поступках* и т.п., – то перед нами процесс, аналогичный экстрапозиции посессора, ср. [7, 8] (в более широком смысле – субъекта). В случае экстрапозиции посессора, посессор, зависящий от актанта некоторого предиката, переподчиняется этому предикату: *Он поцарапал мой телефон → Он поцарапал мне телефон*. В нашем случае предикат ситуации, так сказать, переподчиняется актанту этой ситуации. С другой стороны, в преобразовании субстантивной конструкции можно усмотреть сходство с диатезой, когда предикат перемещается от иерархически более высокого участника ситуации к иерархически более низкому, который тем самым попадает в коммуникативно и синтаксически престижную позицию подлежащего, ср. *Гусар звенел шпорами – Шпоры звенели* [9].

В этом случае валентностью прилагательного является ситуация, обозначенная существительным, – *движения степенны, поступки благоразумны*, а лицо Х – валентность этого существительного и участник этой ситуации, т. е. у прилагательного есть валентность на ситуацию, а у лица Х нет соответствующего признака (нельзя утверждать *Х степенный, Х благоразумный*, основываясь только на одной конкретной ситуации). Поэтому конструкция *А в Пр* – результат смещения предиката (*благоразумный, степенный*) к Х-у, но это актуальная характеристика Х-а в пределах данной ситуации.

Если взять глагольно-наречную модель, прилагательное соответствует субъектному наречию: *Х аккуратен в посещении занятий – Х аккуратно посещает занятия*. Аккуратен можно считать характеристикой Х-а, но это его характеристика в рамках реализации ситуации «посещение занятий» – ситуативный признак, который не распространяется за пределы ситуации, не является постоянной характеристикой Х-а (т. е. здесь не передается значение ‘Х – аккуратный человек’). В позиции сказуемого *аккуратен* остается ситуативным признаком, обозначение ситуации (посещать) сохраняется и выступает в виде зависимой от прилагательного предложно-падежной формы. Таким образом, прилагательное в конструкции *А в Пр* – синтаксическая транспозиция наречия и валентность на ситуацию – валентность наречия. Прилагательное *аккуратный* применительно к лицу (*аккуратный мальчик*) не имеет валентности на ситуацию, оно само обозначает ситуацию (статическую) – свойство лица, и у такого прилагательного одна валентность – лицо-носитель признака.

Таким образом, мы возвращаемся к первой интерпретации нашей конструкции, где говорилось, что разный состав валентностей предполагает разные

значения, – но с некоторым уточнением. Речь идет не о двух разных значениях, а о двух разных единицах с разными свойствами – одна соответствует постоянной характеристике человека, другая – поведению человека в рамках некоторой ситуации, и у этой второй единицы есть валентность на ситуацию. Основной репрезентацией этой единицы является наречие, но у нее есть и другая репрезентация, совпадающая с прилагательным – что неудивительно, так как у прилагательного и наречия есть деривационная связь. Эта ситуация отнюдь не уникальна для языка, ср. предикатив, который формально совпадает с кратким прилагательным среднего рода, но имеет другие свойства и другую семантику.

Наречие – это второстепенный, дополнительный предикат лица X в ситуации «посещать занятия». В результате преобразования наречная характеристика X-а становится главным предикатом X-а, а форма прилагательного – это своего рода предикативная репрезентация наречия, она служит для того, чтобы превратить наречие в сказуемое: хотя у наречия нет своей предикативной формы, но оно может выступать в позиции основного предиката в форме прилагательного. А исходное сказуемое, которое заполняло семантическую валентность наречия, но синтаксически подчиняло его, превращается в подчиненную предложную группу.

И эта конструкция – способ поменять акценты, превратить наречие в сказуемое, т. е. в основной предикат (как бы в глагол), а глагол сделать второстепенным, подчиненным предикатом.

Такая перестановка акцентов одновременно похожа и на диатезу, когда акценты переставляются и актанты передвигаются в рамках одной ситуации, и на экстрапозицию посессора, когда в рамках более широкой, комплексной ситуации актант имени переподчиняется главному глаголу. По сути, и в том и в другом случае речь идет о повышении синтаксического, а значит, и коммуникативного, и прагматического статуса элемента исходной конструкции: в случае субстантивной конструкции повышается статус актанта существительного, в случае глагольно-наречной – статус наречия как второстепенного предиката (оно становится главным – в форме прилагательного). Здесь важно заметить, что саму субстантивную конструкцию можно считать результатом транспозиции наречно-глагольной конструкции: *аккуратно посещает* → *аккуратные посещения*.

Итак, у прототипических прилагательных (характеристики предмета или лица) нет валентности на ситуацию. Этим они отличаются от прилагательных, которые тяготеют к предикативной позиции и имеют «глагольные валентности», в том числе и валент-

ность на ситуацию, ср., *готов помочь*. Если предложные группы с субъектным значением (*неожиданный для нас* = ‘мы не ожидали’) являются валентностями самого прилагательного, то предложная группа в *Пр*, за которой скрывается ситуация, является валентностью наречия из исходной глагольно-наречной конструкции. Если же в качестве исходной рассматривать субстантивную конструкцию (*аккуратные посещения*), то прилагательное в этой конструкции не является признаком лица, а является признаком ситуации (т. е. соотносится с наречием) и тем самым имеет валентность на ситуацию.

СПИСОК ЛИТЕРАТУРЫ

1. Русская грамматика. В 2-х тт. Т. 1. – М., 1980.
2. Шаляпина З.М. Трёхмерная стратификационная модель языка и его функционирования: к общей теории лингвистических моделей. – М.: Восточная литература РАН, 2007.
3. Богуславский И.М. Исследования по синтаксической семантике. – М.: ЯСК, 1985.
4. Богуславский И.М. Сфера действия лексических единиц. – М.: ЯСК, 1996.
5. Кустова Г.И. Синтаксические зависимые прилагательных (конструкция **А для Род S**) // Труды Института русского языка им. В.В. Виноградова. Вып. 10. – М., 2016. – С. 132–149.
6. Гращенков П.В., Кобозева И.М. Семантические классы и управление прилагательных // Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог-2017». – Вып. 16 (23). Том 2. – М.: Изд-во РГГУ, 2017. – С. 123–138.
7. Кибрик А.Е. Внешний посессор в русском языке // Кибрик А.Е. Константы и переменные языка. – СПб.: Алетея, 2003. – С. 307–319.
8. Кибрик А.Е., Брыкина М.М., Леонтьев А.П., Хитров А.Н. Русские посессивные конструкции в свете корпусно-статистического исследования // Вопросы языкознания. – 2006. – № 1. – С. 16–45.
9. Падучева Е.В. Динамические модели в семантике лексики. – М.: ЯСК, 2004.

Материал поступил в редакцию 21.09.17.

Сведения об авторе

КУСТОВА Галина Ивановна – доктор филологических наук, ведущий научный сотрудник Института русского языка им. В.В. Виноградова РАН, Москва
e-mail: galinak03@gmail.com

Российская академия наук
Федеральное государственное бюджетное учреждение науки
ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК

предлагает научным работникам, аспирантам и другим специалистам в области естественных, точных и технических наук, желающим быстро и эффективно опубликовать результаты своей научной и научно-производственной деятельности, использовать способ публикации своих работ через *систему депонирования*.

Депонирование (передача на хранение) – особый метод публикации научных работ (отдельных статей, обзоров, монографий, сборников научных трудов, материалов научных конференций, симпозиумов, съездов, семинаров), разрешенных в установленном порядке к открытому опубликованию.

Подготовка и передача на депонирование научных работ происходит в соответствии с «Инструкцией о порядке депонирования научных работ по естественным, техническим, социальным и гуманитарным наукам» (М., 2014).

Депонированные научные работы находятся на хранении в депозитарии ВИНТИ РАН, копии работ предоставляются заинтересованным организациям и специалистам на бумажном и электронном носителях и являются официальной публикацией.

Информация о депонированных научных работах включается в информационные издания ВИНТИ РАН: Реферативный журнал, Базу данных и Аннотированный библиографический указатель «Депонированные научные работы».

Направить научную работу на депонирование можно, обратившись в Группу депонирования ЦНИО ВИНТИ РАН по адресу:

125190, Москва, ул. Усиевича, 20.

ВИНТИ РАН, Группа депонирования ЦНИО

Тел.: 499-155-43-28, 499-155-43-76, 499-155-42-43, Факс: 499-943-00-60,

e-mail: cnio@viniti.ru, dep@viniti.ru

С инструкцией о порядке депонирования можно ознакомиться на сайте ВИНТИ РАН:
<http://www.viniti.ru>

**Центр (Отдел) научно-информационного обслуживания
(ЦНИО) ВИНИТИ РАН**

**предлагает услуги по предоставлению
информационно-аналитических обзоров**

ВИНИТИ РАН осуществляет подготовку информационно-аналитических обзоров по инновационным и приоритетным направлениям научных исследований в области точных, естественных и технических наук. Обзоры готовятся ведущими специалистами ВИНИТИ, работающими в определенных областях науки и техники. Аналитические материалы содержат результаты анализа и обобщения информации по актуальным научным проблемам, а в некоторых случаях – и прогностические выводы. Основой для составления обзоров служит отечественная и зарубежная научно-техническая литература, доступная ВИНИТИ РАН: фонд НТЛ, включающий более 2,5 млн. отечественных и иностранных журналов, книг, депонированных рукописей, авторефератов диссертаций и другой научной литературы, ретроспектива – с 1991 года. Имеется доступ к базам данных и Интернет-ресурсам: БД ВИНИТИ (разработка ВИНИТИ), БД SCOPUS, БД Questel (патенты) и другим. Кроме того, ВИНИТИ доступны зарубежные электронные платформы ряда ведущих научных издательств, издающих основную часть академических и других рецензируемых изданий, в полнотекстовом варианте.

Основные тематические направления предлагаемых обзоров:

- Науки о жизни;
- Физико-математические науки;
- Химия и науки о материалах;
- Индустрия наносистем и материалов;
- Науки о Земле;
- Рациональное природопользование;
- Информационно-телекоммуникационные системы;
- Энергетика, энергоэффективность, энергосбережение;
- Транспортные, авиационные и космические системы;
- Производственные технологии.

Предлагается подготовка и заказ информационно-аналитических обзоров и материалов по тематике заказчика. Такие обзоры могут относиться к упомянутым выше тематическим направлениям, но могут иметь и междисциплинарный характер. В этом случае обзоры отражают актуальную научную информацию и научные достижения, происходящие на стыке наук.

Более подробная информация о приобретении, заказе и цене обзоров представлена на сайте ВИНИТИ www.viniti.ru

Приобретение и заказ обзоров от юридических лиц проводится на договорной основе.

Обращаться в ЦНИО ВИНИТИ:

- адрес: 125190, Россия, г. Москва, ул. Усиевича, 20;
- телефоны: 499-155-42-43, 499-155-42-17;
- эл. почта cnio@viniti.ru.

ВНИМАНИЮ ЧИТАТЕЛЕЙ!

НОВОЕ ИЗДАНИЕ УДК

УНИВЕРСАЛЬНАЯ ДЕСЯТИЧНАЯ КЛАССИФИКАЦИЯ
АЛФАВИТНО-ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ
в 2-х томах

Алфавитно-предметный указатель (АПУ) к 4-му полному изданию УДК на русском языке:

Том I содержит АПУ от буквы А до Н;

Том II содержит АПУ от буквы М до Я и указатель латинских наименований к классам УДК 56 Палеонтология, 57 Биологические науки, 58 Ботаника, 49 Зоология, 61 Медицинские науки.

АПУ содержит около 100 000 понятий, представленных в полных таблицах УДК.

При его составлении были учтены изменения, опубликованные в Выпусках № 1 – 6 «Изменения и дополнения к УДК»

Для подписки необходимо направить заявку для оформления счета по адресу:

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

Телефоны: 499 155-42-52, 499 155-42-85, 499 151-78-61

E-mail: feo@viniti.ru

<http://www.udcc.ru>