

# НАУЧНЫЕ КОММУНИКАЦИИ КАК РАСПРЕДЕЛЕННЫЕ ИНФОРМАЦИОННЫЕ СТРУКТУРЫ

Бурий А.С., Балванович А.В. \*

**Ключевые слова:** научные коммуникации, информационные структуры, анализ цитируемости, фактор влияния, база цитирования, темпоральная структура, индексы цитирования, информетрия.

## Аннотация.

**Цель работы:** совершенствование научной и методической базы в области наукометрии для сетевых коммуникаций в научном сообществе.

**Метод:** информационно-статистический анализ, экспертное оценивание и прогноз исследовательской деятельности в наукометрии на основе мониторинга и структурного анализа научных коммуникаций.

**Результаты:** обоснована структурно-функциональная концепция анализа баз цитирования как распределенных информационных структур и необходимость применения методов интеллектуального анализа слабоструктурированных данных, технологий семантических сетей, а также комплексирования традиционных метрик цитирования с индикаторами альтметрики, обеспечивающих сокращение времени информационного поиска в наукометрических базах данных.

Показано место и обоснованность задачи анализа цитируемости в цифровом пространстве для оценки тенденций развития науки, выявления перспективных направлений и разработки стратегий научно-технического совершенствования.

DOI: 10.21681/1994-1404-2020-2-17-27

## Введение

Развитие науки связано с повышением спроса на научные высококачественные результаты и инновации, основанные на внедрении современных эффективных инструментов и формировании цифровых платформ и технологий для повышения качества научных исследований [13]. Это во многом зависит от финансовых и кадровых ресурсов, наличия действующих научных школ<sup>1</sup>, создания информационно-коммуникационной инфраструктуры и благоприятной среды, способствующей развитию глубоких корпоративных связей сфер науки, технологий, включая сетевые возможности, потенциал эффективности управления в

этой сфере на всех уровнях. По данным рейтинга Всемирного экономического форума в 2018 г. Россия занимала 25-е место по направлению «Проникновение информационно-коммуникационных технологий» из 140 стран, попавших в аналитику [17].

Все чаще для анализа научного уровня полученных авторами и организациями научных результатов (достижений) применяются и разрабатываются новые информационно-статистические методы и алгоритмы [12], в том числе с привлечением современных методов интеллектуального анализа данных (ИАД).

Анализом и выявлением общих закономерностей развития наук, научных направлений, их связей с социальными институтами занимается науковедение. Одной из дисциплин науковедения является наукометрия, которая оценивает состояние науки в целом, выявляя связи и взаимовлияние между отраслями знаний, путем логической обработки информации о тенденциях в научных направлениях. Развитие современных методов и алгоритмов статистического оценивания публикаций, аналитико-синтетической обработки текстовой информации, анализ сетевых научных коммуникаций составляют понятие *информетрии*.

<sup>1</sup> Например: Системная информатизация управления силами и средствами ракетных и космических комплексов // Летопись Военной академии им. Петра Великого». Том 5. Научные школы. – М.: Издво «ВА РВСН», 2010. – С. 212 – 216.; Пинчук А. В. 20 лет научной школе «Системной информатизации управления силами и средствами ракетных и космических комплексов» // Петровский вестник. – 2011. – № 5(46). – С. 14 – 15.

\* **Бурий Алексей Сергеевич**, доктор технических наук, эксперт РАН, директор департамента ФГУП «Российский научно-технический центр информации по стандартизации, метрологии и оценке соответствия», Российская Федерация, г. Москва.  
E-mail: a.s.burij@gostinfo.ru

**Балванович Александр Витальевич**, кандидат экономических наук, заместитель директора департамента ФГУП «Российский научно-технический центр информации по стандартизации, метрологии и оценке соответствия», Российская Федерация, г. Москва.  
E-mail: a.v.balvanovich@gostinfo.ru

Формы представления текстовой информации могут быть следующими: печатные, т.е. бумажные в виде книг, статей, отчетов, описаний к патентам и др., и электронные (сетевые) в виде различных текстовых форматов, например, *doc*, *rtf*, *pdf*, графических и табличных форматов, а также видео- и аудио форматов. Одним из целевых приложений информетрии является *библиометрия* [3], которая направлена на разработку библиотечных баз, баз цитирования, развитие методов поиска информации, построение информационных коммуникаций на основе применения математических и информационно-статистических методов, а также методов ИАД.

Исходя из этого, науку можно представлять как самоорганизующуюся систему<sup>2</sup> формирования знаний об окружающем мире путем информационного отображения объективных законов природы и общества в наблюдаемые явления, и которая благодаря информационным связям, приобретает социальную, гносеологическую, ценностную и нормативную функции. К коммуникативной и креативной функциям науки будем относить *информационное взаимодействие*, проявляющееся в системах научного цитирования.

Авторитетность современного ученого все чаще сопоставляется с его публикациями [14], которые призваны отразить его научные результаты, являющиеся весомым аргументом в ходе их защиты на диссертационных советах, а также расширить аудиторию и информативность мирового сообщества о достижениях российских научных специалистов. Эффективность или фактор влияния опубликованных научных работ, исследуемое в наукометрии, определяется выбором научного издания (журнала), формой представления (качеством самого материала: его структурой, наглядностью результатов, аттрактивностью), предметной областью исследования и рядом других факторов, составляющих понятие «качества цитирования» [18].

Исследования наукометрических показателей цитируемости научных работ ведутся уже не одно десятилетие. Однако с развитием новых информационных теорий, методов и подходов к анализу данных<sup>3</sup> [9 – 11], информационно-распределенных структур под конкретно решаемые задачи [4, 7] объективно возникла необходимость вновь взглянуть на проблему цитирования, представления и поиска публикаций в коммуникационно-информационной среде [20], с целью сокращения времени поиска необходимых исследователю литературных источников, выбора наиболее авторитетных работ, рассматривая при этом цитируемость как

коллаборативный показатель влияния и научной аттрактивности. Именно рассмотрению указанной проблемы посвящена данная работа.

### Сущность цитирования в современной информационной среде

Для совершенствования процедур технологической поддержки науки в рамках глобальных трендов развития Российской Федерации<sup>4</sup> в контексте научно-технической политики, нацеленной на решение социально-экономических задач, повышение требований к продуктивности научной деятельности с учетом перспектив развития информационных систем и ресурсов в целом и информационных структур в частности, включая сетевые структуры различного уровня пространственного представления (вузовский, региональный, глобальный), ставит новые вопросы в, казалось бы, уже решенной на практике *задачи индексирования* научных публикаций. Однако анализ работ в данной области позволяет выделить следующие развивающиеся методы и подходы к исследованию вопросов цитирования публикаций применительно для задач оценивания как научных журналов (**Ж**), так и самих авторов (**А**), которые включают:

- выбор базы (баз) цитирования (**Ж**) – касается только журналов;
- разработка методов интеллектуального анализа данных в рамках технологии больших данных – применимо для (**Ж**) и (**А**);
- когнитивные карты пространственных связей структур «публикации – цитирования» (**А**);
- разработка модельных структур научных коммуникаций относительно цитируемых публикаций, отражающих их жизненный цикл, включая опубликование, распространение, активное использование в предметной области и архивирование [20];
- совершенствование метрик оценки показателей в зависимости от предметной области, сетей распространения, возможностей и ресурсов сети Интернет (**А, Ж**).

Процесс цитирования можно представить, как некоторый «незримый колледж» или *invisible college* (англ.)<sup>5</sup> – воображаемое академическое сообщество формально не связанных друг с другом членом, которым процесс познания окружающей действительности, а также написания научных работ представляется как обмен знаниями в рассматриваемой среде, в которой существует влияние отдельных научных школ, методологий и методов исследования.

<sup>2</sup> Евстигнеев В. А. Наукометрические исследования в информатике // Новосибирская школа программирования. Переключки времени. – Новосибирск, 2004. – URL: [http://www.computer-museum.ru/books/n\\_collection/investigat.htm](http://www.computer-museum.ru/books/n_collection/investigat.htm) (дата обращения: 16.02.2020).

<sup>3</sup> См., например: Ловцов Д. А. Введение в информационную теорию АСУ. – М.: ВА им. Петра Великого, 1996. – 434 с.; Информационные аспекты комплексного подхода к исследованию систем управления // Научно-техническая информация. Сер. 2. Информ. процессы и системы. – 1997. – № 5. – С. 10 – 17.

<sup>4</sup> Глобальные тренды и перспективы научно-технологического развития Российской Федерации // Докл. к XVIII Междунар. науч. конф. по проблемам развития экономики и общества. – URL: [https://issek.hse.ru/data/2017/04/21/1168819587/Gohberg\\_2017.pdf](https://issek.hse.ru/data/2017/04/21/1168819587/Gohberg_2017.pdf) (дата обращения 23.01.2020).

<sup>5</sup> Здесь – клуб английских учёных-натурфилософов и интеллектуалов 1640 – 1650 гг. (от лат. collegium, давшего русские слова «колледж» и «коллегия»).

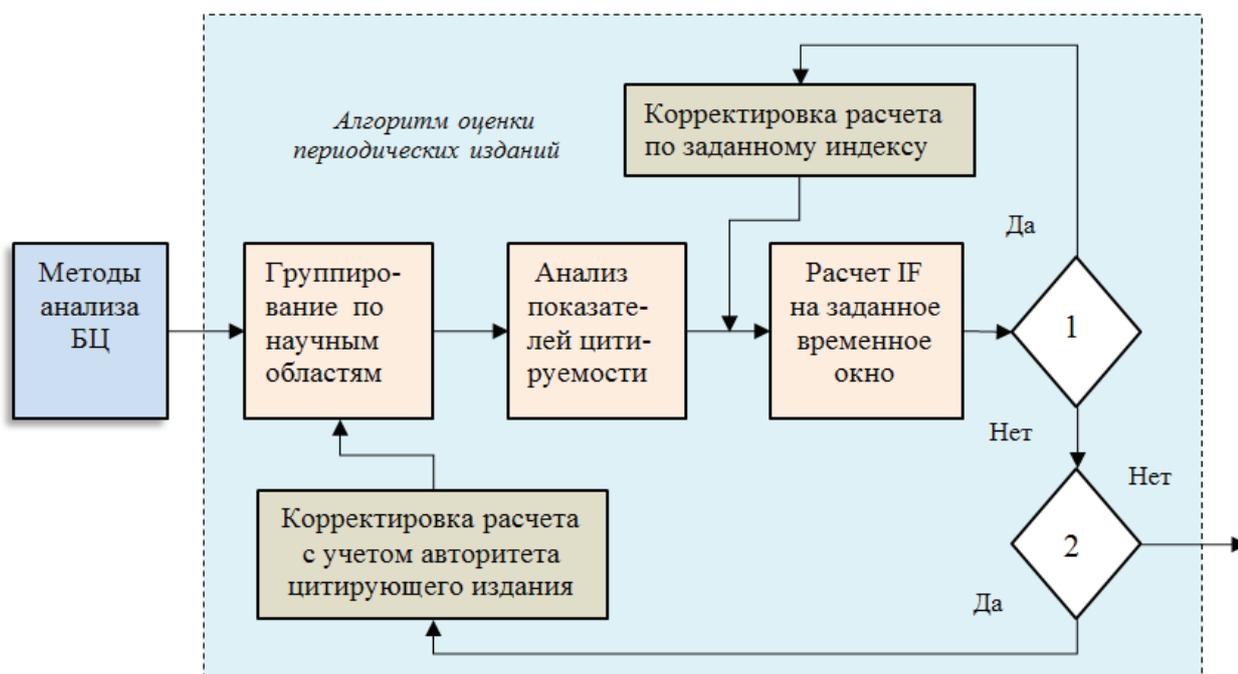


Рис. 1. Методическая схема анализа цитируемости научных публикаций

«Незримый колледж» образует механизм структурирования пространства знаний, позволяя выделить как передовые исследования, так и академический андеграунд [8], а также предметную область научных интересов для рассматриваемой сети цитирования, в которой отражаются научные коммуникации, число которых напрямую отражает интерес к конкретно выбранной работе.

### Методические подходы к анализу баз цитирования

Целесообразно рассматривать научные коммуникации как распределенные информационные системы (РИС). Но в отличие от обычных информационных систем, решающих задачи контроля, регистрации, управления сложными объектами и ряд других [4, 7], информационные системы научной коммуникации обеспечивают выполнение функций передачи результатов научной деятельности академическому сообществу, а также являются средством информационной поддержки организационной деятельности научного сообщества в ходе присуждения научных званий, грантов, подготовки молодых ученых [20].

Анализ существующих систем цитирования, применяемых основных метрик показывает:

- для коллективных публикаций не учитывается доля участия каждого соавтора;
- фиксируются только количественные связи (факты цитирования), не отражающие семантические составляющие;
- все элементы записей цитат имеют одинаковую важность; не ясна мотивация цитирования, которая основывается на контексте, практически не объясняется причина, когда с одним контекстом

автор связывает от двух до 4 – 6 работ, а с другим – не пытается найти и одной.

На рис. 1 представлена методическая схема анализа цитируемости публикаций на основе разработки новых или совершенствования существующих методов исследования систем информационного поиска. Важным элементом схемы является блок методов анализа баз цитирования (БЦ). Он предназначен для исследования новых методов анализа данных, среди которых выделим:

1) способность выявлять новые знания, что при доступе к данным о публикациях сокращает время на получение информации о цитировании. Это актуально в ходе анализа больших данных при исследовании цитируемости журнала или библиотечного фонда по спланированным сценариям, что позволяет получить существенный временной выигрыш<sup>6</sup>;

2) аналитику больших данных для задач комплексного анализа публикаций, учитывая категории научных статей, ключевые слова, уровень научного журнала [21, 23], а также с необходимостью учета в ряде случаев слабоструктурированной информации;

3) приложение идей научного картирования как формы представления данных для получения пространственной картины связей объектов цитирования, предметных (научных) областей, объединяя пространственный и темпоральный<sup>7</sup> анализ, включая тематику исследований (предметную область) и информационно-сетевые связи публикаций [2];

<sup>6</sup> White P. B. Using Data Mining for Citation Analysis // College & Research Libraries, [S.l.]. – 2019. – Vol. 80(1). – URL: <https://crl.acrl.org/index.php/crl/article/view/16892/19368> (дата обращения 18.02.2020).

<sup>7</sup> Темпоральность (от англ. *tempora*) – временные особенности, связанные с социальным воздействием публикации, выраженными фактами ее цитирования.

4) облачные технологии как рациональное сочетание информационных технологий, тенденций развития распределенных узлов обслуживания и постоянно растущего разнообразия запросов потребителей [6];

5) альтернативные подходы к применению метрик цитирования, получивших название *альтметрики* [22]. Источниками этих данных выступают как официальные ресурсы, например, *eLibrary.ru*, где можно анализировать количество просмотров, скачиваний статьи, число полученных отзывов, так и неофициальные: медиа-среда (социальные сети, блоги), информация о публикациях, размещаемая на домашних страницах исследователя, что может учитываться как дополнительные факторы влияния публикаций, придавая им дополнительный вес при ранжировании;

6) анализ цитирования на основе технологии семантических сетей путем выявления онтологий связываемых данных, структурируя предметную область системой связей или отношений, что особенно важно для слабоструктурированной информации, например, при выборе объектов по качественным оценкам [5, 15].

Наряду с этапом теоретического исследования с целью формирования новых наукометрических и статисти-

ческих признаков для информационных объектов (документов), реализуется традиционный *алгоритм* оценивания научных периодических изданий на основе библиометрических данных (см. *рис. 1*). Окончательные результаты корректируются в зависимости от тематики изданий, временного окна исследования фактора влияния [18] (импакт-фактора, *IF*) журнала (2-х или 5-летнего), за счет первого и второго логических блоков.

В настоящее время наукометрия в основном ассоциируется с количественными оценками, признаками (метриками), в большинстве случаев совпадающими с концепцией библиометрии и включающими такие понятия как цитирование, совместное цитирование, библиографические связи, ключевые слова и др. Одновременное воздействие указанных элементов будем представлять в виде *координат* некоторого интеллектуального вектора воздействия  $I_0$  для анализируемой публикации  $P_0$ . Вершина вектора  $I_0$  направлена в пространство области знаний, которой посвящена публикация  $P_0$ . Портрет или образ связей «публикация  $P_0$  – цитирование» представляет собой иерархическую структуру с числом связей, равным числу цитирующих работ (см. *рис. 2*).

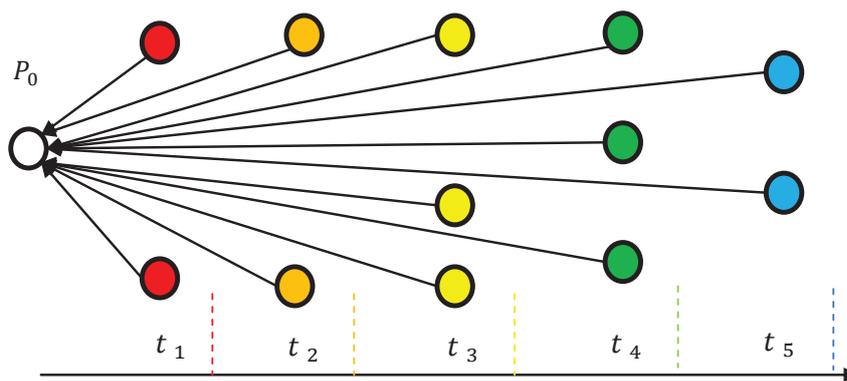


Рис. 2. Структура цитирования публикации  $P_0$  по годам

Временные страты ( $t_1, t_2, \dots, t_5$ ) здесь соответствуют годам цитирования или временным срезам.

С целью анализа темпорального расширения структуры области цитирования первичной публикации  $P_0 \in I_0$  можно показать, как факторы цитирования публикации  $P_0 \in I_0$  трансформируют ее, уже в составе библиографического источника публикации  $P_t$ , принадлежащей новой области знаний  $P_t \in I_t$  или, так называемому, многодисциплинарному пространству знаний (МПЗ). МПЗ формируется путем пересечения отдельных областей знаний.

Таким образом, параметр времени  $t, t_i \in [0, T]$  является меткой для идентификации структуры цитирования в базе данных (БД), в чем заключается суть темпоральной БД, которая расширяется с каждым годом относительно первичной публикации  $P_0$ .

Так, по результатам первого года (года публикации работы  $P_0$ ) число вершин графа определяется количеством цитирований  $X_i, (i = \overline{1, N})$  для  $i$ -го временно-

го среза. Для представленного на *рис. 2* графа количество вершин определим, как

$$\bar{X}^N = 1 + \sum_{i=1}^N X_i.$$

Данный подход уместен для работ, написанных на актуальные темы, способных заинтересовать научное сообщество, а временное окно зависит от направленности научной работы. Обычно более важным считается долгосрочное воздействие публикаций [18] (*Database*), т.е. работы, цитируемость которых продолжается на интервале более 5 лет.

При цитировании конкретного документа существует возможность перехода на связанные документы, т.е. на другие ресурсы, создавая сеть совместного цитирования<sup>8</sup> (*рис. 3-а*). Здесь *A, B, C* – авторы исходных

<sup>8</sup> Trujillo C.M., Long T.M. Document co-citation analysis to enhance transdisciplinary research // Science Advances. – 2018. – Vol. 4. – No. 1:1-9. DOI: 10.1126/sciadv.1701130

документов, маленькими буквами **a – h** обозначены цитируемые источники, различная окраска которых соответствует различным предметным областям исследования. Таким образом, авторы **A, B, C** совместно процитировали работы **c** и **d** по три раза, а работу **e** проци-

тировали два раза в документах **A** и **B** (см. рис. 3-б), где узлы – это цитируемые документы. Подход позволяет на основе трансформации графов показывать динамику совместного цитирования с учетом области знаний, к которой принадлежит научная работа.

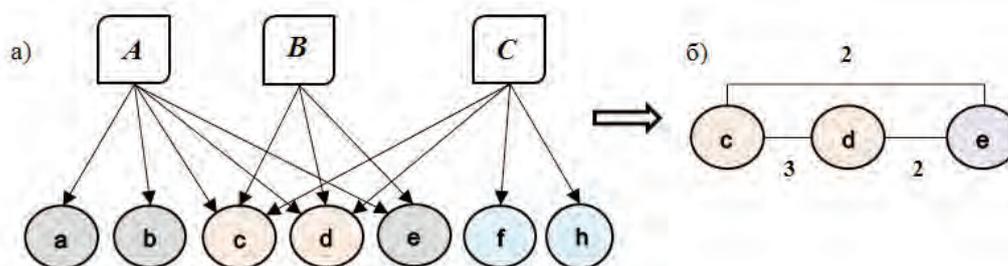


Рис. 3. Структура совместного цитирования

### Организация цитируемости журнала

Большинство индексных баз формируются под выбранные предметные области знаний, объединяя как издателей, так и авторов, а вместе с ними и читателей или пользователей в определенной научной сфере. Так, известными информационными ресурсами для своих направлений являются:

- [pravo.gov.ru](http://pravo.gov.ru) – интернет-портал правовой информации;
- [edirc.repec.org/data/derasru.html](http://edirc.repec.org/data/derasru.html) – база данных публикаций по экономике;
- [uisrussia.msu.ru](http://uisrussia.msu.ru) – электронная библиотечная база данных для университетских исследований в области управления, социологии, философии и других гуманитарных наук;
- [math-net.ru](http://math-net.ru) – для журналов математической направленности;
- [www.gost.ru/portal/gost/home/activity/metrology](http://www.gost.ru/portal/gost/home/activity/metrology) – база данных Федерального агентства по техническому регулированию и «Росстандарт»;
- [www.sciencedirect.com](http://www.sciencedirect.com) – база журналов по математике, информатике, экономике, социальным наукам издательства Elsevier (более 1500 журналов);
- *zbMATH (Database Zentralblatt MATH)* – БД реферативных записей статей (без полных текстов) математического журнала, основанного издательством «Шиллингер» (сотрудничает более чем с 2300 журналами).

Научные журналы по праву становятся основным способом осуществления научных коммуникаций [1] в современной научной среде. С ростом научных направлений постоянно увеличивается и число научных изданий, готовых опубликовать результаты авторских исследований. С 2008 по 2018 гг. объем публикаций вырос с 1,8 млн до 2,6 млн статей. В 2018 г. доля работ ученых Китая составила 21%, представителей США – 17%, отмечается рост международного сотрудничества в этой сфере. Так, в 2018 г. каждая пятая статья написана соавторами из разных стран<sup>9</sup>.

На примере всемирно известного журнала «Автоматика и телемеханика» (АиТ) покажем взаимодействие баз цитирования (рис. 4). Большинство из указанных ресурсов широко известны в научном сообществе, но некоторые следует пояснить. Рассматриваемый журнал – АиТ публикует работы по теоретическим и прикладным проблемам управления. Исторически его переводная версия издается под названием «Automation and Remote Control». Соответственно индексные базы, с которыми взаимодействует издательство журнала, можно представить двумя отдельными ветвями. На практике количество коммуникаций, связанных с журналом информационных ресурсов значительно больше, что, главным образом, объясняется высоким российским и международным авторитетом данного издания – АиТ, являющегося журналом Российской Академии наук.

Кроме того, взаимодействие с рядом ресурсов является обязательным, например, *ROAD*, РИНЦ, Российская государственная библиотека (для ряда изданий), *Crossref*. В некоторые другие БД журнал попадает практически автоматически, достигнув определенного уровня научного авторитета, например, попав в ядро РИНЦ, после чего он регистрируется в *Web of Science (WoS)*, а в *Math-Net.ru*, если входит в число более 130 ведущих российских журналов на математическом портале. С большинством оставшихся научных БД издатели подписывают договорные отношения для реализации выше отмеченных задач. Переводная версия журнала АиТ зарегистрирована в *Scopus*, а, следовательно, индексируется в ряде баз данных (см. рис. 4), например, в *WoS*, *DOAJ*, *WorldCat* (международная межбиблиотечная БД), *Google Scholar* – БД научных публикаций, созданная компанией Google.

В БД *Scopus* используется ряд инструментов для оценки качества журналов, организаций, ученых, а также стран. К наиболее популярным показателям относят систему рейтинга журналов – показатель *SCImago Journal Ranking (SJR)* и *Journal Citation Reports (JCR)* – для оценки показателей цитируемости журналов, позволяя авторам выбирать журналы для заданной области исследований. Примером показателя *JCR*

<sup>9</sup>См.: Publications Output: U.S. Trends and International Comparisons // Science & Engineering Indicators. – URL: <https://nces.nsf.gov/pubs/nsb20206/> (дата обращения 09.01.2020).

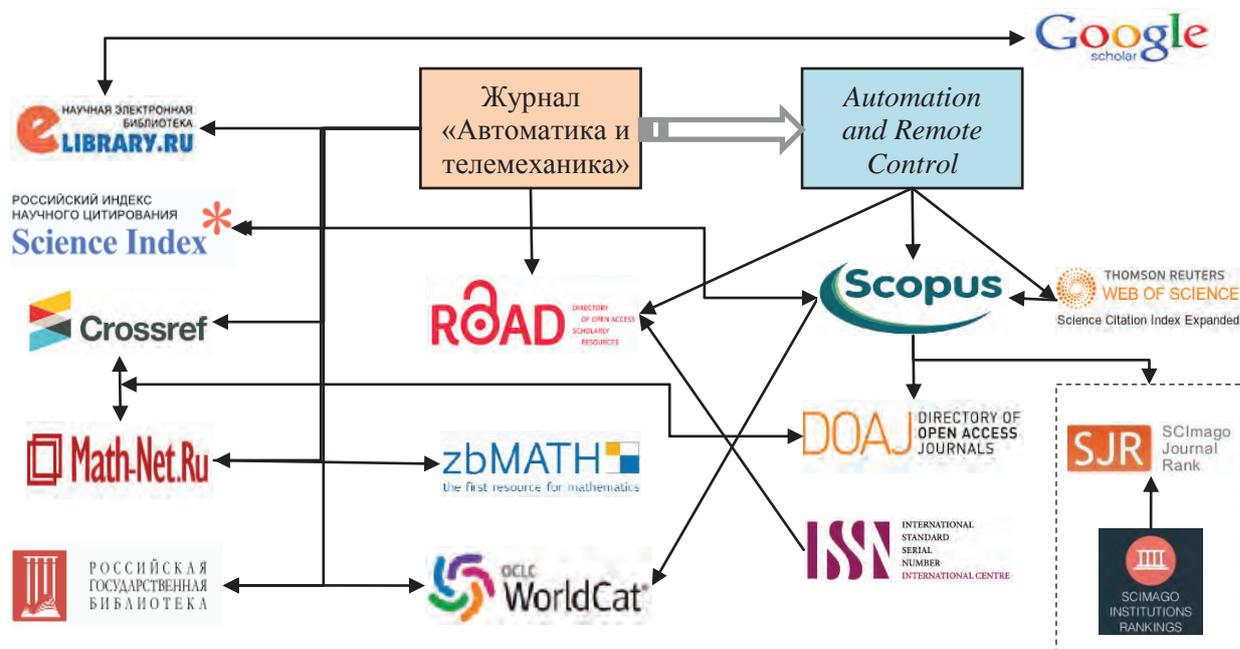


Рис. 4. Коммуникации межбазового информационного взаимодействия

является фактор влияния или импакт-фактор журнала (*JIF*) как отношение общего количества цитирований для заданного временного окна в прошлом к общему числу статей, опубликованных в журнале за этот же период [8].

Для каждого из 200 научных тематических направлений выделяют 4 квартиля ( $Q1 - Q4$ ), на которые делятся (ранжируются) журналы. Данная категория отражает уровень востребованности научного журнала в академической среде. Первые 25% списка составляют  $Q1$  – самый высокий квартиль; от 25% до 50% – квартиль  $Q2$  и т.д. В частности, для приведенного примера журналу “Automation and Remote Control” соответствует квартиль  $Q2$  в предметной области – «Системы управления и проектирования», показатель  $SJR = 0,35$ , а индекс Хирша ( $h$ -индекс) соответствует 28. Полный набор наукометрических показателей для журналов с регулярной издательской деятельностью представлен на сайте *elibrary.ru*. В последнее время большинство журналов на своих сайтах размещают информацию о связях издания с базами цитирования.

#### Совершенствование индексации цитирования

Проблема поиска наиболее приемлемого индекса для определения научного уровня специалиста по-прежнему актуальна, так как основной показатель –  $h$ -индекс основывается на наиболее цитируемых работах и практически не учитывает те из них, которые вышли не так давно или которые имеют единичные цитирования. При этом надо отдавать себе отчет, опираясь на высказывание одного из основоположников наукометрии, автора работ по созданию системы индексации и цитирования научных работ Ю. Гарфильда, что «цитирования не указывают на четкость, важность,

качество или значимость» работы, они лишь «являются индикаторами полезности или влияния»<sup>10</sup>.

На рис. 5 представлена диаграмма формирования  $h$ -индекса Хирша.

Уровень  $h$ -индекса практически не зависит от максимально и минимально цитируемых работ таблицы «Публикации/Цитирования», кроме того, как показывают обзоры [16, 19], не учитывается число авторов и возраст публикации, когда она еще остается востребованной в научном мире, общее количество публикаций автора. Указанные причины явились причиной появления ряда дополнительных индексов, представленных в табл. 1.

Если  $h$ -индекс отталкивается от хорошо цитируемых статей автора, то  $g$ -индекс учитывает статьи с наибольшим цитированием. Стремление повысить «вес» наименее цитируемым статьям вызвало появление  $f$ -индекса, в выражении в табл. 1 для которого  $cit_p$  – количество цитирований  $p$ -й статьи. Характерно, что величина  $f$ -индекса лежит в пределах  $h \leq f \leq g$  [19]. Мы рекомендуем обзор [19], где представлены более ста различных индексов оценки цитируемости, учитывающие и особенности предметной области, и количество авторов работ, и коллаборативные отношения между публикациями и ряд других метрических категорий и подходов.

Семантический анализ содержания текстов применяется для автоматизированного анализа публикаций. Анализ целесообразно проводить по полным публикациям<sup>11</sup>, что обеспечит формирование новых

<sup>10</sup> Garfield E. Introducing index to social sciences and humanities proceedings: more help in locating and acquiring proceedings // Current Contents. – 1978. – No 33. – P. 5 – 9.

<sup>11</sup> Knoth P, Herrmannova D. Towards Semantometrics: A New Semantic Similarity Based Measure for Assessing a Research Publication's Contribution // D-Lib Magazine. – 2014. – Vol. 20. – No 11/12. DOI: 10.1045/november14-knoth

## Определение индекса Хирша

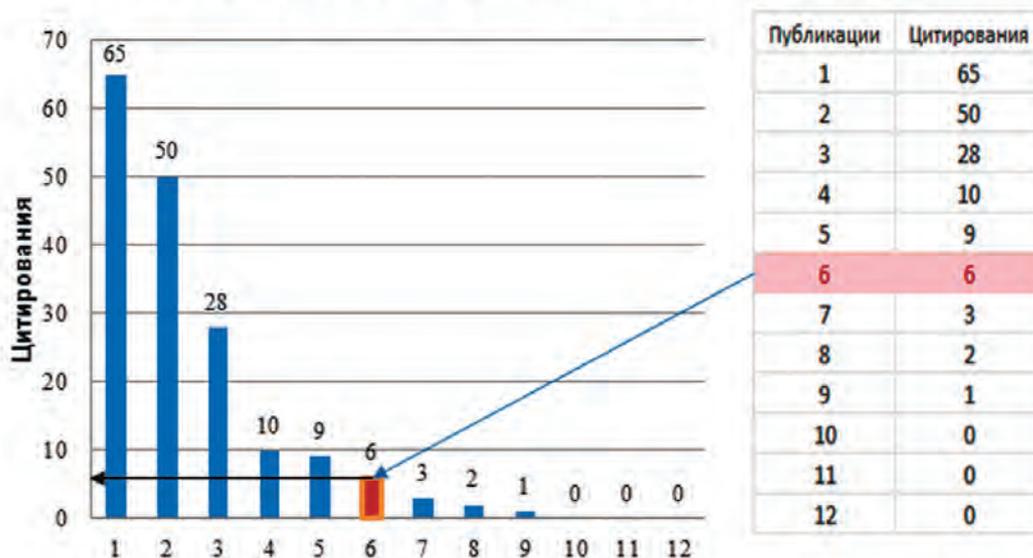


Рис. 5. Пример определения h-индекса

знаний относительно цитируемых публикаций из первоисточников. Вклад цитируемых публикаций оценивается путем вычисления *семантических расстояний* между рассматриваемыми документами или узлами

семантической сети цитирования РИС. В частности, для косинусной меры сходства используются признаки в виде частот появления терминов в сравниваемых документах.

Таблица 1  
Альтернативные оценки цитирования

Индекс	Определение
<i>g</i> -индекс	Это наибольшее число <i>g</i> публикаций, которые суммарно составили не менее $g^2$ цитирований. Позволяет учитывать работы с наибольшим цитированием
$h_2$ -индекс	Это наибольшее число работ автора, причем каждая из $h_2$ (самых популярных публикаций) процитирована не менее $(h_2)^2$ раз
<i>f</i> -индекс	Представляет собой гармоническое среднее, придавая больший вес менее цитируемым работам $f = \max_f \left( f^{-1} \sum_{p=1}^f (cit_p)^{-1} \right)^{-1} \geq f$
<i>r</i> -индекс	Для учета высокоцитируемых работ автора, $r = \left( \sum_{p=1}^h cit_p \right)^{1/2}$
<i>ar</i> -индекс	Это модификация <i>h</i> -индекса, увеличивается или уменьшается с течением времени, $ar = \left[ \sum_{j=1}^h (cit_j/t_j) \right]^{1/2}$ где $cit_j$ – число цитирований <i>j</i> -й статьи, а $t_j$ – время, прошедшее с момента ее опубликования

С целью сокращения времени *извлечения информации*<sup>12</sup> в современных информационных технологиях разрабатываются методы, направленные на *структурирование* данных – результатов анализа, на основе

логической обработки слабоструктурированных документов (текстов публикаций). Это актуально для любой области знаний. Например, подход активно используется при идентификации опорных выражений в юридических текстах для сопоставления и поиска информации при сетевом анализе [24].

<sup>12</sup>Соответствует термину "information extraction" (англ.).

Анализ частоты появления терминов или ключевых слов по данным ресурсов *eLibrary.ru* приведен в табл. 2, где сравниваются известные «информационные» научно-практические журналы России «Правовая информатика», «Вопросы кибернетики», «Информационное

право» по доле статей, использующих указанные термины в составе ключевых слов. Это позволяет оценивать тематическую направленность журнала, а также проследить динамику ее изменения на интервале исследования.

Таблица 2

Сравнение «информационных» научно-практических журналов России по ключевым словам

Ключевые слова	«Правовая информатика»	«Вопросы кибернетики»	«Информационное право»
Качество	0,24	0,17	0,06
Информация	0,46	0,44	0,33
Безопасность	0,26	0,57	0,18
Право	0,24	0,03	0,43
<b>Всего статей</b>	<b>245</b>	<b>352</b>	<b>641</b>

### Выбор журнала для публикации

В условиях широких возможностей цифровых технологий число вариантов для опубликования своих научных результатов значительно возросло. Появились издательства с сомнительной в научной среде репутацией, поэтому следует придерживаться ряда правил и критериев при выборе научного журнала. При этом основными *требованиями* являются: соответствие тематики журнала планируемой публикации и научная репутация издания (по отзывам наставников, коллег). В качестве критериев выбора журнала для публикации можно использовать следующие:

1) качество представления материала (структуризация текста статьи), оформление иллюстративного материала, списка литературных источников, приводимых в качестве обоснования актуальности темы исследования и полемирования, прозрачность и воспроизводимость полученных результатов;

2) состав и авторитетность редакционного состава, уровень рецензирования работ известными учеными, что можно оценить по заголовкам, подробности рефератов опубликованных статей журнала;

3) процесс рецензирования, сроки его проведения и порядок разрешения возникающих конфликтов интересов;

4) репутация журнала в научном мире, четкое освещение, например, на сайте журнала тематики исследований, которые могут быть опубликованы в нём, форма и прозрачность закрепления авторских прав;

5) статус индексирования журнала (библиографические базы данных и базы цитирования, в которых будет размещена информация опубликованных работ), что обеспечивает повышение видимости и доступности журнала для читателей, а также повышение его метрических показателей (фактора влияния).

### Заключение

Рассматриваемые в работе организационные вопросы литературно-издательской и публикационной деятельности ученых и соискателей ученых степеней направлены как на развитие методических инструментов (привлечение методов интеллектуального анализа структурированных данных, семантических сетевых структур, темпорального анализа и др.) для исследования цитируемости научных публикаций, так и практических рекомендаций применительно к повседневной работе издательств научной периодики для повышения факторов влияния в академической среде.

Учет авторами проектов публикаций существующих индексов цитирования (возможных значений) позволит им обоснованно оценивать свой вклад в развитие выбранного научного направления или исследуемой предметной области. А учет известных информационных ресурсов для возможного представления результатов своих научных исследований позволит определить более рациональный и достойный способ доведения соответствующей информации до научной общественности.

Рецензент: **Емелин Николай Михайлович**, доктор технических наук, профессор, заслуженный деятель науки и техники РСФСР, главный научный сотрудник Государственного научно-методического центра Минобрнауки РФ, г. Москва.

**Литература**

1. Абрамов Е.Г. Признание модели журналов открытого доступа как способ повышения качества научных публикаций в России // Научная периодика: проблемы и решения. 2013. № 2(14). С. 4-8.
2. Акоев М. А. Картирование науки и технологии, прогноз развития // Руководство по наукометрии: индикаторы развития науки и технологии. Екатеринбург: Издательство Уральского университета, 2014. С. 164-184.
3. Бредихин С. В., Кузнецов А. Ю., Щербакова Н. Г. Анализ цитирования в библиометрии. Новосибирск: ИВМ и МГ СО РАН, НЭИКОН, 2013. 344 с.
4. Бурый А.С. Информационно-поисковые социотехнические системы: термины и определения. М.: Горячая линия – Телеком, 2018. 166 с.
5. Бурый А.С. Информационно-математическое обеспечение контроля качества компьютерных программ // Правовая информатика. 2019. № 2. С. 15-25. DOI: 10.21681/1994-1404-2019-2-15-25.
6. Бурый А.С., Саков А.А., Слепынцева Л.И. Разработка и актуализация нормативных документов и стандартов при управлении облачными сервисами // Транспортное дело России. 2015. № 6. С. 79-81.
7. Бурый А.С., Шевкунов М.А. Суррогатное моделирование распределенных информационных систем по большим данным // Информационно-экономические аспекты стандартизации и технического регулирования. 2019. № 5(51). С. 43-50.
8. Калимуллин Т.Р. Российский рынок диссертационных услуг // Экономическая социология. 2005. Т. 6. № 4. С. 14-38.
9. Ловцов Д. А. Информационная теория эргасистем: Тезаурус. М.: Наука, 2005. 248 с. ISBN 5-02-033779-X.
10. Ловцов Д. А. Информационная теория эргасистем: основные положения // Правовая информатика. 2019. № 3. С. 4-20. DOI 10.21681/1994-1404-2019-3-04-20.
11. Ловцов Д. А. Теория информационного права: базисные аспекты // Государство и право. 2011. № 11. С. 43-51.
12. Ловцов Д. А., Богданова М. В., Паршинцева Л. С. Пакеты прикладных программ для многоаспектного анализа судебной статистической информации // Правовая информатика. 2017. № 1. С. 28-36. DOI 10.21681/1994-1404-2017-1-28-36.
13. Ловцов Д. А., Ниесов В. А. Об организации научных исследований и разработки основ обеспечения единства судебной системы России в информационной сфере судопроизводства // Российское правосудие. 2010. № 8. С. 102-105.
14. Миркин Б. Г. О понятии научного вклада и его измерителях // Управление большими системами. 2013. № 44. С. 292-307.
15. Онтологический и нечеткий анализ слабоструктурированных информационных ресурсов / Т.В. Афанасьева, В.С. Мошкин, А.М. Наместников и др.; под науч. ред. Н. Г. Ярушкиной. Ульяновск: УлГТУ, 2016. 130 с.
16. Цыганов А.В. Краткое описание наукометрических показателей, основанных на цитируемости // Управление большими системами. 2013. № 44. С. 248-261.
17. Что такое цифровая экономика? Тренды, компетенции, измерение: докл. к XX Апр. междунар. науч. конф. по проблемам развития экономики и общества, Москва, 9-12 апр. 2019 г./ Абдрахманова Г.И., Вишневецкий К.О., Гохберг Л.М. и др. М.: Изд. дом Высшей школы экономики, 2019. 82 с.
18. Aksnes D.W., Langfeldt L., Wouters P. Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories // SAGE Open. 2019. Jan-Mar. P. 1-17. DOI: 10.1177/2158244019829575.
19. Bihari A., Tripathi S., Deepak A. h-index and its alternative: A Review. 2018. URL: <https://arxiv.org/pdf/1811.03308v1.pdf> (дата обращения 20.02.2020).
20. Björk B.-C. A model of scientific communication as a global distributed information system // Information Research – An International Electronic Journal. 2007. Vol. 12. No. 2.
21. Buryi A.S., Lomakin M.I., Sukhov A.V. Quality Assessment of “Stress-Strength” Models in the Conditions of Big Data // International Journal of Innovative Technology and Exploring Engineering. Jan.020. Vol. 9. No 3. P. 3276-3281. DOI: 10.35940/ijitee.C8982.019320
22. Huang W., Wang P., Wu Q. A correlation comparison between Altmetric Attention Scores and citations for six PLOS journals // Plos One. 2018. No. 5. P. 1-15. DOI: 10.1371/journal.pone.0194962.
23. Kalantari A., Kamsin A., Kamaruddin H. S., Ebrahim N. A., Gani A., Ebrahimi A. and Shamshirband S. A bibliometric approach to tracking big data research trends // Journal of Big Data. 2017. Vol. 4, No. 30.
24. Sakhaee N., Wilson M.C. Information extraction framework to build legislation network // Artificial Intelligence and Law. 2018. P. 1-24. DOI: 10.1007/s10506-020-09263-3

# SCIENTIFIC COMMUNICATIONS AS DISTRIBUTED INFORMATION STRUCTURES

**Aleksei Buryi**, Dr.Sc. (Technology), expert of the Russian Academy of Sciences, Director of a department of the Russian Scientific and Technical Centre of Information on Standardization, Metrology and Conformity Assessment. Moscow, Russian Federation

**E-mail:** [a.s.burij@gostinfo.ru](mailto:a.s.burij@gostinfo.ru)

**Aleksandr Balvanovich**, Ph.D. (Economics), Deputy Director of a department of the Russian Scientific and Technical Centre of Information on Standardization, Metrology and Conformity Assessment. Moscow, Russian Federation

**E-mail:** [a.v.balvanovich@gostinfo.ru](mailto:a.v.balvanovich@gostinfo.ru)

**Keywords:** scientific communications, information structures, citation analysis, impact factor, citation base, temporal structure, citation indexes, inform-metrics.

## Abstract.

**Purpose of the work:** improving the scientific and methodological base in the field of science-metrics to ensure network communications in the scientific community.

**Method used:** information and statistical analysis, expert assessment and forecast of research activities in science-metrics based on monitoring and structural analysis of scientific communications.

**Results:** it is substantiated the structural and functional concept of analyzing citation databases as distributed information structures, and the need to apply methods of intellectual analysis of weakly structured data, semantic network technologies, as well as the integration of traditional citation metrics with altmetrics indicators that reduce the time of information search in scientific databases.

The place and validity of the task of citation analysis in the digital space for assessing trends in the development of science, identifying promising areas and developing strategies for scientific and technical perfection is shown.

## References:

1. Abramov E.G. Priznanie modeli zhurnalov otkry`togo dostupa kak sposob pov`sheniia kachestva nauchny`kh publikatsii` v Rossii // Nauchnaia periodika: problemy` i resheniia. 2013. № 2(14). S. 4-8.
2. Akoev M. A. Kartirovanie nauki i tekhnologii, prognoz razvitiia // Rukovodstvo po naukometrii: indikatory` razvitiia nauki i tekhnologii. Ekaterinburg: Izdatel`stvo Ural`skogo universiteta, 2014. S. 164-184.
3. Bredihin S. V., Kuznetcov A. Iu., Shcherbakova N. G. Analiz tciirovaniia v bibliometrii. Novosibirsk: IVM i MG SO RAN, NE`IKON, 2013. 344 s.
4. Buryi` A.S. Informatcionno-poiskovy`e sotciotekhnicheskie sistemy`: terminy` i opredeleniia. M.: Goriachaia liniia – Telekom, 2018. 166 s.
5. Buryi` A.S. Informatcionno-matematicheskoe obespechenie kontroliia kachestva komp`iuterny`kh programm // Pravovaia informatika. 2019. № 2. S. 15-25. DOI: 10.21681/1994-1404-2019-2-15-25.
6. Buryi` A.S., Sakov A.A., Slepyn`tceva L.I. Razrabotka i aktualizatsiia normativny`kh dokumentov i standartov pri upravlenii oblachny`mi servisami // Transportnoe delo Rossii. 2015. № 6. S. 79-81.
7. Buryi` A.S., Shevkunov M.A. Surrogatnoe modelirovanie raspredelenny`kh informatcionny`kh sistem po bol`shim dannym // Informatcionno-e`konomicheskie aspekty` standartizatsii i tekhnicheskogo regulirovaniia. 2019. № 5(51). S. 43-50.
8. Kalimullin T.R. Rossii`skii` ry`nok dissertatsionny`kh uslug // E`konomicheskaiia sotciologiiia. 2005. T. 6. № 4. S. 14-38.
9. Lovtcov D. A. Informatcionnaia teoriia e`rgasistem: Tezaurus. M.: Nauka, 2005. 248 c. ISBN 5-02-033779-X.
10. Lovtcov D. A. Informatcionnaia teoriia e`rgasistem: osnovny`e polozheniia // Pravovaia informatika. 2019. № 3. S. 4-20. DOI 10.21681/1994-1404-2019-3-04-20.
11. Lovtcov D. A. Teoriia informatcionnogo prava: bazisny`e aspekty` // Gosudarstvo i pravo. 2011. № 11. S. 43-51.
12. Lovtcov D. A., Bogdanova M. V., Parshintceva L. S. Pakety` prikladny`kh programm dlia mnogoaspektного analiza sudebnoi` statisticheskoi` informatsii // Pravovaia informatika. 2017. № 1. S. 28-36. DOI 10.21681/1994-1404-2017-1-28-36.
13. Lovtcov D. A., Niesov V. A. Ob organizatsii nauchny`kh issledovaniia i razrabotki osnov obespecheniia edinstva sudebnoi` sistemy` Rossii v informatcionnoi` sfere sudoproizvodstva // Rossii`skoe pravosudie. 2010. № 8. S. 102-105.
14. Mirkin B. G. O poniatii nauchnogo vclada i ego izmeriteliakh // Upravlenie bol`shimi sistemami. 2013. № 44. S. 292-307.

15. Ontologicheskii i nechetkii analiz slabostrukturirovannykh informatcionnykh resursov / T.V. Afanas`eva, V.S. Moshkin, A.M. Namestneykov i dr.; pod nauch. red. N. G. Iarushkinoi`. Ul`ianovsk: UIGTU, 2016. 130 s.
16. TCy`ganov A.V. Kratkoe opisanie naukometricheskikh pokazatelei`, osnovannykh na tcitiruemosti // Upravlenie bol`shimi sistemami. 2013. № 44. S. 248-261.
17. Chto takoe tcifrovaia e`konomika? Trendy`, kompetentcii, izmerenie: docl. k XX Apr. mezhdunar. nauch. konf. po problemam razvitiia e`konomiki i obshchestva, Moskva, 9-12 apr. 2019 g./ Abdrakhmanova G.I., Vishnevskii` K.O., Gokhberg L.M. i dr. M.: Izd. dom Vy`shei` shkoly` e`konomiki, 2019. 82 s.
18. Aksnes D.W., Langfeldt L., Wouters P. Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories // SAGE Open. 2019. Jan-Mar. P. 1-17. DOI: 10.1177/2158244019829575.
19. Bihari A., Tripathi S., Deepak A. h-index and its alternative: A Review. 2018. URL: <https://arxiv.org/pdf/1811.03308v1.pdf> (data obrashcheniia 20.02.2020).
20. Björk B.-C. A model of scientific communication as a global distributed information system // Information Research – An International Electronic Journal. 2007. Vol. 12. No. 2.
21. Buryi A.S., Lomakin M.I., Sukhov A.V. Quality Assessment of "Stress-Strength" Models in the Conditions of Big Data // International Journal of Innovative Technology and Exploring Engineering. Jan.020. Vol. 9. No 3. P. 3276-3281. DOI: 10.35940/ijitee.C8982.019320
22. Huang W., Wang P., Wu Q. A correlation comparison between Altmetric Attention Scores and citations for six PLOS journals // Plos One. 2018. No. 5. P. 1-15. DOI: 10.1371/journal.pone.0194962.
23. Kalantari A., Kamsin A., Kamaruddin H. S., Ebrahim N. A., Gani A., Ebrahimi A. and Shamshirband S. A bibliometric approach to tracking big data research trends // Journal of Big Data. 2017. Vol. 4, No. 30.
24. Sakhaee N., Wilson M.C. Information extraction framework to build legislation network // Artificial Intelligence and Law. 2018. P. 1-24. DOI: 10.1007/s10506-020-09263-3