

НАУКОМЕТРИЯ. БИБЛИОМЕТРИЯ

УДК 002:51

И. А. Мосичева

НИУ «Московский энергетический институт»

**С. Л. Парфенова, В. Н. Долгова, К. А. Безроднова,
Е. А. Лягушкина, В. В. Богатов, Н. В. Халтакшинова,
В. Я. Коробатов**

*Российский НИИ экономики, политики и права
в научно-технической сфере (Москва)*

И. В. Михайленко

ГПНТБ России

Метод прогнозирования числа публикаций на основе интегрального показателя по данным Web of Science и Scopus

Цель представленного в статье исследования – формирование релевантного инструментария расчёта и прогноза совокупного числа публикаций российских исследователей на основе данных двух глобальных систем научного цитирования – Web of Science и Scopus – с дедубликацией по Цифровым идентификаторам объекта, используемым в обеих системах. Определены задачи, которые необходимо решить для достижения этой цели: выбор способа прогнозирования числа научных статей, индексируемых в Web of Science в рамках метода «прогнозирование назад»; разработка интегрального показателя расчёта числа публикаций на основе данных Web of Science и Scopus; расчёт и прогноз числа научных статей и публикаций по странам G7 и BRICS до 2021 г. с помощью разработанного наукометрического инструментария на основе данных Web of Science и Scopus. В статье предложен принципиально новый способ расчёта интегрального показателя числа публикаций по данным Web of Science и Scopus, исключающий дублирование при учёте публикаций. Статья содержит расчёт по странам G7 и BRICS совокупного числа научных статей и публикаций (по данным Web of Science и Scopus) за период с 2007 г. по 2016 г. и прогноз до 2021 г.

Исследование выполнено в рамках государственного задания Министерства образования и науки Российской Федерации «Мониторинг и анализ публикационной и патентной активности российской и мировой науки в целях обеспечения реализации государственной научно-технической политики» (проект № 28.12616.2018/12.1).

Ключевые слова: метод экстраполяции, интегральный показатель, число публикаций, Web of Science, Scopus, прогнозирование числа публикаций.

Irina Mosicheva

University «Moscow Power Engineering Institute», Moscow, Russia

**Svetlana Parfenova, Vladislava Dolgova, Karina Bezrodnova,
Elena Lyagushkina, Victor Bogatov, Nadezhda Khaltakshinova,
Vladimir Korobatov**

*Russian Research Institute for Economy, Policy
and Law in Sci-tech Sphere, Moscow, Russia*

Irina Mikhailenko

*Russian National Public Library for Science and Technology,
Moscow, Russia*

Forecasting the number of publication based on Web of Science and Scopus data integral index

The study is completed within the framework of the state order of the Ministry of Education and Science of the Russian Federation “Monitoring and analysis of publication and patent activities in Russian and global science to support the national policy in science and technology” (Project No. 28.12616.2018/12.1). The authors aim at building relevant instruments for assessing and forecasting the total amount of Russian science publication based on the data of two global science citation systems, i. e. Web of Science and Scopus, with de-duplicating by digital object identifier used in both systems. The related tasks are defined: selection of the method to forecast the number of Web of Science indexed articles within the “back forecast” method; development of integral index to calculate the number of publication based on Web of Science and Scopus data; calculation and forecast of the number of articles and publications by G7 and BRICS states up to the year 2021 with scientometric methods based on Web of Science and Scopus data. The authors suggest a conceptually new method to calculate the integral index of publications based on Web of Science and Scopus data to exclude duplication. The total number of academic articles and publications (based on Web of Science and Scopus data) is calculated for the period 2007—2016, and the forecast is provided up to the year 2021.

Keywords: extrapolation method, integral index, publication number, Web of Science, Scopus, forecast.

The goal is to create a tool for calculating and forecasting the cumulative number of publications of Russian researchers based on the Web of Science and Scopus data, with the dedication of the digital object identifier (DOI) used in both systems. The objectives are: a choice of a way of forecasting of number of the scientific articles indexed in Web of Science, within the framework of a method

«forecasting back»; development of an integrated indicator of the calculation of the number of publications based; calculation and forecast of the number of scientific articles and publications on the G7 and BRICS countries until 2021. The most common method used is extrapolation, based on determining the patterns of development within the dynamic series and preserving the found trend beyond the limits of the analyzed period. Advantages of this method are the small laboriousness of the calculating apparatus and the universal possibility of application. The disadvantages include the need for strict adherence to the rule – the anticipated period should be three times greater than the period of anticipation. The article presents the integral index developed by the authors, which allows to calculate the aggregate number of publications according to the Web of Science and Scopus data, excluding duplication in the case of publications. The method of "forecasting backwards" is based on the choice of the method of forecasting the number of publications, namely, the method of extrapolation according to the average growth rate. A forecast was made up to 2021 on the aggregate number of scientific articles and publications (according to Web of Science and Scopus) in G7 and BRICS. The forecast showed that in the next five years the number of publications will grow in China (average growth rate – 1.06), India (average growth rate – 1.09), Russia (average growth rate – 1.07), Brazil (average rate growth – 1.04) and Germany (average growth rate – 1.03). The cumulative number of scientific articles and publications of US researchers in the same period will decline (the average growth rate is 0.98).

Одно из ключевых направлений интеграции российской науки в мировую – представление релевантных и значимых научных результатов, полученных отечественными исследователями, в авторитетных международных и российских научных изданиях, которые отражаются в глобальных индексах научного цитирования *Web of Science* и *Scopus*. Чтобы оценить включённость российской науки в международное научное пространство, проводятся различные наукометрические измерения, основанные на библиометрических данных о публикациях.

Наукометрия в России – относительно молодая наука [1]: она находится на этапе становления и развития. Сегодня активно разрабатываются и внедряются инструменты наукометрии, ориентированные на принятие управленческих решений в сфере науки и технологий и обоснование научно-технической политики, в том числе форсайт-инструменты, которые позволяют проводить сравнительную оценку публикационных потоков российских и зарубежных исследователей в среднесрочной и долгосрочной перспективах [2].

Отметим, что методики расчёта показателей публикационной активности, рассматриваемые в научной литературе, оперируют данными какой-либо одной из глобальных систем научного цитирования – *Web of Science* или *Scopus*. Это не позволяет проводить объективную оценку публикационного потока страны по всем областям науки, так как их представленность в каждой из систем различна. Разработка методического инструментария для определения совокупного «чистого» числа публикаций на основе данных *Web of Science* и *Scopus* и построения прогноза открывает новые возможности использования наукометрических измерений. Поэтому обозначенная тема исследования чрезвычайно актуальна.

Авторы исследования поставили перед собой цель – сформировать релевантный инструментарий расчёта и прогноза совокупного числа публикаций российских исследователей на основе данных *Web of Science* и *Scopus*, с дедубликацией по Цифровым идентификаторам объекта (ЦИО, *digital object identifier, DOI*), используемым в обеих системах. При таком методе дедубликации потери (равномерно с 2008 г.) публикаций, которым не присвоены *DOI*, составляют около 10%.

Для достижения этой цели определены задачи:

выбор способа прогнозирования числа научных статей, индексируемых в *Web of Science*, в рамках метода «прогнозирование назад»;

разработка интегрального показателя расчёта числа публикаций на основе данных *Web of Science* и *Scopus*;

расчёт и прогноз числа научных статей и публикаций по странам G7 и BRICS до 2021 г. с помощью разработанного наукометрического инструментария (по данным *Web of Science* и *Scopus*).

Обзор публикаций

Методы прогнозирования числа публикаций исследованы в научных работах по наукометрии зарубежных и российских авторов. Приведём несколько примеров.

Авторы статьи «*Forecasting the trend of international scientific collaboration*» [3] рекомендуют использовать и комбинировать статистические методы для построения прогноза по библиометрическим данным в зависимости от целей и задач исследования. Полученные результаты могут быть использованы специалистами, принимающими управленческие решения: в части финансовой поддержки исследований, при создании новых технологий, построении дорожных карт по их разработке и замещении старых отработанных технологий.

В статье «*Forecasting trends of development of psychology from a bibliometric perspective*» [4] предложено строить прогноз числа публикаций на упреждающий период (10 лет) методом экспоненциального сглаживания на

основе ретроспективного временного ряда, содержащего сведения о числе публикаций за 40 лет. Число публикаций рекомендуется определять в специализированных базах данных (например, *PsycINFO*, *PSYINDEX*) [4, 5]. Это связано с тем, что публикации по психологии в глобальных индексах цитирования представлены недостаточно.

Авторы публикации «*The evolution of the use of Foresight methods: a scientometric analysis of global FTA research output*» [6] провели анализ наиболее часто используемых методов Форсайта. Инструменты наукометрии предложено использовать для анализа прошлых тенденций и прогнозирования будущих направлений развития на основе метаданных публикаций и патентов путём изучения, организации и объяснения возникающих тенденций и закономерностей [7]. Библиометрический и патентный анализ являются одними из наиболее часто используемых методов Форсайта и позволяют синтезировать количественные и качественные оценки.

В статье «*Time-aware link prediction to explore network effects on temporal knowledge evolution*» [8] изложены методы прогнозирования, основанные на топологических оценках подобия и методах временных рядов прогнозирования ссылок с целью определения будущих тенденций развития научных концепций.

Анализ методов прогнозирования, представленных в научных работах российских исследователей, позволил классифицировать методы на основе следующих аспектов:

формализация – интуитивные (качественные) и формализованные (количественные) [9];

информационное основание метода (фактографические, комбинированные и экспертные) [Там же];

применяемая методология (нормативные или целевые, исследовательские или поисковые и комплексные) [Там же];

степень разброса прогнозируемых значений показателя (точечные и интервальные) [Там же];

особенности объектов прогнозирования (экономические, научно-технические, демографические, социологические и др.) [10].

Самый распространённый из применяемых методов – экстраполяция, основанная на определении закономерностей развития внутри динамического ряда и сохранении найденной тенденции за пределами анализируемого периода [11]. Достоинствами этого метода являются небольшая трудоёмкость расчётного аппарата и универсальная возможность применения. К недостаткам можно отнести необходимость строгого соблюдения правила – предпрогнозный период должен быть в три раза больше, чем период упреждения [10].

Наиболее универсальные способы экстраполяции основаны на использовании: среднего уровня ряда, среднего абсолютного прироста, среднего темпа роста, скользящей средней, функции аналитического выравнивания [12].

Таким образом, перед наукометрией ставятся новые задачи, связанные с развитием её инструментария. Важная роль отводится совершенствованию прогнозирования на основе статистических методов с использованием различных информационных полей, содержащих сведения о публикациях, по традиционным направлениям (прогнозирование числа публикаций в разрезе стран, тематик исследований, журналов) и по новым синтезированным направлениям, базирующимся на количественных и качественных оценках (Форсайт-технологии по определению вновь возникающих научных концепций, технологических направлений и др.).

Методика проведения исследования

С целью выбора релевантного метода прогнозирования числа публикаций проведён анализ методов экстраполяции способом «прогнозирование назад». На основе ретроспективных данных глубиной 15 лет (с 1997 г. по 2011 г.) сделан прогноз на пять лет (с 2012 г. по 2016 г.) различными методами. Полученные прогнозные ряды сопоставлены с фактическими данными за этот период. Выбран метод, имеющий наименьшую ошибку.

Для проверки достоверности методов прогнозирования в качестве исходных взяты данные о числе научных статей, индексируемых в *Web of Science* по трём индексам научного цитирования (*Science Citation Index Expanded – SCI-EXPANDED*, *Social Sciences Citation Index – SSCI*, *Arts & Humanities Citation Index – A&HCI*). Фактическое число статей за период с 1997 г. по 2016 г. приведено в табл. 1.

Таблица 1

Число научных статей в мировых научных журналах, индексируемых в *Web of Science* за период 1997–2016 гг.

Период	Число научных статей в мировых научных журналах, тыс. ед.	Темп роста, в коэффициентах
1997	757	1,01
1998	771	1,02
1999	784	1,02
2000	798	1,02
2001	797	1

Период	Число научных статей в мировых научных журналах, тыс. ед.	Темп роста, в коэффициентах
2002	814	1,02
2003	849	1,04
2004	885	1,04
2005	929	1,05
2006	977	1,05
2007	1 024	1,05
2008	1 095	1,07
2009	1 139	1,04
2010	1 186	1,04
2011	1 262	1,06
2012	1 323	1,05
2013	1 396	1,06
2014	1 434	1,03
2015	1 481	1,03
2016	1 517	1,02

(Источник: *Web of Science – SCI-EXPANDED, SSCI, A&HCI*; дата обращения: 02.11.2017 г.)

Метод среднего уровня ряда является точечным прогнозом и считается самым простым из вышеперечисленных. Он хорошо подходит для изучения социально-экономических явлений, которые носят стационарный характер. Выдвинута гипотеза о том, что значение прогнозируемого уровня динамического ряда равно среднему значению уровней анализируемого периода [13]:

$$\hat{y}_{t+1} = \bar{y}, \quad (1)$$

где

\hat{y}_{t+1} – прогнозное значение на период упреждения $t+1$;

\tilde{y} – средний уровень ряда за 15 лет.

Средний уровень ряда рассчитан по формуле средней арифметической.

Определён доверительный интервал прогноза, в котором предположительно будет находиться прогнозируемая средняя, по формуле (2) [11]:

$$\hat{y}_t \pm t \times \Delta_{\tilde{y}}, \quad (2)$$

где

\tilde{y} – средний уровень ряда,

t – коэффициент доверия,

$\Delta_{\tilde{y}}$ – предельная ошибка выборки.

Рассчитано прогнозное значение среднего уровня ряда на 2016 г. по формуле средней арифметической – 1 106 тыс. ед.

Среднее квадратическое отклонение (σ) равно 231, средняя ошибка выборки (μ) – 62, а предельная ошибка выборки ($\Delta_{\tilde{y}}$) (при условии, что $t = 2$) будет составлять 123. Таким образом, доверительный интервал имеет следующий вид:

$$(\tilde{y} - \Delta_{\tilde{y}} \leq \hat{y}_{t+1} \leq \tilde{y} + \Delta_{\tilde{y}}), \text{ т.е. } 1\,106 - 123 \leq \hat{y}_t \leq 1\,106 + 123,$$

т.е. прогнозируемое среднее значение количества научных статей на 2016 г. находится в интервале от 983 до 1 229 ед.

На основе ретроспективной выборки за 15 лет (1997–2011 г.) спрогнозированы значения числа научных статей на период с 2012 г. по 2016 г., проведено сравнение полученных значений с фактическими (табл. 2).

Прогнозирование числа научных статей методом среднего уровня ряда

Период, год	Фактическое число статей, тыс. ед. (Y_t)	Прогнозируемое число статей, тыс. ед. (\hat{Y}_t)	Доверительный интервал ($\tilde{y} - \Delta_{\tilde{y}} \leq \hat{y}_t \leq \tilde{y} + \Delta_{\tilde{y}}$)	$ly_t - \hat{Y}_t l$	$ly_t - \hat{Y}_t l / y_t$
2012	1 323	938	$794 \leq \hat{y}_{t+1} \leq 1 082$	385	0,291156
2013	1 396	976	$840 \leq \hat{y}_{t+1} \leq 1 111$	420	0,301194
2014	1 436	1 017	$888 \leq \hat{y}_{t+1} \leq 1 146$	419	0,291643
2015	1 481	1 061	$936 \leq \hat{y}_{t+1} \leq 1 185$	420	0,283817
2016	1 517	1 106	$983 \leq \hat{y}_{t+1} \leq 1 229$	411	0,270798
Итого:					1,438609

(Источник: Web of Science – SCI-EXPANDED, SSCI, A&HCI; дата обращения: 02.11.2017 г.)

Полученные расчёты (табл. 2) показывают, что ни одно из фактических значений не попало в рассчитанный доверительный интервал, а прогнозируемое число научных статей намного меньше фактического, о чём свидетельствует высокая ошибка аппроксимации (29%). Следовательно, этот метод не является корректным для решения поставленной задачи.

Метод среднего абсолютного прироста, как и предыдущий, отличается небольшой трудоёмкостью и используется в тех случаях, когда предполагается, что уровни динамического ряда изменяются в геометрической прогрессии, а тенденция развития изучаемого явления выражается линейной моделью аналитического выравнивания.

Прогнозные значения, исходя из среднего абсолютного прироста, определены по формуле (3) [11]:

$$\hat{y}_t = a_0 + a_1 \cdot t, \quad (3)$$

где \hat{y}_t – прогнозируемый уровень ряда динамики;

a_0 – последний из известных уровней ряда y_n ;

a_1 – средний абсолютный прирост, рассчитанный по формуле средней арифметической;

t – период упреждения.

Метод прогнозирования проверен с помощью средней ошибки аппроксимации, которая рассчитана по формуле (4) [13]:

$$A = \frac{1}{n} \cdot \sum \frac{|y_t - \hat{y}_t|}{y_t} \cdot 100, \quad (4)$$

где Y_t – фактический уровень ряда динамики,

n – количество ошибок аппроксимации (от $i = 1, 2 \dots$ до n).

Если средняя ошибка аппроксимации меньше 15%, то анализируемый метод прогнозирования может быть использован для исследуемого социально-экономического явления.

На основе среднего абсолютного прироста ($a_1 = 21$ тыс. научных статей) научных статей, индексируемых в *Web of Science (SCI-EXPANDED, SSCI, A&HCI)* за 15 лет (1997–2011 гг.), спрогнозировано число научных статей на период с 2012 г. по 2016 г. (табл. 3).

Таблица 3

Прогноз числа научных статей методом среднего абсолютного прироста

Период, год	Фактическое число научных статей, тыс. ед. (y_t)	Прогнозируемое число научных статей, тыс. ед. (\hat{y}_t)	$y_t - \hat{y}_t$	$ y_t - \hat{y}_t / y_t$
2012	1 323	1 284	39	0,029686006
2013	1 396	1 305	91	0,06536504
2014	1 436	1 327	109	0,076137456
2015	1 481	1 348	133	0,089772765
2016	1 517	1 445	73	0,04791935
Итого:				0,308880615

(Источник: *Web of Science – SCI-EXPANDED, SSCI, A&HCI*; дата обращения: 02.11.2017 г.)

Средняя ошибка аппроксимации (\bar{A}) составила 6,2% (что меньше 15%), следовательно, этот метод прогнозирования может быть использован для изучения рассматриваемого явления.

Метод среднего темпа (или коэффициента) роста используется, если выдвигается гипотеза, что уровни динамического ряда изменяются в геометрической прогрессии, а явление подчиняется законам экспоненциальной кривой. Прогнозное значение уровня, исходя из среднего коэффициента роста, рассчитано по формуле (5) [11]:

$$\hat{y}_t = a_0 \cdot \bar{k}^t, \quad (5)$$

где \bar{k} – средний коэффициент роста, рассчитанный по формуле средней геометрической простой –

$$\bar{k} = \sqrt[n-1]{\prod_{i=1}^{n-1} k_i};$$

n – число коэффициентов роста (от $i = 1, 2, \dots$ до n).

Метод прогнозирования проверен с помощью расчёта средней ошибки аппроксимации (табл. 4).

Таблица 4

Прогноз числа научных статей методом среднего темпа роста

Период, год	Фактическое число статей, тыс. ед. (y_t)	Прогнозируемое число статей, тыс. ед. (\hat{y}_t)	$y_t - \hat{y}_t$	$ y_t - \hat{y}_t /y_t$
2012	1 323	1 307	16	0,01205
2013	1 396	1 353	43	0,03081
2014	1 436	1 401	35	0,02403
2015	1 481	1 451	30	0,02017
2016	1 517	1 502	15	0,00982
Итого:				0,09689

(Источник: *Web of Science – SCI-EXPANDED, SSCI, A&HCI*; дата обращения: 02.11.2017 г.)

Средний темп роста (\bar{k}) составил 104% (или 1,04), средняя ошибка аппроксимации (\bar{A}) – 1,9%. По сравнению с предыдущим методом, у которого ошибка аппроксимации была 6,2%, этот метод представляется более корректным.

Метод скользящей средней – самый распространённый метод сглаживания динамических рядов – используется как один из этапов построения математической модели прогнозирования. Но его также можно применять в качестве самостоятельного метода прогнозирования. Он «состоит в замене фактических значений показателя их усреднёнными величинами, расчёт которых проводят путём последовательного смещения начала отсчёта на единицу времени (скольжения), т.е. постепенно исключают из интервала первые уровни и включают последующие. Полученная средняя относится к середине укрупнённого интервала» [11. С. 153]. Усреднение можно проводить по 3, 5 и более «точкам» (уровням ряда). Однако получаемый при этом сглаженный ряд будет сильно укорочен, поэтому чаще всего на практике применяются трёхчленные средние – формула (6):

$$\bar{y}_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}, \quad (6)$$

где y_{t-1}, y_{t+1} – соответственно предыдущий и последующий уровни ряда от выбранного уровня ряда динамики.

Недостаток такого метода – укорачивание исходного ряда динамики на $(n - 2)$ уровней, где n – число уровней первоначального ряда, поэтому на практике он чаще всего используется только для краткосрочного прогнозирования на один, максимум два периода упреждения.

Расчёт прогнозируемого значения по трём «точкам» осуществлялся по формуле (7) [14]:

$$\hat{y}_{t+1} = \bar{y}_{t-1} + \frac{1}{3} \cdot (y_t - y_{t-1}), \quad (7)$$

где \bar{y}_{t-1} – скользящая средняя за два периода до прогнозного;

y_t – фактическое значение исследуемого явления за предшествующий период;

y_{t-1} – фактическое значение исследуемого явления за два периода, предшествующих прогнозируемому.

Прогноз числа научных статей методом скользящей средней по трём «точкам» представлен в табл. 1, расчётные данные – в табл. 5.

Таблица 5

Прогноз числа научных статей методом скользящей средней по трём «точкам»

Период, год	Фактическое количество статей без ESCI, тыс. ед. (y_t)	Скользящая средняя по трём «точкам», тыс. ед. (\bar{y}_{t-1})	Прогнозируемое количество статей без ESCI, тыс. ед. (\hat{y}_{t+1})	$y_t - \bar{y}_{t-1}$	$y_t - \bar{y}_{t-1} / y_t$
2010	1 186	1 196	–	–	–
2011	1 262	1 223	–	–	–
2012	1 323	1 231	1 221	102	0,077035
2013	1 396	1 219	1 210	187	0,133816
2014	1 436	1 221	1 227	209	0,145359
2015	1 481	1 221	1 225	256	0,172704
2016	1 517	–	1 220	297	0,196005
Итого:					0,724919

(Источник: Web of Science – SCI-EXPANDED, SSCI, A&HCI; дата обращения: 02.11.2017 г.)

Средняя ошибка аппроксимации (\bar{A}) составила 14%, что говорит о существенном отклонении прогнозируемых значений от фактических. Следовательно, этот метод по сравнению с двумя предыдущими наилучшим образом отражает тенденцию изменения числа научных статей в мировых научных журналах, индексируемых в *Web of Science (SCI-EXPANDED, SSCI, A&HCI)*.

Метод аналитического выравнивания довольно широко распространён среди математических методов. В ряде классификаций методов прогнозирования он выделяется в отдельную группу и противопоставляется методам экстраполяции, а в некоторых случаях относится к ним.

Суть метода «заключается в том, что находится уравнение $\hat{y}_t = f(t)$, график которого наилучшим образом отражает основную тенденцию ряда динамики. Аналитическое выравнивание позволяет не только определить основную тенденцию развития явления во времени, но и выполнять расчёты для таких периодов, по которым нет информации» [11. С. 153].

«Достоинствами данного метода является универсальность его применения, широкий набор функциональных зависимостей; возможность включения в статистическую модель в качестве самостоятельной переменной фактора времени и пр.» [15. С. 21]. Его недостатки – высокая трудоёмкость подбора параметров уравнений, а также сложность интерпретации полученной математической модели в практических целях. Для подбора типа функциональной зависимости использован пакет *Excel*, а для проверки качества подобранной модели – коэффициент детерминации R^2 (чем больше это значение, тем качественнее подобранная прогнозная модель).

Для прогноза числа научных статей лучше всего подходят несколько функциональных зависимостей (коэффициентов детерминации):

степенная функция ($R^2 = 0,9763$),

экспоненциальная функция ($R^2 = 0,9766$),

полиномиальная функция ($R^2 = 0,9926$).

Однако все три функции в дальнейшем прогнозируют чрезмерно высокий рост числа научных статей, индексируемых в *Web of Science (SCI-EXPANDED, SSCI, A&HCI)*, что не подтверждается фактическими значениями динамического ряда. Поэтому использовать метод аналитического выравнивания для построения прогноза числа публикаций нецелесообразно.

Анализ ошибки аппроксимации (табл. 6) всех представленных методов позволяет рекомендовать для прогнозирования числа научных статей метод среднего темпа роста.

Сводная таблица значений показателя «ошибка аппроксимации»

Наименование метода прогнозирования	Значение ошибки аппроксимации, %
Метод среднего уровня ряда	29,0
Метод среднего абсолютного прироста	6,2
Метод среднего темпа (или коэффициента) роста	1,9
Метод скользящей средней	14,0

Прогноз числа научных статей предложено построить на основе ретроспективных данных о количестве научных статей, индексируемых в *Web of Science* и *Scopus* за 10 лет. Для исключения дублирования научных статей, одновременно индексируемых в *Web of Science* и *Scopus*, разработан интегральный показатель, формула (8):

$$N = (X_w + Y_S) + Z_{wS}, \quad (8)$$

где

X_w – число научных статей, индексируемых в *Web of Science* (без дублирования в *Scopus*);

Y_S – число научных статей, индексируемых в *Scopus* (без дублирования в *Web of Science*);

Z_{wS} – число научных статей, одновременно индексируемых в *Web of Science* и *Scopus*.

Число научных публикаций, индексируемых в *Web of Science* без дублирования в *Scopus*, получено путём дедубликации по цифровым идентификаторам объектов (*DOI*), используемых в глобальных системах научного цитирования, как и число публикаций в *Scopus* без дублирования в *Web of Science*.

Как видно из формулы (8), число научных статей, индексируемых в *Web of Science* и *Scopus*, учитывается один раз, что позволяет исключить дублирование этих статей в двух системах научного цитирования.

Алгоритм расчёта числа научных статей, одновременно индексируемых в *Web of Science* и *Scopus*, сводится к следующим шагам:

формирование исходного массива научных статей, индексируемых только в *Web of Science*, на основе информетрического анализа сведений о цифровых идентификаторах объектов (X_w);

формирование исходного массива научных статей, индексируемых только в *Scopus*, на основе информетрического анализа сведений об их цифровых идентификаторах объектов (Y_S);

формирование массива научных статей, индексируемых одновременно в *Web of Science* и *Scopus*, на основе информетрического анализа сведений об их цифровых идентификаторах объектов (Z_{wS});

расчёт совокупного числа научных статей, индексируемых в *Web of Science* и *Scopus*, путём суммирования научных статей, индексируемых только в *Web of Science* (X_w), и статей, индексируемых только в *Scopus* (Y_S);

учёт числа научных статей (один раз без дублирования), одновременно индексируемых в *Web of Science* и *Scopus*.

На основе формулы (8) можно рассчитать удельный вес научных статей любой страны на основе данных *Web of Science* и *Scopus*, формула (9):

$$D = \frac{N}{(W_w + S_S) + R_{wS}} \times 100, \quad (9)$$

где

W_w – общемировой поток публикаций, индексируемых только в *Web of Science* (без дублирования в *Scopus*);

S_S – общемировой поток публикаций, индексируемых только в *Scopus* (без дублирования в *Web of Science*);

R_{wS} – общемировой поток публикаций, одновременно индексируемых в *Web of Science* и *Scopus*.

Первый показатель (формула 8) отражает соответствие уровня исследований фундаментальной науки конкретной страны мировому уровню, а второй показатель (формула 9) – вклад от результатов исследований анализируемой страны в мировую науку.

Результаты исследования

В соответствии с поставленной в исследовании задачей сделан расчёт и построен прогноз числа научных статей и публикаций на основе данных *Web of Science* и *Scopus* по странам G7 и BRICS до 2021 г.

Количество научных статей и публикаций по странам G7 и BRICS вычислено по формуле (8) на основе данных *Web of Science* и *Scopus* (табл. 7, 8). Число научных статей по данным *Web of Science* определено по четырём индексам научного цитирования (*SCI-EXPANDED*, *SSCI*, *A&HCI*, *ESCI*), тип публикации – «научная статья»; число публикаций – по восьми индексам научного цитирования (*SCI-EXPANDED*, *SSCI*, *A&HCI*, *ESCI*, *CPCI-S*, *CPCI-SSH*, *BKCI-S*, *BKCI-SSH*), тип публикации: «все типы публикаций».

Прогноз построен методом среднего темпа роста на основе двух типов динамических рядов, первый – по научным статьям, второй – по всем типам публикаций. Прогнозы числа научных статей и публикаций стран G7 в мировых научных журналах, индексируемых *Web of Science* и *Scopus*, приведены соответственно на рис. 1 и 2.

Таблица 7

**Интегральный показатель стран G7 (2007–2016 гг.), ед.
(П – все типы публикаций, НС – научные статьи)**

Страна	Тип публикации	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Велико-британия	П	294 571	228 647	235 641	241 893	248 249	260 651	267 814	265 130	275 676	271 010
	НС	167 115	108 871	110 648	112 658	116 429	119 707	129 052	130 884	137 493	139 327
Германия	П	250 753	197 217	201 671	206 921	208 217	215 499	215 752	222 817	221 923	223 552
	НС	96 452	97 677	99 522	99 992	107 629	110 818	119 577	116 964	123 364	127 721
Италия	П	141 296	110 369	117 574	119 168	119 760	126 427	132 484	139 797	147 373	144 944
	НС	86 943	59 066	61 385	60 638	63 608	66 755	73 478	75 852	80 598	80 703
Канада	П	152 159	120 620	124 041	125 800	127 514	133 127	133 788	137 598	139 423	138 195
	НС	92 324	63 899	65 695	65 919	67 462	69 418	72 905	75 264	78 204	77 633
США	П	1 045 741	829 027	847 924	859 879	874 715	913 057	906 410	915 288	921 876	905 089
	НС	607 098	420 730	432 502	439 620	458 120	457 134	475 338	489 133	492 951	488 879
Франция	П	170 096	133 946	139 639	141 755	142 118	146 357	151 587	153 566	153 805	151 806
	НС	109 263	75 065	77 420	77 487	79 874	81 564	85 832	87 092	88 652	89 845
Япония	П	228 336	170 494	172 363	169 875	167 009	170 362	170 177	168 273	160 923	164 990
	НС	152 362	101 646	100 917	95 754	97 615	95 975	98 106	96 339	95 997	97 044

Таблица 8

Интегральный показатель стран BRICS (2007–2016 гг.), ед.
(П – все типы публикаций, НС – научные статьи)

Страна	Тип публикации	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Бразилия	П	67 421	64 327	69 447	74 185	77 784	82 821	84 088	84 412	91 629	94 803
	НС	50 341	46 299	50 037	51 917	55 457	58 162	58 760	58 712	65 323	68 433
Россия	П	55 545	51 524	50 067	51 231	54 141	54 899	58 021	68 900	83 927	100 008
	НС	41 731	38 124	36 807	36 711	39 851	38 272	41 444	46 214	58 327	64 837
Индия	П	91 519	84 493	90 342	101 312	118 641	130 626	141 897	161 968	183 647	195 258
	НС	69 725	58 140	62 431	69 967	79 983	84 573	90 555	101 262	116 541	118 043
Китай	П	360 280	350 895	395 687	431 033	472 249	492 090	529 772	575 677	568 958	606 467
	НС	254 641	219 015	237 785	247 086	270 025	299 142	341 061	378 320	405 419	423 706
ЮАР	П	17 791	15 124	16 707	17 602	19 650	21 966	22 770	25 617	27 535	29 091
	НС	12 083	9 461	10 208	10 960	12 569	13 293	14 379	16 988	18 036	19 067

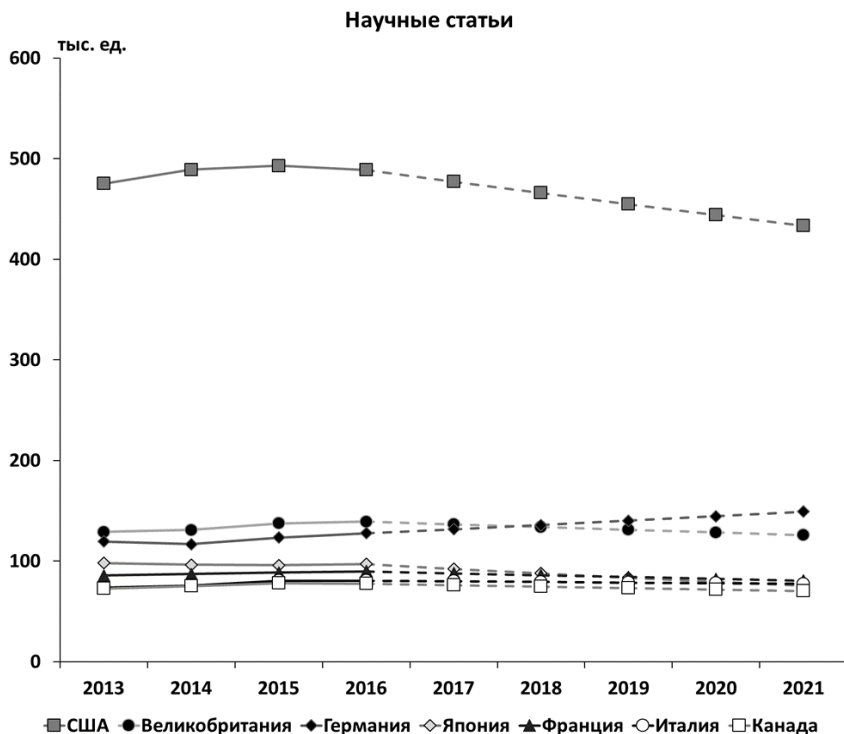


Рис. 1. Прогноз числа научных статей стран G7 по данным *Web of Science* и *Scopus* (дата обращения: 05.02.2018 г.)

Рост числа научных статей в пятилетней перспективе (2017–2021 гг.) по данным *Web of Science* и *Scopus* прогнозируется только в Германии (средний темп роста – 1,03) (рис. 1). Выявлена тенденция сокращения совокупного числа научных статей исследователей США. По данным прогноза, число научных статей в США в 2021 г. сократится по сравнению с 2016 г. на 11% (средний темп роста – 0,98). В других странах G7 число научных статей сохранится на том же уровне.

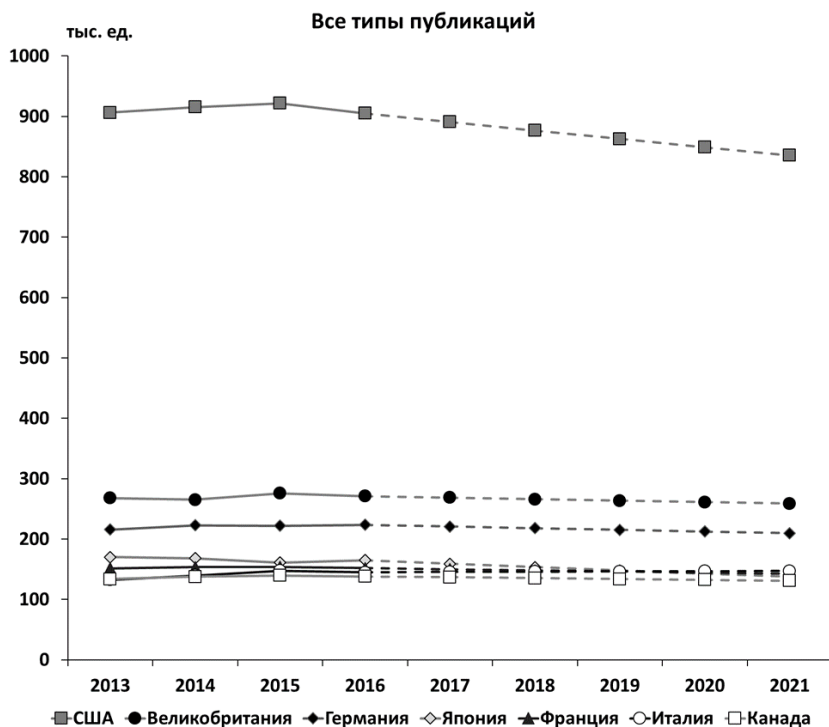


Рис. 2. Прогноз числа публикаций в странах G7 по данным *Web of Science* и *Scopus* (дата обращения: 05.02.2018 г.)

Совокупное число публикаций стран G7 (по данным *Web of Science* и *Scopus*) в 2017–2021 гг. практически не изменится. Исключение составляют США: согласно прогнозу, совокупное число публикаций, аффилированных с США, в 2021 г. сократится по сравнению с 2016 г. на 8% (средний темп роста – 0,98).

Прогнозы количества научных статей и публикаций стран BRICS в мировых научных журналах, индексируемых *Web of Science* и *Scopus*, приведены соответственно на рис. 3 и 4.

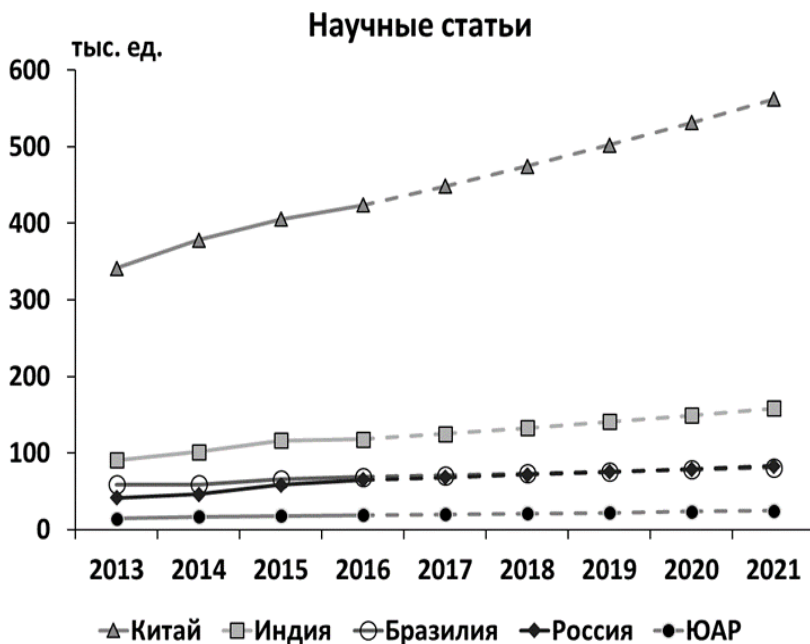


Рис. 3. Прогноз числа научных статей в странах BRICS по данным *Web of Science* и *Scopus* (дата обращения: 05.02.2018 г.)

Согласно прогнозу (рис. 3), значительно возрастет совокупное число научных статей исследователей из Китая. В 2021 г. количество статей, аффилированных с Китаем, увеличится по сравнению с 2016 г. на 33% (средний темп роста – 1,06). Положительная динамика наблюдается в Индии, России и Бразилии, а именно: число научных статей в 2021 г. увеличится по сравнению с 2016 г. в Индии – на 34% (средний темп роста – 1,06), в России – на 28% (средний темп роста – 1,05), в Бразилии – на 19% (средний темп роста – 1,03).

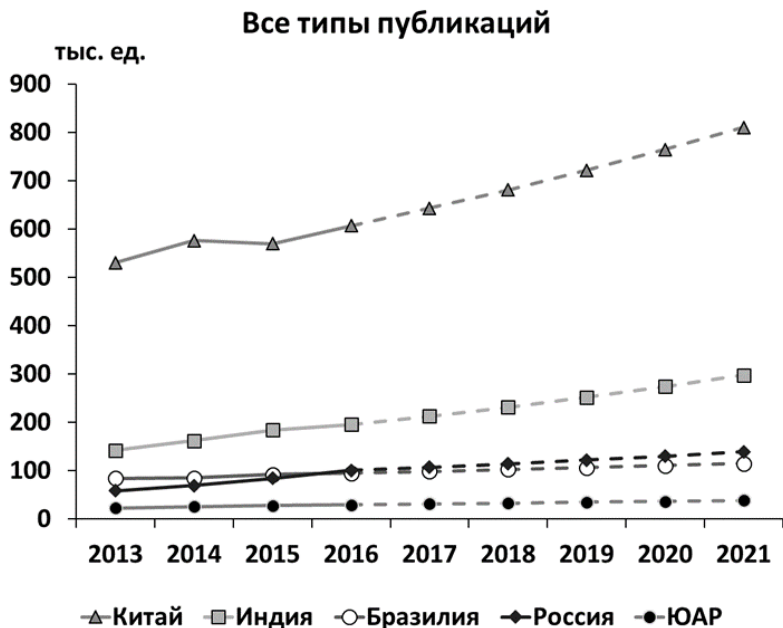


Рис. 4. Прогноз числа публикаций в странах BRICS по данным *Web of Science* и *Scopus* (дата обращения: 05.02.2018 г.)

Прогноз относительно совокупного числа публикаций (по данным *Web of Science* и *Scopus*) по странам BRICS такой: с 2017 г. по 2021 г. число публикаций исследователей Китая вырастет на 34% (средний темп роста – 1,06), Индии – на 52% (средний темп роста – 1,09), России – на 39% (средний темп роста – 1,07), Бразилии – на 21% (средний темп роста – 1,04).

Выводы

В статье представлен разработанный авторами интегральный показатель, который позволяет осуществлять расчёт совокупного числа публикаций по данным *Web of Science* и *Scopus*, исключая дублирование при учёте публикаций. Способом «прогнозирования назад» обоснован выбор метода прогнозирования числа публикаций, а именно – метод экстраполяции по среднему темпу роста.

Сделан прогноз до 2021 г. относительно совокупного числа научных статей и публикаций (по данным *Web of Science* и *Scopus*) в странах G7 и BRICS. Прогноз показал, что в ближайшие пять лет число публикаций бу-

дет расти в Китае (средний темп роста – 1,06), Индии (средний темп роста – 1,09), России (средний темп роста – 1,07), Бразилии (средний темп роста – 1,04) и Германии (средний темп роста – 1,03). Совокупное число научных статей и публикаций (по данным *Web of Science* и *Scopus*) исследователей США в этот же период будет сокращаться (средний темп роста – 0,98).

Представленный инструментарий предназначен для проведения меж-
страновых сопоставлений и наукометрического анализа фактических и про-
гнозируемых показателей публикационной активности.

СПИСОК ИСТОЧНИКОВ

1. **Налимов В. В., Мульченко З. М.** Наукометрия. – Москва : Наука, 1969. [Электрон-
ный ресурс]. – Режим доступа: <http://xn----8sbec3arien.xn--p1ai/books/128/490> (дата обращения:
12.03.2018).

*Nalimov V. V., Muľchenko Z. M. Naukometriia. – Moskva : Nauka, 1969. [E`lektronny`i
resurs]. – <http://xn----8sbec3arien.xn--p1ai/books/128/490> (data obrashcheniia: 12.03.2018).*

2. **Земсков А. И.** Библиометрия, вебметрики, библиотечная статистика : учеб. пособие. –
Москва : ГПНТБ России, 2017. – 136 с.

*Zemskov A. I. Bibliometriia, vebmetriki, bibliotechnaia statistika : ucheb. posobie. – Moskva :
GPNTB Rossii, 2017. – 136 s*

3. **Varun Shrivats S., Bhattacharya S.** Forecasting the trend of international scientific collab-
oration. *Scientometrics*, 101, 1941–1954. – DOI: 10.1007/s11192-014-1364-x

4. **Krampen G., von Eye A., Schui G.** Forecasting trends of development of psychology from
a bibliometric perspective. *Scientometrics*, 87, 687–694. – DOI: 10.1007/s11192-011-0357-2

5. **Tsai H.-H.** Global data mining: An empirical study of current trends, future forecasts and
technology diffusions. *Expert Systems with Applications*, 39, 8172–8181. – DOI: 10.1016/
j.eswa.2012.01.150

6. **Saritas O., Burmaoglu S.** The evolution of the use of Foresight methods: a scientometric
analysis of global FTA research output. *Scientometrics*, 105, 497–508. – DOI: 10.1007/s11192-015-
1671-x

7. **Hess D. J.** Science studies: An advanced introduction. – New York : New York University
Press.

8. **Choudhury N., Uddin S.** Time-aware link prediction to explore network effects on tem-
poral knowledge evolution. *Scientometrics*, 108, 745–776. – DOI: 10.1007/s11192-016-2003-5

9. **Крук Д. М., Лукин В. С., Мосин Е. А. и др.** Основы экономического и социального
прогнозирования. – Москва : Высшая школа, 1985. – 200 с.

*Kruk D. M., Lukin V. S., Mosin E. A. i dr. Osnovy ekonomicheskogo i sotsialnogo prognozi-
rovaniya. – Moskva : Vysshaya shkola, 1985. – 200 s.*

10. **Денискин В. В.** Основы экономического прогнозирования в пищевой промышленности. – Москва : Колос, 1993. – 192 с.

Deniskin V. V. Osnovy ekonomicheskogo prognozirovaniya v pishchevoy promyshlennosti. – Moskva : Kolos, 1993. – 192 s.

11. **Долгова В. Н., Медведева Т. Ю.** Статистика : учеб. и практикум для бакалавров. – 2-е изд., перераб. и доп. – Москва : Юрайт, 2014. – 627 с.

Dolgova V. N., Medvedeva T. Yu. Statistika : ucheb. i praktikum dlya bakalavrov. – 2-e izd., pererab. i dop. – Moskva : Yurayt, 2014. – 627 s.

12. **Салин В. Н., Чурилова Э. Ю.** Курс теории статистики для подготовки специалистов финансово-экономического профиля : учеб. / В. Н. Салин, Э. Ю. Чурилова. – Москва : Финансы и статистика. – 2006. – 480 с.

Salin V. N., Churilova E. Yu. Kurs teorii statistiki dlya podgotovki spetsialistov finansovo-ekonomicheskogo profilya : ucheb. / V. N. Salin, E. Yu. Churilova. – Moskva : Finansy i statistika. – 2006. – 480 s.

13. **Репова М. Л.** Эконометрика : метод. указания и задания к выполнению контрольной работы / М. Л. Репова. – Архангельск : Архангельский государственный технологический университет, 2010. – 32 с.

Repova M. L. Ekonometrika : metod. ukazaniya i zadaniya k vypolneniyu kontrolnoy raboty / M. L. Repova. – Arhangelsk : Arhangelskiy gosudarstvennyy tehnologicheskiy universitet, 2010. – 32 s.

14. **Новикова Н. В., Поздеева О. Г.** Прогнозирование национальной экономики : учеб.-метод. пособие. – Екатеринбург : Издательство Уральского государственного экономического университета, 2007. – 137 с.

Novikova N. V., Pozdeeva O. G. Prognozirovaniye natsionalnoy ekonomiki : ucheb.-metod. posobie. – Ekaterinburg : Izdatelstvo Uralskogo gosudarstvennogo ekonomicheskogo universiteta, 2007. – 137 s.

15. **Кобулов Б. А., Паластина И. П.** Прогнозирование принятия управленческих решений : учеб.-практ. пособие. – Москва : МГТА, 2004. – 76 с.

Kobulov B. A., Palastina I. P. Prognozirovaniye prinyatiya upravlencheskih resheniy : ucheb.-prakt. posobie. – Moskva : MGTA, 2004. – 76 s.

Irina Mosicheva, Cand. Sc. (Technology), Associate Professor, National Research University «Moscow Power Engineering Institute»;

mosicheva1@mail.ru

14, Krasnokazarmennaya st., 111250 Moscow, Russia

Svetlana Parfenova, Cand. Sc. (Economics), First Deputy Director, Head of the Department for Sci-tech Policy Problems and Science Development, Russian Research Institute for Economy, Policy and Law in Sci-tech Sphere (Moscow);

parfyonova.s.l@yandex.ru

50A, bldg. 6, Zemlyanoy Val st., 105064 Moscow, Russia

Vladislava Dolgova, Cand. Sc. (Economics), Associate Professor, Head,

Sector for Socio-economic Problems of Sci-tech Sphere Development, Russian Institute for Economics, Policy and Law in Science and Technology;

vlada8@bk.ru

50A, bldg. 6, Zemlyanoy Val st., 105064 Moscow, Russia

Karina Bezrodnova, *Research Engineer, Russian Institute for Economics, Policy and Law in Science and Technology;*

.karina.@mail.ru

50A, bldg. 6, Zemlyanoy Val st., 105064 Moscow, Russia

Elena Lyagushkina, *Researcher, Russian Institute for Economics, Policy and Law in Science and Technology;*

elen.boon@gmail.com

50A, bldg. 6, Zemlyanoy Val st., 105064 Moscow, Russia

Victor Bogatov, *Researcher, Russian Institute for Economics, Policy and Law in Science and Technology;*

d-benz.92@mail.ru

50A, bldg. 6, Zemlyanoy Val st., 105064 Moscow, Russia

Nadezhda Khaltakshinova, *Researcher, Russian Research Institute for Economy, Policy and Law in Sci-tech Sphere;*

nadya0195@gmail.com

50A, bldg. 6, Zemlyanoy Val st., 105064 Moscow, Russia

Vladimir Korobotov, *Researcher, Russian Institute for Economics, Policy and Law in Science and Technology;*

vkorobotov@mail.ru

50A, bldg. 6, Zemlyanoy Val st., 105064 Moscow, Russia

Irina Mikhailenko, *Head of Prospective Research and Special Projects Department, Russian National Public Library for Science and Technology,*

irimikhaylenko@gmail.com,

17, 3rd Khoroshevskaya st., 123436 Moscow, Russia