

МЕЖСТАТЕЙНЫЕ МЕТРИКИ КАЧЕСТВА ДЛЯ АВТОМАТИЗАЦИИ АНАЛИЗА НАУЧНЫХ ПУБЛИКАЦИЙ

Установлено, что недостатком известных метрик качества на уровне статей является то, что они характеризуют свойства отдельных статей без учета их связи с другими публикациями, недостатком метрик на уровне авторов – то, что они не учитывают взаимосвязь публикаций разных авторов, а метрик на уровне журналов – то, что они неприменимы для отдельных статей и для отдельных ученых. Впервые предложен комплекс межстатейных метрик научных публикаций, позволяющих количественно оценивать свойства научных публикаций, содержащий метрики, учитывающие взаимосвязь статей и их лексику, метрики, учитывающие цитируемость и лексику статей, а также метрики на уровне авторов и метрики на уровне журналов, позволяющие устранить недостатки известных метрик. Использование предложенных метрик позволяет автоматизировать анализ и сравнение научных публикаций и периодических изданий.

Ключевые слова: автоматизация, наукометрия, библиометрия, качество, статья, научная публикация, метрика, анализ цитируемости, важность статьи.

НОМЕНКЛАТУРА

$annot_p$ – p -е слово из словаря аннотации i -й публикации;
 $aut(paper_j)$ – функция, возвращающая список кодов (номеров) авторов j -й публикации в базе публикаций $Base$;

$Base$ – набор научных публикаций (статей, тезисов, диссертаций, авторефератов и т.п.);

$constr$ – ограничения на журналы и временные периоды их издания;

d_j^{rev} – дата получения редакцией доработанной версии j -й публикации от автора;

d_j^{rec} – дата получения редакцией исходной версии j -й публикации от автора;

d_j^{pub} – дата публикации (выхода журнала) доработанной версии j -й публикации;

el_{ji} – формальное обозначение i -го элемента j -й работы;

f_{ji} – частота j -го слова из словаря V_i в тексте i -й публикации;

$h_{k \rightarrow i}$ – коэффициент, характеризующий наименьший уровень иерархии ссылок по которым от i -й публикации можно перейти к k -й публикации (т. е. i -я публикация содержит ссылку на k -ю публикацию непосредственно или через ссылки на другие публикации в базе $Base$);

$I_{i,j}^*$ – условное обозначение показателя влияния i -й публикации на j -ю публикацию;

is_k^b – функция, возвращающая значение “1”, если k -й журнал входит в b -ю базу, и “0” – в противном случае;

I_j^q – q -й показатель качества j -й публикации;

\hat{I}_p^t – максимальный показатель качества номеров (выпусков) журнала за период t ;

\check{I}^{qt} – минимальный показатель качества номеров (выпусков) журнала за период t ;

jrn_k – множество индексов (номеров) статей, опубликованных в k -м журнале за период времени t ;

lat_i – широта места (города) издания (или написания) i -й публикации;

lon_i – долгота места (города) издания (или написания) i -й публикации;

N – число научных работ в базе;

N' – число публикаций, удовлетворяющих ограничениям $constr$;

N_{aut} – общее число авторов в базе;

N_{elj} – число элементов j -й работы;

N_i – объем словаря i -й публикации – общее число слов (включая повторы) в тексте i -й публикации;

N_{ij} – число раз, которое j -е слово из словаря V_i встречается в тексте i -й публикации в любой форме;

N_{iss} – число статей, опубликованных в соответствующем номере журнала;

N_{jrn} – общее число журналов в базе;

N_{*k}^{pu} – общее число публикаций из k -го журнала за все годы;

$N_{jrn_k}^t$ – общее число статей, опубликованных в k -м журнале за период времени t ;

N_{annot_i} – размер словаря аннотации i -й публикации;

N_{text_i} – размер словаря аннотации i -й публикации;

N_b^{pu} – число публикаций в b -й наукометрической базе;

$N_{b,k}^{pu}$ – число публикаций из k -го журнала в b -й наукометрической базе;

N_b^{jr} – число наименований сериальных и продолжающихся изданий в b -й базе;

N_{ref_j} – число ссылок в перечне ссылок j -й публикации;

$paper_j$ – j -я научная работа;

Q – набора показателей качества, характеризующих свойства публикации $paper_j$;

$q(p)$ – номер публикации в базе $Base$ для p -й ссылки в перечне источников i -й публикации (т.е. i -я публикация содержит ссылку на k -ю публикацию);

$ref(j)$ – ссылка на j -ю публикацию;

R_j – набор ссылок в перечне источников j -й публикации;
 $syn(v_i)$ – функция, возвращающая список слов-синонимов для слова v_i , а также их переводы (с их синонимами) на все доступные языки публикаций в базе $Base_2$;
 t – период времени;
 $text_j$ – словарь текста j -й публикации;
 $time$ – функция, возвращающая год, том, номер для публикации-аргумента;
 t_j – временная характеристика выхода j -й публикации;
 V_i – словарь i -й публикации;
 ρ – штрафной коэффициент.

ВВЕДЕНИЕ

Для автоматизации анализа и сравнения научных публикаций необходимо использовать метрики качества – показатели, позволяющие количественно оценивать свойства публикаций.

Программная реализация метрик качества в интеграции с базой научных публикаций могут обеспечить автоматизацию анализа отдельных публикаций, их групп, периодических изданий, содержащих публикации, а также влияния авторов публикаций на другие работы.

Известные метрики качества научных публикаций [1–18], как правило, характеризуют свойства отдельных статей без учета их связи с другими публикациями и дают лишь внешнюю оценку использования материала публикации, не учитывают взаимосвязь публикаций разных авторов и неприменимы для отдельных статей и для отдельных ученых.

Целью данной статьи являлась разработка комплекса показателей, позволяющих устранить недостатки известных метрик на уровне статей, авторов и журналов.

1 ПОСТАНОВКА ЗАДАЧИ

Пусть мы имеем набор научных публикаций (статей, тезисов, диссертаций, авторефератов и т.п.) $Base = \{paper_j\}, j = 1, 2, \dots, N, paper_j = \{el_{ji}\}, i = 1, 2, \dots, N_{elj}$.

Тогда задача оценивания качества научной публикации $paper_j$ по базе $Base$ заключается в определении набора показателей $Q = \{I_j^q\}$, характеризующих свойства публикации $paper_j$.

Показатели качества публикации разделим на абсолютные – те, которые определены в исходных единицах измерения, а также относительные – те, которые соотносены со значениями соответствующих показателей других статей.

2 ОБЗОР ЛИТЕРАТУРЫ

Известные показатели [1–17], характеризующие научные публикации, можно классифицировать как метрики на уровне статей [1] (характеризуют свойства отдельных статей без учета их связи с другими публикациями и дают лишь внешнюю оценку использования материала статьи), метрики на уровне журналов [3, 8, 15–17] (характеризуют статьи, опубликованные за определенный период в одном журнале) и метрики на уровне авторов [9–17] (характеризуют влияние авторов на научные публикации, имеющиеся в базе).

Недостатки известных метрик на уровне статей [1] частично устраняются в предложенном автором комплексе индивидуальных метрик качества, дающих внут-

ренние оценки представления статьи, учитывающие их лексику и библиографию, соответствие их структуры международным стандартам [18]. Нерешенной задачей является разработка показателей, характеризующих свойства отдельных публикаций с учетом их связи с другими публикациями по библиографии, цитируемости, времени, пространству, лексике, а также создание групповых показателей качества для наборов публикаций.

Недостатком метрик на уровне авторов [9–17] является то, что они не учитывают взаимосвязь публикаций разных авторов, а метрик на уровне журналов [3, 8, 15–17] – то, что они неприменимы для отдельных статей и для отдельных ученых.

Поэтому актуальной является задача устранения данных недостатков известных групп метрик.

3 МАТЕРИАЛЫ И МЕТОДЫ

На основе набора публикаций в базе $Base$ определим межстатейные показатели, отражающие взаимосвязь публикаций и их влияние друг на друга.

Показатель цитируемости j -й публикации в базе $Base$ за период времени t определим по формуле (1):

$$I_{cit_j} = \frac{\sum_{k=1}^N \{1 | t_j \in t, ref(j) \in R_k\}}{\max_{p=1,2,\dots,N} \left\{ \sum_{k=1}^N \{1 | ref(p) \in R_k\} \right\}}. \quad (1)$$

Показатель (1) будет принимать значения в диапазоне [0, 1]: его значение будет тем больше, чем более цитируемой в базе является публикация по сравнению с другими публикациями в той же базе.

Показатель прямого влияния i -й публикации на j -ю публикацию определим по формуле (2):

$$I_{infl}(i, j) = \begin{cases} 0, & ref(i) \notin R_j; \\ 1, & ref(i) \in R_j. \end{cases} \quad (2)$$

Показатель (2) будет принимать значение “1” если в перечне ссылок j -й публикации имеется ссылка на i -ю публикацию (если явно указана взаимосвязь публикаций), и “0” – в противном случае.

Определить рекурсивную функцию $hierarchy$ с начальным значением уровня иерархии $h_0=0$ для получения показателя $h_{k \rightarrow i}$.

Функция $h_{k \rightarrow i} = hierarchy(k, i, h_0)$: если $\exists p, p=1, \dots, |R_i|$: $ref(q(p))=ref(k)$, то вернуть: $h_{k \rightarrow i} = h_0 + 1$, в противном случае: если $\neg \exists p, p=1, \dots, |R_i|$: $paper_{q(p)} \in Base$ (т.е. в базе $Base$ отсутствуют публикации, на которые ссылается i -я публикация), тогда вернуть: $h_{k \rightarrow i} = 0$, в противном случае для каждого p -го элемента из набора ссылок i -й публикации, $p=1, \dots, |R_i|$ определить кортеж $h_{k \rightarrow q} = hierarchy(k, q(p), h_0+1)$, после чего определить: $z = \arg \min_{p=1,2,\dots,|R_i|} \{h_{k \rightarrow q(p)}\}$, а затем вернуть $h_{k \rightarrow i} = h_{k \rightarrow q(z)}$.

На основе введенных коэффициентов определим показатель косвенного влияния i -й публикации на j -ю пуб-

ликацию по формуле (3):

$$I_{indifl}(i, j) = \begin{cases} 1, ref(i) \in R_j; \\ 0, h_{i \rightarrow j} = 0; \\ \frac{1}{h_{i \rightarrow j}}, ref(i) \notin R_j, h_{i \rightarrow j} > 0. \end{cases} \quad (3)$$

Данный показатель будет принимать значения в диапазоне [0, 1]: значение показателя (3) будет тем больше, чем короче и прямее связь i -й и j -й публикаций в базе *Base*.

На практике многие публикации основываются на результатах предшествующих им работ. При этом авторы чаще склонны использовать публикации последних лет, содержащие более свежие результаты. Поэтому на конкретную публикацию большее воздействие имеют шанс оказывать непосредственно предшествующие ей по времени публикации.

Временной потенциал прямого возможного влияния i -й публикации на j -ю публикацию определим по формуле (4):

$$\Phi_{i,j}^t = \begin{cases} 0, t_i \geq t_j; \\ \frac{1}{t_j - t_i}, t_i < t_j. \end{cases} \quad (4)$$

Данный показатель будет принимать значения в диапазоне [0, 1]: значение показателя (4) будет тем больше, чем более близкими по времени издания являются i -я и j -я публикации, при условии, что i -я публикация опубликована раньше j -й публикации.

Объединив показатель прямого влияния i -й публикации на j -ю публикацию (2) и временной потенциал прямого возможного влияния i -й публикации на j -ю публикацию (4), получим временной потенциал прямого влияния i -й публикации на j -ю публикацию, определяемый по формуле (5):

$$I_{i,j}^{t,inf} = I_{infl}(i, j) \Phi_{i,j}^t. \quad (5)$$

Данный показатель будет принимать значения в диапазоне [0, 1]: его значение будет тем больше, чем более близкими по времени издания являются i -я и j -я публикации, при условии, что i -я публикация опубликована раньше j -й публикации и в перечне ссылок j -й публикации имеется ссылка на i -ю публикацию (явно указана взаимосвязь публикаций).

Временной потенциал косвенного влияния i -й публикации на j -ю публикацию оценим как (6):

$$I_{i,j}^{t,indifl} = I_{indifl}(i, j) \Phi_{i,j}^t. \quad (6)$$

Показатель (6) будет принимать значения в диапазоне [0, 1]: его значение будет тем больше, чем более близкими по времени издания являются i -я и j -я публикации, при условии, что i -я публикация опубликована раньше j -й публикации и по ссылкам от j -й публикации можно перейти по базе *Base* к i -й публикации (неявная связь пуб-

ликаций). Причем чем меньше будет длина пути от j -й публикации к i -й, тем больше будет значение показателя.

Наряду со временем издания на публикации может оказывать влияние место их издания или написания. Очевидно, что на авторы некоторой публикации могут быть лучше знакомы с публикациями своих коллег, работающих в том же городе или стране, либо же публикующих публикации в одном с ними журнале.

Пространственный потенциал прямого возможного влияния i -й публикации на j -ю публикацию определим по формуле (7):

$$\Phi_{i,j}^g = e^{-6,372 \arccos(\sin(lat_i) \sin(lat_j) + \cos(lat_i) \cos(lat_j) \cos(lon_i - lon_j))} \quad (7)$$

Заметим, что значения северной широты и восточной долготы считают положительными, а значения южной широты и западной долготы – отрицательными. Для определения координат мест издания можно руководствоваться данными [19].

Данный показатель будет принимать значения в диапазоне [0, 1]: значение показателя (7) будет тем больше, чем более близкими по расстоянию между местами издания (написания) являются i -я и j -я публикации.

Объединив показатель прямого влияния i -й публикации на j -ю публикацию (2) и пространственный потенциал прямого возможного влияния i -й публикации на j -ю публикацию (7), получим пространственный потенциал прямого влияния i -й публикации на j -ю публикацию, определяемый по формуле (8):

$$I_{i,j}^{g,inf} = I_{infl}(i, j) \Phi_{i,j}^g. \quad (8)$$

Показатель (8) будет принимать значения в диапазоне [0, 1]: его значение будет тем больше, чем более близкими по месту издания являются i -я и j -я публикации, при условии, что в перечне ссылок j -й публикации имеется ссылка на i -ю публикацию (явно указана взаимосвязь публикаций).

Пространственный потенциал косвенного влияния i -й публикации на j -ю публикацию зададим как (9):

$$I_{i,j}^{g,indifl} = I_{indifl}(i, j) \Phi_{i,j}^g. \quad (9)$$

Данный показатель будет принимать значения в диапазоне [0, 1]: значение показателя (9) будет тем больше, чем более близкими по расстоянию между местами издания (написания) являются i -я и j -я публикации, при условии, что по ссылкам от j -й публикации можно перейти по базе *Base* к i -й публикации (неявная связь публикаций). Причем чем меньше будет длина пути от j -й публикации к i -й, тем больше будет значение показателя.

Пространственно-временной потенциал возможного влияния i -й публикации на j -ю публикацию определим по формуле (10):

$$\Phi_{i,j}^{tg} = \Phi_{i,j}^t \Phi_{i,j}^g. \quad (10)$$

Показатель (10) будет принимать значения в диапазоне [0, 1]: его значение будет тем больше, чем более близкими по времени издания являются i -я и j -я публикации, при условии, что i -я публикация опубликована раньше

j -й публікації, і чим більше близькими по відстані між місцями видання (написання) є i -я і j -я публікації.

Просторово-часовий потенціал прямого впливу i -ї публікації на j -ю публікацію визначимо за формулою (11):

$$I_{i,j}^{tg,inf} = I_{inf}(i,j) \varphi_{i,j}^{tg}. \quad (11)$$

Показатель (11) буде приймати значення в діапазоні $[0, 1]$: його значення буде тим більше, чим більше близькими по часу видання є i -я і j -я публікації, при умові, що i -я публікація опублікована раніше j -ї публікації, і чим більше близькими по відстані між місцями видання (написання) є i -я і j -я публікації, при умові, що i -я публікація опублікована раніше j -ї публікації і в переліку посилань j -ї публікації існує посилання на i -ю публікацію (явно вказана взаємозв'язок публікацій).

Просторово-часовий потенціал косвенного впливу i -ї публікації на j -ю публікацію визначимо за формулою (12):

$$I_{i,j}^{tg,indifl} = I_{indifl}(i,j) \varphi_{i,j}^{tg}. \quad (12)$$

Даний показник буде приймати значення в діапазоні $[0, 1]$: значення показника (12) буде тим більше, чим більше близькими по часу видання, а також по відстані між місцями видання (написання) є i -я і j -я публікації, при умові, що по посиланню від j -ї публікації можна перейти по базі *Base* до i -ї публікації (неявна зв'язок публікацій). Причому чим менше буде довжина шляху від j -ї публікації до i -ї, тим більше буде значення показника (12).

Аналіз взаємозв'язку публікацій і їх подібності за тематикою, очевидно, передбачає врахування їх лексики. При цьому слід врахувати варіативність лексики (синоніми), а також можливість написання публікацій на різних мовах.

За словником тексту ми будемо розуміти набір слів, що містяться в тексті, представлених в нормалізованій формі (наприклад, іменнік і прикметник в однині іменного роду, дієслово – в неопределеної формі і однині іменного роду і т.п.) без повторень.

Показник точного співпадіння словників i -ї і k -ї публікацій визначимо за формулою (13):

$$I_{eq}(V_i, V_k) = \frac{\sum_{j=1}^{|V_i|} \sum_{p=1}^{|V_k|} \{1 | v_{ij} = v_{kp}\}}{|V_i| |V_k|}. \quad (13)$$

Даний показник буде приймати значення в діапазоні $[0, 1]$: чим більше буде значення даного показника, тим сильніше співпадають словники, чим менше – тим слабше.

Показник приблизного співпадіння словників i -ї і k -ї публікацій визначимо як (14):

$$I_{sim}(V_i, V_k) = \frac{\sum_{j=1}^{|V_i|} \sum_{p=1}^{|V_k|} \{1 | v_{ij} \in \text{syn}(v_{kp})\}}{|V_i| |V_k|}. \quad (14)$$

Даний показник буде приймати значення в діапазоні $[0, 1]$: чим більше буде значення показника (14), тим сильніше співпадають словники з врахуванням синонімів і перекладів слів, чим менше – тим слабше.

Показник різниці частот (частотного точного неспівпадіння) лексики i -ї і k -ї публікацій визначимо за формулою (15):

$$I_{dfr}(V_i, V_k) = \frac{\sum_{j=1}^{|V_i|} \sum_{p=1}^{|V_k|} \{|f_{r_{ij}} - f_{r_{kp}}| | v_{ij} = v_{kp}\}}{\max\{|V_i|, |V_k|\}}. \quad (15)$$

Даний показник буде приймати значення в діапазоні $[0, 1]$: чим більше буде значення показника (15), тим сильніше не співпадають частоти однакових слів в словниках, чим менше – тим слабше.

Відповідно, визначимо показник частотного точного співпадіння лексики i -ї і k -ї публікацій, написаних на одній мові, як (16):

$$I_{fr}(V_i, V_k) = 1 - I_{dfr}(V_i, V_k). \quad (16)$$

Даний показник буде приймати значення в діапазоні $[0, 1]$: чим більше буде значення показника (16), тим слабше не співпадають частоти однакових слів в словниках, чим менше – тим сильніше.

Показник приблизного частотного неспівпадіння лексики i -ї і k -ї публікацій з врахуванням синонімів і різних мов визначимо за формулою (17):

$$I_{dsfr}(V_i, V_k) = \frac{\sum_{j=1}^{|V_i|} \sum_{p=1}^{|V_k|} \{|f_{r_{ij}} - f_{r_{kp}}| | v_{ij} \in \text{syn}(v_{kp})\}}{\max\{|V_i|, |V_k|\}}. \quad (17)$$

Даний показник буде приймати значення в діапазоні $[0, 1]$: чим більше буде значення показника (17), тим сильніше не співпадають частоти семантично еквівалентних слів (з врахуванням синонімів і перекладів) в словниках, чим менше – тим слабше.

Відповідно, визначимо показник абсолютного частотного співпадіння лексики i -ї і k -ї публікацій з врахуванням синонімів і перекладів за формулою (18):

$$I_{sfr}(V_i, V_k) = 1 - I_{dsfr}(V_i, V_k). \quad (18)$$

Даний показник буде приймати значення в діапазоні $[0, 1]$: чим більше буде значення показника (18), тим слабше не співпадають частоти семантично еквівалентних слів (з врахуванням синонімів і перекладів) в словниках, чим менше – тим сильніше.

Для більш комплексного аналізу взаємозв'язку публікацій і їх подібності цілорозумно врахувати їх взаємозв'язок через бібліографічні посилання, а також подібність їх лексики.

Обобщений показник відповідності лексики анотацій літературних джерел лексиці тексту цитуємої їх публікації визначимо за формулою (19):

$$I_{artj} = \frac{1}{N_{refj}} \sum_{i=1}^{N_{refj}} \left(\frac{1}{N_{annoti}} \sum_{p=1}^{N_{annoti}} \{1 | \text{syn}(\text{annot}_{q(i)p}) \cap \text{text}_j \neq \emptyset\} \right). \quad (19)$$

Даний показник буде приймати значення в діапазоні $[0, 1]$: чим більше буде значення показника (19),

тем сильнее совпадает лексика текста j -й публикации и лексика аннотаций цитируемых ею публикаций.

Обобщенный показатель соответствия лексики текста литературных источников лексике текста цитирующей их публикации определим по формуле (20):

$$I_{irtj} = \frac{1}{N_{ref}} \sum_{i=1}^{N_{refj}} \left(\frac{1}{N_{texti}} \sum_{p=1}^{N_{texti}} \{1 | \text{syn}(\text{text}_{q(i)p}) \cap \text{text}_j \neq \emptyset\} \right). \quad (20)$$

Данный показатель будет принимать значения в диапазоне $[0, 1]$: чем больше будет значение показателя (20), тем сильнее совпадает лексика текста j -й публикации и лексика аннотаций цитируемых ею публикаций.

Рассмотрим метрики на уровне авторов.

Число публикаций i -го автора в базе публикаций за период времени t определим как (21):

$$N_{Ai}^t = \sum_{j=1}^N \{1 | i \in \text{aut}(\text{paper}_j), t_j \in t\}. \quad (21)$$

Число ссылок на публикации i -го автора в базе публикаций за период времени t определим по формуле (22):

$$N_{Ai}^{cit,t} = \sum_{j=1}^N \sum_{k=1}^N \{1 | \text{ref}(k) \in R_j, i \in \text{aut}(\text{paper}_k), t_j \in t\}. \quad (22)$$

Число цитирований i -м автором собственных публикаций за период времени t в базе публикаций определим по формуле (23):

$$N_{Ai}^{sc,t} = \sum_{j=1}^N \sum_{k=1}^N \{1 | \text{ref}(k) \in R_j, i \in \text{aut}(\text{paper}_k), i \in \text{aut}(\text{paper}_j), t_j \in t\}. \quad (23)$$

На основе введенных показателей определим относительный показатель цитируемости публикаций i -го автора в базе публикаций за период времени t по формуле (24):

$$I_{citAi}^{rel,t} = \frac{N_{Ai}^{cit,t} - N_{Ai}^{sc,t}}{\max_{j=1,2,\dots,N_{aut}} \{N_{Aj}^{cit,t} - N_{Aj}^{sc,t}\}}. \quad (24)$$

Показатель (24) будет принимать значения в диапазоне $[0, 1]$: его значение будет тем больше, чем большим автором цитат в базе.

Показатель прямого влияния публикаций i -го автора на публикации j -го автора за период времени t оценим по формуле (25):

$$I_{Ai \rightarrow Aj}^t = \frac{1}{\max\{N_{Ai}^t, N_{Aj}^t\}^2} \sum_{k=1}^N \sum_{p=1}^N \{1 | i \in \text{aut}(\text{paper}_k), j \in \text{aut}(\text{paper}_p), t_k \in t, t_p \in t, k \neq p, \text{ref}(k) \in R_p\}. \quad (25)$$

Показатель (25) будет принимать значения в диапазоне $[0, 1]$: его значение будет тем больше, чем больше публикаций j -го автора ссылается на большее число публикаций i -го автора.

Показатель косвенного влияния публикаций i -го автора на публикации j -го автора за период времени t оценим по формуле (26):

$$I_{Ai \rightarrow Aj}^{indifl,t} = \frac{1}{N^2} \sum_{k=1}^N \sum_{p=1}^N \{I_{indifl}(k,p) | i \in \text{aut}(\text{paper}_k), j \in \text{aut}(\text{paper}_p), t_k \in t, t_p \in t, k \neq p\}. \quad (26)$$

Показатель (26) будет принимать значения в диапазоне $[0, 1]$: его значение будет тем больше, чем больше публикаций j -го автора ссылается на большее число публикаций i -го автора, либо чем к большему числу публикаций i -го автора можно перейти от публикаций j -го автора по цитируемым публикациям в базе.

Рассмотрим метрики на уровне журналов.

На основе каждого абсолютного показателя качества публикаций, опубликованных за определенный период (определяется годом, томом, номером / выпуском) в журнале, определим, соответственно:

– относительный показатель качества номеров (выпусков) журнала за период t (27):

$$I_j^q = \frac{I_j^q}{\bar{I}^q}, \quad (27)$$

где средний показатель качества номеров (выпусков) журнала за период t оценим из (28):

$$\bar{I}^q = \frac{1}{N_{iss}} \sum_{j=1}^{N_{iss}} \{(I_j^q) | \text{time}(I_j^q) \in t\}; \quad (28)$$

– нормированный показатель качества номеров (выпусков) журнала за период t (29):

$$\tilde{I}^{qt} = \frac{I^{qt} - \bar{I}^{qt}}{\bar{I}^{qt} - \tilde{I}^{qt}}. \quad (29)$$

где \tilde{I}^{qt} – минимальный показатель качества номеров (выпусков) журнала за период t (30):

$$\tilde{I}^{qt} = \min_{j=1,2,\dots,N_{iss}} \{I_j^q | \text{time}(I_j^q) \in t\}, \quad (30)$$

\hat{I}_p^t – максимальный показатель качества номеров (выпусков) журнала за период t (31):

$$\hat{I}_p^t = \max_{j=1,2,\dots,N_{iss}} \{I_j^q | \text{time}(I_j^q) \in t\}. \quad (31)$$

Общий показатель индексируемости k -го журнала по набору баз B определим по формуле (32):

$$I_B(k) = \frac{1}{N_B} \sum_{b=1}^{N_B} \omega_b \Psi_k^b, \quad (32)$$

$$\omega_b = \begin{cases} 0, N_b^{jr} = 0, N_b^{pu} = 0; \\ \frac{N_b^{jr}}{\max_{p=1,\dots,N_B} \{N_p^{jr}\}}, N_b^{jr} > 0, N_b^{pu} = 0; \\ \frac{N_b^{pu}}{\max_{p=1,\dots,N_B} \{N_p^{pu}\}}, N_b^{pu} > 0, \end{cases}$$

$$\Psi_k^b = \begin{cases} \frac{N_{b,k}^{pu}}{N_{*k}^{pu}}, is_k^b = 1, N_{b,k}^{pu} > 0; \\ is_k^b, N_{b,k}^{pu} = 0. \end{cases}$$

Данный показатель будет принимать значения в диапазоне [0, 1]. Его значение будет тем больше, чем в большем числе баз представлен журнал и чем больше публикаций из журнала представлено в базах.

Показатель оперативности рецензирования и доработки публикаций в k -м журнале определим как (33):

$$I_o^{rew} = \frac{1}{N^r} \sum_{j=1}^N \left\{ \frac{1}{d_j^{rev} - d_j^{rec}} \mid j \in constr, d_j^{rev} > d_j^{rec}, paper_j \in jrn_k \right\}. \quad (33)$$

Данный показатель принимает значения в диапазоне [0, 1]: чем больше будет его значение, тем оперативнее рецензенты рассматривают публикации и быстрее авторы дорабатывают публикации в k -м журнале.

Показатель длительности рецензирования и доработки публикаций в k -м журнале определим из (34):

$$I_t^{rew} = 1 - I_o^{rew}. \quad (34)$$

Данный показатель принимает значения в диапазоне [0, 1]: чем меньше будет его значение, тем оперативнее рецензенты рассматривают публикации и быстрее авторы дорабатывают публикации в k -м журнале.

Показатель оперативности рассмотрения и обработки публикаций редакцией в k -м журнале оценим из (35):

$$I_o^{pub} = \frac{1}{N^r} \sum_{j=1}^N \left\{ \frac{1}{d_j^{pub} - d_j^{rec}} \mid j \in constr, d_j^{pub} > d_j^{rec}, paper_j \in jrn_k \right\}. \quad (35)$$

Показатель (35) будет принимать значения в диапазоне [0, 1]: чем больше будет его значение, тем оперативнее публикуются публикации в k -м журнале.

Показатель длительности рассмотрения и обработки публикаций редакцией определим как (36):

$$I_t^{pub} = 1 - I_o^{pub}. \quad (36)$$

Показатель (36) будет принимать значения в диапазоне [0, 1]: чем меньше будет его значение, тем оперативнее публикуются публикации в k -м журнале.

Обобщенно влияние i -й публикации на публикации в k -м журнале по некоторой категории показателей “***” за определенный период времени t оценим по формуле (37):

$$I_{i,(k)}^* = \frac{1}{N_{jrn_k}^t} \sum_{j=1}^N \{I_{i,j}^* \mid paper_j \in jrn_k, t_j \in t\}, \quad (37)$$

где $N_{jrn_k}^t$ определяется по формуле (38):

$$N_{jrn_k}^t = \sum_{j=1}^N \{1 \mid paper_j \in jrn_k, t_j \in t\}, \quad (38)$$

а в качестве $I_{i,j}^*$ могут выступать $I_{infl}(i, j)$, $I_{indifl}(i, j)$, $\Phi_{i,j}^t$, $I_{i,j}^{t,infl}$, $I_{i,j}^{t,indifl}$, $\Phi_{i,j}^g$, $I_{i,j}^{g,infl}$, $I_{i,j}^{g,indifl}$, $\Phi_{i,j}^{tg}$, $I_{i,j}^{tg,infl}$, $I_{i,j}^{tg,indifl}$.

Показатель (37) будет принимать значения в диапазоне [0, 1]: его значение будет тем больше, чем больше в среднем значение показателей влияния публикаций на публикации в k -м журнале по заданной категории показателей за определенное время.

Обобщенно влияние k -го журнала на p -й журнал по некоторой категории показателей “***” за определенный период времени t оценим по формуле (39):

$$I_{(k)(p)}^* = \frac{1}{N_{jrn_k}^t N_{jrn_p}^t} \sum_{i=1}^N \sum_{j=1}^N \{I_{i,j}^* \mid paper_i \in jrn_k, paper_j \in jrn_p, t_i \in t, t_j \in t\}. \quad (39)$$

Показатель (39) будет принимать значения в диапазоне [0, 1]: его значение будет тем больше, чем больше в среднем значение показателей влияния публикаций, опубликованных в k -м журнале, по заданной категории показателей за определенное время на публикации, опубликованные в p -м журнале.

Показатель значимости k -го журнала в базе научных публикаций *Base* определим по формуле (40):

$$I_{jrn_k} = \frac{N_{jrn_k}^t \left(\sum_{j=1}^N \{I_{cit_j} \mid t_j \in t, paper_j \in jrn_k\} - N_{scj_k} \right)}{\max_{p=1,2,\dots,N_{jrn}} \left\{ N_{jrn_p}^t \right\} \max_{p=1,2,\dots,N_{jrn}} \left\{ \sum_{j=1}^N \{I_{cit_j} \mid t_j \in t, paper_j \in jrn_p\} - N_{scj_p} \right\}}. \quad (40)$$

Показатель (40) будет принимать значения в диапазоне [0, 1]: его значение будет тем больше, чем больше общее число статей, опубликованных в k -м журнале за период времени t , чем больше в базе имеется цитирований статей из k -го журнала и чем меньше имеется самоцитирований статей в k -м журнале.

Влияние публикаций автора на журнал охарактеризуем с помощью следующих показателей.

Число статей i -го автора, опубликованных в k -м журнале за период времени t , определим по формуле (41):

$$N_{A_i}^{jrn_k} = \sum_{j=1}^N \{1 \mid i \in aut(paper_j), t_j \in t, paper_j \in jrn_k\}. \quad (41)$$

Число цитирований статей i -го автора в k -м журнале за период времени t определим по формуле (42):

$$N_{A_i}^{cj_k} = \sum_{j=1}^N \sum_{q=1}^N \{1 \mid i \in aut(paper_q), ref(q) \in R_j, t_j \in t, paper_j \in jrn_k\}. \quad (42)$$

Число самоцитирования статей i -м автором в k -м журнале за период времени t оценим по формуле (43):

$$N_{scj_k} = \sum_{j=1}^N \sum_{q=1}^N \{1 \mid ref(q) \in R_j, t_j \in t, paper_q \in jrn_k, paper_j \in jrn_k\}. \quad (43)$$

Число цитирований i -м автором в k -м журнале собственных статей из того же журнала за период времени t определим по формуле (44):

$$N_{A_i}^{sc_k} = \sum_{j=1}^N \sum_{q=1}^N \{1 \mid \text{ref}(q) \in R_j, i \in \text{aut}(\text{paper}_q), i \in \text{aut}(\text{paper}_j), t_j \in t, \text{paper}_q \in \text{jrn}_k, \text{paper}_j \in \text{jrn}_k\}. \quad (44)$$

Число самоцитирований i -го автора в k -м журнале за период времени t определим по формуле (45):

$$N_{A_i}^{scjk} = \sum_{j=1}^N \sum_{q=1}^N \{1 \mid \text{ref}(q) \in R_j, i \in \text{aut}(\text{paper}_q), i \in \text{aut}(\text{paper}_j), t_j \in t, \text{paper}_j \in \text{jrn}_k\}. \quad (45)$$

Показатель влияния (ценности) i -го автора на k -й журнал за период времени t оценим по формуле (46):

$$I_{A_i}^{jrn_k} = \left(1 - e^{-N_{A_i}^{jrn_k}}\right) \left(1 - e^{-\left(N_{A_i}^{cj_k} - \rho N_{A_i}^{scjk}\right)}\right), \quad (46)$$

где $0 < \rho \leq 1$. Значение ρ следует выбирать, исходя из того, что авторы с большой долей самоцитирования менее ценны для журнала. Например, можно рекомендовать принять $\rho = 0,5$.

Показатель (46) будет принимать значения в диапазоне $[0, 1]$: его значение будет тем больше, чем больше число статей i -го автора, опубликованных в k -м журнале, чем больше цитирований статей i -го автора в k -м журнале, чем меньше самоцитирований i -го автора в k -м журнале.

Показатель значимости публикаций i -го автора в базе *Base* оценим по формуле (47):

$$I_{A_i} = \frac{I_{cit_{A_i}} \left(\sum_{j=1}^{N_{aut}} I_{A_i \rightarrow A_j} \right) \sum_{k=1}^{N_{jrn}} I_{A_i}^{jrn_k}}{\max_{j=1,2,\dots,N_{aut}} \left\{ \sum_{k=1}^{N_{aut}} I_{A_i \rightarrow A_j} \right\} \max_{j=1,2,\dots,N_{aut}} \left\{ \sum_{k=1}^{N_{jrn}} I_{A_j}^{jrn_k} \right\}}. \quad (47)$$

Показатель (47) будет принимать значения в диапазоне $[0, 1]$: его значение будет тем больше, чем цитирований статей i -го автора в базе, чем больше влияние i -го автора на публикации и журналы в базе.

4 ЭКСПЕРИМЕНТЫ

Рассмотрим пример расчета чести предложенных показателей для искусственной базы, характеристики статей которых приведены в табл. 1–2. В табл. 1–2 коды цитируемых статей приведены относительно содержащихся в базе статей (остальные цитируемые источники опущены). Примем период времени $t = 1990 - 1995$.

Таблица 1 – Характеристики работ из экспериментальной базы публикаций

Код публикации	Коды авторов	Код журнала	Коды цитируемых статей	d_j^{rev}	d_j^{rec}	d_j^{pub}
P1	A1	J1	–	10.02.1990	19.10.1989	12.04.1990
P2	A2	J2	P1	27.11.1994	03.07.1994	28.01.1995
P3	A3	J1	P1, P2	12.03.1992	14.10.1991	23.05.1992
P4	A2, A4	J3	P3	06.12.2011	23.05.2011	16.02.2012

Таблица 2 – Характеристики журналов экспериментальной базы публикаций

Код журнала	Место издания	
	Широта	Долгота
J1	55°45'07" с.ш.	37°36'56" в.д.
J2	50°27'16" с.ш.	30°31'25" в.д.
J3	47°49'22" с.ш.	35°11'25" в.д.

5 РЕЗУЛЬТАТЫ

Результаты расчета характеристик базы публикаций приведены в табл. 3, метрик качества уровня статей – в табл. 4, метрик уровня авторов – в табл. 5, уровня журналов – в табл. 6.

6 ОБСУЖДЕНИЕ

Результаты проведенных расчетов подтверждают работоспособность и практическую применимость предложенного комплекса метрик качества статей.

В отличие от известных известных метрик на уровне статей [1], предложенные метрики позволяют характеризовать свойства отдельных публикаций с учётом их связи с другими публикациям по библиографии, цитируемости, времени, пространству и лексике.

Таблица 3 – Общие характеристики экспериментальной базы публикаций

Метрика	Значение
N	4
N_{aut}	4
N_{jrn}	3
$N_{jrn_k}^t$	3

Таблица 4 – Метрики уровня статей

Метрика	Код публикации в базе			
	P1	P2	P3	P4
I_{cit_j}	1	0,5	0,5	0
$h_{k \rightarrow i} (i=1)$	0	1	1	2
$I_{infl}(i, j) (i=1)$	0	1	1	0
$I_{indifl}(i, j) (i=1)$	0	1	1	0,5
$\phi_{i,j}^t (i=1)$	0	0,000579374	0,00131406	0,000127162
$I_{i,j}^{t,infl} (i=1)$	0	0,000579374	0,00131406	0
$I_{i,j}^{t,indifl} (i=1)$	0	0,000579374	0,00131406	0,000063581
$\phi_{i,j}^g (i=1)$	1	0,469667632	1	0,408003307
$I_{i,j}^{g,infl} (i=1)$	0	0,469667632	1	0
$I_{i,j}^{g,indifl} (i=1)$	0	0,469667632	1	0,204002
$\phi_{i,j}^{tg} (i=1)$	0	0,00027211	0,00131406	0,00005188
$I_{i,j}^{tg,infl} (i=1)$	0	0,00027211	0,00131406	0
$I_{i,j}^{tg,indifl} (i=1)$	0	0,00027211	0,00131406	0,102001

Таблица 5 – Метрики уровня авторов

Метрика	Код автора в базе			
	A1	A2	A3	A4
$N_{A_i}^t$	1	1	1	0
$N_{A_i}^{cit,t}$	2	1	0	0
$N_{A_i}^{sc,t}$	1	0	0	0
$I_{A_i}^{rel,t}$	1	1	0	0
$I_{A_i \rightarrow A_j}^t (i=1)$	0	1	1	0

Таблица 6 – Метрики уровня журналов

Метрика	Код журнала в базе		
	J1	J2	J3
I_O^{rew}	0,007883	0,006944	0,005181
I_t^{rew}	0,992117	0,993056	0,994819
I_O^{pub}	0,005173	0,004878	0,003802
I_t^{pub}	0,994827	0,995122	0,996198
$N_{jrn_k}^t$	2	1	0
I_{jrn_k}	0,75	0,125	0
$N_{A_i}^{jrn_k} (i=1)$	1	0	0
$N_{A_i}^{cj_k} (i=1)$	0	1	0
$N_{scj_k} (i=1)$	0	0	0
$N_{A_i}^{sc_k} (i=1)$	0	0	0
$N_{A_i}^{scj_k} (i=1)$	0	0	0
$I_{A_i}^{jrn_k} (i=1)$	0	0	0
$I_{A_i} (i=1)$	0	0	0

В отличие от известных метрик на уровне авторов [9–17], предложенные метрики позволяют учитывать взаимосвязь публикаций разных авторов, а в отличие от известных метрик на уровне журналов [3, 8, 15–17], предложенные метрики могут применяться для отдельных статей и для отдельных ученых.

ЗАКЛЮЧЕНИЕ

В работе решена задача разработки показателей для автоматизации анализа научных публикаций.

Впервые предложен комплекс межстатейных метрик научных публикаций, позволяющих количественно оценивать свойства научных публикаций, содержащий метрики, учитывающие взаимосвязь статей и их лексику, метрики, учитывающие цитируемость и лексику статей, а также метрики на уровне авторов и метрики на уровне журналов, позволяющие устранить недостатки известных метрик.

Дальнейшие исследования могут быть направлены на интеграцию предложенных показателей в математи-

ческое обеспечение наукометрических и реферативных баз научных публикаций, программное обеспечение, автоматизирующее процессы аттестации научных кадров и оценивание научных исследований, а также на экспериментальное исследование предложенных показателей в задачах анализа важности научных публикаций и изданий, оценивании трудов авторов, экспертизе научных и квалификационных работ.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке международного проекта «Centers of Excellence for young REsearchers» программы «Темпус» Европейской Комиссии (№ 544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES) в рамках госбюджетных научно-исследовательских тем Запорожского национального технического университета «Интеллектуальные информационные технологии диагностирования и автоматической классификации» и «Интеллектуальные методы диагностирования систем управления удаленными техническими объектами».

СПИСОК ЛІТЕРАТУРИ

1. Article-Level Metrics. A SPARC Primer. [Electronic resource]. – Access mode: <http://www.sparc.arl.org>
2. The PageRank Citation Ranking: Bringing Order to the Web [Electronic resource] / Page L., Brin S., Motwani R., Winograd T. – Stanford : Stanford InfoLab, 1999. – Access mode: <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>
3. Moed H. F. Measuring contextual citation impact of scientific journals / H. F. Moed // Journal of Informetrics. – 2010. – Vol. 4, № 3. – P. 265–277.
4. González-Pereira B. A new approach to the metric of journals' scientific prestige: The SJR indicator / B. González-Pereira, V. P. Guerrero-Bote, F. Moya-Anegón // Journal of informetrics. – 2010. – Vol. 4, № 3. – P. 379–391.
5. Guerrero-Bote V. P. A further step forward in measuring journals' scientific prestige: The SJR2 indicator / V. P. Guerrero-Bote, F. Moya-Anegón // Journal of Informetrics. – 2012. – Vol. 6, № 4. – P. 674–688.
6. Butler D. Free journal-ranking tool enters citation market / D. Butler // Nature. – 2008. – Vol. 451, № 6. – P. 6.
7. Falagas M. E. Comparison of SCImago journal rank indicator with journal impact factor / M. E. Falagas, V. D. Kouranos, R. Arencibia-Jorge, D. E. Karageorgopoulos // The FASEB Journal. – 2008. – Vol. 22, № 8. – P. 2623–2628.
8. Bergstrom C. T. The Eigenfactor Metrics / C. T. Bergstrom, J. D. West, M. A. Wiseman // Journal of Neuroscience. – 2008. – Vol. 28, № 45. – P. 11433–11434.
9. Hirsch J. E. An index to quantify an individual's scientific research output / J. E. Hirsch // Proceedings of The National Academy of Sciences. 2005. – Vol. 102, № 46. – P. 16569–16572.
10. Egghe L. Theory and practise of the g-index / L. Egghe // Scientometrics. – 2006. – Vol. 69, № 1. – P. 131–152.
11. Woeginger G. J. An axiomatic analysis of Egghe's g-index / G. J. Woeginger // Journal of Informetrics. – 2008. – Vol. 2, № 4. – P. 364–368.
12. Tol R. S. J. A rational, successive g-index applied to economics departments in Ireland / R. S. J. Tol // Journal of Informetrics. – 2008. – Vol. 2, № 2. – P. 149–155.
13. Kosmulski M. I – a bibliometric index / M. Kosmulski // Forum Akademickie. – 2006. – Vol. 11. – P. 31.
14. Prathap G. Hirsch-type indices for ranking institutions' scientific research output / G. Prathap // Current Science. – 2006. – Vol. 91, № 11. – P. 1439.
15. Игра в цифирь, или как теперь оценивают труд ученого (сборник статей о библиометрике). – М. : МЦНМО, 2011. – 72 с.

16. Цыганов А. В. Краткое описание наукометрических показателей, основанных на цитируемости / А. В. Цыганов // Управление большими системами : сб. тр. Спец. вып. 44 – Наукометрия и экспертиза в управлении наукой. – М.: ИПУ РАН, 2013. – С.248–261.
17. Штовба С. Д. Обзор наукометрических показателей для оценки публикационной деятельности ученого / С. Д. Штовба, Е. В. Штовба // Управление большими системами : сб. тр. Спец. вып. 44 – Наукометрия и экспертиза в управлении наукой. – М.: ИПУ РАН, 2013. – С. 262–278.
18. Субботин С. О. Модель та індивідуальні метрики якості наукових публікацій / С. О. Субботин // Радіоелектроніка, інформатика, управління. – 2015. – № 3. – С. 30–39.
19. Global Gazetteer [Electronic resource]. – Access mode: <http://www.fallingrain.com/world/>

Стаття надійшла до редакції 29.07.2015.

Після доробки 18.08.2015.

Субботин С. О.

Д-р техн. наук, професор, професор кафедри програмних засобів Запорізького національного технічного університету, Запоріжжя, Україна

МІЖСТАТЕЙНІ МЕТРИКИ ЯКОСТІ ДЛЯ АВТОМАТИЗАЦІЇ АНАЛІЗУ НАУКОВИХ ПУБЛІКАЦІЙ

Встановлено, що недоліком відомих метрик якості на рівні статей є те, що вони характеризують властивості окремих статей без урахування їхнього зв'язку з іншими публікаціям, недоліком метрик на рівні авторів – те, що вони не враховують взаємозв'язки публікацій різних авторів, а метрик на рівні журналів – те, що вони незастосовні для окремих статей і для окремих учених. Уперше запропонований комплекс міжстатейних метрик наукових публікацій, які дозволяють кількісно оцінювати властивості наукових публікацій, що містить метрики, які враховують взаємозв'язок статей і їхню лексику, метрики, що враховують цитованість і лексику статей, а також метрики на рівні авторів і метрики на рівні журналів, що дозволяють усунути недоліки відомих метрик. Використання запропонованих метрик дозволяє автоматизувати аналіз і порівняння наукових публікацій і періодичних видань.

Ключові слова: автоматизація, наукометрия, бібліометрія, якість, стаття, наукова публікація, метрика, аналіз цитованості, важливість статті.

Subbotin S. A.

Dr.Sc., Professor, Professor of department of software tools, Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine

INTER-PAPER QUALITY METRICS FOR SCIENTIFIC PUBLICATION ANALYSIS AUTOMATION

The disadvantages of known publication quality metrics are found: the metric level papers characterize the properties of individual articles without regard to their connection with other publications, the author level metrics did not take into account the relationship of publications of different authors, and journal level metrics are not applicable to individual articles and for individual scientists. The complex of scientific publications inter-paper metrics is proposed firstly. It allow to quantify the properties of scientific publications. The complex contains metrics that take into account the relationship of articles and their vocabulary, metrics taking into account the citation and vocabulary items, as well as at the author and journal level metrics allowing to obviate the disadvantages of the known metrics. The use of the proposed metrics allows to automate the analysis and comparison of scientific publications and periodicals.

Keywords: automation, scientometrics, bibliometrics, quality, article, scientific publication, metric, citation analysis, article importance.

REFERENCES

- Article-Level Metrics. A SPARC Primer. [Electronic resource]. Access mode: <http://www.sparc.arl.org>
- Page L., Brin S., Motwani R., Winograd T. The PageRank Citation Ranking: Bringing Order to the Web [Electronic resource]. Stanford, Stanford InfoLab, 1999. Access mode: <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>
- Moed H. F. Measuring contextual citation impact of scientific journals, *Journal of Informetrics*, 2010, Vol. 4, No. 3, pp. 265–277.
- González-Pereira B., Guerrero-Bote V. P., Moya-Anegón F. A new approach to the metric of journals' scientific prestige: The SJR indicator, *Journal of Informetrics*, 2010, Vol. 4, No. 3, pp. 379–391.
- Guerrero-Bote V. P., Moya-Anegón F. A further step forward in measuring journals' scientific prestige: The SJR2 indicator, *Journal of Informetrics*, 2012, Vol. 6, No. 4, pp. 674–688.
- Butler D. Free journal-ranking tool enters citation market, *Nature*, 2008, Vol. 451, No. 6, P. 6.
- Falagas M. E., Kouranos V. D., Arencibia-Jorge R., Karageorgopoulos D. E. Comparison of SCImago journal rank indicator with journal impact factor, *The FASEB Journal*, 2008, Vol. 22, No. 8, pp. 2623–2628.
- Bergstrom C. T., West J. D., Wiseman M. A. The Eigenfactor Metrics, *Journal of Neuroscience*, 2008, Vol. 28, No. 45, pp. 11433–11434.
- Hirsch J. E. An index to quantify an individual's scientific research output, *Proceedings of The National Academy of Sciences*. 2005, Vol. 102, No. 46, pp. 16569–16572.
- Egghe L. Theory and practise of the g-index, *Scientometrics*, 2006, Vol. 69, No. 1, pp. 131–152.
- Woeginger G. J. An axiomatic analysis of Egghe's g-index, *Journal of Informetrics*, 2008, Vol. 2, No. 4, pp. 364–368.
- Tol R. S. J. A rational, successive g-index applied to economics departments in Ireland, *Journal of Informetrics*, 2008, Vol. 2, No. 2, pp. 149–155.
- Kosmulski M. I – a bibliometric index, *Forum Akademickie*, 2006, Vol. 11, pp. 31.
- Prathap G. Hirsch-type indices for ranking institutions' scientific research output, *Current Science*, 2006, Vol. 91, No. 11, pp. 1439.
- Igra v cyfir', ili kak teper' ocenivayut trud uchenogo (cbornik statej o bibliometrike). – Moscow, MCNMO, 2011, 72 p.
- Cyganov A. V. Kраткое описание наукометрических показателей, основанных на цитируемости, *Управление большими системами : сб. тр. Спец. вып. 44, Наукометрия и экспертиза в управлении наукой*. Moscow, IPU RAN, 2013, pp. 248–261.
- Shtovba S. D., Shtovba E. V. Обзор наукометрических показателей для оценки публикационной деятельности ученого, *Управление большими системами : сб. тр. Спец. вып. 44, Наукометрия и экспертиза в управлении наукой*. Moscow, IPU RAN, 2013, pp. 262–278.
- Subbotin S. O. Model' ta indyvidual'ni metryky jakosti naukovykh publikacij, *Radioelektronika, informatyka, upravlinnja*, 2015, No. 3, pp. 30–39.
- Global Gazetteer [Electronic resource]. – Access mode: <http://www.fallingrain.com/world/>