

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
**«Южно-Уральский государственный университет
(национальный исследовательский университет)»**
Высшая школа электроники и компьютерных наук
Кафедра системного программирования

РАБОТА ПРОВЕРЕНА
Рецензент
программист ООО «Прикладные
технологии»

_____ Т.В. Речкалов
« ____ » _____ 2018 г.

ДОПУСТИТЬ К ЗАЩИТЕ
Заведующий кафедрой
д.ф.-м.н., профессор

_____ Л.Б. Соколинский
« ____ » _____ 2018 г.

**РАЗРАБОТКА СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ
В УПРАВЛЕНИИ НАУЧНЫМИ ИССЛЕДОВАНИЯМИ
НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА
НАУКОМЕТРИЧЕСКИХ ДАННЫХ**

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
ЮУрГУ – 02.04.02.2018.308-474.ВКР

Научный руководитель
к.ф.-м.н., доцент

_____ М.Л. Цымблер
« ____ » _____ 2018 г.

Автор работы
студент группы КЭ-220

_____ Д.В. Валько
« ____ » _____ 2018 г.

Ученый секретарь
(нормоконтролер)

_____ О.Н. Иванова
« ____ » _____ 2018 г.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	4
1. АНАЛИЗ И ИСПОЛЬЗОВАНИЕ НАУКОМЕТРИЧЕСКИХ ДАННЫХ	11
1.1. Развитие наукометрических показателей и источников наукометрических данных	11
1.2. Обзор подходов к оценке публикационной активности	18
1.3. Обзор программных решений в области наукометрии	23
2. МЕТОДИЧЕСКИЙ ИНСТРУМЕНТАРИЙ АНАЛИЗА ПУБЛИКАЦИОННОЙ АКТИВНОСТИ	27
2.1. Методика построения рейтинга публикационной активности	27
2.2. Методика оценки публикационного потенциала ученого и принципы формирования индивидуальных рекомендаций	32
3. ПРОЕКТИРОВАНИЕ И РЕАЛИЗАЦИЯ ПРОГРАММНОЙ СИСТЕМЫ ..	38
3.1. Проектирование программной системы	38
3.2. Разработка и тестирование программной системы	45
3.3. Оценка публикационного потенциала ученых Южно-Уральского государственного университета	47
ЗАКЛЮЧЕНИЕ	51
СПИСОК ЛИТЕРАТУРЫ	52

ВВЕДЕНИЕ

Актуальность темы

В мировой практике функционирования института науки, в качестве основы оценки научной мысли и продуктивности научной деятельности используются два подхода – экспертный (качественный) и наукометрический (количественный) [50]. В последние годы в качестве инструмента оценки эффективности деятельности российских ученых, исследовательских организаций, отечественной науки в целом стали активно использоваться данные об уровне и числе публикаций, числе и качестве цитирований, представленных разнообразными наукометрическими показателями. Публикационные и цитатные наукометрические показатели рассматриваются как целевые индикаторы состояния науки в «Стратегии инновационного развития Российской Федерации на период до 2020 года», утвержденной распоряжением Правительства РФ [48] и в Указе Президента РФ «О мерах по реализации государственной политики в области образования и науки» от 7 мая 2012 года [49].

Несмотря на то, что нет объективных оснований считать, что наукометрические показатели, основанные на измерении публикаций и цитат, способны напрямую отражать уровень и качество научных исследований, как таковых [54], многие исследователи отмечают, что разработка наукометрических измерений и их сопоставление открывают новые грани научной деятельности, дают возможность скорректировать ее направления, структуру и содержание, дать ей оценку, сравнить ее результаты с результатами других исследований [64].

Сегодня исследователю для решения задачи наиболее эффективного опубликования результатов исследований, в числе прочего, необходимо ответить на следующие вопросы:

- в каком виде и объеме опубликовать новые научные результаты;
- какие издания могут быть рекомендованы для опубликования научных результатов;

- как выбрать авторитетный научный журнал;
- как на эти вопросы отвечают наиболее успешные коллеги.

Очевидно, что не всякий исследователь обладает свободным временем и компетенциями, чтобы отвечать на эти вопросы, при этом от ответов на них во многом зависит его репутация и доступность результатов исследования для мирового научного сообщества. Особенно, данная проблема обостряется на уровне руководства научными коллективами, лабораториями и исследовательскими организациями в целом, когда руководителю необходимо принимать управленческие решения по поводу финансирования тех или иных исследований, обеспечения кадровыми и иными ресурсами. Руководитель нуждается в аналитической информации относительно перспективности тех или иных исследований, исследовательского и публикационного потенциала тех или иных исследователей.

Некоторую информацию для руководителя предоставляют международные реферативные и библиографические базы данных и их аналитические приложения. Рассчитываемые в них индивидуальные наукометрические показатели играют ключевую роль при изучении динамики научных исследований, построении международных рейтингов исследовательских и образовательных организаций, прохождении аттестации исследователями и принятии решения о выделении исследовательских грантов.

На современном этапе развития информационных технологий важную роль приобретают интеллектуальные рекомендательные системы. *Рекомендательные системы* – это программы, которые пытаются предсказать, какие объекты (фильмы, музыка, книги, публикации, веб-сайты) будут интересны пользователю, имея определенную информацию о его профиле [17]. Разумеется, и наукометрический профиль исследователя или исследовательской организации в целом в той или иной наукометрической базе данных вполне отвечает необходимым условиям разработки рекомендаций.

Таким образом, мы полагаем, что в настоящее время целесообразна разработка программной системы, призванной помочь руководителям исследовательских групп, а также индивидуальным исследователям принимать решения по поводу улучшения публикационной активности, наукометрической результативности, определять слабые и сильные стороны, направления и конкретные решения в виде автоматизированных рекомендаций по повышению величины и репрезентативности наукометрических показателей.

Рассматриваемая тема приобретает особую актуальность для российских университетов, включенных в государственную программу «Проект 5-100», направленную на повышение конкурентоспособности университетов среди ведущих мировых научно-образовательных центров.

Степень разработанности темы

В области наукометрии и инфометрии в целом наработан обширный базис математических методов для исследования динамики научных исследований и публикационной активности (см. работы В.В. Налимова, З.М. Мульченко, Л. Блэкерта (L. Blackert), С. Зигеля (S. Siegel), О. Наке (O. Naske) и др.). Известны в международной практике аналитические методы и показатели, наиболее полно отвечающие задаче индивидуальной наукометрической оценки автора (Дж. Хирш (J. Hirsch)), статьи и журнала (Ю. Гарфилд (E. Garfield)) и т.п. Совершенствуется аналитический инструментарий (InCites, SciVal etc.) в программных системах международных реферативно-библиографических баз данных Scopus, Web of Science и др.

Вместе с тем, недостаточное внимание, как в теории, так и в практике уделено поддержке принятия управленческих решений в сфере менеджмента научной деятельности: отсутствуют универсальные методики оценки публикационной активности и построения публикационного рейтинга в зависимости от управленческих задач; недостаточно изучены вопросы формирования и управления публикационным потенциалом отдель-

ных исследователей и исследовательских коллективов; не проработаны вопросы работы с «большими» наукометрическими данными с применением методов интеллектуального анализа; отсутствуют программные разработки в области поддержки принятия решения и разработке индивидуальных рекомендаций по улучшению публикационной активности, наукометрической результативности и реализации исследовательского и публикационного потенциала.

Цель и задачи работы

Целью работы является разработка системы поддержки принятия решений в управлении научными исследованиями, формирующей индивидуальные рекомендации по улучшению публикационной активности.

Данная цель определяет следующий круг задач:

- 1) обзор работ в области оценки публикационной активности ученых;
- 2) разработка методики построения публикационного рейтинга на основе профилей ученых в Scopus;
- 3) разработка программной системы для формирования рекомендаций на основе анализа наукометрических данных;
- 4) проведение экспериментов на данных профилей ученых ЮУрГУ.

Теоретическая и практическая значимость работы

Теоретическая ценность работы состоит в том, что в ней сформулировано определение публикационного потенциала ученого; дано формальное описание методов интеллектуального анализа наукометрических данных для решения задачи оценки публикационного потенциала ученого и разработки индивидуальных рекомендаций по улучшению публикационной активности.

Практическая ценность работы заключается в том, что на базе предложенных подходов и алгоритмов разработана программная система, призванная помогать руководителям исследовательских групп, а также исследователям принимать решения по поводу повышения публикационной ак-

тивности, наукометрической результативности, определять слабые и сильные стороны, направления и конкретные решения в виде автоматизированных рекомендаций по повышению величины и репрезентативности наукометрических показателей.

Методология и методы исследования

Методологической основой исследования является реляционная алгебра и методы интеллектуального анализа данных (datamining). Вычислительные алгоритмы реализованы на базе интеллектуальной аналитической системы KNIME Analytics Platform. При разработке программной системы применялись методы объектно-ориентированного проектирования и язык UML.

Апробация результатов исследования

Основные положения диссертационной работы, разработанные методы, алгоритмы и результаты экспериментов докладывались автором на следующих международных и всероссийских научных мероприятиях.

1. Всероссийская молодежная научно-практическая конференция «Компьютерные технологии и телекоммуникации 2016 (КТиТК-2016)», г. Грозный, 20-23 декабря 2016 г.

2. IV Всероссийский конкурс научных работ и научных проектов «Научный прорыв – 2017», г. Челябинск, 15 февраля 2017 г.

3. VI Всероссийская научно-практическая конференция студентов, аспирантов и молодых ученых «Актуальные проблемы современной науки: взгляд молодых – 2017», г. Челябинск, 25 апреля 2017 г.

4. X Международная научно-практическая конференция «Инновационное развитие российской экономики», г. Москва, 25-27 октября 2017 г.

5. 13-я Международная молодежная научно-техническая конференция «Современные проблемы радиоэлектроники и телекоммуникаций, РТ-2017», г. Севастополь, 20-24 ноября 2017 г.

Публикации по теме работы

Основные результаты диссертации опубликованы в следующих научных работах:

Публикации в журналах из перечня ВАК

1. Рекомендательная система на основе интеллектуального анализа наукометрического профиля исследователя // Программные продукты и системы. 2018. №2(31). С. 561-566.

2. О методике систематической оценки и формирования рейтинга публикационной активности исследователей // Модели, системы, сети в экономике, технике, природе и обществе. 2017. № 3(23). С. 17-30.

Публикации в других изданиях

3. К вопросу об эффективной публикационной активности исследователя // Вестник Совета молодых ученых и специалистов Челябинской области. 2017. №1(16). С. 56-58

4. Идентификация наукометрического паттерна университетов в проекте «5-100» на основе кластерного анализа // Управление в современных системах. 2018. № 1(17). С. 44-48.

5. Интеллектуальный поиск ассоциативных правил в наборе наукометрических данных // Компьютерные технологии и телекоммуникации: Сборник трудов IV Всерос. молодежной науч.-практ. конф. Махачкала: УГНТ, 2017. С. 123-126.

6. Роль наукометрических показателей в оценке научных результатов // Актуальные проблемы современной науки: взгляд молодых: сборник трудов VI Всерос. науч.-практ. конф. студ., аспирантов и молодых ученых. Челябинск: ЮУИУиЭ, 2017. С. 478-481.

7. Архитектура рекомендательной системы на основе интеллектуального анализа наукометрического профиля исследователя / Современные проблемы радиоэлектроники и телекоммуникаций «РТ – 2017»: материалы 13-й междунар. молодежной науч.-техн. конф. Севаст.: СевГУ, 2017. С. 216.

Свидетельства о регистрации программ для ЭВМ

1. Свидетельство о государственной регистрации программы для ЭВМ «Система поддержки принятия решений в сфере менеджмента научных исследований на основе интеллектуального анализа наукометрических данных» № 2018613068 от 02.03.2018.

Структура и объем работы

Диссертация состоит из введения, трех глав, заключения и библиографии. Объем диссертации составляет 59 страниц, объем библиографии – 68 наименований.

Краткое содержание работы

Раздел «Анализ и использование наукометрических данных» посвящен рассмотрению истории развития и практики применения наукометрических показателей и источников наукометрических данных, обзору подходов к оценке публикационной активности ученых, а также обзору отечественных и зарубежных программных решений в области наукометрии.

Раздел «Методический инструментарий анализа публикационной активности» содержит формализацию задачи оценки публикационной активности и авторские разработки в этом направлении.

Раздел «Проектирование и реализация программной системы» посвящен проектированию, реализации и тестированию программной системы, построенной на основе методических разработок автора.

1. АНАЛИЗ И ИСПОЛЬЗОВАНИЕ НАУКОМЕТРИЧЕСКИХ ДАННЫХ

1.1. Развитие наукометрических показателей и источников наукометрических данных

Наукометрия, как дисциплина

В энциклопедии Института философии РАН [41] *наукометрия* определена как дисциплина, изучающая эволюцию науки через многочисленные измерения и статистическую обработку научной информации. Наукометрия, возникла из *статистической библиографии* (данный термин ввел библиограф Е. Хамл в 1923 г., применив его к ранжированию стран по числу журнальных статей в определенных областях) и в настоящее время наряду с *библиометрией* и *вебометрией*, является составляющей частью *инфометрии*.

Первые попытки количественного изучения публикаций были сделаны А. Шторхом и Ф. Аделунгом [67], ими в 1810 г. был составлен библиографический указатель, отражающий отечественные публикации 1801-1806 гг. по разным параметрам. Отечественный исследователь П. Вальден [32] в 1911 г. впервые применил метод анализа цитирования для изучения вклада ученых отдельных стран в развитие химии. Затем, в 1917 г. была опубликована работа Ф. Коула (F. Cole) и Н. Ильса (N. Eales) [7], в которой был проведен статистический анализ научных публикаций по сравнительной анатомии.

Базовый термин «библиометрия» впервые встречается в работах основоположника теории информатики и разработчика универсального десятичного классификатора (УДК) П. Отле (P. Otlet) в 1934 г. С этого времени интенсивно развивается применение математических и статистических методов к изданиям и различным публикациям, а также обнаруживаются базовые закономерности, используемые в *инфометрии* (распределение Ципфа [27], закон рассеяния Бредфорда [4], закон обратного квадрата Лотки [18] и др.).

В конце 50-х годов XX века в связи с экспоненциальным ростом числа научных журналов и публикаций было обращено внимание на изменение характера публикационной активности и значимости научных исследований [55]. Однако, наукометрия как отдельное направление формируется позже – в 60-е годы, при этом сам термин «наукометрия» (получивший в дальнейшем распространение как *scientometrics*) введен в работе В.В. Налимова и З.М. Мульченко [44] только в 1969 г.

В 1979 г. в статьях Л. Блэкерта (L. Blackert), С. Зигеля (S. Siegel) [3] и О. Наке (O. Nacke) [19] появился термин «инфометрия», который был определен как комплекс математических методов для исследования объектов информационной науки, описания и анализа их свойств, а также закономерностей в целях оптимизации этих объектов при принятии решений. В последствии *инфометрия* становится более широким понятием, включающим *наукометрию* [21].

Развитие методов наукометрии

В основе наукометрического инструментария лежат закономерности и подходы, сформулированные в рамках *библиометрии*, но одних их недостаточно для исследования научных публикаций [11]. По этой причине возникли аналитические методы и показатели, наиболее полно отвечающие задаче индивидуальной наукометрической оценки: импакт-фактор, индекс Хирша и т.п.

Импакт-фактор

По мере глобализации науки и стремительного роста объема научной периодики возникла необходимость дифференцирования научных журналов по значимости и качеству. Идея формирования некоего рейтинга научных журналов и придания ему количественного выражения принадлежит Ю. Гарфилду (E. Garfield) [9]. С 1964 г. в возглавляемом им Институте научной информации (ISI) стал выпускаться, так называемый *Science Citation Index (SCI)* [10], систематизирующий информацию о цитировании

научных статей по библиографической базе данных¹. SCI стал прообразом библиографического ресурса Web of Knowledge компании Thomson Reuters, включающего, в частности, базу библиографических данных Web of Science (WoS) с приложением Journal Citation Reports (JCR).

По Гарфилду *импакт-фактор* научного журнала – это средняя цитируемость опубликованных в нем статей, характеризующая важность и авторитетность данного журнала [31]. Классический *импакт-фактор* определяется как число ссылок, полученных в данном году из всех имеющихся в библиографической базе журналов, на статьи, опубликованные в данном журнале в течение *двух* предыдущих лет, отнесенное к числу этих статей. В дальнейших работах ISI показано, что для определенных задач² (дифференциации новых журналов, анализа передовых областей исследования и т.п.) требуется модификация расчета импакт-фактора, например, увеличение периода учета цитат³.

Индекс Хирша

Для наукометрической характеристики отдельного исследователя (по аналогии с импакт-фактором для журналов) Дж. Хиршем (J. Hirsch) [15] в 2005 г. был предложен и, в дальнейшем, модифицирован [16] специальный *h-индекс*. *h-индекс* определяется следующим образом: исследователь, опубликовавший *N* статей, имеет индекс равный *h*, если *h* его статей цитируются не менее чем *h* раз каждая, а любая из оставшихся (*N – h*) статей цитируется не более чем *h* раз. Данный индекс позиционируется его автором, как характеризующий актуальность, востребованность и накопленное влияние индивидуальных научных результатов [15], несмотря на то, что данные свойства *h-индекса* неоднократно подвергались и подвергаются конструктивной критике [23].

В частности, поскольку при достижении исследователем больших значений индекса проявляется его инерционность (он может годами оста-

¹ Библиографическая база данных – совокупность библиографических описаний в структурированном виде

² The Thomson Reuters Impact Factor. URL: <http://wokinfo.com/essays/impact-factor/>

³ Методика расчета импакт-фактора в РИНЦ. URL: http://elibrary.ru/help_title_if.asp

ваться постоянным), для отслеживания и прогнозирования публикационной динамики многими авторами предлагаются так называемые рациональные модификации, например *hrat*-индекс [24, 12]. Целая часть этого показателя эквивалентна обычному индексу Хирша, а дробная часть показывает, насколько исследователь приблизился к следующему значению индекса Хирша.

Кроме того, *h*-индекс сильно варьирует в разных областях науки (с различной нормой цитирования и соавторства), поэтому для обеспечения сопоставимости наукометрических результатов П. Батистой (P. Batista) и др. [2], предложен дополненный индекс Хирша (h'), вычисляемый по формуле $h' = h^2/N$, где N – число соавторов для данной публикации. Разнообразные по целям и свойствам модификации индекса Хирша представлены и в отечественных работах, например [37, 38].

Условие равенства числа статей и числа, полученных каждой из этих статей цитат (ссылок) является в определении индекса Хирша достаточно произвольным [59]. Имея это в виду, Н. Ван Эйк (N. Van Eck) и Л. Валтман (L. Waltman) [8] «обобщили» определение Хирша, назвав его h_α -индексом. Индекс равен h_α , если каждая из h_α статей массива публикаций получила не менее αh цитирований, а каждая из остальных – не более αh цитирований. Основным смыслом данного индекса состоит в том, что, изменяя α можно изменять акцент оценки: при увеличении α на первое место выходит цитируемость, наличие высоко цитируемых работ, а при уменьшении важнее становится продуктивность, большое число публикаций.

Ранжирование и нормирование

При использовании наукометрических показателей для оценки необходимо обеспечивать сопоставимость. При подсчете необходимо четкое обозначение как публикационного окна, так и окна цитирования, в противном случае корректная становится невозможной. Кроме того, известно, что значения абсолютных наукометрических показателей сильно варьируют в зависимости от области исследования, периода времени, географии и др.

Еще одним подходом при проведении кросс-дисциплинарных сопоставлений является отказ от использования собственных значений показателей и сопоставление их рангов, т.е. мест или позиций в рейтинге. Например, после сортировки списков исследуемых объектов (исследователей, журналов, статей) в двух различных дисциплинах в порядке убывания какого-либо выбранного наукометрического показателя проводится анализ только занятых порядковых мест.

Суть рангового метода – разбиение полученного упорядоченного списка-рейтинга на n равных частей и определение, в какую из этих частей попадают исследуемые объекты. Если n принимается равным 4, тогда говорят о квартилях. Например, в ранговой системе считается, что журналы, попавшие в первый квартиль, выше журналов, попавших во второй, не только в этой, но и в любой другой дисциплине. При этом важна корректность и полнота составления списка для ранжирования. Преимущество ранговых методов – защита от сильной асимметричности распределений по данному наукометрическому показателю.

Данный метод широко распространен формируются, как мировые наукометрические рейтинги университетов (World University Rankings, Academic Ranking of World Universities, Leiden Ranking), так и отдельных исследователей (например по базе данных Web of Science Core Collection компания Clarivate Analytics публикует рейтинг⁴ 250 самых цитируемых ученых по 22 предметным областям). При этом наукометрические показатели (количество статей, их цитируемость, импакт-факторы и др.) являются важной составляющей взвешенной оценки, а в некоторых рейтингах – единственной.

Библиографические и реферативные базы данных

Библиографической называется база данных, содержащая совокупность библиографических описаний изданий и отдельных публикаций в структурированном (ссылочном) виде. *Реферативные* базы данных содер-

⁴ Highly Cited Researchers. URL: <http://hcr.stateofinnovation.com/>

жат помимо библиографических данных также реферативные данные (мета-данные) об изданиях и отдельных публикациях, в частности аннотации.

Как правило, библиографические базы данных являются реферативными, но могут содержать в той или иной форме первичные тексты (например, в РИНЦ полные тексты хранятся для отображения контекста цитирования и контекстного поиска, но при этом могут быть не доступны пользователям по прямому назначению).

На сегодняшний день существует около сотни международных и национальных библиографических и реферативных баз данных, дифференцированных как по отраслям знания, так и по объему и сфере использования. Среди них в качестве основных как по общему мнению, так и на основе критериев Министерства образования и науки РФ, упоминаются РИНЦ, Scopus и Web of Science – это базы данных по научным публикациям, в которых обрабатываются библиографии публикаций, аннотации к публикациям и списки используемой литературы в публикациях.

На основании информации из этих баз данных формируются основные наукометрические показатели (индекс Хирша, показатели цитируемости, импакт-факторы) научной эффективности организаций, ученых и научных периодических изданий. Именно возникновение и развитие реферативно-библиографических баз данных на основе информационных технологий обусловило интенсивное развитие методов наукометрии и их практическую значимость.

Scopus

Scopus (Sciverse Scopus)⁵ – крупнейшая в мире мультидисциплинарная библиографическая и реферативная база данных, созданная издательской корпорацией «Elsevier». Scopus охватывает свыше 18 тыс. научных журналов (52 % из них европейские) от 5 000 научных издательств мира, включая около 13 млн патентов США, Европы и Японии, а также материалы научных конференций. Данные о публикациях из Scopus используются

⁵ Официальный сайт Scopus. URL: <http://www.scopus.com>

при составлении международных и отечественных рейтингов университетов: QS World/BRICS/EECA, Times Higher Education, Эксперт РА, Интерфакс.

Web of Science

Web of Science (WoS)⁶ – мультидисциплинарная реферативно-библиографическая база данных Института научной информации США (Institute for Scientific Information, ISI), представленная на платформе Web of Knowledge компании «Thompson Reuters». Web of Science охватывает свыше 12 000 журналов и 148 000 материалов конференций в области естественных, общественных, гуманитарных наук и искусства. Данные о публикациях из Web of Science используются при составлении международных и отечественных рейтингов университетов: U-Multirank, Round Ranking, US News, Global Institutional Profiles Project.

РИНЦ

Российский индекс научного цитирования (РИНЦ)⁷ – национальная информационно-аналитическая система, аккумулирующая более 4,7 миллиона публикаций российских авторов, а также информацию о цитировании этих публикаций из более 4 000 российских журналов. На показатели РИНЦ опираются рейтинговые и информационно-аналитические агентства Эксперт-РА, Интерфакс. Кроме того, Министерство образования и науки РФ при проведении ежегодного мониторинга вузов в качестве ключевых источников информации об эффективности научной деятельности использует данные из Web of Science и РИНЦ.

На сегодняшний день, наукометрические базы данных и рассчитываемые по ним индивидуальные наукометрические показатели играют ключевую роль при изучении динамики научных исследований, построении международных рейтингов исследовательских и образовательных органи-

⁶ Официальный сайт Web of Science. URL: <http://www.webofknowledge.com>

⁷ Официальный сайт РИНЦ. URL: <http://www.elibrary.ru>

заций, прохождении аттестации исследователями и принятии решения о выделении исследовательских грантов.

1.2. Обзор подходов к оценке публикационной активности

Наукометрические показатели позволяют в той или иной мере решать две основные задачи:

- во-первых, оценить вклад в обмен научными результатами, осуществить мониторинг публикационной активности и выполнить ее сопоставление по отдельным исследователям, научным коллективам, исследовательским организациям;

- во-вторых, управлять публикационной активностью, устанавливать тактические и стратегические наукометрические ориентиры, поддерживать баланс публикационной активности в научной деятельности.

Исходя из этих, как правило, последовательно решаемых задач можно обобщенно сформулировать ряд принципов, позволяющих, с одной стороны – обозначить их таксономию, а с другой стороны – определить критерии их применения в контексте каждой задачи. В числе основных принципов формирования системы наукометрических показателей и интегральной оценки публикационной активности, можно назвать следующие.

1. Некомпенсаторность [1] оценки – невозможность компенсации низкого значения одного из используемых критериев высокими значениями по другим критериям в целях сокращения дисбаланса и возможности искусственного манипулирования.

2. Дифференцированность оценки – учет отрасли науки или предметной области, жизненного цикла научного коллектива и научных результатов, при агрегировании и сопоставлении, поскольку различные публикационные кластеры обладают существенно различающейся статистикой и нормой числа соавторов, интенсивности цитирований публикаций и т.п. [34]

3. Консервативность и сбалансированность оценки – игнорирование

флуктуаций и нивелирование ошибок оценивания, ориентация на сбалансированность и устойчивость публикационной активности.

4. Интегративность (аддитивность) оценки – возможность объединения индивидуальных интегральных оценок в обобщенную оценку для характеристики научных коллективов разного масштаба.

5. Институциональная обоснованность – содержательное соответствие применяемых критериев и пороговых значений нормам права, государственным стратегическим ориентирам, лицензионным [53], аккредитационным [52] критериям и параметрам мониторинга эффективности [51] организаций высшего образования Минобрнауки РФ.

6. Комплексность оценки – исчерпывающий охват наукометрических показателей и источников наукометрических данных для всесторонней оценки. В частности, ВАК РФ считает признанными международными реферативными системами Web of Science, Scopus, Web of Knowledge, Astrophysics, PubMed, Mathematics, Chemical Abstracts, Springer, Agris, GeoRef, MathSciNet, BioOne, Compendex, CiteSeerX, а также РИНЦ.

Использование наукометрической оценки для задач управления выдвигает на первый план еще один стратегически важный принцип – обеспечение гибкого использования для системы мотивации научных сотрудников. Всякая оценка должна удовлетворять информационным потребностям основных пользователей, а в данном случае основной пользователь – руководство научного коллектива или индивидуальный исследователь.

В рамках настоящей работы целесообразно остановиться на рассмотрении наиболее актуальных и общих методик наукометрической оценки и уделить внимание их краткой характеристике в разрезе нижеследующих групп.

К *первой группе* следует отнести разнообразные модификации канонических индивидуальных показателей, в частности индекса Хирша [16] и импакт-фактора [11]. Их наиболее полный обзор представлен С.Д. Штовба и Е.В. Штовба [66], а также Б.И. Бедным и Ю.М. Сорокиным [31]. В ука-

занной группе наибольшего внимания заслуживает модификация индекса Хирша отмеченная М.П. Разиным [57], которая на основе корректирующих коэффициентов учитывает отрасль науки и возраст исследователя. Другая интересная методика [25] предлагает корректировать индекс Хирша с помощью взвешивания числа цитат каждой публикации на ее возраст.

Также внимания заслуживает альтернативная дробная модификация индекса Хирша, предложенная С.В. Мавриным [38], учитывающая количество авторов цитируемых статей и иллюстрирующая в дробной части близость автора к следующей ступени индекса Хирша. Аналогичный смысл заложен в методику П. Батисты (Batista P.D.) и др. [2] – предлагается вычислять дополненный индекс Хирша (h') по формуле $h' = h^2 / N$, где N – число соавторов для данной публикации. Показано, что такой индекс устойчив при сравнении наукометрических результатов в разных областях науки (с различной нормой цитирования и соавторства).

В *отдельную подгруппу* следует выделить ряд методик, предназначенных не столько для непосредственной индивидуальной наукометрической оценки, сколько для определения ее состоятельности. Ярким примером является модификация индекса Хирша, предложенная Е.В. Орловым и др. [37]. Она устойчива к манипулированию, т.к. игнорирует искажения в окрестности базового значения Хирша (поскольку опирается на его теоретическое значение, полученное по лучшему тренду).

Ко *второй группе* нами отнесены комплексные или групповые показатели, представляющие собой сводные индексы, показатели на основе усреднения или свертки исходных индивидуальных показателей. В их многообразии можно выделить методику В.Н. Николенко и др. [47], в которой выделены группы показателей продуктивности, авторитетности и результативности в расчете на одного автора.

Подобные методики носят широкую применимость и в оценке эффективности публикационной активности и в использовании для формирования фонда стимулирования труда. Например, в работе Б.И. Бедного и др.

[30], поддержанной Минобрнауки РФ, показано применение взвешенных наукометрических индексов для оценки эффективности подготовки научных кадров в аспирантуре. Затем, Б.Г. Ильясовым и др. [35] предложена интеграция такого рода оценки в информационную систему оценки эффективности научной школы.

В рамках настоящей работы заслуживает внимание методика М.П. Разина [56], для расчета показателя публикационной активности с учетом показателей доступных в РИНЦ, а также возраста ученого и средневзвешенного импакт-фактора журналов, в которых опубликованы и цитируются статьи. Вместе с тем, анализ подобных методик и показателей показывает, что, как правило, применение того или иного способа агрегирования диктуется лишь предпочтениями исследователя и недостаточно обосновывается. В частности, широко применяемая взвешенная сумма значений критериев требует теоретического обоснования возможности суммирования и выбора весов. Для рассматриваемой задачи такого обоснования нет, следовательно, нельзя быть уверенным, что суммирование взвешенных значений наукометрических показателей может дать сколь угодно удовлетворительный результат.

Третья группа объединяет квалиметрические методики на основе пороговой агрегации, балльной и ранговой оценки, а также разнообразные методики ранжирования наборов индивидуальных показателей для оценки публикационной активности и научной продуктивности.

Использование квалиметрических (интервальных, порядковых и абсолютных) шкал для интерпретации наукометрических показателей весьма распространено, особенно в области измерения научного труда и его стимулирования. Подобные методики балльной оценки и ранжирования систем показателей сравнительно недавно предложены Ю.Т. Шарабчиевым [65], О.В. Москалевой [43], С.С. Неустроевым [45] и др. Например, в методике, предложенной В.Г. Никитушкиным и др. [46], индивидуальный рейтинг оценивается по ряду совокупных показателей, в первую очередь, по

степени публикационной активности и цитируемости, а также по соотношению цитируемых публикаций ко всем написанным.

Наиболее любопытной из всех методик в данной группе являются методика порогового агрегирования, дополненная многомерным ранжированием, предложенная Ф.Т. Алескеровым и др. [1]. В данной методике предлагается построение агрегированного рейтинга, по различным наукометрическим критериями. Показано, что при многокритериальном ранжировании невозможна ситуация, когда низкие ранги по отдельным показателям могут быть «заретушированы» высокими оценками по другим. Чем больше критериев используется, тем более статистически значимы различия между присвоенными рангами.

Наша оценка представленных методик в разрезе рассмотренных выше принципов представлена в табл. 1.

Табл. 1. Реализация принципов оценки в различных методиках

Автор и год методики	Некомпенсаторность	Дифференцированность	Консервативность	Интегративность	Обоснованность	Комплексность охвата
Бедный Б.И. и др. (2010) [31, 30]	–	+	–	–	+	–
Разин М.П. и др. (2013) [57]	–	–/+	+	–/+	–	–/+
Алескеров Ф.Т. и др. (2013) [1]	+	–	–	–	–/+	–/+
Николенко В.Н. и др. (2014) [47]	–	+	–	+	–/+	–/+
Маврин С.В. и др. (2015) [38]	–	–/+	–	–/+	–	–/+
Ильясов Б.Г. и др. (2015) [35]	–	–/+	–	–	–/+	+

–/+ – частичная реализация

Анализ таблицы показывает, что корректным решением в описанных условиях может стать комплексная методика, включающая нормирование, поправки на период научной и публикационной активности, а также пороговую агрегацию и ранжирование.

1.3. Обзор программных решений в области наукометрии

В настоящее время в рассматриваемой области преимущественно применяются встроенные аналитические приложения к реферативным базам данных. Каждое такое приложение является проприетарным (InCites, SciVal и др.) и работает только со своей наукометрической базой, в то время как обстоятельный анализ требует сопоставления показателей из РИНЦ, Scopus, Web of Science и внутриуниверситетской информационной системы, в том числе базы портфолио и т.п. Достижения в области информационных технологий ориентируют на открытость наукометрических аналитических приложений (например, Google Scholar, Microsoft Academic Search) и данные сервисы наращивают потенциал [5]. Попытка интеграции нескольких источников данных на принципах открытости предпринята, например, в проекте «Карта российской науки»⁸, но ее нельзя признать состоятельной, поскольку данный проект предполагает одновременную работу только с одним источником данных (в режиме переключения между источниками), а также не предоставляет серьезных аналитических возможностей.

Рассмотрим далее ряд актуальных зарубежных прикладных разработок в данной области.

В работе Ю. Хе (He Y.) и др. [14] предложен алгоритм автоматизации процесса анализа цитирования соавторов и структура соответствующей базы данных. Исходные данные предварительно располагаются в специальной WebCitation-базе данных. В основе разработанного программного инструмента лежат методы кластерного анализа, с помощью которых в результате формируется карта кластеров связанных авторов по той или иной тематике.

В работе Р. Падрос-Кюксарт (Padrós-Cuxart R.) и др. [20] представлен программный инструмент анализа библиометрических и наукометрических данных, который на основе расчета ряда стандартных показателей

⁸ Официальный сайт сервиса «Карта российской науки». URL: <https://mapofscience.ru/>

предоставляет нормированные результаты для исследователей, научно-исследовательских групп, исследовательских центров, факультетов или университета в целом. Инструмент разработан на Java и использует централизованную университетскую MySQL базу данных, в которую администратором подгружаются необходимые сведения об авторах, статьях и журналах.

Программный продукт «Publish or Perish Bibliometric Data Tools», предложенный А. Харцинг (Harzing A.W.) и др. [13], интегрирует набор утилит, позволяющих преобразовывать исходные данные Google Scholar и Microsoft Academic Search, а затем на основе их анализа рассчитать ряд показателей⁹ (рис. 1):

- общее количество публикаций, а также общее количество цитирований;
- среднее число цитат на публикацию, на автора, число цитат в год и др.;
- базисный h -индекс и его основные модификации (в т.ч. среднегодовой прирост индекса).

Среди отечественных научных и прикладных разработок в данной области выделяется интеллектуальная измерительная система «Эйдос»¹⁰, адаптированная к анализу данных РИНЦ [36]. Система является отечественным лицензионным программным продуктом и предназначена для углубленного анализа результатов тестирования, включающего ранговый (информационный) и кластерно-конструктивный анализ эталонных описаний классов распознавания и признаков, а также анализ достоверности заполнения исходных данных.

В рамках задач настоящей работы, система «Эйдос» на основе согласования мнений экспертов и статистического анализа данных формирует классификатор (рис. 2) и определяет статус отдельного автора (ученого).

⁹ Официальный сайт программного продукта «Publish or Perish». URL: <http://www.harzing.com/resources/publish-or-perish>

¹⁰ Официальный сайт программного продукта «Эйдос». URL: http://lc.kubagro.ru/aidos/_Aidos-X.htm

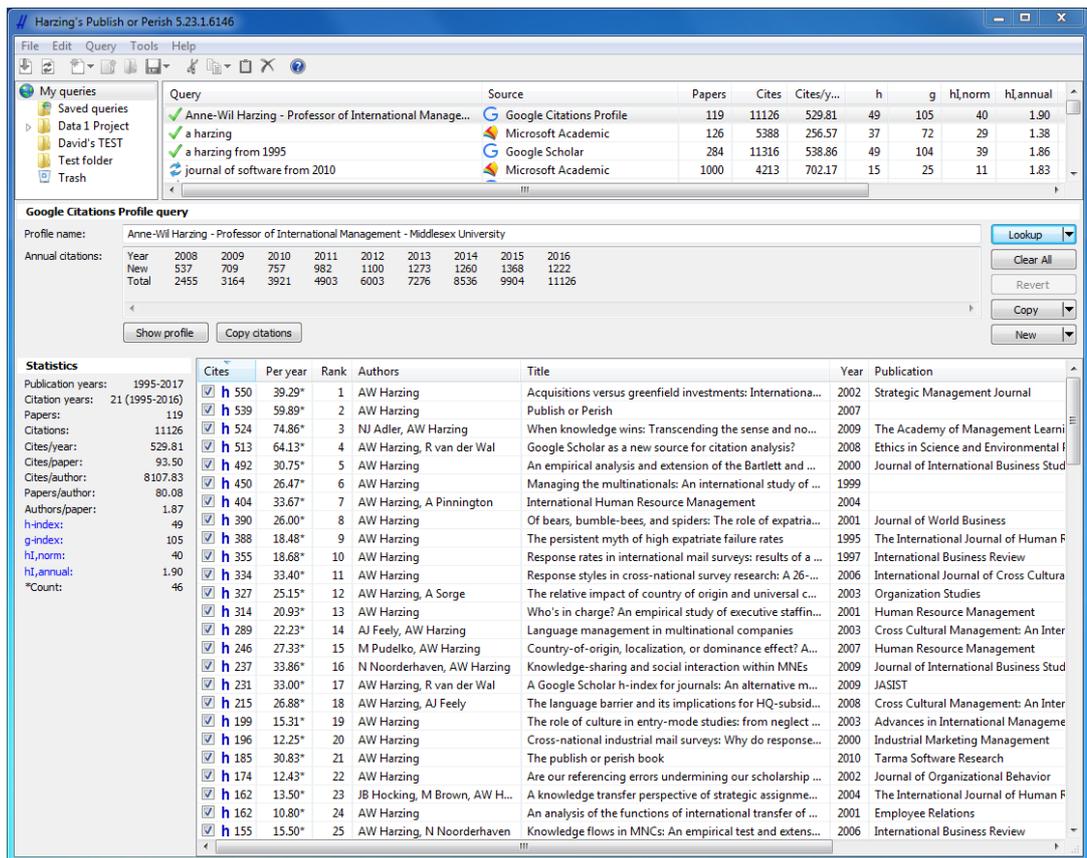


Рис. 1. Рабочая форма PoP Tools

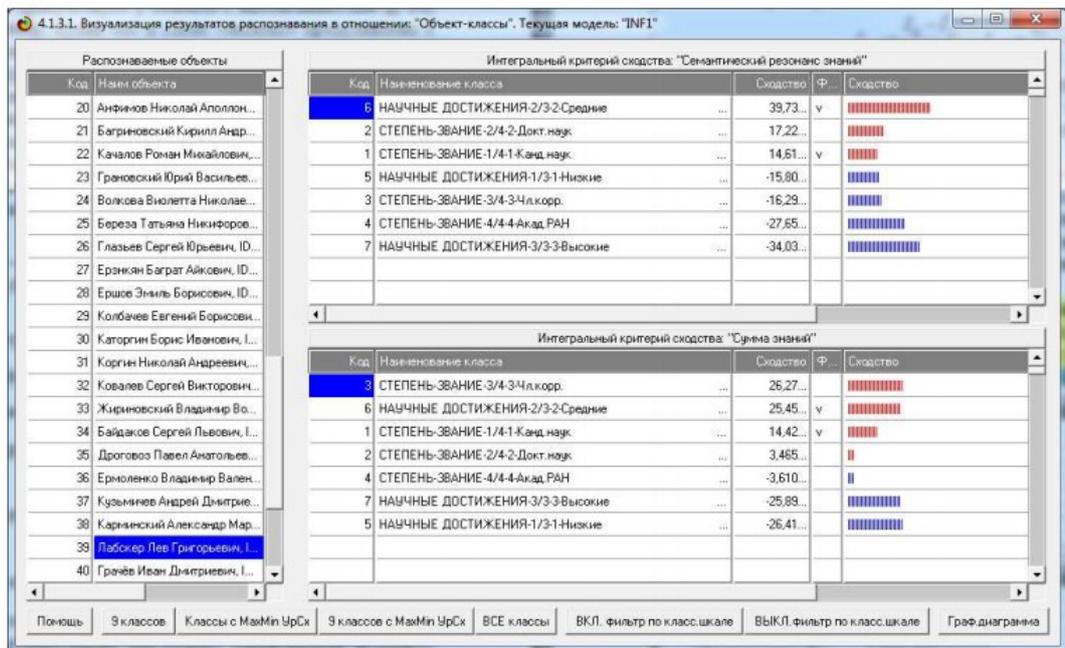


Рис. 2. Рабочая форма системы Эйдос

В работе А.А. Сеницына и др. [60] представлена модульная архитектура системы анализа публикационной активности сотрудников научно-

образовательной организации, включающая модуль для управления пространством сотрудников научно-образовательной организации, модуль для управления областью публикационных объектов и модуль для вычислений и анализа. К сожалению, в данной и подобных работах [61] представлено лишь описание концептуальных блоков и их общее взаимодействие в рамках системы, не ориентированное на какую-либо конкретную реализацию или варианты использования.

В работах И.А. Мбого и др. [40] обсуждаются вопросы разработки инструментов автоматизации и интеграции научной информации в пространстве разнородных информационных систем, для этой цели разработан способ автоматизированного сбора метаданных о публикациях в репозиториях и некоторые шаблоны обработки данных [39].

2. МЕТОДИЧЕСКИЙ ИНСТРУМЕНТАРИЙ АНАЛИЗА ПУБЛИКАЦИОННОЙ АКТИВНОСТИ

2.1. Методика построения рейтинга публикационной активности

Системное и операционализованное решение задачи анализа публикационной активности и оценки индивидуального вклада в обмен научными результатами для целей управления научной деятельностью в целом требует учета базовых принципов в рамках научно-технической политики научной организации или университета.

Прежде всего, следует обеспечить *дифференцированность* оценки. Очевидно, что для решения данной задачи в разрезе различных научных направлений или аффилированных организаций достаточно делать выборку (например, по ключевым словам, принадлежности к организации и др.).

Важным элементом дифференцированного подхода, по нашему мнению, является учет продолжительности публикационной активности. Действительно, сопоставление наукометрических показателей исследователя, работающего в своей области в течение многих лет и исследователя, недавно защитившего диссертацию, является затруднительным. Учет данного замечания обычно выражается во введении поправочного коэффициента пропорционального продолжительности публикационной активности или возрасту исследователя. Мы считаем достаточным взвешивание посредством деления на продолжительность публикационной активности. В данном случае может решаться и другая важная задача – стимулирование молодых ученых. Для научного коллектива, в случае необходимости, в качестве оценки продолжительности публикационной активности следует брать максимальную индивидуальную продолжительность.

Важной задачей является разработка подхода по обеспечению *консервативности* оценки. В ряде работ в этой связи предлагается использовать функции с убывающей производной. Обеспечение убывающего вклада обусловлено необходимостью ограничить спекуляцию наукометрическими показателями и учесть некоторую специфику их динамики (напри-

мер, количество цитирований статьи линейно зависит от числа соавторов [26], а динамика индекса Хирша аппроксимируется сигмоидальной функцией [63]).

В частности, выбор подобной функции (например в [1], $\sqrt[4]{N^2/n}$ где N – число цитат, n – число публикаций) позволяет сделать учет цитирования более консервативным, а вклад каждой следующей цитаты заметно меньше вклада предыдущей. Другим вариантом является использование логарифмического преобразования, при этом логарифмирование наукометрических данных также решает задачу приближения распределения к нормальному.

Принципы *комплексности* и *обоснованности* оценки могут быть реализованы использованием полных наборов данных по показателям из трех основных, закрепленных требованиями мониторинга Минобрнауки РФ [51], наукометрических баз – Web of Science, Scopus и РИНЦ. В зависимости от стратегических задач по стимулированию реферирования в той или иной базе данных, наборам показателей для оценки необходимо присвоить соответствующие веса (в случае ранжирования – приоритетные ранги).

Важным критерием выбора тех или иных наборов показателей, по нашему мнению, является возможность их автоматизированного извлечения из доступных источников и интеграция наукометрической оценки в информационную систему организации (университета).

Опираясь на описанные выше основания, нами предлагается оценка публикационной активности на основе формирования агрегированного рейтинга. Основные этапы методики состоят в следующем (на примере Scopus).

1. Отбор наукометрических показателей, участвующих в построении рейтинга, исходя из принципов оценки и управленческих задач.
2. Формирование массива и выгрузка показателей, общедоступных для проверки в системе Scopus, в разрезе областей исследования.

3. Корректировка значений показателей (взвешивание):

$$x'_i = x_i / (y_i + 1), \quad (1)$$

где:

x_i – исходный показатель по i -му автору;

y_i – корректировка по i -му автору (например, число лет от первой публикации).

Корректировка позволяет получить удельные, сопоставимые по авторам, величины, а также позволяет формировать определенные стимулы публикационной активности, в частности:

– отнесение количества публикаций и индекса Хирша к числу лет от первой публикации с одной стороны обуславливает необходимость поддержания публикационной активности на должном уровне, а с другой – позволяет стимулировать более молодых авторов;

– отнесение общего числа цитат к числу публикаций ставит приоритет на публикации малого числа крупных цитируемых работ по сравнению с многочисленными маловлиятельными работами (предупреждает дробление научных результатов на несколько публикаций);

– отнесение числа публикаций к числу соавторов с одной стороны ориентирует на привлечение продуктивных соавторов и участие в значимых работах, а с другой – предупреждает номинальное соавторство.

4. Преобразование значений показателей в ранги в результате отображения на ранговое множество (сортировка по убыванию):

$\langle x'_i \rangle \xrightarrow{\text{sort} \downarrow} \langle 1, 2 \dots N \rangle$, ранг i -го автора определен, как:

$$r_i = \text{sort} \downarrow (x'_i), \quad r_i \in 1, 2 \dots N. \quad (2)$$

5. Ранжирование показателей по значимости для оценки индивидуальной публикационной активности, затем пороговое агрегирование [33], которое, как ранее отмечено, обеспечивает *некомпенсаторность* итоговой оценки.

На данном этапе следует разделить наукометрические показатели на

две группы: к *первой* относятся показатели, которые априори можно ранжировать по значимости (например, вхождение в top-10 % более значимо, чем в top-25 % и т.п.), а ко *второй* относятся показатели *условно* равной значимости.

Пусть n -мерный вектор рангов всех n показателей (атрибутов профиля автора в системе Scopus) i -го автора, называемый *альтернативой*, определен, как:

$$a_i = \langle r_i^1, r_i^2, \dots, r_i^n \rangle, a_i \in A. \quad (3)$$

Тогда, для *первой* группы показателей ранжирование следует производить, исходя из значимости показателя, т.е. более *предпочтительной* является альтернатива имеющая больший ранг по более значимому показателю (например, вектор с большим рангом по показателю вхождения top-10 %, находится в ранжировании выше, чем вектор с большим рангом по показателю вхождения top-25% и т.п.).

Решение задачи ранжирования наукометрических показателей по значимости может опираться на внутриорганизационную научно-техническую политику, государственные стратегические ориентиры и т.п. Например, настоящая методика разрабатывается для обеспечения реализации Проекта 5-100 Минобрнауки России¹¹ в Южно-Уральском государственном университете, т.е. в данном случае обозначен ряд конкретных показателей публикационной активности научно-педагогических работников, которым отдается предпочтение.

Общее решение может быть также найдено на основе анализа экспертных оценок. Для этого нами проведено анкетирование 20 экспертов, в число которых вошли руководители научной работой разного уровня, что позволило определить согласованную сравнительную значимость наукометрических показателей, доступных в Scopus и РИНЦ (табл. 2). Из числа экспертов 100 % имеют ученую степень и 50 % имеют ученое звание. 30 % имеют опыт руководства научным проектом федерального значения, 60 %

¹¹ Федеральный проект 5/100. – URL: <http://5top100.ru/documents/regulations/>

руководили научным проектом на уровне организации или подразделения. 60 % экспертов осуществляют фактическое и систематическое руководство научным коллективом/проектом.

Табл. 2. Итоги ранжирования наукометрических показателей по сравнительной значимости

Ранг	Наименование показателя	Источник
1	Число публикаций в top-10% журналов	Scopus
2	Число публикаций в top-25% журналов	Scopus
3	Число статей в зарубежных журналах	РИНЦ
4	Число цитирований в зарубежных журналах	РИНЦ
5	Импакт-фактор журналов, в которых были опубликованы статьи	РИНЦ, Scopus
6	Импакт-фактор журналов, в которых были процитированы статьи	РИНЦ
7	Число статей из журналов перечня ВАК	РИНЦ
8	Число цитирований из журналов перечня ВАК	РИНЦ
9	Индекс Хирша	РИНЦ, Scopus
10	Число публикаций автора	РИНЦ, Scopus
11	Число цитирований автора	РИНЦ, Scopus
12	Число соавторов	РИНЦ, Scopus

Коэффициент конкордации (W) для данного ранжирования равен 28,8, что гарантирует статистическую значимость ранжирования при $\alpha=0,05$. При этом эксперты также оценили значимость источников наукометрических данных (реферативных баз) для оценки индивидуального вклада в обмен научными результатами: на первом и втором месте Scopus и Web of Science, соответственно, затем РИНЦ и Google Scholar.

В свою очередь, для *второй* группы показателей необходимо применить пороговое правило ранжирования, поскольку, известно, что такой агрегированный рейтинг будет удовлетворять условию *некомпенсаторности* и Парето-оптимальности ранжирования [29].

Итак, обозначим через $v_m(a)$ – количество рангов m в векторе a , $0 \leq v_m(a) \leq n$. Тогда альтернатива $a_x \in A$ считается (строго) более *предпочтительной*, чем альтернатива $a_y \in A$ (кратко $a_x \succ a_y$), если найдется такой ранг $1 \leq t \leq n$, что $v_m(a_x) = v_m(a_y)$ для всех рангов $1 \leq m \leq t - 1$ и $v_t(a_x) > v_t(a_y)$. Таким образом, результирующий ранг ρ_i i -го ав-

тора из z при многомерном ранжировании с учетом n наукометрических показателей и условия $a_i > a_{i+1}, i \in 1..z$ очевидно равен $\rho_i = i$ (т.е. вектор с большим числом единиц располагается в ранжировании выше, чем вектор с меньшим числом единиц, иначе сравнивается количество двоек и т.п.).

По описанному алгоритму осуществляется отображение множества векторов альтернатив A на ранговое множество: $A \rightarrow \langle 1, 2 .. N \rangle$.

Фрагмент полученного ранжирования на массиве экспериментальных данных представлен на рис. 3. Полученное таким образом ранжирование также является статистически значимым при $\alpha=0,05$ [29].

ID	Итоговый ранг	Ранг по числу публикаций в топ-10% журналов	Ранг по числу публикаций в топ-25% журналов	Ранг по средневзвешенному импакт-фактору	Ранг по индексу Хирша	Ранг по числу публикаций автора	Ранг по числу цитирований автора	...
616541	1	1	1	1	3	5	5	...
777564	2	1	1	1	5	5	5	...
652380	3	1	1	1	10	5	10	...
598902	4	1	1	3	1	2	1	...
656836	5	1	1	6	2	1	1	...
49809	6	1	1	8	3	1	1	...
660255	7	1	2	4	7	10	6	...
16588	8	1	2	5	6	3	4	...
148095	9	1	2	7	5	6	2	...
839852	10	10	1	1	4	5	6	...
819540	11	10	1	1	10	9	10	...
565054	12	10	1	3	1	1	1	...
119431	13	10	1	3	6	9	6	...
119441	14	10	1	5	5	7	4	...
24507	15	10	1	6	2	2	1	...
...

Рис. 3. Пример ранжирования по предложенной методике

Предложенная методика, включая алгоритм сбора, предобработки и ранжирования наукометрических данных представляет собой основу методического инструментария анализа и оценки публикационной активности.

2.2. Методика оценки публикационного потенциала ученого и принципы формирования индивидуальных рекомендаций

В общем смысле, под *публикационным потенциалом* автора следует понимать его возможности по улучшению публикационной активности и,

в частности, по повышению наукометрических показателей без снижения качества. Для целей настоящей работы определение индивидуального публикационного потенциала на основе публикационного рейтинга, рассмотренного выше, опирается на квартильную меру.

Квартили – значения, которые делят таблицу данных (или ее часть) на четыре группы, содержащие приблизительно равное количество кортежей (значений). Общий объем делится на четыре равные части: 25 %, 50 %, 75 % 100 %.

В нашем случае, принадлежать к первому квартилю (или верхнему, q_1) будут авторы, относящиеся к верхним 25 % мест в публикационном рейтинге ($R_{25\%}$); ко второму квартилю q_2 – относящиеся к следующим 25 % мест ($R_{50\%}$) и так далее. Иными словами, в первом квартиле будут находиться 25 % авторов, занимающих лучшие места в публикационном рейтинге, а в четвертом квартиле (q_4) – 25 % занимающих худшие места, соответственно.

С учетом этого, *публикационный потенциал* автора определяется нами, как прогнозная оценка возможностей данного автора по переходу из своего квартиля по рейтингу публикационной активности к лучшему квартилю. Необходимые условия для подобного перехода в тривиальном случае представляют собой такой прирост наукометрических показателей автора, который бы обеспечил существенный подъем в публикационном рейтинге до места, занимаемого последним автором в лучшем квартиле.

Поскольку при построении публикационного рейтинга предпочтение отдается более молодым авторам (по году первой публикации), то определение публикационного потенциала на его основе, сохраняет данную особенность. Это, впрочем, согласуется со спецификой жизненного цикла исследователя: предполагается, что более молодой автор располагает большим публикационным потенциалом.

В целях оценки публикационного потенциала нами предлагается адаптация алгоритма интеллектуального анализа взаимосвязей (поиска ас-

социативных правил) *Apriori* [1], модифицированного К. Боргельтом (С. Borgelt) [6] к наукометрическим данным в системе Scopus. С помощью данного алгоритма, как будет показано далее, могут быть сформулированы некоторые наборы ассоциативных правил, пригодные для прогнозирования вероятных будущих публикационных результатов автора.

Для дальнейшего рассмотрения введем некоторые определения.

Атрибуты профиля ученого (например, индекс Хирша, число публикаций, список журналов и др.): $A_i = \langle a_i : w_i \rangle$, где a_i – название атрибута, w_i – вес атрибута, $1 \leq i \leq m$, $i \in \mathbb{N}$.

Профиль ученого: $P = \{\langle A_1 : v_1 \rangle, \langle A_2 : v_2 \rangle, \dots, \langle A_m : v_m \rangle\}$, $v_i \in \mathbb{R}$.

Рейтинг-лист – последовательность из P , упорядоченная по убыванию значения атрибутов, взятых в соответствии с убыванием их весов: $R = (P_1, P_2, \dots, P_n)$.

Квартильная функция профиля – ставит в соответствие заданному профилю ученого $P \in R$ его квартиль Q по рейтинг-листу R : $Q: R \rightarrow \mathbb{Q}$, $\mathbb{Q} = \{Q_1, Q_2, Q_3, Q_4\}$, $\forall i, j Q_i > Q_j \Leftrightarrow i < j$.

Квартильная функция атрибута – ставит в соответствие заданному профилю ученого $P \in R$ его квартиль q по атрибуту A_i : $q: R \times \mathbb{N} \rightarrow \mathbb{Q}$, $\mathbb{Q} = \{q_1, q_2, q_3, q_4\}$, $\forall i, j q_i > q_j \Leftrightarrow i < j$.

Оценка профиля ученого – функция, которая формирует множество кварталей соответствующих заданному профилю $P \in R$ и всем его атрибутам: $\mathcal{V}: R \rightarrow \mathbb{Q} \cup \mathbb{Q}$, $\mathcal{V}(P) = \bigcup_{i=1}^m q_i(P, i) \cup Q(P)$.

Множество оценок всех профилей ученых: $V = \bigcup_{j=1}^n V_j$, $V_j = \mathcal{V}(P_j)$.

Шаблон публикационной активности – правило вида ЕСЛИ *ant* ТО *cons*, для которого выполняются следующие условия:

- 1) $ant \subseteq \mathbb{Q} \wedge ant \neq \emptyset \wedge cons \in \mathbb{Q} \wedge cons \neq \emptyset$;
- 2) $\forall V_j \in V (ant \subseteq V_j \Rightarrow cons \in V_j)$;
- 3) $\sigma \geq \text{minsupp} \wedge \delta \geq \text{minconf}$.

Поддержка шаблона σ – условная вероятность события, в котором

оценка профиля содержащая $cons$, так же содержит $ant : \sigma = P(cons|ant)$.

Достоверность шаблона δ – отношение вероятности события в котором оценка профиля содержит ant и $cons$, к вероятности события в котором оценка профиля содержит $ant : \delta = P(ant \cup cons)/P(ant)$.

Множество всех шаблонов публикационной активности: \mathcal{AR} .

Множество шаблонов публикационной активности, подходящих профилю ученого $P \in R : ar(P) = \{T | ant \subseteq \mathcal{V}(P), T \in \mathcal{AR}\}$.

Публикационный потенциал ученого с профилем $P \in R : pt(P) = \max_{cons} ar(P)$.

Таким образом, в соответствии с предпосылками применяемого метода интеллектуального анализа, методика оценки публикационного потенциала и алгоритм выявления авторов имеющих такой потенциал, будут состоять в следующем.

1. Для заданного набора авторов строится публикационный рейтинг на основе предложенной методики оценки публикационной активности.

2. На основе полученного публикационного рейтинга набор авторов разбивается на четыре квантили. Первый квантиль (первые 25 % авторов) интерпретируется как лучший на основе оценки публикационной активности, второй квантиль – как хороший и т.д. Каждый автор приобретает синтетический атрибут, содержащий метку принадлежности к квантилю.

3. По каждому атрибуту профиля (наукометрическому показателю), участвующему в построении рейтинга производится квантильная дискретизация аналогичная предыдущему пункту, т.е. формируется набор синтетических атрибутов, представляющих собой метки вида «attrib_QX», где attrib – имя атрибута, а «X» – признак квантиля, например, «h_index_Q1».

Кроме того, значения атрибута эквивалентные минимуму (чаще всего для наукометрических показателей это ноль) заведомо приобретают метку «Q4». Данное коррекционное правило приводит к искажению квантильной структуры по некоторым атрибутам, но обеспечивает верную интерпретацию наукометрического профиля автора.

4. Для полученных наборов синтетических атрибутов по каждому автору, представляющих собой транзакцию, выполняется поиск ассоциативных правил. Затем производится выборка устойчивых правил, обнаруживающих переход авторов из собственных квартилей в лучшие по сочетанию наукометрических атрибутов.

5. На основе полученных правил выявляется набор авторов им соответствующих, т.е. имеющих потенциал перехода к лучшему квартилю рейтинга публикационной активности. При этом прогнозной мерой публикационного потенциала автора является мера достоверности подходящего устойчивого правила.

Разумеется, данная методика может быть легко реализована в рамках информационной системы научно-образовательной организации и использована для задач управления научной деятельностью. Полная автоматизация может быть достигнута интеграцией модуля автоматического сбора (выгрузки) наукометрических данных [42] из используемых реферативных баз.

В заключение рассмотрим принципы формирования индивидуальных рекомендаций. Под *рекомендацией* нами понимается комбинация из набора альтернатив по улучшению показателей публикационной активности до определенного уровня. Фактически, рекомендация для автора представляет собой тезисную интерпретацию результатов наукометрического анализа (и в частности, оценки публикационного потенциала).

Формализованные определения выглядят следующим образом:

Индивидуальная рекомендация – для данного профиля $P_j \in R$ и его целевого квартиля $target \in \mathbb{Q}$ представляет собой *список журналов* для публикации и *список ученых* для сотрудничества: $\langle J, S \rangle$.

Функция кластеризации – разбивает заданный набор профилей на $1 \leq k \leq n$ кластеров и возвращает номер кластера, к которому принадлежит данный профиль P_j : $\rho: \mathbb{N}^2 \rightarrow \{1, 2, \dots, n\}$.

Список профилей ученых целевого квартиля в одном кластере с P_j :
 $\mathcal{S} = \{P_i | Q(P_i) = target, \rho(j, k) = \rho(i, k)\}, P_i \in R, 1 \leq i \leq n .$

Список журналов, в которых публиковались ученые из списка \mathcal{S} :
 $\mathcal{J} = \bigcup_{i=1}^{|\mathcal{S}|} P_i \cdot journals.$

Формирование рекомендации происходит на основе следующих альтернативных шаблонов:

1) ...сконцентрироваться на улучшении следующих показателей (в порядке понижения значимости): X, Y, Z...

2) ...по увеличению X: следует достигнуть уровня – не менее Y, т.е. повысить свой показатель на Z...

3) ...необходимый прирост к следующему отчетному периоду составляет +F...

4) ...по X следует сохранить положительную динамику (превышение на Y)...

5) ...рекомендуются следующие журналы для публикации: X, Y, Z...

3. ПРОЕКТИРОВАНИЕ И РЕАЛИЗАЦИЯ ПРОГРАММНОЙ СИСТЕМЫ

3.1. Проектирование программной системы

Программная система предназначена для интеллектуального анализа наукометрических данных – некоторой совокупности значений наукометрических показателей публикационной активности для набора *авторов*.

Пользователь программной системы представлен двумя ролями (актерами): *автор* (преподаватель, исследователь, имеющий публикации) – использует основной функционал программной системы для получения индивидуальных рекомендаций и *руководитель* – использует программную систему для анализа публикационной активности авторского коллектива, а также готовит программную систему к работе с *автором*.



Рис. 4. Диаграмма вариантов использования

Опишем функциональные возможности (варианты использования) программной системы (рис. 4).

1. Формирование *публикационного рейтинга*. Прецедент, предоставляющий пользователю возможность просмотра *таблицы рейтинга* авторов, построенной по наукометрическим данным, ранее загруженным в

систему (на основе оригинальной методики оценки публикационной активности).

2. Оценка *публикационного потенциала*. Позволяет пользователю ознакомиться с результатами прогноза возможностей авторов по переходу к *лучшему квартилю* в публикационном рейтинге (на основе поиска ассоциативных правил).

3. Поиск групп публикационной активности. В рамках данного прецедента выполняется формирование групп авторов, близких по своей публикационной активности с выделением их типичных представителей (на основе кластеризации с выделением оптимального числа кластеров или с разбиением на заданное число кластеров).

4. Получение индивидуальных *рекомендаций*. Позволяет пользователю получить набор рекомендаций по продвижению в рейтинге публикационной активности и реализации публикационного потенциала (на основе анализа места автора в публикационном рейтинге относительно лучшего, типичного и худшего авторов, а также подбора журналов, в которых опубликованы работы авторов).

5. Подготовка наукометрических данных. Обобщенный прецедент, доступный только руководителю, реализующий программные возможности по подготовке и *предобработке* наукометрических данных авторов, их *загрузке* в программную систему и, по необходимости, *обновлению*.

6. Настройка аналитических моделей. Выполняется руководителем. Прецедент, реализующий программные возможности по настройке рабочих методик, а также параметров моделей интеллектуального анализа наукометрических данных, в дальнейшем используемых системой для работы. В частности, сюда относится выбор источника данных, контроль предобработки данных, подбор минимальной поддержки и достоверности для шаблонов публикационной активности, выбор способа определения числа кластеров и др.

Диаграмма классов, проектируемой программной системы представ-

лена на рис. 5.

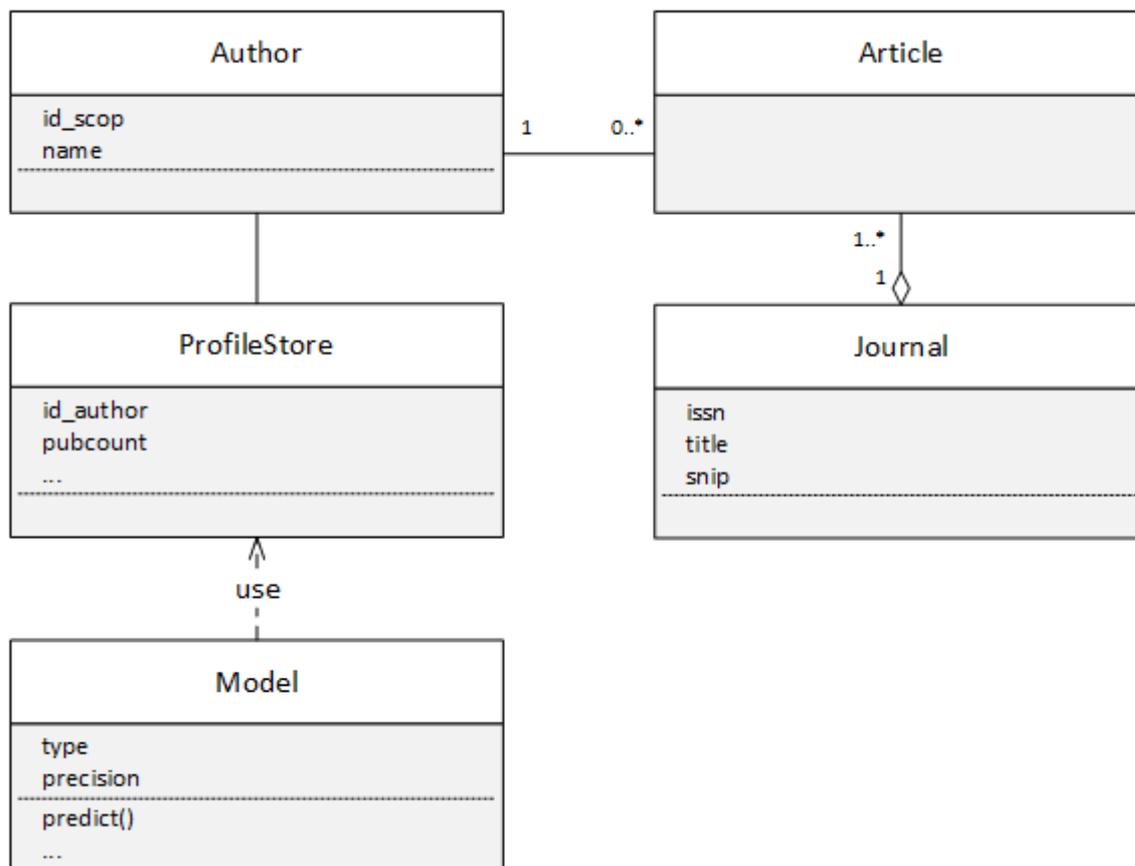


Рис. 5. Диаграмма классов проектируемой программной системы

Для программной системы на данном уровне детализации проектирования существенны классы предметной области: автор (*Author*), статья (*Article*) и журнал (*Journal*). Данные классы ассоциированы в соответствии с логикой предметной области и обладают существенными для программной системы атрибутами.

Автор в программной системе работает с поименованным набором наукометрических данных предоставляемых хранилищем наукометрических профилей (*ProfileStore*) и аналитических моделей (*Model*). Пользователь может создавать, удалять и изменять имеющиеся в системе наборы данных и моделей. Набор создается путем отбора данных и моделей из загруженных и подготовленных руководителем (модератором) ранее, в зависимости от области исследований и других критериев. На основе имеющихся наборов данных, определенных атрибутами наукометрического

профиля автора в системе Scopus, программная система строит *публикационный рейтинг*.

Аналитическая модель (*Model*) представляет собой, созданный и настроенный модератором интеллектуальный объект, который используется для оценки и прогнозирования принадлежности автора к той или иной *публикационной группе*. Предполагается, что автор может быть отнесен к четырем основным публикационным группам, соответствующим квартилям публикационного рейтинга, а также к одной из естественных публикационных групп (например, в случае кластеризации с выделением оптимального числа кластеров).

Перечень атрибутов, хранимых в *ProfileStore*, представлен в табл. 4, 5.

Табл. 4. Перечень наукометрических атрибутов открытого профиля автора в системе Scopus

№	Наименование атрибута	Описание
1	pubcount	Число публикаций автора
2	citcount	Число цитат автора
3	hindex	Индекс Хирша
4	authcount	Число соавторов
5	yearstart	Год первой публикации

При этом в соответствии с методикой для формирования рейтинга и реализации других вариантов использования применяется предобработка и вычисление дополнительных атрибутов (табл. 5).

На основе выбранной модели и имеющихся в хранилище данных, методом *predict ()* производится оценка и прогнозирование принадлежности автора к *публикационной группе*, оценка публикационного потенциала.

Программная система состоит из трех основных компонентов, а также использует внешнюю интеллектуальную систему, обрабатывающую данные на платформе KNIME Analytics Platform¹².

¹² Официальный сайт KNIME Analytics Platform. – URL: www.knime.org/knime-analytics-platform

Табл. 5. Перечень преобработанных наукометрических атрибутов

№	Наименование атрибута	Характеристика
1	top10_sum	Общее количество публикаций в TOP10% изданий по импакт-фактору отнесенное к числу лет от первой публикации
2	top25_sum	Общее количество публикаций в TOP25% изданий по импакт-фактору отнесенное к числу лет от первой публикации
3	snip_mean	Средневзвешенный импакт-фактор журналов в которых имеются публикации
4	h_index	Индекс Хирша отнесенный к числу лет от первой публикации
5	doc_count	Общее количество публикаций отнесенное к числу лет от первой публикации
6	cit_count	Общее количество цитат отнесенное к числу публикаций
7	doc_per_auth	Число публикаций отнесённое к числу соавторов

Основные компоненты предоставляют и используют соответствующие интерфейсы (рис. 6):

- пользовательский интерфейс программной системы и/или интерфейс программирования (компонент *Gateway*);
- компонент хранения данных (компонент *DataSystem*);
- базовая вычислительная подсистема (компонент *BaseSystem*), реализующая взаимодействие с хранилищем данных, вычисления и обработку результатов работы внешней интеллектуальной системы KNIME Analytics Platform.

Развертывание (см. рис. 7) прототипа программной системы предусматривается на одном вычислительном узле (персональном компьютере) с предустановленной актуальной версией KNIME Analytics Platform.

Базовая вычислительная подсистема (компонент *BaseSystem*) предусматривает модульную организацию таким образом, что каждый модуль независимо реализует один из вариантов использования. Структура основных модулей показана на рис. 8.

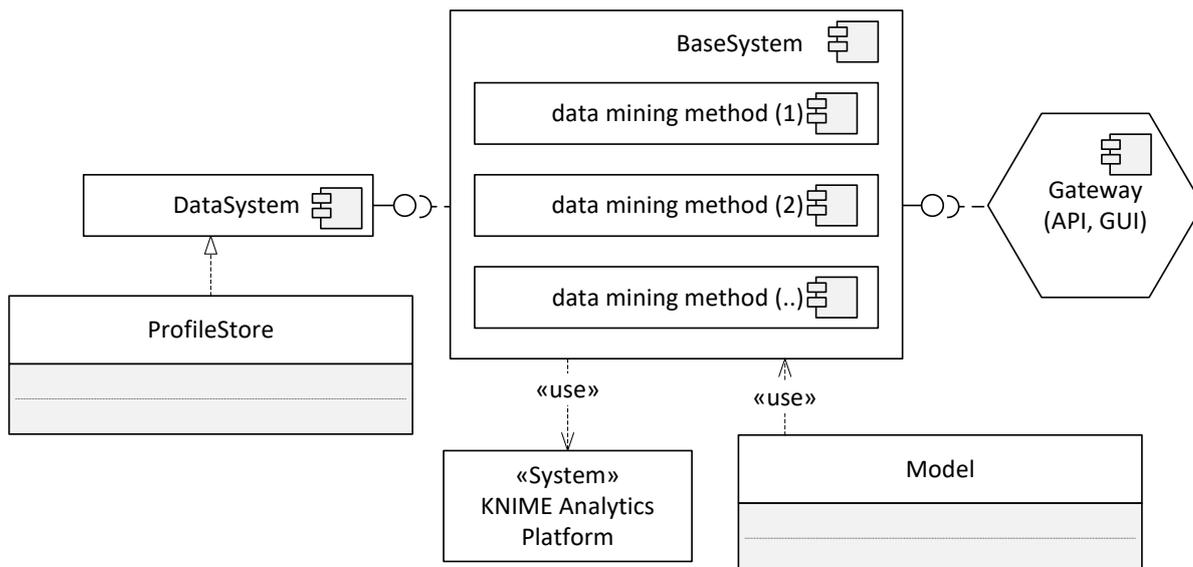


Рис. 6. Диаграмма компонентов программной системы

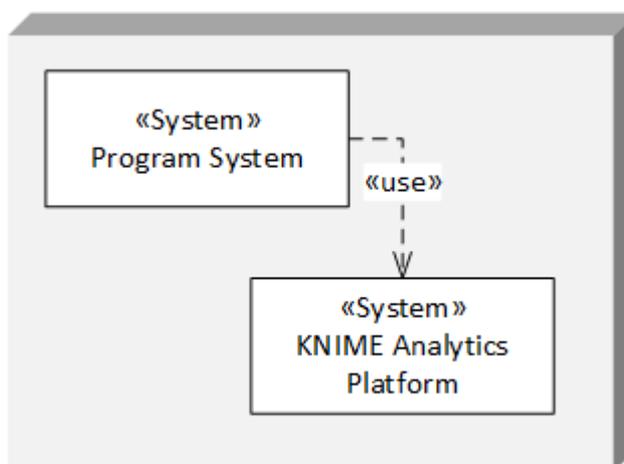


Рис. 7. Диаграмма развертывания программной системы

Как видно из рисунка, ядром функционального модуля является вычислительный поток (*KNIME workflow*), отдельный экземпляр которого обрабатывает данные на платформе KNIME. Из наименования модулей ясно, что: *scopus_rating*, реализует формирование публикационного рейтинга на основе данных системы Scopus; *scopus_rules* – оценку публикационного потенциала; *scopus_cluster* – поиск групп публикационной активности; а *scopus_preparing* – подготовку и предобработку наукометрических данных.

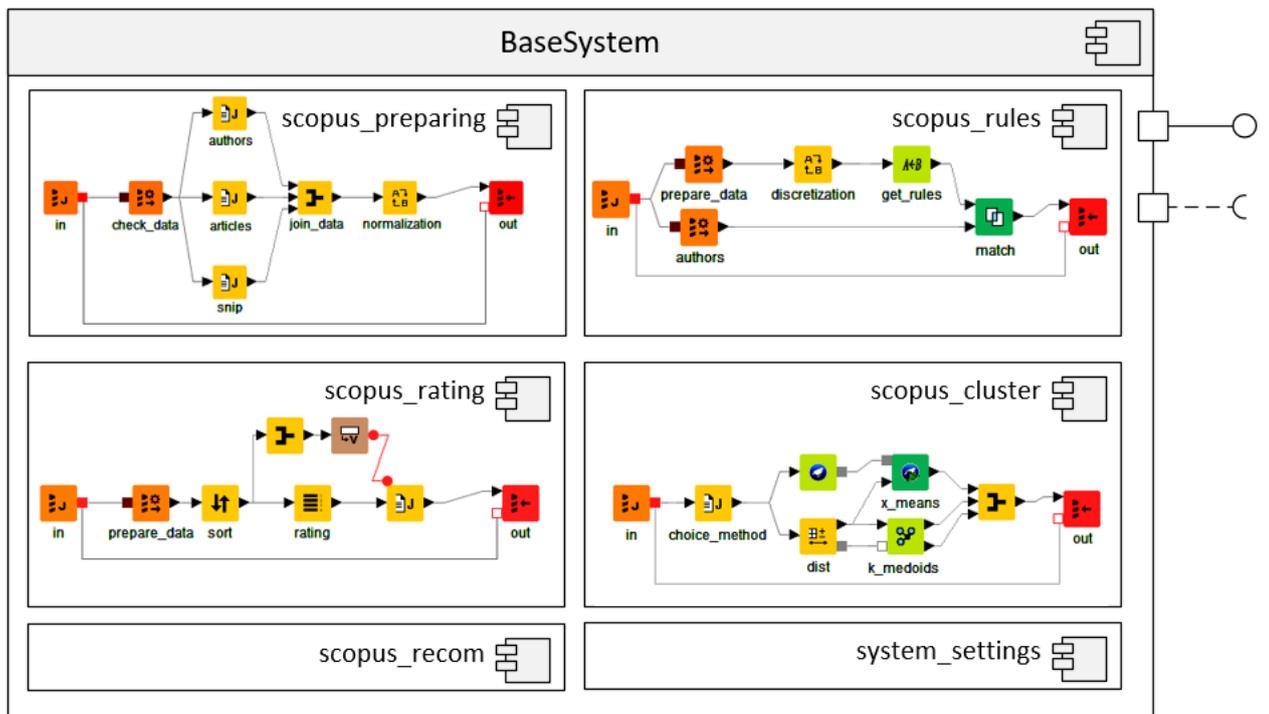


Рис. 8. Схема модульной организации программной системы

В заключение данного раздела представим некоторые соображения по организации пользовательского интерфейса. Схема организации рабочего окна представлена на рис. 9.



Рис. 9. Схема рабочего окна пользовательского интерфейса

Исходя из привычного пользовательского дизайна распространенных операционных систем линейки Windows, рабочее окно должно содержать главное меню (1), предоставляющее доступ к основным параметрам программной системы, справке и помощи, а также область статуса (5), в которой будет отображен лог основных событий и прочая служебная информация.

Основная рабочая область (3) окна предоставляет пользователю различные способы отображения результатов наукометрического анализа, в частности – таблицы, текстовые сведения, графическую информацию. Информация, представленная в рабочей области должна предоставляться в зависимости от выбранного варианта использования и примененных инструментов, при этом кнопки инструментария размещаются на инструментальной панели (4).

Выбор доступных вариантов использования системы осуществляется путем переключения вкладок (2), каждая из которых отвечает за свой вариант использования и предоставляет свой инструментарий на панели (4).

Разумеется, данная схема интерфейса есть только достаточный вариант представления функциональных возможностей системы. В силу модульной организации системы и в случае её реализации в виде, например, микросервисов пользовательский интерфейс может быть выполнен на основе независимого web-представления.

3.2. Разработка и тестирование программной системы

В соответствии с рассмотренными выше проектными материалами, в рамках настоящей работы представлен реализованный прототип программной системы.

В качестве интегрированной среды разработки выбрана Embarcadero Delphi 10¹³. Delphi предоставляет самый быстрый способ написания, компиляции, сборки и развертывания системных и кросс-платформенных при-

¹³ Официальный сайт среды разработки. URL: www.embarcadero.com/ru/products/delphi

ложений, сочетает гибкость современного языка Object Pascal, с native-компиляторами и библиотеками компонентов для быстрой разработки для Windows, macOS, iOS, Android и Linux на базе единого кода.

Выбор данной среды обусловлен минимальными временными затратами на разработку прототипа программной системы и возможностями дальнейшей доработки прототипа до полноценного кроссплатформенного программного комплекса.

Из тех же соображений, простоты и скорости разработки, в качестве инструментария работы с хранилищем данных выбрана технология ActiveX (ADO), включающая клиентские приложения для доступа и управления данными из различных источников через поставщика OLE DB. С учетом того, что ADO поддерживает основные функции для создания клиента и сервера веб-приложений, также предусматривается дальнейшая доработка прототипа хранилища до веб-сервиса.

Поскольку, в сущности, графический интерфейс и хранилище являются лишь способами представления результатов и хранения данных вычислений, а основным вычислительным ядром являются рабочие потоки KNIME, представленные независимыми исполнимыми файлами, разумно использовать два этапа тестирования системы.

В частности, для тестирования разработаны три набора данных, для которых заранее, в полуавтоматическом режиме, проведены все процедуры согласно алгоритмам построения рейтинга, оценки публикационного потенциала и расчета метрик для формирования индивидуальных рекомендаций.

Кроме того, корректность самих файлов рабочих потоков проверялась с помощью встроенных средств контроля целостности и связности потоков данных для KNIME. Результаты тестов методом черного ящика представлены в таблице ниже и могут быть признаны удовлетворительными и достаточными для положительного решения о готовности системы.

Табл. 6. Результаты функционального тестирования

№ теста	№ тестового набора	Корректность вычислений в подсистеме KNIME	Корректность обработки и представления в программной системе	Результат теста
1	1	Верно	Верно	пройден
2	2	Верно	Верно	пройден
3	3	Верно	Верно	пройден

Несистематические ошибки, связанные с качеством самих входных данных и их пригодностью для вычисления, не могут быть предусмотрены заранее и будут оценены нами в период опытной эксплуатации программной системы. Таким образом, функционально и технически разработанный программный продукт удовлетворяет предъявляемым требованиям и его можно допустить в эксплуатацию.

3.3. Оценка публикационного потенциала ученых Южно-Уральского государственного университета

В рамках апробации программной системы в части реализованных методик построения публикационного рейтинга и оценки публикационного потенциала рассмотрим результаты работы на примере Южно-Уральского государственного университета (табл. 7). Для этой цели в хранилище программной системы импортированы данные наукометрических профилей авторов, отнесенных Scopus к области исследования «computer science» и аффилированных с университетом на 1 июля 2017 г.

Табл. 7. Общая характеристика объекта исследования

Международное наименование университета	Общий QS-рейтинг за 2016	QS-research in 2016	Число авторов в области «computer science»
South Ural State University (SUSU)	151-200	Medium	228

Подготовительные этапы интеллектуального анализа, в соответствии с предложенной методикой, состоят в следующем.

1. Строится публикационный рейтинг авторов и на его основе выделяются четыре квартиля, затем авторы, входящие в каждый квартиль, получают определенную метку по следующему правилу: верхний квартиль соответствует метке «rank_Q1» (высокий уровень публикационной активности), нижний квартиль соответствует метке «rank_Q4» (базовый уровень публикационной активности), две центральных квартиля «rank_Q2» и «rank_Q3», соответственно. При этом авторы, занимающие одно место в рейтинге, получают одинаковую метку.

2. По каждому показателю, участвующему в публикационном рейтинге производится аналогичная дискретизация, а показатели приобретают метки вида «factor_Qx», где factor – имя показателя, а «x» – признак соответствующего квартиля, например, «top10_sum_Q1». Кроме того, минимальные значения показателя (чаще всего нуль) приобретают метку «Q4». Данное коррекционное правило приводит к некоторому искажению квартильной структуры по ряду показателей, но обеспечивает верную интерпретацию данных (табл. 8).

Табл. 8. Структура исходных данных в результате дискретизации

Наименование показателя	Доля авторов в квартиле, %			
	Q1	Q2	Q3	Q4
top10_sum	2,68	0,0	0,0	97,32
top25_sum	5,98	0,0	0,0	94,02
snip_mean	25,12	12,93	0,0	61,95
h_index	25,12	25,0	25,0	24,88
doc_count	25,12	25,0	25,0	24,88
cit_count	25,12	25,0	25,0	24,88
doc_per_auth	25,12	25,0	25,0	24,88
Публикационный рейтинг	23,78	24,15	25,12	26,95

Интеллектуальный анализ (в случае оценки публикационного потенциала – метод анализа потребительской корзины) выполнялся нами с помощью компонента вычислительного потока Association Rule Learner (Borgelt) для KNIME Analytics Platform.

Полученные ассоциативные правила имеют следующий вид (см. рис. 10). Для поиска правил анализируются наборы с не менее чем 5 элементами. Это, с одной стороны, гарантирует гибкость правил (поскольку соответствие 7 из 7 элементов – обеспечивает сам публикационный рейтинг), а с другой стороны – отсекает огромное множество заведомо недостоверных правил с малым числом элементов.

Поддержка набора (support)	Достоверность правила (confidence)	Значимость правила (lift)	Результирующая группа (consequent)	Исходный набор уровней показателей (antecedent)		
				h_index_Q3	snip_mean_Q1	..
5,12%	93,33%	392,47%	rank_Q1	<--		
5,61%	88,50%	371,99%	rank_Q1	<--		
...			

Рис. 10. Схема набора полученных ассоциативных правил

В целом программная система на имеющихся данных строит 4163 правила при минимальной поддержке и достоверности в 5,0 %. Выборка с надежностью не менее 50,0 % обнаруживает 78 правил. Однако, для перехода авторов в публикационные квартили рейтинга «Q1» и «Q2», а именно они нас в большей степени интересуют с точки зрения анализа публикационного потенциала, выделено 20 лучших правил, отвечающих заданным критериям достоверности. В соответствии с выделенными правилами 28 авторов могут быть потенциально (с разной степенью достоверности) переквалифицированы в лучший квартиль. Из этических соображений нами не публикуются ID авторов в системе Scopus, поэтому в таблице представлены групповые оценки (табл. 9).

Табл. 9. Публикационный потенциал авторов в области «computer science»

Исходный квартиль	Число авторов	Средняя поддержка правил по квартилю, %	Средняя достоверность правил по квартилю, %	Целевой квартиль (consequent)
Q2	13	6,0	90,2	Q1
Q3	48	9,0	65,8	Q2
Q4	4	7,6	60,2	Q2

Детализация оценок публикационного потенциала в разрезе отдельных авторов может стать основой для принятия управленческих решений, в частности о целевом стимулировании авторов. Таким образом, проведенная в настоящей работе оценка публикационного потенциала ученых Южно-Уральского государственного университета является самостоятельным научно-практическим результатом, пригодным для внедрения в деятельность университета.

ЗАКЛЮЧЕНИЕ

В ходе исследования показано, что в настоящее время целесообразна разработка программной системы, призванной помочь руководителям исследовательских групп, а также индивидуальным исследователям принимать решения по поводу улучшения публикационной активности, наукометрической результативности, определять слабые и сильные стороны, направления и конкретные решения в виде автоматизированных рекомендаций по повышению величины и репрезентативности наукометрических показателей. В рамках настоящей работы достигнуты следующие основные результаты.

1. Выполнен обзор научных работ и программных решений в области оценки публикационной активности ученых.

2. Разработана методика построения публикационного рейтинга на основе профилей ученых в Scopus.

3. Разработана программная система для формирования рекомендаций на основе интеллектуального анализа наукометрических данных.

4. Проведены эксперименты на данных профилей ученых ЮУрГУ.

Основные положения и результаты, представлены на 2 международных и 2 всероссийских научных конференциях, а также опубликованы в 7 изданиях, учитываемых в РИНЦ.

Таким образом, поставленная цель достигнута и задачи решены в полном объеме. В качестве направлений дальнейших исследований можно выделить следующие:

- адаптация методов и алгоритмов интеллектуального анализа для нереляционной модели данных, например, в графовом представлении;
- разработка архитектуры программной системы для noSQL-хранилища данных, например, графового хранилища, как наиболее адекватного для библиографических связей;
- расширение возможностей рекомендательной подсистемы на основе нейронной сети глубокого обучения.

СПИСОК ЛИТЕРАТУРЫ

1. Agrawal R., Srikant R. Fast Algorithms for Mining Association Rules / In Proc. of the 20th Int'l Conference on Very Large Databases, Santiago, Chile, September 1994.
2. Batista P.D., Campiteli M.G., Kinouchi O., et al. Is it possible to compare researchers with different scientific interests? // *Scientometrics*. – 2006. – Vol. 68(1). – P. 179-189. – DOI: 10.1007/s11192-006-0090-4
3. Blackert L. Ist in ger wissenschaftlichen information Platz fur die Informetrie? / L. Blackert, S. Siegel // *Wiss. Zeitschrift Tech. Hochschule*. – 1979. Bd 25, Hf 6. – P. 187-199.
4. Bradford S.C. *Documentation* / S.C. Bradford. – London: Crosly Lockwood, 1948. – 156 p.
5. Butler D. Computing giants launch free science metrics // *Nature*. – Vol 476. – 2011. – P. 18.
6. Borgelt C. Efficient Implementations of Apriori and Eclat // Workshop of Frequent Item Set Mining Implementations (FIMI 2003, Melbourne, FL, USA). – URL: http://www.borgelt.net/papers/fimi_03.pdf (дата обращения: 02.11.2017).
7. Cole F.J. The history of comparative anatomy. Part 1. A statistical analysis of the literature / F.J. Cole, N.B. Eales // *Science Progress*. – 1917. Vol. 10, 11. – P. 1229-1234.
8. Eck N. J. van, Waltman L. Generalizing the h- and g-indices // *Journal of Informetrics*. – 2008. – Vol. 2. – Iss. 4. – P. 263–271.
9. Garfield E. Citation indexes to science: a new dimension in documentation through association of ideas // *Science*. – 1955. – Vol. 122. – P. 108-111.
10. Garfield E., Sher I.H. New factors in the evaluation of scientific literature through citation indexing // *American Documentation*. – 1963. – Vol. 14. – No. 3. – P. 195–201.
11. Garfield E. The intended consequences of Robert K. Merton / E. Garfield // *Scientometrics*. – 2004. – Vol. 60. – P. 51-61.

12. Guns R., Rousseau R. Real and rational variants of the h-index and the g-index // *Journal of Informetrics*. – 2009. – Vol. 3. – P. 64-71.
13. Harzing A.W., Alakangas S. Publish or Perish // *Scientometrics*. – 2016. – Vol. 106, 787. – DOI:10.1007/s11192-015-1798-9
14. He Y., Hui S.C. Mining a Web Citation Database for author co-citation analysis // *Information Processing and Management*. – 2002. – Vol. 38. – P. 491-508.
15. Hirsch J.E. An index to quantify an individual's scientific research output // *Proc. Nat. Acad. Sci.* – 2005. – Vol. 102(46). – P. 16569-16572. – URL: <http://arxiv.org/abs/physics/0508025>
16. Hirsch J.E. An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship // *Scientometrics*. – 2010. – Vol. 85. – P. 741.
17. Koenigstein N., Dror G., Yehuda Koren Y. Yahoo! music recommendations: modeling music ratings with temporal dynamics and item taxonomy / *Proceedings of the fifth ACM conference on Recommender systems*. – ACM, 2011. – P. 165-172.
18. Lotka A.J. The frequency distribution of scientific productivity / A.J. Lotka // *J. Washington Acad. Sci.* – 1926. – Vol. 16. – P. 317-323.
19. Nacke O. Informetria: Name fur eine neue Disziplin / O. Nacke // *Nach. Dok.* – 1979. – Bd 30, Hf 6. – P. 219-226.
20. Padrós-Cuxart R., Riera-Quintero C., March-Mir F. Bibliometrics: a Publication Analysis Tool // *Proc. of the 3rd Workshop on Bibliometric-enhanced Information Retrieval (BIR 2016)*. – 2016. – P. 44-53.
21. Pindlowa W. Wokol informetrii, bibliometrii i naukometrii / W. Pindlowa // *Aktual. Probl. Inf. i dok.* – 1989. – Vol. 34. – P. 3-7.
22. Price D.J. de S. *Science since Babylon*. New Haven. – Yale University Press, 1961.
23. Rousseau R., Garcia-Zorita C., Sanz-Casado E. The h-bubble // *Journal of Informetrics*. – 2013. – Vol. 7. – P. 294-300.

24. Ruane F., Tol R. Rational (successive) h-indices: An application to economics in the Republic of Ireland // *Scientometrics*. – 2008. – Vol. 75. – P. 395-405.
25. Jin B. The AR-index: complementing the h-index / B. Jin, L. Liang, R. Rousseau, et al // *Chinese SCI bull.* – 2007. – Vol. 3(1). – P. 6-52. – DOI:10.1007/s11434-007-0145-9
26. Vieira E.S, Gomes J.A.N.F. Citations to Scientific Articles: Its Distribution and Dependence on the Article Features // *Journal of Informetrics*. – 2010. – Vol. 4. – No. 1. – P. 1-13. – DOI: 10.1016/j.joi.2009.06.002
27. Zipf G.K. Human behavior and the principle of least effort: an introduction to human ecology (1902-1950) / G.K. Zipf. – Cambridge: Addison-Wesley Press, 1949. – 573 p.
28. Алескеров Ф.Т., Катаева Е.С., Писляков В.В., Якуба В.И. Оценка вклада научных работников методом порогового агрегирования // *Управление большими системами*. – 2013. – № 44. – С. 172-189.
29. Алескеров Ф.Т., Якуба В.И. Метод порогового агрегирования трехградационных ранжировок // *ДАН*. – 2007. – Т. 413. – № 2. – С. 181-183.
30. Бедный Б.И., Миронос А.А., Серова Т.В. Методика оценки эффективности подготовки научных кадров в аспирантуре // *Вестник ННГУ*. – 2010. – № 5-1. – С.11-19.
31. Бедный Б.И., Сорокин, Ю.М. О показателях научного цитирования и их применении // *Высшее образование в России*. – 2012. – № 3. – С. 17-28.
32. Вальден П.И. О развитии химии в России // *Дневники 2 Менделеевского съезда (21-28 дек. 1911 г.)*. – СПб.: Б.г. – № 4-8. – С. 124-141.
33. Ван Вухт Ф., Вестерхайден Д. Многомерное ранжирование: новый инструмент прозрачности в области высшего образования // *Вестник международных организаций: образование, наука, новая экономика*. – 2012. – №1. – С.9-33.

34. Игра в цифирь, или как теперь оценивают труд ученого (сборник статей по библиометрике). – М.: МЦНМО, 2011. – 72 с.

35. Ильясов Б.Г., Карамзина А.Г., Фазлетдинова Ю.Р. Моделирование системных показателей оценки эффективности научных школ // Программные продукты и системы. – 2015. – № 2 (110). – С. 5-12.

36. Луценко Е.В., Орлов А.И., Глухов В.А. Наукометрическая интеллектуальная измерительная система по данным РИНЦ на основе АСК-анализа и системы "Эйдос" // Научный журнал КубГАУ. – 2016. – № 122(08). – С. 1-56.

37. Луценко Е.В., Орлов А.И. Количественная оценка степени манипулирования индексом Хирша и его модификация, устойчивая к манипулированию [Электронный ресурс] // Научный журнал КубГАУ. – 2016. – № 121(07). – URL: <http://ej.kubagro.ru/2016/07/pdf/05.pdf> (дата обращения: 02.11.2017).

38. Марвин С.В. Альтернативная дробная модификация индекса Хирша, учитывающая количество авторов цитируемых статей // УБС. – 2015. – № 56. – С. 108-122.

39. Мбого И.А., Прокудин Д.Е., Чугунов А.В. Разработка инструментов интеграции научной информации в пространстве разнородных информационных систем // Научный сервис в сети Интернет: труды XVIII Всероссийской научной конференции (19-24 сентября 2016 г., г. Новороссийск). – М.: ИПМ им. М.В. Келдыша, 2016. – С. 249-258. – DOI:10.20948/abrau-2016-44

40. Мбого И.А., Прокудин Д.Е. Подходы к развитию инструментов автоматизации и интеграции ресурсов информационного пространства поддержки междисциплинарного научного направления // Сборник научных статей XVIII Объединенной конференции «Интернет и современное общество» IMS-2015 (23-25 июня 2015 г.). – Санкт-Петербург: 2015. – С. 290-302.

41. Мирский Э.М. Наукометрия // Новая философская энциклопедия / Ин-т философии РАН; Нац. обществ.-науч. фонд; Предс. научно-ред. совета В. С. Степин, заместители предс.: А. А. Гусейнов, Г. Ю. Семигин, уч. секр. А. П. Огурцов. – М.: Мысль, 2010. – ISBN 978-5-244-01115-9.

42. Модуль агрегации наукометрических данных открытых сервисов в сети Интернет / Свидетельство о государственной регистрации программы для ЭВМ № 2016619028 от 11 августа 2016 г. «Модуль агрегации наукометрических данных открытых сервисов в сети Интернет» / Д.В. Валько, А.С. Колташев // Программы для ЭВМ. Базы данных. Топологии интегральных микросхем: Официальный бюллетень Федеральной службы по интеллектуальной собственности (РОСПАТЕНТ). – № 9. – 2016. (дата опубликования 20.09.2016).

43. Москалева О.В. Можно ли оценивать труд ученых по библиометрическим показателям? // УБС. – 2013. – № 44. – С. 308-331.

44. Налимов В.В. Наукометрия: Изучение развития науки как информационного процесса / В.В. Налимов, З.М. Мульченко. – М.: Наука, 1969. – 192 с.

45. Неустроев С.С., Предыбайло В.А. Совершенствование внутренней системы оценивания качества научных исследований: от качества процесса к качеству результатов // Управление образованием: теория и практика. – 2016. – № 2 (22). – С. 5-14.

46. Никитушкин В.Г., Германов Г.Н., Корольков А.Н. Рейтинг институтов МГПУ по данным публикационной активности сотрудников // Ученые записки университета Лесгафта. – 2015. – № 6 (124). – С. 148-155.

47. Николенко В.Н., Вялков А.И., Мартынчик С.А., Глухова Е.А. Подходы к оценке эффективности и способы стимулирования публикационной активности в крупном медицинском вузе // Высшее образование в России. – 2014. – № 10. – С. 18-25.

48. Об утверждении Стратегии инновационного развития Российской Федерации на период до 2020 года: распоряжение Правительства РФ

от 08.12.2011 №2227-р [Электронный ресурс]. – URL: base.consultant.ru/cons/cgi/online.cgi?req=doc; base=LAW;n=123444 (дата обращения: 22.09.2017)

49. О мерах по реализации государственной политики в области образования и науки: Указ Президента Российской Федерации от 7 мая 2012 г. №599 // Российская газета. – 2012. – 7 мая [Электронный ресурс]. – URL: www.rg.ru/2012/05/09/nauka-dok.html (дата обращения: 22.09.2017)

50. Питерс Д., Марш Р. Rate my research dot com: измеряем то, что ценим, ценим, что измеряем // Научная периодика: проблемы и решения. – 2011. – № 1. – С. 40-45.

51. Письмо Минобрнауки РФ от 10 марта 2015 г. № АК-571/05 «О проведении мониторинга эффективности образовательных организаций высшего образования в 2015 году». – URL: <http://минобрнауки.рф/документы/5267> (дата обращения: 02.11.2017).

52. Положение о государственной аккредитации образовательной деятельности [Утверждено постановлением Правительства РФ от 18.11.2013 № 1039 (в ред. от 26.12.2014 г.)] // Справочно-правовая система «КонсультантПлюс».

53. Положение о лицензировании образовательной деятельности [Утверждено постановлением Правительства Российской Федерации от 28 октября 2013 г. № 966 (в ред. от 27.11.2014 г.)] // Справочно-правовая система «КонсультантПлюс».

54. Поляк Б.Т. Наукометрия: кого мы лечим? // УБС. – 2013. – № 44. – С. 161-170.

55. Прайс Д. Малая наука, большая наука // Наука о науке, М.: Изд-во «Прогресс», 1966.

56. Разин М. П. К вопросам наукометрии... и не только // Вятский медицинский вестник. – 2013. – № 2. – С. 44-47.

57. Разин М.П. Недостатки индекса Хирша в медицинской науке и возможные пути их преодоления // Детская хирургия. – 2013. – № 3. – С. 58-59.
58. Редькина Н.С. Формализованные методы анализа документальных информационных потоков // Библиосфера. – 2005. – № 2. – С. 51-59.
59. Руководство по наукометрии: индикаторы развития науки и технологии: [монография] / М.А. Акоев, В.А. Маркусова, О.В. Москалева, В.В. Писляков; [под. ред. М.А. Акоева]. – Екатеринбург: Изд-во Урал. ун-та, 2014. – 250 с.
60. Сеницын А.А., Никифоров О.Ю., Андреев М.А. Концепция и структура информационно-аналитической системы анализа публикационной активности сотрудников научно-образовательной организации // Фундаментальные исследования. – 2014. – № 11. – С. 1276-1280.
61. Сеницын А.А., Никифоров О.Ю., Андреев М.А. Особенности применения информационно-аналитической системы для оценки направления поддержки по созданию результатов интеллектуальной деятельности научно-образовательной организации // Фундаментальные исследования. – 2014. – № 11-6. – С. 1271-1275.
62. Солтон Дж. Динамические библиотечно-информационные системы. – Пер. с англ. В.Р. Хисамутдинова. – М.: Мир, 1979. – 557 с.
63. Тарасевич Ю.Ю., Шиняева Т.С. Временная динамика индекса Хирша // Вестник ЮУрГУ. Серия: Математическое моделирование и программирование. – 2016. – № 1 – С. 32-45.
64. Третьякова О.В., Кабакова Е.А. Возможности и перспективы использования индексов цитирования в оценке результатов деятельности научного учреждения // Экономические и социальные перемены: факты, тенденции, прогноз. – 2013. – № 6 (30). – С. 189-200.
65. Шарабчиев Ю.Т. Использование наукометрических методов для мониторинга продуктивности научной деятельности // Медицинские новости. – 2013. – №6 (225). – С. 13-19.

66. Штовба С.Д., Штовба Е.В. Обзор наукометрических показателей для оценки публикационной деятельности ученого // УБС. – 2013. – № 44. – С. 262-278.

67. Шторх А. Систематическое обозрение литературы в России (1801-1806 гг.) / А. Шторх, Ф. Аделунг. – СПб., 1810. Ч.1. – 352 с.

68. Яблонский А.И. Модели и методы исследования науки. – М.: Эдиториал УРСС, – 2001. – 400 с.