АКУСТИЧЕСКИЙ ЖУРНАЛ, 2020, том 66, № 1, с. 104–116

ИНФОРМАЦИЯ

УДК 001.92;002.6;004.91

РУССКОЯЗЫЧНОЕ НАПРАВЛЕНИЕ РАБОТЫ РОССИЙСКИХ ИНФОРМАЦИОННЫХ СЛУЖБ

© 2020 г. В. Г. Шамаев[a], *, А. Б. Горшков[b]

[a]Московский государственный университет им. М.В. Ломоносова, физический ф-т, Ленинские горы, ГСП-1, Москва, 119991 Россия

[b]Московский государственный университет им. М.В. Ломоносова, Государственный астрономический ин-т им. П.К. Штернберга, Москва Россия

*e-mail: shamaev08@gmail.com

Поступила в редакцию 26.06.2019 г.
После доработки 26.06.2019 г.
Принята к публикации 09.07.2019 г.

Рассматривается проблема отражения научных печатных и электронных источников информации в информационном поле, которое сейчас тесно связано с Интернетом. В настоящее время можно говорить только о первом этапе агрегирования основных русскоязычных научных электронных ресурсов. Рассматриваются политематические ресурсы, среди них Банк данных ВИНИТИ РАН, Научная электронная библиотека, “Истина” Московского государственного университета им. М.В. Ломоносова, Scopus и несколько тематических ресурсов, которые имеют большое будущее даже в условиях ограничения доступных их создателям материальных ресурсов. Обращается внимание на необходимость создания государственной наукометрической системы для объективной оценки научной работы как учреждений, так и отдельных научных коллективов и их сотрудников. Отдельным блоком на примере архива “Акустического журнала” описывается технология наложения текстового слоя на изображения статей журнала, количество которых достигает 10 000. Это может быть полезным при оцифровке ретроспективных печатных изданий.

Ключевые слова: русскоязычные научные ресурсы, национальный индекс цитирования, импакт-фактор русскоязычных журналов, тематические информационные продукты, технология полнотекстовой оцифровки научных журналов

DOI: 10.31857/S0320791919060157

ВВЕДЕНИЕ

Одной из важных задач на нынешнем этапе развития информационных технологий и выпуска информационных продуктов является получение полной информации о русскоязычных публикациях как в нашей стране, так и в окружающем нас русскоязычном пространстве. В консолидированном виде информации об этих публикациях нет ни у нас, разве что за последние пару десятков лет в Научной электронной библиотеке, ни тем более за рубежом. Для западного читателя, не владеющего русским языком, их как бы и не существует. Отсюда низкий уровень получаемого по данным Web of Science (WoS) или Scopus значения импакт-факторов почти всех русскоязычных журналов, а соответственно, и индекса цитирования наших научных сотрудников, которые в подавляющей массе пишут в наши журналы и, конечно, на русском языке. По этой же причине многие публикации ученых, пишущих на русском языке, остаются неизвестными на западе, и их исследования в лучшем случае воспроизводятся заново, а в худшем – публикуются на английском языке с другими авторами. Началось это не сейчас (см. http://www.akzh.ru/pdf/1978_1_156.pdf), но в последнее время приняло массовый характер.

ЗАДАЧИ РАЗВИТИЯ РУССКОЯЗЫЧНОГО НАПРАВЛЕНИЯ

В недавней статье и.о. директора ВИНИТИ РАН Ю.Н. Щуко в журнале “Научно-техническая информация. Сер. 1. Организация и методика информ. работы” об аспектах развития института проведен краткий анализ деятельности ВИНИТИ предшествующего периода и сформулированы задачи текущего [1]. Автор называет их тактическими. Среди них мы выделим очевидные, касающиеся темы уже этой статьи и которые давно необходимо было решить:

1. Провести оценку текущего состояния Банка данных ВИНИТИ (БнД ВИНИТИ) как основно-

Таблица 1. Количество научных журналов по различным тематикам в Scopus

Тематика	Всего журналов по тематике	Российские журналы
Биология	1903	26
Математика	1272	20
Информатика. Компьютерные науки	1378	7
Химия	802	28
Технические науки	2338	24
Физика. Механика. Астрономия	992	43

го на сегодняшний день информационного ресурса в стране, претендующего на роль национального.

Ранее, до начала 1990-х гг., таким национальным ресурсом был Реферативный журнал ВИНИТИ (РЖ) и, как одна из его тематических частей, – выпуск РЖ “Акустика”. Другие информационные продукты ВИНИТИ – как “Экспресс-информация”, “Сигнальная информация”, да и нынешний Банк данных – даже во времена СССР не только не дотягивали до уровня национального продукта, но и никогда не рассматривались в этом качестве даже самыми горячими головами. Разве что “Итоги науки и техники” были, так же как Реферативный журнал, достаточно известны. За последние четверть века Итоги, как и Экспресс-, и Сигнальная информация, практически закончили свое существование [1–3].

2. Сформулировать пути решения задач по оценке перспективных направлений развития науки и техники и традиционного информационного обеспечения научных исследований.

Мы бы добавили в этот список еще и:

3. Обеспечение в БнД ВИНИТИ полного покрытия русскоязычных публикаций.

4. Создание национального индекса цитирования и ежегодная публикация импакт-факторов русскоязычных журналов.

5. Формирование тематических и проблемно-ориентированных информационных ресурсов и продуктов и создание на их основе информационно-поисковых систем. Разработка механизма взаимодействия между ними. И, в первую очередь, связь между их рубрикаторами и ключевыми словами.

6. Разработка единой технологии полнотекстовой оцифровки научных журналов с выкладкой их в Интернете. Нами предполагается, что эту роль мог бы выполнять ВИНИТИ РАН как наиболее подготовленный к этой работе информационный центр.

Вышеприведенные пункты расположены не в порядке их важности – они все важны, так же как и еще три пункта, которые есть в вышеуказанной статье, но которые мы не приводим. Необходима параллельная и оперативная работа по всем этим направлениям. Слишком много потеряно времени.

Если с первым пунктом все ясно, то с оценкой перспективных направлений и традиционным информационным обеспечением научных исследований все гораздо сложнее. Основываясь на нашем опыте работы в ВИНИТИ, можно предположить, что выполнение этого пункта возможно только с привлечением ведущих специалистов из научных учреждений РАН, Минобрнауки и других профильных специалистов по каждой тематической области.

Далее в этом же номере журнала следует интересная статья Р.С. Гиляревского и Е.В. Мельниковой “О разработке концепции государственной наукометрической системы...” [4], которая хорошо коррелирует с предыдущей статьей журнала [1] и добавленными нами пунктами. Это, вместе с изложенным в [5] направлением развития БнД ВИНИТИ, как нам кажется, является началом реализации в ВИНИТИ основных задач в области информации о русскоязычных исследованиях, публикуемых в русскоязычных научных изданиях и, тем самым, весомым подспорьем для научных и технических работников.

К этому мы еще вернемся, а сейчас отметим, что в статье Р.С. Гиляревского и Е.В. Мельниковой четко формулируется, что такое наукометрическая система и цель ее создания на уровне государства. Отмечается то, что и нас беспокоит [6], а именно, что повсеместное применение публикуемых материалов Web of Science и Scopus для оценки работы наших научных учреждений и отдельных научных работников создает проблемы. Одна из них заключается в том, что в этих информационно-поисковых системах (их базах данных) недостаточно отражаются русскоязычные публикации. Сошлемся здесь на доклад О.В. Кирилловой на конференции РИНЦ в Австрии (см. табл. 1) [7].

Отметим также, что в настоящее время большинство научных журналов распределено по квартилям – категориям научных журналов, каждую из которых определяют библиометрические показатели, отражающие уровень цитируемости журнала. Всего квартилей четыре, начиная от Q1 (самый высокий, к которому принадлежат наиболее авторитетные журналы) до Q4 (самый низкий). Предполагается, что система квартилей позволяет наиболее объективно оценить качество – уровень журнала вне зависимости от предметной области.

Приведем суммарное распределение по квартилям российских журналов, обрабатываемых в Scopus (см. табл. 2) [7].

И здесь, как и количество отражаемых в Scopus, так и качество наших научных журналов не является удовлетворительным.

Еще отметим, что в статье [4] делается, на наш взгляд, правильный вывод: “Наукометрический анализ нельзя строить на базах данных только двух–трех мировых корпораций, владеющих системами индексирования и цитирования. Это может привести к полной монополизации мирового рынка научно-информационных услуг и искажению результирующей информации в интересах монополиста”. Мы это и видим на приведенном примере одного такого монополиста как Scopus, который практически полностью игнорирует русскоязычные журналы. Подобный же вывод делается в работе И.В. Зибаревой [8] (см. рис. 1).

Таблица 2. Распределение российских журналов по квартилям и тематическим областям в Scopus

Тематика	Q1	Q2	Q3	Q4
Биология	0	2	10	14
Математика	1	3	11	5
Информатика. Компьютерные науки	0	1	6	0
Химия	1	1	17	9
Технические науки	0	6	12	6
Физика. Механика. Астрономия	1	10	25	7

ПУТИ РАЗВИТИЯ РУССКОЯЗЫЧНОГО НАПРАВЛЕНИЯ

Политематические информационные ресурсы

Как мы видим, потребность в заполнении научной ниши с русскоязычными публикациями велика, в том числе и в Интернете. Именно поэтому, на наш взгляд, важной работой является актуализация Банка данных ВИНИТИ с упором на этот аспект. Мы имеем в виду как распространение его русскоязычной части в ретроспективную область, так и дополнение документами, изъятыми в процессе переработки по разным причинам из полных журнальных комплектов. Отметим, что в последнее десятилетие наполнение БнД ВИНИТИ характеризуется увеличением доли статей из периодических изданий в общем потоке его наполнения. Так, для разных тематических фрагментов этот показатель варьируется от 52 до 98% [1]. Хорошо ли это, так как при этом, видимо, игнорируются книжные издания, а также труды конференций, семинаров и т.д.

В то же время в БнД ВИНИТИ стало увеличиваться процентное соотношение русскоязычных работ. Такой вывод можно сделать, анализируя табл. 3. В 1990 г., собственно, последнем году “советского” ВИНИТИ, ситуация была иной, что и видно из таблицы. Проценты дают нам, надо сказать, цифры условные, т. к. политика ВИНИТИ с

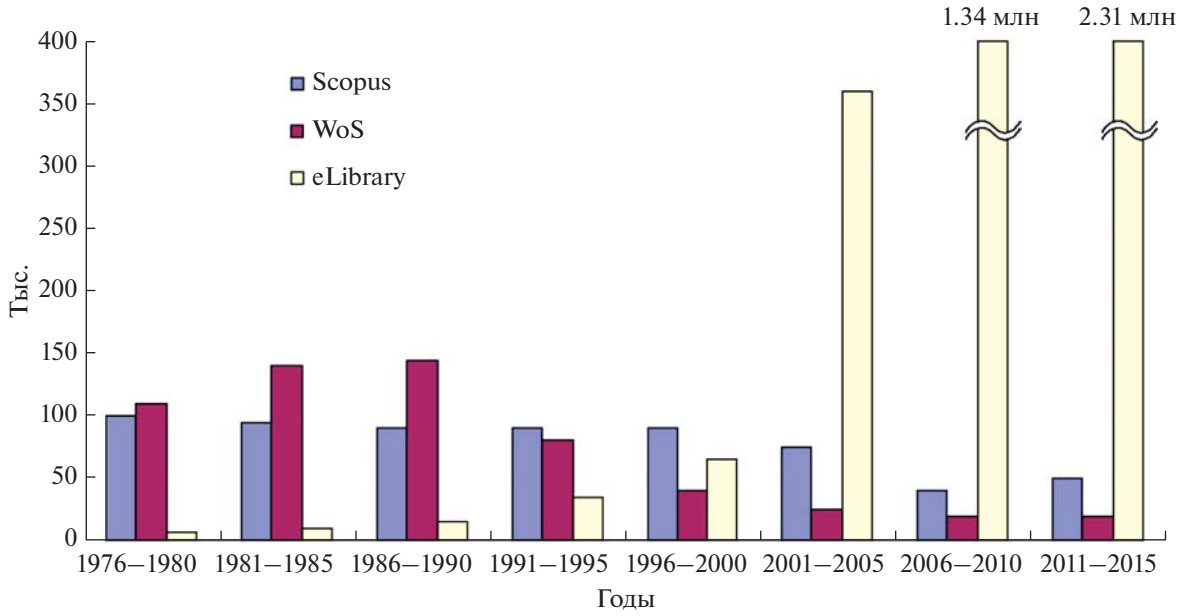

Рис. 1. Количество русскоязычных публикаций в различных информационных системах в период 1976–2015 гг. (для 2006–2010 и 2011–2015 гг. по данным Научной электронной библиотеки – 1.34 и 2.31 млн соответственно).

Таблица 3. Количество документов на русском языке в некоторых тематических фрагментах БнД ВИНИТИ

Тематика	Кол-во русскоязычных документов в % к общему кол-ву			
	1990	2010	2017	2018
АиРЭ	21.5	41.8	53.2	54.0
Астрономия	27.1	27.8	42.5	46.5
Геология	46.5	57.4	56.5	63.6
Математика	–	38.3	36.9	37.1
Машиностроение	32.5	36.5	48.0	47.4
Механика	40.4	42.3	38.8	41.8
Физика	22.6	11.8	14.0	16.2
Химия	27.8	24.5	29.7	30.1

1990-х гг. заключалась в уменьшении плановых показателей, чтобы казалось, что выполняется 100% плана (см. табл. 1 и 2 в [1]). Если же рассматривать цифры в абсолютном значении, то ситуация выглядит, наверное, удручающе, но не совсем безнадежно для русскоязычных публикаций, по крайней мере для большинства из приведенных тематических фрагментов. К сожалению, по физике, а следовательно, и акустике число отраженных русскоязычных публикаций уменьшается, несмотря на увеличение их реального количества, что можно посмотреть в информационно-поисковой системе “Акустика” [9].

Какова же на самом деле в БнД ВИНИТИ ситуация в области отражения акустических исследований – она приведена в [10], и в связи с этим еще в 2012 г. нами начата работа, которая названа “Акустика. Русскоязычные источники”. В рамках этого проекта создана информационно-аналитическая система, позволяющая собрать в одном месте публикации акустической тематики и оценивать как количественную, так и качественную сторону научной деятельности русскоязычного сегмента информационного поля в области акустики. Ведь главное сейчас – это представить в наиболее полном объеме русскоязычные публикации и довести их до пользователя – научного работника. Реализацией портала “Акустика. Русскоязычные источники” мы решаем в области акустики третью задачу из выделенных в начале статьи – задачу обеспечения полного покрытия русскоязычных документов.

Подробное обсуждение пути решения четвертой задачи, а именно создания национального индекса цитирования и получения импакт-факторов русскоязычных журналов, проведено в статье [4]. Выполнение пятой задачи, сформулированной нами, но и затронутой в той же статье, требует пояснения. Поэтому мы остановимся на функциональной структуре системы формирования тематических и проблемно-ориентированных информационных ресурсов. На наш взгляд, их список в статье [4] выглядит внушительным, но далеко не однородным. И это понятно, координации между ресурсами, о необходимости которой говорится в этой статье, нет и в помине [10]. Каждый создает свой ресурс, не обращая внимания на остальных. О БнД ВИНИТИ и его претензии на звание национального ресурса мы рассказали ранее. Приведем пример с ресурсом Киберленинка (https://cyberleninka.ru), преподносимом его создателями как “научная электронная библиотека, построенная на парадигме открытой науки (Open Science)”. Ресурс в предлагаемом виде очень слабый. Нет приемлемого интерфейса для пользователя, поисковые возможности многоступенчатые с использованием "древовидной” структуры, что является весьма архаичным. Выдачи скупые в ответ на запрос и избыточные по обильно появляющейся на экране сторонней информации. Даже неоднократно критикуемые нами ресурсы интернета по диссертациям (https://www.dissercat.com, https://www.twirpx.com или http://www.dslib.net) гораздо удобнее в использовании, хотя и сильно загрязнены в результате плохой работы корректоров [11]. Или корректуры совсем нет, что ближе к истине. Но, чтобы уж совсем не впасть в пессимизм, отметим прекрасный реферат в Киберленинке Е.Г. Гребенщиковой нашей статьи “Навигация по русскоязычным источникам научной информации” из “Вестника РАН” [12].

Неплохим примером среди политематических ресурсов выступает Научная электронная библиотека (https://elibrary.ru). Ресурс с хорошим интерфейсом в плане показа информации по конкретному журналу (рис. 2) и огромен по объему, собранному за относительно небольшой промежуток времени. К сожалению, сам поиск в этой системе крайне перегружен (рис. 3), да и выдача его результатов не кажется нам продуманной.

Рис. 2. Интерфейс “Акустического журнала” в Научной электронной библиотеке (Дата обращения 28.06.2019).

Примерно с 2012 г. мы наблюдаем за развитием портала “Истина” Московского государственного университета им. М.В. Ломоносова (Интеллектуальная Система Тематического Исследования НАукометрических данных – https://istina.msu.ru/) [13], предназначенного для учета и анализа научной деятельности сотрудников, а также в помощь научным сотрудникам. На этот портал обращают внимание и авторы статьи [14], и пользователи, работающие за рубежом. “Истина”, кроме статистических данных для отчета учебно-научных подразделений МГУ, дает возможность и самим пользователям вести систематический учет своей деятельности, к тому же система внесения записей слегка формализована, что облегчает работу. Тем самым, в эту информационную систему введен элемент унификации. Но у “Истины” есть существенный изъян, заключающийся в том, что сведения по публикациям вносят сами пользователи. В результате встречается много ошибок во всех элементах библиографии, начиная с названия работ, занесения данных по авторам, присутствия многих вариантов написания источников. Встречаются многочисленные ошибки в выходных данных как номеров выпусков, так и страниц публикаций и т.д. “Истине”, на наш взгляд, не хватает экспертной группы, а точнее редакторов, которые бы осуществляли контроль корректности вносимых данных. Поиск в системе также примитивен, впрочем, такой задачи авторы разработки, видимо, и не ставили. Также в системе слишком большое внимание уделяется публикациям в престижных журналах [15].

Приведенные ресурсы политематические, что привлекает к ним внимание многих пользователей, да и при поиске в Интернете, как правило, они появляются в выдаче в числе первых.

Тематические информационные ресурсы

Из тематических информационно-поисковых систем, имеющих в основе хорошую базу данных русскоязычных публикаций, отметим следующие:

1. общероссийский математический портал Math-Net.Ru (рис. 4), который создан и развивается Математическим институтом им. В.А. Стеклова РАН совместно с Отделением математических наук РАН;

2. портал “Акустика. Русскоязычные источники” AkData.Ru (рис. 5), создан и развивается на кафедре акустики физического факультета МГУ им. М.В. Ломоносова совместно с редакцией “Акустического журнала”;

3. портал издательства Сибирского отделения РАН с полнотекстовыми архивами 24-х журналов (http://sibran.ru/journals);

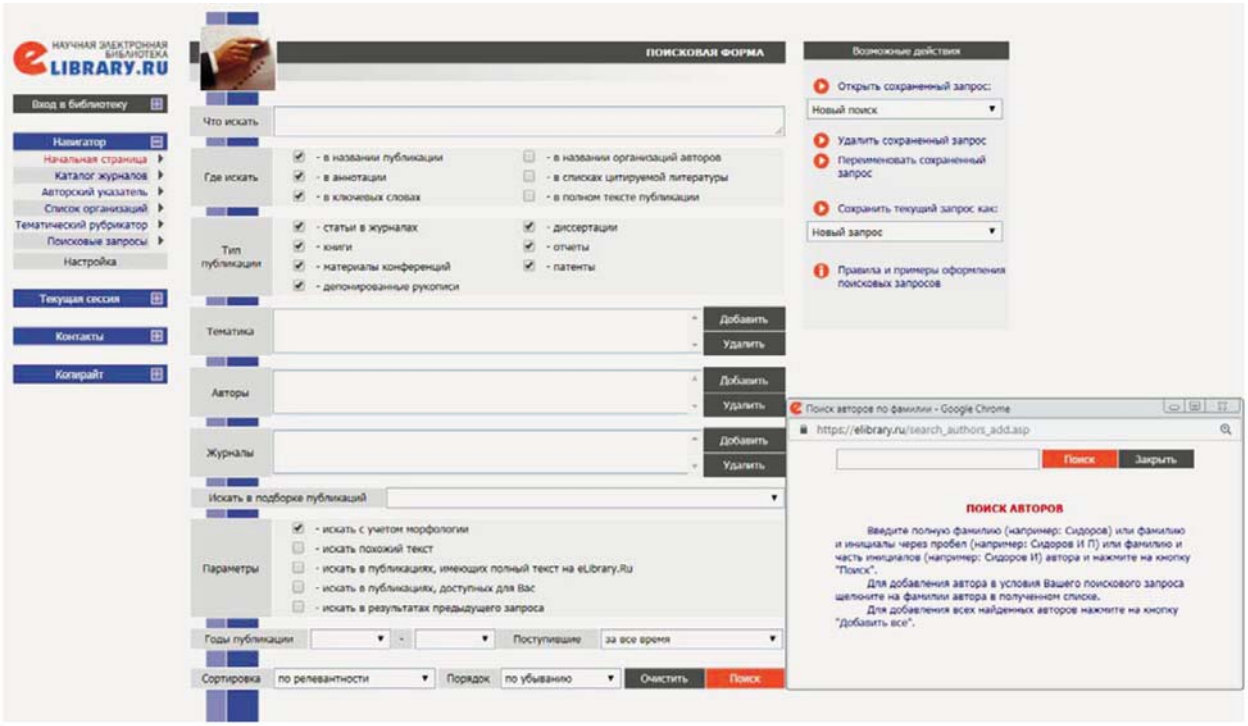

Рис. 3. Интерфейс поискового запроса в Научной электронной библиотеке.

4. а также портал журнала “Успехи физических наук” как пример правильной и законченной системы представления научного журнала в Интернете (https://ufn.ru).

Почему именно эти? Мы считаем, что они должны быть в первую очередь взяты в рассмотрение при реализации предложений по пятому пункту – “формирование тематических и проблемно-ориентированных информационных продуктов и разработка единой технологии полнотекстовой оцифровки научных журналов”.

Общероссийский математический портал Math-Net.Ru, что следует из названия, посвящен математике, лишь с небольшим вкраплением физических журналов. Он прекрасно организован, имеет продуманную структуру и удобный интерфейс, обладает хорошим наполнением и поисковыми возможностями. Тут же можно прочитать тексты большинства статей. Авторы проекта подумали даже о такой необходимой функции, как поиск по пристатейным спискам литературы, которая раньше нам особенно не встречалась, а является крайне востребованной. Такой поиск является дополнительной функцией получения полезной информации, функцией продуманной и тщательно выполненной авторами разработки. На основе своей базы данных математических работ составляется импакт-фактор журналов и авторский индекс цитирования [16]. Отмечается, что “далеко не все ссылки из вполне достойных журналов попадают в числитель IF ISI (Impact factor, Institute for Scientific Information)”, т.е. WoS занижает импакт-факторы российских журналов, а следовательно, и индексы цитирования наших авторов [17]. На это же мы обратили внимание при анализе отраженной в WoS информации по “Акустическому журналу”, который на западе выходит под названием “Acoustical Physics”.

На портале Информационной системы “Акустика. Русскоязычные источники” AkData.Ru обрабатывается около 800 журналов, и наполнение его БД составляет более 55 тыс. статей [9]. Портал использует рубрикатор, составленный на основе PACS, и является тематическим, что означает выборку статей акустической тематики из журналов. На нем мы видим улучшение информативности вследствие как рубрицирования каждой статьи, так и возможности поиска по пяти параметрам, включая и рубрики статей. Ключевые параметры – источники, авторы, рубрики – связаны гиперссылками. На портале также помещен полнотекстовый архив “Акустического журнала” с момента его организации в 1955 г. и присутствует “Сигнальная информация” по акустике, которая выходит шесть раз в год.

Таким образом, пятый пункт (задача) по подготовки тематических продуктов из списка, сформу-

Рис. 4. Портал Math-Net.Ru (Дата обращения 03.07.2019).

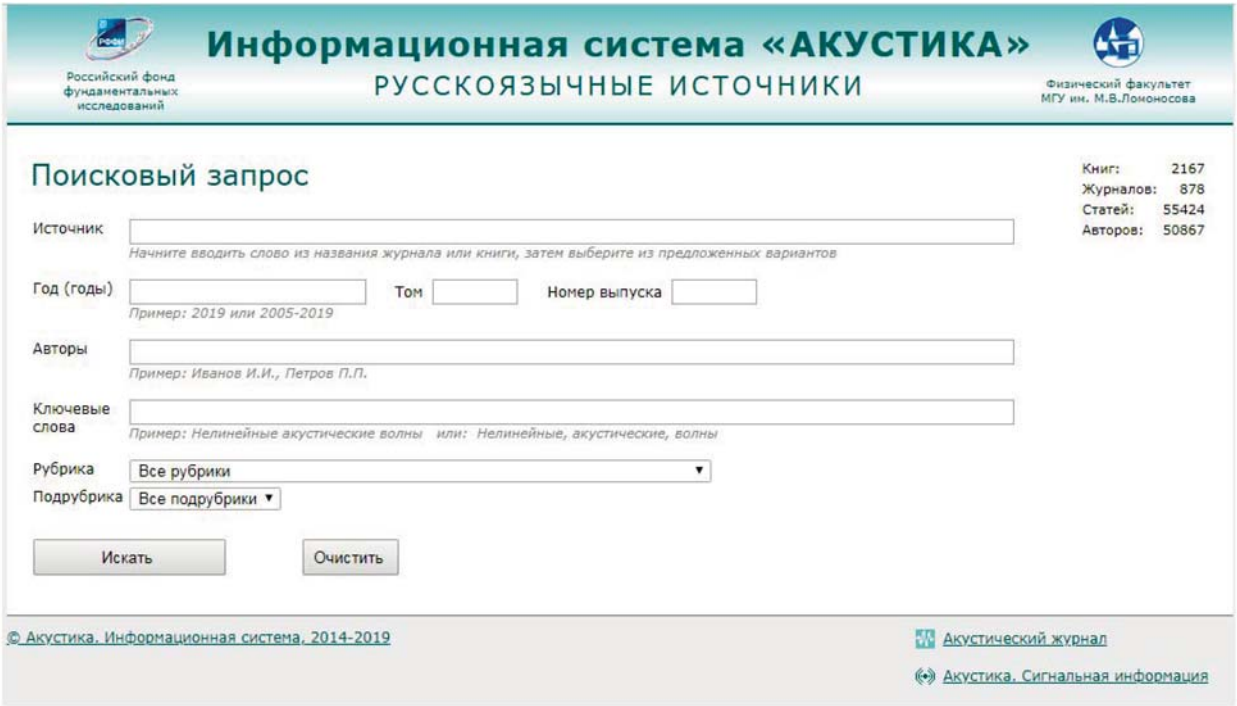

Рис. 5. Портал AkData.Ru (Дата обращения 03.07.2019).

лированного в начале статьи, имеет хороший задел по технологии и способам ее реализации.

Перейдем теперь к последнему пункту – разработке единой технологии полнотекстовой оцифровки научных журналов. В этой области тоже есть хорошо разработанные технологии, например, как использованная при представлении в Интернете журнала “Успехи физических наук”, так и технология подготовки близкого нам сайта “Акустического журнала”. Есть еще технологии вывода в Интернет журналов издательства Сибирского отделения РАН и других, например, журналов Физико-технического института им. А.Ф. Иоффе РАН. Но мы бы остановились на первых двух, т.к. сайты последних, несмотря на хорошее выполнение, являются лишь интернет-проекциями печатных изданий. Не будем все это описывать, каждый может сам посмотреть. Отметим лишь, что сайты интересных нам журналов представляют собой информационные системы с большими поисковыми возможностями, и как раз на их основе может разрабатываться технология полнотекстовой оцифровки ретроспективных номеров печатных научных журналов, например, на базе ВИНИТИ как организации РАН. А лучше всего привлечь самих авторов технологий к этой работе.

Нами же выполнена подобная работа по “Акустическому журналу”, о фрагменте технологии которой в части реализации наложения текстового слоя мы кратко и расскажем.

ФОРМИРОВАНИЕ ПОЛНОТЕКСТОВОГО АРХИВА “АКУСТИЧЕСКОГО ЖУРНАЛА”

Электронный полнотекстовый архив “Акустического журнала” (АкЖ) представляет собой набор PDF-файлов с изображениями страниц журнала (каждый файл содержит отдельную статью), размещенный на интернет-сайте архива akzh.ru. Сайт предоставляет пользователю доступ к статьям журнала посредством списка статей по каждому выпуску журнала (с резюме), а также по рубрикатору и авторскому указателю [18].

На середину 2019 г. архив содержал более 350 выпусков АкЖ в ~10000 файлах формата PDF общим объемом порядка 50000 страниц или 20 Гб.

Изначально архив содержал PDF-файлы без текстового слоя, что не позволяло поисковым системам (Google, Yandex и т.п.) производить индексацию этих файлов и ограничивало возможности поиска нужной статьи в архиве только ее названием, списком авторов и резюме. Возможность поиска по текстам статей отсутствовала. Для расширения комфортности работы с архивом было принято решение добавить в PDF-файлы архива текстовый слой.

При создании текстового слоя использовалась программа ABBY FineReader 12 Professional как дающая на наш взгляд наиболее качественный результат. В то же время, эта версия программы не поддерживает пакетную обработку файлов и не воспринимает параметры командной строки. Чтобы избежать необходимости запускать вручную обработку каждой из порядка 10 000 статей, в функциональность программы Pub2Site (программа нашей разработки, создающая HTML-файлы сайта архива АкЖ на основе информации в базе данных) были добавлены функции “сборки” группы PDF-файлов в один документ и, соответственно, “разборки” этого сборного документа на исходные файлы. Добавленные функции использовали для непосредственной обработки PDF-файлов свободно распространяемый пакет PDFtk Free (https://www.pdflabs.com/tools/pdftk-the-pdf-toolkit/).

Алгоритм работы получился следующий:

– Программа Pub2Site. “Сборка” выбранной группы PDF-файлов в один сборный PDF-файл, создание “журнала” с информацией об имени каждого исходного файла и его позиции в сборном файле.

– Программа FineReader. Добавление текстового слоя в сборный PDF-файл.

– Pub2Site. “Разборка” полученного PDF-файла на файлы отдельных статей.

Опишем подробнее.

В связи с ограничениями, связанными с возможностями компьютера и отдельных программ, данный алгоритм применялся последовательно к ограниченным порциям архива объемом примерно 0.5 Гб каждая.

Объединение PDF-файлов статей в один сборный PDF-файл:

1. Поместить файлы, предназначенные для сборки, в отдельную папку.

2. Запустить программу Pub2Site, открыть вкладку “Нарезка PDF”. В поле “Путь к PDF выпусков” указать адрес папки.

3. Нажать кнопку “Собрать PDF” (рис. 6).

По данной команде программа Pub2Site:

– составляет список PDF-файлов в папке;

– создает подпапку “sborka”;

– последовательно (по одному) добавляет PDF-файлы в сборный PDF-файл, записывая в отдельный текстовый файл (журнал) имя исходного файла и его положение в сборном файле (начальная и конечная страницы).

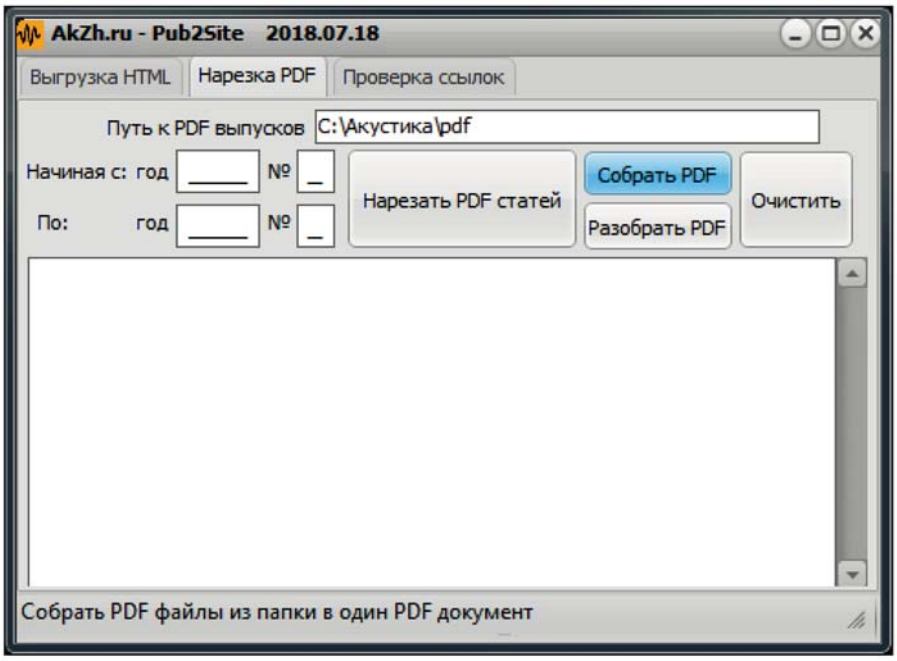

Рис. 6. АРМ программы Pub2Site в режиме сборки PDF-файлов.

Собственно добавление PDF-файла производится при помощи вызова утилиты pdftk.exe с соответствующими параметрами командной строки.

Имя сборного файла задается автоматически в виде sborkaNNNN.pdf, где NNNN – порядковый номер. Каждый раз при запуске сборки программа проверяет имеющиеся сборные файлы и создает новый с уникальным номером NNNN. Журнал сборки содержится в текстовом файле с именем sborkaNNNN.txt. Пример его содержимого:

1955_1_12-22.pdf%1%12

1955_1_23-30.pdf%13%20

1955_1_3-11.pdf%21%29

1955_1_31-39.pdf%30%38

1955_1_40-47.pdf%39%46

1955_1_48-57.pdf%47%56

1955_1_58-69.pdf%57%68

1955_1_70-77.pdf%69%76

1955_1_78-88.pdf%77%87

1955_1_89-95.pdf%88%94

1955_1_96.pdf%95%95

1955_1_cover.pdf%96%102

Каждая строка соответствует одному исходному PDF-файлу и содержит его имя, а также положение (начальную и конечную страницы) в сборном PDF-файле, разделенные символом процента.

Создание в сборном PDF-файле текстового слоя при помощи ABBY FineReader:

1. Запустить FineReader, в стартовом меню выбрать “Adobe PDF” (рис. 7).

2. Выбрать формат выходного файла “PDF”, язык документа “Русский и английский”.

3. Настроить качество изображения (рис. 8).

4. Нажать на панель “Файл изображения в PDF” (рис. 7).

5. Выбрать с помощью появившегося диалогового окна PDF файл, предназначенный для добавления текстового слоя.

FineReader загрузит выбранный файл и произведет его распознавание. Созданный текстовый слой будет размещен под изображением страниц в новом, “выходном” PDF-файле, который изначально имеет имя вида sndt8d14.pdf (комбинация букв и цифр в имени файла может отличаться) и располагается во временной папке FineReader по адресу: C:\Windows\Temp\FineReader12.00. Вместо C:\Windows\Temp может быть другая папка – в зависимости от текущего значения системной переменной TEMP.

Далее удобнее всего поступить так: не закрывая FineReader и не удаляя в нем распознанные страницы, переименовываем исходный сборный файл sborkaNNNN.pdf, например, в sborkaNNNN_0.pdf, а на его место копируем полученный временный PDF-файл из E:\Temp\FineRead-

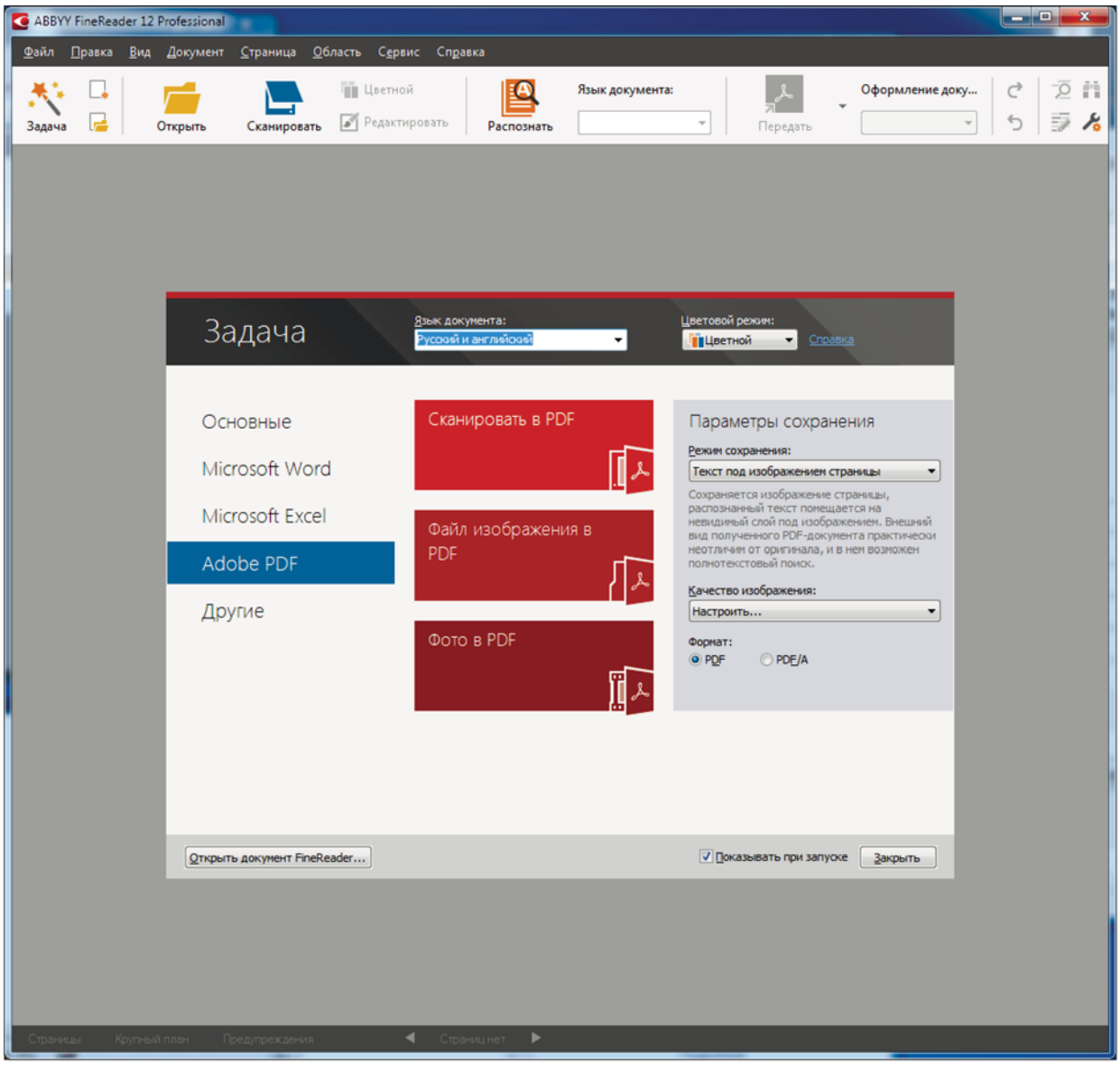

Рис. 7. Стартовое меню программы FineReader.

er12.00, переименовывая его из sndt8d14.pdf в sborkaNNNN.pdf.

После этого можно закрыть FineReader или удалить в нем распознанные страницы, готовясь к обработке следующего файла.

Извлечение PDF-файлов статей из сборного PDF-файла:

Поместить сборные файлы, предназначенные для разборки, в одну папку.

В эту же папку поместить файлы журналов, созданных на стадии сборки. Эти файлы должны иметь расширение.txt и имя, совпадающее с именем соответствующего сборного PDF-файла.

Запустить программу Pub2Site, открыть вкладку “Нарезка PDF”. В поле “Путь к PDF выпусков” указать адрес папки. Наличие косой черты в конце адреса папки не обязательно.

Нажать кнопку “Разобрать PDF” (рис. 9).

По данной команде программа Pub2Site:

– составляет список TXT-файлов в папке, рассматривая их в качестве журналов сборки;

– создает подпапку “razborka”;

– последовательно (по одному) обрабатывает TXT-файлы журналов сборки. Если обнаруживает соответствующий PDF-файл (с тем же именем, что и у журнала, но расширением .pdf), то запускает процедуру его разборки – извлекает из сбор-

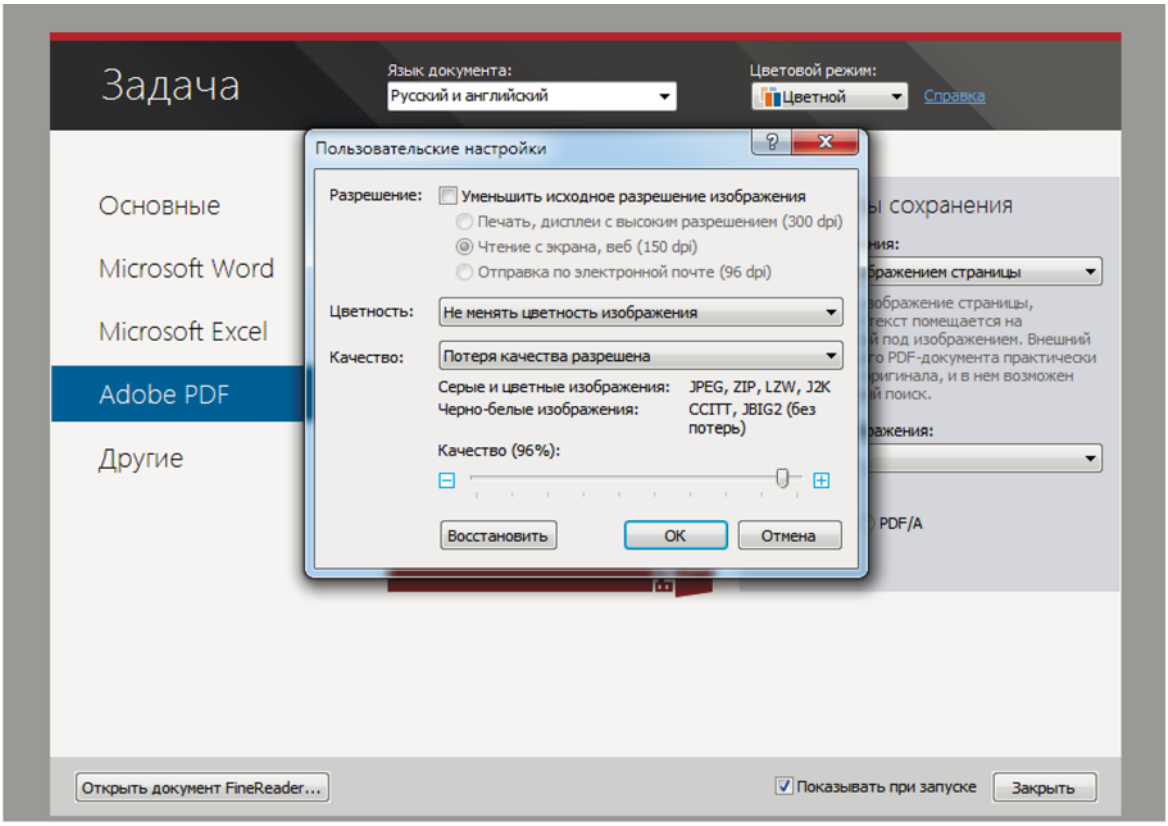

Рис. 8. Панель выбора выходного файла, языка распознавания и настройки изображения.

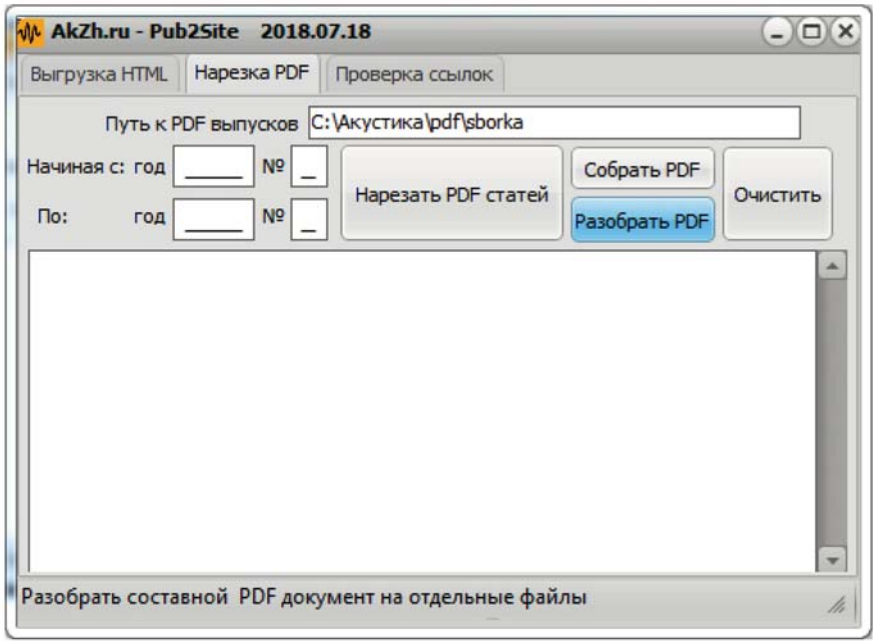

Рис. 9. АРМ программы Pub2Site в режиме разборки PDF-файлов.

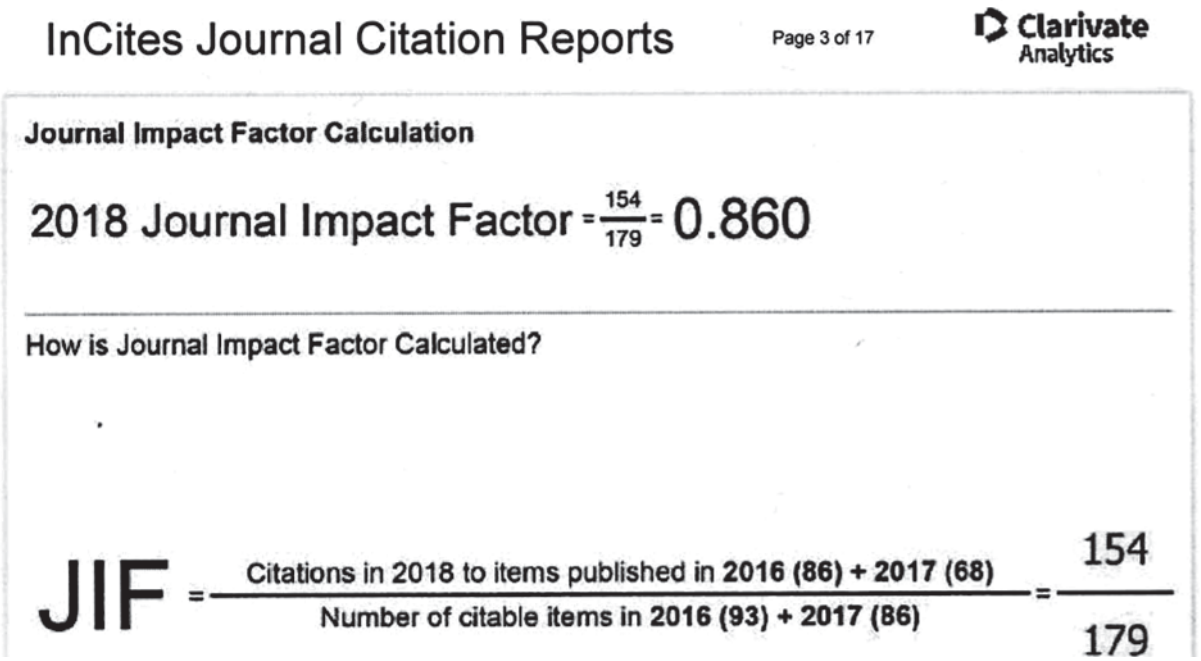

Рис. 10. Импакт-фактор “Акустического журнала” на 2018 г. по данным Clarivate Analytics.

ного PDF-файла файлы статей согласно записям в журнале об их имени и положении.

Собственно извлечение каждого PDF-файла производится при помощи вызова утилиты pdftk.exe с соответствующими параметрами командной строки.

Если при извлечении очередного файла в подпапке “razborka” уже существует файл статьи с таким именем, то программа замещает его новым.

На этом процесс наложения на изображения статей текстового слоя заканчивается.

ЗАКЛЮЧЕНИЕ

Надо отметить, что работы в информационном поле русскоязычных публикаций активно ведутся и другими организациями см. например, [19–21]. Естественно, это можно только приветствовать. В целом, тема сбора русскоязычных публикаций, перевода их в электронный вид и доступа к ним является, на наш взгляд, важной как со стороны их сохранения, ликвидации повторных разработок и выявления плагиата, так и со стороны подтверждения при необходимости приоритета наших исследователей. Последнее, по мнению многих, является проблемой из-за слабого знакомства в англоязычном мире с русскоязычными публикациями. Таким образом, основной посыл нашей статьи заключается в необходимости облегчения доступа к русскоязычным источникам информации и наращивания их полноты в электронном виде.

В заключение отметим, что отраженный в статье аспект наукометрической деятельности в современных условиях является чрезвычайно важным. “По несчастью или к счастью”, чиновничество обратило на него свое внимание в попытке оценки перспективных направлений развития науки и техники, формализации оценки деятельности как научных сотрудников, так и научных и образовательных учреждений. Поэтому мы обращаем внимание читателей на более тщательное использование пристатейной литературы, не забывая при этом “Акустический журнал”, а в особенности цитирования его статей за предшествующие два года. Ведь именно эти ссылки дают как импакт-фактор нашего журнала, так и индексы цитирования наших авторов. На рис. 10 мы приводим скан страницы компании Clarivate Analytics по “Акустическому журналу” (Acoustical Physics) с информацией по импакт-фактору журнала на 2018 г. Сделай мы с вами на 25 ссылок больше, и импакт-фактор стал бы более единицы.

СПИСОК ЛИТЕРАТУРЫ

1. *Щуко Ю.Н.* Некоторые аспекты развития Всероссийского института научной и технической информации // НТИ. Сер. 1. Организация и методика информ. работы. 2018. № 9. С. 1–6.

2. *Семенов В.В.* Нынешние реалии Реферативного журн. // Вестник Российской академии наук (РАН). 2010. Т. 80. № 4. С. 337–341.
3. *Шамаев В.Г.* Реферативный журнал “Физика” ВИНИТИ: проблемы и решения // Вестник РАН. 2011. Т. 85. № 5. С. 430–435.
4. *Гиляревский Р.С., Мельникова Е.В.* О разработке концепции государственной наукометрической системы и методики ее функционирования // НТИ. Сер. 1. Организация и методика информ. работы. 2018. № 9. С. 7–12.
5. *Шамаев В.Г., Щуко Ю.Н.* Банк данных ВИНИТИ РАН. Проблемы и перспективы развития // НТИ. Сер. 1. Организация и методика информ. работы. 2019. № 9. С. 1–8.
6. *Шамаев В.Г., Горшков А.Б.* Русскоязычные публикации по акустике: фрагменты инфометрического анализа // Ученые записки физического факультета Московского Университета. 2018. № 5. С. 1850501-1–1850501-6.
7. *Кириллова О.И.* Российские журналы в международном пространстве: перспективы признания и повышения авторитета // Конференция РИНЦ Science Online XXI “Электронные информационные ресурсы для науки и образования”, 27 января–3 февраля 2018 г., Австрия. https://elibrary.ru/projects/conference/austria2018/presentations/KirillovaRussianJournals.pdf (Дата обращения 03.07.2019).
8. *Зибарева И.В.* Российская научная периодика в глобальных информационно-аналитических ресурсах: вчера и сегодня // Научное издание международного уровня–2017: мировая практика подготовки и продвижения публикаций. Материалы 6-й Международ. науч.-практ. конф. М., 18–21 апреля 2017 г. С. 43–53. https://doi.org/10.24069/2017.978-5-7996-2227-5.07
9. *Шамаев В.Г., Горшков А.Б.* Открытая система информационного обеспечения акустики // Акуст. журн. 2017. Т. 63. № 4. С. 449–458.
10. *Шамаев В.Г., Горшков А.Б.* Система информационного обеспечения и поддержка научных исследований в области физико-математических наук. М: ВИНИТИ, 2017. 272 с. ISBN 978-5-9002-4251-4.
11. *Шамаев В.Г., Горшков А.Б.* О новых информационных ресурсах и авторефератах диссертаций по акустике и смежным дисциплинам, опубликованных за 2007–2017 гг. // Акуст. журн. 2019. Т. 65. № 2. С. 241–288.
12. *Шамаев В.Г., Горшков А.Б.* Навигация по русскоязычным источникам научной информации // Вестник Российской академии наук. 2017. Т. 87. № 7. С. 650–654.
13. Интеллектуальная система тематического исследования научно-технической информации. Ред. Садовничий В.А. М.: Изд-во МГУ, 2014. 262 с. https://istina.msu.ru/media/publications/book/4cd/546/7375366/Istina-book.pdf (Дата обращения 12.04.2019).
14. *Шамаев В.Г., Горшков А.Б., Гущина Л.Г., Якименко В.И.* Анализ информационно-поисковых систем по физике: проблема поиска в Интернете на примере акустики // Ученые записки физического ф-та МГУ. 2017. № 4. С. 1740801-1–1740801-9.
15. *Некрылов Н.* Q1 не то, что кажется // Троицкий вариант. 2019. № 281. С. 4–5.
16. *Жижченко А.Б., Изаак А.Д.* Информационная система Math-Net.Ru. Современное состояние и перспективы развития. Импакт-факторы российских математических журналов // Успехи мат. наук. 2009. Т. 64. № 4. С. 195–204.
17. *Чебуков Д.Е.* Поиск потерянных цитирований в Web of Science. Исправление ошибок в списках литературы Web of Science // Научный сервис в сети Интернет: труды XIX Всероссийской научной конференции. Новороссийск, 18–23 сентября 2017 г. М.: ИПМ им. М.В. Келдыша, 2017. С. 461–467.
18. *Шамаев В.Г., Горшков А.Б., Якименко В.И.* Полнотекстовый архив “Акустического журнала” в Интернете (http://www.akzh.ru). Опыт первых пяти лет // Акуст. журн. 2017. Т. 63. № 5. С. 573–580.
19. *Атаева О.М., Серебряков В.А.* Онтология цифровой семантической библиотеки LibMeta // Информатика и ее применение. 2018. Т. 12. № 1. С. 2–10.
20. *Атаева О.М., Серебряков В.А.* Персональная открытая семантическая цифровая библиотека LibMeta. Конструирование контента. Интеграция с источниками LOD // Информатика и ее применение. 2017. Т. 11. № 2. С. 85–100.
21. *Огальцов А.В., Бахтеев О.Ю.* Автоматическое извлечение метаданных из научных PDF-документов // Информатика и ее применение. 2018. Т. 12. № 2. С. 75–82.