



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 187 за 2018 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Богданова В.М.,
Горбунов-Посадов М. М.,
Китаев Е. Л., Кузьмичев Д.Л.,
Слепенков М.И.

Экспорт метаданных
препринтов ИПМ им.
М.В.Келдыша

Рекомендуемая форма библиографической ссылки: Экспорт метаданных препринтов ИПМ им. М.В.Келдыша / В.М.Богданова [и др.] // Препринты ИПМ им. М.В.Келдыша. 2018. № 187. 20 с.
<http://doi.org/10.20948/prepr-2018-187>
URL: <http://library.keldysh.ru/preprint.asp?id=2018-187>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

**В.М. Богданова, М.М. Горбунов-Посадов,
Е.Л. Китаев, Д.Л. Кузьмичев, М.И. Слепенков**

**Экспорт метаданных
препринтов ИПМ им. М.В.Келдыша**

Москва — 2018

УДК 004.915

Богданова В.М., Горбунов-Посадов М.М., Китаев Е.Л., Кузьмичев Д.Л., Слепенков М.И.

Экспорт метаданных препринтов ИПМ им. М.В.Келдыша

Метаинформация о препринтах поступает в базу данных электронной библиотеки публикаций ИПМ и далее экспортируется во внешние хранилища, такие как eLibrary и CrossRef. Для обеспечения эффективной работы поисковых систем строится карта сайта препринтов. Кроме того, дополнительные усилия прилагаются для корректного индексирования препринтов поисковой системой Google Scholar.

Ключевые слова: электронная библиотека публикаций, экспорт метаданных, DOI, карточка публикации, карта сайта

Vera Mikhailovna Bogdanova, Mikhail Mikhailovich Gorbunov-Possadov, Evgeny L'vovich Kitaev, Dmitry Leonidovich Kuzmichev, Mikhail Ivanovich Slepencov

Export of metadata of Keldysh Institute Preprints

The meta-information about preprints is entered the database of the electronic library of Keldysh Institute publications and then exported to external storages, such as eLibrary and CrossRef. To ensure the effective operation of search engines a site map of preprints is built. Additional efforts are made for the correct indexing of preprints by the Google Scholar search engine.

Key words: electronic publications library, export of metadata, DOI, publication card, sitemap

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 16-01-00158-а.

Оглавление

Введение	3
1. Карточка публикации.....	3
2. Экспорт метаданных в РИНЦ	9
3. Получение DOI препринта	12
4. Карта публичного библиотечного сайта ИПМ	14
5. Метаданные в препринтах для Google Scholar.....	17
6. Метаданные живой публикации	19
Заключение.....	20
Библиографический список.....	20

Введение

В ИПМ разработана и успешно эксплуатируется библиотечная система, обслуживающая размещение на сайте полных текстов публикаций сотрудников ИПМ и авторов других организаций, а также обеспечивающая регистрацию и хранение метаданных о размещаемых публикациях.

Библиотечная система ИПМ состоит из двух веб-приложений: автоматизированное рабочее место администратора (АРМ администратора) и публичный сайт электронной библиотеки. В настоящей работе будет описана подготовка данных в административной системе.

Для обеспечения широкой доступности препринтов через поисковики (такие как Google, Yandex и т. п.) в АРМ администратора и на публичном библиотечном сайте реализованы механизмы экспорта метаданных во внешние системы. Данные, необходимые для экспорта, берутся из карточки публикации, ввод информации в которую выполняется администратором.

1. Карточка публикации

Прежде чем погружаться в детали ведения препринтов, важно отметить, что веб-представительство ИПМ в Интернет имеет некоторую особенность. Электронная библиотека и публичный сайт ИПМ физически размещены на двух разных веб-серверах — <http://library.keldysh.ru> и <http://www.keldysh.ru> (<http://keldysh.ru>) соответственно. Такая конфигурация сложилась исторически, поскольку эти компоненты разрабатывались разными командами программистов. При этом следует подчеркнуть, что карточки публикаций (метаданные) и полные тексты публикаций хранятся на разных серверах, что повлекло за собой некоторые технические проблемы при реализации экспорта метаданных, решение которых будет описано в настоящей работе.

Основным объектом хранения в базе данных системы является электронный документ, вводимый через карточку публикации (рис. 1). Введенные администратором данные отображаются в дальнейшем на публичном сайте электронной библиотеки ИПМ и будут доступны посетителю этого сайта в виде аналогичной карточки публикации для просмотра, где ряд значений оформлен в виде гипертекстовых ссылок [1-3].

В представленной ниже карточке для ввода публикаций в АРМ администратора мы не будем акцентировать внимание на всех полях, необходимых для экспорта метаданных, но очевидных при их заполнении. В описании опущены такие поля как: Год издания, Номер препринта, Язык публикации, Количество страниц, УДК, Грант, Название, Аннотация, Ключевые слова. Остановимся только на тех полях, которые имеют особенности заполнения.

Регистрационная карточка публикации	
Тип публикации:	Препринт ИПМ
Предметная рубрика:	Теоретические и прикладные задачи механики
Год издания: 2017	Месяц издания:
Номер препринта:	20 (заполняется только для препринтов)
Авторы:	<input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> Вашковьяк М.А./Vashkov'yak M.A. <input type="checkbox"/> отсортировать по алфавиту
	<input type="button" value="Добавить (ИПМ)"/> <input type="button" value="Добавить (не ИПМ)"/>
Язык публикации:	русский
Издание:	
Признак живой публикации:	<input type="checkbox"/>
Место издания:	Москва (рус) Moscow (eng)
Кол-во страниц:	32 (рус) (eng)
Название (рус):	О компланарном интегрируемом случае двукратно-осредненной задачи Хилла с учетом сжатия центрального тела
Название (анг):	On coplanar integrable case of double-averaged Hill's problem taking into account the oblateness of central body
Аннотация (рус):	Рассматривается двукратно-осредненная задача Хилла с учетом сжатия центральной планеты в случае, когда ее экваториальная плоскость совпадает с плоскостью орбитального движения относительно возмущающего тела. Качественное исследование этого так
Аннотация (анг):	Double-averaged Hill's problem taking into account the oblateness of central planet is considering in the case, when its equatorial plane coincides with the plane of orbital motion relatively perturbing body. A qualitative investigation of this so named
Ключевые слова (рус):	спутниковые орбиты, вековые возмущения
Ключевые слова (анг):	satellite orbits, secular perturbations
DOI (рус):	10.20948/prepr-2017-20 <input type="button" value="..."/> для препринтов при нажатии кнопки поле будет заполнено автоматически
DOI (анг):	<input type="button" value="..."/> для препринтов при нажатии кнопки поле будет заполнено автоматически
ББК:	22.21
УДК:	521.186
Гранты (рус.):	
Файл с текстом публикации (рус):	prep2017_20.pdf (рус)
Файл с текстом публикации (анг):	(анг)
Рецензия:	Настоящая работа выполнена в рамках плановых исследований, ведущихся в отделе. Она является продолжением и развитием исследования задачи о вековых возмущениях орбиты спутника пренебрежимо малой массы при совместном влиянии сжатия
ФИО рецензента:	Зав. отделом Голубев Ю.Ф.

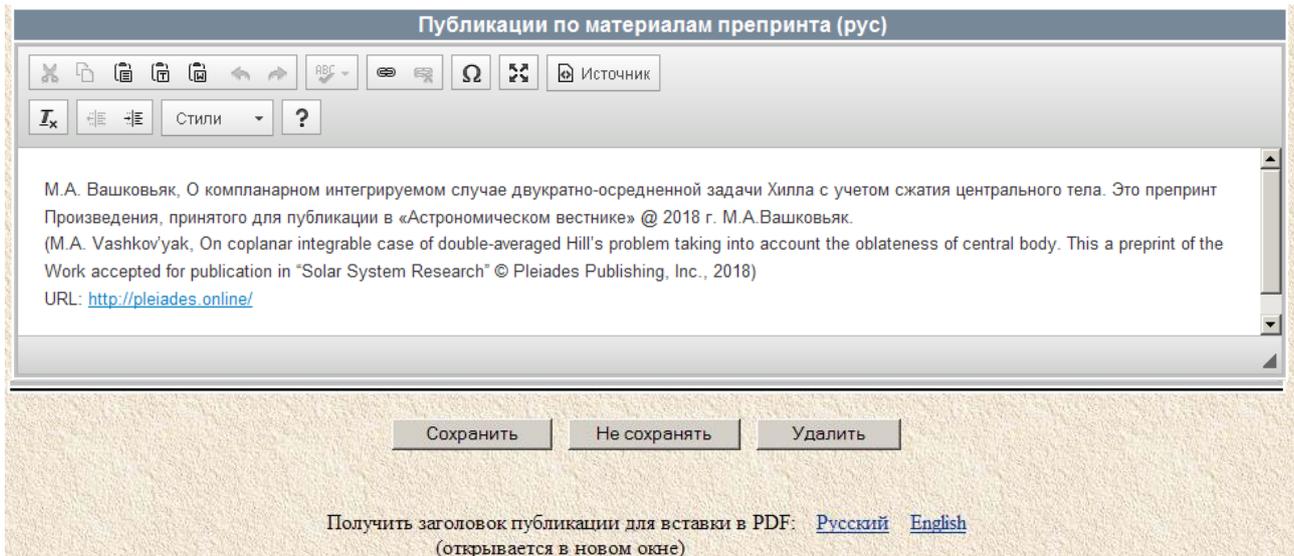


Рис. 1. Карточка для ввода публикации в АРМ администратора

Данные вводятся в карточку администратором на основании заявки авторов на издание препринта.

Типы публикаций

Карточка позволяет вводить различные типы публикаций, такие как препринт, статья, сборник, монография и т.п. В настоящей работе основной акцент делается на публикациях типа «препринт ИПМ».

Поддержка языков

Публичный сайт библиотеки ИПМ поддерживает двуязычие, в связи с этим в карточке публикаций все основные поля дублируются на английском языке. Двуязычие обеспечивается универсальной функцией, которая представляет все тексты публичного сайта на выбранном пользователем языке (русском или английском).

Вставка авторов

При заполнении поля Авторы различается ввод авторов из ИПМ и авторов из других организаций. Ввод авторов из ИПМ выполняется на основе справочника, который хранится в отдельной таблице БД. В справочнике представлены не только ФИО автора на русском и английском языках, но и персональные данные сотрудника, такие как E-mail, телефоны, SPIN, ORCID и т.п. Часть данных ведется отделом кадров ИПМ и экспортируется в библиотечную систему.

Авторы сторонних организаций вводятся через отдельную карточку (рис. 2), в которой помимо ФИО автора на русском и английском языках и персональных данных автора указывается его место работы. В базе данных авторы сторонних организаций хранятся при публикации в отдельной таблице БД.

Форма для ввода других авторов (не из ИПМ) -- Webpage Dialog

ФИО (рус) / ФИО (анг)

Окунев С.К. / Okunev S.K.

Полное ФИО (рус): Окунев Сергей Константинович

Полное ФИО (анг): Okunev Sergey Konstantinovich

Если фамилия состоит из нескольких слов, то после фамилии надо ставить "-" (восклицательный знак)

Название организации: АО "МЦСТ"

Адрес:

E-mail: sergey.k.okunev@mcstru

Персональная страница (URL):

SPIN:

ORCID: ORCID.ORG/

Изменить

Рис. 2. Форма ввода внешних авторов

При заполнении карточки публикаций имеются средства перестановки ФИО авторов в том порядке, который указали сами авторы, а также возможность отсортировать ФИО введенных авторов по алфавиту.

Публикации по материалам препринта

По желанию автора в карточке могут быть указаны данные о публикациях по материалам препринта. Эти данные включают в себя URL публикации, ее DOI, название, авторов и другие реквизиты публикации. Публикации по материалам препринта оформляются в виде HTML-фрагментов, которые вводятся с помощью специализированного текстового редактора, встроенного в веб-страницу карточки публикаций. Публикаций по материалам препринта может быть введено несколько.

Поле DOI

Поле DOI представлено в двух вариантах, для препринтов на русском и английском языке. Если препринт двуязычный, то заполняются оба поля. В поле DOI по кнопке автоматически генерируется DOI препринта на основании года препринта, его номера и языка написания. Собственно в DOI препринта значение данного поля превратится после регистрации его на сайте Crossref, что будет описано ниже.

Ввод списка литературы

Для взаимодействия с eLibrary.ru, о чем также будет рассказано ниже, администратор вводит для каждого препринта список цитируемой литературы. Этот список вводится в отдельную форму регистрации списка литературы (рис. 3).

Ввод сотрудников • Вставка ID автора • Поиск сотрудников • Ввод публикаций • Поиск публикаций

Регистрация списка литературы

Список литературы для публикации: **Препринт ИПМ №135, год выпуска:2018**
 Название: **Алгебры бернуллиевских распределений с единственной предельной точкой**

Колпаков Р. М. Критерий порождения множеств рациональных вероятностей в классе булевых функций // Салимов Ф. И. Об одной системе образующих для алгебр над случайными величинами // Изв. вузов. Мат. Схиртладзе Р. Л. О методе построения булевой случайной величины с заданным распределением вероятности Яшунский А. Д. Преобразования бернуллиевских распределений булевыми функциями из замкнутых классов Яшунский А. Д. Конечные алгебры бернуллиевских распределений // Дискретная математика. 2018. Т. 30

Зарегистрированный список литературы:

1. Колпаков Р. М. Критерий порождения множеств рациональных вероятностей в классе булевых функций // Дискретный анализ и исследование операций. 1999. Т. 6, No 2. С. 41–61.
2. Салимов Ф. И. Об одной системе образующих для алгебр над случайными величинами // Изв. вузов. Матем., 1981, No 5. С. 78–82.
3. Схиртладзе Р. Л. О методе построения булевой случайной величины с заданным распределением вероятностей // Дискретный анализ: Сб. науч. тр. Новосибирск: Ин-т математики СО АН СССР, 1966. Вып. 7. С. 71–80.
4. Яшунский А. Д. Преобразования бернуллиевских распределений булевыми функциями из замкнутых классов // Препринты ИПМ им. М.В.Келдыша. 2016. No 38. 23 с. doi:10.20948/prepr-2016-38 URL: <http://library.keldysh.ru/preprint.asp?id=2016-38>
5. Яшунский А. Д. Конечные алгебры бернуллиевских распределений // Дискретная математика. 2018. Т. 30, No 2. С. 148–161.

Рис. 3. Форма ввода списка литературы

По кнопке из списка автоматически удаляется нумерация, проставленная авторами, и назначается нумерация стандартного вида. В базе данных список литературы хранится при публикации в отдельной таблице БД.

Поле Файл с текстом публикации

В поле Файл с текстом публикации администратор вводит имя PDF-файла, под которым данная публикация сохранена на сервере ИПМ. Если авторы размещают двуязычную публикацию, то заполняются два поля Файл: PDF-файлы препринта на русском и английском языках. На основании данного поля автоматически формируется ссылка на полный текст публикации, расположенная в карточке публикации на публичном сайте.

Формирование PDF-файла и ввод метаатрибутов

Все публикации ИПМ размещаются в виде PDF-файлов. Если исходный текст публикации был передан авторами в другом формате, то перевод его в PDF-формат выполняется администратором.

Чаще всего обнаруживается, что авторы не указали метаатрибуты публикации с помощью средства (Word, Tex), которым был подготовлен ее текст. В таком случае администратор вводит метаатрибуты в свойства PDF-файла, как переданного авторами, так и полученного в результате перевода. В качестве метаатрибутов публикации берутся название, ФИО авторов и ключевые слова препринта. Наличие в публикации метаатрибутов надежно обеспечивает корректный результат ее индексирования и последующего нахождения поисковыми системами. Конечно, публикация будет найдена и при отсутствии в ней метаатрибутов. Однако если взять произвольный фрагмент из основного текста такой публикации (без метаатрибутов) и поискать его в Google, то вероятнее всего, что в полученном результате в качестве названия публикации будет фигурировать название нашего института — Google сочтет, что названием является первое предложение файла, набранное прописными буквами.

Генерация нулевой страницы препринта

К PDF-файлу с текстом препринта администратор присоединяет нулевую страницу (рис. 4), которая генерируется из карточки публикации отдельной программой. Нулевая страница оформлена в дизайне сайта ИПМ, в ней присутствуют гиперссылки на главную страницу сайта ИПМ и разделы онлайн-библиотеки. В нулевую страницу помещаются метаданные препринта из базы данных публикаций. В первом блоке выводится ISSN, авторы и название публикации. На данной странице, равно как и на всех страницах библиотечного сайта ИПМ, фамилии авторов из ИПМ оформлены в виде гипертекстовых ссылок. Эти ссылки приводят на страницу, где представлена контактная информация и список публикаций данного автора с разбивкой по годам.

Во втором блоке страницы размещена рекомендуемая форма библиографической ссылки на данную работу. Ссылка на публикацию формируется в каноническом виде, позволяющем использовать ее в списке литературы в других работах. URL публикации генерируется автоматически на основании года и номера препринта с учетом языка его написания и представляет собой ссылку на карточку метаданных препринта на публичном сайте.

Во втором блоке страницы может быть также представлен список публикаций по материалам препринта.



ИПМ им.М.В.Келдыша РАН • Электронная библиотека
Препринты ИПМ • Препринт № 20 за 2017 г.

ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Вашковьяк М.А.

О компланарном интегрируемом случае двукратно-осредненной задачи Хилла с учетом сжатия центрального тела

Рекомендуемая форма библиографической ссылки

Вашковьяк М.А. О компланарном интегрируемом случае двукратно-осредненной задачи Хилла с учетом сжатия центрального тела // Препринты ИПМ им. М.В.Келдыша. 2017. № 20. 32 с.
doi:[10.20948/prepr-2017-20](https://doi.org/10.20948/prepr-2017-20)
URL: <http://library.keldysh.ru/preprint.asp?id=2017-20>

Публикации по материалам препринта

М.А. Вашковьяк, О компланарном интегрируемом случае двукратно-осредненной задачи Хилла с учетом сжатия центрального тела. Это препринт Произведения, принятого для публикации в «Астрономическом вестнике» @ 2018 г. М.А.Вашковьяк.
(M.A. Vashkov'yak, On coplanar integrable case of double-averaged Hill's problem taking into account the oblateness of central body. This a preprint of the Work accepted for publication in "Solar System Research" © Pleiades Publishing, Inc., 2018)
URL: <http://pleiades.online/>

Рис. 4. Нулевая страница препринта

Помимо описанных выше полей в карточке публикации автоматически заполняются необходимые для экспорта метаданных поля Дата создания объекта, Дата изменения объекта.

2. Экспорт метаданных в РИНЦ

Одна из важных задач при ведении препринтов ИПМ — взаимодействие с системой Российского индекса научного цитирования (РИНЦ). РИНЦ — это национальная информационно-аналитическая система, аккумулирующая в настоящий момент метатрибуты более 12 миллионов публикаций российских ученых, а также информацию о цитировании этих публикаций из около 6 тысяч российских журналов. Она предназначена не только для оперативного

обеспечения научных исследований актуальной справочно-библиографической информацией, но и выполняет функцию инструмента для оценки учёных или научных организаций на основе цитирования. РИНЦ, в отличие от основных международных систем цитирования, находится в открытом доступе.

РИНЦ реализуется на базе Научной электронной библиотеки eLibrary.ru, куда в результате обработки выпусков изданий попадают основные атрибуты статей, пристатейные библиографические списки, а также полные тексты статей либо ссылки на полные тексты, размещенные на сайтах издательств. eLibrary.ru предоставляет пользователю возможность поиска статьи или издания по заданному набору атрибутов. Предусмотрены также поисковые запросы, позволяющие находить как публикации, цитируемые в отдельно взятой статье, так и публикации, цитирующие эту статью.

На основе анализа пристатейных списков литературы для размещенных в базе статей РИНЦ формирует рейтинг цитируемых авторов, а также подсчитывает импакт-фактор отдельного журнала — интегральный показатель цитируемости опубликованных в журнале научных статей, который позволяет выстраивать рейтинги изданий [4].

ИПМ в 2012 году заключил с Научной электронной библиотекой eLibrary.ru лицензионный договор на регулярное размещение в РИНЦ метаданных как издающихся, так и выпущенных в предыдущие годы препринтов. В РИНЦ передаются также ссылки на полные тексты препринтов, размещенные на сайте ИПМ.

Для подготовки метаданных eLibrary.ru предлагает использовать онлайн-инструмент XML-разметки «Артикулус», на выходе из которого получается XML необходимого формата. Однако такой ручной ввод информации в eLibrary.ru по сути дублировал бы деятельность администратора по вводу метаданных публикации в БД библиотечной системы ИПМ. Во избежание ручного труда, нами была разработана программа автогенерации требуемого XML. Поскольку сам формат XML «Артикулуса» нигде не описан, для создания программы пришлось разметить несколько препринтов вручную с помощью «Артикулуса» и проанализировать форматы сформированных при этом XML. Программа автогенерации XML запускается администратором после того, как для препринта заполнена карточка публикации. В программе выполняется также и упаковка сгенерированного XML, как того требует eLibrary.ru.

В eLibrary.ru для каждого препринта передается в виде XML следующая информация.

- Год препринта
- Номер препринта
- Количество страниц
- Язык публикации
- Название препринта на русском и английском языках

- Аннотация на русском и английском языках
- Ключевые слова на русском и английском языках
- Сведения об авторах

Для каждого автора указываются:

- фамилия, имя, отчество полностью на русском и английском языках
- полное название организации — места работы

Опционально для автора могут быть указаны:

- адрес организации
- e-mail
- персональная страница
- SPIN-код
- Тематическая рубрика — код УДК
- Список литературы
- URL публикации

Опционально для препринта могут быть указаны:

- DOI
- Финансовая поддержка работы (грант)

В передаваемом в eLibrary.ru XML-файле перечисленные выше метаданные препринта предваряются информацией (ISSN, название и т.д.) о журнале, содержащем препринты. Журнал, зарегистрированный в eLibrary.ru нашим институтом, называется «Препринты ИПМ им. М.В.Келдыша». Каждый препринт передается как отдельный номер журнала.

Чтобы передать сгенерированный XML в eLibrary.ru, администратор заходит в Систему подготовки научных изданий РИНЦ и выполняет загрузку XML для каждого препринта (рис. 5).

The screenshot shows a web browser window with the URL http://intra.elibrary.ru/projects/intra/system2/issue_edit_publisher.asp. The page title is 'Журналы - Windows Internet Explorer'. The main content area is titled 'Добавление журнала' (Add journal). There are buttons for 'Сохранить' (Save) and 'Отменить' (Cancel). The form contains the following fields:

- Форма журнала: Электронная версия
- Тип выпуска: Выпуск уже размеченный (с xml файлом)
- Издание: Препринты ИПМ им. М.В. Келдыша

Below the form, there is a section 'Введите выпуски, которые вы хотите добавить:' (Enter issues you want to add:). It contains a table with the following data:

№	Год	Тон	Номер	Страниц
	2018			<--Шаблон
1	2018		156	Страниц: 19
2	2018		157	Страниц: 16
3	2018		158	Страниц: 20

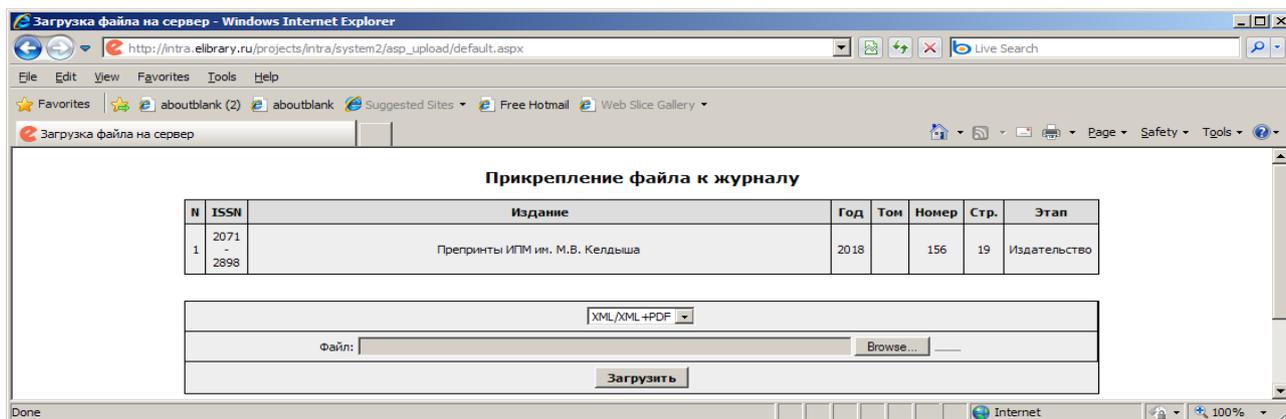


Рис. 5. Загрузка XML препринта в eLibrary.ru

3. Получение DOI препринта

DOI (Digital Object Identifier) — это уникальный цифровой идентификатор документа. DOI электронного документа содержит указатель местонахождения документа, его название и ассоциированный с документом набор описывающих его метаданных в структурированном виде.

DOI гарантирует переход на актуальное местонахождение определенной публикации. Объект в интернете может поменять свой адрес, после чего прямые ссылки на него перестают работать. Если же меняет адрес объект, ссылки на который реализованы посредством DOI, то издатель легко обеспечит корректность таких ссылок, поменяв в системной записи соответствующий DOI конкретный адрес.

DOI служит эффективным средством поиска нужных статей в периодических изданиях. Это более надежный метод поиска, чем библиографическая ссылка, поскольку DOI распознается без ошибки. Использование DOI повышает авторитет научного издания и является стандартом, принятым всеми ведущими издательствами мира [5].

ИПМ в 2016 году заключил договор с организацией Crossref — агрегатором и регистратором DOI, а также международной базой научных статей и их метаданных, на получение DOI для препринтов журнала “Препринты ИПМ им. М.В.Келдыша”.

Для присвоения и регистрации DOI для научной работы необходимо разместить её метаданные в системе Crossref. Аналогично ситуации с eLibrary.ru, требовался ручной ввод информации, уже введенной администратором в карточку публикации библиотечной системы ИПМ. Очевидным решением тут также стала разработка нами программы автогенерации XML для размещения метаданных в системе Crossref.

В Crossref при регистрации DOI для каждого препринта передается XML со следующей информацией.

- Год препринта
- Номер препринта
- Количество страниц

- Название препринта на русском и английском языках
- Сведения о каждом из авторов:
 - фамилия, имя, отчество полностью на русском и английском языках
 - место работы на английском языке
 - ORCID (опционально)
- Ссылка на полный текст препринта
- Номер DOI

Структура номера DOI состоит из префикса и суффикса. Префикс — это идентификатор издателя. Префиксы распределяются регистрационными агентствами, такими как Crossref. Издания ИПМ получили префикс "10.20948/". Суффикс присваивается публикации издательством и является уникальным в рамках префикса. Суффикс препринтов нашего института генерируется в виде `prepr-YYYY-N[-e]`, где `YYYY` — год выпуска препринта, `N` — номер препринта. Буква `e` добавляется для DOI препринта на английском языке.

Перед метаданными препринта в XML помещается блок информации о журнале, содержащем препринты. В нашем случае это журнал, зарегистрированный в Crossref как «Keldysh Institute Preprints». Метаданные журнала содержат, помимо полного и краткого названия журнала, его ISSN, DOI и URL. URL нашего журнала указывает на страницу публичного сайта ИПМ <http://library.keldysh.ru/preprints/>, где расположен список всех препринтов по годам издания.

Чтобы зарегистрировать DOI, администратор сначала должен выполнить генерацию XML для Crossref. Программы автогенерации XML для eLibrary.ru и Crossref объединены в одну утилиту (рис. 6), запускаемую администратором после заполнения карточки публикации препринта.

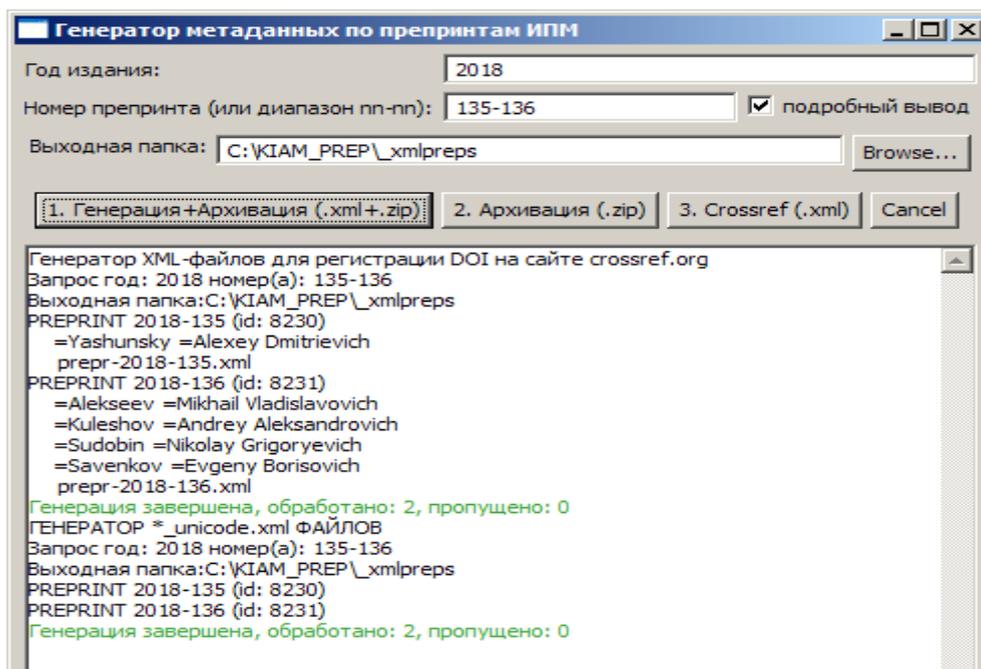


Рис. 6. Утилита генерации метаданных по препринтам ИПМ

Утилита может генерировать XML для заданного диапазона номеров препринтов. Подробный вывод ФИО авторов служит для контроля наличия в метаданных препринта полных имен и отчеств на английском языке, что не было обязательным при ведении препринтов ИПМ, но является жестким требованием Crossref.

Далее для регистрации сгенерированного XML в Crossref администратор заходит на сайт <https://doi.crossref.org> и выполняет загрузку XML (рис. 7) для каждого препринта.

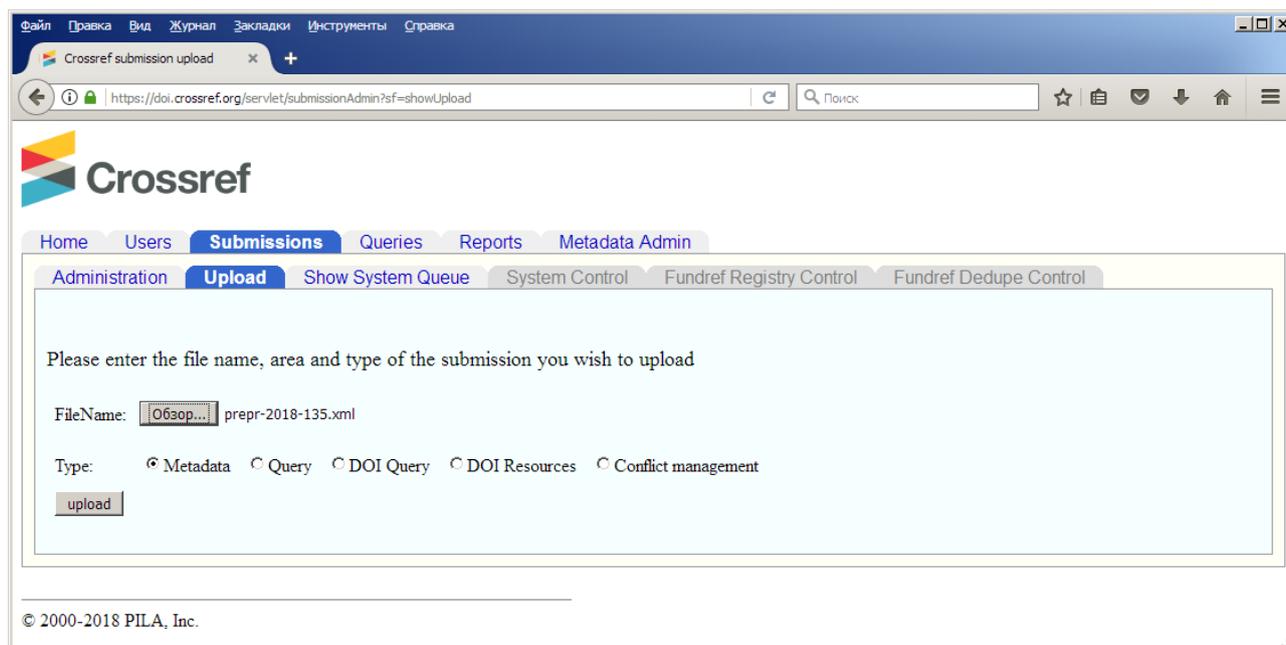


Рис. 7. Загрузка XML препринта в Crossref

4. Карта публичного библиотечного сайта ИПМ

Для публичного библиотечного сайта ИПМ нами была разработана карта сайта. Карта сайта (sitemap) — один из важнейших инструментов оптимизации сайта, способный ускорить его индексацию поисковыми системами примерно в 3 раза. Карта сайта представляет собой html страницу сайта или специальный xml файл, в котором отражены ссылки на все важные страницы сайта. Карта сайта помогает посетителю, будь это человек или поисковый робот, быстро найти любую страницу сайта, совершая минимальное число переходов.

Поисковые системы индексируют сайт постепенно, уровень за уровнем, начиная с главной страницы. Если на сайте много уровней вложенности страниц или большое количество страниц, то индексация всех страниц сайта занимает много времени. Для таких сайтов необходимо создать карту сайта, на которой будут представлены ссылки на все важные страницы сайта сразу, чтобы робот не тратил время на переходы по уровням вложенности.

Библиотечный сайт ИПМ <http://library.keldysh.ru> выполнен в парадигме интерфейса поисковых систем. Когда посетитель заходит на сайт, ему выдается

форма запроса, через которую формируется критерий поиска публикаций. Результатом выполнения запроса будет порционный список, в котором представлены все релевантные запросу объекты. Кликнув на любой элемент списка, посетитель попадает на карточку публикации, где, в свою очередь, имеется ссылка на полный текст публикации. На библиотечном сайте ИПМ к настоящему моменту размещено более 6,5 тысячи публикаций, и их количество увеличивается практически ежедневно. Таким образом, разработка карты для библиотечного сайта ИПМ являлась настоятельной необходимостью.

Карты сайта делятся на два основных вида или формата: когда карта сайта оформлена в html, и когда карта сайта представлена в виде файла в формате xml. html карта сайта — это, как правило, страница сайта, на которой перечислены ссылки. Обычно это ссылки на самые важные разделы и страницы сайта. html карта сайта больше рассчитана на людей, нежели на роботов, и помогает быстрее ориентироваться в основных разделах сайта. На публичном сайте ИПМ аналогом html карты сайта является страница «Каталог препринтов по годам», на которой представлены все года издания в виде ссылок, переход по любой из которых приводит к выводу списка всех препринтов за выбранный год. Названия публикаций в списке оформлены в виде ссылок на карточки препринтов.

Отметим, что для карты сайта, оформленной в виде html-страницы, существуют серьезные ограничения по количеству выводимых здесь ссылок. Если на странице ссылок будет слишком много, далеко не все ссылки могут быть проиндексированы или вовсе страница карты сайта может быть исключена из поиска за чрезмерное количество ссылок. html карта сайта — это не полноценный sitemap. Если на сайте несколько тысяч страниц, то рекомендуется представлять ссылки на них посредством карты сайта в формате xml.

XML карта сайта — это файл в формате xml с именем sitemap.xml, который обычно находится в корне сайта [6]. У карты сайта в формате xml есть множество преимуществ перед html картой сайта. Sitemap.xml — это специальный формат карты сайта, который принимается всеми популярными поисковыми системами, например, Google и Яндекс. В sitemap.xml можно указать до 50.000 ссылок. Более того, в sitemap.xml можно указать относительный приоритет и частоту обновления страниц. Если файл карты сайта включает более 50 тысяч ссылок или размер sitemap.xml превышает 10 Мб, то можно разделить карту сайта на несколько файлов, при этом в объединенной карте сайта можно указать несколько ссылок на разные файлы с картами.

Карта сайта для библиотечного сайта ИПМ была реализована нами в xml формате, который имеет следующий вид:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
<url>
```

```

    <loc>http://www.keldysh.ru/papers/1968/preprint.asp?id=1968-1</loc>
    <lastmod>2012-09-27</lastmod>
</url>
<url>
    <loc>http://www.keldysh.ru/papers/1968/preprint.asp?id=1968-2</loc>
    <lastmod>2012-09-27</lastmod>
</url>
<url>
    <loc>http://www.keldysh.ru/papers/1969/preprint.asp?id=1969-3</loc>
    <lastmod>2012-09-27</lastmod>
</url>
    ...
</urlset>

```

Теги, использованные в карте сайта ИПМ, имеют следующие значения:

- <urlset> — родительский тег, в который заключены все url-адреса;
- <url> — тег, в котором указаны сведения о конкретном url-адресе;
- <loc> — в данном теге указывается непосредственно url-адрес;
- <lastmod> — этот тег включает в себе дату последнего изменения страницы.

В карте библиотечного сайта ИПМ для индексирования перечислены все страницы с карточками публикаций, т. е. каждый url-адрес указывает на конкретную карточку публикации.

При создании карты сайта необходимо не только четко указать, какие страницы сайта нужно индексировать, но и как файл sitemap определяется поисковыми системами. Как было сказано выше, файл карты сайта размещается в корне сайта. Поисковые системы способны самостоятельно обнаружить этот файл. Существует несколько способов указания прямой ссылки на файл карты сайта для более быстрого его обнаружения поисковыми системами. Самый простой способ указать место расположения файла карты сайта — это непосредственное указание ссылки или нескольких ссылок на файлы карт сайта в инструментах для вебмастеров от Yandex и Google. Там же можно проверить карту сайта, провести анализ карты сайта на корректность, соответствие того, какие страницы из карты сайта найдены поисковой системой и сколько из них находится в индексации. Второй способ указать поисковым системам на расположение файла карты сайта — это директива Sitemap в файле robots.txt. Именно этот способ был использован при создании карты сайта для библиотечного сайта ИПМ.

Указание ссылки на карту сайта в robots.txt в нашем случае выглядит так: Sitemap: <http://www.keldysh.ru/papers/sitemap.xml>.

Для создания карты библиотечного сайта ИПМ нами была разработана программа генерации. Эта программа имеет важную особенность — она создает карту сайта динамически, т.е. карта строится в момент обращения робота и поэтому она всегда актуальна. Как описано выше, в карту включаются

URL и дата последнего обновления публикации. Умные роботы поисковиков должны помнить, когда последний раз обновлялась публикация, и не анализировать то, что помечено прежними датами (именно для этого в карте сайта используется тег `<lastmod>`).

Определенная трудность при генерации карты сайта была вызвана тем, что такие поисковые системы, как Google, доверяют метаданным только в том случае, если они расположены на одном сервере с текстом публикации. Между тем, как говорилось выше, в библиотечной системе ИПМ карточки публикаций и их полный текст расположены на разных серверах (в разных доменах). Для преодоления этого ограничения программа генерации выполняет `redirect`, чтобы “подложить” карточку публикации на тот сервер, где содержится текст публикации. При этом надо выполнить это действие так, чтобы посетитель сайта не видел дубль карточки публикации в другом месте.

Рассмотрим подробнее, как нами была решена эта проблема. Именно с головного сайта ИПМ <http://www.keldysh.ru> (<http://keldysh.ru>) стартуют поисковики для индексирования веб-страниц; здесь же размещен файл `robots.txt`, где прописана `http`-ссылка на `sitemap.xml` (<http://www.keldysh.ru/papers/sitemap.xml>) электронной библиотеки ИПМ. Библиотечный сайт ИПМ размещается на другом сервере <http://library.keldysh.ru>, в базе данных которого хранятся публикации, и собственно он отвечает за генерацию `sitemap.xml`. Для разрешения этой ситуации на головном сервере ИПМ в конфигурационном файле нами были написаны два декларативных правила, которые перенаправляют `http`-запросы на библиотечный сайт, делая серверный `redirect`. Первое правило перенаправляет `http`-запрос поисковика, когда тот обращается к файлу <http://www.keldysh.ru/papers/sitemap.xml>, на программу библиотечного сайта, которая генерирует из базы данных `sitemap`. Второе правило тоже адресовано к поисковику, когда тот начнет ходить по ссылкам, переданным ему в `sitemap.xml`. Здесь также происходит подмена ссылки с изменением домена. По этой ссылке выдается карточка публикаций, где, в свою очередь, представлена ссылка на полный текст публикации — основной объект индексирования в поисковых системах.

В заключение следует сказать, что содержимое карты сайта является лишь рекомендацией для поискового робота. Например, если установить для страницы сайта ежегодную частоту обновления, поисковые роботы все равно будут ходить чаще. А если установить частоту обновления страницы ежечасно, это не значит, что роботы будут индексировать страницу каждый час.

5. Метаданные в препринтах для Google Scholar

В 2004 году начала работать система Google Scholar (Google Академия) — бесплатная поисковая система по полным текстам научных публикаций всех форматов и дисциплин, доступная по ссылке <https://scholar.google.ru/>. Индекс Google Scholar включает данные из большинства рецензируемых онлайн

журналов крупнейших научных издательств Европы и Америки. По функциям Google Scholar похож на сайты, предоставляющие доступ к публикациям после оформления платной подписки, например, такие как Scopus и Web of Science.

В результатах поиска Google Scholar выводит ссылки на статьи. Большинство из ссылок ведут на страницы, содержащие краткую информацию о статье; есть также и ссылки к полному тексту статей. Расширенный поиск на этом сайте позволяет искать в конкретных журналах или статьях. Результаты поиска сортируются по рейтингу автора; количеству ссылок на статью; рейтингу статей, ссылающихся на найденную статью; рейтингу журналов, в которых опубликованы ссылающиеся статьи; рейтингу журнала, в котором опубликована найденная статья. Блок «Цитируется в» содержит список статей, в которых цитируется рассматриваемая статья. Эта функция, в частности, обеспечивает индекс цитирования, подобно Scopus и Web of Knowledge. В блоке «Статьи по теме» выводится список статей, похожих по содержанию с рассматриваемой статьёй. Статьи упорядочены по степени сходства с рассматриваемой статьёй и по своей значимости [7].

Для того чтобы публикации ИПМ корректно индексировались, Google Scholar рекомендует помещать в саму публикацию или в карточку описания публикации теги, содержащие метаинформацию о ней. Согласно этим рекомендациям, мы всегда размещаем в карточке публикации следующие метаданные:

```
<meta name="citation_title" content="Управление тремя спутниками
      в групповом полете при помощи электростатических сил">
<meta name="citation_author" content="Шестопёров А.И.">
<meta name="citation_author" content="Ткачев С.С.">
<meta name="citation_publication_date" content="2018">
<meta name="citation_issue" content="5">
<meta name="citation_firstpage" content="1">
<meta name="citation_lastpage" content="17">
<meta name="citation_pdf_url"
      content="http://keldysh.ru/papers/2018/prep2018_5.pdf">
<meta name="citation_journal_title " content="Препринты ИПМ им.
      М.В.Келдыша ">
```

Теги метаинформации имеют следующие значения:

`citation_title` — название публикации;
`citation_author` — ФИО авторов публикации (каждый автор размещается в отдельном теге);
`citation_publication_date` — дата выпуска публикации (полная дата или только год);
`citation_issue` — номер публикации;
`citation_firstpage` — номер первой страницы;
`citation_lastpage` — номер последней страницы;

citation_pdf_url — url полного текста публикации;

citation_journal_title — название журнала, содержащего препринты.

Метаданные размещаются в карточке публикации на выбранном языке (русском или английском) её просмотра.

По соображениям безопасности Google Scholar требует, чтобы полный текст публикации находился в том же каталоге, что и карточка описания этой публикации [6]. Для выполнения этого требования на головном веб-сервере ИПМ в его конфигурационном файле используются те же правила, которые действуют для обслуживания sitemap (упомянутые ранее). Реально обращение по ссылке, указанной в citation_pdf_url, переадресуется к универсальной программе, размещенной на сайте <http://library/keldysh.ru/>, которая выводит карточку публикации.

6. Метаданные живой публикации

Препринты ИПМ позиционируются как живые публикации: после выхода препринта авторы вправе корректировать, т.е. улучшать и актуализировать его содержание. Работа с живыми публикациями требует введения в обращение еще одного метаатрибута — оперативно обновляемой даты редакции его текста. Этот метаатрибут сейчас достаточно рельефно представлен на сайте онлайн-библиотеки ИПМ (рис. 8).

Скриншот веб-страницы поиска публикаций на сайте библиотеки ИПМ. В верхней части страницы логотип библиотеки и заголовок «Найденные публикации». В центре — панель «РЕЗУЛЬТАТЫ ПОИСКА» с подсчетом найденных записей: 46. Ниже — навигационные кнопки: «В начало списка», «Предыдущая порция», «Следующая порция», «В конец списка». Основное содержимое — таблица с тремя результатами поиска.

№	Название	Авторы	Издательство	Дата редакции
11.	Препринт ИПМ № 14, Москва, 2008	Горбунов-Посадов М. М., Корягин А.Н.	ИПМ им.М.В.Келдыша, 2008.	Редакция от ≈ 06.05.2015 ≈
12.	Статья, Москва, 2008	Горбунов-Посадов М. М.	ИПМ им.М.В.Келдыша, 2008.	Редакция от ≈ 04.06.2018 ≈
13.	Статья, Москва, 2008	Горбунов-Посадов М. М.	ИПМ им.М.В.Келдыша, 2008.	Редакция от ≈ 04.06.2018 ≈

Рис. 8. Даты редакции живых публикаций в поисковой выдаче онлайн-библиотеки ИПМ

К сожалению, ни одна из упомянутых выше внешних систем, принимающих метаатрибуты публикаций, с обновляемыми датами работать не умеет, так что об их экспорте пока остается только мечтать.

Заключение

Электронная библиотека ИПМ имеет уже довольно длительную историю. Однако если раньше было достаточно размещать препринты и другие публикации на сайте ИПМ, то современная ситуация предъявляет к ведению сериальных изданий дополнительные требования. Необходимо обеспечить корректное индексирование каждой онлайн-публикации в популярных поисковых системах.

Предпринятые нами усилия в этой области привели к определенным положительным результатам — публикации ИПМ теперь находят и корректно отображают такие поисковики как Google Scholar, Yandex. Размещение метаданных публикаций в системе Crossref и получение DOI делает ссылки на публикации ИПМ более привлекательными. Передача в РИНЦ метаданных препринтов ИПМ и ссылок на их полные тексты усилила позиции нашего издания в научном интернет-сообществе.

Библиографический список

1. Горбунов-Посадов М.М. Препринты ИПМ им. М.В.Келдыша.
URL: <http://keldysh.ru/gorbunov/preprints.htm>
2. Полилова Т.А., Горбунов-Посадов М.М, Ермаков А.В., Симон А.А.
Препринты ИПМ им. М.В.Келдыша: новые реалии // Препринты ИПМ им. М.В.Келдыша. 2013. № 8. С. 1-27.
URL: <http://library.keldysh.ru/preprint.asp?id=2013-8>
3. Богданова В.М., Горбунов-Посадов М.М, Полилова Т.А., Слепенков М.И.
Диссертационные материалы в онлайн-библиотеке // Препринты ИПМ им. М.В.Келдыша. 2015. № 62.
URL: <http://library.keldysh.ru/preprint.asp?id=2015-62>
4. Научная электронная библиотека eLibrary. URL: <http://elibrary.ru>
5. The DOI System. URL: <https://www.doi.org/>
6. Sitemaps XML format. URL: <https://www.sitemaps.org/protocol.html>
URL: <https://scholar.google.com/intl/en/scholar/inclusion.html>
7. Google Scholar Inclusion Guidelines for Webmasters.
URL: <https://scholar.google.com/intl/en/scholar/inclusion.html>