

УДК 004

АНАЛИТИЧЕСКАЯ СТАТИСТИКА О НАУЧНЫХ ПУБЛИКАЦИЯХ КАЗАНСКОГО ФЕДЕРАЛЬНОГО УНИВЕРСИТЕТА НА SCILIT

А. В. Ермаков^[0000-0002-6054-0813]

*Институт прикладной математики им. М.В. Келдыша
Российской академии наук, г. Москва*

ermakov@keldysh.ru

Аннотация

Рассмотрены вопросы, связанные с представлением информации о публикациях научных сотрудников, преподавателей, аспирантов и студентов Казанского федерального университета (КФУ), а также научных изданиях Университета в информационно-аналитических материалах системы Scilit. На конкретных примерах показаны преимущества полного и корректного задания метаданных научных публикаций, а также возникающие проблемы при небрежном обращении с библиографической информацией.

Ключевые слова: *аналитическая статистика, библиографические базы данных, метаданные публикаций, Scilit.*

ВВЕДЕНИЕ

Последние 10 лет показатели публикационной активности российских ученых неуклонно растут. С точки зрения оценки показателей результатов научной деятельности безусловно важны индекс цитируемости публикаций и индекс Хирша. Но с точки зрения автора одним из наиболее важных показателей является интерес читателей (количество просмотров). Многие библиографические системы дают возможность автору увидеть различные статистические показатели не только в статике (общее число просмотров), но и в динамике (за день, месяц, в виде графиков и т. д.).

Работы по исследованию методов оценки веб-аналитики ведутся как в России, так и за рубежом. В российском сегменте это прежде всего проект Science Space от eLibrary [1].

Среди зарубежных следует отметить инструменты Crossref [2] – API от

Crossref (<https://www.crossref.org/documentation/retrieve-metadata/rest-api/>) и, конечно же, разработку швейцарско-китайского издательства MDPI (Multidisciplinary Digital Publishing Institute) – комплексной платформы-агрегатора научных публикаций Scilit.

Издательство MDPI, выпускающее порядка 390 журналов, часто находится на передовой инноваций в области цифровой публикации. Одним из примеров является Scilit [3] – база данных, которая ежедневно собирает, индексирует и извлекает новейшие данные и материалы из CrossRef и PubMed, обеспечивает поиск нужной информации среди 167 миллионов публикаций, 19 миллионов ученых со всего мира, 24 533 издательств, 1 миллиона научных организаций.

Платформа Scilit была разработана в 2014 году. Потенциал этой базы данных позволил достаточно быстро охватить все статьи с цифровым идентификатором объекта (DOI) из Crossref и с идентификатором PMID (Pubmed). Scilit анализирует метаданные Crossref и сохраняет их в своей информационной базе. Затем они индексируются в поисковой системе для быстрого использования в режиме реального времени.

В своих аналитических отчетах Scilit четко указывает, является ли материал открытым или нет, и какие издатели публикуют работы в формате, максимально доступном читателю. Поскольку академический издательский ландшафт смещается в сторону модели открытого доступа, это очень важный индикатор.

SCILIT – КОМПЛЕКСНАЯ ПЛАТФОРМА-АГРЕГАТОР ДЛЯ НАУЧНЫХ ПУБЛИКАЦИЙ

Главное меню Scilit (<https://www.scilit.net/>) предлагает выбрать формат представления аналитической информации (рис. 1):

- по публикации;
- по ученым (авторам);
- по издателям (издательствам);
- по источникам;
- по организациям.

Следует сразу отметить, что исходные материалы иллюстраций являются англоязычными, и мы использовали автоматический переводчик Google.

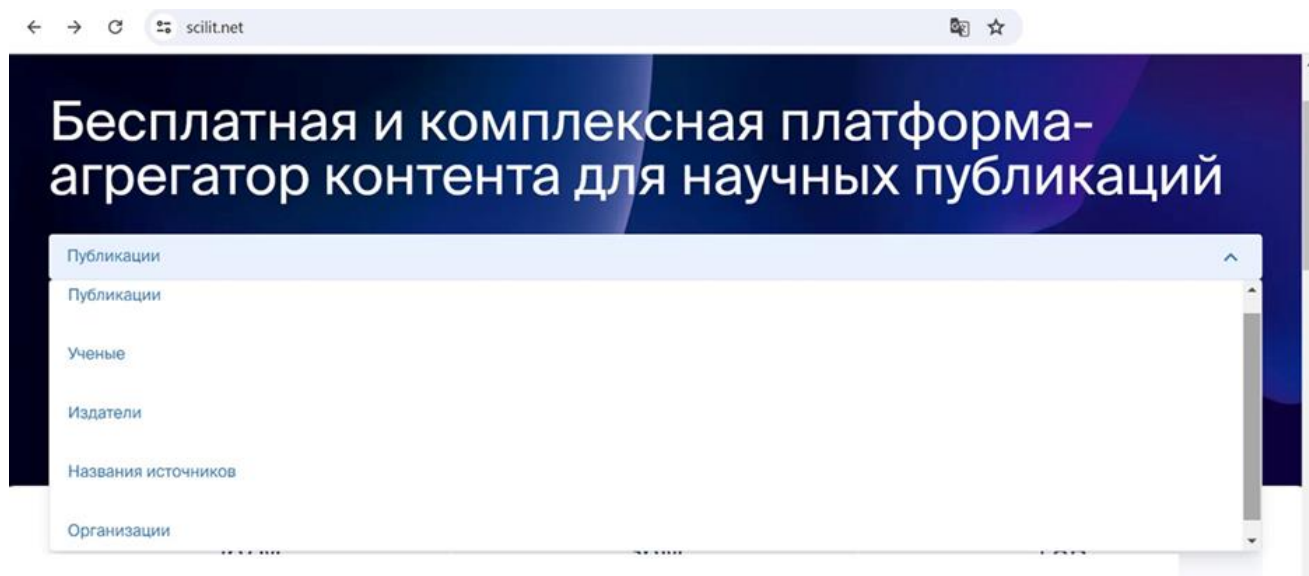


Рис. 1. Главное меню Scilit.

Безусловно, все предлагаемые направления интересны для изучения. Но на наш взгляд, проще и понятней начать рассмотрения с последнего пункта – «Организации».

Выбираем этот пункт и указываем название организации – KFU (Kazan Federal University).

Здесь следует отметить, что в метаданных параметр «ROR» (Research Organization Registry) появился сравнительно недавно, а до этого авторы не всегда точно указывали в статьях названия организаций, да и издатели далеко не всегда утруждали себя заданием корректной аффилиации авторов. В некоторых зарубежных и в очень малом количестве отечественных журналов этот параметр появился в начале 2020-х годов. А до этого использовалось текстовое название организации ({"affiliation":{"name":"Kazan Federal University"}}). Кто-то указывал полное название, кто-то сокращенное, а многие не указывали вовсе. В результате значительное количество публикаций авторов – преподавателей и сотрудников Казанского университета – не попало в статистику, приведенную на рис. 2.

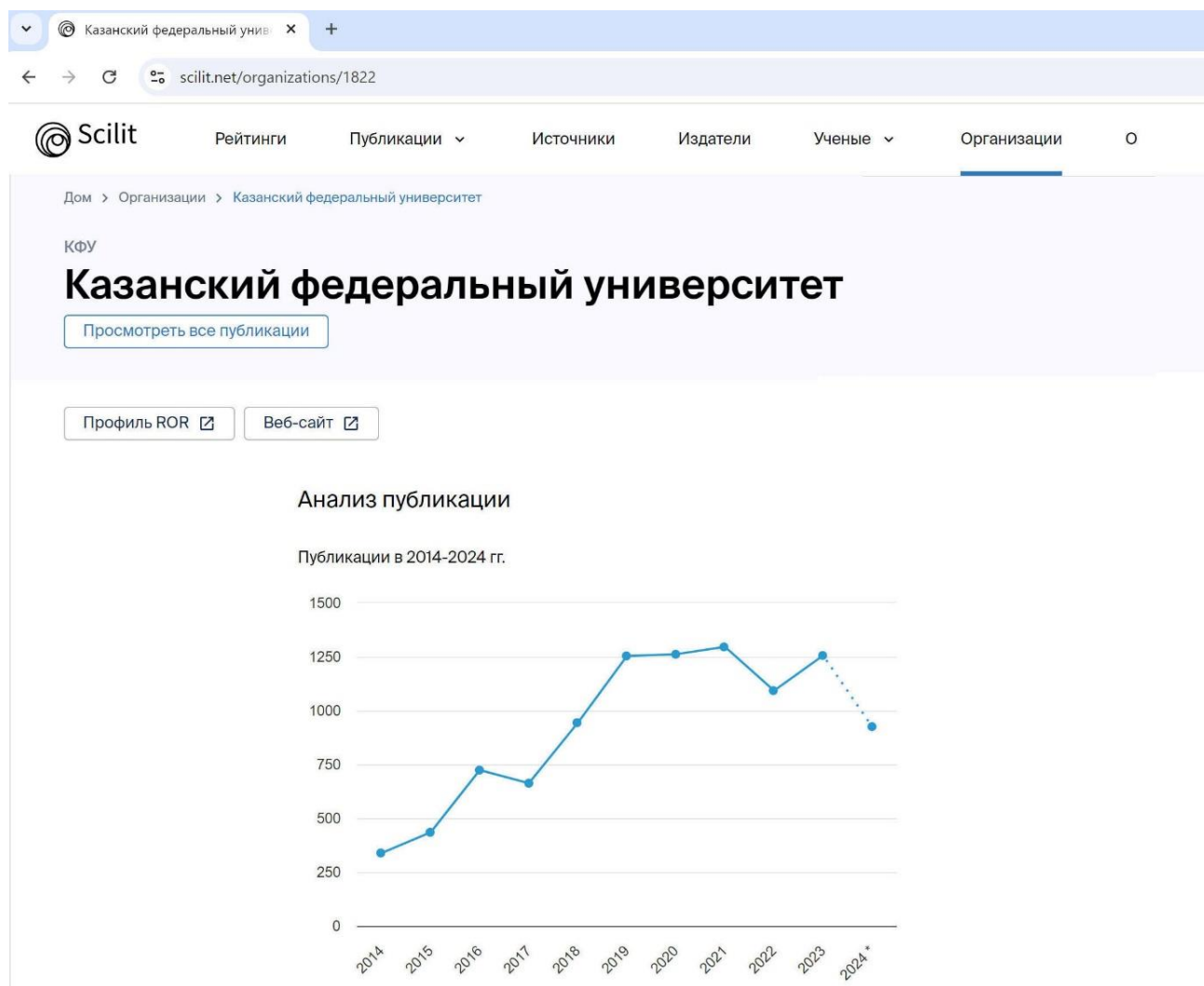


Рис. 2. Анализ публикационной активности сотрудников КФУ.

Как же получается, что авторы и редакции журналов заинтересованы в попадании своих публикационных материалов в отчеты, статистику, аналитику, но информационно-аналитические системы эти материалы не находят?

Первая причина – это небрежность авторов, опечатки в названии или использование транслитерации вместо корректных англоязычных названий своих организаций. Таким образом, некорректная аффилиация автора статьи зачастую автоматически переносится в метаданные публикации и отсылается в библиографические базы данных.

Вторая причина – это работа редакции и издательства.

Возьмем для примера статью в журнале Лобачевского (издательство Pleiades Publishing Ltd).

1 June 2024

Published by Pleiades Publishing Ltd in Lobachevskii Journal of Mathematics

Vol. 45 (6), 2810-2820

<https://doi.org/10.1134/s1995080224602923>

Estimates of Generalized St. Venant Functionals

by F. G. Avkhadiev

<https://doi.org/10.1134/s1995080224602923>

Способы получения метаданных из базы Crossref описаны в [4], здесь покажем эту функцию коротко:

<https://api.crossref.org/works/<DOI>>

В приведенном примере DOI – 10.1134/s1995080224602923, тогда запрос метаданных указанной статьи будет выглядеть так:

<https://api.crossref.org/works/10.1134/s1995080224602923>

Получаем ответ в формате JSON и контекстным поиском находим данные об аффилиации автора:

```
"author": [{"given": "F.G.", "family": "Avkhadiev"},  
"sequence": "first", "affiliation": []}]
```

Приведена «пустая» аффилиация, хотя на сайте Springer и в pdf-версии статьи указаны правильные данные аффилиации – Kazan Federal University, 420008, Kazan, Russia.

Другой пример не совсем корректного подхода к указанию аффилиации автора – издательство MDPI. Регистрируясь на сайте издательства [5], каждый автор заполняет шаблон (рис. 3, 4) – здесь я привожу фрагменты заполнения своего авторского профиля, так как невозможно получить доступ к шаблонам других авторов.

Рис. 3. Шаблон для регистрации автора в издательстве MDPI.

Рис. 4. Поля шаблона, связанные с аффилиацией.

Рассмотрим статью в журнале MDPI «International Journal of Molecular Sciences» (Международный журнал молекулярных наук).

17 July 2024

Published by MDPI AG in International Journal of Molecular Sciences

Vol. 25 (14), 7822

<https://doi.org/10.3390/ijms25147822>

Applications of Microbial Organophosphate-Degrading Enzymes to Detoxification of Organophosphorous Compounds for Medical Countermeasures against Poisoning and Environmental Remediation by Tatiana Pashirova, Rym Salah-Tazdaït, Djaber Tazdaït, Patrick Masson, Tatiana Pashirova, Rym Salah-Tazdaït, Djaber Tazdaït, Patrick Masson, Tatiana

Pashirova, Rym Salah-Tazdaït, et al.

В метаданных статьи указанное издательство в поле аффилиации записывает все обязательные поля своего регистрационного шаблона (рис. 4) – Affiliation, Address1, Zip Code, City, Country.

```
"author":[{"ORCID":"http://orcid.org/0000-0002-2001-1570"},  
"gien":"Tatiana","family":"Pashirova","sequence":"first","affiliation":{"name":"Institute  
of Fundamental Medicine and Biology, Kazan Federal University, 18 Kremlyovskaya St.,  
420008 Kazan, Russia"}].
```

Кстати, сотрудники MDPI, безусловно уважая автора, указывают и его ORCID (в отличие, например, от того же Pleiades Publishing Ltd), что однозначно «связывает» эту статью с конкретным автором.

Конечно, такое указание аффилиации избыточно, но это, безусловно, лучше, чем «пустая» аффилиация. Информационно-аналитическая система при обработке этой информации может узнать и показать читателям не только организацию автора, но и город, и страну проживания.

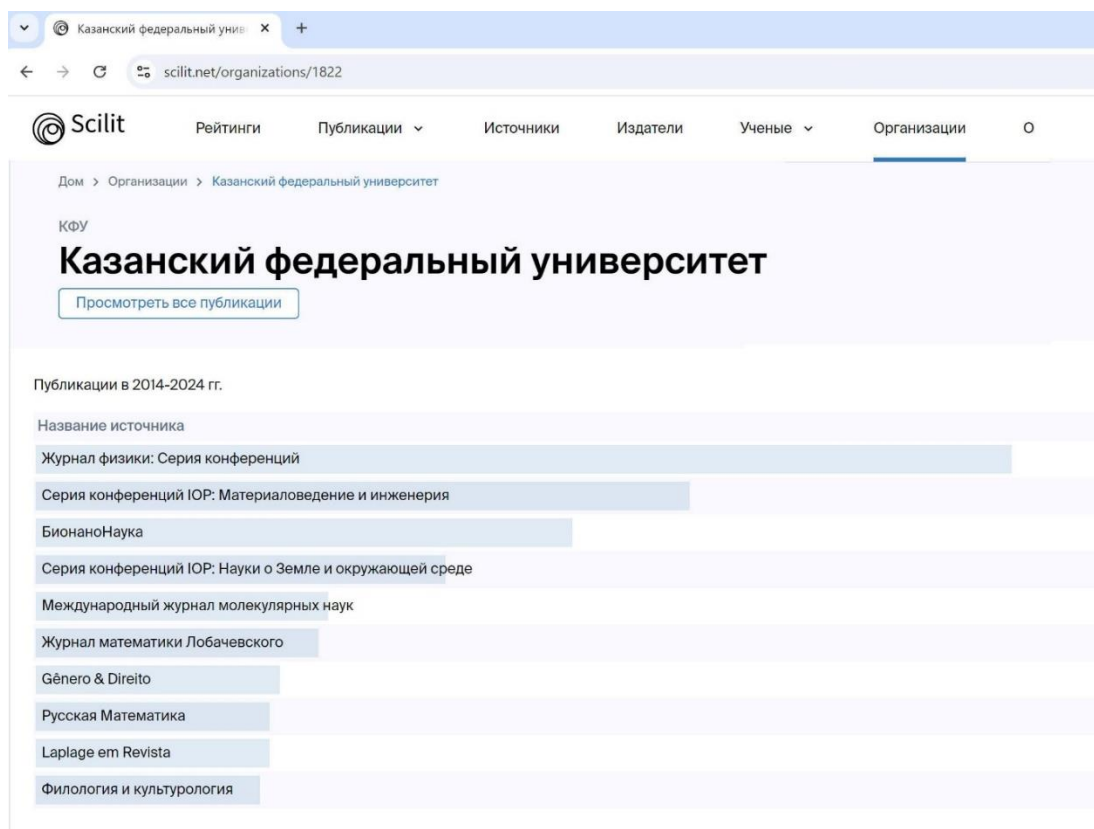


Рис. 5. Источники метаинформации о публикациях сотрудников КФУ.

На рис. 5 показаны «источники» метаинформации – издания, в которых публиковались статьи преподавателей и сотрудников КФУ и которые стали источником метаданных научных публикаций для приведенной статистики.

На рис. 6 показана «тематическая направленность» публикаций. Но в метаданных тематическое направление исследований не задается. На основе модели машинного обучения Scilit автоматически классифицирует публикации по различным темам. Модель классификации тем была обучена только на англоязычных публикациях и может привести к «слабым» результатам для публикаций на других языках (если тематическое направление вообще можно определить).

Таким образом, система сама, возможно по названиям статей или аннотациям, построила таблицу тематических направлений.

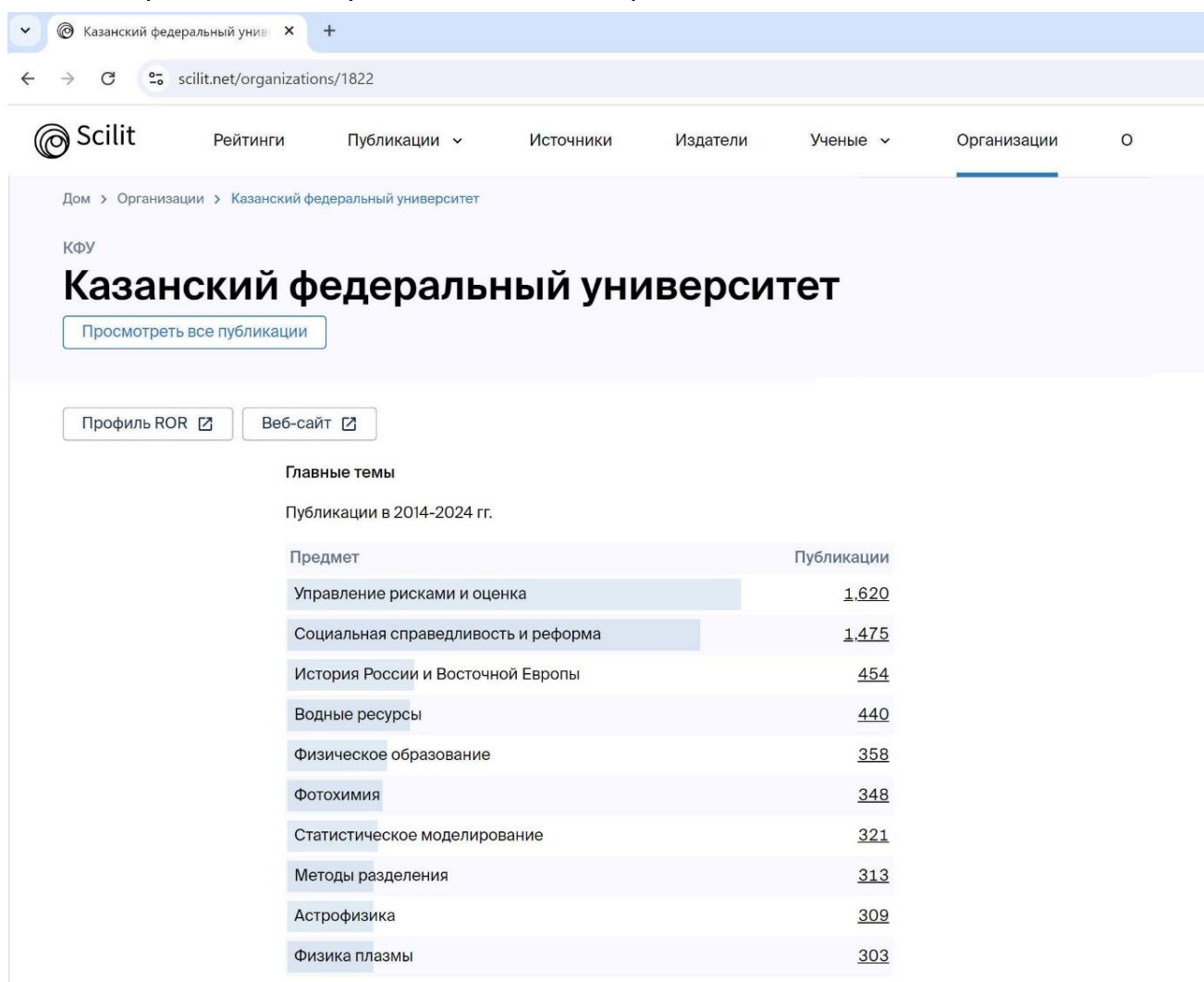


Рис. 6. Тематическая направленность публикаций сотрудников КФУ.

В целом это безусловно хорошо и удобно исследователю. Однако на текущий момент корректность отнесения той или иной публикации к определенной теме вызывает вопросы.

Это легко проверить – справа указано количество публикаций, отнесенное к данной теме. Это поле является «кликабельным», т. е. можно получить весь список статей.

Рассмотрим, например, публикации по первой теме указанного списка – «Управление рисками и оценка». Система ИИ отнесла к этой тематике статью про оборудование и системы управления на предприятиях целлюлозно-бумажной промышленности:

Intelligence of equipment and control systems at pulp and paper industry enterprises by Stanislav Gorobchenko, Dmitriy Kovalev, Sergey Voinash, Ramil Zagidullin, Ildar Khafizov, Taisiya Garbuzova, Svetlana Alekseeva

<https://doi.org/10.1051/e3sconf/202454803003>

Конечно, стремительное развитие систем ИИ в ближайшем будущем сможет исправить такого рода ошибки и даже начнет работать и с русскоязычными аннотациями статей. Но даже нынешнее положения, на наш взгляд, полезно и интересно.

Также следует обратить внимание на статистику по цитированию (рис. 7), которую можно получить за любой диапазон лет с начала наполнения базы данных Scilit (2014 год).

«Кликнув» на вопросительный знак «?» в верхней части рисунка, можно получить подробную информацию по каждой позиции данного статистического отчета.

На рис. 8 показаны данные по цитированию в графическом виде.

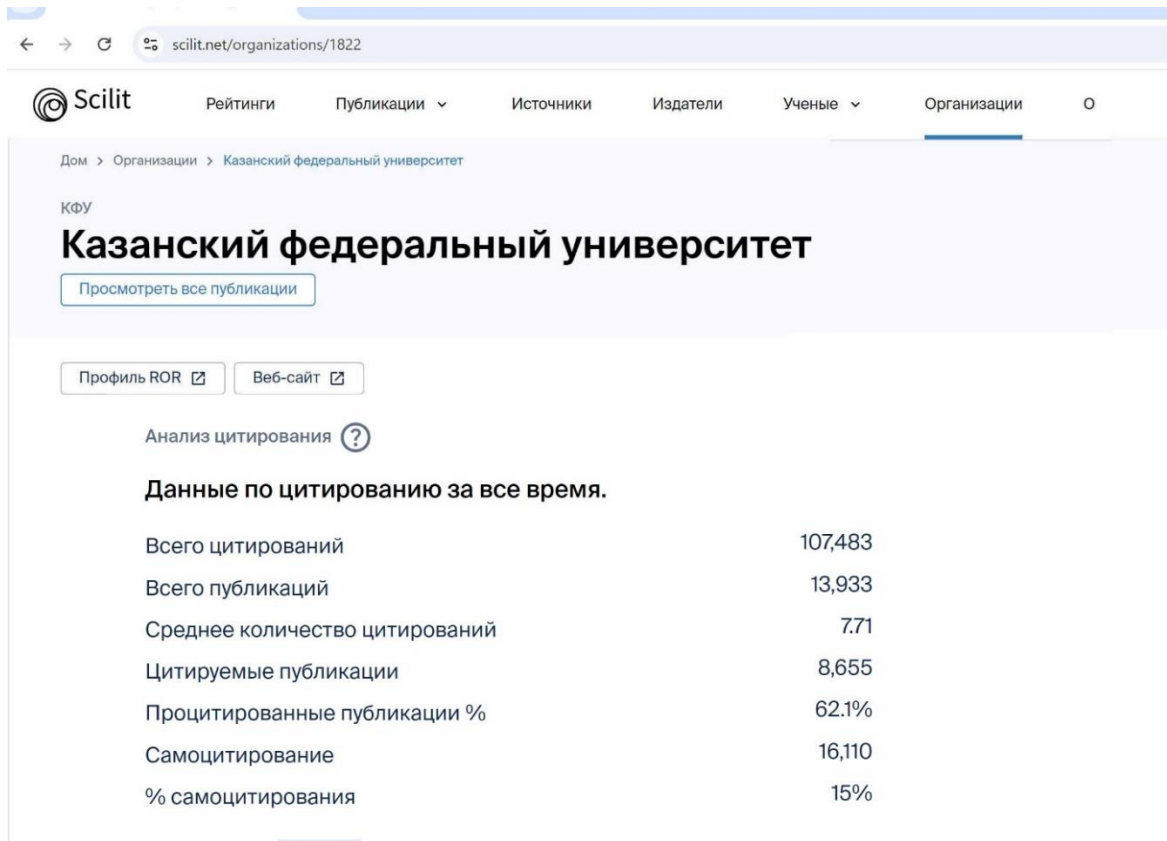


Рис. 7. Статистика по цитированию публикаций сотрудников КФУ.

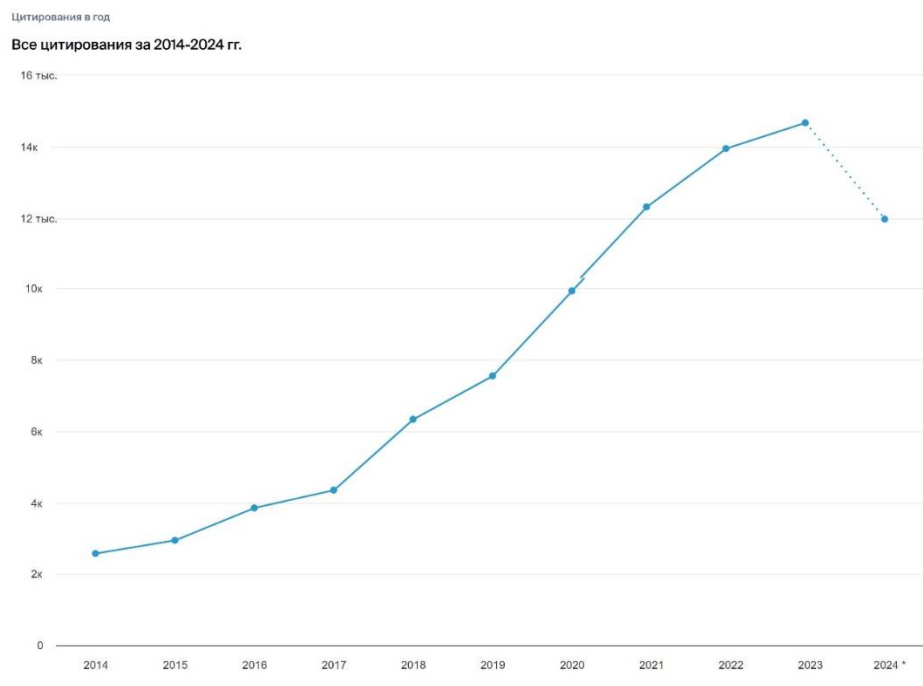


Рис. 8. Графическое представление статистики цитирования публикаций преподавателей и сотрудников КФУ.

Еще одним, интересным на наш взгляд, показателем, являются данные о сотрудничестве – соавторстве с сотрудниками других организаций (рис. 9).

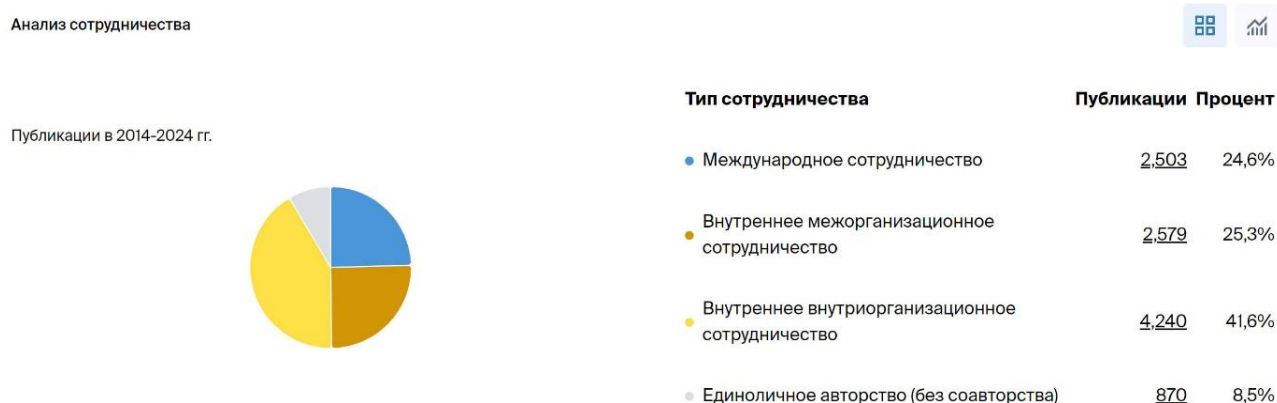


Рис. 9. Данные по научным публикациям, подготовленным в соавторстве с сотрудниками других организаций.

ИЗДАТЕЛЬСТВО КАЗАНСКОГО ФЕДЕРАЛЬНОГО УНИВЕРСИТЕТА

Теперь рассмотрим статистику публикаций, которые вышли в издательстве КФУ, – журналы, монографии, сборники конференций (рис. 10).

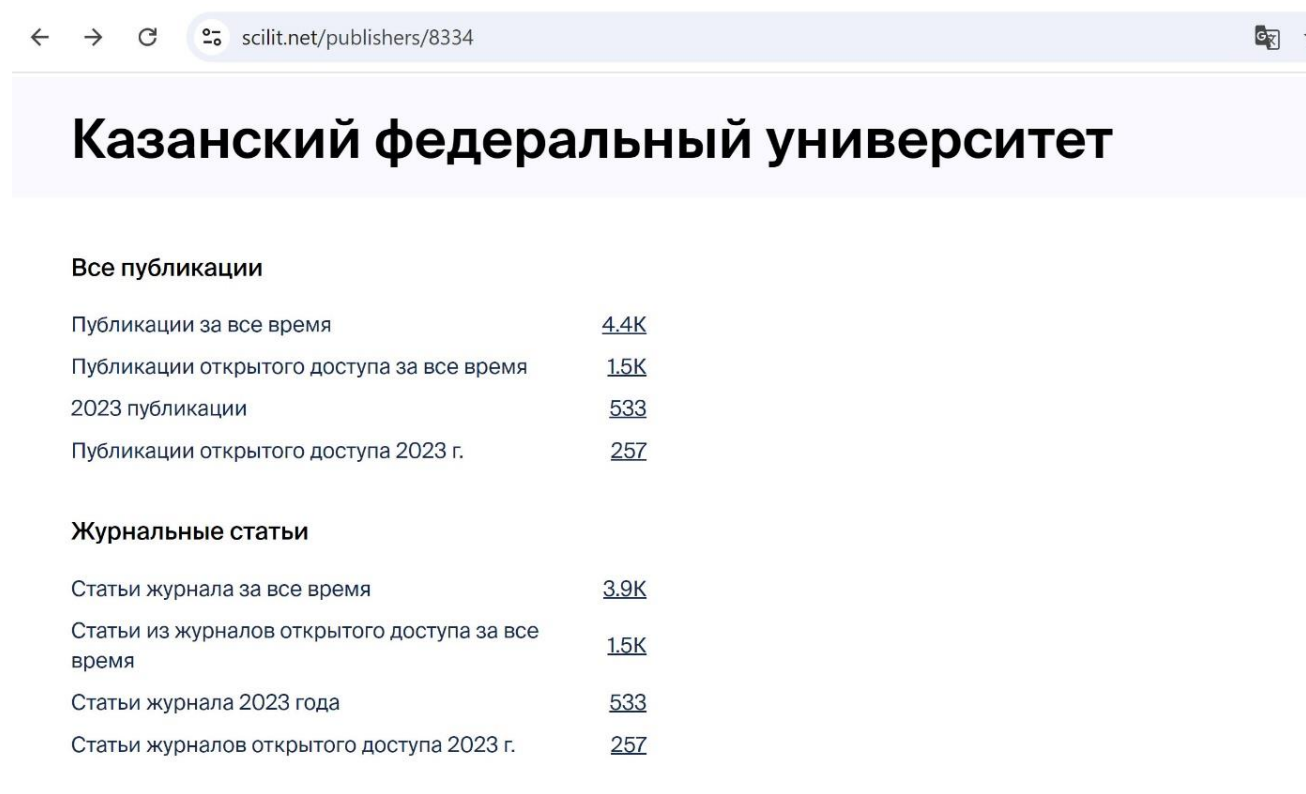


Рис. 10. Статистика по изданиям КФУ.

Далее на рис. 11 представлены издания КФУ (Топ-10).

Публикации в 2014–2024 гг.

Название источника	Публикации
Филология и культурология	782
Известия Высших Учебных Заведений. Математика	589
КАЗАНСКИЙ СОЦИАЛЬНО-ГУМАНИТАРНЫЙ ВЕСТНИК	537
Ученые записки Казанского университета. Серия Гуманитарные Науки	438
Образование и саморазвитие	395
Журнал «Российские электронные библиотеки»	248
Ученые записки Казанского университета. Серия Естественные науки	236
КАЗАНСКИЙ ЛИНГВИСТИЧЕСКИЙ ЖУРНАЛ	190
Ученые записки Казанского университета. Серия Физико-Математические Науки	180
Татарика	153

Рис. 11. Издания КФУ.

По каждому из журналов, приведенных в этой иллюстрации, можно провести отдельное интересное исследование, но мы в следующем разделе ограничимся журналом «Российские электронные библиотеки».

А здесь следует обратить внимание на актуальность представленной информации. Мы уже отмечали, что правое поле в такого рода таблицах Scilit (количество публикаций) является кликабельным, и если кликнуть эту позицию журнала «Филология и культура» (Google переводит «Philology and Culture» как «Филология и культурология»), то получим список всех публикаций данного журнала, начиная с самой последней:

24 октября 2024 г.

Опубликовано Казанский федеральный университет в Филология и культурология № 3, стр. 336–341

<https://doi.org/10.26907/2782-4756-2024-77-3-336-341>

Взаимосвязь и взаимообусловленность развития коммуникативно-речевой и лингвокультурологической компетенций при обучении русскому языку как иностранному

Л. Сянъюй, Л. Москалёва, Л. Сянъюй

Обратите внимание на дату публикации – 24 октября 2024 г. (написано 29 октября 2024 г.), т. е. актуальность информации, представленной Scilit, очень высокая, несмотря на терабайтные объемы базы данных этой системы.

Безусловно интересна тематическая направленность публикаций в изданиях КФУ (рис. 12).

Главные темы

Публикации в 2014–2024 гг.

Предмет	Публикации
История России и Восточной Европы	593
Социальная справедливость и реформа	575
Диагностическое и терапевтическое УЗИ	508
Динамика океана	437
Управление рисками и оценка	381
Физика плазмы	263
Современная история	258
Лингвистика и языкознание	231
Статистическое моделирование	228
Наука о культуре	216

Рис. 12. Темы публикаций в изданиях КФУ.

Эта тематическая таблица также построена при помощи ИИ, и к ней также возникают вопросы. Но, скорее всего, с развитием модели ИИ эти вопросы будут постепенно сниматься.

Однако здесь необходимо сделать одно важное замечание. Если журнал не размещает в метаданных статей англоязычные аннотации, то в аналитических обзорах Scilit эти материалы не будут отражены в тематических рубриках.

Безусловно, интересной является статистика по сотрудничеству ученых КФУ с зарубежными специалистами (рис. 13).

Публикации в 2014–2024 гг.

Страна	Публикации
Россия	865
Узбекистан	30
Германия	12
Беларусь	9
Турция	9
Грузия	8
Китай	7
Соединенные Штаты	6
Азербайджан	5
Канада	4

Рис. 13. Сотрудничество с зарубежными специалистами.

В качестве примера рассмотрим список публикаций узбекских коллег, «кликнув» соответствующее поле в столбце «Публикации»:

6 октября 2024 г.

Опубликовано Казанский федеральный университет в Ученые записки Казанского университета. Серия Физико-Математические Науки Том 166 (3), 437–449

<https://doi.org/10.26907/2541-7746.2024.3.437-449>

Положительные неподвижные точки интегральных операторов Гаммерштейна с вырожденным ядром

Ю.Х. Эшкабилов, Ш.Д. Нодиров, Ю.Х. Эшкабилов

В метаданных публикации нет указания страны, но есть аффилиация авторов:

```
"author":{"given":"Yu.Kh.", "family":"Eshkabilov",  
"sequence":"first", "affiliation":{"name":"Tashkent International University of Financial Management and Technologies"}}
```

И система ИИ по названию университета определила гражданство данного автора. Конечно, это не всегда корректно, но нам чаще всего важны не паспортные данные, а место работы автора.

ЖУРНАЛ «РОССИЙСКИЕ ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ»

Анализ статистики по изданиям Казанского Университета покажем на примере журнала «Российские электронные библиотеки». Хотя все уже привыкли называть его просто «Электронные библиотеки», и на сайте rdl-journal.ru указано привычное нам название, однако в метаданных, загружаемых в Crossref, задано: "container-title":["Russian Digital Libraries Journal"]

Именно под этим названием информационно-аналитическая система Scilit видит этот журнал, и именно его зададим в поле ввода на главной странице системы.

На рис. 14 представлена графическая информация за все годы сотрудничества этого журнала с Crossref (метаданные «Электронных библиотек» стали загружаться в Crossref с 2019 года).

Журнал «Российские электронные библиотеки»

Опубликовано [Казанский федеральный университет](#)

EISSN 1562-5419

Охват базы данных

Сцилит

Журнальные статьи

Статьи журнала за все время

248

Статьи журнала 2023 года

6



Рис. 14. Статистический анализ по публикациям журнала «Российские электронные библиотеки».

Резкое падение числа публикаций на графике в районе 2023–2024 гг. связано с тем, что DOI опубликованным статьям присваивались, но метаданные в Crossref не пересылались. Это можно увидеть, пролистнув аналитические таблицы и графики Scilit до раздела «Последние публикации». Последняя опубликованная с точки зрения Scilit статья:

20 января 2023 г.

Опубликовано Казанский федеральный университет в Журнал «Российские электронные библиотеки» Том 25 (6), 600–615

<https://doi.org/10.26907/1562-5419-2022-25-6-600-615>

Использование графа соавторства для тематического поиска конференций по наукометрическим данным

Александр Сергеевич Козицын, Сергей Александрович Афонин, Дмитрий Алексеевич Шачнев

Несмотря на то, что в метаданных журнала систематически размещаются англоязычные аннотации, тематический каталог выглядит красиво, но очень странно (рис. 15). Если посмотреть публикации, соответствующие каждому разделу, то «тематических совпадений» будет очень немного.

Публикации в 2014–2024 гг.

Предмет	Публикации
Социальная справедливость и реформа	<u>113</u>
Управление рисками и оценка	<u>107</u>
Дистанционное обучение и исследования	<u>47</u>
Наукометрия и исследовательская этика	<u>28</u>
Разработка программного обеспечения	<u>26</u>
Информатика и библиотечное дело	<u>22</u>
Физическое образование	<u>22</u>
Управление операциями	<u>21</u>
Взаимодействие человека и машины	<u>20</u>
Образование и педагогика	<u>18</u>

Рис. 15. Тематический каталог журнала «Российские электронные библиотеки»

Другие разделы аналитического исследования, представленного системой Scilit, также представляют определенный интерес, но они уже были рассмотрены нами в предыдущих разделах статьи.

ЗАКЛЮЧЕНИЕ

В данной работе мы не ставили задачу максимально подробно рассказать о всех возможностях информационно-аналитической системы Scilit. Нам было важно показать на конкретных примерах некоторые, наиболее интересные, на наш взгляд, ее возможности. Показать, какие преимущества дает аккуратное и максимально полное задание метаданных научных публикаций и какие проблемы возникают при небрежном отношении авторов и/или издателей к этой информации.

СПИСОК ЛИТЕРАТУРЫ

1. Информационно-аналитическая система SCIENCE SPACE.
URL: https://elibrary.ru/projects/science_space/science_space_info.asp
2. API унифицированных ресурсов CrossRef.
URL: <https://api.crossref.org/swagger-ui/index.html>
3. Платформа-агрегатор для научных публикаций Scilit.
URL: <https://www.scilit.net>
4. *Ермаков А.В.* Некоторые проблемы при регистрации даты публикации в Crossref // Научный сервис в сети Интернет: труды XXIV Всероссийской научной конференции (19–22 сентября 2022 г., онлайн). М.: ИПМ им. М.В. Келдыша, 2022. С. 179–182. <https://doi.org/10.20948/abrau-2022-21>,
URL: <https://keldysh.ru/abrau/2022/theses/21.pdf>
5. Регистрационная форма автора в издательстве MDPI.
URL: <https://susy.mdpi.com/user/edit>

ANALYTICAL STATISTICS ABOUT SCIENTIFIC PUBLICATIONS OF THE KAZAN FEDERAL UNIVERSITY ON SCILIT

Alexey V. Ermakov^[0000-0002-6054-0813]

Ermakov@Keldysh.ru

Keldysh Institute of Applied Mathematics, Moscow, Russia

Abstract

The paper examines issues related to the presentation of information about publications of KFU researchers, teachers, graduate students and students, as well as about the University's scientific sources in information and analytical materials of the Scilit system. Specific examples show the advantage of complete and correct setting of metadata for scientific publications, as well as the problems that arise when handling bibliographic information carelessly.

Keywords: *analytical statistics, bibliographic databases, publication metadata, Scilit.*

REFERENCES

1. Information and analytical system SCIENCE SPACE.
URL: https://elibrary.ru/projects/science_space/science_space_info.asp
2. Crossref Unified Resource API.
URL: <https://api.crossref.org/swagger-ui/index.html>
3. Content aggregator platform for scholarly publications.
URL: <https://www.scilit.net>
4. *Ermakov A.V.* Some problems when registering publication date in Crossref // Conference material: "Scientific service & Internet: proceedings of the 24th All-Russian Scientific Conference (September 19-22, 2022, online)".
<https://doi.org/10.20948/abrau-2022-21>,
URL: <https://keldysh.ru/abrau/2022/theses/21.pdf>
5. Registration form of the author in the publishing house MDPI.
URL: <https://susy.mdpi.com/user/edit>

СВЕДЕНИЯ ОБ АВТОРЕ



ЕРМАКОВ Алексей Викторович – старший научный сотрудник Институт прикладной математики им. М.В. Келдыша Российской академии наук

Alexey Viktorovich ERMAKOV – Senior Researcher, Keldysh Institute of Applied Mathematics.

email: ermakov@keldysh.ru,

ORCID: 0000-0002-6054-0813

Материал поступил в редакцию 28 октября 2024 года