ЭЛЕКТРОННЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ

УДК 025.2 – 028.27 + 001.83 – 047.44
https://doi.org/10.33186/1027-3689-2025-11-182-202

Применение метода «Exploratory Data Analysis» в библиометрии: на примере научных журналов из «Белого списка»

Н. А. Моисеева

Омский государственный технический университет,
Омск, Российская Федерация,
nat_lion@mail.ru, https://orcid.org/0000-0002-9502-3891

Аннотация. Уровень научного журнала – ключевой показатель в современной национальной системе оценки результативности исследований. Категорирование научных журналов – одна из основных задач библиометрии, постоянный рост объёма данных об изданиях и публикациях требует эффективного анализа соответствия между декларируемыми и наблюдаемыми паттернами категорирования. Цель статьи заключается в апробации исследовательского анализа данных (Exploratory Data Analysis, EDA) для изучения распределения научных журналов по уровням в «Белом списке» и выявления статистических закономерностей, связывающих уровень издания с его индексацией в наукометрических базах данных (БД). Объект исследования – открытые данные о научных журналах из «Белого списка». В работе применён статистический подход к анализу данных об изданиях, реализованный в цифровой среде Google Colaboratory с использованием библиотек для EDA (Pandas, Matplotlib, Seaborn) на языке программирования Python. Результаты позволили провести количественный анализ соответствия эмпирических данных правилам категорирования изданий «Белого списка» и выявить различия в паттернах индексации журналов разных уровней. Практическая значимость заключается в интеграции цифровых технологий EDA в библиометрический инструментарий, что открывает возможности для верификации систем категорирования и развития методов библиометрии в условиях цифровизации науки.

Ключевые слова: цифровизация, библиометрия, научный журнал, категорирование, «Белый список», наука о данных, исследовательский анализ данных

Для цитирования: Моисеева Н. А. Применение метода «Exploratory Data Analysis» в библиометрии: на примере научных журналов из «Белого списка» // Научные и технические библиотеки. 2025. № 11. С. 182–202. https://doi.org/10.33186/1027-3689-2025-11-182-202

DIGITAL INFORMATION RESOURCES

UDC 025.2 – 028.27 + 001.83 – 047.44
https://doi.org/10.33186/1027-3689-2025-11-182-202

Application of Exploratory Data Analysis in bibliometrics: A case study of scientific journals on the “White List”

Natalya A. Moiseeva

Omsk State Technical University, Omsk, Russian Federation, nat_lion@mail.ru, https://orcid.org/0000-0002-9502-3891

Abstract. In the context of scientific and technological progress, the level of scientific journal is a key element of the modern national system for assessing research performance. Categorizing scientific journals is one of the primary tasks of bibliometrics, and the constant growth in the volume of data on publications and journals requires credible analysis of the alignment between declared and observed categorization patterns. The purpose of this article is to test the application of Exploratory Data Analysis (EDA) to study the distribution of scientific journals across the “White List” levels and to identify statistical patterns linking a journal's level to its indexing in scientometric databases (DB). The research subject is open data on scientific journals on the “White List”. The study employs statistical approach to analyzing publication data, implemented in the Google Collaboratory digital environment using Python programming language libraries for EDA (Pandas, Matplotlib, Seaborn). The results enabled a quantitative analysis of the alignment between empirical data and the “White List” journal categorization rules, revealing differences in indexing patterns across different journal levels. The practical significance lies in integrating EDA digital technologies into the bibliometric toolkit, opening

opportunities for verifying categorization systems and advancing bibliometric methods amid the digitalization of science.

Keywords: digitalization, bibliometrics, scientific journal, “White List”, Data science, Exploratory data analysis (EDA)

Cite: Moiseeva N. A. Application of Exploratory Data Analysis in bibliometrics: A case study of scientific journals on the “White List” // Scientific and technical libraries. 2025. No. 11, pp. 182–202. https://doi.org/10.33186/1027-3689-2025-11-182-202

Введение

Компетентная оценка научных исследований и инновационных разработок в условиях цифровой экономики подразумевает несколько ключевых аспектов: 1) публикация информации о теоретической и практической значимости исследований; 2) освещение использования результатов для развития научных знаний; 3) описание процесса внедрения разработок в промышленные образцы и их интеграцию в образовательные процессы при подготовке специалистов [**1, 2**]. Эти элементы формируют комплексную систему, способствующую пониманию вклада науки в развитие цифровой экономики. При этом ключевым каналом распространения научных знаний выступают авторитетные журналы, индексируемые в специализированных БД.

В рамках системы оценки науки категоризация научных журналов является одной из ключевых задач библиометрии, обеспечивающей объективную оценку исследовательской деятельности. В частности, в России эту задачу выполняет «Белый список»[1], который включает авторитетные научные журналы и присваивает им один из четырёх уровней[2]. Важным шагом в развитии системы стало обновление платформы eLIBRARY.RU в конце 2024 г.: была улучшена визуализация данных о

---

[1] Межведомственная рабочая группа Минобрнауки РФ утвердила Методику категорирования российских и международных научных изданий «Белого списка» по уровням российских и международных научных изданий «Белого списка». URL: https://podpiska.rcsi.science/storage/202305_metodika.pdf

[2] Утверждены правила распределения по категориям научных изданий «Белого списка». URL: https://minobrnauki.gov.ru/press-center/news/nauka/68029/

«Белом списке», включая новые атрибуты, такие как «Белый список» (да/нет) и «Уровень в Белом списке» (У1–У4), а также добавлена возможность фильтрации публикаций по уровню журнала. Кроме того, на веб-сайте появилась ссылка для доступа к интерфейсу программирования приложений (Application Programming Interface, API) eLIBRARY[3], который предназначен для автоматических запросов и извлечения библиометрических показателей авторов и журналов, а также библиографических записей публикаций из БД РИНЦ. Используя библиометрический анализ на основе извлечённых данных, можно получить необходимую аналитику, что позволит интегрировать результаты публикационной активности профессорско-преподавательского состава в систему оценки научной деятельности университетов.

Анализ зарубежного опыта [**3–5**] показывает, что во многих странах также формируются белые (или избранные) списки и осуществляется категорирование изданий. Например, в странах Северной Европы это основано на экспертной оценке специалистов. В Индии [**4, 5**] существует избранный список журналов, однако он требует пересмотра из-за недостаточной эффективности действующих критериев отбора. Ряд российских авторов [**6–8**] отмечает, что разработка иерархии научных журналов в рамках «Белого списка» по-прежнему вызывает обсуждения, а состав и методы формирования «Белого списка» остаются предметом активных дискуссий в научном сообществе.

Актуальная версия «Белого списка», включающего около 30 тыс. научных журналов, доступна на веб-сайте Российского центра научной информации (РЦНИ). Карточки изданий регулярно обновляются с добавлением новых наукометрических показателей (например, индексация в наукометрической БД), формируя комплексную эмпирическую базу [**9**]. Учитывая необходимость регулярного мониторинга соответствия правилам категорирования в условиях динамично обновляемых данных, в настоящей работе EDA[4] предлагается в качестве современ-

---

[3] Интерфейс программирования приложений API. URL: https://www.elibrary.ru/projects/api/api_info.asp

[4] Исследовательский анализ данных (EDA) предполагает поиск и всестороннее исследование скрытой в данных информации для изучения их структуры и содержания. EDA позволяет визуализировать, обобщать и интерпретировать данные, что способствует выявлению закономерностей, аномалий и взаимосвязей [**10, 11**].

ного инструмента библиометрии. Это обусловлено способностью EDA выявлять скрытые статистические закономерности, визуализировать распределения и корреляции в больших массивах данных.

Таким образом, цель статьи заключается в изучении данных о научных журналах с применением статистических методов EDA и соответствующих фреймворков, реализованных в цифровой среде Google Colaboratory на языке программирования Python. Основное внимание уделяется анализу распределения научных журналов по уровням и верификации правил категорирования изданий в «Белом списке» через исследование взаимосвязей между уровнем журнала и его наукометрическими показателями (индексация в Scopus, Web of Science, RSCI).

EDA как один из важных методов Data Science

Вопросы о формировании новых подходов и применении инновационных инструментов в библиометрических исследованиях активно исследуются такими авторами, как Н. Б. Баканова, Р. С. Гиляревский, С. С. Захарова, О. В. Сюнтюренко, H. Maliha, J. A. Moral-Muñoz, E. Herrera-Viedma, A. Santisteban-Espejo, V. Pessin, L. Vilker, L. Yamane, C. N.-L. Tan, M. Fauzi, M. и др. [**2, 12–20**].

В научной статье [**19**] для выполнения библиометрического анализа используется библиометрический цифровой инструмент VOSviewer[5]. В другой работе авторы [**20**] применяют алгоритмы глубокого обучения для классификации документов на основе прямого цитирования. Переходя к современным решениям, в статье [**18**] описывается Smart Bibliometrics как интегрированная облачная методология, объединяющая классический библиометрический анализ, научное картирование и методы Data Science (обработка данных, машинное обучение, интерактивная визуализация) в единую автоматизированную систему. Технологически это переход от традиционных библиометрических инструментов к облачным интеллектуальным решениям.

---

[5] VOSviewer – программное обеспечение для построения и визуализации наукометрических сетей на основе алгоритмов Data Science. Преобразует данные о цитированиях, соавторстве или терминах публикаций в интерактивные карты науки (Science Mapping), используя кластеризацию, сетевую визуализацию, NLP для интеллектуальной обработки текста. URL: https://www.vosviewer.com/

Таким образом, на сегодняшний день особое внимание уделяется интеграции инструментов Data Science (DS)[6] в библиометрические исследования. Ярким примером являются аналитическая платформа «EC3metrics»[7] и курс «Craft Bibliometrics through Data Science»[8], представляющий практическое введение в методы DS и их интеграцию в библиометрию. Примечательно, что центральное место в курсе отводится EDA, который пронизывает все модули этой программы. Это объясняется тем, что EDA позволяет исследователям не только понять структуру и содержание своих данных, но и подготовить их к дальнейшему анализу.

EDA является ключевым инструментом для специалистов в области обработки данных, позволяя им анализировать и исследовать данные, а также обобщать их основные характеристики. Этот подход занимает центральное место в DS на этапе анализа данных [**21**]. В процессе EDA исследователь имеет возможность: 1) обобщить структуру и содержание данных; 2) оценить качество данных, выявляя пропуски, ошибки и аномалии; 3) обнаружить скрытые зависимости в больших объёмах информации [**10, 11**]. Эти задачи включают в себя проверку качества данных, расчёт статистических показателей, визуализацию и применение более сложных методов анализа, таких как анализ главных компонентов или корреляционный анализ. Завершив этап EDA и сделав выводы, исследователь может использовать полученные ре-

---

[6] Data Science – это область на стыке статистики, информатики и предметной экспертизы, которая фокусируется на сборе, обработке, анализе, интерпретации и визуализации больших объёмов данных [**10**]. Ключевым инструментом современной DS являются методы ИИ, особенно машинное обучение и глубокое обучение. Эти методы позволяют автоматически выявлять сложные закономерности, строить прогнозные модели, классифицировать информацию, генерировать новые данные и принимать решения на основе данных [**11**]. Наиболее популярными языками программирования в области DS, применяемыми для реализации EDA, являются Python и язык статистического программирования R [**22, 23**].

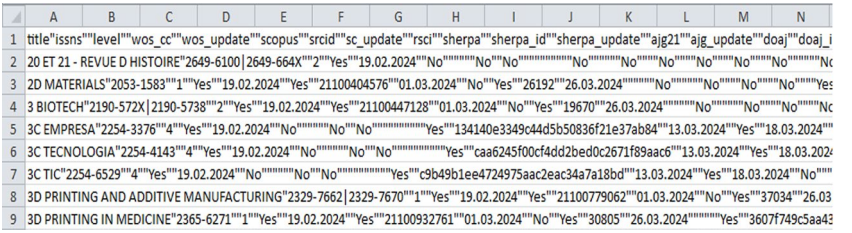
[7] EC3metrics -- это платформа, специализирующаяся на оценке научного и инновационного потенциала исследовательских проектов, библиометрии и альтметриии, а также научное картирование коллабораций (https://ec3metrics.com/home).

[8] Курс «Craft Bibliometrics through Data Science». URL: https://ec3metrics.com/courses/craft-bibliometrics-through-data-science

зультаты для более глубокого анализа данных, включая методы машинного обучения.

Рассмотрим примеры применения EDA. Коллектив авторов [**24**] использовал EDA для изучения данных о психических расстройствах, другие исследователи [**25**] применяли EDA для анализа и визуализации паттернов, трендов и аномалий в глобальных киберугрозах с целью проверки гипотез о доминировании атак DDoS и фишинга. В контексте библиометрии можно привести следующие примеры.

В данном исследовании EDA применялся для анализа датасета journalrank.csv, содержащего данные о научных журналах из «Белого списка». Датасет был получен с веб-сайта РЦНИ[9], где также доступен дополнительный файл формата CSV с метаданными научных журналов [**9**]. Исследуемый датасет включает 29554 записи и 96 признаков. Каждая запись описывает научный журнал, используя соответствующие метаданные и дополнительные данные, связанные с его индексацией в специализированных БД (см. табл.). На рис. 1 представлен фрагмент датасета до проведения предобработки.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	title"issns""level""wos_cc""wos_update""scopus""srcid""sc_update""rsci""sherpa""sherpa_id""sherpa_update""ajg21""ajg_update""doaj""doaj_i													
2	20 ET 21 - REVUE D HISTOIRE"2649-6100|2649-664X""2""Yes""19.02.2024""No""""""No""No""""""""""No""""""No""""No""""No""""No""""""No													
3	2D MATERIALS"2053-1583""1""Yes""19.02.2024""Yes""21100404576""01.03.2024""No""Yes""26192""26.03.2024""""""No""""""No""""No""""No""""Yes													
4	3 BIOTECH"2190-572X|2190-5738""2""Yes""19.02.2024""Yes""21100447128""01.03.2024""No""Yes""19670""26.03.2024""""""No""""""No""""No""""No													
5	3C EMPRESA"2254-3376""4""Yes""19.02.2024""No""""""No""No""""""""""Yes""134140e3349c44d5b50836f21e37ab84""13.03.2024""Yes""18.03.2024""													
6	3C TECNOLOGIA"2254-4143""4""Yes""19.02.2024""No""""""No""No""""""""""Yes""caa6245f00cf4dd2bed0c2671f89aac6""13.03.2024""Yes""18.03.2024													
7	3C TIC"2254-6529""4""Yes""19.02.2024""No""""""No""No""""""""""Yes""c9b49b1ee4724975aac2eac34a7a18bd""13.03.2024""Yes""18.03.2024""No""													
8	3D PRINTING AND ADDITIVE MANUFACTURING"2329-7662|2329-7670""1""Yes""19.02.2024""Yes""21100779062""01.03.2024""No""Yes""37034""26.03													
9	3D PRINTING IN MEDICINE"2365-6271""1""Yes""19.02.2024""Yes""21100932761""01.03.2024""No""Yes""30805""26.03.2024""""""Yes""3607f749c5aa43													

Рис. 1. Фрагмент исследуемого датасета

Во время предобработки датасета было установлено, что отсутствуют дубликаты и пропущенные значения. Для удобства последующего анализа данные были предварительно обработаны: приведены к корректным типам, а значения признаков, связанных с индексацией в специализированных БД, унифицированы. Часть категориальных переменных преобразована в бинарный формат (True/False заменены на 1 и 0 соответственно).

---

[9] Актуальная версия «Белого списка» публикуется на веб-сайте РЦНИ в файле формата CSV. URL: https://journalrank.rcsi.science/ru/record-sources/

Для систематизации анализа в таблице представлены ключевые группы специализированных БД с кратким пояснением по каждой из них. В этих БД индексируются издания из «Белого списка», названия которых совпадают с признаками в исследуемом датасете. Группы БД выделены по критерию их специализации и целевой научной аудитории. Данная классификация способствует систематизации БД по их функциональности и области применения. Эта классификация помогает систематизировать их функциональные особенности, что позволяет оценить авторитетность научного журнала и сформировать его профиль.

Классификация БД, в которых индексируются научные журналы, включённые в «Белый список»

Наименование БД	Краткая характеристика
Научные БД: ориентированы на предоставление доступа к научной информации, включая статьи, книги, диссертации; используются учёными и исследователями	*Research Citation Index (RSCI):* научные публикации и информация об их цитировании в научных журналах
	Scopus: крупнейшая в мире библиографическая и реферативная БД; содержит инструмент анализа цитирования и метрик
	WoS (Web of Science): объединяет несколько библиографических и реферативных БД рецензируемых научных работ; основана на индексе цитирования научных работ, разработанном одним из основателей наукометрии Ю. Гарфилдом в 1960-е гг.
	DOAJ (Directory of Open Access Journals): доступ к высококачественным рецензируемым журналам
Специализированные БД: узкая тематика или область (например, медицинские БД); предназначены для профессионалов в этих областях	*MEDLINE:* медицинские науки
	Embase: биомедицина и фармакология
	AGRICOLA: сельское хозяйство
	CAB Abstracts: агрономия, экология и смежные области
	INSPEC: физика, электроника
	GeoRef: геонаука, включая геологию, геофизику и экологию
	GeoBase: науки о Земле
	EconLit: экономика и бизнес
Технические и справочные БД: доступ к технической информации, справочным данным, инструкциям	*European Reference Index for the Humanities (ERIH):* гуманитарные науки
	DBLP: компьютерные науки
	Wikidata: открытая база знаний с информацией из различных областей; используется для создания связанной информации в Википедии и других проектах

Окончание таблицы

Наименование БД	Краткая характеристика
и стандартам (например, БД по материалам, программному обеспечению)	*Fatcat:* научные публикации с открытым доступом
	FSTA (Food Science and Technology Abstracts): наука о пище и технологиям пищевой промышленности
	ecbiz (EconBiz): бизнес и экономика
	cpx: поиск информации по техническим параметрам
Общие БД и сервисы: могут использоваться широкой аудиторией, охватывают большой круг тем, включая общедоступные информационные ресурсы (например, каталоги библиотек, поисковые системы)	*SHERPA:* оценка политики открытого доступа журналов и поддержки авторов в соблюдении требований к публикациям
	Scilit: поиск научных публикаций и мониторинг их цитирования
	SUDOC: каталогизация библиотек во Франции, доступ к ресурсам университетских библиотек
	CREF: образование и финансы
	msn: содержит информацию по мультимедийному контенту и технологиям
	zbm: охватывает различные темы и области, предоставляя доступ к множеству публикаций
	oax: доступ к научным статьям без платных подписок
Специальные БД и коллекции: включают уникальные или редкие материалы, которые могут быть собраны для определённой цели (например, архивы, коллекции личных фондов)	*Chim:* химические исследования
	Zoorec: зоология и биология
	Biabs: биология и смежные дисциплины
	Petro: нефтегазовые исследования
	DNB: Национальная библиотека Германии
	CAS (Chemical Abstracts Service) Core: химия
	job: публикации по исследованиям о науке и труде
	HSABS: академическая деятельность и высшее образование
	WPSA (Web of Political Science Abstracts): политические науки
	SOCABS (Social Science Abstracts): социальные науки

Каждая БД, представленная в таблице, имеет уникальные особенности и предназначена для определённых целей в научном сообществе. Совокупность тех или иных специализированных БД, в которых индексируется научный журнал, представляет собой профиль научного журнала. Например, профиль журнала «Информатика и образование», включённого в «Белый список», имеет Уровень 3 (У3) (рис. 2).

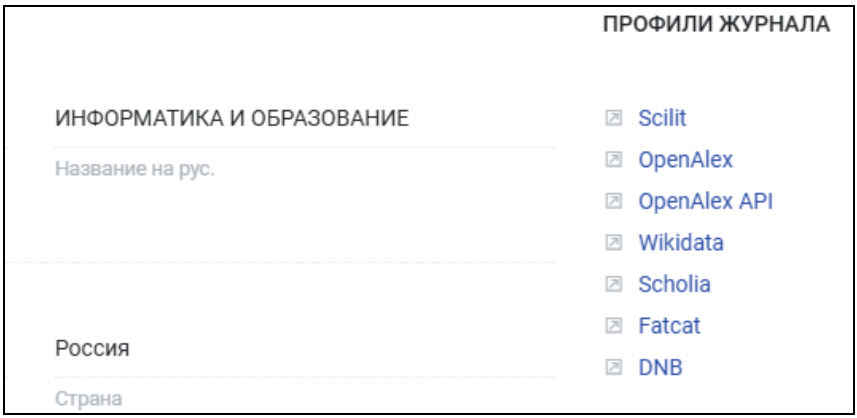


Рис. 2. Профиль научного журнала «Информатика и образование»

Публикации в этом журнале полезны для работников образования и научных сотрудников в области образовательных технологий и информатики. Его идентификация в БД (рис. 2) подтверждает научную значимость и доступность для исследователей и практиков. Индексирование в этих БД обеспечивает открытый доступ и метаданные для поиска, способствуя распространению знаний. Выбор подходящей БД для индексации зависит от области исследования и особенностей издания.

Практика применения EDA для исследования данных о научных журналах из «Белого списка»

В соответствии с классификацией типов EDA и соответствующими методами для каждого из них [**10, 11**] проведём анализ информационной базы, основываясь на открытых данных из «Белого списка» научных журналов. Для реализации EDA использовался специализированный облачный сервис Google Colaboratory, который предоставляет мощные инструменты для анализа данных и машинного обучения. Ниже представлены необходимые цифровые инструменты, включая фреймворки для проведения EDA:

высокоуровневый язык программирования: Python;
предобработка данных, EDA: Numpy, Pandas, StatsPy;
визуализация EDA: Matplotlib, Seaborn.

После предобработки данных перейдем к применению методов EDA, начиная с одномерного анализа.

Одномерный анализ. Это тип количественного анализа данных, используемый для описания, обобщения и поиска закономерностей в данных из одной переменной. Этот тип анализа выполняется с помощью описательной статистики, частотных распределений (например, гистограмма, круговая диаграмма).

Начнём с изучения признака level в виде круговой диаграммы для анализа распределения уровней научных журналов, представленных в исходных данных (рис. 3а) и после исключения изданий (рис. 3б), не соответствующих критериям Межведомственной Рабочей группы по формированию «Белого списка» (решение от 11.07.2024, протокол ДС/25-пр).

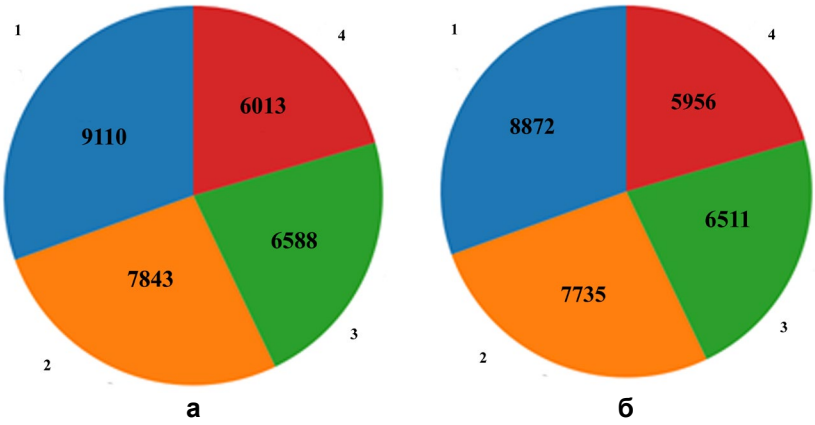


Рис. 3. Распределение научных журналов по уровням: до (а) и после (б) исключения из «Белого списка»

Анализируя распределение уровней научных изданий, следует отметить, что Уровень 1 (У1) остаётся преобладающим как до, так и после исключения журналов из «Белого списка», составляя примерно 30% от общего количества журналов. Уровень 2 (У2) удерживает около 25,8% и 25,7% соответственно, что тоже является значительной долей. Распределение уровней не претерпело значительных изменений с точки зрения пропорций: У1 и У2 по-прежнему доминируют, что указыва-

ет на то, что высококачественные и авторитетные научные журналы сохраняются в большинстве. Общее распределение уровней можно назвать неравномерным, поскольку У1 и У2 имеют значительно более высокие значения, чем У3 и У4, однако в пределах У3 и У4 соотношение довольно равномерное. Таким образом, несмотря на небольшое снижение общего количества научных журналов, распределение уровней остаётся схожим, не показывая значительных изменений в пропорциях.

Двумерный анализ. Это один из видов статистического анализа, в котором исследуется связь между двумя переменными (часто обозначаемыми как *x* и *y*). Он используется для поиска эмпирических связей между двумя величинами. Этот тип анализа выполняется с помощью диаграммы рассеяния, регрессионного анализа, коэффициентов корреляции (КК), кросс-табуляции[10].

Выполним двумерный анализ данных с помощью таких методов, как кросс-табуляция и визуализация. Анализ кросс-табуляции позволяет увидеть, как распределены значения индексации в БД RSCI в зависимости от уровня научного журнала level (рис. 4): У1 имеет наименьшую долю индексирования в БД RSCI 0,83%; У3 показывает наибольшую долю индексирования журналов 5,81% в сравнении с У2 и У4, которые имеют значения 2,93% и 4,22% соответственно.

rsci	0	1	All
level			
1	9034	76	9110
2	7613	230	7843
3	6205	383	6588
4	5759	254	6013
All	28611	943	29554

Рис. 4. Кросс-табуляция признаков level и rsci

[10] Кросс-табуляция – это эффективный метод исследования данных, который помогает обнаруживать и анализировать связи между переменными, относящимися к различным категориям. Анализ кросс-табуляции (также известный как анализ таблиц сопряжённости, таблица контингентности, факторная таблица в статистике) чаще всего используется для анализа категориальных данных с помощью Pandas.

Для научных журналов У3 и У4 наблюдается тенденция к увеличению доли индексированных публикаций, что может свидетельствовать о высоком качестве научных трудов или о большем внимании к их индексированию в БД RSCI. Рекомендуется дальнейшее изучение факторов, влияющих на индексирование журналов, а также анализ причин низкой индексируемости научных журналов У1 и У2 в БД RSCI.

Теперь воспользуемся визуализацией для более глубокого понимания взаимосвязи между индексированием в БД RSCI и уровнем научного журнала. Столбчатая диаграмма (рис. 5) позволяет наглядно оценить тренд: по мере уменьшения уровня журнала (слева направо по оси *X*) количество индексируемых журналов увеличивается, это говорит о положительной зависимости между журналами У2, У3, У4 и вероятностью их индексации в БД RSCI.

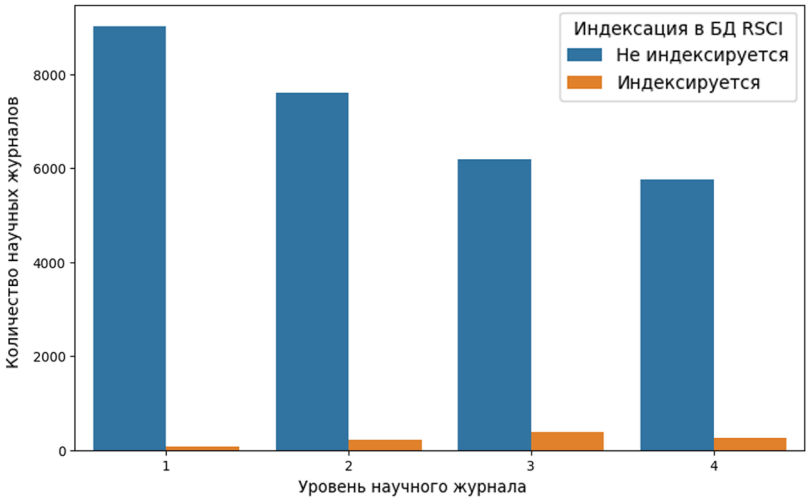


Рис. 5. Распределение индексации в БД RSCI по уровням журнала

Многомерный анализ. Это статистический метод понимания связей между несколькими переменными одновременно**;** фокусируется на понимании сложных взаимодействий и зависимостей между несколькими переменными [**22**]. Этот тип анализа выполняется с помощью корреляционного анализа, тепловой карты[11].

На рис. 6 представлена тепловая карта с использованием корреляционной матрицы между признаками level, wos_cc, scopus, rsci и doaj. Анализ тепловой карты позволил выявить взаимосвязанные и нейтральные признаки, основываясь на значениях КК. Необходимо отметить, что между рассматриваемыми признаками существуют незначительные корреляции. Например, признаки wos_cc и level (КК = -0.39), scopus и level (КК = -0.33) имеют умеренно отрицательную связь – это значит, что при преобладании одного признака, например, при повышении порядкового номера уровня научного журнала (level) наблюдается снижение другого wos_cc (то есть уменьшение количества индексации в БД WoS).

Обнаруженная обратная корреляция между уровнем журнала и индексацией в RSCI свидетельствует о региональной ориентации изданий У3–У4, тогда как доминирование Scopus у журналов У1–У2 подтверждает их международный статус.

Рассмотрим нейтральные признаки: wos_cc и scopus (КК = 0.05), rsci и doaj (КК = -0.02) – отсутствие связи между признаками; doaj с признаками level, wos_cc, scopus, rsci: КК находятся в пределах от -0.00 до 0.08, что указывает на отсутствие значимых взаимосвязей. В целом, можно сделать вывод, что между признаками level, wos_cc и scopus наблюдается умеренно отрицательная связь, в то время как признаки doaj и rsci имеют слабую или почти нулевую корреляцию с остальными признаками.

---

[11] Тепловая карта (англ. Heatmap) в Seaborn – это специализированный инструмент для визуализации многомерных корреляций, где значения матрицы взаимосвязей преобразуются в цветовые градиенты. Она особенно эффективна для анализа корреляционных матриц, позволяя мгновенно оценить силу и направление зависимостей между большим количеством переменных в многомерных наборах данных.

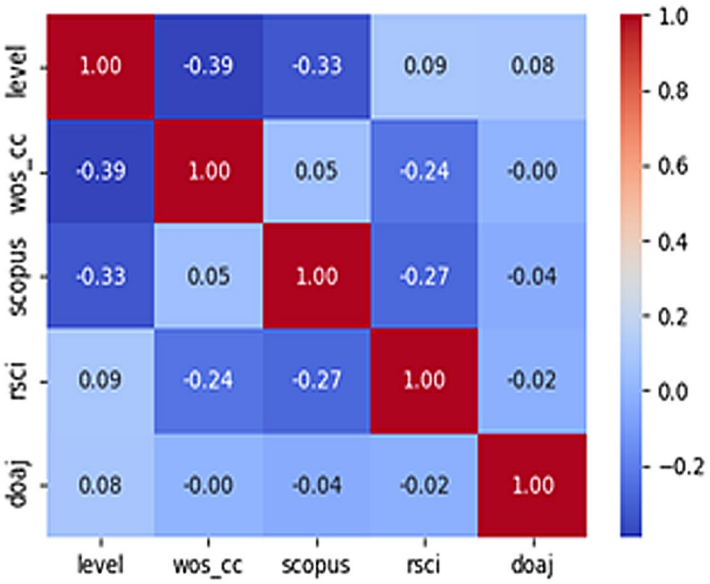


Рис. 6. Тепловая карта признаков level, wos_cc, scopus, rsci и doaj

Вычислим, сколько раз научные журналы определённого уровня индексируется в таких БД как Scopus, RSCI и WoS. На основе данных сводной таблицы о количестве индексации журналов в этих БД можно сделать несколько выводов и выявить закономерности, касающиеся уровней журналов (рис. 7).

Снижение индексации с понижением уровня журнала: по мере снижения уровня журнала (с У1 до У4) наблюдается значительное уменьшение количества индексаций в международных БД. Это указывает на то, что более авторитетные журналы (например, У1) индексируются большее количество раз, чем журналы У4.

Сравнение индексации научных журналов по БД: наивысшее количество индексаций наблюдается в БД Scopus для всех уровней. Это может свидетельствовать о большей международной репутации Scopus по сравнению с другими БД. WoS показывает схожую динамику с БД Scopus, но количество индексаций журнала в WoS всегда меньше, чем в Scopus.

level	rsci	scopus	wos_cc
1	76	8949	8415
2	230	7396	6420
3	383	5705	4224
4	254	4098	2765

Рис. 7. Сводная таблица индексации научных журналов в БД RSCI, Scopus, Web of Science

Характеристика уровней научных журналов: У1 – самый высокий уровень, характеризующийся наибольшим количеством индексаций в международных БД. Журналы этого уровня, вероятно, являются научными изданиями с серьёзным влиянием на свои области. Журналы У2 тоже имеют значительный авторитет, но их индексация ниже, чем у У1. Они заслуживают внимания, но их влияние меньше. Журналы У3 реже присутствуют в научных индексах, что может указывать на более узкую специализацию или менее высокий статус в академическом сообществе. У4 – наиболее низкий уровень из представленных, с минимальным количеством индексаций, что может быть следствием меньшего внимания к таким изданиям со стороны международного исследовательского сообщества.

В заключение отметим, что данные сводной таблицы (рис. 7) чётко демонстрируют, что высокий уровень журналов сопровождается большей индексацией в авторитетных международных БД Scopus и WoS, что служит индикатором их качества и значимости в научной среде.

Выводы

1. С помощью EDA был проведён количественный анализ соответствия эмпирических данных заявленным правилам категорирования. Выявленные статистические закономерности отражают различия в паттернах индексации для научных журналов различных уровней из «Белого списка».

2. Результаты анализа, выполненного с использованием методов EDA на цифровой платформе Google Colaboratory, подтвердили эффективность индекса цитирования как одного из ключевых критериев для категорирования научных журналов. Обнаружена обратная корреляция между уровнем научного журнала и индексацией в RSCI, что указывает на региональную ориентацию изданий уровней У3 и У4. В то же время преобладание Scopus среди научных журналов уровней У1 и У2 свидетельствует о международной значимости изданий высших уровней. Кроме того, в распределении данных о научных журналах выявлены статистические аномалии, которые могут указать на направления для оптимизации системы категорирования. Например, для изданий уровня У2 рекомендуется нормализация порогов индексации в WoS с учётом нелинейного распределения, а для журналов уровней У3 и У4 может быть предложен механизм стимулирования входа в DOAJ, что повысит видимость их публикаций.

3. EDA подтвердил свою ценность как современный инструмент верификации, позволяющий выявлять скрытые зависимости в наукометрических показателях. В методологии EDA для библиометрических исследований целесообразно использовать следующие методы: одномерный анализ (для изучения распределения уровней), двумерный анализ (кросс-табуляция «уровень – индексация»), многомерный анализ (тепловые карты корреляций). Адаптация EDA к верификации «Белого списка» способствует развитию аналитических подходов в библиометрии в эпоху цифровизации науки.

Список источников

1. **Котляров И. Д.** Проблемы функционирования российской системы научных публикаций и пути их решения // Образование и наука. Известия УрО РАО. 2011. № 1 (80). С. 92–101.

2. **Сюнтюренко О. В., Гиляревский Р. С.** Использование методов наукометрии и сопоставительного анализа данных для управления научными исследованиями по тематическим направлениям // Научно-техническая информация. Сер. 2. 2016. № 12. С. 1–12.

3. **Медведева О. О., Дьяченко Е. Л.** Белые списки журналов: международный опыт составления и роль в управлении наукой [Презентация]. 9-я Международная научно-практическая конференция «Научное издание международного уровня: мировые тенденции и национальные приоритеты», г. Москва; 24–27 мая 2021 г. URL: https://rassep.ru/academy/biblioteka/106130/ (дата обращения: 11.11.2025).

4. **Patwardhan B., Nagarkar Sh., Gadre Sh., Lakhotia S., Katoch V., Moher D.** A Critical Analysis of the 'UGC-Approved List of Journals' // Current science. 2018; 114 (6): 1299–1303. DOI 10.18520/cs/v114/i06/1299-1303.

5. **Singh J.** Indian LIS Journal: Current Status and Scenario as seen through UGC CARE List in the year 2019-2020. 2023. DOI 10.1729/Journal.36020.

6. **Кочетков Д. М.** Белый список российских журналов: вопросы, ждущие ответа // Научный редактор и издатель. 2022. Т. 7, № 2. С. 185–190.

7. **Горелкин В. А.** Регламентация издания научных журналов в России: проблемы и предложения // Научный редактор и издатель. 2022. Т. 7, № S1. С. 6–15.

8. **Полилова Т. А.** Рейтинги библиографической базы и «белые списки» // Электронные библиотеки. 2022. Т. 25, № 6. С. 640–670.

9. **«Белый список»** журналов на веб-сайте Российского центра научной информации. URL: https://journalrank.rcsi.science/ru (дата обращения: 11.11.2025).

10. **Скиен С.** Наука о данных: учебный курс / пер. с англ. Санкт-Петербург : ООО «Диалектика», 2020. 544 с.

11. **Дэви С.** Основы Data Science и Big Data. Python и наука о данных. Санкт-Петербург : Питер, 2017. 336 с.

12. **Баканова Н. Б.** Анализ данных публикационной активности для исследования направлений научного сотрудничества организации // Научные и технические библиотеки. 2024. № 11. С. 31–47.

13. **Баканова Н. Б.** Многокритериальная оценка публикационной результативности научных подразделений организации // Искусственный интеллект и принятие решений. 2022. № 3. С. 88–95. DOI 10.14357/20718594220307.

14. **Захарова С. С.** Библиометрическая аналитика в результатах научных разработок // НТИ-2022. Научная информация в современном мире: глобальные вызовы и приоритеты экономики. Москва : Всероссийский институт научной и технической информации РАН, 2022. С. 430–433. DOI 10.36535/2022-9785945770829-74.

15. **Pessin V., Vilker L., Yamane L., Siman R.** Smart bibliometrics: an integrated method of science mapping and bibliometric analysis. Scientometrics. 2022. No. 127. DOI 10.1007/s11192-022-04406-6.

16. **Maliha H.** A Review on Bibliometric Application Software. Sci. Lett. 2023. No. 1. DOI 10.58968/sl.v1i1.458.

17. **Moral-Muñoz J. A., Herrera-Viedma E., Santisteban-Espejo A., Cobo M. J.** Software tools for conducting bibliometric analysis in science: An up-to-date review. El profesional de la información. 2020. v. 29, No. 1, e290103. DOI 10.3145/epi.2020.ene.03.

18. **Pessin V., Vilker L., Yamane L., Siman R.** Smart bibliometrics: an integrated method of science mapping and bibliometric analysis. Scientometrics. 2022. No. 127. DOI 10.1007/s11192-022-04406-6.

19. **Tan C. N.-L., Fauzi M**. The bibliometric overview of research on healthcare information systems using big data analytics. International Journal of Data Science and Big Data Analytics. 2023. No. 3. Pp. 45–57. DOI 10.51483/IJDSBDA.3.1.2023.45-57.

20. **Rivest M., Vignola-Gagné E., Archambault É.** Article-level classification of scientific publications: A comparison of deep learning, direct citation and bibliographic coupling. PLOS ONE. 2021. No. 16. e0251493. DOI 10.1371/journal.pone.0251493.

21. **Rao A. S., Vardhan B. V., Shaik H.** Role of Exploratory Data Analysis in Data Science. Proc. 6th Int. Conf. Commun. Electron. Syst. ICCES 2021. 2021. No. 7. Pp. 1457–1461. DOI 10.1109/ICCES51350.2021.9488986.

22. **Практическая** статистика для специалистов Data Science / пер. с англ. П. Брюс, Э. Брюс, П. Гедек. 2-е изд., перераб. и доп. Санкт-Петербург : БХВ-Петербург, 2021. 352 с.

23. **МакКинни У.** Python и анализ данных: Первичная обработка данных с применением Pandas, NumPy и Jupiter / пер. с англ. А. А. Слинкина. 3-е изд. Москва : МК Пресс, 2023. 536 с.

24. **Simangunsong J., Simanjuntak M., Simanjuntak N.** Mental disorder classification with exploratory data analysis (EDA). Journal of Intelligent Decision Support System (IDSS). 2024. No. 7. Pp. 210–217. DOI 10.35335/idss.v7i3.252.

25. **Karunia R., Hidayati N.** Analisis Data dan Visualisasi Pola Ancaman Siber Global (2015–2024) menggunakan Exploratory Data Analysis (EDA). Dinamik. 2025. No. 30. Pp. 203–211. DOI: 10.35315/dinamik.v30i2.10136.

References

1. **Kotliarov I. D.** Problemy` funktcionirovaniia rossii`skoi` sistemy` nauchny`kh publikatcii` i puti ikh resheniia // Obrazovanie i nauka. Izvestiia UrO RAO. 2011. № 1 (80). S. 92–101.

2. **Siuntiurenko O. V., Giliarevskii` R. S.** Ispol`zovanie metodov naukometrii i sopostavitel`nogo analiza danny`kh dlia upravleniia nauchny`mi issledovaniiami po tematicheskim napravleniiam // Nauchno-tekhnicheskaia informatciia. Ser. 2. 2016. № 12. S. 1–12.

3. **Medvedeva O. O., D`iachenko E. L.** Bely`e spiski zhurnalov: mezhdunarodny`i` opy`t sostavleniia i rol` v upravlenii naukoi` [Prezentatciia]. 9-ia Mezhdunarodnaia nauchno-prakticheskaia konferentciia «Nauchnoe izdanie mezhdunarodnogo urovnia: mirovy`e tendentcii i natcional`ny`e prioritety`», g. Moskva; 24–27 maia 2021 g. URL: https://rassep.ru/academy/biblioteka/106130/ (дата обращения: 11.11.2025).

4. **Patwardhan B., Nagarkar Sh., Gadre Sh., Lakhotia S., Katoch V., Moher D.** A Critical Analysis of the 'UGC-Approved List of Journals' // Current science. 2018; 114 (6): 1299–1303. DOI 10.18520/cs/v114/i06/1299-1303.

5. **Singh J.** Indian LIS Journal: Current Status and Scenario as seen through UGC CARE List in the year 2019-2020. 2023. DOI 10.1729/Journal.36020.

6. **Kochetkov D. M.** Bely`i` spisok rossii`skikh zhurnalov: voprosy`, zhdushchie otveta // Nauchny`i` redaktor i izdatel`. 2022. T. 7, № 2. S. 185–190.

7. **Gorelkin V. A**. Reglamentatciia izdaniia nauchny`kh zhurnalov v Rossii: problemy` i predlozheniia // Nauchny`i` redaktor i izdatel`. 2022. T. 7, № S1. S. 6–15.

8. **Polilova T. A.** Rei`tingi bibliograficheskoi` bazy` i «bely`e spiski» // E`lektronny`e biblioteki. 2022. T. 25, № 6. S. 640–670.

9. **«Bely`i` spisok»** zhurnalov na veb-sai`te Rossii`skogo centra nauchnoi` informatcii. URL: https://journalrank.rcsi.science/ru (data obrashcheniia: 11.11.2025).

10. **Skien S.** Nauka o danny`kh: uchebny`i` kurs / per. s angl. Sankt-Peterburg : OOO «Dialektika», 2020. 544 s.

11. **De`vi S.** Osnovy` Data Science i Big Data. Python i nauka o danny`kh. Sankt-Peterburg : Peter, 2017. 336 s.

12. **Bakanova N. B.** Analiz danny`kh publikatcionnoi` aktivnosti dlia issledovaniia napravlenii` nauchnogo sotrudnichestva organizatcii // Nauchny`e i tekhnicheskie biblioteki. 2024. № 11. S. 31–47.

13. **Bakanova N. B.** Mnogokriterial`naia ocenka publikatcionnoi` rezul`tativnosti nauchny`kh podrazdelenii` organizatcii // Iskusstvenny`i` intellekt i priniatie reshenii`. 2022. № 3. S. 88–95. DOI 10.14357/20718594220307.

14. **Zaharova S. S.** Bibliometricheskaia analitika v rezul`tatakh nauchny`kh razrabotok **//** NTI-2022. Nauchnaia informatciia v sovremennom mire: global`ny`e vy`zovy` i prioritety` e`konomiki. Moskva : Vserossii`skii` institut nauchnoi` i tekhnicheskoi` informatcii RAN, 2022. S. 430–433. DOI 10.36535/2022-9785945770829-74.

15. **Pessin V., Vilker L., Yamane L., Siman R.** Smart bibliometrics: an integrated method of science mapping and bibliometric analysis. Scientometrics. 2022. No. 127. DOI 10.1007/s11192-022-04406-6.

16. **Maliha H.** A Review on Bibliometric Application Software. Sci. Lett. 2023. No. 1. DOI 10.58968/sl.v1i1.458.

17. **Moral-Muñoz J. A., Herrera-Viedma E., Santisteban-Espejo A., Cobo M. J.** Software tools for conducting bibliometric analysis in science: An up-to-date review. El profesional de la información. 2020. v. 29, No. 1, e290103. DOI 10.3145/epi.2020.ene.03.

18. **Pessin V., Vilker L., Yamane L., Siman R.** Smart bibliometrics: an integrated method of science mapping and bibliometric analysis. Scientometrics. 2022. No. 127. DOI 10.1007/s11192-022-04406-6.

19. **Tan C. N.-L., Fauzi M**. The bibliometric overview of research on healthcare information systems using big data analytics. International Journal of Data Science and Big Data Analytics. 2023. No. 3. Pp. 45–57. DOI 10.51483/IJDSBDA.3.1.2023.45-57.

20. **Rivest M., Vignola-Gagné E., Archambault É.** Article-level classification of scientific publications: A comparison of deep learning, direct citation and bibliographic coupling. PLOS ONE. 2021. No. 16. e0251493. DOI 10.1371/journal.pone.0251493.

21. **Rao A. S., Vardhan B. V., Shaik H.** Role of Exploratory Data Analysis in Data Science. Proc. 6th Int. Conf. Commun. Electron. Syst. ICCES 2021. 2021. No. 7. Pp. 1457–1461. DOI 10.1109/ICCES51350.2021.9488986.

22. **22. Prakticheskaia** statistika dlia spetcialistov Data Science / per. s angl. P. Brius, E`. Brius, P. Gedek. 2-e izd., pererab. i dop. Sankt-Peterburg : BKHV-Peterburg, 2021. 352 s.

23. **23. MakKinni U.** Python i analiz danny`kh: Pervichnaia obrabotka danny`kh s primeneniem Pandas, NumPy i Jupiter / per. s angl. A. A. Slinkina. 3-e izd. Moskva : MK Press, 2023. 536 s.

24. **Simangunsong J., Simanjuntak M., Simanjuntak N.** Mental disorder classification with exploratory data analysis (EDA). Journal of Intelligent Decision Support System (IDSS). 2024. No. 7. Pp. 210–217. DOI 10.35335/idss.v7i3.252.

25. **Karunia R., Hidayati N.** Analisis Data dan Visualisasi Pola Ancaman Siber Global (2015–2024) menggunakan Exploratory Data Analysis (EDA). Dinamik. 2025. No. 30. Pp. 203–211. DOI: 10.35315/dinamik.v30i2.10136.

Информация об авторе / Author

Моисеева Наталья Александровна – канд. пед. наук, доцент кафедры «Прикладная математика и фундаментальная информатика» Омского государственного технического университета, Омск, Российская Федерация

nat_lion@mail.ru

Natalya A. Moiseeva – Cand. Sc. (Pedagogy), Associate Professor, Applied Mathematics and Basic Informatics Chair, Omsk State Technical University, Omsk, Russian Federation

nat_lion@mail.ru