
























OpenAlex и библиометрия

Мохначева Ю. В.

Основные факты

- Создана компанией OurResearch в 2022 г., заменив Microsoft Academic.
- OpenAlex – бесплатный и полностью открытый каталог научных метаданных с открытым исходным кодом, позволяющий проводить комплексный библиографический анализ.
- OpenAlex придерживается принципа наибольшей широты охватываемого контента, не используя экспертные оценки при отборе источников.
- Индексация на уровне документов, а не журналов.
- Собирает и стандартизирует данные из таких источников как: MAG; Crossref; ORCID, ROR, DOAJ, Unpaywall, Pubmed, Pubmed Central, The ISSN International Centre, различных репозиториях.
- Лейденский университет в Нидерландах с 2024 г. начал использовать OpenAlex для составления рейтинга своих научно-исследовательских учреждений – CWTS Leiden Ranking на 2024 г.

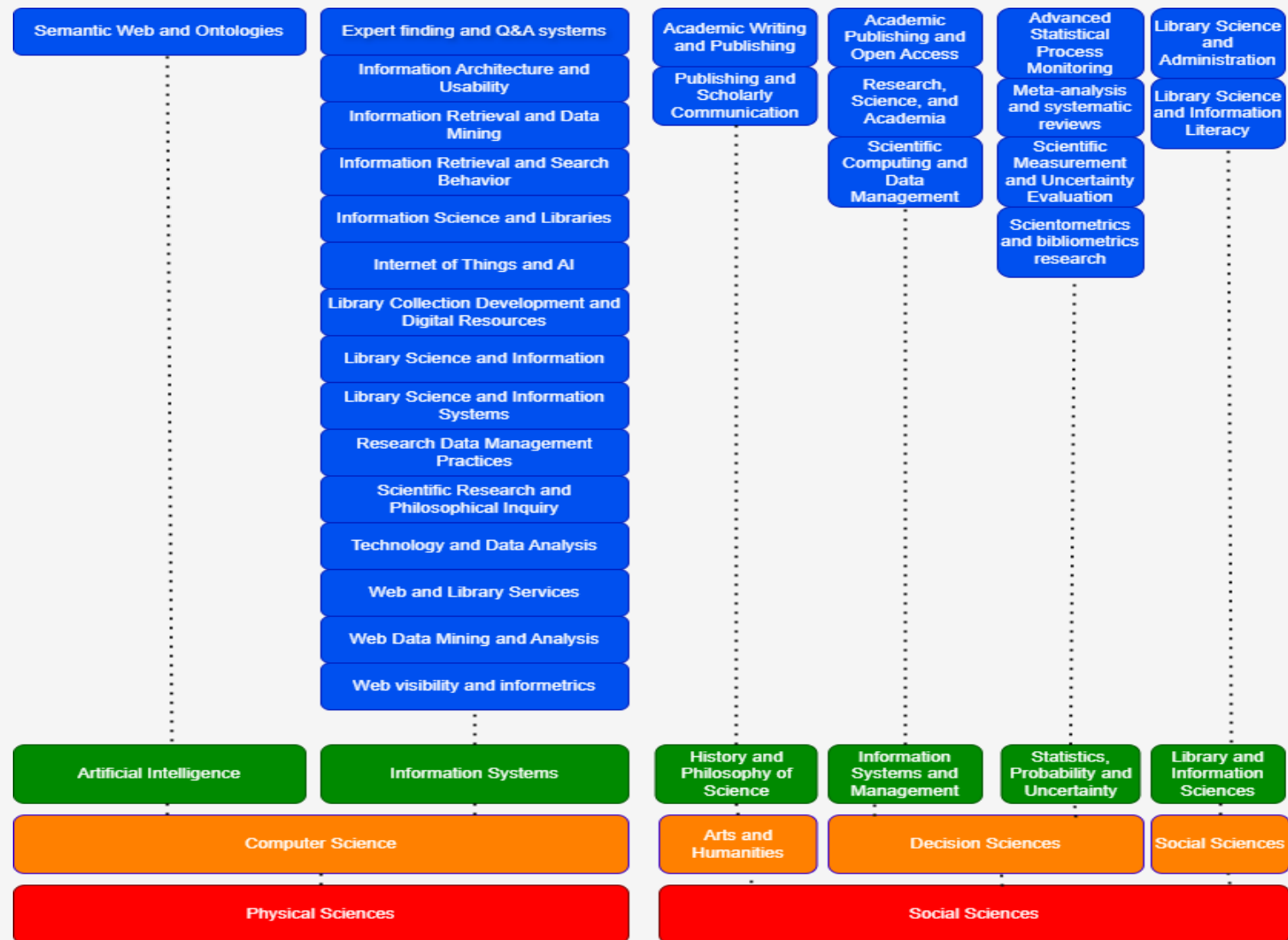
Статистические данные по ресурсу на 24.09.2026

 271M Works 	 105M Authors 	 261K Sources 
62M Open Access works	8M with ORCIDs	45K that are Open Access
34M from the Global South	17M from the Global South	
8M datasets		
 11K Publishers 	 32K Funders 	 118K Institutions 
 65K Concepts 	 73K Keywords 	 5K Topics 
 252 Subfields	 26 Fields	 4 Domains
 17 Sdgs	 247 Countries	 7 Continents
 184 Languages	 20 Types	 6 Source-Types
 8 Institution-Types	 13 Licenses	

Иерархия классификации в OpenAlex

- около **5000 тем** сгруппированы в **252 подобласти**, которые, в свою очередь, сгруппированы уже в **26 областей знания**, которые, в конечном итоге, группируются в **4 научные сферы**.
- Документы помечаются тегами с помощью автоматизированной системы, которая учитывает имеющуюся информацию о работе: название, аннотацию, наименование источника (издания) и ссылки.
- Произведениям с использованием искусственного интеллекта назначаются темы с помощью модели, присваивающей им баллы: тема с наибольшим количеством баллов становится «основной темой» произведения.
- У каждой темы имеется одна подобласть, одна область знания (поле) и одна научная сфера (домен).

Отобранные 27 тем в OpenAlex, сопряженные с библиотечно-информационной сферой: распределение по классификатору



Темы и ключевые слова в OpenAlex

- Совместно с Центром исследований науки и технологий (CWTS) Лейденского университета, **для тем** был разработан **метод классификации с использованием искусственного интеллекта. Но!** Некоторые работы остались без тем из-за недостаточности связанных с ними данных, что делает эти работы невидимыми при тематическом поиске. Возможные причины: большие потери метаданных документов при индексации.
- **Ключевые слова** в интерпретации OpenAlex — короткие слова или фразы, присваиваемые произведениям с использованием искусственного интеллекта. Алгоритм присвоения ключевых слов OpenAlex заключается в присвоении ключевых слов на основе тем — более 73000 ключевых слов (по состоянию данных на август 2025 г.).
- **Технология присвоения ключевых слов:** сначала подготавливается список из 10 ключевых слов для каждой темы. Для того, чтобы присвоить ключевые слова работам, берутся не более трех тем, связанных с этой работой, выбираются ключевые слова, связанные с этими темами — на данный момент не более 30 ключевых слов, а затем определяется сходство ключевого слова с названием и аннотацией. Пять лучших ключевых слов отбираются до момента, пока показатель сходства не превысит определенный порог.

Достоинства и недостатки технологии присвоения ключевых слов

Достоинства:

- Поиск разработчиками **путей совершенствования унификации ключевых слов**, позволяющих устранять такие погрешности как: различия в регистре/пробелах/знаках препинания и т.п. для того, чтобы похожие ключевые слова не присваивались одной и той же работе.
- Удобство, т.к. множество синонимов, сокращений и т.п. объединяются в один термин.

Недостатки:

- Самый существенный — недостаточное количество отобранных ключевых слов для тем.
- Ошибочная/чрезмерная унификация терминов (вероятно вызвана огрехами искусственного интеллекта). Ведет к утере терминологической специфики ключевых слов и к другим более значимым ошибкам (Пример на след. слайде).

Пример: ключевое слово «Impact factor» в теме: «Scientometrics and bibliometrics research»

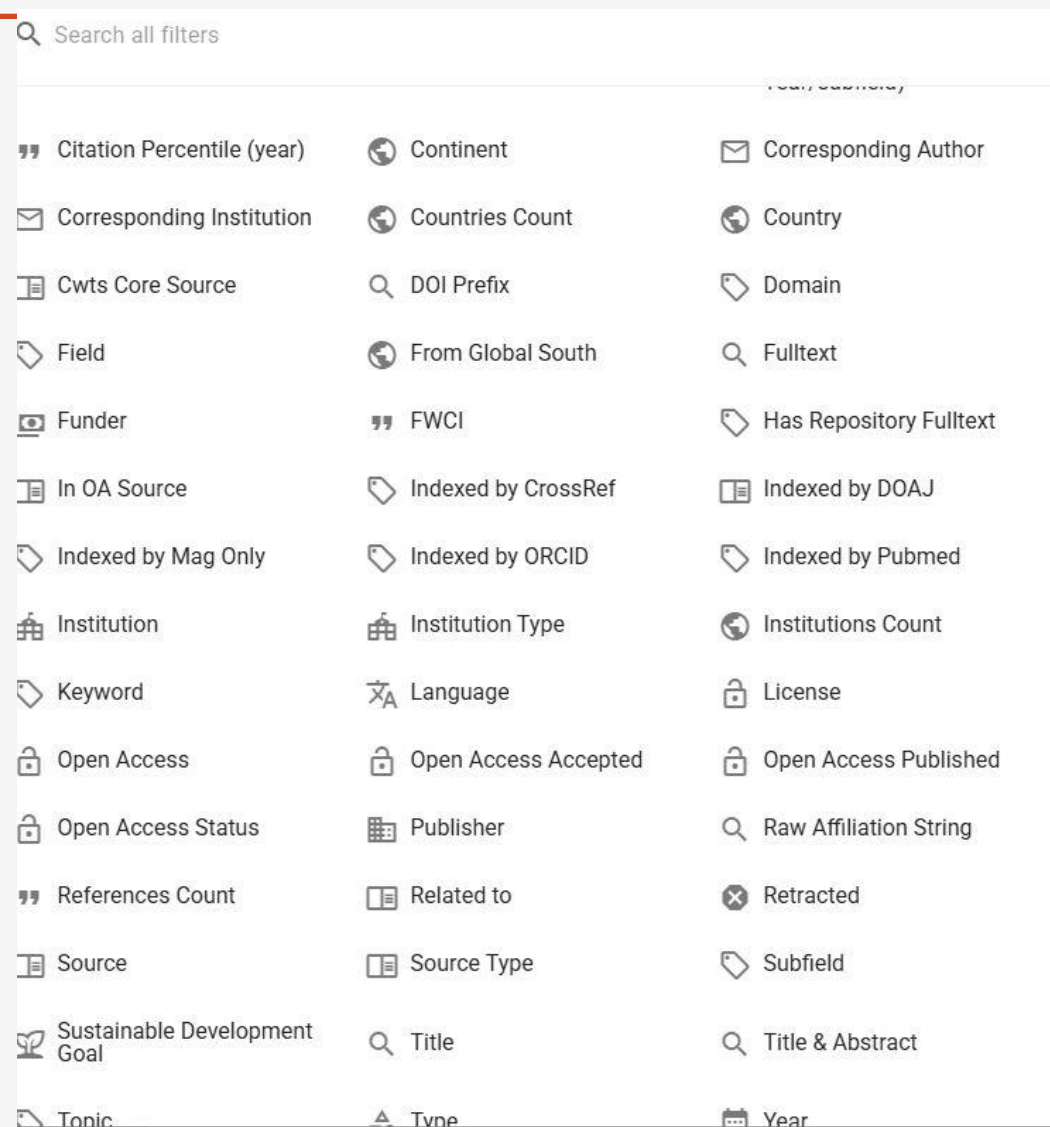
За 2022-2023 гг. выявлено 320 публикаций по теме «Scientometrics and bibliometrics research», которым OpenAlex присвоила ключевое слово «Impact factor».

Диапазон набора разных ключевых терминов в отобранных документах варьировался от 1 до 9-ти:

- тридцати трем публикациям (более 10%) данный термин был присвоен в качестве единственного;
- два термина были присвоены 75-ти работам;
- три – 87-ми;
- четыре – 59-ти; пять – 37-ми; шесть – 14-ти; семь – 11-ти; восемь и девять – двум документам соответственно.

Обнаружено, что например, по каким-то причинам системой не классифицируются как самостоятельные ключевые слова: «Scopus», «Google Scholar», «Medline» и др. Данные термины ассоциируются системой с терминами: «Bibliographic database», «Data source», «Citation Index», «Citation database» и, что самое странное – с «Web of Science», который, в отличие от других баз данных, все же присутствует в качестве самостоятельного ключевого слова в OpenAlex (виноват ИИ?)

51 фильтр для поиска документов. Например:




Варианты экспорта данных

Из веб-интерфейса

Export results

☒ Spreadsheet (.csv)
☐ Endnote format (.ris)
☐ Text format (.txt)
☐ Shorten column values for Excel compatibility?

 Since there are many records, the export will take up to fifteen minutes.

Cancel [Start export](#)

+ по API;
возможность загружать данные в аналитические системы

Пример результата выгрузки данных в таблицу Excel

Всего: 182 поля с метаданными и статистикой

1	id	doi	title	display_name	publi	publi	langu	type	type	indexed_in	countries_dis	institutions	corresponding_author_ids
2	https://openalex.org/W2913767564	https://doi.org/10.1056/nejmoa1815671	Darolutamide in Nonmetastatic, Castration-	Darolutamide in Nonmetastatic, Castration-	2019	2019-02	en	article	journal-	crossref pubme	10	14	
3	https://openalex.org/W2586384977	https://doi.org/10.1093/ckj/sfz011	The European Renal Association – European	The European Renal Association – European	2019	2019-01	en	article	journal-	crossref pubme	26	41	https://openalex.org/A5022293730
4	https://openalex.org/W2972707815	https://doi.org/10.1200/jco.19.00456	Vemurafenib for Refractory Multisystem La	Vemurafenib for Refractory Multisystem La	2019	2019-09	en	article	journal-	crossref pubme	14	37	
5	https://openalex.org/W2964281414	https://doi.org/10.1111/jnc.14829	The energetic brain – A review from studen	The energetic brain – A review from studen	2019	2019-07	en	review	journal-	crossref pubme	14	36	https://openalex.org/A5057678663
6	https://openalex.org/W2930184070	https://doi.org/10.3390/nano9040494	Fe3O4 Nanoparticles for Complex Targeted	Fe3O4 Nanoparticles for Complex Targeted	2019	2019-03	en	article	journal-	crossref doaj p	3	8	https://openalex.org/A5054312816
7	https://openalex.org/W2943933477	https://doi.org/10.1016/j.jallcom.2019.05.015	Immobilization of boron-rich compound on	Immobilization of boron-rich compound on	2019	2019-05	en	article	journal-	crossref	3	7	https://openalex.org/A5054312816
8	https://openalex.org/W2914616192	https://doi.org/10.1021/acsp Photonics.8b01153	Experimental Demonstration of Ultrafast TH	Experimental Demonstration of Ultrafast TH	2019	2019-02	en	article	journal-	crossref pubme	5	7	https://openalex.org/A5028737213
9	https://openalex.org/W2979539159	https://doi.org/10.1183/13993003.01522-2019	Surveillance of adverse events in the treatr	Surveillance of adverse events in the treatr	2019	2019-10	en	article	journal-	crossref pubme	28	62	
10	https://openalex.org/W2996311965	https://doi.org/10.1186/s13017-019-0278-6	Duodeno-pancreatic and extrahepatic biliar	Duodeno-pancreatic and extrahepatic biliar	2019	2019-12	en	review	journal-	crossref pubme	18	30	
11	https://openalex.org/W2951865226	https://doi.org/10.1038/s41598-019-44341-3	Brown bear attacks on humans: a worldwid	Brown bear attacks on humans: a worldwid	2019	2019-06	en	article	journal-	crossref pubme	21	42	https://openalex.org/A5090391158
12	https://openalex.org/W2996944141	https://doi.org/10.1016/j.jaci.2019.12.896	Long-term outcome of LRBA deficiency in 7	Long-term outcome of LRBA deficiency in 7	2019	2019-12	en	article	journal-	crossref pubme	19	44	https://openalex.org/A5027980182
13	https://openalex.org/W2991978356	https://doi.org/10.1007/978-3-030-32072-0	The Theory of Queuing Systems with Correl	The Theory of Queuing Systems with Correl	2019	2019-12	en	book	book	crossref	2	2	
14	https://openalex.org/W2989976849	https://doi.org/10.3390/nano9121720	Functional Magnetic Composites Based on f	Functional Magnetic Composites Based on f	2019	2019-12	en	article	journal-	crossref pubme	3	5	https://openalex.org/A5103003041
15	https://openalex.org/W2996762807	https://doi.org/10.1016/j.jallcom.2019.15353	Peculiarities of the magnetic structure and r	Peculiarities of the magnetic structure and r	2019	2019-12	en	article	journal-	crossref	6	12	https://openalex.org/A5103001143
16	https://openalex.org/W4230432148	https://doi.org/10.6028/nist.sp.330-2019	The international system of units (SI):	The international system of units (SI):	2019	2019-08	en	report	report	crossref	86	0	
17	https://openalex.org/W2924302487	https://doi.org/10.1149/2.1001904jes	Control of Growth Mechanism of Electrode	Control of Growth Mechanism of Electrode	2019	2019-01	en	article	journal-	crossref	3	6	
18	https://openalex.org/W2980499137	https://doi.org/10.1021/acs.jpcc.9b06997	Features of the Growth Processes and Magn	Features of the Growth Processes and Magn	2019	2019-10	en	article	journal-	crossref	3	4	https://openalex.org/A5023391021
19	https://openalex.org/W2990183389	https://doi.org/10.1016/j.jmmm.2019.16619	Effect of treatment conditions on structure	Effect of treatment conditions on structure	2019	2019-11	en	article	journal-	crossref	3	6	https://openalex.org/A5048158694
20	https://openalex.org/W2925327111	https://doi.org/10.1016/j.jallcom.2019.03.333	Features of crystal and magnetic structure c	Features of crystal and magnetic structure c	2019	2019-03	en	article	journal-	crossref	5	9	https://openalex.org/A5073302795
21	https://openalex.org/W2981424704	https://doi.org/10.1016/j.colsurfa.2019.12474	'Green' approach for obtaining stable pecti	'Green' approach for obtaining stable pecti	2019	2019-10	en	article	journal-	crossref	3	5	https://openalex.org/A5007285434
22	https://openalex.org/W2918148752	https://doi.org/10.1103/physrevb.99.054422	Route toward semiconductor magnonics: Li	Route toward semiconductor magnonics: Li	2019	2019-02	en	article	journal-	crossref	2	3	
23	https://openalex.org/W2950154582	https://doi.org/10.1016/j.physletb.2019.13474	Asymptotically flat spinning scalar, Dirac an	Asymptotically flat spinning scalar, Dirac an	2019	2019-08	en	article	journal-	arxiv crossref d	3	4	https://openalex.org/A5023153460
24	https://openalex.org/W2982340107	https://doi.org/10.1186/s40425-019-0747-1	Intratumoral immunoglobulin isotypes pred	Intratumoral immunoglobulin isotypes pred	2019	2019-10	en	article	journal-	crossref pubme	4	7	
25	https://openalex.org/W2954008651	https://doi.org/10.1016/j.jallcom.2019.07.040	Formation and corrosion properties of Ni-b	Formation and corrosion properties of Ni-b	2019	2019-07	en	article	journal-	crossref	3	4	https://openalex.org/A5054312816
26	https://openalex.org/W2955535322	https://doi.org/10.3390/ijerph16132297	The Prevalence of Cigarette and E-cigarette	The Prevalence of Cigarette and E-cigarette	2019	2019-06	en	article	journal-	crossref doaj p	6	12	https://openalex.org/A5011512085
27	https://openalex.org/W2901529951	https://doi.org/10.1038/s41436-019-0566-2	PEDIA: prioritization of exome data by imag	PEDIA: prioritization of exome data by imag	2019	2019-06	en	article	journal-	crossref pubme	17	45	https://openalex.org/A5013713722
28	https://openalex.org/W2921217660	https://doi.org/10.3390/s19051250	Structure-Property Relationships in Graph	Structure-Property Relationships in Graph	2019	2019-03	en	review	journal-	crossref doaj p	3	5	https://openalex.org/A5063901096
29	https://openalex.org/W2916447240	https://doi.org/10.1038/s41565-019-0492-0	Optical valley Hall effect for highly valley-c	Optical valley Hall effect for highly valley-c	2019	2019-07	en	article	journal-	arxiv crossref d	5	6	
30	https://openalex.org/W3123733457	https://doi.org/10.1007/jhep11(2019)003	Top-quark physics at the CLIC electron-posit	Top-quark physics at the CLIC electron-posit	2019	2019-11	en	article	journal-	arxiv crossref d	15	30	
31	https://openalex.org/W2994780760	https://doi.org/10.1021/acsp Photonics.9b01153	Nanoscale Tunable Optical Binding Mediate	Nanoscale Tunable Optical Binding Mediate	2019	2019-12	en	article	journal-	crossref	3	6	https://openalex.org/A5011860899
32	https://openalex.org/W3011333185	https://doi.org/10.1093/ptep/ptz125	COMET Phase-I technical design report	COMET Phase-I technical design report	2019	2019-10	en	article	journal-	arxiv crossref d	17	45	https://openalex.org/A5107899212
33	https://openalex.org/W2920825177	https://doi.org/10.1016/j.memsci.2019.03.030	Effect of Pluronic F127 on porous and dense	Effect of Pluronic F127 on porous and dense	2019	2019-03	en	article	journal-	crossref	3	4	https://openalex.org/A5050964575
34	https://openalex.org/W2916601221	https://doi.org/10.1016/j.jssc.2019.02.023	Study of WO3-In2O3 nanocomposites for h	Study of WO3-In2O3 nanocomposites for h	2019	2019-02	en	article	journal-	crossref	2	4	https://openalex.org/A5055517747
35	https://openalex.org/W2913908916	https://doi.org/10.1038/s41598-018-38474-7	Nuclear nanomedicine using Si nanoparticle	Nuclear nanomedicine using Si nanoparticle	2019	2019-02	en	article	journal-	crossref pubme	5	7	

Выводы

1. Ресурс обладает всем необходимым для сбора и обработки библиографических данных по самым разнообразным критериям: от публикационной активности на всех уровнях (макро-; мезо-; микро) до терминологического анализа документов. Особо выделяется среди других ресурсов следующими качествами:

- Удобство и универсальность экспорта данных для последующей аналитики.
- Широкий охват документов со всего мира по полному спектру научных дисциплин.
- Широкий охват языков документов.
- Индексация на уровне документов, а не журналов.
- Максимально широкий спектр выгружаемых данных о документах. Помимо стандартных метаданных выгружаются: нормированные показатели цитируемости; число авторов, стран, организаций; число библиографических ссылок; динамика ссылок на документы по годам; присутствие публикации в топ-1%, или топ-10 по перцентилю нормированной цитируемости и др.
- Стабильность и скорость работы из-за лаконичности интерфейса.

Основной недостаток: большой процент утери ключевых метаданных. Особенно это касается стран и мест работы несмотря на корректные аффилиации в базе данных. Результат – неполные результаты поиска по данным критериям.

2. OpenAlex вполне может служить в качестве базового ресурса в информационном сопровождении научных исследований, включая библиометрию, особенно на фоне имеющихся ограничений доступа к проприетарным системам.

Благодарю за внимание!

Мохначева Юлия Валерьевна

Канд. пед. наук, вед.н.с. БЕН РАН,
ИО за НИО ИОН

mohnacheva@benran.ru