

Формирование междисциплинарных тематических коллекций с использованием информационных ресурсов открытого доступа



**Солошенко Н.С.,
Федорец О.В**

ВИНИТИ РАН

Методика формирования тематических коллекций научных публикаций

Открытые библиографические БД

- **The Lens** (Австралия) – **287,1 млн** научных публикаций;
- **Dimensions** (Free) (Великобритания), бесплатная версия – **155,4 млн** публикаций;
- **OpenAlex** – более 250 млн научных публикаций.

Отбор публикаций в метаданных

- **The Lens** – Title, Abstract, Keyword or Field of Study;
- **Dimensions** – Title, Abstract;
- **OpenAlex**: Title, Abstract



Тематика:

Перспективная композитная керамика (ПКК)

- термостойкие материалы,
- изностостойкие и антикоррозионные покрытия,
- термоэлектрики,
- катализаторы,
- многофункциональные мембраны



Содержание и полнота тематической выборки статей

- **состав профильных источников;**
- **набор классификационных рубрик;**
- **полнота загрузки выпусков профильных источников в БД;**
- **полнота метаданных (наличие рефератов, авторских ключевых слов)**

Поисковые запросы для отбора статей по ПМК (2018-2024) в БД The Lens

ceramic* AND (composite* OR poros* OR porous* OR membrane* OR "high entropy" OR "ultra high temperature")

353 Journal Subjects

**метаданные описаний
СИ в Crossref**

697.9 тыс. Fields of Study

**парсинг всех
доступных текстовых
данных в описаниях
статей**

The Lens (Subjects):

**Materials Chemistry,
Ceramics and
Composites;
General Materials
Science,
Electronic, Optical and
Magnetic Materials;
Surfaces, Coatings and
Films;**

The Lens Fields of Study:

**Materials science,
Ceramic,
Composite material,
Porosity,
Chemical engineering,
Microstructure,
Membrane,
Coating,
Chemistry,
Spark plasma sintering**

Поисковые запросы для отбора статей по ПКК (2018-2024) в БД Dimensions и OpenAlex

**ceramic AND (composite OR poros OR porous OR
membrane OR "high entropy" OR "ultra high
temperature")**

Dimensions

**Fields of Research (ANZRSC
2020) :**

4016 Materials Engineering

34 Chemical Sciences

4004 Chemical Engineering

OpenAlex*

4 Domains

26 Fields

193 Subfields

200 Topics

Статистика распределения количества статей ПКК по годам

Год	The Lens		Dimensions	OpenAlex
	Subjects	Fields of Study		
2018	2006	3717	2272	2363
2024	⁺ 2152	5343	4103	2905
Всего 2018-2024	15497	41996	21997	18855

Доля метаданных с наличием рефератов в БД The Lens и OpenAlex

Количество статей	The Lens (Subjects)		The Lens (Fields of Study)		OpenAlex	
	2018	2024	2018	2024	2018	2024
Всего	2002	2152	3717	5343	2363	2905
с рефератом	1892	1481	3602	4456	1818	2141
Доля статей с рефератом, %	94,5	68,8	96,9	83,4	76,9	73,7



Перечни источников

The Lens	Dimensions	OpenAlex
100 наиболее релевантных СИ	+	только по API

Покрытие зон Брэдфорда публикациями ТОР 100 СИ

Количество статей	The Lens (Subjects)		The Lens (Fields of Study)		Dimensions	
	Годы		Годы		Годы	
	2018	2024	2018	2024	2018	2024
Количество статей , всего	2002	2152	3717	5343	273	4103
ТОР 100 СИ, статей	1703	1851	2546	3035	1556	2667
Доля статей в ТОР 100 СИ,%	85	86	68,5	56,8	68	65

######

Степень пересечения перечней сериальных изданий в БД

Коэффициент Жаккара: $KJ = c/(a+b-c)$

где а – количество журналов в 1-м списке,

b – количество журналов во 2-м списке,

с - количество журналов, общих для 1-го и 2-го списков

Количество СИ	The Lens (Subjects)	The Lens (Fields of Study)	Dimensions
Всего	100	100	100
Общих , 2018-2024	68	47	56
Коэффициент Жаккара, %	51,52	30,72	38,89

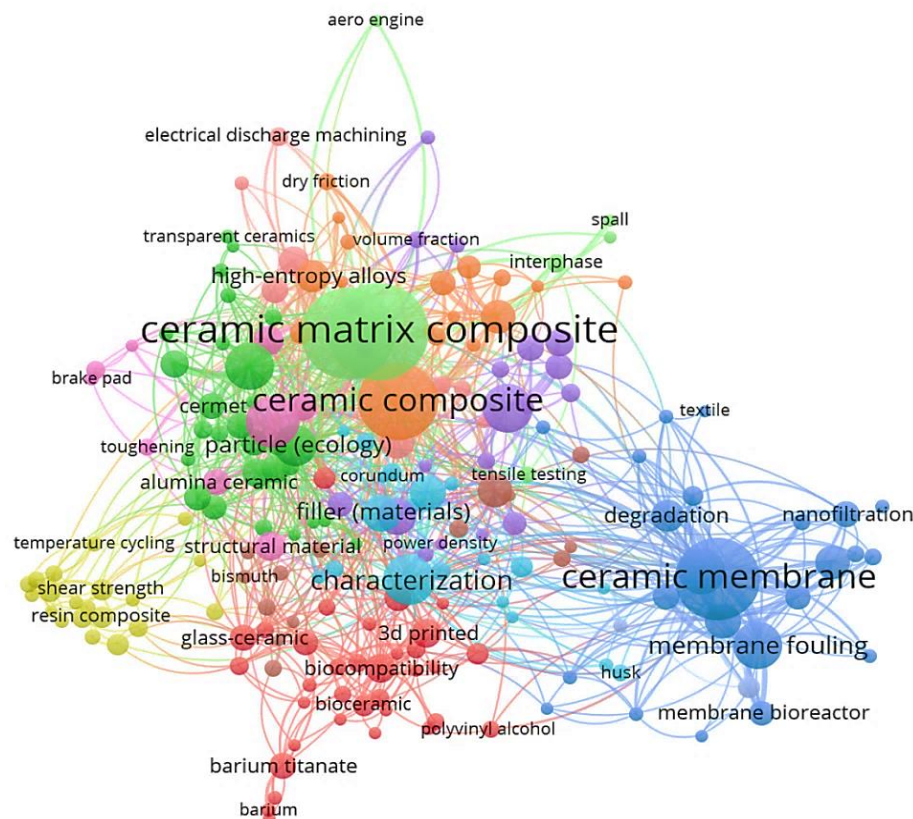
Наличие авторских КС в метаданных статей по ПКК (программа визуализации данных VOSviewer)

Количество	The Lens (Subjects)		The Lens (Fields of Study)		OpenAlex		SciDir	
	2018	2024	2018	2024	2018	2024	2018	2024
Документов в выборке	2002	2150	3717	5343	2363	2905	605	1069
Ключевых слов (КС), частота ≥ 1	919	1389	1187	945	1297	1214	1833	3104
КС / документов	0,46	0,65	0,32	0,18	0,55	0,42	3,03	2,90
КС ≥ 3 (SciDir), КС ≥ 5 (OpenAlex)	—	—	—	—	200	178	153	281

Ключевые термины по ПКК в OpenAlex с ростом общей силы связи (VOSviewer)

КС	Перевод КС на русский язык	Частота 2018	Частота 2024	ОСС 2018	ОСС 2024	Прирост частоты %	Прирост ОСС, %
matrix (chemical analysis)	матрица (хим. анализ)	114	238	174	327	108,8	87,9
ceramic composite	композиц. керамика	37	177	31	127	378,4	309,7
fast ion conductor	быстрый ионный проводник	21	69	4	44	228,6	1000,0
spark plasma sintering	искровое плазм. спекание	65	65	83	100	0,0	20,5

Карта визуализации (VOSviewer) значимости терминов в тематической коллекции ПКК в OpenAlex (2024)

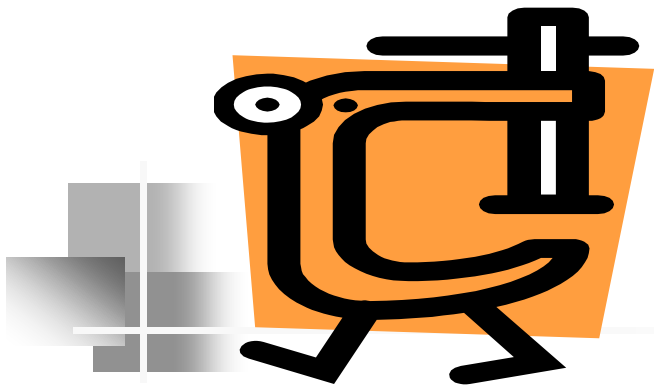


Заключение

Характеристики БД	The Lens	Dimensions	OpenAlex
Набор фильтров	+	+	+
Выбор предметных областей	области исследования (Fields of Study) и предметные рубрики (Subjects)	области исследования (Fields of Research) (ANZSRC)	26 предметных областей (fields) 193 предметных категорий (subfields)
Возможность выгрузки метаданных	полностью	max 2500 документов	полностью
Статистика кол-ва статей по годам	+	+	+
Перечень источников	100 наиболее релевантных	+	только по API

Заключение (2)

Характеристики БД	The Lens	Dimensions	OpenAlex
Полнота метаданных (наличие реф.)	2024 : $\approx 76\%$ (Subjects)	--	2024 г.: $\approx 74\%$
Полнота метаданных (наличие авторских КС)	<u>Subjects</u>: 0,46 (2018); 0,65 (2024) <u>FoS</u>: 0,32 (2018); 0,18 (2024)	—	0,55 (2018); 0,42 (2024)



Солошенко Н.С., solns@viniti.ru
Федорец О.В., ovf@viniti.ru