

数据格式说明

最建议的方式：自己去可视化界面看看！

[kibana: ES可视化分析工具](#)

用户名：elastic

密码为：IIZsZ3id5WueUPa2_53t

(无论如何，找样例都是要上的)

[neo4j门户](#)

用户名：neo4j

密码为：afuvqtepid

如果你叛逆期到了：

Author

Es

字段	解释	说明
display_name	学者姓名	String
domain	学者所在的领域	String, 格式为：领域名 & 领域ID & 领域层级 & 领域活跃度 下一条领域数据
institution	学者所在的机构	String, 格式为 机构名 & 机构ID, 只有当前所在的机构
most_cited_work	学者代表作	String, 只有名字没有ID
cited_by_count	引用量	integer
summay_stats	各种统计信息	String, 格式是json
counts_by_year	每年的作品量和引用量	String, 格式是json
id	学者的id	String
orcid	另一个id	String

Neo4j

字段	解释
ID	String, 拿数字搜搜不到
Name	String, 是Author数据集里的名字, 与论文作者署名不一定相同 (一个作者名字有不同表示)

Concept

ES

字段	解释	说明
display_name	领域名字	String
description	描述	String
id	领域id	String
level	几级学科	范围[0,5]
counts_by_year	每年的统计信息	String, Json数组
image_url	领域图片	前端应该可以直接用的图片
created_date	创建时间	理解错了, 这个创建时间是openalex数据库中这个领域数据的创建时间, 没什么意义
summary_stats	统计信息	String, Json格式, 像总论文数, h指数等
ancestors	父领域	多级学科中的学科与子学科的关系, 例如有机化学和化学, String, Json数组, 每一个数据都有
sons	子领域	子学科, 只有0级学科有子领域的数据

Institution

ES

字段	解释	说明
display_name	名字	
display_name_acronyms	简称	可能为空
country_code	国家编码	
image_url	机构图片	前端可以直接用
id	id	String
summary_stats	统计信息	String, Json格式, 像总论文数, h指数等
counts_by_year	每年的统计信息	String, Json数组

字段	解释	说明
homepage_url	机构主页	
type	机构类型	school/funder这类的，学校，基金会，研究所这些
geo	机构地址	Json，国家，城市，街道，有的字段没有
ror	另一种机构id	
associated_institutions	相关机构	Json数组，包括机构和机构间关系，例如子机构
domain	领域	机构所处领域，跟上面的一样

Work

因为这个数据最多，所以希望能少占一点空间，存储方式部分比较奇葩qaq

Es

字段	解释	说明
title	标题	
id	id	
abstract	摘要	就是一大段文字，不过也可以按关键词搜索，可能没有
domain	领域	跟上面的领域都一样
cited_count	引用量	integer

字段	解释	说明
author_all	全部作者	String, 格式为作者名 & 作者id 下一条数据
pdf_url	下载地址	String, 前端可以直接通过这个地址下载, 可能为空
landing_page_url	发布页	String, 可能没有
source	来源	指期刊, 学报之类的
publication_date	发表日期	Date
counts_by_year	每年引用量	String, Json数组
corresponding_author_ids	相关作者id	String, Json数组, 比全部作者数量少
corresponding_institution_ids	相关机构id	String, Json数组, 可以用来搜索每个机构的论文
referenced_works	引用作品id	String, Json数组, 用来在详情页返回引用论文

字段	解释	说明
related_works	相关作品id	String, json数组, 用来在详情页推荐相似论文
type_num	筛选字段	如果前端有需求的话, 这个字段可以用来筛选是否有来源, 是否有摘要, 是否有发布页, 是否能下载, integer

```
if abstract is not None:
    type_num += 1000
if source is not None:
    type_num += 100
if landing_page_url is not None:
    type_num += 10
if pdf_url is not None:
    type_num += 1
```

type_num具体计算方法, 根据筛选条件计算可能的type_num然后扔进去搜就能筛选了