OpenAlex Analysis

Generated by Doxygen 1.9.1

1 OpenAlex Analysis	1
1.1 Examples	. 1
1.1.1 Basic	. 1
1.1.2 Concepts yearly count	. 2
1.2 Configure the library	. 2
1.2.1 Default settings	. 3
2 Data folder	5
3 Hierarchical Index	7
3.1 Class Hierarchy	. 7
4 Class Index	9
4.1 Class List	. 9
5 Class Documentation	11
5.1 openalex_analysis.analysis.entities_analysis.AnalysisConfig Class Reference	. 11
5.1.1 Detailed Description	. 12
5.2 openalex_analysis.analysis.entities_analysis.AuthorsAnalysis Class Reference	. 13
5.3 openalex_analysis.plot.entities_plot.AuthorsPlot Class Reference	. 14
5.4 openalex_analysis.analysis.entities_analysis.ConceptsAnalysis Class Reference	. 15
5.5 openalex_analysis.plot.entities_plot.ConceptsPlot Class Reference	. 16
5.6 openalex_analysis.names.entitie_names.EntitieNames Class Reference	. 17
5.6.1 Detailed Description	. 17
5.7 openalex_analysis.analysis.entities_analysis.EntitiesAnalysis Class Reference	. 17
5.7.1 Detailed Description	. 19
5.7.2 Constructor & Destructor Documentation	. 19
5.7.2.1init()	. 19
5.7.3 Member Function Documentation	. 20
5.7.3.1 add_average_combined_concept_score_to_multi_concept_entitie_df()	. 20
5.7.3.2 auto_remove_databases_saved()	. 20
5.7.3.3 create_multi_concept_filters_entities_dataframe()	. 20
5.7.3.4 get_api_query()	. 21
5.7.3.5 get_count_entities_matched()	. 21
5.7.3.6 get_database_file_name()	. 21
5.7.3.7 get_df_filtered_entities_selection_threshold()	. 22
5.7.3.8 get_entitie_string_name()	. 22
5.7.3.9 get_entitie_type_from_id()	. 22
5.7.3.10 get_name_of_entitie()	. 23
5.7.3.11 get_number_of_entities_selected()	. 23
5.8 openalex_analysis.plot.entities_plot.EntitiesPlot Class Reference	. 24
5.8.1 Detailed Description	. 25
5.8.2 Member Function Documentation	. 25
5.8.2.1 get_figure_entities_of_a_concept_color_country()	. 25

Index

5.8.2.2 get_figure_entities_selection_threshold()	ì
5.8.2.3 get_figure_time_series_element_used_by_entities()	26
5.9 openalex_analysis.analysis.entities_analysis.InstitutionsAnalysis Class Reference	27
5.9.1 Detailed Description	35
5.9.2 Member Function Documentation	35
5.9.2.1 filter_and_format_entitie_data_from_api_response()	35
5.9.2.2 get_sum_concept_scores()	35
5.10 openalex_analysis.plot.entities_plot.InstitutionsPlot Class Reference	26
5.10.1 Detailed Description	30
5.10.2 Member Function Documentation	30
5.10.2.1 get_figure_institutions_multi_concepts_filtered()	30
5.10.2.2 getCustomData()	30
5.10.2.3 getHoverTemplate()	31
5.11 openalex_analysis.analysis.entities_analysis.PublishersAnalysis Class Reference	32
5.12 openalex_analysis.plot.entities_plot.PublishersPlot Class Reference	33
5.13 openalex_analysis.analysis.entities_analysis.SourcesAnalysis Class Reference	34
5.14 openalex_analysis.plot.entities_plot.SourcesPlot Class Reference	35
5.15 openalex_analysis.analysis.entities_analysis.WorksAnalysis Class Reference	36
5.15.1 Detailed Description	37
5.15.2 Member Function Documentation	38
5.15.2.1 add_statistics_to_element_count_array()	38
5.15.2.2 create_element_used_count_array()	38
5.15.2.3 filter_and_format_entitie_data_from_api_response()	36
5.15.2.4 get_country_code()	35
5.15.2.5 get_institution_name()	36
5.15.2.6 get_works_concepts_count()	łC
5.15.2.7 get_works_references_count()	łC
5.15.2.8 sort_count_array()	łC
5.16 openalex_analysis.plot.entities_plot.WorksPlot Class Reference	<b>l</b> 1
5.16.1 Detailed Description	<b>l</b> 1
5.16.2 Member Function Documentation	<b>l</b> 1
5.16.2.1 get_figure_nb_time_referenced()	ł 1
5.16.2.2 getCustomData()	12
5.16.2.3 getHoverTemplate()	12

43

# **Chapter 1**

# **OpenAlex Analysis**

Scientific literature analysis using the OpenAlex API.

This repo provides classes and methods to extract statistics, plots and graphs, as well as examples in Jupyter Notebooks.

```
Install with:
pip install openalex-analysis
```

A web app based on the library is available here.

Documentation: https://romain894.github.io/openalex-analysis

OpenAlex documentation Explore OpenAlex in a GUI

Licence: GPL V3

# 1.1 Examples

More examples can be found in the notebooks  $Works\_examples.ipynb$  and  $Concepts\_Works\_examples.ipynb$ 

### 1.1.1 Basic

In the example, we create a dataset with the works about sustainability.

This dataset can be used as it, it is stored in a parquet file (more optimized than CSV) on the computer and can be simply imported as a dataframe with Pandas.

After getting this dataset, we continue by extracting the most cited articles by the dataset. For that, we extract all the references of the articles present in the dataset and rank these references.

```
from openalex_analysis.plot import WorksPlot concept_sustainability_id = 'c66204764' # get the works about sustainability wplt = WorksPlot(concept_sustainability_id) print("\nFirst entities in the dataset:") print(wplt.entities_df[['id', 'title']].head(3)) # compute the most cited works by the dataset previously downloaded wplt.create_element_used_count_array('reference') print("\nMost cited work within the dataset:") print(wplt.element_count_df.head(3))
```

2 OpenAlex Analysis

```
Loading dataframe of works of the concept C66204764
Loading the list of entities from a parquet file...
First entities in the dataset:
                                 id
                                                                                 title
  https://openalex.org/W2101946146 Asset Stock Accumulation and Sustainability of...
  https://openalex.org/W1999167944 Planetary boundaries: Guiding human developmen...
 https://openalex.org/W2122266551 Agricultural sustainability and intensive prod...
Getting name of C66204764 from the OpenAlex API (cache disabled)...
Creating the works references count of works C66204764...
Most cited work within the dataset:
                                  C66204764 Sustainability
element
https://openalex.org/W2026816730
                                                       2.62
https://openalex.org/W2096885696
https://openalex.org/W2103847341
                                                       203
```

## 1.1.2 Concepts yearly count

In this example, we will create two datasets: one with the articles about sustainability of the SRC (Stockholm Resilience Centre) and one with the articles about sustainability of the UTT (University of Technology of Troyes).

We will then plot the yearly usage of the concept sustainability by these institutions (in this case it's equal to the number of articles in the dataset, as the dataset contains only the articles about sustainability).

We could also plot the yearly usage of other concepts or of the references by changing the parameters of the functions  $create\_element\_used\_count\_array()$  and  $get\_figure\_time\_series\_element\_\leftrightarrow$ 

```
used_by_entities().
from openalex_analysis.plot import InstitutionsPlot, WorksPlot
concept_sustainability_id = 'C66204764'
# create the filter for the API to get only the articles about sustainability
sustainability_concept_filter = {"concepts": {"id": concept_sustainability_id}}
# set the years we want to count
count_years = list(range(2004, 2024))
institution_ids_list = ["I138595864", "I140494188"]
institution_names_list = ["Stockholm Resilience Centre", "University of Technology of Troyes"]
# create a list of dictionaries with each dictionary containing the ID, name and filter for each institution
entities_ref_to_count = [None] * len(institution_ids_list)
for i in range(len(institution_ids_list)):
    entities_ref_to_count[i] = {'entitie_from_id': institution_ids_list[i],
                                       'extra_filters': sustainability_concept_filter,
'entitie_name': institution_names_list[i]}
wplt = WorksPlot()
wplt.create_element_used_count_array('concept', entities_ref_to_count, count_years = count_years)
wplt.add_statistics_to_element_count_array(sort_by = 'sum_all_entities', min_concept_level = 2)
wplt.get_figure_time_series_element_used_by_entities().write_image("Plot_yearly_usage_sustainability_SRC_UTT.svg",
        width=1200)
wplt.get_figure_time_series_element_used_by_entities()
```

# 1.2 Configure the library

By default, the library will run out of the box. Nevertheless, some optional configurations can be done to improve the performance and to fit best the use case.

Setting up the email address allows you to use the polite pool from OpenAlex which is faster than the default one. from openalex\_analysis.plot import config, InstitutionsPlot config.email = "email@example.com"
InstitutionsPlot()

The notebook Setup\_example.ipynb contains more setup examples.

#### 1.2.1 Default settings

```
config.email = None
config.api key = None
config.openalex_url = "https://api.openalex.org"
config.allow_automatic_download = True
config.disable_tqdm_loading_bar = False
config.n_max_entities = 10000
config.project_datas_folder_path = "data".
config.parquet_compression = "brotli"
config.max_storage_percent = 95
config.redis_enabled = False
 Uncomment the following lines if you want to use Redis cache
 config.redis_client = StrictRedis(host=os.environ.get('DOCKER_REDIS_URL', "localhost"),
                                   decode_responses=True,
                                   port=6379.
                                   db=2,)
 config.redis_cache = RedisCache(redis_client=config_redis_client)
 Don't forget to add the following two lines with all the imports
  from redis import StrictRedis
 from redis_cache import RedisCache
```

- email The email address is need to access the polite pool from OpenAlex which is faster than the default one.
- api\_key Optional, if you have one from OpenAlex
- openalex\_url OpenAlex URL
- allow\_automatic\_download Allow the library to download dataset from OpenAlex if not already
  present on the disk
- disable\_tqdm\_loading\_bar If set to True, it will disable the loading bar in the terminal output when downloading data from the OpenAlex API.
- n\_max\_entities When downloading a list of entities from the API (eg a list of works), the maximum number of entities to download. Set to None to have no limitation. This number must be a multiple of 200 (the is the number of element per page used by the library)
- project\_datas\_folder\_path Path to store the data downloaded from the API. The data will be stored as parquet files, with each file corresponding to one request.
- parquet\_compression By default, the parquet files are compressed. The compression can be disabled
  by setting with parquet\_compression = None. For other parquet compression algorithms, see the pandas
  documentation. Compressing reduces by 2 to 10 the file size while needing a negligeable time to compress
  or decompress. Disabling the compression is usefull if you want to read the parquet files with an external
  software.
- max\_storage\_percent Maximum storage usage percentage on the disk before starting to delete data stored in project\_datas\_folder\_path. The parquet file with the oldest last read data will be deleted first.
- redis\_enabled Whenever Redis cache is enabled or not
- redis\_client The Redis client configuration. Don't forget to add from redis import Strict← Redis where the configuration is defined.
- redis\_cache The Redis cache configuration. Don't forget to add from redis\_cache import RedisCache where the configuration is defined.

Romain Thomas 2023

4 OpenAlex Analysis

# Chapter 2

**Data folder** 

6 Data folder

# **Chapter 3**

# **Hierarchical Index**

# 3.1 Class Hierarchy

This inheritance list is sorted roughly, but not completely, alphabetically:

dict
openalex_analysis.analysis.entities_analysis.AnalysisConfig
openalex_analysis.names.entitie_names.EntitieNames
openalex_analysis.plot.entities_plot.EntitiesPlot
openalex_analysis.plot.entities_plot.AuthorsPlot
openalex_analysis.plot.entities_plot.ConceptsPlot
openalex_analysis.plot.entities_plot.InstitutionsPlot
openalex_analysis.plot.entities_plot.PublishersPlot
openalex_analysis.plot.entities_plot.SourcesPlot
openalex_analysis.plot.entities_plot.WorksPlot
Authors
openalex_analysis.analysis.entities_analysis.AuthorsAnalysis
openalex_analysis.plot.entities_plot.AuthorsPlot
Concepts
openalex_analysis.analysis.entities_analysis.ConceptsAnalysis
openalex_analysis.plot.entities_plot.ConceptsPlot
EntitieNames
openalex_analysis.analysis.entities_analysis.EntitiesAnalysis
openalex_analysis.analysis.entities_analysis.AuthorsAnalysis
openalex_analysis.analysis.entities_analysis.ConceptsAnalysis
openalex_analysis.analysis.entities_analysis.InstitutionsAnalysis
openalex_analysis.plot.entities_plot.InstitutionsPlot
openalex_analysis.analysis.entities_analysis.PublishersAnalysis
openalex_analysis.plot.entities_plot.PublishersPlot
openalex_analysis.analysis.entities_analysis.SourcesAnalysis
openalex_analysis.plot.entities_plot.SourcesPlot
openalex_analysis.analysis.entities_analysis.WorksAnalysis
openalex_analysis.plot.entities_plot.WorksPlot
Institutions
openalex_analysis.analysis.entities_analysis.InstitutionsAnalysis
Publishers
openalex_analysis.analysis.entities_analysis.PublishersAnalysis
Sources
openalex_analysis.analysis.entities_analysis.SourcesAnalysis
Works
openalex_analysis.analysis.entities_analysis.WorksAnalysis

8 Hierarchical Index

# **Chapter 4**

# **Class Index**

# 4.1 Class List

Here are the classes, structs, unions and interfaces with brief descriptions:

openalex_analysis.analysis.entities_analysis.AnalysisConfig	
TODO	11
openalex_analysis.analysis.entities_analysis.AuthorsAnalysis	13
openalex_analysis.plot.entities_plot.AuthorsPlot	14
openalex_analysis.analysis.entities_analysis.ConceptsAnalysis	15
openalex_analysis.plot.entities_plot.ConceptsPlot	16
openalex analysis.names.entitie names.EntitieNames	17
openalex_analysis.analysis.entities_analysis.EntitiesAnalysis	
OpenAlexAnalysis class which contains generic methods to do analysis over OpenAlex entities	17
openalex analysis.plot.entities plot.EntitiesPlot	
EntitiesPlot class which contains generic methods to do plots of OpenAlex entities	24
openalex_analysis.analysis.entities_analysis.InstitutionsAnalysis	
This class contains specific methods for Institutions concepts analysis	27
openalex_analysis.plot.entities_plot.InstitutionsPlot	
This class contains specific methods for Institutions concepts plot	29
openalex_analysis.analysis.entities_analysis.PublishersAnalysis	32
openalex_analysis.plot.entities_plot.PublishersPlot	33
openalex_analysis.analysis.entities_analysis.SourcesAnalysis	34
openalex_analysis.plot.entities_plot.SourcesPlot	35
openalex_analysis.analysis.entities_analysis.WorksAnalysis	
This class contains specific methods for Works concepts analysis	36
openalex analysis.plot.entities plot.WorksPlot	-
This class contains specific methods for Works concepts plot	41
This states sentante openine methode for trente denotate plat 1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.	- "

10 Class Index

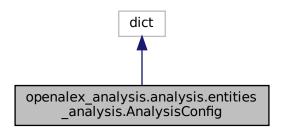
# **Chapter 5**

# **Class Documentation**

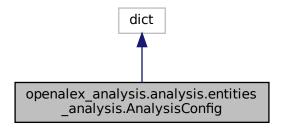
# 5.1 openalex\_analysis.analysis.entities\_analysis.AnalysisConfig Class Reference

TODO.

Inheritance diagram for openalex\_analysis.analysis.entities\_analysis.AnalysisConfig:



 $Collaboration\ diagram\ for\ openalex\_analysis.analysis.entities\_analysis.AnalysisConfig:$ 



# **Public Member Functions**

- def \_\_getattr\_\_ (self, key)
- def \_\_setattr\_\_ (self, key, value)

# 5.1.1 Detailed Description

TODO.

#### **Parameters**

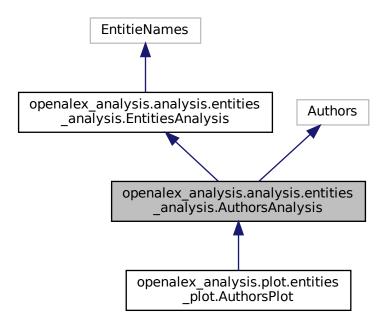
email	Email
api_key	API key
openalex_url	OpenAlex URL
http_retry_times	HTTP retry times
allow_automatic_download	The allow automatic download (True/False)
disable_tqdm_loading_bar	The disable tqdm loading bar (True/False)
n_max_entities	TODO
project_datas_folder_path	TODO
parquet_compression	TODO
max_storage_percent	TODO
redis_parameters	TODO TO UPDATE DOC

The documentation for this class was generated from the following file:

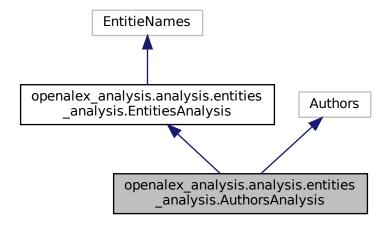
• openalex\_analysis/analysis/entities\_analysis.py

# 5.2 openalex\_analysis.analysis.entities\_analysis.AuthorsAnalysis Class Reference

Inheritance diagram for openalex\_analysis.analysis.entities\_analysis.AuthorsAnalysis:



Collaboration diagram for openalex\_analysis.analysis.entities\_analysis.AuthorsAnalysis:



## **Static Public Attributes**

• EntitieOpenAlex = Authors

#### **Additional Inherited Members**

The documentation for this class was generated from the following file:

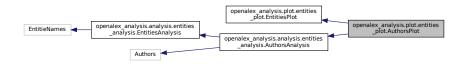
• openalex\_analysis/analysis/entities\_analysis.py

# 5.3 openalex\_analysis.plot.entities\_plot.AuthorsPlot Class Reference

Inheritance diagram for openalex\_analysis.plot.entities\_plot.AuthorsPlot:



Collaboration diagram for openalex\_analysis.plot.entities\_plot.AuthorsPlot:



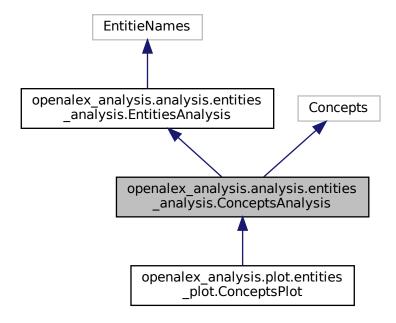
#### **Additional Inherited Members**

The documentation for this class was generated from the following file:

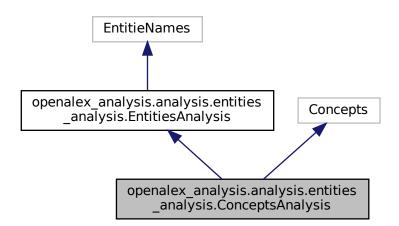
• openalex\_analysis/plot/entities\_plot.py

# 5.4 openalex\_analysis.analysis.entities\_analysis.ConceptsAnalysis Class Reference

Inheritance diagram for openalex\_analysis.analysis.entities\_analysis.ConceptsAnalysis:



Collaboration diagram for openalex analysis.analysis.entities analysis.ConceptsAnalysis:



## **Static Public Attributes**

• EntitieOpenAlex = Concepts

#### **Additional Inherited Members**

The documentation for this class was generated from the following file:

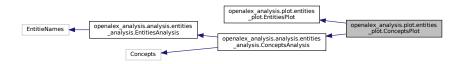
· openalex\_analysis/analysis/entities\_analysis.py

# 5.5 openalex\_analysis.plot.entities\_plot.ConceptsPlot Class Reference

Inheritance diagram for openalex\_analysis.plot.entities\_plot.ConceptsPlot:



 $Collaboration\ diagram\ for\ openalex\_analysis.plot.entities\_plot.Concepts Plot:$ 



#### **Additional Inherited Members**

The documentation for this class was generated from the following file:

• openalex\_analysis/plot/entities\_plot.py

# 5.6 openalex\_analysis.names.entitie\_names.EntitieNames Class Reference

#### **Static Public Attributes**

- string concepts\_parquet\_file\_path = "list\_all\_concepts.parquet"
   CONCEPTS #####.
- concepts\_df = pd.read\_parquet(BytesIO(pkgutil.get\_data(\_\_name\_\_, concepts\_parquet\_file\_path)))
- concepts\_names = concepts\_df[['openalex\_id', 'display\_name']].set\_index('openalex\_id')['display\_
  name'].to\_dict()
- concepts\_levels = concepts\_df[['openalex\_id', 'level']].set\_index('openalex\_id')['level'].to\_dict()
- concepts\_normalized\_names = concepts\_df[['openalex\_id', 'normalized\_name']].set\_index('openalex\_⇔ id')['normalized\_name'].to\_dict()
- string institutions\_parquet\_file\_path = "list\_all\_institutions.parquet"
   INSTITUTIONS #####.
- institutions\_df = pd.read\_parquet(BytesIO(pkgutil.get\_data(\_\_name\_\_, institutions\_parquet\_file\_path)))
- institutions\_names = institutions\_df[['id', 'display\_name']].set\_index('id')['display\_name'].to\_dict()

## 5.6.1 Detailed Description

Class to manage the concept and institution names

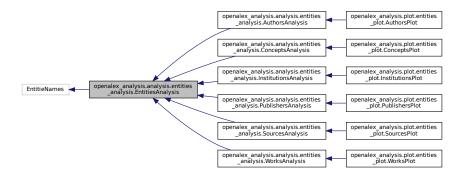
The documentation for this class was generated from the following file:

• openalex\_analysis/names/entitie\_names.py

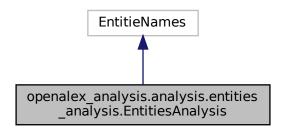
# 5.7 openalex\_analysis.analysis.entities\_analysis.EntitiesAnalysis Class Reference

OpenAlexAnalysis class which contains generic methods to do analysis over OpenAlex entities.

Inheritance diagram for openalex analysis.analysis.entities analysis.EntitiesAnalysis:



Collaboration diagram for openalex\_analysis.analysis.entities\_analysis.EntitiesAnalysis:



#### **Public Member Functions**

- def \_\_init\_\_ (self, entitie\_from\_id=None, extra\_filters=None, database\_file\_path=None, create\_
   dataframe=True, entitie name=None, load only columns=None)
- def get\_count\_entities\_matched (self, query\_filters)

Gets and return the number of entities which match the query fitlers.

def get\_api\_query (self)

Gets the api query from the parameters of the instance.

def download\_list\_entities (self)

Downloads the entities which match the parameters of the instance, and store the dataset as a parquet file.

def load\_entities\_dataframe (self)

Loads an entities dataset from file (or download it if needed and allowed by the instance) to the dataframe of the instance.

• def auto\_remove\_databases\_saved (self)

Remove databases files (the data downloaded from OpenAlex) if the storage is full.

def get\_df\_filtered\_entities\_selection\_threshold (self, df\_filters)

Gets df filtered which contains the entities of self entities df fitting the filters in df filters.

- def get\_number\_of\_entities\_selected (self, x\_threshold, y\_threshold, cited\_by\_threshold, x\_datas, y\_datas)

  Gets the number of entities selected on the plot.
- def create\_multi\_concept\_filters\_entities\_dataframe (self, concepts\_from, concepts\_filters, thresholds, x\_← datas, x\_threshold, cited\_by\_threshold)

Creates the multi concept filters entities dataframe.

def add\_average\_combined\_concept\_score\_to\_multi\_concept\_entitie\_df (self, concepts\_from)

Adds a column with the average combined concept score to the multi concept entities dataframe.

def get\_database\_file\_name (self, entitie\_from\_id=None, entities\_type=None, db\_format="parquet", extra
text=None)

Gets the database file name according to the parameters of the ojbect or the arguments given.

• def get\_entitie\_string\_name (self, entitie=None)

Gets the entitie type string name.

• def get\_entitie\_type\_from\_id (self, entitie=None)

Gets the entitie type from the entitie id.

• def get name of entitie (self, entitie=None, allow download from API=True)

Gets the name of entitie.

def get\_info\_about\_entitie (self, entitie, infos=["display\_name"], return\_as\_pd\_serie=True, allow\_
 download\_from\_API=True)

#### **Public Attributes**

- per\_page
- · entitie\_from\_id
- entitie\_from\_type
- · extra\_filters
- · database\_file\_path
- entitie\_name
- · load\_only\_columns
- · entities df
- · entities\_multi\_filtered\_df
- · element\_count\_df
- · entitie\_downloading\_progress\_percentage
- · create\_element\_count\_array\_progress\_percentage
- · create\_element\_count\_array\_progress\_text
- · count\_element\_type
- · count\_element\_years
- · count\_entities\_cols

# **Static Public Attributes**

• **cc** = coco.CountryConverter()

## 5.7.1 Detailed Description

OpenAlexAnalysis class which contains generic methods to do analysis over OpenAlex entities.

## 5.7.2 Constructor & Destructor Documentation

```
5.7.2.1 __init__()
```

entitie_from_id	The entitie identifier (eg an institution id) from which to take the entities (eg the works) to analyse (str)
filters	Optional additionnal filters, refer to the documentation of openalex and pyalex for the format (dict)
database_file_path	The database file path to force the analyse over datas in a specific file (str)
create_dataframe	Create the dataframe at the initialisation (and download the data if needed and allowed)
entitie name Generated by Doxygen	To specify the name of the entitie to avoid downloading it via the API if needed

#### 5.7.3 Member Function Documentation

#### 5.7.3.1 add\_average\_combined\_concept\_score\_to\_multi\_concept\_entitie\_df()

```
def openalex_analysis.analysis.entities_analysis.EntitiesAnalysis.add_average_combined_ \leftarrow concept_score_to_multi_concept_entitie_df ( self, concepts_from )
```

Adds a column with the average combined concept score to the multi concept entities dataframe.

#### **Parameters**

#### 5.7.3.2 auto\_remove\_databases\_saved()

```
\label{lem:condition} \begin{tabular}{ll} def openalex\_analysis.analysis.analysis.entities\_analysis.EntitiesAnalysis.auto\_remove\_databases\_saved \\ ( & self ) \end{tabular}
```

Remove databases files (the data downloaded from OpenAlex) if the storage is full.

It keeps the last accessed files

#### 5.7.3.3 create\_multi\_concept\_filters\_entities\_dataframe()

Creates the multi concept filters entities dataframe.

Combines different datasets and filters them.

concepts_from	The concept datasets to import and on which the filters will be applied (list of str)	
concepts_filters	The concepts which will be used to filter (list of str)	
thresholds	The thresholds attached to each concepts to filter (list of float or int)	
x_datas	The dataframe key of the global filter (eg the number of works) (str)	
x_threshold	The threshold for the global filter (float or int)  Generated	y Doxygen
cited_by_threshold	The cited by threshold (another global filter) (float or int)	

#### 5.7.3.4 get api query()

```
def openalex_analysis.analysis.entities_analysis.EntitiesAnalysis.get_api_query ( self \ )
```

Gets the api query from the parameters of the instance.

#### Returns

The api query (dict)

#### 5.7.3.5 get\_count\_entities\_matched()

```
def openalex_analysis.analysis.entities_analysis.EntitiesAnalysis.get_count_entities_matched ( self, query\_filters )
```

Gets and return the number of entities which match the query fitlers.

#### **Parameters**

query_filters	The query filters (dict)
---------------	--------------------------

## Returns

The count entities matched (int)

#### 5.7.3.6 get\_database\_file\_name()

Gets the database file name according to the parameters of the ojbect or the arguments given.

entitie_from↔ _id	The identifier of the entitie (eg a concept id) which was used to filter the entities (eg works) in the database (str)
entities_type	The entities type in the database (eg works) (EntitieOpenAlex)
db_format	The database file format (str)
Ge®¥ <i>ṫ≴6</i> e <u>d</u> İb6y∕1Óroxygen	Extra text to add to the file name (str)

#### Returns

The database file name (str)

## 5.7.3.7 get\_df\_filtered\_entities\_selection\_threshold()

```
def openalex_analysis.analysis.entities_analysis.EntitiesAnalysis.get_df_filtered_entities_\leftrightarrow selection_threshold ( self, \\ df_filters )
```

Gets df\_filtered which contains the entities of self.entities\_df fitting the filters in df\_filters.

#### **Parameters**

df_filters	The filters in a dictionnary with for the key for the data to filter and for the value the minimum	1
	threshold (dict)	

#### Returns

df\_filtered, corresponding the the entities fitting the thresholds (pandas DataFrame)

## 5.7.3.8 get\_entitie\_string\_name()

Gets the entitie type string name.

#### **Parameters**

```
entitie The entitie, if not provided, the instance entitie id will be used (BaseOpenAlex)
```

#### Returns

The entitie type name (str)

#### 5.7.3.9 get\_entitie\_type\_from\_id()

```
def openalex_analysis.analysis.entities_analysis.EntitiesAnalysis.get_entitie_type_from_id ( self, \\ entitie = None \ )
```

Gets the entitie type from the entitie id.

#### **Parameters**

#### Returns

The entitie type (BaseOpenAlex)

#### 5.7.3.10 get\_name\_of\_entitie()

Gets the name of entitie.

#### **Parameters**

entitie	The entitie id, if not provided, use the one from the instance (str)
allow_download_from_API	Allow to download the entitie name from the OpenAlex API (bool)

#### Returns

The name of entitie (str)

## 5.7.3.11 get\_number\_of\_entities\_selected()

Gets the number of entities selected on the plot.

x_threshold	The x threshold (float or int)
y_threshold	The y threshold (float or int)
cited_by_threshold	The cited by threshold (float or int)
x_datas	The x datas key on the dataframe (str)
y_datas	The y datas key on the dataframe (str)

#### Returns

The number of entities selected (int)

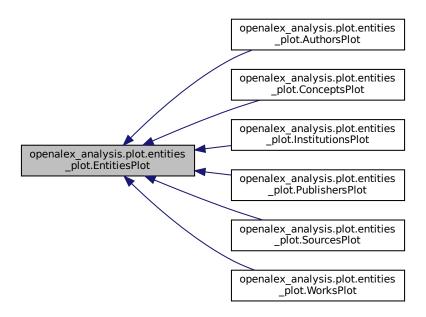
The documentation for this class was generated from the following file:

· openalex analysis/analysis/entities analysis.py

# 5.8 openalex\_analysis.plot.entities\_plot.EntitiesPlot Class Reference

EntitiesPlot class which contains generic methods to do plots of OpenAlex entities.

Inheritance diagram for openalex\_analysis.plot.entities\_plot.EntitiesPlot:



#### **Public Member Functions**

def get\_figure\_entities\_of\_a\_concept\_color\_country (self, concept, plot\_parameters=None)

Gets the figure with the entities of a concept, and with the country as color.

• def get\_figure\_entities\_selection\_threshold (self, concept, plot\_parameters, x\_threshold=0, y\_threshold=0, cited\_by\_threshold=0, display\_only\_selected\_entities=None, display\_threshold\_lines=None, entity\_to\_← highlight=None)

Gets the figure with the entities of a concept and the selection threshold lines (optional)

• def get\_figure\_time\_series\_element\_used\_by\_entities (self, element=None, plot\_title=None, x\_datas='year', x\_legend="Year", y\_datas=None, color\_legend="Entities")

Gets the figure with the time series usage of a element (eg reference, concept) by entities.

## 5.8.1 Detailed Description

EntitiesPlot class which contains generic methods to do plots of OpenAlex entities.

#### 5.8.2 Member Function Documentation

#### 5.8.2.1 get\_figure\_entities\_of\_a\_concept\_color\_country()

Gets the figure with the entities of a concept, and with the country as color.

#### **Parameters**

concept	The concept (str)
plot_parameters	The plot parameters (dict)

#### Returns

The figure (fig)

#### 5.8.2.2 get\_figure\_entities\_selection\_threshold()

Gets the figure with the entities of a concept and the selection threshold lines (optional)

concept	The concept (str)
plot_parameters	The plot parameters (dict)

## **Parameters**

x_threshold	The x threshold (float or int)
y_threshold	The y threshold (float or int)
cited_by_threshold	The cited by threshold (float or int)
display_only_selected_entities	The display only selected entities (bool)
display_threshold_lines	The display threshold lines (bool)
entity_to_highlight	The entity to highlight on the plot (str)

#### Returns

The figure (fig)

#### 5.8.2.3 get\_figure\_time\_series\_element\_used\_by\_entities()

Gets the figure with the time series usage of a element (eg reference, concept) by entities.

#### **Parameters**

element	The element (default first in the dataframe) (str)
plot_title	The plot title (str)
x_datas	The x datas (default: year) (str)
x_legend	The x legend (str)
y_datas	The y datas (the entities to plot, the default is all entities in the dataframe) (list[str])
color_legend	The color legend (str)

#### Returns

The figure (fig)

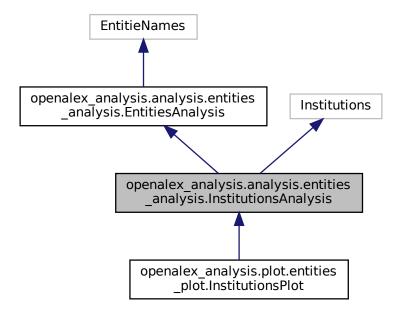
The documentation for this class was generated from the following file:

• openalex\_analysis/plot/entities\_plot.py

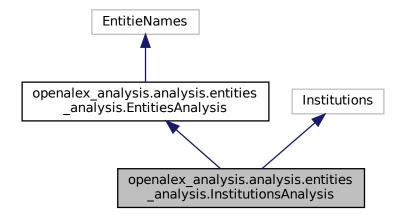
# 5.9 openalex\_analysis.analysis.entities\_analysis.InstitutionsAnalysis Class Reference

This class contains specific methods for Institutions concepts analysis.

Inheritance diagram for openalex\_analysis.analysis.entities\_analysis.InstitutionsAnalysis:



Collaboration diagram for openalex\_analysis.analysis.entities\_analysis.InstitutionsAnalysis:



#### **Public Member Functions**

• def filter\_and\_format\_entitie\_data\_from\_api\_response (self, entitie)

Filter and format the institutions data downloaded from the API.

def get\_sum\_concept\_scores (self, institutions, concept\_links)

Gets the sum of the concept scores of the concepts in the list concepts.

#### **Public Attributes**

· entitie\_from\_type

## **Static Public Attributes**

• EntitieOpenAlex = Institutions

## 5.9.1 Detailed Description

This class contains specific methods for Institutions concepts analysis.

#### 5.9.2 Member Function Documentation

#### 5.9.2.1 filter\_and\_format\_entitie\_data\_from\_api\_response()

Filter and format the institutions data downloaded from the API.

#### **Parameters**

```
entitie The institutions data from the API (dict)
```

#### Returns

The institutions datas (dict)

#### 5.9.2.2 get\_sum\_concept\_scores()

```
\label{lem:concept_scores} \mbox{ def openalex\_analysis.analysis.entities\_analysis.InstitutionsAnalysis.get\_sum\_concept\_scores \mbox{ ( } self, \mbox{ } self, \mbox{ } \mbo
```

```
institutions,
concept_links )
```

Gets the sum of the concept scores of the concepts in the list concepts.

#### **Parameters**

Institutions	The institution (list of dict)
concept_links	The concept links

#### Returns

The sum of the concept scores

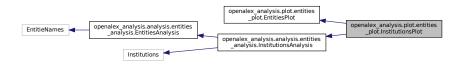
The documentation for this class was generated from the following file:

openalex\_analysis/analysis/entities\_analysis.py

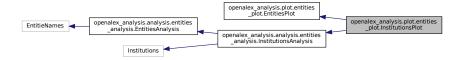
# 5.10 openalex\_analysis.plot.entities\_plot.InstitutionsPlot Class Reference

This class contains specific methods for Institutions concepts plot.

Inheritance diagram for openalex\_analysis.plot.entities\_plot.InstitutionsPlot:



Collaboration diagram for openalex\_analysis.plot.entities\_plot.InstitutionsPlot:



## **Public Member Functions**

- def getCustomData (self, concept)
   Gets the custom data for the plot.
- def getHoverTemplate (self, concept)

Gets the hover template for the plot.

• def get\_figure\_institutions\_multi\_concepts\_filtered (self, plot\_parameters, concepts\_from, concepts\_filters, thresholds, x\_threshold, cited\_by\_threshold, institution\_to\_highlight)

Gets the figure with the institutions of multiple concepts and filtered.

## **Additional Inherited Members**

## 5.10.1 Detailed Description

This class contains specific methods for Institutions concepts plot.

## 5.10.2 Member Function Documentation

## 5.10.2.1 get\_figure\_institutions\_multi\_concepts\_filtered()

Gets the figure with the institutions of multiple concepts and filtered.

#### **Parameters**

plot_parameters	The plot parameters (dict)
concepts_from	The concepts to import to create the dataset (list of str)
concepts_filters	The concepts to use to filter the institutions (list of str)
thresholds	The thresholds for each concept filter (list of float or int)
x_threshold	The global threshold (eg nb of works), usually corresponding to the x data (float or int)
cited_by_threshold	The cited by threshold (float or int)
institution_to_highlight	The institution to highlight on the plot (str)

#### Returns

The figure (fig)

#### 5.10.2.2 getCustomData()

```
def openalex_analysis.plot.entities_plot.InstitutionsPlot.getCustomData ( self, \\ concept \ )
```

Gets the custom data for the plot.

#### **Parameters**

concept	The concept (str)
---------	-------------------

#### Returns

The custom data (list of str)

#### 5.10.2.3 getHoverTemplate()

```
def openalex_analysis.plot.entities_plot.InstitutionsPlot.getHoverTemplate ( self, \\ concept \ )
```

Gets the hover template for the plot.

#### **Parameters**

concept	The concept (str)
---------	-------------------

#### Returns

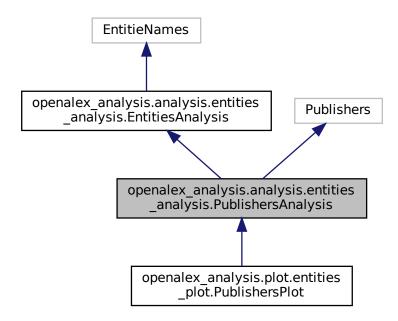
The hover template (list of str)

The documentation for this class was generated from the following file:

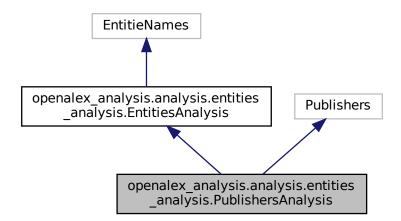
• openalex\_analysis/plot/entities\_plot.py

# 5.11 openalex\_analysis.analysis.entities\_analysis.PublishersAnalysis Class Reference

Inheritance diagram for openalex\_analysis.analysis.entities\_analysis.PublishersAnalysis:



Collaboration diagram for openalex analysis.analysis.entities analysis.PublishersAnalysis:



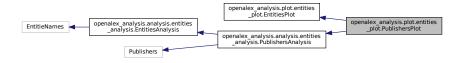
## **Additional Inherited Members**

The documentation for this class was generated from the following file:

• openalex\_analysis/analysis/entities\_analysis.py

# 5.12 openalex\_analysis.plot.entities\_plot.PublishersPlot Class Reference

Inheritance diagram for openalex\_analysis.plot.entities\_plot.PublishersPlot:



Collaboration diagram for openalex\_analysis.plot.entities\_plot.PublishersPlot:



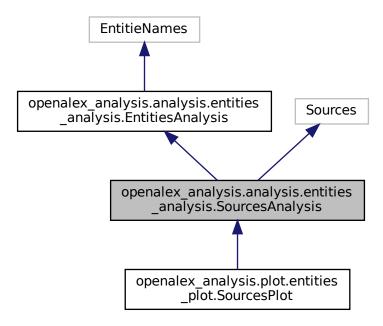
## **Additional Inherited Members**

The documentation for this class was generated from the following file:

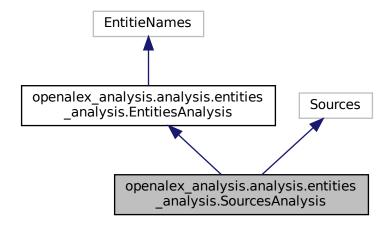
· openalex analysis/plot/entities plot.py

## 5.13 openalex\_analysis.analysis.entities\_analysis.SourcesAnalysis Class Reference

Inheritance diagram for openalex\_analysis.analysis.entities\_analysis.SourcesAnalysis:



Collaboration diagram for openalex\_analysis.analysis.entities\_analysis.SourcesAnalysis:



## **Static Public Attributes**

• EntitieOpenAlex = Sources

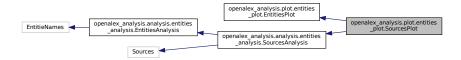
## **Additional Inherited Members**

The documentation for this class was generated from the following file:

· openalex\_analysis/analysis/entities\_analysis.py

## 5.14 openalex\_analysis.plot.entities\_plot.SourcesPlot Class Reference

Inheritance diagram for openalex analysis.plot.entities plot.SourcesPlot:



Collaboration diagram for openalex\_analysis.plot.entities\_plot.SourcesPlot:



#### **Additional Inherited Members**

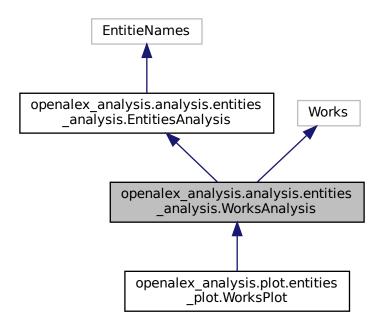
The documentation for this class was generated from the following file:

• openalex\_analysis/plot/entities\_plot.py

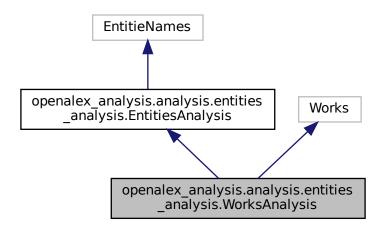
# 5.15 openalex\_analysis.analysis.entities\_analysis.WorksAnalysis Class Reference

This class contains specific methods for Works concepts analysis.

Inheritance diagram for openalex analysis.analysis.entities analysis.WorksAnalysis:



Collaboration diagram for openalex analysis.analysis.entities analysis.WorksAnalysis:



#### **Public Member Functions**

def filter\_and\_format\_entitie\_data\_from\_api\_response (self, entitie)

Filter and format the works data downloaded from the API.

• def get\_country\_code (self, entitie)

Gets the country code from an entitie.

• def get\_institution\_name (self, entitie)

Gets the institution name from an entitie.

def get\_works\_references\_count (self, count\_years=[])

Gets the works references count of the works list of the instance.

def get\_works\_concepts\_count (self, count\_years=[])

Gets the concepts count of the works list of the instance.

- def get\_element\_count (self, element\_type, count\_years=[])
- def create\_element\_used\_count\_array (self, element\_type, entities\_from=[], out\_file\_name=None, save\_← out\_array=False, count\_years=[])

Creates the element used count array.

· def sort count array (self, sort by='h used all I use main', sort by ascending=False)

Sort the dataframe with the count array (element\_count\_df)

 def add\_statistics\_to\_element\_count\_array (self, sort\_by='h\_used\_all\_l\_use\_main', sort\_by\_ascending=False, min\_concept\_level=None)

Adds a statistics to the element count array (statistics between the main entitie to compare (second column in the dataframe) and the sum of the other entities)

def add statistics to references works count array (self)

Adds a statistics to the references works count array (statistics between the main entitie to compare (second column in the dataframe) and the sum of the other entities)

def add\_statistics\_to\_concept\_count\_array (self, min\_concept\_level=None)

Adds a statistics to the concepts count array (statistics between the main entitie to compare (second column in the dataframe) and the sum of the other entities)

## **Public Attributes**

- entitie\_from\_type
- · count element type
- · count element years
- · count entities cols
- · entitie\_from\_id
- · element count df
- · create\_element\_count\_array\_progress\_percentage
- · create\_element\_count\_array\_progress\_text

## **Static Public Attributes**

• EntitieOpenAlex = Works

## 5.15.1 Detailed Description

This class contains specific methods for Works concepts analysis.

## 5.15.2 Member Function Documentation

## 5.15.2.1 add\_statistics\_to\_element\_count\_array()

Adds a statistics to the element count array (statistics between the main entitie to compare (second column in the dataframe) and the sum of the other entities)

#### **Parameters**

sort_by	The key to sort the dataframe (str)	
sort_by_ascending	Whenever to sort the dataframe ascending (bool)	
min_concept_level	In case the element is a concept, this is the minimum level of the concepts we will keep (aka remove the lower (= more global) concepts)	

#### 5.15.2.2 create\_element\_used\_count\_array()

Creates the element used count array.

Count the number of times each element (eg reference, concept..) is used and save the array as CSV (optional)

## **Parameters**

element_type	The element type	
entities_from	The extra entities to which to count the concepts (list of str)	
out_file_name	The out CSV file name, if not provided, an appropriate name is generated (str)	
save_out_array	Save out array (bool)	
count_years	If given, it will compute the count for each year (list[int])	

#### 5.15.2.3 filter\_and\_format\_entitie\_data\_from\_api\_response()

```
def openalex_analysis.analysis.entities_analysis.WorksAnalysis.filter_and_format_entitie_ \leftrightarrow data_from_api_response ( self, entitie )
```

Filter and format the works data downloaded from the API.

#### **Parameters**

entitie	The works data from the API (dict)
---------	------------------------------------

## Returns

The works datas (dict)

## 5.15.2.4 get\_country\_code()

```
def openalex_analysis.analysis.entities_analysis.WorksAnalysis.get_country_code ( self, \\ entitie \ )
```

Gets the country code from an entitie.

## **Parameters**

```
entitie The entitie (dict)
```

#### Returns

The country code (str)

## 5.15.2.5 get\_institution\_name()

```
def openalex_analysis.analysis.entities_analysis.WorksAnalysis.get_institution_name ( self, \\ entitie \ )
```

Gets the institution name from an entitie.

## **Parameters**

```
entitie The entitie (dict)
```

#### Returns

The institution name (str)

## 5.15.2.6 get\_works\_concepts\_count()

```
def openalex_analysis.analysis.entities_analysis.WorksAnalysis.get_works_concepts_count ( self, \\ count\_years = [] \ )
```

Gets the concepts count of the works list of the instance.

#### **Parameters**

count only	vear If differ	ent than None.	count only the co	ncepts of the wo	rks of the given	vear (int)
------------	----------------	----------------	-------------------	------------------	------------------	------------

#### Returns

The concept count (pandas Serie)

## 5.15.2.7 get\_works\_references\_count()

Gets the works references count of the works list of the instance.

## Returns

The works references count (pandas Serie)

## 5.15.2.8 sort\_count\_array()

Sort the dataframe with the count array (element\_count\_df)

#### **Parameters**

sort_by	The key to sort the dataframe (str)
sort_by_ascending	Whenever to sort the dataframe ascending (bool)

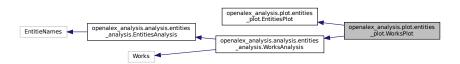
The documentation for this class was generated from the following file:

· openalex\_analysis/analysis/entities\_analysis.py

## 5.16 openalex\_analysis.plot.entities\_plot.WorksPlot Class Reference

This class contains specific methods for Works concepts plot.

Inheritance diagram for openalex analysis.plot.entities plot.WorksPlot:



Collaboration diagram for openalex\_analysis.plot.entities\_plot.WorksPlot:



#### **Public Member Functions**

• def getCustomData (self, concept)

Gets the custom data for the plot.

• def getHoverTemplate (self, concept)

Gets the hover template for the plot.

def get\_figure\_nb\_time\_referenced (self, element\_type)

Gets the figure with the number of time each reference is used in a list of works.

## **Additional Inherited Members**

## 5.16.1 Detailed Description

This class contains specific methods for Works concepts plot.

## 5.16.2 Member Function Documentation

## 5.16.2.1 get\_figure\_nb\_time\_referenced()

Gets the figure with the number of time each reference is used in a list of works.

#### **Parameters**

element_type	The element type
--------------	------------------

## Returns

The figure works number of time referenced (fig)

## 5.16.2.2 getCustomData()

Gets the custom data for the plot.

#### **Parameters**

concept	The concept (str)
---------	-------------------

#### Returns

The custom data (list of str)

## 5.16.2.3 getHoverTemplate()

Gets the hover template for the plot.

#### **Parameters**

concept	The concept (str)
---------	-------------------

## Returns

The hover template (list of str)

The documentation for this class was generated from the following file:

• openalex\_analysis/plot/entities\_plot.py

# Index

```
init
                                                                                                                                     get_figure_entities_selection_threshold
            openalex analysis.analysis.entities analysis.EntitiesAnalysispenalex analysis.plot.entities plot.EntitiesPlot,
                                                                                                                                      get_figure_institutions_multi_concepts filtered
add_average_combined_concept_score_to_multi_concept_entitipedialex_analysis.plot.entities_plot.InstitutionsPlot,
            openalex_analysis.analysis.entities_analysis.EntitiesAnalysis, 30
                                                                                                                                      get figure nb time referenced
add statistics to element count array
                                                                                                                                                 openalex_analysis.plot.entities_plot.WorksPlot, 41
            openalex_analysis.analysis.entities_analysis.WorksAngelysisgure_time_series_element_used_by_entities
                                                                                                                                                 openalex analysis.plot.entities plot.EntitiesPlot,
auto_remove_databases_saved
            ____
openalex_analysis.analysis.entities_analysis.EntitiesAnalysistitution_name
                                                                                                                                                 openalex_analysis.analysis.entities_analysis.WorksAnalysis,
                                                                                                                                                             39
create_element_used_count_array
            openalex_analysis.analysis.entities_analysis.WorksAnalysis,
                                                                                                                                                 openalex_analysis.analysis.entities_analysis.EntitiesAnalysis,
create_multi_concept_filters_entities_dataframe
           openalex_analysis.analysis.entities_analysis.EntitiesAnalysis, ____of_entities_selected
                                                                                                                                                 openalex_analysis.analysis.entities_analysis.EntitiesAnalysis,
                                                                                                                                      get sum concept scores
filter and format entitie data from api response
            openalex_analysis.analysis.entities_analysis.WorksAnsatysis_rks_concepts_count
                                                                                                                                                 openalex analysis.analysis.entities analysis.WorksAnalysis,
                       38
get_api_query
                                                                                                                                      get works references count
            openalex\_analysis. analysis. entities\_analysis. Entities Analysis fenalex\_analysis. analysis. entities\_analysis. Works Analysis, analysis. The substitution of the s
                                                                                                                                                             40
get_count_entities_matched
                                                                                                                                     getCustomData
            openalex\_analysis. analysis. entities\_analysis. Entities Analysis\_enalex\_analysis. plot. entities\_plot. Institutions Plot, analysis\_enalex\_analysis. entities\_plot. entite plot. entit plo
get country code
                                                                                                                                                 openalex_analysis.plot.entities_plot.WorksPlot, 42
            openalex_analysis.analysis.entities_analysis.WorksAngelylsisverTemplate
                                                                                                                                                 openalex_analysis.plot.entities_plot.InstitutionsPlot,
get database file name
            openalex_analysis.analysis.entities_analysis.EntitiesAnalysispenalex_analysis.plot.entities_plot.WorksPlot, 42
get_df_filtered_entities_selection_threshold
                                                                                                                                      openalex_analysis.analysis.entities_analysis.AnalysisConfig,
            openalex analysis.analysis.entities analysis.EntitiesAnalysis, 11
                                                                                                                                     openalex_analysis.analysis.entities_analysis.AuthorsAnalysis,
get_entitie_string_name
            openalex_analysis.analysis.entities_analysis.Entities_analysis.Entities_analysis.analysis.entities_analysis.ConceptsAnalysis,
get entitie type from id
                                                                                                                                     openalex analysis.analysis.entities analysis.EntitiesAnalysis,
            openalex analysis.analysis.entities analysis.EntitiesAnalysis, 17
                                                                                                                                                      init , 19
                                                                                                                                                 add average combined concept score to multi concept entitie di
get figure entities of a concept color country
            openalex_analysis.plot.entities_plot.EntitiesPlot,
                                                                                                                                                 auto_remove_databases_saved, 20
```

44 INDEX

```
create_multi_concept_filters_entities_dataframe,
    get_api_query, 21
     get_count_entities_matched, 21
     get_database_file_name, 21
    get df filtered entities selection threshold, 22
    get entitie string name, 22
     get entitie type from id, 22
     get name of entitie, 23
     get_number_of_entities selected, 23
openalex_analysis.analysis.entities_analysis.InstitutionsAnalysis,
     filter_and_format_entitie_data_from_api_response,
     get_sum_concept_scores, 28
openalex_analysis.analysis.entities_analysis.PublishersAnalysis,
openalex analysis.analysis.entities analysis.SourcesAnalysis,
openalex_analysis.analysis.entities_analysis.WorksAnalysis,
     add statistics to element count array, 38
     create_element_used_count_array, 38
     filter_and_format_entitie_data_from_api_response,
          38
    get_country_code, 39
     get_institution_name, 39
     get works concepts count, 40
     get works references count, 40
     sort count array, 40
openalex_analysis.names.entitie_names.EntitieNames,
openalex analysis.plot.entities plot.AuthorsPlot, 14
openalex_analysis.plot.entities_plot.ConceptsPlot, 16
openalex_analysis.plot.entities_plot.EntitiesPlot, 24
     get_figure_entities_of_a_concept_color_country,
     get_figure_entities_selection_threshold, 25
     get_figure_time_series_element_used_by_entities,
          26
openalex analysis.plot.entities plot.InstitutionsPlot, 29
     get figure institutions multi concepts filtered, 30
     getCustomData, 30
     getHoverTemplate, 31
openalex analysis.plot.entities plot.PublishersPlot, 33
openalex_analysis.plot.entities_plot.SourcesPlot, 35
openalex_analysis.plot.entities_plot.WorksPlot, 41
     get figure nb time referenced, 41
     getCustomData, 42
     getHoverTemplate, 42
sort count array
     openalex analysis.analysis.entities analysis.WorksAnalysis,
          40
```