The Patstat library - Lesson 4

This notebook expands on the second lesson about Patstat. We will learn to work with nested queries, also known as subqueries, can be very useful for more complex data retrieval tasks.

Example scenario

In lesson three we built a query with a double join, to display granted European patents filed this decade, the name of the inventor, and the amount of families that cite each application. Then we aggregated the results, summing the total citations for each inventor, as a proxy for the most influential inventors of the decade.

We will then use this query as a filter for an outer query that finds the patent applications for an inventor in the list, based on their ranking. This could be useful e.g. if you want to write a news article about those influential inventors.

Subqueries in PATSTAT

As we have already seen, the PATSTAT library is an implementation of SQLAlchemy.

In SQLAlchemy, when you create a subquery using the `subquery()` method, the resulting subquery object can be referenced in the outer query. This is particularly useful for nested queries where the result of one query is used as a filter or condition in another.

We will appli the `subquery()` method to the query of last example, limiting the query to the first entry to only get the first name in the list of top inventors. But first we need to initalize our ORM client for PATSTAT.

In [1]:
```
# Importing the patstat client
from epo.tipdata.patstat import PatstatClient

# Initialize the PATSTAT client
patstat = PatstatClient()

# Access ORM
db = patstat.orm()

# Importing tables as models
from epo.tipdata.patstat.database.models import TLS201_APPLN, TLS
207_PERS_APPLN, TLS206_PERSON
```

The subquery

We are going to take the query from lesson 3, and limit the result to the top inventor by adding `limit(1)` after the `query()` method. This way we will have only the top ranked `person_id`. This will be our subquery, that we will use later as a filter. In order to wrap this query in an outer query, we need to add the method `subquery()` at the end of the query.

In [6]:
```
# Importing the func model
from sqlalchemy import func

# Defining the subquery for finding the top inventors
inner = db.query(
    TLS206_PERSON.person_id,  # inventor's name
    func.sum(TLS201_APPLN.nb_citing_docdb_fam).label('total_citat
ions')  # sum of families citing patents by a given inventor
).join(
    TLS207_PERS_APPLN, TLS201_APPLN.appln_id == TLS207_PERS_APPL
N.appln_id
).join(
    TLS206_PERSON, TLS207_PERS_APPLN.person_id == TLS206_PERSON.p
erson_id
).filter(
    TLS201_APPLN.appln_filing_year >= 2020,
    TLS201_APPLN.appln_auth == 'EP',
    TLS201_APPLN.granted == 'Y',
    TLS207_PERS_APPLN.invt_seq_nr > 0  # filter to include only i
nventors
).group_by(
    TLS206_PERSON.person_id
).order_by(
    func.sum(TLS201_APPLN.nb_citing_docdb_fam).desc()  # order by
total citations in descending order
).limit(1).subquery()
```

The outer query

Now that we have a subquery that gives us the top rated `person_id` in terms of citations for their granted patents of this decade, we can use it as a filter to find the granted European patents that mention this `person_id` as an inventor. We use the line `TLS206_PERSON.person_id == inner.c.person_id` as a filter, to ensure that we only find applications that mention the top `person_id` as an inventor.

In [3]:
```
# Creating the outer query
outer_query = db.query(
    TLS201_APPLN.appln_id,
    TLS201_APPLN.appln_nr,
    TLS206_PERSON.person_name,
    TLS206_PERSON.person_id
).join(
    TLS207_PERS_APPLN, TLS201_APPLN.appln_id == TLS207_PERS_APPL
N.appln_id
).join(
    TLS206_PERSON, TLS207_PERS_APPLN.person_id == TLS206_PERSON.p
erson_id
).filter(
    TLS206_PERSON.person_id == inner.c.person_id,
    TLS201_APPLN.appln_auth == 'EP',
    TLS201_APPLN.granted == 'Y',
    TLS207_PERS_APPLN.invt_seq_nr > 0
)

# Creating a dataframe with the results
patents_df = patstat.df(outer_query)

# Display the dataframe with detailed information about the paten
ts of the selected inventor
patents_df
```

Out[3]:

	appln_id	appln_nr	person_name	person_id
0	545974287	21157430	HARRIS, Jason L.	53448894
1	468445884	16186383	HARRIS, Jason L.	53448894
2	475146383	17155675	HARRIS, Jason L.	53448894
3	543439205	20217600	HARRIS, Jason L.	53448894
4	487783532	17209358	HARRIS, Jason L.	53448894
...	...	...	...	...
102	450638724	16162067	HARRIS, Jason L.	53448894
103	469713345	16190171	HARRIS, Jason L.	53448894
104	507904622	19158219	HARRIS, Jason L.	53448894
105	470969018	16196387	HARRIS, Jason L.	53448894
106	543216412	20215721	HARRIS, Jason L.	53448894

107 rows × 4 columns

In [10]:
```
for index, row in patents_df.iterrows():
    appln_nr = row['appln_nr']
    print (f"https://register.epo.org/application?number=EP{appln
_nr}")
```

```
https://register.epo.org/application?number=EP17155675
https://register.epo.org/application?number=EP16207245
https://register.epo.org/application?number=EP16186383
https://register.epo.org/application?number=EP21157430
https://register.epo.org/application?number=EP16185375
https://register.epo.org/application?number=EP16157574
https://register.epo.org/application?number=EP18275254
https://register.epo.org/application?number=EP16162058
https://register.epo.org/application?number=EP16162048
https://register.epo.org/application?number=EP16162059
https://register.epo.org/application?number=EP17209378
https://register.epo.org/application?number=EP16186414
https://register.epo.org/application?number=EP16185892
https://register.epo.org/application?number=EP17164408
https://register.epo.org/application?number=EP16185871
https://register.epo.org/application?number=EP16185387
https://register.epo.org/application?number=EP19158301
https://register.epo.org/application?number=EP19212938
https://register.epo.org/application?number=EP15186913
https://register.epo.org/application?number=EP16185857
https://register.epo.org/application?number=EP17209673
https://register.epo.org/application?number=EP18275253
https://register.epo.org/application?number=EP19188666
https://register.epo.org/application?number=EP17177896
https://register.epo.org/application?number=EP20217600
https://register.epo.org/application?number=EP16162067
https://register.epo.org/application?number=EP16190171
https://register.epo.org/application?number=EP16157535
https://register.epo.org/application?number=EP18275227
https://register.epo.org/application?number=EP17155713
https://register.epo.org/application?number=EP16186436
https://register.epo.org/application?number=EP15186936
https://register.epo.org/application?number=EP16185376
https://register.epo.org/application?number=EP19152828
https://register.epo.org/application?number=EP16185368
https://register.epo.org/application?number=EP18180571
https://register.epo.org/application?number=EP16185858
https://register.epo.org/application?number=EP16186382
https://register.epo.org/application?number=EP19213362
https://register.epo.org/application?number=EP18180482
https://register.epo.org/application?number=EP18203367
https://register.epo.org/application?number=EP18275246
https://register.epo.org/application?number=EP17155691
https://register.epo.org/application?number=EP16157544
https://register.epo.org/application?number=EP16185899
https://register.epo.org/application?number=EP16163036
https://register.epo.org/application?number=EP16185863
```

https://register.epo.org/application?number=EP16191672
https://register.epo.org/application?number=EP17177853
https://register.epo.org/application?number=EP19188719
https://register.epo.org/application?number=EP20150529
https://register.epo.org/application?number=EP16191673
https://register.epo.org/application?number=EP19158223
https://register.epo.org/application?number=EP16190158
https://register.epo.org/application?number=EP20152602
https://register.epo.org/application?number=EP19186773
https://register.epo.org/application?number=EP16185370
https://register.epo.org/application?number=EP16191736
https://register.epo.org/application?number=EP16158830
https://register.epo.org/application?number=EP16162045
https://register.epo.org/application?number=EP15828707
https://register.epo.org/application?number=EP17209685
https://register.epo.org/application?number=EP16186861
https://register.epo.org/application?number=EP17177844
https://register.epo.org/application?number=EP16158837
https://register.epo.org/application?number=EP19196725
https://register.epo.org/application?number=EP16158807
https://register.epo.org/application?number=EP16185898
https://register.epo.org/application?number=EP17209358
https://register.epo.org/application?number=EP16191706
https://register.epo.org/application?number=EP17166586
https://register.epo.org/application?number=EP17209435
https://register.epo.org/application?number=EP17164216
https://register.epo.org/application?number=EP16186395
https://register.epo.org/application?number=EP18180376
https://register.epo.org/application?number=EP16186416
https://register.epo.org/application?number=EP16190194
https://register.epo.org/application?number=EP16207255
https://register.epo.org/application?number=EP17155701
https://register.epo.org/application?number=EP17164412
https://register.epo.org/application?number=EP16186366
https://register.epo.org/application?number=EP20157723
https://register.epo.org/application?number=EP17209648
https://register.epo.org/application?number=EP16186381
https://register.epo.org/application?number=EP16186877
https://register.epo.org/application?number=EP16186384
https://register.epo.org/application?number=EP18180458
https://register.epo.org/application?number=EP16186789
https://register.epo.org/application?number=EP16185883
https://register.epo.org/application?number=EP20181956
https://register.epo.org/application?number=EP16196332
https://register.epo.org/application?number=EP15186914
https://register.epo.org/application?number=EP19165947
https://register.epo.org/application?number=EP17177857
https://register.epo.org/application?number=EP16163032
https://register.epo.org/application?number=EP20182100
https://register.epo.org/application?number=EP16196245
https://register.epo.org/application?number=EP17209653
https://register.epo.org/application?number=EP16196387
https://register.epo.org/application?number=EP16186373

```
https://register.epo.org/application?number=EP19158306
https://register.epo.org/application?number=EP20215721
https://register.epo.org/application?number=EP19171920
https://register.epo.org/application?number=EP18157200
https://register.epo.org/application?number=EP19158219
https://register.epo.org/application?number=EP16186417
https://register.epo.org/application?number=EP16162064
```

In [ ]: