

# Crawling from Flickr (intro)

Analysis of Social Media Contents

Alessandro Ortis - University of Catania

1 - Install flickrapi:

```
pip install flickrapi
```

(more details about python flickr api: <https://stuvel.eu/flickrapi-doc/> (<https://stuvel.eu/flickrapi-doc/>))

List of Flickr API calls explored in this example:

- Photo's API
  - flickr.photos.getInfo*: get information about a photo
  - flickr.photos.search*: a list of photos matching some criteria

More details about each of the above API functions are available at <https://www.flickr.com/services/api/> (<https://www.flickr.com/services/api/>)

## Objective

Explore the Flickr API.

```
In [15]: import flickrapi
import configparser
```

```
In [16]: config = configparser.ConfigParser()
config.read('FlickrConfig.ini')

auth_info = config['AO_Tutorial']
api_key = auth_info['flickr_api_key']
api_secret = auth_info['flickr_api_secret']
```

Create a FlickrAPI instance to perform the API calls.

```
In [17]: flickr = flickrapi.FlickrAPI(api_key, api_secret, format='parsed-json')
```

Search the pictures with tag 'nature', restrict the request to 4 pictures and require the URL at the original resolution.

We are going to use the [flickr.photos.search\(\)](https://www.flickr.com/services/api/flickr.photos.search.html) (<https://www.flickr.com/services/api/flickr.photos.search.html>) API call.

```
In [18]: response = flickr.photos.search(api_key= api_key, tags='nature', extras='url_o', per_page=4)
print("request result:\t"+response['stat'])

print(response)
```

```
request result: ok
{'photos': {'page': 1, 'pages': 26290, 'perpage': 4, 'total': 105157, 'photo': [{'id': '52673122100', 'owner': '51483961@N03', 'secret': '4f7444f6e2', 'server': '65535', 'farm': 66, 'title': '', 'ispublic': 1, 'isfriend': 0, 'isfamily': 0, 'url_o': 'https://live.staticflickr.com/65535/52673122100_5c9024519a_o.jpg', 'height_o': 1125, 'width_o': 2000}, {'id': '52672977489', 'owner': '188303638@N07', 'secret': '3d905bf78c', 'server': '65535', 'farm': 66, 'title': 'Snowy blanc', 'ispublic': 1, 'isfriend': 0, 'isfamily': 0, 'url_o': 'https://live.staticflickr.com/65535/52672977489_a71d4caf5d_o.jpg', 'height_o': 4672, 'width_o': 3500}, {'id': '52672971299', 'owner': '195397489@N07', 'secret': 'ae5e8d44fa', 'server': '65535', 'farm': 66, 'title': 'Canard malla rd / Anas Platyrhynchos.', 'ispublic': 1, 'isfriend': 0, 'isfamily': 0, 'url_o': 'https://live.staticflickr.com/65535/52672971299_8f091fec0d_o.jpg', 'height_o': 7600, 'width_o': 7680}, {'id': '52673136543', 'owner': '145425803@N04', 'secret': '990888fdfb', 'server': '65535', 'farm': 66, 'title': 'West side', 'ispublic': 1, 'isfriend': 0, 'isfamily': 0, 'url_o': 'https://live.staticflickr.com/65535/52673136543_2930f4fe9d_o.jpg', 'height_o': 1356, 'width_o': 2048}]}}, 'stat': 'ok'}
```

Take a look at the output.

```
In [19]: print(response.keys())

dict_keys(['photos', 'stat'])
```

```
In [6]: print(response['stat'])

ok
```

```
In [20]: print(response['photos'].keys())

dict_keys(['page', 'pages', 'perpage', 'total', 'photo'])
```

```
In [21]: for key in response['photos']:
         value = str(response['photos'][key])
         print('\n' + key + '\t-->>\t' + value)
```

```
page    -->>    1
```

```
pages   -->>   26290
```

```
perpage -->>    4
```

```
total   -->>  105157
```

```
photo   -->>  [{ 'id': '52673122100', 'owner': '51483961@N03', 'secret': '4f7444f6e2', 'server': '65535', 'farm': 66, 'title': '', 'ispublic': 1, 'isfriend': 0, 'isfamily': 0, 'url_o': 'https://live.staticflickr.com/65535/52673122100_5c9024519a_o.jpg', 'height_o': 1125, 'width_o': 2000}, { 'id': '52672977489', 'owner': '188303638@N07', 'secret': '3d905bf78c', 'server': '65535', 'farm': 66, 'title': 'Snowy blanc', 'ispublic': 1, 'isfriend': 0, 'isfamily': 0, 'url_o': 'https://live.staticflickr.com/65535/52672977489_a71d4caf5d_o.jpg', 'height_o': 4672, 'width_o': 3500}, { 'id': '52672971299', 'owner': '195397489@N07', 'secret': 'ae5e8d44fa', 'server': '65535', 'farm': 66, 'title': 'Canard mallard / Anas Platyrhynchos.', 'ispublic': 1, 'isfriend': 0, 'isfamily': 0, 'url_o': 'https://live.staticflickr.com/65535/52672971299_8f091fec0d_o.jpg', 'height_o': 7600, 'width_o': 7680}, { 'id': '52673136543', 'owner': '145425803@N04', 'secret': '990888fdfb', 'server': '65535', 'farm': 66, 'title': 'West side', 'ispublic': 1, 'isfriend': 0, 'isfamily': 0, 'url_o': 'https://live.staticflickr.com/65535/52673136543_2930f4fe9d_o.jpg', 'height_o': 1356, 'width_o': 2048}]
```

The list of pictures is stored in *response*

```
'photos'
'photo'
```

Look at one example element.

```
In [22]: a_photo = response['photos']['photo'][3]
         print(a_photo)
         # URL of the post
         print("http://flickr.com/photo.gne?id="+a_photo['id'])
```

```
{ 'id': '52673136543', 'owner': '145425803@N04', 'secret': '990888fdfb', 'server': '65535', 'farm': 66, 'title': 'West side', 'ispublic': 1, 'isfriend': 0, 'isfamily': 0, 'url_o': 'https://live.staticflickr.com/65535/52673136543_2930f4fe9d_o.jpg', 'height_o': 1356, 'width_o': 2048}
http://flickr.com/photo.gne?id=52673136543
```

```
In [23]: for key in a_photo:
          value = str(a_photo[key])
          print('\n' + key + '\t-->\t' + value)
```

```
id      --> 52673136543
owner   --> 145425803@N04
secret  --> 990888fdfb
server  --> 65535
farm    --> 66
title   --> West side
ispublic --> 1
isfriend --> 0
isfamily --> 0
url_o   --> https://live.staticflickr.com/65535/52673136543_2930f4fe9d_o.
jpg
height_o --> 1356
width_o --> 2048
```

Now write a summary for all the retrieved pictures.

```
In [24]: photo_ls = response['photos']['photo']
print('Number of photos:\t'+str(len(photo_ls)))

for ph in photo_ls:
    print('\nTitle:\t'+ph['title'])
    print('Owner:\t'+ph['owner'])
    if 'url_o' in ph:
        print('URL:\t'+ph['url_o'])
        print("Post:\thttp://flickr.com/photo.gne?id="+ph['id'])

    else:
        print('URL:\tnot available')
```

Number of photos: 4

Title:

Owner: 51483961@N03

URL: https://live.staticflickr.com/65535/52673122100\_5c9024519a\_o.jpg

Post: http://flickr.com/photo.gne?id=52673122100

Title: Snowy blanc

Owner: 188303638@N07

URL: https://live.staticflickr.com/65535/52672977489\_a71d4caf5d\_o.jpg

Post: http://flickr.com/photo.gne?id=52672977489

Title: Canard mallard / Anas Platyrhynchos.

Owner: 195397489@N07

URL: https://live.staticflickr.com/65535/52672971299\_8f091fec0d\_o.jpg

Post: http://flickr.com/photo.gne?id=52672971299

Title: West side

Owner: 145425803@N04

URL: https://live.staticflickr.com/65535/52673136543\_2930f4fe9d\_o.jpg

Post: http://flickr.com/photo.gne?id=52673136543

Now exploit the [flickr.photos.getInfo\(\)](https://www.flickr.com/services/api/flickr.photos.getInfo.html) (<https://www.flickr.com/services/api/flickr.photos.getInfo.html>) API call to get more information about a specific photo.

```
In [25]: photo_id = a_photo['id']
print("\nProcessing photo with FlickrId:\t" +photo_id)

# GETTING PHOTO INFO
print( "Getting photo info...")
response = flickr.photos.getInfo(api_key = api_key, photo_id=photo_id)
print( "request result:\t"+response['stat'])
photo_info = response['photo']

print(photo_info)
```

Processing photo with FlickrId: 52673136543

Getting photo info...

request result: ok

```
{'id': '52673136543', 'secret': '990888fdfb', 'server': '65535', 'farm': 66,
'dateuploaded': '1675696479', 'isfavorite': 0, 'license': '0', 'safety_level': '0', 'rotation': 0, 'originalsecret': '2930f4fe9d', 'originalformat': 'jpg', 'owner': {'nsid': '145425803@N04', 'username': 'javiborbon', 'realname': 'Javi Borbón', 'location': '', 'iconserver': '65535', 'iconfarm': 66, 'path_alias': 'javiborbon', 'gift': {'gift_eligible': True, 'eligible_durations': ['year', 'month', 'week'], 'new_flow': True}}, 'title': {'_content': 'West side'}, 'description': {'_content': 'Nerín, Huesca, Aragón, Spain'}, 'visibility': {'ispublic': 1, 'isfriend': 0, 'isfamily': 0}, 'dates': {'posted': '1675696479', 'taken': '2023-01-29 18:11:19', 'takengranularity': 0, 'takenunknown': '0', 'lastupdate': '1675700751'}, 'views': '61', 'editability': {'cancomment': 0, 'canaddmeta': 0}, 'publiceditability': {'cancomment': 1, 'canaddmeta': 0}, 'usage': {'candownload': 1, 'canblog': 0, 'canprint': 0, 'canshare': 0}, 'comments': {'_content': '1'}, 'notes': {'note': [{'id': '72157721248375800', 'photo_id': '52673136543', 'author': '145425803@N04', 'authorname': 'javiborbon', 'authorrealname': 'Javi Borbón', 'authorispro': 1, 'authorisdeleted': 0, 'x': '261', 'y': '267', 'w': '37', 'h': '17', '_content': 'Sercué, pro_badge': 'standard'}]}, 'people': {'haspeople': 0}, 'tags': {'tag': [{'id': '145393664-52673136543-1905497', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'fanlo', '_content': 'fanlo', 'machine_tag': False}, {'id': '145393664-52673136543-3540', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'huesca', '_content': 'huesca', 'machine_tag': False}, {'id': '145393664-52673136543-18149', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'aragon', '_content': 'aragon', 'machine_tag': False}, {'id': '145393664-52673136543-2815', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'spain', '_content': 'spain', 'machine_tag': False}, {'id': '145393664-52673136543-178912', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'sobrarbe', '_content': 'sobrarbe', 'machine_tag': False}, {'id': '145393664-52673136543-69627', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'pyrenees', '_content': 'pyrenees', 'machine_tag': False}, {'id': '145393664-52673136543-6214', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'nieve', '_content': 'nieve', 'machine_tag': False}, {'id': '145393664-52673136543-412', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'snow', '_content': 'snow', 'machine_tag': False}, {'id': '145393664-52673136543-199154', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'cañon', '_content': 'cañon', 'machine_tag': False}, {'id': '145393664-52673136543-9812', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'canyon', '_content': 'canyon', 'machine_tag': False}, {'id': '145393664-52673136543-6817', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'pueblo', '_content': 'pueblo', 'machine_tag': False}, {'id': '145393664-52673136543-2267', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'town', '_content': 'town', 'machine_tag': False}, {'id': '145393664-52673136543-6061', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'mountaintown', '_content': 'mountaintown', 'machine_tag': False}, {'id': '145393664-52673136543-1886', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'atardecer', '_content': 'atardecer', 'machine_tag': False}, {'id': '145393664-52673136543-223', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'sunset', '_content': 'sunset', 'machine_tag': False}, {'id': '145393664-52673136543-14304', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'cielo', '_content': 'cielo', 'machine_tag': False}, {'id': '145393664-52673136543-282', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'sky', '_content': 'sky', 'machine_tag': False}, {'id': '145393664-52673136543-100895', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'sunsetsky', '_content': 'sunsetsky', 'machine_t
```

```
ag': False}, {'id': '145393664-52673136543-318165', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'cumbres', '_content': 'cumbres', 'machine_tag': False}, {'id': '145393664-52673136543-691349', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'summits', '_content': 'summits', 'machine_tag': False}, {'id': '145393664-52673136543-101595', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'montañas', '_content': 'montañas', 'machine_tag': False}, {'id': '145393664-52673136543-499', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'mountains', '_content': 'mountains', 'machine_tag': False}, {'id': '145393664-52673136543-204721', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'geologia', '_content': 'geologia', 'machine_tag': False}, {'id': '145393664-52673136543-25697', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'geology', '_content': 'geology', 'machine_tag': False}, {'id': '145393664-52673136543-361', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'naturaleza', '_content': 'naturaleza', 'machine_tag': False}, {'id': '145393664-52673136543-791', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'nature', '_content': 'nature', 'machine_tag': False}, {'id': '145393664-52673136543-21538', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'invierno', '_content': 'invierno', 'machine_tag': False}, {'id': '145393664-52673136543-201', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'winter', '_content': 'winter', 'machine_tag': False}, {'id': '145393664-52673136543-8948', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'paisaje', '_content': 'paisaje', 'machine_tag': False}, {'id': '145393664-52673136543-1695', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'landscape', '_content': 'landscape', 'machine_tag': False}}], 'urls': {'url': [{'type': 'photopage', '_content': 'https://www.flickr.com/photos/javiborbon/52673136543/'}]}, 'media': 'photo'}
```



```
In [26]: for key in photo_info:
          value = str(photo_info[key])
          print('\n' + key + '\t-->\t' + value)
```

id -->> 52673136543

secret -->> 990888fdfb

server -->> 65535

farm -->> 66

dateuploaded -->> 1675696479

isfavorite -->> 0

license -->> 0

safety\_level -->> 0

rotation -->> 0

originalsecret -->> 2930f4fe9d

originalformat -->> jpg

owner -->> {'nsid': '145425803@N04', 'username': 'javiborbon', 'realname': 'Javi Borbón', 'location': '', 'iconserver': '65535', 'iconfarm': 66, 'path\_alias': 'javiborbon', 'gift': {'gift\_eligible': True, 'eligible\_durations': ['year', 'month', 'week'], 'new\_flow': True}}

title -->> {'\_content': 'West side'}

description -->> {'\_content': 'Nerín, Huesca, Aragón, Spain'}

visibility -->> {'ispublic': 1, 'isfriend': 0, 'isfamily': 0}

dates -->> {'posted': '1675696479', 'taken': '2023-01-29 18:11:19', 'takengranularity': 0, 'takenunknown': '0', 'lastupdate': '1675700751'}

views -->> 61

editability -->> {'cancomment': 0, 'canaddmeta': 0}

publiceditability -->> {'cancomment': 1, 'canaddmeta': 0}

usage -->> {'candownload': 1, 'canblog': 0, 'canprint': 0, 'canshare': 0}

comments -->> {'\_content': '1'}

notes -->> {'note': [{'id': '72157721248375800', 'photo\_id': '52673136543', 'author': '145425803@N04', 'authorname': 'javiborbon', 'authorrealname': 'Javi Borbón', 'authorispro': 1, 'authorisdeleted': 0, 'x': '261', 'y': '267', 'w': '37', 'h': '17', '\_content': 'Sercué', 'pro\_badge': 'standard'}]}

people -->> {'haspeople': 0}

tags -->> {'tag': [{'id': '145393664-52673136543-1905497', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'fanlo', '\_content': 'fanlo', 'machine\_tag': False}, {'id': '145393664-52673136543-3540', 'author': '14

5425803@N04', 'authorname': 'javiborbon', 'raw': 'huesca', '\_content': 'huesca', 'machine\_tag': False}, {'id': '145393664-52673136543-18149', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'aragon', '\_content': 'aragon', 'machine\_tag': False}, {'id': '145393664-52673136543-2815', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'spain', '\_content': 'spain', 'machine\_tag': False}, {'id': '145393664-52673136543-178912', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'sobrarbe', '\_content': 'sobrarbe', 'machine\_tag': False}, {'id': '145393664-52673136543-69627', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'pyrenees', '\_content': 'pyrenees', 'machine\_tag': False}, {'id': '145393664-52673136543-6214', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'nieve', '\_content': 'nieve', 'machine\_tag': False}, {'id': '145393664-52673136543-412', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'snow', '\_content': 'snow', 'machine\_tag': False}, {'id': '145393664-52673136543-199154', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'cañon', '\_content': 'cañon', 'machine\_tag': False}, {'id': '145393664-52673136543-9812', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'canyon', '\_content': 'canyon', 'machine\_tag': False}, {'id': '145393664-52673136543-6817', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'pueblo', '\_content': 'pueblo', 'machine\_tag': False}, {'id': '145393664-52673136543-2267', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'town', '\_content': 'town', 'machine\_tag': False}, {'id': '145393664-52673136543-6061', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'mountaintown', '\_content': 'mountaintown', 'machine\_tag': False}, {'id': '145393664-52673136543-1886', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'atardec er', '\_content': 'atardec er', 'machine\_tag': False}, {'id': '145393664-52673136543-223', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'sunset', '\_content': 'sunset', 'machine\_tag': False}, {'id': '145393664-52673136543-14304', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'cielo', '\_content': 'cielo', 'machine\_tag': False}, {'id': '145393664-52673136543-282', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'sky', '\_content': 'sky', 'machine\_tag': False}, {'id': '145393664-52673136543-100895', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'sunset sky', '\_content': 'sunset sky', 'machine\_tag': False}, {'id': '145393664-52673136543-318165', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'cumbres', '\_content': 'cumbres', 'machine\_tag': False}, {'id': '145393664-52673136543-691349', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'summits', '\_content': 'summits', 'machine\_tag': False}, {'id': '145393664-52673136543-101595', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'montañas', '\_content': 'montañas', 'machine\_tag': False}, {'id': '145393664-52673136543-499', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'mountains', '\_content': 'mountains', 'machine\_tag': False}, {'id': '145393664-52673136543-204721', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'geologia', '\_content': 'geologia', 'machine\_tag': False}, {'id': '145393664-52673136543-25697', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'geology', '\_content': 'geology', 'machine\_tag': False}, {'id': '145393664-52673136543-361', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'naturaleza', '\_content': 'naturaleza', 'machine\_tag': False}, {'id': '145393664-52673136543-791', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'nature', '\_content': 'nature', 'machine\_tag': False}, {'id': '145393664-52673136543-21538', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'invierno', '\_content': 'invierno', 'machine\_tag': False}, {'id': '145393664-52673136543-201', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'winter', '\_content': 'winter', 'machine\_tag': False}, {'id': '145393664-52673136543-8948', 'author': '145425803@N04', 'authorname': 'javiborbon', 'raw': 'paisaje', '\_content': 'paisaje', 'machine\_tag': False}, {'id': '145393664-52673136543-1695', 'author': '145425803@

```
N04', 'authorname': 'javiborbon', 'raw': 'landscape', '_content': 'landscape', 'machine_tag': False}}]
```

```
urls --> {'url': [{'type': 'photopage', '_content': 'https://www.flickr.com/photos/javiborbon/52673136543/'}]}
```

```
media --> photo
```

### **TODO: take a look at the documentation**

- [flickr.photos.getRecent\(\)](https://www.flickr.com/services/api/flickr.photos.getRecent.html) (<https://www.flickr.com/services/api/flickr.photos.getRecent.html>)
- [flickr.stats.getPopularPhotos\(\)](https://www.flickr.com/services/api/flickr.stats.getPopularPhotos.html) (<https://www.flickr.com/services/api/flickr.stats.getPopularPhotos.html>)
- [flickr.photos.search\(\)](https://www.flickr.com/services/api/flickr.photos.search.html) (<https://www.flickr.com/services/api/flickr.photos.search.html>)

### **Exercise 0:**

Create a Flickr account and get API tokens to use the APIs.

### **Exercise 1:**

Get a list of Flickr's photos and crawl the related information.

### **Exercise 2:**

get the list of the photos with the most views on Flickr and crawl the related information.

In [ ]: