Analysis of unstructered data

Lecture 5 - Data collection

Janusz Szwabiński

Overview:

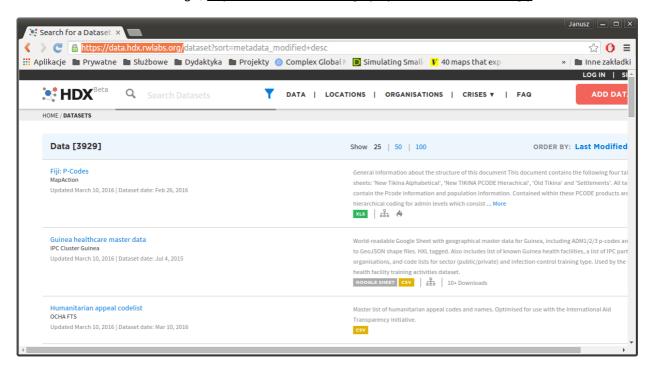
- · Where to look for data?
- Reading files in different formats
- · Screen scrapping
- · Using of existing APIs
- · Platforms for extracting data
- Online surveys

Where to look for data?

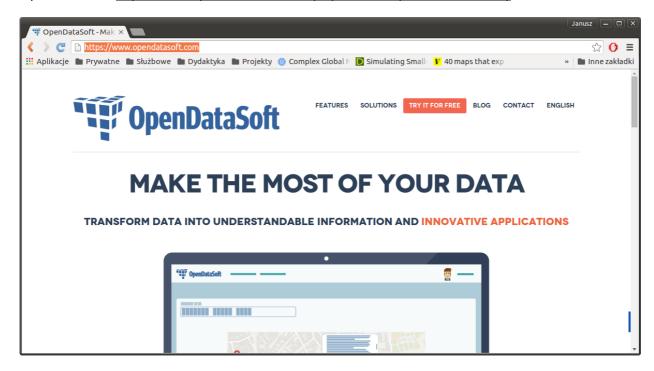
- in short everywhere
- in particular:
 - documents, maps, image files online
 - databases and warehouses
 - government portals
 - NGO portals
 - data scientists' webpages
 - other webpages
 - blogs

Worth a visit

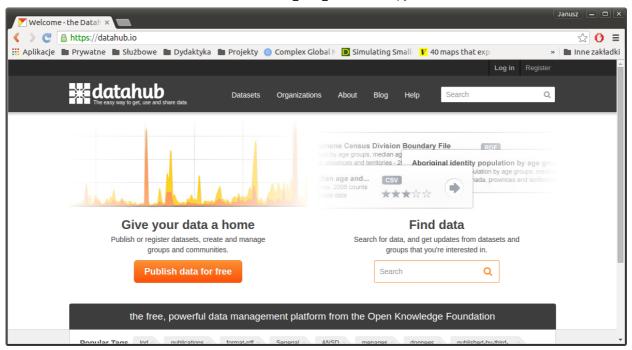
• Humanitarian Data Exchange, https://data.humdata.org/)



• OpenDataSoft, https://www.opendatasoft.com/ (https://www.opendatasoft.com/)



• Datahub.io, https://datahub.io/)



How to collect data?

- · file downloads
- · screen scrapping
- existing APIs
- custom methods (surveys, sensors, crowdsourcing)

From Wikipedia:

Crowdsourcing – a specific sourcing model in which individuals or organizations use contributions from Internet users to obtain needed services or ideas.

Reading files in different formats

If our file...

... is a CSV file

In [1]:

!head bikes.csv

- structured data
- some cleaning usually required
- easy to analyze
- Python modules available

In [2]:

```
import csv
with open('bikes.csv', 'r', encoding='latin1') as csvfile: #we know already that
  the file is in latin1 encoding
    reader = csv.reader(csvfile, delimiter=';')
    for row in reader:
        print(row)
```

```
['Date', 'Berri 1', 'Brébeuf (données non disponibles)', 'Côte-Saint
e-Catherine', 'Maisonneuve 1', 'Maisonneuve 2', 'du Parc', 'Pierre-D
upuy', 'Rachell', 'St-Urbain (données non disponibles)']
                '35',
                            '0', '38', '51', '26', '10', '16',
['01/01/2012',
                                 '68', '153', '53',
                            '1',
['02/01/2012',
                                                      '6',
                                                            '43'
                '83'
                             '2',
                                  '104',
                                          '248',
['03/01/2012',
                '135',
                                                  '89', '3', '58', '']
                                   '116',
                                          '318',
                             '1',
                                                  '111',
                                                          '8',
['04/01/2012'
                '144'
                                                                '61',
                                         '330',
                             '2',
                                   '124',
                                                   '97',
                                                         '13'
['05/01/2012',
                                                                '95'
                 '197'
                             '0',
                                   '98',
                                         '244',
                                                        '4',
                                                 '86',
                                                              75',
                '146',
                         1 1
['06/01/2012',
                            '2',
                                                '53',
                 '98',
                       1.1
                                  '80',
                                        '108',
                                                       '6',
                                                             '54',
['07/01/2012'
                       1 1
                                  '62'
                                        '98',
                                                            '63'
                            '1',
                                              '64',
                                                      '11'
                 '95'
['08/01/2012'
                             '2',
                                           '432',
['09/01/2012',
                 '244',
                                                   '198',
                                                                 '173',
                                   '165',
                                                           '12',
                             '3',
                                   '238',
                                          '563',
                                                          '18',
                                                                 '241',
['10/01/2012',
                '397'
                                                  '275',
                             '0',
                                                          '12',
                                   '182',
                                          '443',
                                                   '258',
                                                                 '194'
['11/01/2012'
                 '273'
                                                               '63',
                             '1'
                                   '134'
                                           '261', '137'
                                                          '9',
                 ' 157 '
['12/01/2012'
                                                ,
'64',
                                                            ´'0',
                            '0',
                                                      '2',
                                        '105',
['13/01/2012',
                                  '41',
                 '75'
                            '0',
                                                            '1',
                                        '56',
                                                      '0',
                 '32',
                                  '54',
                                               '19',
['14/01/2012',
                                                                 '']
                                                      '0',
                                                           '0',
                       1 1
                                  '33'
                                              , '18'
                 '54'
                            '0'
                                        '60'
['15/01/2012'
                                          '312',
                             '2',
                                                          '1',
['16/01/2012',
                 '168',
                                  '136',
                                                   '137',
                             '0',
                                  '86',
                                         '256',
                                                 '74',
                                                        '0',
['17/01/2012',
                ' 155 '
                             '0',
                                          188 , '248', ', '67',
                                  '66', '188', '68',
                                                        '3'
                '139'
['18/01/2012'
                                   '104',
                                                         '3'
                 '191'
                             '1'
['19/01/2012'
                                   '96',
                             '4',
                                                        '1',
['20/01/2012',
                                         '217',
                 '161'
                            '0',
                                  '47',
                                                      '1',
                 '53',
                                        '70',
                                               '32',
['21/01/2012',
                       1 1
                 '71'
                            '0'
                                  '41'.
                                        '73'
                                               '35'
                                                      '5'
['22/01/2012'
                                          '357',
                             '6',
                                   '114',
                                                   '91'
                 '210'
['23/01/2012'
                                   '189',
                                           '444',
                             '1',
['24/01/2012',
                 '299'
                                                   '174'
                             '1',
                                   '217',
                                           '453',
                                                  '180',
                                                          '4',
['25/01/2012',
                 '334'
                             '0',
                                         ' 495 '
                                                  '191'
                                                          '0'
['26/01/2012'
                 '306'
                                   '215',
                            '5',
                                  '79',
['27/01/2012',
                                        '204',
                 '91'
                                                 '65', '0',
                            '1',
                                        '123',
                                                       '9',
                       1 1
                                                '33',
                 '80',
                                  '61',
['28/01/2012',
                                                             '1'1
                       1 1
                 '87'
                            '1',
                                  '65',
                                        '132'
                                                '40'
                                                       '7'
['29/01/2012'
                             '0',
                                                   '152',
                                   '146',
                                          '371',
['30/01/2012'
                 '219'
                                                          '0',
                             '1',
                                   '109',
                                           '324',
                                                   '122',
['31/01/2012',
                 '186'
                                          '271',
                                                         '5'
['01/02/2012',
                             '0',
                                  '100',
                '138'
                                                   '71',
                             '5',
                                  '134',
                                          '345',
                                                          '2',
                 '217'
                                                  '128',
['02/02/2012'
                                                           '1'
                             '1'
                                   '103', '301', '111',
                 '174'
['03/02/2012'
                            '0',
                                                '44',
 '04/02/2012',
                                 '53',
                                        '119',
                 '84'
                                                       '8',
                       1 1
                            '0',
                                                '54'
['05/02/2012'
                 '72'
                                  '46'
                                        '133'
                                                       '7'
                             '1',
                                   '136',
                                           '425',
                                                   '167'
                 '248'
                                                           '10',
['06/02/2012'
['07/02/2012',
                             '0',
                                   '209',
                                           '516',
                                                           '9',
                 '316'
                                                   '225'
                             '0',
                                   '202',
                                           '503',
['08/02/2012',
                 '271'
                                                   '215',
                                                           '4',
                                                                '1']
                 '342',
                             '0',
                                  '227',
                                          '471',
                                                  '231',
['09/02/2012',
                                                          '11',
                             '1'
                                   '206',
                                                  '216'
                                                           '6', '2'1
                 '303'
                                          '478',
['10/02/2012'
                            '0',
                                                ,
'49'
                                        '112',
                                 '63',
['11/02/2012'
                 '71'
                        1 1
                            '0',
                                  '36',
                                               '53',
                                        '91',
['12/02/2012'
                 '78',
                             '0',
                                           '408',
                                                           '4',
['13/02/2012'
                                   '175',
                 '211'
                                                   '207'
                             '0',
                                   '186',
                                           '504',
                                                   '243'
                                                           '9'
                 '318'
                                                                '1']
['14/02/2012'
                             '0',
                                           '491',
                                   '180',
['15/02/2012',
                 '307'
                                                   '232'
                                                           '5'
                             '1',
                                           '569',
                                   '212',
                                                   '295',
                                                           '10',
['16/02/2012'
                 '386'
                         1.1
                             '3'
                                   '237'
                 '332'
                                           '496'
                                                   '260'
                                                           '11'
                                                                 '1'1
['17/02/2012'
                             '1'
                                   '159'
                                           '280',
                 220
                                                   '134'
                                                           '30'
 '18/02/2012'
                             '3',
                         1 1
                                   '110',
                                           '205',
['19/02/2012',
                                                   '113',
                                                           '13'
                 '169'
                             '2',
                                   '224',
                        1.1
                                           '470',
                                                   '243',
                                                           '14',
['20/02/2012'
                 '303'
                         1.1
                             '5'
                                   '292'
                                           '503'
                                                   '286'
                                                           '10'
                 '441'
['21/02/2012'
                                                           12',
                             '10',
                                            '501',
                 '375'
                                    '263',
                                                    '264',
['22/02/2012'
                             '10',
                                           '528',
                                                    '329',
                 '397'
                                    '293',
                                                            '22',
['23/02/2012'
                             '4',
                         1.1
                                  '187', '313', '137',
                                                          '4'.
['24/02/2012'
                 '243'
                                                                '13']
                            '0',
                                 '48',
                                        '52', '18', '2'
                 '62'
  25/02/2012
                                '76',
                       1 1
                           '0',
                                       '47'
['26/02/2012',
                 '78'
                                              '35',
                                                     '4'
                '119', '', '4', '94', '298', '134', '3', '5']
['27/02/2012',
```

```
'0',
               '195',
                                        '350',
                                 '158',
                                                '168', '9',
                       1.1
['28/02/2012',
               '242',
                           '0',
                       1.1
                                        '446', '190', '7', '4'Î
['29/02/2012',
                                 '164',
['01/03/2012',
               '92', ''
                          '0', '56',
                                      '199', '63', '0'
                                                       , '14']
                                              '77',
                                      '283',
['02/03/2012',
                '143',
                           '0', '56',
                                                    '7'
                          '0',
                      1.1
                                '65',
                                      '133',
                                             '47',
['03/03/2012',
                '82',
                '107',
                       1.1
                                '79',
                                       '138', '60', '15',
                           '2',
['04/03/2012'
                                        '363',
['05/03/2012',
                                                '157', '8'
                '155'
                            '1',
                                 '114',
                                        '437',
                            '0',
                                 '192',
                                                       '10',
                                                '170',
['06/03/2012',
                '269'
                           '3',
                                 '290',
                                        '715',
['07/03/2012',
                '438',
                                                '266',
                                                       '28',
                       1.1
                           '6',
                                        '530',
                                                '270',
['08/03/2012'
                '348'
                                 '238',
                                                       '9', '9']
['09/03/2012',
                                 '279',
                            '8'
                                                '268',
                '371'
                                         '575',
                                                       '8'
                                                 <sup>'</sup>165',
                                                        '0',
                '182',
                       1 1
                           '13', '162', '296',
['10/03/2012',
                       1.1
['11/03/2012',
                           '60',
                                  '253',
                                         '540',
                                                        '0',
                '380'
                                                 '289'.
['12/03/2012',
                       1.1
                            '179', '618',
                                          '1265', '747', '0', '548']
                '802'
                                   '321',
['13/03/2012',
                                          '769', '425', '0', '325']
'739', '451', '0', '336']
                       1.1
                '442'
                            '145',
                                   '313',
['14/03/2012',
                '469',
                       1.1
                           '146',
                       1.1
                                          '1021', '631', '63', '565']
                '724'
                           '244',
                                   '562',
['15/03/2012'
                                          '695', '41.
'695', '468',
'1(
                                   '305',
                                                 '419', '9', 422
'468', '334', '1008']
['16/03/2012',
                            '149',
                '423'
                       1 1
                                  '422',
['17/03/2012',
                            '287',
                '681'
['18/03/2012', '1940', '', '856', '1036', '1923', '1021', '1128', '2
477']
['19/03/2012', '1821', '', '1024', '1278', '2581', '1609', '506', '2
058'1
['20/03/2012', '2481', '', '1261', '1709', '3130', '1955', '762', '2
609']
['21/03/2012', '2829', '', '1558', '1893', '3510', '2225', '993', '2
846']
['22/03/2012', '2195', '', '1030', '1640', '2654', '1958', '548', '2
['23/03/2012', '2115', '', '1143', '1512', '2955', '1791', '663', '2
['24/03/2012', '753', '', '336', '517', '1001', '635', '277', '103
5'1
['25/03/2012', '520', '', '243', '309', '691', '427', '145', '723']
['26/03/2012', '968', '', '564', '729', '1493', '965', '130', '116
8']
['27/03/2012', '1049', '', '517', '774', '1576', '972', '163', '120
7']
['28/03/2012', '435', '', '179', '329', '709', '486', '28', '529']
['29/03/2012', '878', ''
                          '406', '646', '1264', '807', '78', '937']
['30/03/2012', '1157', '', '529', '910', '1596', '957', '196', '128
8'1
['31/03/2012', '980', '', '499', '587', '1083', '706', '524', '137
0']
['01/04/2012', '662', ''
                           '341', '442', '824', '471', '168', '1086']
['02/04/2012', '1937', '', '967', '1537', '2853', '1614', '394', '21
22']
['03/04/2012', '2416', '', '1078', '1791', '3556', '1880', '513', '2
450'1
['04/04/2012', '2211', '', '933', '1674', '2956', '1666', '274', '22
['05/04/2012', '2424', '', '1036', '1823', '3273', '1699', '355', '2
463 ' 1
['06/04/2012', '1633', '', '650', '1045', '1913', '975', '621', '213
8'1
['07/04/2012', '1208', '', '494', '739', '1445', '709', '598', '156
6'1
['08/04/2012', '1164', '', '560', '621', '1333', '704', '792', '153
3'1
['09/04/2012', '828', ''
                           '298', '560', '1048', '605', '65', '1001']
['10/04/2012', '2183', '', '909', '1588', '2932', '1736', '252', '21
```

```
08'1
['11/04/2012', '2328', '', '1049', '1765', '3122', '1843', '330', '2
311']
['12/04/2012', '3064', '', '1483', '2306', '4076', '2280', '590', '3
213']
['13/04/2012', '3341', '', '1505', '2565', '4465', '2358', '922', '3
728']
['14/04/2012', '2890', '', '1072', '1639', '2994', '1594', '1284',
 '3428']
['15/04/2012', '2554', '', '1210', '1637', '2954', '1559', '1846',
 '3604'1
['16/04/2012', '3643', '', '1841', '2723', '4830', '2677', '1061',
 '3616'1
['17/04/2012', '3539', '', '1616', '2636', '4592', '2450', '544', '3
333']
['18/04/2012', '3570', '', '1751', '2759', '4655', '2534', '706', '3
542'1
['19/04/2012', '4231', '', '2010', '3235', '5311', '2877', '1206',
 3929 ' 1
['20/04/2012', '2087', '', '800', '1529', '2922', '1531', '170', '20
65']
['21/04/2012', '533', '', '212', '398', '710', '408', '50', '755']
['22/04/2012', '1853', '', '487', '1224', '1331', '654', '198', '177
['23/04/2012', '623', '', '315', '544', '1076', '612', '27', '846']
['24/04/2012', '1810', '', '720', '1355', '2379', '1286', '188', '17
53']
['25/04/2012', '2966', '', '1023', '2228', '3444', '1800', '445', '2
454'1
['26/04/2012', '2751', '', '1069', '2196', '3546', '1789', '381', '2
438 ' 1
['27/04/2012', '1986', '', '743', '1526', '2586', '1298', '139', '18
99'1
['28/04/2012', '1684', '', '628', '1190', '1908', '931', '523', '232
3'1
['29/04/2012', '1970', '', '765', '1212', '2077', '1062', '702', '24
93']
['30/04/2012', '3610', '', '1572', '2825', '4675', '2446', '851', '3
541'1
['01/05/2012', '1986', '', '815', '1722', '2766', '1516', '195', '19
60'1
['02/05/2012', '3724', '', '1677', '2885', '4731', '2508', '876', '3
501']
['03/05/2012', '3698', '', '1618', '3001', '4943', '2577', '731', '3
603'1
['04/05/2012', '2511', '', '1163', '2058', '3717', '1823', '380', '2
631'1
['05/05/2012', '3492', '', '1366', '2106', '3696', '1779', '1677',
 '4108'1
['06/05/2012', '3411', '', '1525', '1815', '3346', '1879', '2036',
 '4633']
['07/05/2012', '5552', '', '2573', '3959', '6355', '3416', '1848',
 '5253'1
['08/05/2012', '1241', '', '625', '991', '1729', '1007', '119', '138
3'1
['09/05/2012', '3297', '', '1545', '2700', '4343', '2340', '737', '3
['10/05/2012', '2755', '', '1227', '2130', '4056', '2075', '399', '2
437']
['11/05/2012', '4639', '', '1803', '3663', '5713', '2888', '1260',
 '4499'1
```

```
['12/05/2012', '3854', '', '1457', '2429', '3894', '1805', '2268',
 '4855'1
['13/05/2012', '2741', '', '1223', '1703', '3086', '1592', '1394',
 '3496'1
['14/05/2012', '6189', '', '2709', '4402', '7006', '3868', '2215',
 '5775'1
['15/05/2012', '3964', '', '1773', '3144', '5088', '2650', '975', '3
765']
['16/05/2012', '4947', '', '2178', '3681', '5882', '3057', '1332',
 '4348'1
['17/05/2012', '5351', '', '2441', '4182', '6551', '3408', '1631',
 '4988'1
['18/05/2012', '5980', '', '2241', '4415', '6646', '3196', '1711',
 '5273']
['19/05/2012', '4732', '', '1454', '2807', '4673', '1966', '2914',
 '5293'1
['20/05/2012', '5255', '', '1663', '2730', '4462', '2182', '4241',
 '5539']
['21/05/2012', '5129', '', '1646', '2672', '4169', '2044', '3413',
 '5053'1
['22/05/2012', '2315', '', '938', '1847', '2599', '1610', '251', '20
['23/05/2012', '5974', '', '2650', '4407', '7281', '3737', '1826',
 '4798'1
['24/05/2012', '6485', '', '2653', '4600', '7600', '3792', '2062',
 5209 ' 1
['25/05/2012', '5697', '', '2205', '4096', '6734', '3341', '1953',
 '5174'1
['26/05/2012', '4974', '', '1622', '2936', '4991', '2373', '3455',
 '5443']
['27/05/2012', '4396', '', '1525', '2578', '4587', '2073', '2886',
 '5168']
['28/05/2012', '4268', '', '1962', '3449', '5798', '2898', '1027',
 '3894'1
['29/05/2012', '3154', '', '1184', '2325', '3904', '1933', '475', '2
731']
['30/05/2012', '6459', '', '2722', '4806', '7632', '3817', '2454',
 '5172']
['31/05/2012', '5104', '', '2177', '3985', '6631', '3205', '1389',
 '4410'1
['01/06/2012', '6097', '', '2604', '4110', '7175', '3895', '1692',
 '6595'1
['02/06/2012', '943', '', '392', '630', '1289', '628', '71', '1436']
['03/06/2012', '2755', '', '1897', '2020', '3768', '2324',
 '4936'1
['04/06/2012', '2717', '', '1408', '2095', '4276', '2168', '333', '3
090'1
['05/06/2012', '5842', '', '2721', '3927', '7302', '3786', '1232',
 '5348'1
['06/06/2012', '6037', '', '2724', '4273', '7822', '3987', '1223',
 '5269'1
['07/06/2012', '6246', '', '2607', '4670', '8222', '3972', '1019',
 '5724'1
['08/06/2012', '4169', '', '1833', '3303', '5881', '3001', '623', '4
474']
['09/06/2012', '5164', '', '1672', '3418', '5557', '2451', '1551',
 '6026'1
['10/06/2012', '5112', '', '1812', '2905', '4798', '2328', '1833',
 '6170'1
['11/06/2012', '6206', '', '2577', '4333', '7015', '3757', '1264',
 '5721']
```

```
['12/06/2012', '3361', '', '1556', '2362', '4230', '2381', '589', '3
['13/06/2012', '6180', '', '2792', '4374', '7297', '3825', '1968',
 '5887'1
['14/06/2012', '6908', '', '2978', '4809', '7934', '4223', '2386',
 '6243'1
['15/06/2012', '7077', '', '2469', '4999', '7663', '4053', '2293',
 '6491']
['16/06/2012', '5421', '', '1651', '3099', '4909', '2337', '3280',
 '5537'1
['17/06/2012', '4638', '', '1552', '2082', '4108', '2020', '3453',
 '5128'1
['18/06/2012', '5921', '', '2703', '3582', '6824', '3960', '2455',
 '5496']
['19/06/2012', '5382', '', '2360', '3447', '6679', '3438', '1711',
 '5015'1
['20/06/2012', '5713', '', '2402', '3524', '6848', '3510', '2363',
 '4852']
['21/06/2012', '5183', '', '2195', '3297', '6684', '3272', '2265',
 '4957'1
['22/06/2012', '5398', '', '1934', '3612', '6659', '3252', '2009',
 '5263'1
['23/06/2012', '3753', '', '1326', '1842', '3992', '1910', '2585',
 '4824'1
['24/06/2012', '3341', '', '1152', '1423', '2989', '1510', '3634',
 '4581'1
['25/06/2012', '2245', '', '1077', '1374', '3031', '1440', '1349',
 '3283'1
['26/06/2012', '3327', '', '1435', '2139', '4318', '2193', '620', '3
223']
['27/06/2012', '3141', '', '1322', '2142', '4211', '2134', '536', '2
897 ' 1
['28/06/2012', '6064', '', '2381', '4411', '6864', '3523', '2415',
 '5175'1
['29/06/2012', '5770', '', '1935', '4004', '6656', '3090', '2109',
 '5332']
['30/06/2012', '4738', '', '1359', '3007', '4226', '1822', '2870',
 '4527']
['01/07/2012', '4758', '', '1343', '2911', '3935', '1777', '3732',
 '4522'1
['02/07/2012', '4144', '', '1397', '2658', '3998', '1883', '2783',
 '4464'1
['03/07/2012', '6712', '', '2634', '4398', '7416', '3896', '2606',
 '5462']
['04/07/2012', '5153', '', '2038', '3591', '5814', '3070', '1246',
 '4493'1
['05/07/2012', '6672', '', '2603', '4830', '7764', '3816', '2746',
 '5153'l
['06/07/2012', '5958', '', '2095', '4137', '6942', '3387', '2248',
 '5610'1
['07/07/2012', '5420', '', '1289', '3802', '4978', '2132', '3156',
 '4939'1
['08/07/2012', '4756', '', '1497', '2407', '4495', '2089', '4386',
 '5535']
['09/07/2012', '5661', '', '2330', '3524', '6769', '3577', '2599',
 '4987'1
['10/07/2012', '6500', '', '2625', '4064', '7436', '3749', '2822',
 '5952']
['11/07/2012', '6424', '', '2548', '3921', '7374', '3781', '2779',
 '5474']
['12/07/2012', '6179', '', '2371', '3772', '7200', '3687', '2433'.
```

```
'5140']
['13/07/2012', '5518', '', '2005', '3525', '6492', '3088', '2219',
 '5185']
['14/07/2012', '4206', '', '1244', '2441', '4284', '1562', '2900',
 '4252'1
['15/07/2012', '3035', '', '1060', '1798', '3437', '1449', '2034',
 '3545']
['16/07/2012', '4827', '', '2092', '3141', '6049', '2965', '1703',
 '4379']
['17/07/2012', '2756', '', '1046', '2001', '3802', '1700', '569', '2
641'1
['18/07/2012', '6180', '', '2589', '3750', '7311', '3654', '2813',
 '5884'1
['19/07/2012', '6309', '', '2397', '3996', '7699', '3777', '2647',
 '5665']
['20/07/2012', '5813', '', '2088', '3744', '6959', '3316', '2538',
 '5775'1
['21/07/2012', '5092', '', '1419', '2957', '4658', '1821', '3835',
 '4848'1
['22/07/2012', '4971', '', '1238', '2683', '4104', '1788', '2960',
 '4787']
['23/07/2012', '3877', '', '1660', '2420', '4937', '2534', '1461',
 '3475'1
['24/07/2012', '5243', '', '2318', '3465', '6721', '3317', '2170',
 '4694']
['25/07/2012', '6104', '', '2593', '3786', '7073', '3622', '3132',
 '5059']
['26/07/2012', '5560', '', '2043', '4174', '6476', '3208', '1595',
 '4383'1
['27/07/2012', '5759', '', '1948', '3658', '6851', '3275', '2653',
 '5295']
['28/07/2012', '4105', '', '1345', '2504', '4540', '1767', '2468',
 '4631']
['29/07/2012', '4230', '', '1332', '1791', '4235', '1905', '3731',
 '4906'1
['30/07/2012', '5225', '', '2285', '2975', '6219', '3196', '2564',
 5100'1
['31/07/2012', '5415', '', '2185', '3145', '6705', '3222', '2505',
 '4468']
['01/08/2012', '4638', '', '1842', '2783', '5846', '2810', '1481',
 '3889'1
['02/08/2012', '5715', '', '2361', '3340', '6956', '3329', '2643',
 '4828']
['03/08/2012', '5577', '', '1915', '3306', '6427', '2789', '3195',
 '5031']
['04/08/2012', '4223', '', '1305', '2019', '4425', '1769', '2939',
 '4467']
['05/08/2012', '1864', '', '630', '1100', '2327', '970', '1066', '22
68'1
['06/08/2012', '5240', '', '2277', '3193', '6375', '3187', '2274',
 '4604']
['07/08/2012', '5997', '', '2410', '3622', '6840', '3480', '2780',
 '4903']
['08/08/2012', '5498', '', '2313', '3365', '6862', '3403', '2226',
 '4771']
['09/08/2012', '4127', '', '1802', '2601', '5377', '2656', '1471',
['10/08/2012', '2414', '', '934', '1727', '3323', '1641', '440', '24
56'1
['11/08/2012', '2453', '', '815', '1589', '2628', '1043', '1340', '2
608'1
```

```
['12/08/2012', '3995', '', '1338', '1937', '3427', '1719', '2988',
 '4178']
['13/08/2012', '4931', '', '2298', '3124', '5936', '3151', '2090',
 '4842'1
['14/08/2012', '5333', '', '2322', '3571', '6332', '3260', '1456',
 '4362'1
['15/08/2012', '4297', '', '1882', '2936', '4948', '2686', '970', '3
877']
['16/08/2012', '6062', '', '2538', '3898', '6549', '3366', '1549',
 '4943'1
['17/08/2012', '4236', '', '1777', '3099', '4979', '2581', '586', '4
154'1
['18/08/2012', '4427', '', '1333', '2505', '3866', '1864', '1229',
 '4359']
['19/08/2012', '4644', '', '1408', '2605', '3780', '1758', '1683',
 '4478'1
['20/08/2012', '5335', '', '2474', '3521', '6088', '3332', '1439',
 '5201'1
['21/08/2012', '5792', '', '2508', '4063', '6687', '3369', '2353',
 '5640'1
['22/08/2012', '6529', '', '2671', '4513', '7065', '3774', '2391',
 '5154'l
['23/08/2012', '5437', '', '2482', '3807', '6773', '3573', '1864',
 '5010'1
['24/08/2012', '5690', '', '2394', '3778', '6703', '3312', '2178',
 5299 ' 1
['25/08/2012', '4242', '', '1305', '2358', '4126', '1726', '2558',
 '4833'1
['26/08/2012', '3964', '', '1318', '2118', '3558', '1750', '2932',
 '4536'1
['27/08/2012', '4314', '', '2273', '3098', '5196', '2762', '1010',
 '3598']
['28/08/2012', '5817', '', '2794', '3966', '6313', '3420', '1629',
 '4720'1
['29/08/2012', '6327', '', '3092', '4370', '7009', '3759', '2152',
 '5342']
['30/08/2012', '6003', '', '2910', '4199', '6781', '3481', '1796',
 '5202']
['31/08/2012', '4106', '', '1868', '2999', '5054', '2541', '875', '4
188'1
['01/09/2012', '3934', '', '1225', '2371', '3917', '1887', '2801',
 '3845'1
['02/09/2012', '3698', '', '1232', '2093', '3827', '1788', '2959',
 '4247']
['03/09/2012', '4179', '', '1532', '2233', '3718', '1620', '3783',
 '4709'1
['04/09/2012', '4225', '', '1934', '3057', '5275', '3006', '894', '3
868 ' 1
['05/09/2012', '5655', '', '2722', '3861', '7110', '4061', '1383',
 '4491']
['06/09/2012', '5883', '', '2883', '3967', '7511', '4014', '1498',
 '4732']
['07/09/2012', '6186', '', '2720', '4153', '7323', '3886', '2051',
 '6104'1
['08/09/2012', '2155', '', '1192', '1467', '2739', '898', '801', '26
26'1
['09/09/2012', '3312', '', '1810', '1828', '3458', '2668', '2118',
 '4355'1
['10/09/2012', '5077', '', '2792', '3718', '6644', '3849', '1271',
 '4984']
['11/09/2012', '6015', '', '2913', '4081', '7148', '4314', '1484',
```

```
'5451']
['12/09/2012', '6349', '', '3124', '4209', '7292', '4510', '1981',
 '5697'1
['13/09/2012', '6520', '', '3076', '4369', '7514', '4494', '1986',
 '5742']
['14/09/2012', '5216', '', '2257', '3480', '6104', '3574', '1524',
 '4918']
['15/09/2012', '3341', '', '1100', '1900', '3099', '1593', '1268',
 '3426']
['16/09/2012', '3635', '', '1231', '1614', '3040', '1852', '2348',
 '3696'1
['17/09/2012', '5299', '', '2658', '3670', '6582', '4076', '1704',
 '4788'1
['18/09/2012', '2530', '', '1277', '1695', '3254', '2275', '388', '2
377']
['19/09/2012', '4653', '', '2351', '3176', '5746', '3738', '840', '4
189'1
['20/09/2012', '5260', '', '2496', '3615', '6348', '3991', '1391',
 5119'1
['21/09/2012', '4022', '', '1821', '2911', '4745', '2945', '889', '3
886']
['22/09/2012', '1521', '', '656', '1073', '1682', '971', '365', '197
['23/09/2012', '2314', '', '1093', '1496', '2572', '1470', '984', '3
321'1
['24/09/2012', '4553', '', '2246', '3117', '5738', '3705', '1100',
 '4683']
['25/09/2012', '5038', '', '2236', '3273', '5861', '3712', '1007',
 '4420'1
['26/09/2012', '3948', '', '1873', '2720', '4755', '3119', '650', '3
721'1
['27/09/2012', '5119', '', '2288', '3465', '5983', '3704', '1152',
 '4598']
['28/09/2012', '4652', '', '2134', '3209', '5372', '3309', '1133',
['29/09/2012', '1896', '', '755', '1206', '1963', '1062', '442', '22
67']
['30/09/2012', '876', ''
                         '359', '513', '957', '520', '133', '1162']
['01/10/2012', '3255', '', '1576', '2184', '3830', '2582', '397', '2
987 ' 1
['02/10/2012', '5139', '', '2525', '3281', '5845', '3732', '1229',
 '4497'1
['03/10/2012', '4685', '', '2377', '3191', '5475', '3415', '858', '3
941'1
['04/10/2012', '4034', '', '2025', '2705', '4850', '3066', '555', '3
418'1
['05/10/2012', '4151', '', '1977', '2799', '4688', '2844', '1035',
 '4088'1
['06/10/2012', '1304', '', '469', '933', '1589', '776', '236', '177
5'1
['07/10/2012', '1580', '', '660', '922', '1629', '860', '695', '205
2'1
['08/10/2012', '1854', '', '880', '987', '1818', '1040', '1115', '25
02'1
['09/10/2012', '4787', '', '2210', '3026', '5138', '3418', '927', '4
078']
['10/10/2012', '3115', '', '1537', '2081', '3681', '2608', '560', '2
703'1
['11/10/2012', '3746', '', '1857', '2569', '4694', '3034', '558', '3
457 ' ]
['12/10/2012', '3169', '', '1460', '2261', '4045', '2564', '448', '3
```

```
224'1
['13/10/2012', '1783', '', '802', '1205', '2113', '1183', '681', '23
09'1
['14/10/2012', '587', '', '287', '443', '852', '503', '65', '952']
['15/10/2012', '3292', '', '1678', '2165', '4197', '2754', '560', '3
183'1
['16/10/2012', '3739', '', '1858', '2684', '4681', '2997', '554', '3
593']
['17/10/2012', '4098', '', '1964', '2645', '4836', '3063', '728', '3
834 1
['18/10/2012', '4671', '', '2292', '3129', '5542', '3477', '1108',
 '4245'1
['19/10/2012', '1313', '', '597', '885', '1668', '1209', '111', '148
6']
['20/10/2012', '2011', '', '748', '1323', '2266', '1213', '797', '22
43'1
['21/10/2012', '1277', '', '609', '869', '1777', '898', '242', '164
8'1
['22/10/2012', '3650', '', '1819', '2495', '4800', '3023', '757', '3
721'1
['23/10/2012', '4177', '', '1997', '2795', '5216', '3233', '795', '3
554 ' 1
['24/10/2012', '3744', '', '1868', '2625', '4900', '3035', '649', '3
622 ' 1
['25/10/2012', '3735', '', '1815', '2528', '5010', '3017', '631', '3
767 ' 1
['26/10/2012', '4290', '', '1987', '2754', '5246', '3000', '1456',
 '4578'1
['27/10/2012', '1857', '', '792', '1244', '2461', '1193', '618', '24
71'1
['28/10/2012', '1310', '', '697', '910', '1776', '955', '387', '187
6'1
['29/10/2012', '2919', '', '1458', '2071', '3768', '2440', '411', '2
795 ' 1
['30/10/2012', '2887', '', '1251', '2007', '3516', '2255', '338', '2
790']
['31/10/2012', '2634', '', '1294', '1835', '3453', '2220', '245', '2
570']
['01/11/2012', '2405', '', '1208', '1701', '3082', '2076', '165', '2
461'1
['02/11/2012', '1582', '', '737', '1109', '2277', '1392', '97', '188
8'1
['03/11/2012', '844', '', '380', '612', '1137', '713', '105', '130
2'1
['04/11/2012', '966', '', '446', '710', '1277', '692', '197', '137
['05/11/2012', '2247', '', '1170', '1705', '3221', '2143', '179', '2
430'1
```

In [3]:

```
import pandas as pd
bikes = pd.read_csv('bikes.csv', sep=';', encoding='latin1', parse_dates=
['Date'], dayfirst=True, index_col='Date')
bikes[:3]
```

Out[3]:

	Berri 1	Brébeuf (données non disponibles)	Côte- Sainte- Catherine	Maisonneuve 1	Maisonneuve 2	du Parc	Pierre- Dupuy	Rach
Date								
2012- 01-01	35	NaN	0	38	51	26	10	16
2012- 01-02	83	NaN	1	68	153	53	6	43
2012- 01-03	135	NaN	2	104	248	89	3	58

... contains a lot of curly brackets, {}

Most probably it is a file in JSON format:

In [4]:

- Python modules available
- · easy to transform to other formats

In [5]:

In [6]:

```
df = pd.read_json('employees.json')
df
```

Out[6]:

	employees
0	{'firstName': 'John', 'lastName': 'Doe'}
1	{'firstName': 'Anna', 'lastName': 'Smith'}
2	{'firstName': 'Peter', 'lastName': 'Jones'}

In [7]:

```
from pandas.io.json import json_normalize

with open('employees.json') as data_file:
    data = json.load(data_file)
json_normalize(data['employees'])
```

Out[7]:

	firstName	lastName
0	John	Doe
1	Anna	Smith
2	Peter	Jones

... contains a lot of angle brackets, <>

Most probably it is a file in XML/HTML format:

In [8]:

```
!head kursy.xml
```

• again, there are already Python modules available

In [9]:

```
import xmltodict #not in standard library
with open('kursy.xml',encoding='latin2') as fd:
    doc = xmltodict.parse(fd.read())
```

```
In [10]:
```

```
for k in doc.keys():
    print(k)
```

tabela_kursow

In [11]:

```
for item in doc['tabela kursow']:
   print(item)
```

@typ numer_tabeli data publikacji pozycja

In [12]:

```
from xml.dom import minidom
xmldoc = minidom.parse('kursy.xml')
itemlist = xmldoc.getElementsByTagName('pozycja')
print(len(itemlist))
for s in itemlist:
    kraj = s.childNodes[1]
    print(kraj.childNodes[0].data)
```

26 Australia Austria Belgia Czechy Dania Estonia Finlandia Francia Grecja Hiszpania

Holandia

Irlandia Japonia

Kanada

Luksemburg

Norwegia

Portugalia

RFN **USA**

Szwajcaria

Szwecja Węgry

W. Brytania

Włochy UGW

MFW (SDR)

... is a PDF file

Unit Count and Square Footage

First Quarter, Fiscal Year 2014 As of April 30, 2013

Format	New	Closed	Relocation/ Expansion/ Conversion	Ending Square Footage	Total Locations
Walmart Discount Stores	1		(8)	58,277,925	554
Walmart Supercenters	16	-	8	574,058,937	3,182
Neighborhood Markets	17	-	-	11,700,094	284
Neighborhood Markets	17	-	-	10,948,346	258
Amigo	-			672,148	24
Supermercado	_	-	_	79,600	2
Small Formats	4			294,308	23

- pdftables (commercial)
- Tabula (https://github.com/tabulapdf/tabula (https://github.com/tabulapdf/tabula))
- pdftotext
- pdf-table-extract (https://github.com/ashima/pdf-table-extract (https://github
- pdfplumber (https://github.com/jsvine/pdfplumber (https://github.com/jsvine/pdfplumber))
- in case the above programs do not work:
 - OCR (e.g. https://github.com/tesseract-ocr (https://github.com/tesseract-ocr))
 - manual rewriting

In [13]:

```
import pandas as pd
import pdfplumber

with pdfplumber.open("data/example.pdf") as pdf:
    page = pdf.pages[0]
    table = page.extract_table()

df = pd.DataFrame(table[1:], columns=table[0])
df
```

Out[13]:

	Format	New	Closed	Relocation/ Expansion/ Conversion*	Ending Square Footage	Total Locations	
0	Walmart Discount Stores	1		(8)	58,277,925	5 54	
1	Walmart Supercenters	16	ı	8	574,058,937	3 ,182	
2	Neighborhood Markets	17	ı	-	11,700,094	2 84	
3	3 Neighborhood Markets		-	-	10,948,346	2 58	
4	Amigo	-	-	-	672,148	2 4	
5	Supermercado	-	-	-	79,600	2	
6	Small Formats	4	-	-	294,308	2 3	
7	7 Marketside		-	-	-	-	
8	Super Ahorros	-	-	-	46,349	3	
9	Walmart Express	4	-	-	239,702	17	
10	Walmart on Campus	-	-	-	8,257	3	
11	Walmart U.S.	38	-		644,331,264	4 ,043	
12	Sam's Club		-	-	82,669,348	6 20	
13	Total U.S.	38	-	-	727,000,612	4 ,663	
14	International	50	(3)	(1)	349,900,213	6 ,194	
15	Total Walmart	88	(3)	(1)	1,076,900,825	1 0,857	

In [14]:

```
for column in df.columns:
    df[column] = df[column].str.replace(" ", "")
df
```

Out[14]:

	Format	New	Closed	Relocation/ Expansion/ Conversion*	Ending Square Footage	Total Locations
0	WalmartDiscountStores	1		(8)	58,277,925	554
1	WalmartSupercenters	16	-	8	574,058,937	3,182
2	NeighborhoodMarkets	17	-	-	11,700,094	284
3	NeighborhoodMarkets	17	-	-	10,948,346	258
4	Amigo	-	-	-	672,148	24
5	Supermercado	-	-	-	79,600	2
6	SmallFormats	4	-	-	294,308	23
7	Marketside	-	-	-	-	-
8	SuperAhorros	-	-	-	46,349	3
9	WalmartExpress	4	-	-	239,702	17
10	WalmartonCampus	-	-	-	8,257	3
11	WalmartU.S.	38	-		644,331,264	4,043
12	Sam'sClub		-	-	82,669,348	620
13	TotalU.S.	38	-	-	727,000,612	4,663
14	International	50	(3)	(1)	349,900,213	6,194
15	TotalWalmart	88	(3)	(1)	1,076,900,825	10,857

... is a map



- shapefiles (.shp), geojson (.json), geotiff (.tiff)
- can contain interesting data
- Quantum GIS (http://www.ggis.org/pl/site/))
- ogr2ogr (http://www.gdal.org/ (http://www.gdal.org/))

In [16]:

!ogr2ogr -f CSV florida.csv data/florida/intersection.shp -lco GEOMETRY=AS_XYZ

In [17]:

!head florida.csv

X,Y,Z,ROADWAY,INTSEC_ROA,INTSEC_DIR,ROAD_SURF_,OLD_KEY,surface_de,in
tsec des,BEGIN POST

465256.85986718442291,3176275.000456310808659,0,77000032,HEARTHSTONE LN,2,,,,INTERSECTING STREET 90 DEGREES TO THE LEFT,1.7560

577439.59264366235584,2858089.087327975779772,0,87000337,NW 18 AV,5,A,28,ASPHALTIC CONCRETE,INTERSECTING STREET 90 DEGREES TO THE RIGHT, 0.4690

577892.025412083603442,2858107.435359993949533,0,87000337,NW 14 PL, 2,A,28,ASPHALTIC CONCRETE,INTERSECTING STREET 90 DEGREES TO THE LEF T,0.7500

580471.267142485827208,2858212.333280090242624,0,87000337,NW MIAMI C T,2,A,28,ASPHALTIC CONCRETE,INTERSECTING STREET 90 DEGREES TO THE LE FT.2.3520

581147.522503116168082,2858238.537824114784598,0,87000337,NE 3 AV,5,A,28,ASPHALTIC CONCRETE,INTERSECTING STREET 90 DEGREES TO THE RIGHT, 2.7720

578320.828612483106554,2861939.633763562887907,0,87000338,NW 108 TE R,5,A,28,ASPHALTIC CONCRETE,INTERSECTING STREET 90 DEGREES TO THE RI GHT,0.3480

551164.5057711917907,2816250.52012101188302,0,87000416,SOUTH POINT C ROSSING,5,A,28,ASPHALTIC CONCRETE,INTERSECTING STREET 90 DEGREES TO THE RIGHT,0.6610

551425.308459434658289,2816252.892217013984919,0,87000416,NW 7 PL,5, A,28,ASPHALTIC CONCRETE,INTERSECTING STREET 90 DEGREES TO THE RIGHT, 0.8240

561397.268276722170413,2854967.945693070068955,0,87000417,RAMP 87471 534,4,A,28,ASPHALTIC CONCRETE,INTERSECTING STREET 45 DEGREES TO THE RIGHT,0.0890

In [18]:

roads = pd.read_csv('florida.csv', sep=',')
roads[:3]

Out[18]:

	х	Υ	Z	ROADWAY	INTSEC_ROA	INTSEC_DIR	ROAD_
0	465256.859867	3.176275e+06	0	77000032	HEARTHSTONE LN	2	NaN
1	577439.592644	2.858089e+06	0	87000337	NW 18 AV	5	А
2	577892.025412	2.858107e+06	0	87000337	NW 14 PL	2	А

In [19]:

roads['X'].count()

Out[19]:

207133

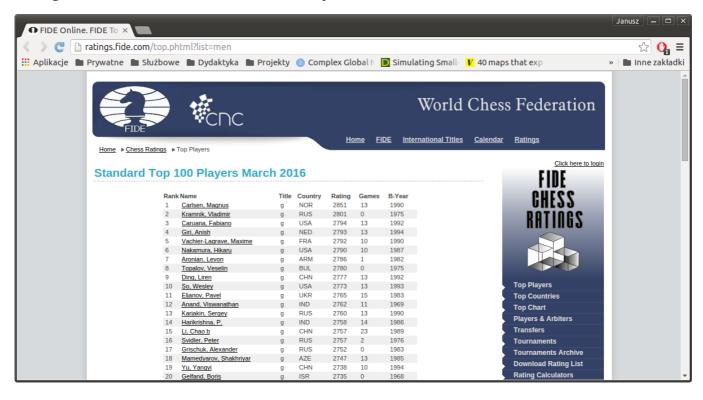
Screen scraping

From Wikipedia:

Screen scraping - a technique in which a computer program extracts data from human-readable output coming from another program.

The key element that distinguishes data scraping from regular parsing is that the output being scraped was intended for display to an end-user, rather than as input to another program, and is therefore usually neither documented nor structured for convenient parsing. Data scraping often involves ignoring binary data (usually images or multimedia data), display formatting, redundant labels, superfluous commentary, and other information which is either irrelevant or hinders automated processing

Google Docs as a basic screen scraper

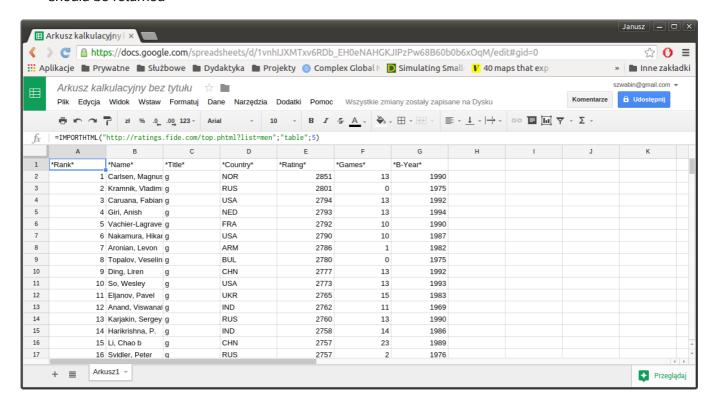


We are going to use the IMPORTHTML function to extract the table,

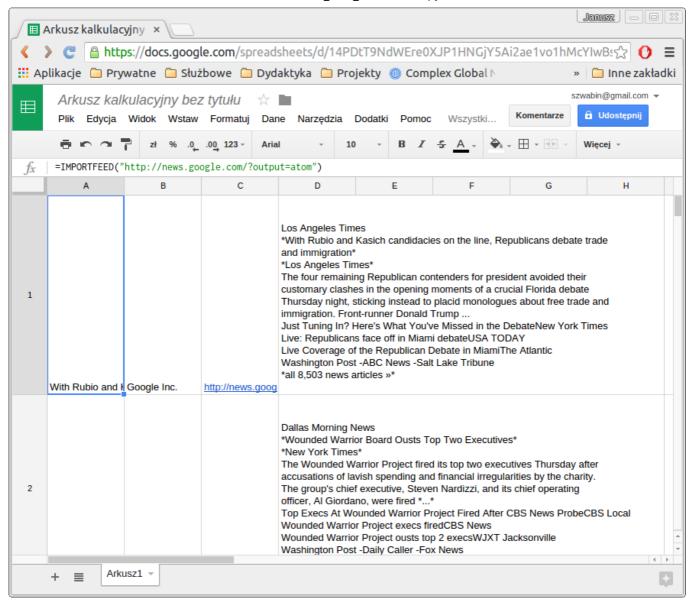
IMPORTHTML(url;type;index)

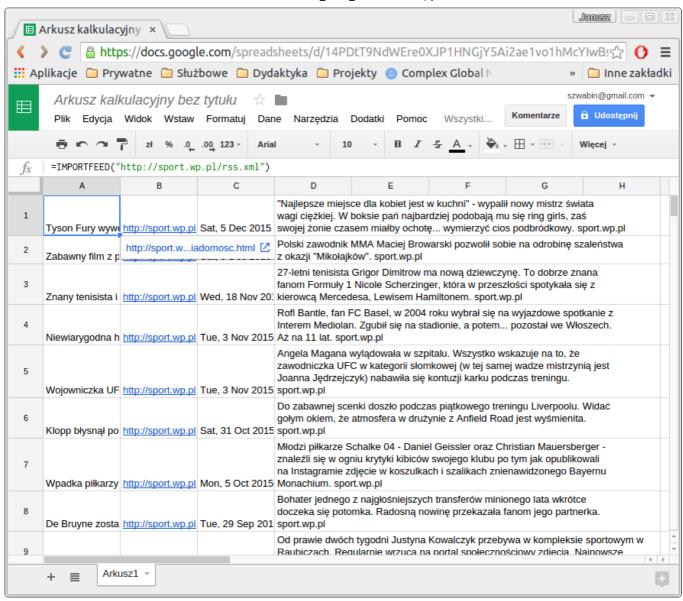
where:

- url URL of the page to examine, including protocol (e.g. http://)
- type either "list" or "table" depending on what type of structure contains the desired data.
- index the index, starting at 1, which identifies which table or list as defined in the HTML source should be returned



Similarly, we can use IMPORTXML or IMPORTFEED to read data from xml files and RSS feeds:





Custom scrapers

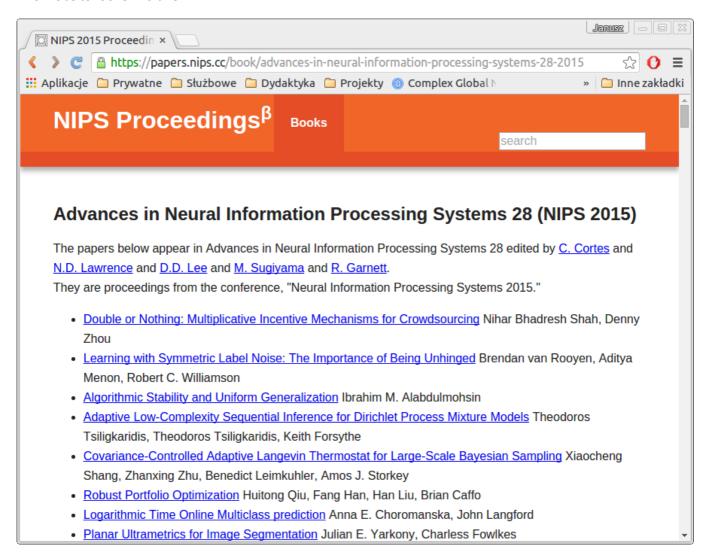
Python offers several libraries that may be used for scraping:

- reguests (http://docs.python-reguests.org/en/master/) (http://docs.python-reguests.org/en/master/))
- BeautifulSoup (http://www.crummy.com/software/BeautifulSoup/)
- scrapy (http://scrapy.org/)

Example - Requests and BeautifulSoup

In the 1st lecture, we talked (among other things) about papers from the NIPS 2015 conference. Our goal is now to download all these papers in pdf.

The website looks like this:



First of all we have to look at the source of the webpage in order to gain insight into the structure of the HTML code:

In [20]:

```
import requests

base_url = "http://papers.nips.cc"
index_url = "http://papers.nips.cc/book/advances-in-neural-information-processin
g-systems-28-2015"

r = requests.get(index_url)
print(r.text)
```

```
<!DOCTYPE html>
    <!--[if lt IE 7]>
                          <html class="no-js lt-ie9 lt-ie8 lt-ie7">
<![endif]-->
    <!--[if IE 7]>
                           <html class="no-js lt-ie9 lt-ie8"> <![end
if]-->
   <!--[if IE 8]>
                          <html class="no-js lt-ie9"> <![endif]-->
   <!--[if qt IE 8]><!--> <html class="no-js"> <!--<![endif]-->
    <head>
        <meta charset="utf-8">
        <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=</pre>
1">
        <title>NIPS 2015 Proceedings</title>
        <meta name="description" content="Eletronic Proceedings of N</pre>
eural Information Processing Systems">
        <meta name="viewport" content="width=device-width">
       <link rel="stylesheet" href="//media.nips.cc/nipsbooks/nipsp</pre>
apers/css/normalize.min.css">
        <link rel="stylesheet" href="//media.nips.cc/nipsbooks/nipsp</pre>
apers/css/main.css">
        <script src="//media.nips.cc/nipsbooks/nipspapers/js/vendor/</pre>
modernizr-2.6.2-respond-1.1.0.min.js"></script>
    </head>
    <body>
        <!--[if lt IE 7]>
            You are using an <strong>outdated
</strong> browser. Please <a href="http://browsehappy.com/">upgrade
your browser</a> or <a href="http://www.google.com/chromeframe/?red"
irect=true">activate Google Chrome Frame</a> to improve your experie
nce.
       <![endif]-->
        <div class="header-container">
            <header class="wrapper clearfix">
                    <h1 class="sitename"><a href="/">NIPS Proceeding
s</a><sup><math>\beta</sup></h1>
                <form action="/search/" class="search">
                    <input name="q" value="" class="search" placehol</pre>
der="search" />
                </form>
                <nav>
                    ul>
                        <a href="/">Books</a>
                    </nav>
            </header>
        </div>
```

<div class="main-container">
 <div class="main wrapper clearfix">

The papers below appear in Advances in Neural Information Processing Systems 28 edited by C. Cortes and N.D. Lawrence and D.D. Lee and M. Sugiyama and R. Garnett.
br />

They are proceedings from the conference, "Neural Information Processing Systems 2015."

ul>

="author">Jiajun Wu, Ilker Yildirim, Joseph J. Lim, Bill Freeman, Josh Tenenbaum

<a href="/paper/5778-bidirectional-recurrent-convolution
al-networks-for-multi-frame-super-resolution">Bidirectional Recurren
t Convolutional Networks for Multi-Frame Super-Resolution <a hre
f="/author/yan-huang-8094" class="author">Yan Huang, <a href="/a
uthor/wei-wang-4613" class="author">Wei Wang, <a href="/author/l
iang-wang-6455" class="author">Liang Wang

<a href="/paper/5876-streaming-distributed-variational-i
nference-for-bayesian-nonparametrics">Streaming, Distributed Variati
onal Inference for Bayesian Nonparametrics <a href="/author/trev
or-campbell-6631" class="author">Trevor Campbell, <a href="/auth
or/julian-straub-8223" class="author">Julian Straub, <a href="/a
uthor/john-w-fisher-iii-1965" class="author">John W. Fisher III,
Jonathan P. How

<a href="/paper/5791-deeply-learning-the-messages-in-mes"</pre>

sage-passing-inference">Deeply Learning the Messages in Message Pass
ing Inference <a href="/author/guosheng-lin-8116" class="autho
r">Guosheng Lin, <a href="/author/chunhua-shen-4132" class="auth
or">Chunhua Shen, <a href="/author/ian-reid-8117" class="autho
r">Ian Reid, <a href="/author/anton-van-den-hengel-7207" class
="author">Anton van den Hengel

Synaptic Sampling: A B ayesian Approach to Neural Network Plasticity and Rewiring <a hr ef="/author/david-kappel-8311" class="author">David Kappel, <a h ref="/author/stefan-habenschuss-5372" class="author">Stefan Habensch uss, Rob ert Legenstein, Wolfgang Maass

Nonparametric von Mises Estimators for Entropies, Divergences and Mutual Informations Kir thevasan Kandasamy, Akshay Krishnamurthy, Barnabas Poczos, Larry Wasserman, james m. robins

ass="author">Raquel Urtasun

<a href="/paper/6008-algorithms-with-logarithmic-or-subl
inear-regret-for-constrained-contextual-bandits">Algorithms with Log
arithmic or Sublinear Regret for Constrained Contextual Bandits
Huasen Wu, <a hr
ef="/author/r-srikant-8403" class="author">R. Srikant, Xin Liu, Chong Jiang

<a href="/paper/5879-spectral-learning-of-large-structur
ed-hmms-for-comparative-epigenomics">Spectral Learning of Large Stru
ctured HMMs for Comparative Epigenomics <a href="/author/chichen
g-zhang-7424" class="author">Chicheng Zhang, <a href="/author/ji
min-song-8227" class="author">Jimin Song, <a href="/author/kamal
ika-chaudhuri-3963" class="author">Kamalika Chaudhuri, Kevin Chen

<a href="/paper/5925-on-the-global-linear-convergence-of
-frank-wolfe-optimization-variants">On the Global Linear Convergence
of Frank-Wolfe Optimization Variants <a href="/author/simon-laco
ste-julien-7161" class="author">Simon Lacoste-Julien, <a href="/</pre>

author/martin-jaggi-7651" class="author">Martin Jaggi

<a href="/paper/5847-attention-based-models-for-speech-r
ecognition">Attention-Based Models for Speech Recognition <a hre</pre>

f="/author/jan-k-chorowski-8177" class="author">Jan K. Chorowski, Dzmitry Ba hdanau, Dm itriy Serdyuk, Kyunghyun Cho, Yoshua Bengio

Closed-form Estimators for Highdimensional Generalized Linear Models Eunho Yang, Aurelie C. Lozano, Pradeep K. Ravikumar

Online Rank Elicitation for Plack ett-Luce: A Dueling Bandits Approach Balázs Szörényi, Róbert Busa-Fekete, Adil Paul, Eyke Hüllermeier

Tractable Bayesian Netw ork Structure Learning with Bounded Vertex Cover Number Janne H. Korhonen, Pekka Parv iainen

Learning Large-Scale Poisson D AG Models based on OverDispersion Scoring Gunwo ong-park, Garvesh Raskutti

<a href="/paper/5709-monotone-k-submodular-function-maxi
mization-with-size-constraints">Monotone k-Submodular Function Maxim
ization with Size Constraints Class="author">Naoto Ohsaka, Yuichi Yoshida

Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting/a > Xingjian SHI, Zhourong Chen
, Hao Wang, <
a href="/author/dit-yan-yeung-1323" class="author">Dit-Yan Yeung, Wai-kin Wong, Wang-chun WO 0

<a href="/paper/5699-gap-safe-screening-rules-for-sparse
-multi-task-and-multi-class-models">GAP Safe screening rules for spa
rse multi-task and multi-class models <a href="/author/eugene-nd
iaye-7989" class="author">Eugene Ndiaye, <a href="/author/olivie
r-fercoq-7990" class="author">Olivier Fercoq, <a href="/author/a
lexandre-gramfort-4730" class="author">Alexandre Gramfort, <a hr
ef="/author/joseph-salmon-7307" class="author">Joseph Salmon

<a href="/paper/5907-maximum-likelihood-learning-with-ar</pre>

bitrary-treewidth-via-fast-mixing-parameter-sets">Maximum Likelihood Learning With Arbitrary Treewidth via Fast-Mixing Parameter Sets Justin Domke

v G. Gogate

Predtron: A Family of Online Al gorithms for General Prediction Problems Prateek Jain, Nagarajan Natarajan, Ambuj Tewari

Weig hted Theta Functions and Embeddings with Applications to Max-Cut, Cl ustering and Summarization Fredrik D. Johansson, Ankani Chattoraj, Chiranjib Bhattach aryya, De vdatt Dubhashi

<a href="/paper/5878-human-memory-search-as-initial-visi
t-emitting-random-walk">Human Memory Search as Initial-Visit Emittin
g Random Walk <a href="/author/kwang-sung-jun-8224" class="autho
r">Kwang-Sung Jun, <a href="/author/xiaojin-zhu-2978" class="aut
hor">Xiaojin Zhu, <a href="/author/timothy-t-rogers-4348" class
="author">Timothy T. Rogers, <a href="/author/zhuoran-yang-8225"
class="author">Zhuoran Yang, <a href="/author/ming-yuan-8226" c
lass="author">ming yuan

<a href="/paper/5921-convergence-rates-of-active-learnin
g-for-maximum-likelihood-estimation">Convergence Rates of Active Lea
rning for Maximum Likelihood Estimation <a href="/author/kamalik
a-chaudhuri-3963" class="author">Kamalika Chaudhuri, <a href="/a
uthor/sham-m-kakade-2572" class="author">Sham M. Kakade, Praneeth Netrapal
li, Sujay S
anghavi

class="author">David Blei

<a href="/paper/6022-unified-view-of-matrix-completion-u
nder-general-structural-constraints">Unified View of Matrix Completi
on under General Structural Constraints Suriya Gunasekar, <a href="/autho
r/arindam-banerjee-5303" class="author">Arindam Banerjee, <a hre
f="/author/joydeep-ghosh-1501" class="author">Joydeep Ghosh

Online Learning with Adversarial Delays Kent Quanrud, Daniel Khashabi

<a href="/paper/5985-efficient-thompson-sampling-for-onl
ine-matrix-factorization-recommendation">Efficient Thompson Sampling
for Online Matrix-Factorization Recommendation <a href="/autho
r/jaya-kawale-8376" class="author">Jaya Kawale, <a href="/autho
r/hung-h-bui-8377" class="author">Hung H. Bui, <a href="/author/
branislav-kveton-2628" class="author">Branislav Kveton, Long Tran-Thanh,
 Sanjay Chawla/
a>

Improved I teration Complexity Bounds of Cyclic Block Coordinate Descent for Co nvex Problems Ru oyu Sun, Mingy i Hong

<a href="/paper/5670-fast-second-order-stochastic-backpr</pre>

opagation-for-variational-inference">Fast Second Order Stochastic Ba ckpropagation for Variational Inference Kai Fan, A href="/author/ziteng-wang-650">A href="/author/jeff-beck-5318" class="author">Jeff Beck, James Kwok, A href="/author/katherine-a-heller-402"

<a href="/paper/5959-cross-domain-matching-for-bag-of-wo
rds-data-via-kernel-embeddings-of-latent-distributions">Cross-Domain
Matching for Bag-of-Words Data via Kernel Embeddings of Latent Dist
ributions Y
uya Yoshikawa, <a href="/author/tomoharu-iwata-2797" class="auth
or">Tomoharu Iwata, <a href="/author/hiroshi-sawada-7478" class
="author">Hiroshi Sawada, <a href="/author/takeshi-yamada-4351"
class="author">Takeshi Yamada

Mixing Time Estimation in Reversible Markov Chains from a Single Sample Path <a href="/au

thor/daniel-j-hsu-6464" class="author">Daniel J. Hsu, Aryeh Kontorovich, Csaba Szepes vari

Deep Generative Image M odels using a Laplacian Pyramid of Adversarial Networks Emily L. Denton, Soumith Chintala, arthur szlam, Rob Fergus

<a href="/paper/5858-decoupled-deep-neural-network-for-s
emi-supervised-semantic-segmentation">Decoupled Deep Neural Network
for Semi-supervised Semantic Segmentation <a href="/author/seun
ghoon-hong-8189" class="author">Seunghoon Hong, <a href="/autho
r/hyeonwoo-noh-8190" class="author">Hyeonwoo Noh, <a href="/auth
or/bohyung-han-7598" class="author">Bohyung Han

BACKSHIFT: Learning causal cy clic graphs from unknown shift interventions Dominik Rothenhäusler, Christina He inze, Jonas P eters, Nicolai Meinshausen

<a href="/paper/6031-asynchronous-stochastic-convex-opti
mization-the-noise-is-in-the-noise-and-sgd-dont-care">Asynchronous s

tochastic convex optimization: the noise is in the noise and SGD don 't care Sorathan Chaturapruek, John C. Duchi, Christopher Ré

A Tractabl e Approximation to Optimal Point Process Filtering: Application to N eural Encoding Yuval Harel, R on Meir, Manf red Opper

Teaching Machines to Read and Comprehend Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, Phil Blunsom

<a href="/paper/5882-inverse-reinforcement-learning-with
-locally-consistent-reward-functions">Inverse Reinforcement Learning
with Locally Consistent Reward Functions Quoc Phong Nguyen, <a href="/a
uthor/bryan-kian-hsiang-low-8233" class="author">Bryan Kian Hsiang L
ow, Patric
k Jaillet

<a href="/paper/5967-end-to-end-learning-of-lda-by-mirro
r-descent-back-propagation-over-a-deep-architecture">End-to-end Lear

ning of LDA by Mirror-Descent Back Propagation over a Deep Architect ure Jianshu Chen, Jianshu Chen, Yelong Shen, Yelong Shen, Lin Xiao, Xiaodong He, Jianfeng Gao, Xinying Song, Li Deng

<a href="/paper/5687-regressive-virtual-metric-learnin
g">Regressive Virtual Metric Learning <a href="/author/michael-p
errot-7974" class="author">Michaël Perrot, <a href="/author/amau
ry-habrard-7975" class="author">Amaury Habrard

Analysis of Robust PCA via Local Incoherence Huishuai Zhang, Yi Zhou, Yingbin Liang

Max-Margin Deep Generative Models Chongxuan Li, 1" class="author">Jun Zhu, <a href="/author/tianlin-shi-8325" cl</pre>

ass="author">Tianlin Shi, <a href="/author/bo-zhang-3841" class
="author">Bo Zhang

<a href="/paper/5769-biologically-inspired-dynamic-textu
res-for-probing-motion-perception">Biologically Inspired Dynamic Tex

tures for Probing Motion Perception Jonathan Vacher, Andrew Isaac Meso, Laurent U. Perrinet, Gabriel Peyr é

Ro
bust PCA with compressed data <a href="/author/wooseok-ha-7995"
 class="author">Wooseok Ha, Rina Foygel Barber

Sampling from Probabilistic Submodular Models <a hre f="/author/alkis-gotovos-8046" class="author">Alkis Gotovos, Hamed Hassani, Andreas Krause

<a href="/paper/5673-supervised-learning-for-dynamical-s
ystem-learning">Supervised Learning for Dynamical System Learning Ahmed Hefny,
 Carlton Downey
, Geoffr
ey J. Gordon

<a href="/paper/5910-regret-based-pruning-in-extensive-f
orm-games">Regret-Based Pruning in Extensive-Form Games Noam Brown, Tuomas Sandholm

Fast Two-Sample Testing with Analytic Representations of Probability Measures Kacper P. Chwia lkowski, Aa ditya Ramdas, Dino Sejdinovic, Arthur Gretton

ef="/author/piotr-dollar-7400" class="author">Piotr Dollar

Collaborative Filterin g with Graph Information: Consistency and Scalable Methods <a hr ef="/author/nikhil-rao-6320" class="author">Nikhil Rao, Hsiang-Fu Yu, Pradeep K. Ravik umar, I nderjit S. Dhillon

<a href="/paper/5762-competitive-distribution-estimation
-why-is-good-turing-good">Competitive Distribution Estimation: Why i

s Good-Turing Good Alon Orlitsky, Ananda Theertha Suresh

ass="author">Raman Sankaran, Chiranjib Bhattacharyya, Francis Bach

<a href="/paper/5696-on-the-convergence-of-stochastic-gr</pre>

adient-mcmc-algorithms-with-high-order-integrators">On the Convergen ce of Stochastic Gradient MCMC Algorithms with High-Order Integrator s Changyou Chen, Nan Ding, Lawrence Carin in

<a href="/paper/5641-are-you-talking-to-a-machine-datase
t-and-methods-for-multilingual-image-question">Are You Talking to a
Machine? Dataset and Methods for Multilingual Image Question Haoyuan Gao, Junhua Mao, <a hr
ef="/author/jie-zhou-7882" class="author">Jie Zhou, <a href="/au
thor/zhiheng-huang-7883" class="author">Zhiheng Huang, Lei Wang, Wei Xu

Generalization in Adaptive Data Analysis an d Holdout Reuse <a href="/author/cynthia-dwork-8382" class="auth</pre> or">Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toni Pitassi, Omer Reingold, Aaron Roth

Sparse Linear Programming v ia Primal and Dual Augmented Coordinate Descent Ian En-Hsu Yen, Kai Zhong, Cho-Jui Hsieh, Pradeep K. Ravikumar, a>, Inderji t S. Dhillon

Bayesian Active Model Sele ction with an Application to Automated Audiometry Jacob Gardner, Gustavo Malkomes, Roman Garnett, Kilian Q. Weinberger, Dennis Barbour, John P. Cunningham

Learning spatiotemporal tr ajectories from manifold-valued longitudinal data Jean-Baptiste SCHIRA TTI, Stéphanie ALLASSONNIERE, diass="author">Olivier Colliot, Stanley DURRLEMAN

="/author/rajesh-p-rao-2384" class="author">Rajesh P. Rao

End-To-End Memory Networks Sainbayar Sukhbaatar, arthur szlam, Jason Weston, Rob Ferqus

<a href="/paper/5650-a-theory-of-decision-making-under-d
ynamic-context">A Theory of Decision Making Under Dynamic Context Michael S
hvartsman, <a href="/author/vaibhav-srivastava-7902" class="auth
or">Vaibhav Srivastava, <a href="/author/jonathan-d-cohen-347" c
lass="author">Jonathan D. Cohen

<a href="/paper/5960-a-gaussian-process-model-of-quasar-</pre>

spectral-energy-distributions">A Gaussian Process Model of Quasar Spectral Energy Distributions Andrew Miller, Albert Wu, Jeff Regier, Jon McAuliffe, Dustin Lang, Mr. Prabhat, David Schlegel, Ryan P. Adams

7143" class="author">Alan Malek, Peter L. Bartlett, Yasin Abbasi

<a href="/paper/5983-estimating-jaccard-index-with-missi
ng-observations-a-matrix-calibration-approach">Estimating Jaccard In
dex with Missing Observations: A Matrix Calibration Approach Wenye Li

lla

Taming the Wild: A Unified Analysis of Hogwild-Style Algorithms Christopher M. De Sa, Ce Zhang, Kunle Olukotun, Christopher Ré, Christopher Ré, Christopher Ré

Gram
mar as a Foreign Language <a href="/author/oriol-vinyals-5571" c
lass="author">Oriol Vinyals, <a href="/author/ukasz-kaiser-7869"
class="author">Łukasz Kaiser, <a href="/author/terry-koo-7870"
class="author">Terry Koo, <a href="/author/slav-petrov-3753" cl
ass="author">Slav Petrov, <a href="/author/ilya-sutskever-3959"
class="author">Ilya Sutskever, Geoffrey Hinton

Winne
r-Take-All Autoencoders <a href="/author/alireza-makhzani-8100"
class="author">Alireza Makhzani, <a href="/author/brendan-j-fre
y-1370" class="author">Brendan J. Frey

Deep
Poisson Factor Modeling <a href="/author/ricardo-henao-4452" cl
ass="author">Ricardo Henao, <a href="/author/zhe-gan-7916" class
="author">Zhe Gan, <a href="/author/james-lu-8107" class="autho
r">James Lu, <a href="/author/lawrence-carin-3001" class="autho
r">Lawrence Carin

-using-deep-networks-in-atari-games">Action-Conditional Video Prediction using Deep Networks in Atari Games Junhyuk Oh, Xiaoxiao Guo, Honglak Lee, Richard L. Lewis, Satinder Singh

<a href="/paper/5752-distributed-submodular-cover-succin
ctly-summarizing-massive-data">Distributed Submodular Cover: Succinc
tly Summarizing Massive Data <a href="/author/baharan-mirzasolei
man-6553" class="author">Baharan Mirzasoleiman, <a href="/autho
r/amin-karbasi-6295" class="author">Amin Karbasi, <a href="/auth
or/ashwinkumar-badanidiyuru-8058" class="author">Ashwinkumar Badanid
iyuru, Andr
eas Krause

<a href="/paper/5834-structured-estimation-with-atomic-n
orms-general-bounds-and-applications">Structured Estimation with Ato
mic Norms: General Bounds and Applications <a href="/author/shen
g-chen-7218" class="author">Sheng Chen, <a href="/author/arindam
-banerjee-5303" class="author">Arindam Banerjee

en Pasteris,
Shaona Ghosh

Learning Continuous Control Policie s by Stochastic Value Gradients Nicolas Heess, Gregory Wayne, David Silver, Tim Lillicrap, Tom Erez, Yuval Tassa

Para llel Multi-Dimensional LSTM, With Application to Fast Biomedical Vol umetric Image Segmentation href="/author/marijn-f-stollenga-7 189" class="author">Marijn F. Stollenga, Wonmin Byeon, Marcus Liwicki, Juergen Schmidhuber

<a href="/paper/6016-no-regret-learning-in-bayesian-game
s">No-Regret Learning in Bayesian Games <a href="/author/jason-h
artline-8412" class="author">Jason Hartline, <a href="/author/va
silis-syrgkanis-8069" class="author">Vasilis Syrgkanis, Eva Tardos

<a href="/paper/5887-statistical-topological-data-analys
is-a-kernel-perspective">Statistical Topological Data Analysis - A K
ernel Perspective <a href="/author/roland-kwitt-8239" class="aut
hor">Roland Kwitt, <a href="/author/stefan-huber-8240" class="au
thor">Stefan Huber, <a href="/author/marc-niethammer-8241" class
="author">Marc Niethammer, <a href="/author/weili-lin-8242" clas
s="author">Weili Lin, <a href="/author/ulrich-bauer-8243" class
="author">Ulrich Bauer

287" class="author">Quoc V. Le

<a href="/paper/5757-rapidly-mixing-gibbs-sampling-for-a
-class-of-factor-graphs-using-hierarchy-width">Rapidly Mixing Gibbs
Sampling for a Class of Factor Graphs Using Hierarchy Width Christopher
M. De Sa, Ce Zha
ng, Kunle 0
lukotun, Ch
ristopher Ré

<a href="/paper/5823-interpolating-convex-and-non-convex
-tensor-decompositions-via-the-subspace-norm">Interpolating Convex a
nd Non-Convex Tensor Decompositions via the Subspace Norm Qinqing Zheng, Ryota Tomioka

Submodu lar Hamming Metrics Jennifer A. Gillenwater, Rishabh K. Iyer, Bethany Lusch, Rahul Kidambi, Jeff A. Bilmes

r">Volkan Cevher

<a href="/paper/5909-learning-causal-graphs-with-small-i
nterventions">Learning Causal Graphs with Small Interventions Karthikeya
n Shanmugam, <a href="/author/murat-kocaoglu-7411" class="autho
r">Murat Kocaoglu, <a href="/author/alexandros-g-dimakis-7413" c
lass="author">Alexandros G. Dimakis, <a href="/author/sriram-vis
hwanath-8274" class="author">Sriram Vishwanath

<a href="/paper/5877-fixed-length-poisson-mrf-adding-dep
endencies-to-the-multinomial">Fixed-Length Poisson MRF: Adding Depen
dencies to the Multinomial <a href="/author/david-i-inouye-7228"
 class="author">David I. Inouye, <a href="/author/pradeep-k-ravi
kumar-6369" class="author">Pradeep K. Ravikumar, <a href="/autho
r/inderjit-s-dhillon-6368" class="author">Inderjit S. Dhillon

<a href="/paper/5770-large-scale-bayesian-multi-label-le
arning-via-topic-based-label-embeddings">Large-Scale Bayesian MultiLabel Learning via Topic-Based Label Embeddings <a href="/autho
r/piyush-rai-4147" class="author">Piyush Rai, <a href="/author/c
hangwei-hu-8083" class="author">Changwei Hu, <a href="/author/ri
cardo-henao-4452" class="author">Ricardo Henao, <a href="/author/ri
r/lawrence-carin-3001" class="author">Lawrence Carin

<a href="/paper/5748-the-self-normalized-estimator-for-c
ounterfactual-learning">The Self-Normalized Estimator for Counterfac

tual Learning Adith Swaminathan, Thorsten Joachims

Seguy, Marco Cuturi

<a href="/paper/5810-inference-for-determinantal-point-p
rocesses-without-spectral-knowledge">Inference for determinantal poi

nt processes without spectral knowledge Rémi Bardenet, Michalis Titsias RC AUEB

="author">Percy S. Liang, Christopher D. Manning

<a href="/paper/5900-lifted-inference-rules-with-constra
ints">Lifted Inference Rules With Constraints <a href="/author/h
appy-mittal-7499" class="author">Happy Mittal, <a href="/author/
anuj-mahajan-8263" class="author">Anuj Mahajan, <a href="/autho
r/vibhav-g-gogate-7474" class="author">Vibhav G. Gogate, Parag Singla

s Pualo Reis, <a href="/author/gerhard-neumann-3982" class="auth
or">Gerhard Neumann

<a href="/paper/5947-semi-supervised-learning-with-ladde
r-networks">Semi-supervised Learning with Ladder Networks <a hre
f="/author/antti-rasmus-8307" class="author">Antti Rasmus, <a hr
ef="/author/mathias-berglund-7903" class="author">Mathias Berglund, Mikko Honkal
a, Harri Val
pola, Tapani
Raiko

<a href="/paper/5924-a-dual-augmented-block-minimization
-framework-for-learning-with-limited-memory">A Dual Augmented Block
Minimization Framework for Learning with Limited Memory Ian En-Hsu Yen, Shan-Wei Lin, <
a href="/author/shou-de-lin-7476" class="author">Shou-De Lin

<a href="/paper/5792-efficient-learning-of-continuous-ti
me-hidden-markov-models-for-disease-progression">Efficient Learning
of Continuous-Time Hidden Markov Models for Disease Progression
Yu-Ying Liu,
Shuang Li, <a h
ref="/author/fuxin-li-4523" class="author">Fuxin Li, <a href="/a
uthor/le-song-3161" class="author">Le Song, <a href="/author/jam
es-m-rehg-2077" class="author">James M. Rehg

<a href="/paper/5674-expectation-particle-belief-propaga
tion">Expectation Particle Belief Propagation <a href="/author/t</pre>

hibaut-lienart-7953" class="author">Thibaut Lienart, Yee Whye Teh, Arnaud Doucet

```
<aside>
                    <h3>Neural Information Processing Systems (NIPS)
</h3>
                    Papers published at the Neural Information Pr
ocessing Systems Conference.
                </aside>
            </div> <!-- #main -->
        </div> <!-- #main-container -->
        <div class="footer-container">
            <footer class="wrapper">
                <h3>&copy; 1987 - 2017 Neural Information Processing
Systems Foundation, Inc.</h3>
            </footer>
        </div>
        <script src="//media.nips.cc/nipsbooks/nipspapers/js/vendor/</pre>
jquery-1.10.2.min.js"></script>
        <script src="//media.nips.cc/nipsbooks/nipspapers/js/main.j</pre>
s"></script>
<script type="text/javascript">
// using jQuery
function getCookie(name) {
    var cookieValue = null;
    if (document.cookie && document.cookie != '') {
        var cookies = document.cookie.split(';');
        for (var i = 0; i < cookies.length; i++) {
            var cookie = jQuery.trim(cookies[i]);
            // Does this cookie string begin with the name we want?
            if (cookie.substring(0, name.length + 1) == (name + 1)
 '=')) {
                cookieValue = decodeURIComponent(cookie.substring(na
me.length + 1));
                break;
            }
        }
    }
    return cookieValue;
```

The piece of code with the information about the paper is in the following format:

<a href="/paper/5677-double-or-nothing-multiplicative-incentive-mecha
nisms-for-crowdsourcing">Double or Nothing: Multiplicative Incentive Mech
anisms for Crowdsourcing <a href="/author/nihar-bhadresh-shah-7956" c
lass="author">Nihar Bhadresh Shah, <a href="/author/denny-zhou-2649"
class="author">Denny Zhou

We may also check the webpage indicated by this code:

In [21]:

```
paper_ref = "/paper/5677-double-or-nothing-multiplicative-incentive-mechanisms-f
or-crowdsourcing"
paper_url = base_url + paper_ref
r = requests.get(paper_url)
print(r.text)
```

```
<!DOCTYPE html>
    <!--[if lt IE 7]>
                           <html class="no-js lt-ie9 lt-ie8 lt-ie7">
<![endif]-->
    <!--[if IE 7]>
                           <html class="no-js lt-ie9 lt-ie8"> <![end
if]-->
   <!--[if IE 8]>
                           <html class="no-js lt-ie9"> <![endif]-->
    <!--[if qt IE 8]><!--> <html class="no-js"> <!--<![endif]-->
    <head>
        <meta charset="utf-8">
        <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=</pre>
1">
        <title>Double or Nothing: Multiplicative Incentive Mechanism
s for Crowdsourcing</title>
    <meta name="citation title" content="Double or Nothing: Multipli</pre>
cative Incentive Mechanisms for Crowdsourcing">
    <meta name="citation_author" content="Nihar Bhadresh Shah">
    <meta name="citation author" content="Denny Zhou">
    <meta name="citation publication date" content="2015">
    <meta name="citation conference title" content="Advances in Neur</pre>
al Information Processing Systems">
    <meta name="citation firstpage" content="1">
    <meta name="citation lastpage" content="9">
    <meta name="citation pdf url" content="http://papers.nips.cc/pap</pre>
er/5677-double-or-nothing-multiplicative-incentive-mechanisms-for-cr
owdsourcing.pdf">
        <meta name="description" content="Eletronic Proceedings of N</pre>
eural Information Processing Systems">
        <meta name="viewport" content="width=device-width">
        <link rel="stylesheet" href="//media.nips.cc/nipsbooks/nipsp</pre>
apers/css/normalize.min.css">
        <link rel="stylesheet" href="//media.nips.cc/nipsbooks/nipsp</pre>
apers/css/main.css">
        <script src="//media.nips.cc/nipsbooks/nipspapers/js/vendor/</pre>
modernizr-2.6.2-respond-1.1.0.min.js"></script>
    </head>
    <body>
        <!--[if lt IE 7]>
```

<![endif]-->

<div class="header-container">
 <header class="wrapper clearfix">

 $$<\!h1$$ class="sitename">NIPS Proceeding s^{$\beta}</h1>$

Books

<a href="/book/advances-in-neural-information-processing-sys
tems-28-2015">2015

<h2 class="subtitle">Double or Nothing: Multiplicative Incentive Mec hanisms for Crowdsourcing</h2>

Part of: Advances in Neural Information Processing Systems 28 (NIPS 2015)

[PDF]

[BibTeX]

[Supplemental]

<a href="//media.nips.cc/nipsbooks/nipspapers/paper_files/nips2
8/reviews/2.html">[Reviews]

<h3>Authors</h3>

Ni
har Bhadresh Shah

Denny Zhou

<h3>Conference Event Type: Poster</h3>

<h3>Abstract</h3>

Crowdsourcing has gained immense popularity in m achine learning applications for obtaining large amounts of labeled data. Crowdsourcing is cheap and fast, but suffers from the problem of low-quality data. To address this fundamental challenge in crowds ourcing, we propose a simple payment mechanism to incentivize worker s to answer only the questions that they are sure of and skip the re st. We show that surprisingly, under a mild and natural no-free-lunc h requirement, this mechanism is the one and only incentive-compatib le payment mechanism possible. We also show that among all possible incentive-compatible mechanisms (that may or may not satisfy no-fr ee-lunch), our mechanism makes the smallest possible payment to spam Interestingly, this unique mechanism takes a multiplicative f orm. The simplicity of the mechanism is an added benefit. In prelim inary experiments involving over several hundred workers, we observe a significant reduction in the error rates under our unique mechanis m for the same or lower monetary expenditure.

<aside>

<h3>Neural Information Processing Systems (NIPS)

</h3>

```
</div> <!-- #main -->
        </div> <!-- #main-container -->
        <div class="footer-container">
            <footer class="wrapper">
                <h3>&copy; 1987 - 2017 Neural Information Processing
Systems Foundation, Inc.</h3>
            </footer>
        </div>
        <script src="//media.nips.cc/nipsbooks/nipspapers/js/vendor/</pre>
jquery-1.10.2.min.js"></script>
        <script src="//media.nips.cc/nipsbooks/nipspapers/js/main.j</pre>
s"></script>
<script type="text/javascript">
// using jQuery
function getCookie(name) {
    var cookieValue = null;
    if (document.cookie && document.cookie != '') {
        var cookies = document.cookie.split(';');
        for (var i = 0; i < cookies.length; i++) {</pre>
            var cookie = jQuery.trim(cookies[i]);
            // Does this cookie string begin with the name we want?
            if (cookie.substring(0, name.length + 1) == (name +
 '=')) {
                cookieValue = decodeURIComponent(cookie.substring(na
me.length + 1));
                break:
            }
        }
    }
    return cookieValue;
var csrftoken = getCookie('csrftoken');
function csrfSafeMethod(method) {
    // these HTTP methods do not require CSRF protection
    return (/^(GET|HEAD|OPTIONS|TRACE)$/.test(method));
function sameOrigin(url) {
    // test that a given url is a same-origin URL
    // url could be relative or scheme relative or absolute
    var host = document.location.host; // host + port
    var protocol = document.location.protocol;
    var sr origin = '//' + host;
    var origin = protocol + sr origin;
    // Allow absolute or scheme relative URLs to same origin
    return (url == origin || url.slice(0, origin.length + 1) == orig
in + '/') ||
        (url == sr_origin || url.slice(0, sr_origin.length + 1) == s
r origin + '/') ||
        // or any other URL that isn't scheme relative or absolute
i.e relative.
        !(/^(\/\/|http:|https:).*/.test(url));
}
```

```
$.ajaxSetup({
    beforeSend: function(xhr, settings) {
        if (!csrfSafeMethod(settings.type) && sameOrigin(settings.ur
1)) {
            // Send the token to same-origin, relative URLs only.
            // Send the token only if the method warrants CSRF prote
ction
            // Using the CSRFToken value acquired earlier
            xhr.setReguestHeader("X-CSRFToken", csrftoken);
        }
    }
});
</script>
<script type="text/javascript" src="//media.nips.cc/nipsbooks/autoco</pre>
mplete light/autocomplete.js"></script>
<script type="text/javascript" src="//media.nips.cc/nipsbooks/autoco</pre>
mplete light/widget.js"></script>
<script type="text/javascript" src="//media.nips.cc/nipsbooks/autoco</pre>
mplete light/addanother.js"></script>
<script type="text/javascript" src="//media.nips.cc/nipsbooks/autoco</pre>
mplete light/text widget.js"></script>
<script type="text/javascript" src="//media.nips.cc/nipsbooks/autoco</pre>
mplete light/remote.js"></script>
<link rel="stylesheet" type="text/css" href="//media.nips.cc/nipsboo</pre>
ks/autocomplete light/style.css" />
        <script>
          (function(i,s,o,q,r,a,m){i['GoogleAnalyticsObject']=r;i[r]
=i[r]||function(){
          (i[r].q=i[r].q||[]).push(arguments), i[r].l=1*new Date(); a
=s.createElement(o),
          m=s.getElementsByTagName(o)[0];a.async=1;a.src=q;m.parentN
ode.insertBefore(a,m)
          })(window,document,'script','//www.google-analytics.com/an
alytics.js','ga');
          ga('create', 'UA-2097757-2', 'nips.cc');
          ga('send', 'pageview');
        </script>
    <script type="text/javascript" src="//media.nips.cc/nipsbooks/ni</pre>
pspapers/js/MathJax/MathJax.js"></script>
    </body>
</html>
<!-- Do not remove: This comment is monitored to verify that the sit
e is working properly -->
```

Analyzing the above code gives us the source of the pdf file:

```
<a href="/paper/5677-double-or-nothing-multiplicative-incentive-mechanism
s-for-crowdsourcing.pdf">[PDF]</a>
```

So, we will do the scraping in the following steps:

- we extract the subpage, title, id and authors of a paper from the webpage containing the list of all papers
- we build url of the pdf file
- · we download the pdf file

The HTML code may be parsed by the BeautifulSoup module:

In [22]:

```
from bs4 import BeautifulSoup

base_url = "http://papers.nips.cc"
index_url = "http://papers.nips.cc/book/advances-in-neural-information-processin
g-systems-28-2015"

r = requests.get(index_url)

soup = BeautifulSoup(r.content, "lxml")
paper_links = [link for link in soup.find_all('a') if link["href"][:7]=="/pape
r/"]
print("%d Papers Found" % len(paper_links))
```

403 Papers Found

In [23]:

```
for i in paper_links[:4]:
    print(i)
    print()
```

<a href="/paper/5677-double-or-nothing-multiplicative-incentive-mech
anisms-for-crowdsourcing">Double or Nothing: Multiplicative Incentiv
e Mechanisms for Crowdsourcing

<a href="/paper/5941-learning-with-symmetric-label-noise-the-importa
nce-of-being-unhinged">Learning with Symmetric Label Noise: The Impo
rtance of Being Unhinged

<a href="/paper/6019-algorithmic-stability-and-uniform-generalizatio"
n">Algorithmic Stability and Uniform Generalization

<a href="/paper/6035-adaptive-low-complexity-sequential-inference-fo
r-dirichlet-process-mixture-models">Adaptive Low-Complexity Sequenti
al Inference for Dirichlet Process Mixture Models

In [24]:

```
import re

for link in paper_links[:4]:
    paper_title = link.contents[0]
    info_link = base_url + link["href"]
    pdf_link = info_link + ".pdf"
    pdf_name = link["href"][7:] + ".pdf"
    paper_id = re.findall(r"^(\d+)-", pdf_name)[0]
    pdf = requests.get(pdf_link)
    pdf_file = open(pdf_name, "wb")
    pdf_file.write(pdf.content)
    pdf_file.close()
    print(paper_id,paper_title)
```

5677 Double or Nothing: Multiplicative Incentive Mechanisms for Crow dsourcing
5941 Learning with Symmetric Label Noise: The Importance of Being Un hinged
6019 Algorithmic Stability and Uniform Generalization
6035 Adaptive Low-Complexity Sequential Inference for Dirichlet Proc ess Mixture Models

In [25]:

```
!ls *.pdf
```

5677-double-or-nothing-multiplicative-incentive-mechanisms-for-crowd sourcing.pdf
5941-learning-with-symmetric-label-noise-the-importance-of-being-unh inged.pdf
6019-algorithmic-stability-and-uniform-generalization.pdf
6035-adaptive-low-complexity-sequential-inference-for-dirichlet-proc ess-mixture-models.pdf

Having the pdf files, we can transform them to txt format for further analysis:

In [26]:

```
import subprocess
subprocess.call(["pdftotext", pdf_name, "temp.txt"])
```

Out[26]:

In [27]:

```
!head temp.txt
```

Adaptive Low-Complexity Sequential Inference for Dirichlet Process Mixture Models Theodoros Tsiligkaridis, Keith W. Forsythe Massachusetts Institute of Technology, Lincoln Laboratory Lexington, MA 02421 USA ttsili@ll.mit.edu, forsythe@ll.mit.edu

Abstract

We develop a sequential low-complexity inference procedure for Diric hlet process mixtures of Gaussians for online clustering and paramet er estimation when the number of clusters are unknown a-priori. We present an easily computable,

Scrapy

Scrapy (http://scrapy.org/) is an open source and collaborative framework for extracting the data you need from websites.

Before you start scraping, you have to set up a new scrapy project. Enter a directory where you would like to store your code and run:

scrapy startproject iftpapers

This will create the iftpapers directory with the following content:

```
iftpapers/
                           <-- project directory
    scrapy.cfg
                           <-- configuration file
                          <-- project's Python module
    iftpapers/
        init .py
       items.py
                          <-- project items definition file
                          <-- project pipelines file
       pipelines.py
                          <-- project settings file
       settings.py
       spiders/
                          <-- a directory where you'll later put your sp
iders
            __init__.py
```

Now we can create the items.py file:

In [11]:

```
%writefile iftpapers/iftpapers/items.py

# -*- coding: utf-8 -*-
import scrapy

class IftpapersItem(scrapy.Item):
    num = scrapy.Field()
    aut = scrapy.Field()
    tit = scrapy.Field()
```

Overwriting iftpapers/iftpapers/items.py

The actual spider may look like this:

In [12]:

```
%writefile iftpapers/iftpapers/spiders/ift test.py
# -*- coding: utf-8 -*-
import scrapy
class IFTSpider(scrapy.Spider):
    name = "ift"
    allowed domains = ["http://www.ift.uni.wroc.pl"]
    start urls = ["http://www.ift.uni.wroc.pl/publications"]
    def parse(self, response):
        for sel in response.xpath('//table/tr'): #find the table
            num = sel.xpath('td[1]').extract() #first cell contains paper id
            aut = sel.xpath('td[2]/b/text()').extract() #second cell - authors
            tit = sel.xpath('td[3]/i/text()').extract() #third cell - title
            with open("papers.txt", "a") as f:
                                                        #not very efficient - fi
le closed after each found paper
                txt = aut[0] + u", " + tit[0].upper() + u"\n"
                f.write(txt.encode('utf8', 'replace'))
```

Overwriting iftpapers/iftpapers/spiders/ift test.py

Let us start the spider:

In [14]:

```
import os
os.chdir("iftpapers")
print(os.getcwd())
```

/home/szwabin/Dropbox/Zajecia/UnstructuredData/5-data_collection/ift
papers

In [15]:

!scrapy crawl ift

```
2017-09-29 12:09:53 [scrapy.utils.log] INFO: Scrapy 1.4.0 started (b
ot: iftpapers)
2017-09-29 12:09:53 [scrapy.utils.log] INFO: Overridden settings:
 {'BOT NAME': 'iftpapers', 'NEWSPIDER MODULE': 'iftpapers.spiders',
 'SPIDER_MODULES': ['iftpapers.spiders'], 'ROBOTSTXT_OBEY': True}
2017-09-29 12:09:53 [scrapy.middleware] INFO: Enabled extensions:
['scrapy.extensions.memusage.MemoryUsage',
 scrapy.extensions.corestats.CoreStats',
 'scrapy.extensions.logstats.LogStats',
 'scrapy.extensions.telnet.TelnetConsole']
2017-09-29 12:09:53 [scrapy.middleware] INFO: Enabled downloader mid
dlewares:
['scrapy.downloadermiddlewares.robotstxt.RobotsTxtMiddleware',
 'scrapy.downloadermiddlewares.httpauth.HttpAuthMiddleware',
 'scrapy.downloadermiddlewares.downloadtimeout.DownloadTimeoutMiddle
ware',
 scrapy.downloadermiddlewares.defaultheaders.DefaultHeadersMiddlewa
re',
 'scrapy.downloadermiddlewares.useragent.UserAgentMiddleware'.
 'scrapy.downloadermiddlewares.retry.RetryMiddleware',
 'scrapy.downloadermiddlewares.redirect.MetaRefreshMiddleware',
 'scrapy.downloadermiddlewares.httpcompression.HttpCompressionMiddle
ware',
 'scrapy.downloadermiddlewares.redirect.RedirectMiddleware',
 'scrapy.downloadermiddlewares.cookies.CookiesMiddleware',
 'scrapy.downloadermiddlewares.httpproxy.HttpProxyMiddleware',
 'scrapy.downloadermiddlewares.stats.DownloaderStats']
2017-09-29 12:09:53 [scrapy.middleware] INFO: Enabled spider middlew
['scrapy.spidermiddlewares.httperror.HttpErrorMiddleware',
 'scrapy.spidermiddlewares.offsite.OffsiteMiddleware',
 'scrapy.spidermiddlewares.referer.RefererMiddleware',
 'scrapy.spidermiddlewares.urllength.UrlLengthMiddleware',
 'scrapy.spidermiddlewares.depth.DepthMiddleware']
2017-09-29 12:09:53 [scrapy.middleware] INFO: Enabled item pipeline
[]
2017-09-29 12:09:53 [scrapy.core.engine] INFO: Spider opened
2017-09-29 12:09:53 [scrapy.extensions.logstats] INFO: Crawled 0 pag
es (at 0 pages/min), scraped 0 items (at 0 items/min)
2017-09-29 12:09:53 [scrapy.extensions.telnet] DEBUG: Telnet console
listening on 127.0.0.1:6023
2017-09-29 12:09:53 [scrapy.core.engine] DEBUG: Crawled (200) <GET h
ttp://www.ift.uni.wroc.pl/robots.txt> (referer: None)
2017-09-29 12:09:53 [scrapy.core.engine] DEBUG: Crawled (200) <GET h
ttp://www.ift.uni.wroc.pl/publications> (referer: None)
2017-09-29 12:09:53 [scrapy.core.engine] INFO: Closing spider (finis
hed)
2017-09-29 12:09:53 [scrapy.statscollectors] INFO: Dumping Scrapy st
{'downloader/request bytes': 456,
 'downloader/request count': 2,
 'downloader/request method count/GET': 2,
 'downloader/response_bytes': 34518,
 'downloader/response count': 2,
 'downloader/response_status_count/200': 2,
 'finish reason': 'finished',
 'finish time': datetime.datetime(2017, 9, 29, 10, 9, 53, 773515),
 'log count/DEBUG': 3,
 'log count/INFO': 7,
 'memusage/max': 52178944,
```

```
'memusage/startup': 52178944,
'response_received_count': 2,
'scheduler/dequeued': 1,
'scheduler/dequeued/memory': 1,
'scheduler/enqueued': 1,
'scheduler/enqueued/memory': 1,
'scheduler/enqueued/memory': 1,
'start_time': datetime.datetime(2017, 9, 29, 10, 9, 53, 358123)}
2017-09-29 12:09:53 [scrapy.core.engine] INFO: Spider closed (finish ed)
```

In [16]:

! cat papers.txt

- D. E. Alvarez-Castillo, M. A. R. Kaltenborn, D. Blaschke, EXCLUDED V OLUME EFFECTS IN THE HYBRID STAR EOS
- A. Ayriyan, D. E. Alvarez-Castillo, D. Blaschke, H. Grigorian, MASS-RADIUS CONSTRAINTS FOR THE NEUTRON STAR EOS BAYESIAN ANALYSIS
- A. Ayriyan, D. Blaschke, R. Lastowiecki, PHASE DIAGRAM OF THE THREE-FLAVOR COLOR SUPERCONDUCTING PNJL MODEL
- N.-U. Bastian, D. Blaschke, TOWARDS A NEW QUARK-NUCLEAR MATTER EOS F OR APPLICATIONS IN ASTROPHYSICS AND HEAVY-ION COLLISIONS
- D. Blaschke, D. E. Alvarez-Castillo, HIGH-MASS TWINS & RESOLUTION OF THE RECONFINEMENT, MASQUERADE AND HYPERON PUZZLES OF COMPACT STAR IN TERIORS
- D. Blaschke, N. T. Gevorgyan, A. D. Panferov, S. A. Smolyansky, SCHW INGER EFFECT AT MODERN LASER FACILITIES
- A. Borowiec, J. Lukierski, V. N. Tolstoy, REAL AND PSEUDOREAL FORMS OF D=4 COMPLEX EUCLIDEAN (SUPER)ALGEBRAS AND SUPER-POINCARÉ/ SUPER-EUCLIDEAN R-MATRICES
- A. Borowiec, J. Lukierski, V. N. Tolstoy, QUANTUM DEFORMATIONS OF D = 4 EUCLIDEAN, LORENTZ, KLEINIAN AND QUATERNIONIC 0*(4) SYMMETRIES I N UNIFIED 0(4;C) SETTING
- A. Borowiec, A. Stachowski, M. Szydłowski, A. Wojnar, INFLATIONARY C OSMOLOGY WITH CHAPLYGIN GAS IN PALATINI FORMALISM
- A. Dubinin, D. Blaschke, A. Radzhabov, PION AND KAON IN THE BETH-UHL ENBECK APPROACH
- L. Juchnowski, D. Blaschke, T. Fischer, S. A. Smolyansky, NONEQUILIB RIUM MESON PRODUCTION IN STRONG FIELDS
- J. Jędrzejewski, QUANTUM CRITICALITY IN CONDENSED MATTER. PHENOMENA, MATERIALS AND IDEAS IN THEORY AND EXPERIMENT
- A. Jędrzejewski, K. Sznajd-Weron, J. Szwabiński, MAPPING THE QQ-VOTE R MODEL: FROM A SINGLE CHAIN TO COMPLEX NETWORKS
- M. Matyka, J. Gołembiewski, Z. Koza, POWER-EXPONENTIAL VELOCITY DIST RIBUTIONS IN DISORDERED POROUS MEDIA
- K. Redlich, H. Satz, THE LEGACY OF ROLF HAGEDORN: STATISTICAL BOOTST RAP AND ULTIMATE TEMPERATURE
- D. E. Alvarez-Castillo, M. A. R. Kaltenborn, D. Blaschke, EXCLUDED V OLUME EFFECTS IN THE HYBRID STAR EOS
- A. Ayriyan, D. E. Alvarez-Castillo, D. Blaschke, H. Grigorian, MASS-RADIUS CONSTRAINTS FOR THE NEUTRON STAR EOS BAYESIAN ANALYSIS
- A. Ayriyan, D. Blaschke, R. Lastowiecki, PHASE DIAGRAM OF THE THREE-FLAVOR COLOR SUPERCONDUCTING PNJL MODEL
- N.-U. Bastian, D. Blaschke, TOWARDS A NEW QUARK-NUCLEAR MATTER EOS F OR APPLICATIONS IN ASTROPHYSICS AND HEAVY-ION COLLISIONS
- D. Blaschke, D. E. Alvarez-Castillo, HIGH-MASS TWINS & RESOLUTION OF THE RECONFINEMENT, MASQUERADE AND HYPERON PUZZLES OF COMPACT STAR IN TERIORS
- D. Blaschke, N. T. Gevorgyan, A. D. Panferov, S. A. Smolyansky, SCHW INGER EFFECT AT MODERN LASER FACILITIES
- A. Borowiec, J. Lukierski, V. N. Tolstoy, REAL AND PSEUDOREAL FORMS OF D=4 COMPLEX EUCLIDEAN (SUPER)ALGEBRAS AND SUPER-POINCARÉ/ SUPER-EUCLIDEAN R-MATRICES
- A. Borowiec, J. Lukierski, V. N. Tolstoy, QUANTUM DEFORMATIONS OF D = 4 EUCLIDEAN, LORENTZ, KLEINIAN AND QUATERNIONIC 0*(4) SYMMETRIES I N UNIFIED 0(4;C) SETTING
- A. Borowiec, A. Stachowski, M. Szydłowski, A. Wojnar, INFLATIONARY C OSMOLOGY WITH CHAPLYGIN GAS IN PALATINI FORMALISM
- A. Dubinin, D. Blaschke, A. Radzhabov, PION AND KAON IN THE BETH-UHL ENBECK APPROACH
- H. Grigorian, D. N. Voskresensky, D. Blaschke, INFLUENCE OF THE STIF FNESS OF THE EQUATION OF STATE AND IN-MEDIUM EFFECTS ON THE COOLING OF COMPACT STARS
- L. Jakóbczyk, A. Frydryszak, P. Ługiewicz, QUTRIT GEOMETRIC DISCORD
- L. Juchnowski, D. Blaschke, T. Fischer, S. A. Smolyansky, NONEQUILIB

RIUM MESON PRODUCTION IN STRONG FIELDS

- J. Jędrzejewski, QUANTUM CRITICALITY IN CONDENSED MATTER. PHENOMENA, MATERIALS AND IDEAS IN THEORY AND EXPERIMENT
- A. Jędrzejewski, K. Sznajd-Weron, J. Szwabiński, MAPPING THE QQ-VOTE R MODEL: FROM A SINGLE CHAIN TO COMPLEX NETWORKS
- M. Matyka, J. Gołembiewski, Z. Koza, POWER-EXPONENTIAL VELOCITY DIST RIBUTIONS IN DISORDERED POROUS MEDIA
- A. D. Panferov, S. A. Smolyansky, A. Otto ,B. Kampfer, D. B. Blaschk e, Ł Juchnowski, ASSISTED DYNAMICAL SCHWINGER EFFECT: PAIR PRODUCTIO N IN A PULSED BIFREQUENT FIELD
- K. Redlich, H. Satz, THE LEGACY OF ROLF HAGEDORN: STATISTICAL BOOTST RAP AND ULTIMATE TEMPERATURE

Using of existing APIs

Twitter and tweepy

Real-time monitoring of Twitter stream

In []:

```
from tweepy.streaming import StreamListener
from tweepy import OAuthHandler
from tweepy import Stream
import time
import pprint
# Twitter account data
consumer key="XXXX"
consumer_secret="XXXX"
# Generate application data within your Twitter account. Then:
access token="XXXXX"
access token secret="XXXXX"
class StdOutListener(StreamListener):
    """ A listener handles tweets that are received from the stream.
    This is a basic listener that just prints received tweets to stdout.
    def on data(self, data):
        try:
            with open("twitter.json", 'a') as f:
                f.write(data)
                print(data)
                return True
        except BaseException as e:
            print("Error on data: %s" % str(e))
            time.sleep(5)
        return True
    def on error(self, status):
        print(status)
if name == ' main ':
    \overline{l} = S\overline{td}0utListener()
    auth = OAuthHandler(consumer key, consumer secret)
    auth.set_access_token(access_token, access token secret)
    stream = Stream(auth, l)
    stream.filter(track=['python'])
```

{"created at": "Wed Oct 11 12:23:28 +0000 2017", "id": 9180896675845652 50,"id str":"918089667584565250","text":"@python \u30b9\u30c8\u30ec \u30c3\u30ab\u30fc\u53cd\u5fdc\n NowTime: 2017\/10\/11 21:23:28.4773 ownload\/android\" rel=\"nofollow\"\u003eTwitter for Android\u003c\/ a\u003e", "truncated": false, "in reply to status id": null, "in reply to status_id_str":null,"in_reply_to_user_id":765482,"in reply to user id_str":"765482","in_reply_to_screen_name":"Python","user":{"id":244 9488770,"id str":"2449488770","name":"Ferix","screen_name":"nfz6","l ocation":null, "url": "http:\/\/twpf.jp\/nfz6", "description": "Razer\/P UBG\/R6S\/Ryzen \u9593\u9055\u3048\u305f\u9053\u3092\u6b69\u3093\u30 67\u308b\u4eba\u3067\u3059\u3002", "translator type": "regular", "prote cted":false, "verified":false, "followers count":64447, "friends coun t":17, "listed count":0, "favourites count":7682, "statuses count":1112 07, "created at": "Thu Apr 17 10:35:46 +0000 2014", "utc_offset":3240 0, "time_zone": "Seoul", "geo_enabled": false, "lang": "ja", "contributors_ enabled":false,"is translator":false,"profile background color":"000 000", "profile background image url": "http:\/\/abs.twimg.com\/images \/themes\/theme1\/bg.png","profile_background_image_url_https":"http s:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png","profile_backgr ound_tile":false,"profile_link_color":"FF0080","profile sidebar bord er_color":"000000","profile_sidebar_fill_color":"000000","profile_te xt_color":"000000","profile_use_background_image":false,"profile_ima ge url":"http:\/\/pbs.twimg.com\/profile images\/917741971917094913 \/Tna0Rd5R normal.jpg", "profile image url https": "https:\/\/pbs.twim $q.com\profile images\prom{917741971917094913\prom{5Rd5R normal.jpg","pro$ file banner url": "https:\/\/pbs.twimg.com\/profile banners\/24494887 70\/1507702130", "default_profile": false, "default_profile_image": false e, "following": null, "follow request sent": null, "notifications": nul l}, "geo":null, "coordinates":null, "place":null, "contributors":null, "i s quote status":false, "quote count":0, "reply count":0, "retweet coun t":0, "favorite_count":0, "entities":{ "hashtags":[], "urls":[], "user me ntions":[{"screen_name":"Python","name":"Python","id":765482,"id st r":"765482","indices":[0,7]}],"symbols":[]},"favorited":false,"retwe eted":false,"filter level":"low","lang":"ja","timestamp ms":"1507724 608690"}

{"created at": "Wed Oct 11 12:23:29 +0000 2017", "id": 9180896726009528 33, "id str": "918089672600952833", "text": "@ashemedai @dstufft @ncoghl an dev @DRMacIver @hynek @tintvrtkovic i.e. a python app shipped wit h PEX that uses twis\u2026 https:\/\/t.co\/fEuXE6pbUy","display text range":[66,140], "source": "\u003ca href=\"http:\/\/twitter.com\" rel =\"nofollow\"\u003eTwitter Web Client\u003c\/a\u003e","truncated":tr ue, "in reply to status id":918089500235939845, "in reply to status id _str": "918089500235939845", "in_reply_to_user_id": 1143919728, "in_repl y_to_user_id_str":"1143919728","in_reply_to_screen_name":"hawkieow l","user":{"id":1143919728,"id_str":"1143919728","name":"owl sez yes \ud83c\udff3\ufe0f\u200d\ud83c\udf08", "screen_name": "hawkieowl", "loc ation": "melburnia, 'straya", "url": "http:\/\/atleastfornow.net", "desc ription": "Miss Amber Brown, @twistedmatrix's owl-themed release mana gement simulator. Keynote speaker, sometimes. Opinions mine, even th e bad ones. \ud83e\udd89\ud83d\udc26 [she\/her]","translator_typ e": "none", "protected": false, "verified": false, "followers_count": 92 9, "friends_count":308, "listed_count":61, "favourites_count":4807, "sta tuses count":3781, "created at": "Sun Feb 03 01:57:19 +0000 2013", "utc _offset":39600,"time_zone":"Melbourne","geo_enabled":false,"lang":"e n", "contributors enabled":false, "is translator":false, "profile backg round_color":"000000","profile_background_image_url":"http:\/\/pbs.t wimg.com\/profile_background_images\/452780958622560257\/b3lFr0jZ.pn g","profile_background_image_url_https":"https:\/\/pbs.twimg.com\/pr ofile background images\/452780958622560257\/b3lFr0jZ.png", "profile

background tile":true, "profile link color": "FF0000", "profile sidebar _border_color":"000000","profile_sidebar_fill_color":"E6F6F9","profi le text color": "333333", "profile use background image": false, "profil e_image_url":"http:\/\/pbs.twimg.com\/profile_images\/91075056250897 2032\/0VQIs8Kz normal.jpg", "profile image url https": "https:\/\/pbs. twimg.com\/profile_images\/910750562508972032\/0VQIs8Kz_normal.jp g", "profile banner url": "https:\/\/pbs.twimg.com\/profile banners\/1 143919728\/1424238152", "default_profile": false, "default_profile_imag e":false, "following":null, "follow request sent":null, "notification s":null}, "geo":null, "coordinates":null, "place":null, "contributors":n ull, "is_quote_status":false, "extended_tweet":{"full_text":"@ashemeda i @dstufft @ncoghlan dev @DRMacIver @hynek @tintvrtkovic i.e. a pyth on app shipped with PEX that uses twisted.web will still ship the re st of twisted, and such, and its hard to strip it out", "display text range":[66,198], "entities":{ "hashtags":[], "urls":[], "user mention s":[{"screen name":"ashemedai", "name": "Jeroen Ruigrok", "id":19651805 2,"id str":"196518052","indices":[0,10]},{"screen name":"dstufft","n ame": "Donald Stufft", "id": 250397632, "id_str": "250397632", "indices": [11,19]},{"screen_name":"ncoghlan_dev","name":"Nick Coghlan","id":26 1665936, "id_str": "261665936", "indices": [20,33]}, { "screen_name": "DRMa cIver", "name": "DRMacIver's Monster", "id": 14368342, "id str": "1436834 2","indices":[34,44]},{"screen_name":"hynek","name":"Hynek Schlawack $\del{locality} $$\del{locality} \del{locality} $$\del{locality} $$\del{l$ s":[45,51]},{"screen name":"tintvrtkovic","name":"Tin Tvrtkovi\u010 7","id":1380016717,"id str":"1380016717","indices":[52,65]}],"symbol s":[]}}, "quote_count":0, "reply_count":0, "retweet_count":0, "favorite_ count":0,"entities":{"hashtags":[],"urls":[{"url":"https:\/\/t.co\/f EuXE6pbUy","expanded url":"https:\/\/twitter.com\/i\/web\/status\/91 8089672600952833", "display url": "twitter.com\/i\/web\/status\/9\u202 6", "indices":[117,140]}], "user mentions":[{"screen name":"ashemeda i", "name": "Jeroen Ruigrok", "id": 196518052, "id str": "196518052", "indi ces":[0,10]},{"screen name":"dstufft","name":"Donald Stufft","id":25 0397632, "id str": "250397632", "indices": [11,19]}, {"screen name": "ncog hlan dev", "name": "Nick Coghlan", "id": 261665936, "id str": "26166593 6", "indices": [20,33]}, {"screen name": "DRMacIver", "name": "DRMacIver's Monster","id":14368342,"id_str":"14368342","indices":[34,44]},{"scre en_name":"hynek","name":"Hynek Schlawack \ud83c\uddff\ud83c\udde 6","id":14914177,"id_str":"14914177","indices":[45,51]},{"screen_nam e":"tintvrtkovic", "name": "Tin Tvrtkovi\u0107", "id":1380016717, "id st r":"1380016717","indices":[52,65]}],"symbols":[]},"favorited":fals e, "retweeted": false, "filter level": "low", "lang": "en", "timestamp m s":"1507724609886"}

{"created at":"Wed Oct 11 12:23:46 +0000 2017","id":9180897428262461 44, "id str": "918089742826246144", "text": "Hire a Python Developer by \u00a0mp2consult https:\/\/t.co\/amg0zD4mkL","source":"\u003ca href= \"http:\/\/publicize.wp.com\/\" rel=\"nofollow\"\u003eWordPress.com \u003c\/a\u003e","truncated":false,"in_reply_to_status_id":null,"in_ reply_to_status_id_str":null, "in_reply_to_user_id":null, "in_reply_to _user_id_str":null,"in_reply_to_screen_name":null,"user":{"id":45549 24208, "id_str": "4554924208", "name": "Kenith Blaque", "screen_name": "ke nithblaque","location":null,"url":null,"description":null, "translato r_type":"none","protected":false,"verified":false,"followers count": 39, "friends_count":63, "listed_count":0, "favourites_count":0, "statuse s_count":5412,"created_at":"Mon Dec 21 07:57:30 +0000 2015","utc_off set":null,"time_zone":null,"geo_enabled":false,"lang":"en","contribu tors_enabled":false,"is_translator":false,"profile_background_colo r": "F5F8FA", "profile_background_image_url": "", "profile_background_im age_url_https":"","profile_background_tile":false,"profile_link_colo r":"1DA1F2", "profile_sidebar_border_color":"C0DEED", "profile_sidebar _fill_color": "DDEEF6", "profile_text_color": "333333", "profile_use_bac kground_image":true,"profile_image_url":"http:\/\/pbs.twimg.com\/profile_images\/678850288296599552\/zNeNITNs_normal.png","profile_image _url_https":"https:\/\/pbs.twimg.com\/profile_images\/678850288296599552\/zNeNITNs_normal.png","default_profile":true,"default_profile_image":false,"following":null,"follow_request_sent":null,"notifications":null,"geo":null,"coordinates":null,"place":null,"contributors":null,"is_quote_status":false,"quote_count":0,"reply_count":0,"retweet_count":0,"favorite_count":0,"entities":{"hashtags":[],"urls":[{"urls":[{"urls":[{"urls":[*"urls:\/\/t.co\/amg0zD4mkL","expanded_url":"http:\/\/megasmartlife.com\/2017\/10\/11\/hire-a-python-developer-by-mp2consult\/","display_url":"megasmartlife.com\/2017\/10\/11\/hir\u2026","indices":[38,61]}],"user_mentions":[],"symbols":[]},"favorited":false,"retweeted":false,"possibly_sensitive":false,"filter_level":"low","lang":"en","timestamp_ms":"1507724626629"}

{"created at": "Wed Oct 11 12:23:47 +0000 2017", "id": 9180897480607375 41,"id str":"918089748060737541","text":"RT @rochacbruno: Preciso de ajuda para manter a lib #python @pagseguro algumas issues abertas e o @pagseguro n\u00e3o d\u00e1 nenhuma ajuda https:\/\/t\u2026","sou rce":"\u003ca href=\"http:\/\/twitter.com\/download\/iphone\" rel= \"nofollow\"\u003eTwitter for iPhone\u003c\/a\u003e","truncated":fal se, "in_reply_to_status_id":null, "in_reply_to_status_id_str":null, "in reply to user id":null,"in reply to user id str":null,"in reply to screen name":null, "user":{"id":32444222, "id str":"32444222", "nam e":"Igor Souza", "screen name": "igordcsouza", "location": "Rio de Janei ro, Brasil", "url": null, "description": "Dev residente vivendo entre Sy sadmins! - @ Stone Pagamentos | Stone CO", "translator type": "non e", "protected": false, "verified": false, "followers count": 172, "friends _count":90,"listed_count":8,"favourites_count":259,"statuses_count": 2543,"created_at":"Fri Apr 17 15:21:22 +0000 2009","utc_offset":-108 00, "time zone": "Santiago", "geo enabled": false, "lang": "pt", "contribut ors_enabled":false,"is_translator":false,"profile_background_colo r":"1A1B1F","profile_background_image_url":"http:\/\/abs.twimg.com\/ images\/themes\/theme9\/bg.gif","profile background image url http s":"https:\/\/abs.twimg.com\/images\/themes\/theme9\/bg.gif","profil e_background_tile":false,"profile_link_color":"2FC2EF","profile_side bar_border_color":"181A1E","profile_sidebar_fill_color":"252429","pr
ofile_text_color":"6666666","profile_use_background_image":true,"prof ile image url":"http:\/\/pbs.twimg.com\/profile images\/917507145138 130944\/QAKPNDft normal.jpg","profile image url https":"https:\/\/pb s.twimg.com\/profile images\/917507145138130944\/QAKPNDft normal.jp g","profile banner url":"https:\/\/pbs.twimg.com\/profile_banners\/3 2444222\/1475186525", "default_profile": false, "default_profile_imag e":false,"following":null,"follow_request_sent":null,"notification s":null}, "geo":null, "coordinates":null, "place":null, "contributors":n ull, "retweeted status": {"created at": "Wed Oct 11 00:31:44 +0000 201 7","id":917910554899111937,"id_str":"917910554899111937","text":"Pre ciso de ajuda para manter a lib #python @pagseguro algumas issues ab ertas e o @pagseguro n\u00e3o d\u00e1 nenhuma ajuda https:\/\/t.co\/ j6B1mMb7xR", "source": "\u003ca href=\"http:\/\/twitter.com\" rel=\"no follow\"\u003eTwitter Web Client\u003c\/a\u003e", "truncated":fals e, "in_reply_to_status_id":null, "in_reply_to_status_id_str":null, "in_ reply to user id":null,"in reply to user id str":null,"in reply to s creen_name":null,"user":{"id":14347532,"id_str":"14347532","name":"B runo Rocha \ud83d\udc0d\u2699\ufe0f\ud83c\udf31\ud83d\udeb4","screen _name":"rochacbruno","location":"S\u00e3o Paulo, Brasil","url":"http s:\/\/about.me\/rochacbruno", "description": "Herbivore, animalist, Mi nister of the Church of the Flying Spaghetti Monster, Biker, Pythoni sta and Programmer at Red Hat in his spare time!","translator_typ e": "none", "protected": false, "verified": false, "followers count": 320 0, "friends_count": 1377, "listed_count": 212, "favourites_count": 1026

4, "statuses_count": 26529, "created_at": "Thu Apr 10 01:52:50 +0000 200 8","utc offset":-10800,"time_zone":"America\/Sao_Paulo","geo_enable d":true, "lang": "pt", "contributors enabled": false, "is translator": fal se, "profile_background_color": "ACDED6", "profile_background_image_ur l":"http:\/\/pbs.twimg.com\/profile background images\/4616236244372 27520\/2M8jQLBk.jpeg", "profile background image url https": "https:\/ \/pbs.twimg.com\/profile background images\/461623624437227520\/2M8j QLBk.jpeg", "profile_background_tile":true, "profile_link_color": "ABB8 C2", "profile_sidebar_border_color": "FFFFFF", "profile_sidebar_fill_color": "F6F6F6", "profile_text_color": "333333", "profile_use_background_ image":true,"profile image url":"http:\/\/pbs.twimg.com\/profile ima ges\/831153043697602560\/jm3mCUps normal.jpg", "profile image url htt ps":"https:\/\/pbs.twimg.com\/profile images\/831153043697602560\/jm 3mCUps_normal.jpg","profile_banner_url":"https:\/\/pbs.twimg.com\/pr ofile banners\/14347532\/1501092729", "default profile":false, "defaul t profile image":false, "following":null, "follow request sent":nul l, "notifications":null}, "geo":null, "coordinates":null, "place":nul l,"contributors":null,"is_quote_status":false,"quote_count":0,"reply count":6,"retweet count":6,"favorite count":16,"entities":{"hashtag s":[{"text":"python","indices":[35,42]}],"urls":[{"url":"https:\/\/ t.co\/j6B1mMb7xR","expanded_url":"https:\/\/github.com\/rochacbruno $\verb|\python-pagseguro|/issues", "display_url": "github.com|/rochacbruno|/|$ py\u2026","indices":[113,136]}],"user mentions":[{"screen name":"pag seguro", "name": "PagSeguro UOL", "id": 15087174, "id str": "15087174", "in dices":[43,53]},{"screen name":"pagseguro","name":"PagSeguro UOL","i d":15087174,"id_str":"15087174","indices":[81,91]}],"symbols":[]},"f avorited":false, "retweeted":false, "possibly sensitive":false, "filter level":"low","lang":"pt"},"is quote status":false,"quote count": 0, "reply_count":0, "retweet_count":0, "favorite_count":0, "entities": {"hashtags":[{"text":"python","indices":[52,59]}],"urls":[],"user me ntions":[{"screen name":"rochacbruno","name":"Bruno Rocha \ud83d\udc 0d\u2699\ufe0f\ud83c\udf31\ud83d\udeb4","id":14347532,"id str":"1434 7532", "indices": [3,15]}, {"screen_name": "pagseguro", "name": "PagSeguro UOL", "id":15087174, "id_str": "15087174", "indices": [60,70]}, {"screen_n ame":"pagseguro","name":"PagSeguro UOL","id":15087174,"id str":"1508 7174", "indices": [98,108]}], "symbols": []}, "favorited": false, "retweete d":false,"filter_level":"low","lang":"pt","timestamp_ms":"1507724627 877"}

{"created at": "Wed Oct 11 12:23:56 +0000 2017", "id": 9180897840394936 33, "id str": "918089784039493633", "text": "Looking for Python projects to assist with. https:\/\/t.co\/Xu0uuKE8YV \n\nHello I am a self tau ght Python programmer. I am looking for any hel\u2026", "source": "\u0 03ca href=\"https:\/\/ifttt.com\" rel=\"nofollow\"\u003eIFTTT\u003c \/a\u003e", "truncated": false, "in reply to status id": null, "in reply to_status_id_str":null,"in_reply_to_user_id":null,"in_reply_to_user_ id_str":null,"in_reply_to_screen_name":null,"user":{"id":330661011
7,"id_str":"3306610117","name":"Programming Gurgaon","screen_nam e":"programmingncr","location":"Gurgaon, Haryana, India","url":nul l,"description":"Learn Python Programming in Delhi NCR #Python Class es in Gurgaon (Age: 9 to 90 Years) #PythonGurgaon #Programming #Data #Coding #R #Gurgaon #DelhiNCR #Analytics", "translator type": "non e", "protected": false, "verified": false, "followers count": 1273, "friend s_count":98,"listed_count":546,"favourites_count":4,"statuses_coun t":258812,"created_at":"Wed Aug 05 03:47:32 +0000 2015","utc_offse t":-25200, "time_zone": "Pacific Time (US & Canada)", "geo_enabled":fal se, "lang": "en", "contributors enabled": false, "is translator": false, "p rofile_background_color":"FFF04D","profile_background_image_url":"ht tp:\/\/abs.twimg.com\/images\/themes\/theme19\/bg.gif","profile_back ground image url https": "https:\/\/abs.twimg.com\/images\/themes\/th eme19\/bg.gif", "profile_background_tile": false, "profile_link_colo

r":"3B94D9","profile_sidebar_border_color":"000000","profile_sidebar _fill_color":"000000","profile_text_color":"000000","profile_use_bac kground image":true, "profile image url": "http:\/\/pbs.twimg.com\/pro file images\/662602436222193664\/qKuvsfnG normal.jpg","profile image url https":"https:\/\/pbs.twimg.com\/profile images\/66260243622219 3664\/qKuvsfnG_normal.jpg","default_profile":false,"default_profile_ image":false, "following":null, "follow_request_sent":null, "notificati ons":null}, "geo":null, "coordinates":null, "place":null, "contributor s":null, "is quote status":false, "quote count":0, "reply count":0, "ret weet count":0,"favorite count":0,"entities":{"hashtags":[],"urls": [{"url":"https:\/\/t.co\/Xu0uuKE8YV","expanded_url":"https:\/\/www.r eddit.com $\/r\/$ learnprogramming $\/$ comments $\/$ 75ooua $\/$ looking for python _projects_to_assist_with\/","display_url":"reddit.com\/r\/learnprogr am\u2026","indices":[44,67]}],"user_mentions":[],"symbols":[]},"favo rited":false, "retweeted":false, "possibly sensitive":false, "filter le vel":"low","lang":"en","timestamp ms":"1507724636455"}

{"created at": "Wed Oct 11 12:23:58 +0000 2017", "id": 9180897930193264 64,"id str":"918089793019326464","text":"Super Oct 17 https:\/\/t.co SNES #Python #blog #geek", "source": "\u003ca href=\"https:\/\/www.gam er-geek-news.com\/\" rel=\"nofollow\"\u003eRssToTwitterPy3\u003c\/a \u003e","truncated":false,"in reply to status id":null,"in reply to status_id_str":null,"in_reply_to_user_id":null,"in_reply_to_user_id_ str":null, "in reply to screen name":null, "user":{"id":316331833, "id str":"316331833","name":"Gamer Geek","screen_name":"GamerGeekNew s", "location":null, "url": "https:\/\/www.gamer-geek-news.com", "descri ption":"Data guru by day, geek tech gamer junky by night.\n\nHere's a feed of my interests, real person behind the posts", "translator t ype": "none", "protected": false, "verified": false, "followers count": 213 0, "friends count": 2478, "listed count": 123, "favourites count": 1409, "s tatuses_count":228004,"created_at":"Mon Jun 13 08:53:46 +0000 201 1", "utc offset": 46800, "time zone": "Auckland", "geo enabled": false, "la ng": "en", "contributors enabled": false, "is translator": false, "profile background color":"131516","profile background image url":"http:\/ \/abs.twimg.com\/images\/themes\/theme14\/bg.gif","profile backgroun d_image_url_https":"https:\/\/abs.twimg.com\/images\/themes\/theme14 \/bg.gif", "profile background tile":true, "profile link color": "00999 9", "profile sidebar border color": "EEEEEEE", "profile sidebar fill col or": "EFEFEF", "profile text color": "333333", "profile use background i mage":true,"profile image url":"http:\/\/pbs.twimg.com\/profile imag es\/1404245782\/igeek_normal.jpg","profile_image_url_https":"http s:\/\/pbs.twimg.com\/profile images\/1404245782\/igeek normal.jp g","profile_banner_url":"https:\/\/pbs.twimg.com\/profile banners\/3 16331833\/1431495420", "default profile": false, "default profile imag e":false, "following":null, "follow_request_sent":null, "notification s":null, "geo":null, "coordinates":null, "place":null, "contributors":n ull, "is_quote_status":false, "quote_count":0, "reply_count":0, "retweet count":0,"favorite count":0,"entities":{"hashtags":[{"text":"Playst ation", "indices": [37,49]}, {"text": "DataScience", "indices": [50,62]}, {"text":"borgbackup","indices":[63,74]},{"text":"NeonChrome","indice s":[75,86]},{"text":"MiniSNES","indices":[87,96]},{"text":"Pytho ,"indices":[97,104]},{"text":"blog","indices":[105,110]},{"tex t":"geek","indices":[111,116]}],"urls":[{"url":"https:\/\/t.co\/uC9I YZhtVZ", "expanded_url": "http:\/\/www.gamer-geek-news.com\/super-oct-17\/", "display_url": "gamer-geek-news.com\/super-oct-17\/", "indices": [13,36]}], "user mentions":[], "symbols":[]}, "favorited":false, "retwee ted":false, "possibly sensitive":false, "filter level": "low", "lang": "e n","timestamp_ms":"1507724638596"}

{"created at": "Wed Oct 11 12:23:59 +0000 2017", "id": 9180897975577845

76,"id_str":"918089797557784576","text":"Hire a Python Developer by \u00a0mp2consult https:\/\/t.co\/MYu4hVxbE9","source":"\u003ca href= \"http:\/\/publicize.wp.com\/\" rel=\"nofollow\"\u003eWordPress.com \u003c\/a\u003e","truncated":false,"in_reply_to_status_id":null,"in_ reply to status id str":null, "in reply to user id":null, "in reply to _user_id_str":null,"in_reply_to_screen_name":null,"user":{"id":83391 1267899367424, "id_str": "833911267899367424", "name": "Justine Cormie r", "screen_name": "dexterity247onl", "location": null, "url": null, "descr iption": "I am Justine Cormier Graphics designer with years of experi ence in my field which gives me the privilege of a better edge into knowing what the market perfectly","translator_type":"none","protec ted":false, "verified":false, "followers count":33, "friends count":2 1, "listed count":0, "favourites count":0, "statuses count":5466, "creat ed at": "Tue Feb 21 05:28:34 +0000 2017", "utc offset": null, "time zon e":null, "geo enabled":false, "lang": "en", "contributors enabled":fals e, "is translator": false, "profile background color": "F5F8FA", "profile background image_url":"","profile_background_image_url_https":"","p rofile_background_tile":false,"profile_link_color":"1DATF2","profile _sidebar_border_color":"CODEED","profile_sidebar_fill_color":"DDEEF 6","profile_text_color":"333333","profile_use_background_image":tru e, "profile image url": "http:\/\/pbs.twimg.com\/profile images\/83391 3679372500992\/WMUaNe7v_normal.jpg","profile_image_url_https":"http $s:\/\pbs.twimg.com\profile images\/833913679372500992\/WMUaNe7v no$ rmal.jpg", "default profile":true, "default profile image":false, "foll owing":null, "follow request sent":null, "notifications":null}, "geo":n ull, "coordinates":null, "place":null, "contributors":null, "is_quote_st atus":false, "quote count":0, "reply count":0, "retweet count":0, "favor ite count":0,"entities":{"hashtags":[],"urls":[{"url":"https:\/\/t.c o\/MYu4hVxbE9", "expanded url": "http:\/\/megasmartlife.com\/2017\/10 \/11\/hire-a-python-developer-by-mp2consult\/","display url":"megasm ions":[], "symbols":[]}, "favorited":false, "retweeted":false, "possibly sensitive":false,"filter level":"low","lang":"en","timestamp ms":"1 507724639678"}

{"created at": "Wed Oct 11 12:24:08 +0000 2017", "id": 9180898361831219 20,"id str":"918089836183121920","text":"3 Python web scrapers and c rawlers https:\/\/t.co\/KMdUBYWMDp via @opensourceway by @jehb #webs craping #python","source":"\u003ca href=\"http:\/\/twitter.com\" rel =\"nofollow\"\u003eTwitter Web Client\u003c\/a\u003e","truncated":fa lse,"in_reply_to_status_id":null,"in_reply_to_status_id_str":null,"i n reply to user id":null, "in reply to user id str":null, "in reply to screen name":null,"user":{"id":126610866,"id str":"126610866","nam e":"Ra\u00fal Quijada Ferrero","screen_name":"Raulq_f","location":nu ll, "url":null, "description":null, "translator type": "none", "protecte d":false, "verified":false, "followers count":95, "friends count":26 6, "listed count": 24, "favourites count": 45, "statuses count": 234, "crea ted_at":"Fri Mar 26 12:26:49 +0000 2010","utc_offset":null,"time_zon e":null, "geo enabled":false, "lang": "en", "contributors enabled":fals e, "is_translator": false, "profile_background_color": "CODEED", "profile background image url":"http:\/\/abs.twimg.com\/images\/themes\/them e1\/bg.png","profile_background_image_url_https":"https:\/\/abs.twim g.com\/images\/themes\/theme1\/bg.png","profile_background_tile":fal se, "profile_link_color": "1DA1F2", "profile_sidebar_border_color": "COD ,"profile_sidebar_fill_color":"DDEEF6","profile_text_color":"333 333", "profile_use_background_image":true, "profile_image_url": "htt p:\/\/pbs.twimg.com\/profile_images\/522755201665490944\/3qHyKQMy_no rmal.jpeg","profile_image_url_https":"https:\/\/pbs.twimg.com\/profi le_images\/522755201665490944\/3qHyKQMy_normal.jpeg","profile_banner url": "https: $\/\pbs.twimg.com\profile banners\/126610866\/15053697$ 11", "default_profile":true, "default_profile_image":false, "followin

g":null,"follow_request_sent":null,"notifications":null},"geo":nul
l,"coordinates":null,"place":null,"contributors":null,"is_quote_stat
us":false,"quote_count":0,"reply_count":0,"retweet_count":0,"favorit
e_count":0,"entities":{"hashtags":[{"text":"webscraping","indices":
[87,99]},{"text":"python","indices":[100,107]}],"urls":[{"url":"http
s:\/\/t.co\/KMdUBYWMDp","expanded_url":"http:\/\/red.ht\/2fr9lyV","d
isplay_url":"red.ht\/2fr9lyV","indices":[35,58]}],"user_mentions":
[{"screen_name":"opensourceway","name":"Open Source Way","id":805892
55,"id_str":"80589255","indices":[63,77]},{"screen_name":"jehb","nam
e":"Jason Baker\u200b","id":10191332,"id_str":"10191332","indices":
[81,86]}],"symbols":[]},"favorited":false,"retweeted":false,"possibl
y_sensitive":false,"filter_level":"low","lang":"en","timestamp_m
s":"1507724648887"}

{"created at": "Wed Oct 11 12:24:12 +0000 2017", "id": 9180898497222983 69, "id str": "918089849722298369", "text": "RT @taylorndean: what would u rather have, 10,000 for a Louis Vuitton dead python purse or 2,000 for my living python Louis https:\/\/t.co\/CL\u2026","source":"\u003 ca href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow \"\u003eTwitter for iPhone\u003c\/a\u003e","truncated":false,"in rep ly to status id":null, "in reply to status id str":null, "in reply to user_id":null,"in_reply_to_user_id_str":null,"in_reply_to_screen_nam e":null,"user":{"id":479029743,"id_str":"479029743","name":"shannon healy", "screen name": "shannon healy1", "location": "Cincinnati, O H","url":"http:\/\/vsco.co\/shannonnhealy","description":"university of cincinnati | UCMC", "translator_type": "none", "protected": false, "ve rified":false, "followers count":847, "friends count":280, "listed coun t":2, "favourites count":26621, "statuses count":24386, "created at": "M on Jan 30 23:00:15 +0000 2012", "utc_offset":-10800, "time_zone":"Atla ntic Time (Canada)","geo_enabled":true,"lang":"en","contributors_ena bled":false, "is translator":false, "profile background color": "ACDED 6", "profile background image url": "http:\/\/pbs.twimg.com\/profile b ackground images\/673342458\/3ce387b304d6beecd9f7d4b52c2541b8.pn g", "profile background image url https": "https:\/\/pbs.twimg.com\/pr ofile background images\/673342458\/3ce387b304d6beecd9f7d4b52c2541b 8.png", "profile_background_tile":true, "profile_link_color": "03854 3", "profile_sidebar_border_color": "FFFFFF", "profile_sidebar_fill_col or":"F3F3F3","profile text color":"333333","profile use background i mage":true,"profile image url":"http:\/\/pbs.twimg.com\/profile imag es\/917924626319241216\/XysMKjrf_normal.jpg","profile_image_url_http $s":"https:\/\pbs.twimg.com\/profile images\/917924626319241216\/Xys$ MKjrf normal.jpg","profile banner url":"https:\/\/pbs.twimg.com\/pro file_banners\/479029743\/1507603201", "default_profile": false, "defaul t_profile_image":false,"following":null,"follow_request_sent":nul l, "notifications": null}, "geo": null, "coordinates": null, "place": nul l,"contributors":null,"retweeted_status":{"created_at":"Sat Oct 07 2 3:49:41 +0000 2017", "id":916812806917419008, "id str": "91681280691741 9008", "text": "what would u rather have, 10,000 for a Louis Vuitton d ead python purse or 2,000 for my living python Louis https:\/\/t.co \/CLHlSFMezK", "display_text_range":[0,106], "source": "\u003ca href= \"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow\"\u003eTwi tter for iPhone\u003c\/a\u003e", "truncated": false, "in reply to statu s id":null, "in reply to status id str":null, "in reply to user id":nu ll, "in_reply_to_user_id_str":null, "in_reply_to_screen_name":null, "us er":{"id":855106531, "id_str":"855106531", "name": "Taylor Nicole Dea n", "screen_name": "taylorndean", "location": "texas", "url": "https:\/\/y outu.be\/nd1KAwmf2lI", "description": "family friendly pet mom | insta gram & snapchat: taylorndean | newest video below !!!!","translator type": "none", "protected": false, "verified": true, "followers_count": 107 517, "friends count":575, "listed count":608, "favourites count":6163 3, "statuses_count":12702, "created_at": "Sun Sep 30 17:57:46 +0000 201

2","utc_offset":-14400,"time_zone":"Eastern Time (US & Canada)","geo _enabled":true,"lang":"en","contributors_enabled":false,"is_translat or":false, "profile background color": "FFFFFF", "profile background im age url":"http:\/\/pbs.twimg.com\/profile background images\/3788000 00125843992\/SWJQaVvR.jpeg", "profile background image url https": "ht tps:\/\/pbs.twimg.com\/profile_background_images\/378800000125843992 \/SWJQaVvR.jpeg","profile_background_tile":false,"profile_link_colo r":"13C4B5","profile_sidebar_border_color":"FFFFFF","profile_sidebar fill color": "252429", "profile text color": "6666666", "profile use bac kground image":true, "profile image url": "http:\/\/pbs.twimg.com\/pro file images\/916953555445141504\/KHMptV3G normal.jpg","profile image url https":"https:\/\/pbs.twimg.com\/profile images\/91695355544514 1504\/KHMptV3G_normal.jpg","profile_banner_url":"https:\/\/pbs.twim $g.com\profile\profil$ se, "default profile image": false, "following": null, "follow request se nt":null, "notifications":null}, "geo":null, "coordinates":null, "plac e":null, "contributors":null, "is_quote_status":false, "quote_count":30 8, "reply_count":170, "retweet_count":2933, "favorite_count":11940, "ent ities":{"hashtags":[], "urls":[], "user_mentions":[], "symbols":[], "med ia":[{"id":916812795722661888,"id str":"916812795722661888","indice s":[107,130], "media url": "http:\/\/pbs.twimg.com\/media\/DLksYVaVAAA Lj98.jpg", "media url https": "https:\/\/pbs.twimg.com\/media\/DLksYVa VAAALj98.jpg", "url": "https:\/\/t.co\/CLHlSFMezK", "display url": "pic. twitter.com\/CLHlSFMezK","expanded url":"https:\/\/twitter.com\/tayl orndean\/status\/916812806917419008\/photo\/1","type":"photo","size s":{"medium":{"w":675,"h":1200,"resize":"fit"},"small":{"w":383,"h": 680, "resize": "fit"}, "thumb": {"w": 150, "h": 150, "resize": "crop"}, "larg e":{"w":1152,"h":2048,"resize":"fit"}}}]},"extended entities":{"medi a":[{"id":916812795722661888,"id_str":"916812795722661888","indice s":[107,130],"media url":"http:\/\/pbs.twimg.com\/media\/DLksYVaVAAA Lj98.jpg", "media url https": "https:\/\/pbs.twimg.com\/media\/DLksYVa VAAALj98.jpg", "url": "https:\/\/t.co\/CLHlSFMezK", "display url": "pic. twitter.com\/CLHlSFMezK","expanded_url":"https:\/\/twitter.com\/tayl $orndean\sl 916812806917419008\photo\sl "", "type": "photo", "size"$ s":{"medium":{"w":675,"h":1200,"resize":"fit"},"small":{"w":383,"h": 680, "resize": "fit"}, "thumb": {"w": 150, "h": 150, "resize": "crop"}, "larg e":{"w":1152,"h":2048,"resize":"fit"}}},{"id":916812795735187456,"id str":"916812795735187456","indices":[107,130],"media url":"http:\/ \/pbs.twimg.com\/media\/DLksYVdUIAAfuYg.jpg","media url https":"http s:\/\/pbs.twimg.com\/media\/DLksYVdUIAAfuYg.jpg","url":"https:\/\/t. co\/CLHlSFMezK","display url":"pic.twitter.com\/CLHlSFMezK","expande d url":"https:\/\/twitter.com\/taylorndean\/status\/9168128069174190 08\/photo\/1","type":"photo","sizes":{"thumb":{"w":150,"h":150,"resi ze":"crop"}, "medium":{"w":600, "h":800, "resize":"fit"}, "large":{"w":6 00, "h":800, "resize": "fit"}, "small": {"w":510, "h":680, "resize": "fi t"}}}], "favorited":false, "retweeted":false, "possibly_sensitive":fal se, "filter_level": "low", "lang": "en"}, "is_quote_status": false, "quote_ count":0,"reply_count":0,"retweet_count":0,"favorite_count":0,"entit ies":{"hashtags":[],"urls":[],"user_mentions":[{"screen_name":"taylo rndean", "name": "Taylor Nicole Dean", "id":855106531, "id_str": "8551065 31", "indices":[3,15]}], "symbols":[]}, "favorited":false, "retweeted":f alse, "filter_level": "low", "lang": "en", "timestamp_ms": "150772465211 5"}

screen name":null, "user":{"id":491363748, "id str": "491363748", "nam e":"Raymond Falkenhage", "screen_name": "Falkenhage", "location":nul l, "url":null, "description": "Liefhebber van de Efteling. #TeamHugo. I t's a Raymond thing! You wouldn't understand! \u2764\ufe0f @Mandyhon line 10-3-2015. \ud83d\udd34\u26aa", "translator type": "none", "protec ted":false,"verified":false,"followers_count":219,"friends_count":21 6,"listed_count":4,"favourites_count":581,"statuses_count":15305,"cr eated_at":"Mon Feb 13 15:12:37 +0000 2012","utc_offset":3600,"time_z one": "Casablanca", "geo enabled": true, "lang": "nl", "contributors enabl ed":false,"is translator":false,"profile background color":"CODEE D", "profile_background_image_url": "http:\/\/abs.twimg.com\/images\/t hemes\/theme1\/bg.png","profile_background_image_url_https":"http s:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png","profile_backgr ound_tile":false,"profile_link_color":"1DA1F2","profile_sidebar_bord er_color":"CODEED","profile_sidebar_fill_color":"DDEEF6","profile_te
xt_color":"333333","profile_use_background_image":true,"profile_imag e url":"http:\/\/pbs.twimg.com\/profile images\/619085106943291392\/ ye47mqtC_normal.jpg","profile_image_url_https":"https:\/\/pbs.twimg. com\/profile images\/619085106943291392\/ye47mqtC normal.jpg","profi le banner url":"https:\/\/pbs.twimg.com\/profile banners\/491363748 \/1421534969", "default profile":true, "default profile image":fals e, "following": null, "follow_request_sent": null, "notifications": nul l}, "geo":null, "coordinates":null, "place":null, "contributors":null, "i s_quote_status":false,"quote_count":0,"reply_count":0,"retweet_coun t":0, "favorite count":0, "entities":{ "hashtags":[{ "text": "Eftelin g","indices":[25,34]}],"urls":[],"user_mentions":[],"symbols":[],"me dia":[{"id":918089907335237632,"id str":"918089907335237632","indice s":[35,58], "media url": "http:\/\/pbs.twimg.com\/media\/DL216A6WkAA9i Sj.jpg", "media url https": "https:\/\/pbs.twimg.com\/media\/DL216A6Wk AA9iSj.jpg","url":"https:\/\/t.co\/KevwHaC9rH","display_url":"pic.tw itter.com\/KevwHaC9rH","expanded url":"https:\/\/twitter.com\/Falken $hage\sl "1","type":"photo","sizes":$ {"thumb":{"w":150,"h":150,"resize":"crop"},"large":{"w":2048,"h":153 6, "resize": "fit"}, "medium": {"w": 1200, "h": 900, "resize": "fit"}, "smal l":{"w":680,"h":510,"resize":"fit"}}}],"extended entities":{"medi a":[{"id":918089907335237632,"id_str":"918089907335237632","indice s":[35,58], "media_url": "http:\/\/pbs.twimg.com\/media\/DL216A6WkAA9i Sj.jpg", "media url https": "https:\/\/pbs.twimg.com\/media\/DL216A6Wk AA9iSj.jpg", "url": "https:\/\/t.co\/KevwHaC9rH", "display url": "pic.tw itter.com\/KevwHaC9rH","expanded_url":"https:\/\/twitter.com\/Falken hage\/status\/918089926532460544\/photo\/1","type":"photo","sizes": {"thumb":{"w":150,"h":150,"resize":"crop"},"large":{"w":2048,"h":153 6, "resize": "fit"}, "medium": {"w": 1200, "h": 900, "resize": "fit"}, "smal l":{"w":680,"h":510,"resize":"fit"}}}]},"favorited":false,"retweete d":false, "possibly sensitive":false, "filter level": "low", "lang": "n l","timestamp ms":"1507724670428"}

{"created_at":"Wed Oct 11 12:24:48 +0000 2017", "id":9180900026296729 60, "id_str":"918090002629672960", "text": "RT @barrysmyth: How did I m iss this? Great resource of code and examples. The Python Graph Gall ery \u2013 #python #DataScience https:\/\/t.co\/s02A\u2026", "sourc e":"\u003ca href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"no follow\"\u003eTwitter for iPhone\u003c\/a\u003e", "truncated":fals e, "in_reply_to_status_id":null, "in_reply_to_status_id_str":null, "in_reply_to_user_id_str":null, "in_reply_to_s creen_name":null, "in_reply_to_user_id_str":"14286148", "name":"votis", "screen_name":"votis", "location":"Taiwan, Taipei, Tamsui", "url": null, "description":"Animo!", "translator_type":"none", "protected":fal se, "verified":false, "followers_count":47, "friends_count":219, "listed _count":2, "favourites_count":12, "statuses_count":5012, "created_a t":"Wed Apr 02 18:36:42 +0000 2008", "utc_offset":28800, "time_zon

e":"Taipei", "geo_enabled":false, "lang": "en", "contributors_enabled":f alse, "is_translator":false, "profile_background_color": "EBEBEB", "prof ile background image url":"http:\/\/abs.twimg.com\/images\/themes\/t heme7\/bg.gif","profile_background_image_url_https":"https:\/\/abs.t wimg.com\/images\/themes\/theme7\/bg.gif","profile background tile": false, "profile_link_color": "990000", "profile_sidebar_border_colo r":"DFDFDF","profile_sidebar_fill_color":"F3F3F3","profile_text_colo r":"333333","profile_use_background_image":true,"profile_image_ur $l'': http: \/\pbs.twimg.com\profile images \/860663990241828864 \/8Uz$ ltm0_normal.jpg","profile_image_url_https":"https:\/\/pbs.twimg.com \/profile_images\/860663990241828864\/8Uz_ltm0_normal.jpg","profile_ banner url":"https:\/\/pbs.twimg.com\/profile banners\/14286148\/149 4033251", "default_profile":false, "default_profile_image":false, "foll owing":null, "follow_request_sent":null, "notifications":null}, "geo":n ull, "coordinates":null, "place":null, "contributors":null, "retweeted s tatus":{"created at":"Wed Oct 11 08:02:21 +0000 2017","id":918023956 719861760, "id str": "918023956719861760", "text": "How did I miss this? Great resource of code and examples. The Python Graph Gallery \u2013 #python #DataScience https:\/\/t.co\/s02At72BG9","source":"\u003ca h ref=\"http:\/\/www.apple.com\/\" rel=\"nofollow\"\u003e0S X\u003c\/a \u003e","truncated":false,"in_reply_to_status_id":null,"in_reply_to_ status_id_str":null,"in_reply_to_user_id":null,"in_reply_to_user_id_ str":null, "in reply to screen name":null, "user":{"id":16534969, "id s tr":"16534969","name":"barrysmyth","screen_name":"barrysmyth","locat ion":"Dublin, Ireland","url":"http:\/\/barrysmyth.me","descriptio n": "Professor of Computer Science. Founding Director, Insight Centre for Data Analytics. Enterpreneur. Runner. Geek-dad.", "translator typ e":"none", "protected":false, "verified":false, "followers count":222 9, "friends_count":828, "listed_count":160, "favourites_count":483, "sta tuses count":10420, "created at": "Tue Sep 30 21:15:19 +0000 2008", "ut c offset":3600,"time zone":"Dublin","geo enabled":true,"lang":"e n", "contributors enabled":false, "is translator":false, "profile backg round_color":"1A1B1F","profile_background_image_url":"http:\/\/abs.t wimg.com\/images\/themes\/theme9\/bg.gif","profile background image url https":"https:\/\/abs.twimg.com\/images\/theme9\/bg.gi f", "profile_background_tile":false, "profile_link_color": "2FC2EF", "pr ofile_sidebar_border_color":"181A1E","profile_sidebar_fill_color":"2 52429","profile_text_color":"6666666","profile_use_background_image": true,"profile image url":"http:\/\/pbs.twimg.com\/profile images\/42 5673449478946818\/qXKfkxJ7_normal.jpeg","profile_image_url_https":"h ttps:\/\/pbs.twimg.com\/profile images\/425673449478946818\/qXKfkxJ7 _normal.jpeg","profile_banner_url":"https:\/\/pbs.twimg.com\/profile banners\/16534969\/1390323469", "default profile": false, "default pro file_image":false,"following":null,"follow_request_sent":null,"notif ications":null, "geo":null, "coordinates":null, "place":null, "contribu tors":null, "is_quote_status":false, "quote_count":1, "reply_count": 0, "retweet_count":10, "favorite_count":10, "entities": { "hashtags":[{"t ext":"python","indices":[85,92]},{"text":"DataScience","indices":[9 3,105]}], "urls":[{"url":"https:\/\/t.co\/s02At72BG9", "expanded ur l":"https:\/\/python-graph-gallery.com\/","display_url":"python-grap h-gallery.com", "indices": [106,129]}], "user_mentions": [], "symbols": []}, "favorited": false, "retweeted": false, "possibly_sensitive": fals e, "filter_level": "low", "lang": "en"}, "is_quote_status": false, "quote_c ount":0, "reply_count":0, "retweet_count":0, "favorite_count":0, "entiti es":{"hashtags":[{"text":"python","indices":[101,108]},{"text":"Data Science", "indices": [109,121]}], "urls": [], "user_mentions": [{"screen_n ame": "barrysmyth", "name": "barrysmyth", "id": 16534969, "id_str": "165349 69", "indices": [3,14]}], "symbols": []}, "favorited": false, "retweeted": f alse, "filter_level": "low", "lang": "en", "timestamp_ms": "150772468857 1"}

{"created at":"Wed Oct 11 12:24:55 +0000 2017","id":9180900323297771 52, "id str": "918090032329777152", "text": "SickRage: Download Finishe d: Highway Thru Hell - 6x06 - Python Unleashed - SDTV", "source": "\u0 03ca href=\"http:\/\/www.sickbeard.com\" rel=\"nofollow\"\u003eSick Beard\u003c\/a\u003e", "truncated":false, "in reply to status id":nul l,"in_reply_to_status_id_str":null,"in_reply_to_user_id":null,"in_re ply_to_user_id_str":null,"in_reply_to_screen_name":null,"user":{"i d":2358889982, "id str": "2358889982", "name": "Sam's Plex Server", "scre en name":"SamsPlexServer","location":null,"url":null,"description":n ull, "translator type": "none", "protected": false, "verified": false, "fol lowers_count":29, "friends_count":0, "listed_count":1, "favourites_coun t":2, "statuses count":7850, "created at": "Mon Feb 24 03:38:33 +0000 2 014", "utc_offset":null, "time_zone":null, "geo_enabled":false, "lan g":"en","contributors enabled":false,"is translator":false,"profile background color": "CODEED", "profile background image url": "http:\/\/ abs.twimg.com\/images\/themes\/theme1\/bg.png","profile background i mage url https":"https:\/\/abs.twimg.com\/images\/themes\/theme1\/b g.png","profile_background_tile":false,"profile_link_color":"1DA1F 2", "profile_sidebar_border_color": "CODEED", "profile_sidebar_fill_col or":"DDEEF6","profile text color":"333333","profile use background i mage":true,"profile image url":"http:\/\/pbs.twimg.com\/profile imag es\/437794080227999744\/0BVVE1wI_normal.jpeg","profile_image_url_htt ps":"https:\/\/pbs.twimg.com\/profile images\/437794080227999744\/0B VVE1wI_normal.jpeg","default_profile":true,"default_profile_image":f alse, "following": null, "follow request sent": null, "notifications": nul l}, "geo":null, "coordinates":null, "place":null, "contributors":null, "i s quote status":false, "quote count":0, "reply count":0, "retweet coun t":0, "favorite count":0, "entities":{ "hashtags":[], "urls":[], "user me ntions":[],"symbols":[]},"favorited":false,"retweeted":false,"filter
_level":"low","lang":"en","timestamp_ms":"1507724695652"}

{"created at": "Wed Oct 11 12:25:05 +0000 2017", "id": 9180900716848619 52, "id str": "918090071684861952", "text": "Software Engineer Python Ge neralist @Landing jobs Lisbon Portugal https:\/\/t.co\/0CUIsPiTEU #L inux #Machinelearning #Permanent", "source": "\u003ca href=\"http:\/\/ www.hootsuite.com\" rel=\"nofollow\"\u003eHootsuite\u003c\/a\u003 e","truncated":false,"in_reply_to_status_id":null,"in_reply_to_statu s id str":null, "in reply to user id":null, "in reply to user id str": null, "in reply to screen name":null, "user": { "id": 72453431982231142 4,"id str":"724534319822311424","name":"SA Python Jobs","screen nam e":"sa python jobs","location":"Global","url":"http:\/\/www.silicona rmada.com", "description": "Python Jobs", "translator type": "none", "pro tected":false, "verified":false, "followers count":374, "friends coun t":244, "listed_count":426, "favourites_count":14, "statuses_count":118 92, "created at": "Mon Apr 25 09:43:37 +0000 2016", "utc offset": -2520 0, "time zone": "Pacific Time (US & Canada)", "geo enabled": false, "lan g":"en", "contributors enabled":false, "is translator":false, "profile background_color":"F5F8FA","profile_background_image_url":"","profil e_background_image_url_https":"","profile_background_tile":false,"pr ofile_link_color":"1DA1F2","profile_sidebar_border_color":"CODEE D","profile_sidebar_fill_color":"DDEEF6","profile_text_color":"33333 3","profile_use_background_image":true,"profile_image_url":"http:\/ $\protect\pro$ l.jpg","profile_image_url_https":"https:\/\/pbs.twimg.com\/profile_i mages\/724534604376510464\/_gqKwJx__normal.jpg","profile_banner_ur l":"https:\/\/pbs.twimg.com\/profile_banners\/724534319822311424\/14 61577626", "default profile": true, "default profile image": false, "foll owing":null,"follow request sent":null,"notifications":null},"geo":n ull, "coordinates":null, "place":null, "contributors":null, "is_quote_st atus":false, "quote count":0, "reply count":0, "retweet count":0, "favor ite_count":0, "entities":{"hashtags":[{"text":"Linux", "indices":[90,9 6]},{"text":"Machinelearning","indices":[97,113]},{"text":"Permanen t","indices":[114,124]}],"urls":[{"url":"https:\/\/t.co\/0CUIsPiTE U","expanded_url":"http:\/\/bit.ly\/2wluKzJ","display_url":"bit.ly\/2wluKzJ","indices":[66,89]}],"user_mentions":[{"screen_name":"Landin g_jobs","name":"Landing\u2022jobs","id":1555969092,"id_str":"1555969092","indices":[36,49]}],"symbols":[]},"favorited":false,"retweete d":false,"possibly_sensitive":false,"filter_level":"low","lang":"un d","timestamp ms":"1507724705035"}

{"created at": "Wed Oct 11 12:25:09 +0000 2017", "id": 9180900900936089 61,"id_str":"918090090093608961","text":"@kyran_dale we thought you might like to attend our free Intel Code Modernization Workshop, 1-2 Nov, Oxford https:\/\/t.co\/4eAwV1q1sq #Python","source":"\u003ca href=\"http:\/\/www.hootsuite.com\" rel=\"nofollow\"\u003eHootsuite \u003c\/a\u003e","truncated":false,"in reply to status id":null,"in reply to status id str":null, "in reply to user id":2677230157, "in re ply to user id str": "2677230157", "in reply to screen name": "kyran da "user":{"id":113019315,"id str":"113019315","name":"Grey Matte r", "screen name": "GreyMatter News", "location": "Devon, UK", "url": "htt p:\/\/www.greymatter.com","description":"Providing independent and t rusted software supply, services, management and advice for organisa tions seeking development, technical and business solutions.", "trans lator type": "none", "protected": false, "verified": false, "followers cou nt":895, "friends count":691, "listed count":232, "favourites count":52 8, "statuses count":6672, "created at": "Wed Feb 10 12:47:02 +0000 201 ."utc_offset":3600,"time_zone":"London","geo_enabled":true,"lan q":"en", "contributors enabled":false, "is translator":false, "profile background color":"022330","profile background image url":"http:\/\/ pbs.twimg.com\/profile background images\/496224204241391616\/XUXMQe 9E.jpeg", "profile background image url https": "https:\/\/pbs.twimg.c om\/profile background images\/496224204241391616\/XUXMQe9E.jpeg","p rofile background tile":false, "profile link color": "5B5B5B", "profile _sidebar_border_color":"FFFFFF","profile_sidebar_fill_color":"DDEEF 6", "profile text color": "333333", "profile use background image": tru e, "profile image url": "http:\/\/pbs.twimg.com\/profile images\/91478 8013955313664\/7WS-_cNd_normal.jpg","profile_image_url_https":"http $s:\/\pbs.twimg.com\profile_images\/914788013955313664\/7WS-_cNd_no$ rmal.jpg","profile banner url":"https:\/\/pbs.twimg.com\/profile ban ners\/113019315\/1506937428", "default profile": false, "default profil e image":false, "following":null, "follow request sent":null, "notifica tions":null}, "geo":null, "coordinates":null, "place":null, "contributor s":null, "is quote status":false, "quote count":0, "reply count":0, "ret weet count":0,"favorite count":0,"entities":{"hashtags":[{"text":"Py thon", "indices": [132,139]}], "urls": [{"url": "https:\/\/t.co\/4eAwV1q1 sq", "expanded url": "http:\/\/ow.ly\/tIDv30fNsNy", "display url": "ow.l y\/tIDv30fNsNy","indices":[108,131]}],"user_mentions":[{"screen nam e":"kyran dale", "name": "Kyran Dale", "id": 2677230157, "id str": "267723 0157", "indices": [0,11]}], "symbols": []}, "favorited": false, "retweete d":false, "possibly_sensitive":false, "filter_level": "low", "lang": "e n","timestamp_ms":"1507724709424"}

{"created_at":"Wed Oct 11 12:25:23 +0000 2017","id":9180901475135610
89,"id_str":"918090147513561089","text":"@python 5\u670817\u65e5 (\u
65e7\u66a6)\n NowTime: 2017\/10\/11 21:25:22.9451\nTweetCount: 227
3","source":"\u003ca href=\"http:\/\/twitter.com\/download\/android
\" rel=\"nofollow\"\u003eTwitter for Android\u003c\/a\u003e","trunca
ted":false,"in_reply_to_status_id":null,"in_reply_to_status_id_str":
null,"in_reply_to_user_id":765482,"in_reply_to_user_id_str":"76548
2","in_reply_to_screen_name":"Python","user":{"id":2449488770,"id_st
r":"2449488770","name":"Ferix","screen_name":"nfz6","location":nul
l,"url":"http:\/\/twpf.jp\/nfz6","description":"Razer\/PUBG\/R6S\/Ry

zen \u9593\u9055\u3048\u305f\u9053\u3092\u6b69\u3093\u3067\u308b\u4e ba\u3067\u3059\u3002", "translator type": "regular", "protected": fals e, "verified": false, "followers count": 64447, "friends count": 17, "liste d_count":0,"favourites_count":7682,"statuses_count":111212,"created at": "Thu Apr 17 10:35:46 +0000 2014", "utc offset": 32400, "time zon e":"Seoul", "geo enabled":false, "lang": "ja", "contributors enabled":fa lse,"is_translator":false,"profile_background_color":"000000","profi le background image url":"http:\/\/abs.twimg.com\/images\/themes\/th emel\/bg.png","profile background image url https":"https:\/\/abs.tw img.com\/images\/themes\/theme1\/bg.png","profile background tile":f alse, "profile_link_color": "FF0080", "profile_sidebar_border_color": "0 00000", "profile sidebar fill color": "000000", "profile text color": "0 00000", "profile_use_background_image": false, "profile_image_url": "htt $p:\/\pbs.twimg.com\/profile images\/917741971917094913\/Tna0Rd5R no$ rmal.jpg","profile image url https":"https:\/\/pbs.twimg.com\/profil e images $\/917741971917094913\/$ Tna $\/$ Rd5R normal.jpg","profile banner u rl":"https:\/\/pbs.twimg.com\/profile banners\/2449488770\/150770213 0", "default profile": false, "default profile image": false, "followin g":null, "follow request sent":null, "notifications":null}, "geo":nul l, "coordinates":null, "place":null, "contributors":null, "is quote stat us":false, "quote count":0, "reply count":0, "retweet count":0, "favorit e_count":0,"entities":{"hashtags":[],"urls":[],"user_mentions":[{"sc reen name": "Python", "name": "Python", "id": 765482, "id str": "765482", "i ndices":[0,7]}], "symbols":[]}, "favorited":false, "retweeted":false, "f ilter level":"low","lang":"ja","timestamp ms":"1507724723114"}

In [1]:

```
import json
with open('twitter.json', 'r') as f:
    line = f.readline() # only first tweet
    tweet = json.loads(line) # as a dict
    print(json.dumps(tweet, indent=4)) # output formatting
```

```
11.10.2017
                                         5 data collection-Copy1
   {
       "in_reply_to_status_id": null,
       "source": "<a href=\"http://twitter.com\" rel=\"nofollow\">Twitt
   er Web Client</a>",
       "place": null,
       "in reply to status id str": null,
       "favorite count": 0,
       "possibly_sensitive": false,
       "in reply to screen name": null,
       "lang": "da",
       "timestamp ms": "1457702276931",
       "in reply to user id": null,
       "coordinates": null,
       "favorited": false,
       "geo": null,
       "created at": "Fri Mar 11 13:17:56 +0000 2016",
       "contributors": null,
       "user": {
           "notifications": null,
           "name": "georgi mitchew",
           "profile banner url": "https://pbs.twimg.com/profile banner
   s/227091033/1450559987",
           "profile_image_url_https": "https://pbs.twimg.com/profile_im
   ages/469072348629327873/Gg91WkGv_normal.jpeg",
           "profile background tile": false,
           "protected": false,
           "profile_sidebar_border_color": "CODEED",
           "location": "Copenhagen, Denmark",
           "url": null,
           "profile background image url": "http://abs.twimg.com/image
   s/themes/theme1/bg.png",
           "description": "Pro bono tekstforfatter og researcher (freel
   ance). Hjerte banker is\u00e6r for v\u00e6rgel\u00f8se b\u00f8rn, dy
   r og \u00e6ldre. Freder whistleblowing og st\u00f8tter socialt under
   trykte.",
           "default_profile": true,
           "lang": "da",
"following": null,
           "profile sidebar fill color": "DDEEF6",
           "favourites_count": 7542,
           "is translator": false,
           "default_profile_image": false,
           "created at": "Wed Dec 15 21:58:46 +0000 2010",
           "contributors enabled": false,
           "listed_count": 47,
"screen_name": "gmitchew",
           "geo enabled": true,
           "profile_link_color": "0084B4",
           "follow request sent": null,
           "utc offset": 3600,
           "followers_count": 830,
           "id": 227091033,
           "statuses count": 16698,
           "profile background color": "CODEED",
           "profile use_background_image": true,
           "time_zone": "Copenhagen",
           "profile background image url https": "https://abs.twimg.co
   m/images/themes/theme1/bg.png",
           "friends count": 2202,
           "profile_text_color": "333333",
           "verified": false,
```

```
"id str": "227091033"
        "profile image url": "http://pbs.twimg.com/profile images/46
9072348629327873/Gg91WkGv normal.jpeg"
    "entities": {
        "symbols": [],
        "user mentions": [],
        "urls": [
            {
                "indices": [
                     99,
                     122
                ],
                "url": "https://t.co/waCZ39x8Ih",
                "display url": "kino.dk/nyheder/2016/0\u2026",
                "expanded url": "http://www.kino.dk/nyheder/2016/03/
sjoveste-film-nogensinde"
        ],
        "hashtags": [
            {
                "indices": [
                    82,
                    88
                "text": "dkpol"
            },
                "indices": [
                     89,
                     98
                "text": "dkmedier"
            }
        1
    "text": "Sjoveste film nogensinde? Dit bud? Mit - ubetinget: Mon
ty Python's Life of Brian! #dkpol #dkmedier https://t.co/waCZ39x8I
h",
    "truncated": false,
    "id": 708280801398562816,
    "retweet count": 0,
    "in_reply_to_user_id_str": null,
    "is_quote_status": false,
    "filter level": "low",
    "retweeted": false,
    "id str": "708280801398562816"
}
In [3]:
with open('twitter.json', 'r') as f:
    for line in f:
        tweet = json.loads(line) # each tweet as a dict
    print(tweet["text"])
```

RT @QrnaJnyynpr: #100DaysOfCode - Actually I'm still in bed. I promi se, more HTML and Python after just <lies>five</lies> mo re minutes.

Looking for tweets on a given topic

In [4]:

```
import tweepy

consumer_key="XXXX"
consumer_secret="XXXXX"

access_token="XXXXX"

auth = tweepy.auth.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token,access_token_secret)
api = tweepy.API(auth)
search_results = api.search(q="python", count=100)

for i in search_results:
    print(i.created_at, i.text)
    print()
```

2017-10-11 12:31:04 MySQL workbench utilities administer #MySQL with python scripts

https://t.co/004dL5PcEj

2017-10-11 12:31:03 @ncoghlan_dev @ashemedai @dstufft @DRMacIver @hy nek @tintvrtkovic (I also don't have robust internet access, I have... https://t.co/bfEdyc8kAi

2017-10-11 12:31:00 evoke 5.9: a simple and powerful python web fram ework with pythonic "evo" templating https://t.co/Bhkxax77Jb

2017-10-11 12:30:49 A python that cannot dance in sambisa forest, is that one a python?

2017-10-11 12:30:38 Soup, biscuits and Monty Python ⊕

2017-10-11 12:30:38 ほしくうん♥

2017-10-11 12:30:27 RT @HOSHIBACKYARD: ぱいにき///

2017-10-11 12:30:26 RT @analyticbridge: R and Python cheatsheets https://t.co/iauHYE1meZ https://t.co/m97FQOnQOJ

2017-10-11 12:30:25 What are the best GitHub repositories that anyon e can contribute to? https://t.co/XopAWJcsGm #Python

I'd like to take part in DigitalOce...

2017-10-11 12:30:22 try: python brasil*2

except:

outubro.replace('semana das crianças', 'Python Brasil')
#pybr13 #uaipython

2017-10-11 12:30:19 Operation Python Dance in the SE arrested 8 suspected cult members, vandals & drug peddler https://t.co/YdwRxf6xyj Great ROI. BH nko?

2017-10-11 12:30:14 the inventor of your brother's Python code is NP -complete

2017-10-11 12:30:04 @Riklin10 https://t.co/7ZGYkiA0Gr

2017-10-11 12:30:03 barisuke がQiitaに投稿しました.

[python] 複数行の標準入力の処理

Python

https://t.co/IzyVSAVZH9

2017-10-11 12:30:02 nori-dev-akg がQiitaに投稿しました.

Bluetoothシャットダウンボタンを作る #300円でIoTボタン

Python・RaspberryPi・IoT・ダイソー

https://t.co/b1V75mtj9S

2017-10-11 12:30:01 Стрим по майнкрафту в 10 часов вечера по Киеву

2017-10-11 12:30:01 @ncoghlan_dev @ashemedai @dstufft @DRMacIver @hy nek @tintvrtkovic So the theoretical user here is one that both use... https://t.co/NrMvDlRrVL

2017-10-11 12:29:49 RT @lucazav: You can finally query #SQLServer ta

- bles using #regex with #R in 2016 and 2017 (or even #Python in 2017), and #TSQL https://t.m.
- 2017-10-11 12:29:41 (ちょっとそこからよく分かんないけど)同じ号を二冊買ってしまったとかで日経ソフトウェアを一冊貰ったので試しにpythonとやらをさわってみたんだけど、ホントにサンプル動かしてから紙面に乗せてんだろうな…と疑ってしまう。
- 2017-10-11 12:29:41 RT @PythonIreland: Retweet this for a chance to win 1 of 2 free tickets for PyConIE 2017 (closing date : Sat 14th 0 ct Noon) #Python https:/...
- 2017-10-11 12:29:16 RT @abhaga: Hiring a developer at https://t.co/R fsEKy60VT. Please help spread the word. Prior Python experience not necessary. https://t.co...
- 2017-10-11 12:29:11 Bluetoothシャットダウンボタンを作る #300円でIoTボタン - https://t.co/lyfphTrhc4
- 2017-10-11 12:29:01 RT @analyticbridge: R and Python cheatsheets https://t.co/iauHYElmeZ https://t.co/m97FQOnQOJ
- 2017-10-11 12:28:55 RT @BiafraHerald: BIAFRA: " OPERATION PYTHON DAN CE 11" AND WHY NIGERIAN GOVERNMENT/ARMY MUST PROVIDE MAZI NNAMDI KAN U https://t.co/1QZDpw6D...
- 2017-10-11 12:28:44 【定期】沖縄でPythonのイベント開催に興味ある方、コミュニティに参加してみたい方、Python始めてみたいけどどうしたら良いかわからない方、リプライいただければなにかお手伝いできるかもしれないです。気軽にどうぞ。
- 2017-10-11 12:28:39 緋色の大人達の信念と幸せを何よりも誰よりも願う共犯者達の話
- 残党は居るものの実質組織の破壊に成功した後幸せを噛みしめたのも束の間 何も知りもしない身勝手な大人達は知りすぎた青年を危険視。情報を流す事もましてや大切 な人が守る国を裏切るつもりなぞないがお偉い方はそうは思わなかった
- 2017-10-11 12:28:06 ほしくん好きだよこれ
- 2017-10-11 12:28:05 RT @pythonbot_: Hands-on Data Science And Python Machine Learning https://t.co/XIU0ZkzTPT #machine #learning #python@pythonbot_
- 2017-10-11 12:28:00 tf-nightly 1.4.0.dev20171011: TensorFlow helps the tensors flow https://t.co/PHC43bGUAf
- 2017-10-11 12:28:00 snipsskills 0.1.6.21: Snips Skills Manager https://t.co/Pi8cxj0fw7
- 2017-10-11 12:27:36 Hands-on Data Science And Python Machine Learning https://t.co/XIU0ZkzTPT #machine #learning #python @pythonbot_
- 2017-10-11 12:27:35 A Concise Introduction To Programming In Python (chapman & Hall/crc Tex https://t.co/NjgWSuS7EX #software #development #python @pythonbot
- 2017-10-11 12:27:30 @python_klfn 油のかたまりだからね。ゆりさんも気をつけてくださいね⊜⊜⊜
- 2017-10-11 12:27:19 [Python] 请问哪里可以找到在爬虫的 pipelines.py 中将数据写入 MYSQL 数据库的例子?: 想在 pipelines.py 中将爬到的数据写入 MYSQL 数据库,希望能有这样的范例代码可供学习,

- 恳请大家... https://t.co/ra6IdIbL6Z
- 2017-10-11 12:27:02 From bytes to strings in #Python and back again https://t.co/tEcpexGUCW https://t.co/xK1jOwtrkI
- 2017-10-11 12:26:57 RT @analyticbridge: R and Python cheatsheets https://t.co/iauHYElmeZ https://t.co/m97FQOnQOJ
- 2017-10-11 12:26:50 PYTHON EATS A PORCUPINE AND INSTANTLY REGRETS HI S DECISION! https://t.co/jLCT8jpEqe via @YouTube
- 2017-10-11 12:26:49 RT @RNPYM_: 【急募】11/12に絶対絶望少女の併せを行います。此方カメラマン様揃っておりますがレイヤー様が切実に足りておりません。FF内外問わず募集致しておりますので、お気軽にお声掛け下さい。宜しくお願い致します!!#拡散希望 #併せ募集 https://t.co/...
- 2017-10-11 12:26:43 RT @IBMResearch: Developers, check out QISKit! & amp; explore #IBMQ Experience using a Python interface. https://t.co/9wJymvDZX8 https://t.co/tY...
- 2017-10-11 12:26:42 RT @HOSHIBACKYARD: 当時巨峰食べながら無表情無言であのツイートしてました...すみません許してください
- 2017-10-11 12:26:33 RT @HOSHIBACKYARD: 巨峰を食べた人「キョッホーwwwww」
- 2017-10-11 12:26:01 RT @PythonWeekly: PySyft Encrypted Deep Learning Library. https://t.co/hwHsmEUepd #python
- 2017-10-11 12:26:00 Pareidoscope 0.10.2: A collection of tools for d etermining the association between arbitrary linguistic s... https://t.co/M6hX4Ke5b0
- 2017-10-11 12:26:00 aiobfd 0.1: Asynchronous BFD Daemon https://t.c o/OXYWpjR2Hq
- 2017-10-11 12:26:00 django-powerbank 0.2.15: Extra power for include d batteries https://t.co/KPNigP3K0H
- 2017-10-11 12:25:58 RT @thefronthat: Software Developer at YouGov #w orkfromhome #Python #MongoDB https://t.co/DvQ0c7KB5x
- 2017-10-11 12:25:44 @dstufft @ashemedai @ncoghlan_dev @DRMacIver @hy nek @tintvrtkovic ship your python platform installer to install so… https://t.co/KXKrUwVsFo
- 2017-10-11 12:25:09 @kyran_dale we thought you might like to attend our free Intel Code Modernization Workshop, 1-2 Nov, Oxford https://t.co/4eAwVlq1sq #Python
- 2017-10-11 12:25:05 Software Engineer Python Generalist @Landing_job s Lisbon Portugal https://t.co/0CUIsPiTEU #Linux #Machinelearning #Permanent
- 2017-10-11 12:24:55 SickRage: Download Finished: Highway Thru Hell 6x06 Python Unleashed SDTV
- 2017-10-11 12:24:48 RT @barrysmyth: How did I miss this? Great resou rce of code and examples. The Python Graph Gallery #python #DataScience https://t.co/s02A...
- 2017-10-11 12:24:30 Werkzaamheden aan Python #Efteling https://t.co/

KevwHaC9rH

- 2017-10-11 12:24:25 @kalafina8 美味いけど仕方ないね...
- 2017-10-11 12:24:12 RT @taylorndean: what would u rather have, 10,00 0 for a Louis Vuitton dead python purse or 2,000 for my living python Louis https://t.co/CL...
- 2017-10-11 12:24:08 3 Python web scrapers and crawlers https://t.co/KMdUBYWMDp via @opensourceway by @jehb #webscraping #python
- 2017-10-11 12:23:56 Looking for Python projects to assist with. https://t.co/Xu0uuKE8YV
- Hello I am a self taught Python programmer. I am looking for any he l...
- 2017-10-11 12:23:47 RT @rochacbruno: Preciso de ajuda para manter a lib #python @pagseguro algumas issues abertas e o @pagseguro não dá nenhuma ajuda https://t...
- 2017-10-11 12:23:46 Hire a Python Developer by mp2consult https://t.co/amg0zD4mkL
- 2017-10-11 12:23:29 @ashemedai @dstufft @ncoghlan_dev @DRMacIver @hy nek @tintvrtkovic i.e. a python app shipped with PEX that uses twis... https://t.co/fEuXE6pbUy
- 2017-10-11 12:23:27 RT @itzskizy: Nigerian Rappers tempted to diss B uhari but are conscious of python dance like... https://t.co/n1pHwg5 XEt
- 2017-10-11 12:23:14 RT @MBNjobs: Modelling Manager-Customer Contact Bristol https://t.co/KVHmIu6JuK #Modelling #SAS #SQL #R #Python # Analytics #BigData #NoSQ...
- 2017-10-11 12:23:08 Anyone going to see Monty Python Spamalot @gohbe lfast this week? Come join us for our pre-theatre menu.... https://t.co/BCSrsWv08g
- 2017-10-11 12:23:02 Keen coder? Don't miss DigiCode Club tomorrow 4-6pm #Java #Ruby #Python #programming #techtalk @csmcr @UoMLibrary... h ttps://t.co/HFRVit730l
- 2017-10-11 12:23:01 RT @jodirowley: Some lovely #reptiles seen durin g frog surveys on the Northern Tablelands of NSW- Carpet Python (Mor elia spilota) & amp; Pink-to...
- 2017-10-11 12:22:54 RT @MBNjobs: Modelling Manager-Customer Contact Bristol https://t.co/KVHmIu6JuK #Modelling #SAS #SQL #R #Python # Analytics #BigData #NoSQ...
- 2017-10-11 12:22:52 RT @jrhunt: Relevant for my #AWS talk @Serverles sConf tomorrow: we just open sourced the python x-ray sdk https://t.co/3mc93fbyQz
- 2017-10-11 12:22:48 @ashemedai @dstufft @ncoghlan_dev @DRMacIver @hy nek @tintvrtkovic if you have a smart linker you can only use parts... https://t.co/h2yzMNrqq6
- 2017-10-11 12:22:33 @hawkieowl @ashemedai @ncoghlan dev @DRMacIver @

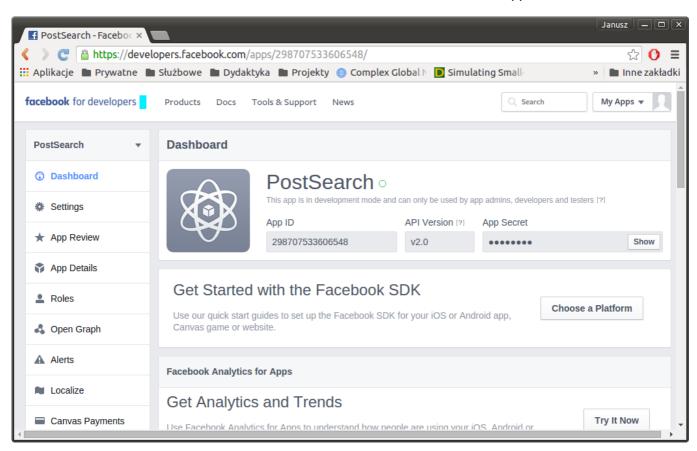
- hynek @tintvrtkovic E.g. you can have Python Core (the interpreter... https://t.co/U41z5lfsRx
- 2017-10-11 12:22:20 Colt Python Double Action Revolver https://t.co/9DC4dv00UE
- 2017-10-11 12:22:02 Análisis de RAID: World War II. Ni los Monty Pyt hon al completo podrían salvar este desaguisado https://t.co/KBBAbUy Phg
- 2017-10-11 12:21:53 @ashemedai @dstufft @ncoghlan_dev @DRMacIver @hy nek @tintvrtkovic the thing about such stdlibs is that they are com... https://t.co/C9nb6Ma7oJ
- 2017-10-11 12:21:39 RT @Timthadon01: Python Dance: Nigerian Army pla nted bombs in Nnamdi Kanu's house IPOB Alleges https://t.co/emV8Wm5jyz https://t.co/BAwC1...
- 2017-10-11 12:21:32 RT @PythonIreland: Retweet this for a chance to win 1 of 2 free tickets for PyConIE 2017 (closing date : Sat 14th 0 ct Noon) #Python https:/...
- 2017-10-11 12:21:15 Depths of an emotional Swamp https://t.co/fZhv7hc80P #python #code #data #datascience #pandas #learn
- 2017-10-11 12:21:03 @python_klfn 特に皮は中性脂肪の大敵なのですよ@@@@
- 2017-10-11 12:20:49 RT @KirkDBorne: 6 Easy Steps to Learn Naive Baye s Algorithm (with #Python) https://t.co/ui8NHHTAly #abdsc #BigData # DataScience #Statistics...
- 2017-10-11 12:20:44 RT @KirkDBorne: #Python vs R: Which programming language is better for #DataScience? https://t.co/0ByswCIsNh #BigData #MachineLearning #Rst...
- 2017-10-11 12:20:40 RT @4orgexcellence: Here's how #bigdata can help your business!
- #analytics #center4oe #iot #datascience #python #hadoop #marketing #
 cmo #c...
- 2017-10-11 12:20:39 SHOCKING: Worlds Largest Snake 49ft. Indonesian Reticulated Python. https://t.co/0KwHErgFec via @YouTube
- 2017-10-11 12:20:24 RT @ahmedjr_16: 7 Best #Programming Language Courses
- https://t.co/m6B7CD0qxj
- #SmallBusiness #Programmer #IIoT #Python #Java #HTML5 #tec...
- 2017-10-11 12:20:20 RT @JohnSnowLabs: Credible Sources of Accurate I nformation About #AI https://t.co/M6QzFZBXNo #datascience #Python #machinelearning #IoT #...
- 2017-10-11 12:20:19 RT @itzskizy: Nigerian Rappers tempted to diss B uhari but are conscious of python dance like... https://t.co/nlpHwg5XEt
- 2017-10-11 12:20:13 R And Python: Two Growing Languages https://

- t.co/7HjKTP8HD8 #CuratedSQL
- 2017-10-11 12:20:12 Checking out "6 Easy Steps to Learn Naive Bayes #Algorithm (with code in #Python" on Data Science Central:... https://t.co/oBnJyaUwhP
- 2017-10-11 12:20:00 @kalafina8 そういえばそんなお話お聞きしたような... 早く制限とれるといいですね♪
- 2017-10-11 12:20:00 csvfilecleaner 0.5: A very simple utility class for cleaning csv files https://t.co/ksr6m0VaTh
- 2017-10-11 12:19:59 RT @torysmasher: The tea lady has been detained Jackie detained for unlawfully making tea, another Monty Python moment from North Yorkshire...
- 2017-10-11 12:19:50 RT @NkirukaNistoran: BREAKING!! Over 30 Soldiers Killed, Time For The ''Python'' To Dance In The North And Leave The South-East(Graphic ... h...
- 2017-10-11 12:19:39 RT @4orgexcellence: Here's how #bigdata can help your business!
- #analytics #center4oe #iot #datascience #python #hadoop #marketing #
 cmo #c...
- 2017-10-11 12:19:30 RT @yutakashino: The Python Graph Gallery https://t.co/BZOUrpDOs2 Pythonにおけるプロット大全. おお,これは素晴らしい. https://t.co/3EzY3LOdMC
- 2017-10-11 12:19:11 @davidfrum @Lawrence Did you say Monty Python sc enario near the end because now I'm worried about killer rabbits?
- 2017-10-11 12:18:53 RT @JohnSnowLabs: Credible Sources of Accurate I nformation About #AI https://t.co/M6QzFZBXNo #datascience #Python #machinelearning #IoT #...
- 2017-10-11 12:18:51 RT @4orgexcellence: Here's how #bigdata can help your business!
- #analytics #center4oe #iot #datascience #python #hadoop #marketing #
 cmo #c...
- 2017-10-11 12:18:49 Reticulated Python/Anaconda Comparison https://t.co/N9akcFN0e4 via @YouTube
- 2017-10-11 12:18:46 Senior Manager, Software Engineering at Comcast (@comcastcareers) [New York, NY] https://t.co/FnhrhuxocR #python
- 2017-10-11 12:18:43 Check out new e-book 'Learn #Python in One Week' from one of our members @trustonailende: https://t.co/nnde5HtgGf... https://t.co/wCXPgFzOGn
- 2017-10-11 12:18:43 人生を楽にするにはPythonを学ぶしかねぇ!
- 2017-10-11 12:18:38 @makoto0218ne56 Python強いですね笑
- 2017-10-11 12:18:19 Job Senior Python Developer Bengaluru, Hyder abad, Mumbai Squareroot Consulting Pvt Ltd. 3 7 Years of https://t.co/LA7Ho9pNLR

2017-10-11 12:18:11 Erstmal meinen Downloads-Ordner mit nem Python-Skript aufgeräumt.

Facebook

As in the case of Twitter, we need a Facebook account. Then, we create a new application:



In [9]:

```
app_id = "XXXX"
app_secret = "XXXX"
access_token = app_id + "|" + app_secret
```

In [10]:

```
page_id = 'wfiauwr'
```

In [11]:

```
import json
import requests

# we want to access the WFA UWr fanpage
# we build the address
base = "https://graph.facebook.com/v2.4"
node = "/" + page_id
parameters = "/?access_token=%s" % access_token
url = base + node + parameters

# and download the data
r = requests.get(url)
print(r.content)
```

b'{"name":"Wydzia\\u0142 Fizyki i Astronomii Uniwersytetu Wroc\\u014 2awskiego","id":"687027717991014"}'

In [21]:

```
#we are looking for posts
base = "https://graph.facebook.com/v2.4"
node = "/" + page_id + "/feed"  # "/feed" added
parameters = "/?access_token=%s" % access_token
url = base + node + parameters

# we download the data
import codecs

r = requests.get(url)
result = r.json()

for i in result["data"]:
    print('-'*20)
    print(i.get(u'message'))
    print()
```

Praca dla fizyków we Wrocławiu (od naszego absolwenta):

Job: Application Scientist — Single Crystal X-ray Diffraction

Location: Wrocław, Poland Salary: Competitive package Closing Date: October 16, 2017

Description:

Rigaku Oxford Diffraction is an international manufacturer of scient ific instrumentation priding itself on being a customer relationship focused business whilst offering the best in engineering and softwar e solutions. The advertised role will be based at Rigaku Polska, loc ated in Wrocław. Poland.

In your role as an application scientist, you will be responsible for training users worldwide on the operation of their instruments and conducting demonstrations at our new demo facility in Poland. Your a bility to produce high quality scientific reports tailored to the exacting standards of customers will be a key part of your job. As an application scientist you will be required to communicate our message to other scientists by way of presenting at conferences and user meetings. The ability for you to travel both in Europe and outside of Europe is essential and it is estimated that travel will be no more than 50% of the time.

You will become the face of Rigaku Oxford Diffraction, promoting thr ough science our single crystal X-ray diffraction equipment. Your ab ility to share with others your enthusiasm will be advantageous to t he business by forging bonds with scientists. Your knowledge of the current market situation for single crystal X-ray diffraction will be beneficial in helping to understand how to engage customers.

Main Duties:

- Working with the sales team to conduct demonstrations for cus tomers
- Produce reports in a timely fashion after the demonstration
- Training users on software and how to best operate the hardware
- Maintaining the instrument(s) in the laboratory
- Maintaining the user forum and responding to queries
- Attending and presenting at conferences and user meetings worldwide
- Providing feedback to the software team
- Assisting with marketing activities
- Managing to completion smaller individual projects

Experience and Skills:

You will have good knowledge of using X-ray crystallography as a research tool and be able to engage customers in a positive and enthusi astic manner. Attention to detail is critical for accurate reporting of scientific facts.

Essential

- Ph.D. in a relevant scientific field
- Experience of operating single crystal X-ray diffraction inst ruments

- Skilled in data processing and structure solution of single c rystal diffraction data
- Excellent people skills to interact with both colleagues and customers
- Accuracy in delivering reports
- Written and spoken fluency in English
- Ability to travel up to 50% of the time

Desirable

- Good knowledge of chemical crystallography preferred
- Experience of both protein and chemical crystallography would be beneficial
- Experience of maintaining single crystal X-ray diffraction in struments
- Proven experience of presenting to a scientific audience
- Written and spoken fluency in German
- Polish language skills would be an advantage

Benefits

- Employment contract in a stable, international company
- Attractive social benefits package (private medical care, B enefit Sport card, group insurance)
- Motivational salary system

Please submit your cover letter and résumé in PDF only to HR@RIGAKU. COM.

http://www.rigaku.com/en/rigakuoxford

Artykuł Prof. Bernarda Jancewicza, emerytowanego pracownika Wydziału Fizyki i Astronomii o ostatnim Zjeździe Fizyków Polskich. Zapraszamy do lektury:

https://wszystkoconajwazniejsze.pl/prof-bernard-jancewicz-laureaci-n
agrody-nobla-we-wroclawiu/

Uniwersytet Wrocławski Gazeta Wrocławska

Garść inspiracji do Bulwaru Fizyków (projekt WBO na nadodrzu).

Głosujemy do 2. października na stronie www.wroclaw.pl/wbo

Nasze projekty to:

Projekt #12 (ogólnomiejski za 1mln PLN)

Projekt #15 (lokalny za 250tys. PLN)

Projekt #263 (lokalny za 750tys. PLN)

(można zagłosować na wszystkie)

Też tam będziemy, a konkretniej Instytut Astronomiczny Uniwersytetu Wrocławskiego

Zapraszamy! http://worldspaceweek.pl/

Nasze działania w sprawie WBOBulwarFizykow wsparł CRZ Krzywy Komin, serdecznie pozdrawiamy i dziękujemy:)

Od naszego absolwenta dostaliśmy taką wiadomość:

"W następną sobotę (7 października) organizujemy wraz ze Stowarzysze niem Wrocławskich Użytkowników R

6-cio godzinne warsztaty wprowadzające do analizy danych w języku R.

Więcej informacji można znaleźć tutaj: https://stwur.github.io/STWU R//blog/eRementarz/

Pewnie część studentów WFiA, zarówno fizyki jak i ISSP mogłaby być t ym zainteresowana."

Serdecznie polecamy, to nie pierwsza impreza organizowana przez Kubę i naprawdę warto!

- MM

Mały remont pracowni na Wydział Fizyki i Astronomii Uniwersytetu Wro cławskiego już prawie gotowy na nowy semestr!

Akcja Miasto wsparła nasz projekt WBOBulwarFizykow, zapraszamy do głosowania! http://www.wroclaw.pl/wbo

Nasze numery projektów to: Projekt #12 (ogólnomiejski za 1mln PLN) Projekt #15 (lokalny za 250tys. PLN) Projekt #263 (lokalny za 750tys. PLN)

U nas #naprawdejestfajnie :-)

Kiermasz Fizyczny 2017 - zapraszamy na Wydział Fizyki i Astronomii U niwersytetu Wrocławskiego jeszcze 26.09 od godz. 9:30-14:30. Pl. M. Borna 9

Kilka słów o projekcie Bulwaru Fizyków. Trochę szczegółów nt. wizji tego miejsca i planów. Zapraszamy do głosowania w internecie: htt p://www.wroclaw.pl/wbo lub w dowolnym UM Wrocławia. Nasze projekty t o:

#12 (projekt ogólnomiejski)
#15 (projekt lokalny, mały)
#263 (projekt lokalny, duży)

Już od wielu wielu lat (sprecyzujemy - od 20 lat!), na naszym Wydzia le odbywa się Cyrk Fizyczny w ramach którego prezentujemy doświadcze nia i pokazy zjawisk fizycznych. W tym roku przygotowaliśmy jubileus zowy pokaz pt. 'The best of..'. W ramach DFN mamy premierę tegoroczn ej edycji, ale zapraszamy przez cały rok! Serdecznie zapraszamy do k ontaktu z nami i zapisania się na ten pokaz. Nie prezentujemy zbyt d użo zdjęć by nie zdracać "co będzie" :-)

Już po raz kolejny odbywa się u nas Kiermasz Fizyczny w ramach Dolno śląskiego Festiwalu Nauki. W tym roku mamy dla Was wiele atrakcji, z apraszamy w godzinach 9-14 w dniach 26-27 września!

Dzisiaj o godzinie 12:00 ruszyła rejestracja na Ogólnopolską Sesję K ół Naukowych Fizyków. Serdecznie Zapraszamy do zapisów! ------

Dzisiaj kolejne wydarzenia Dolnośląskiego Festiwalu Nauki.

Pogoda nie zachęca do spacerowania, za to my zachęcamy do spędzenia popołudnia z fizyką. Zapraszamy na sobotnie wykłady w budynku Wydz iału na pl. Maksa Borna 9.

O godzinie 14, "Mniejsze, gęstsze, gorętsze, jaśniejsze - przepis na najgorętszą substancję we Wszechświecie (Smaller, denser, hotter, br ighter - making the hottest stuff in the Universe)" - dr Pasi Huovin en (wykład po angielsku).

O godzinie 15, "Bardzo zimne i bardzo gorące silnie sprzężone ciecze kwantowe (Strongly coupled quantum fluids - ultra cold and super ho t)", - dr Marcus Bluhm (wykład po angielsku).

Natomiast o godzinie 16 wyjątkowo ciekawie zapowiadające się spojrze nie na przyrodę oczami fizyką, mgr Artur Rokosa - "Jak wysokie mogą być drzewa i co wspólnego ma krewetka z czajnikiem? Czyli o fizyce w świecie zwierząt i roślin".

Niestety wykład dra Tomasza Greczyło "Co Einstein powiedział swojemu fryzjerowi?", został odwołany. Niestety

Miło nam poinformować, że premierę Cyrku Fizycznego w niedzielę o 1 6:00 zobaczy na naszej sali komplet publiczności. Niestety sala nie jest z gumy, więc zamykamy listę osób na tą imprezę.

Dolnośląski Festiwal Nauki 2017 już rozpoczęty na naszym Wydziale. Właśnie trwa wykład dra Tomasza Greczyło "Pobiegaj z fizykiem".

Jeszcze dziś będzie można wziąć udział w wykładach:

"Matematyczne i fizyczne metody szyfrowania" dra hab. Marka Mozrzyma sa - godz. 14

"Maria Skłodowska-Curie i jej promieniotwórczy świat" dra hab. Jana Chojcana - godz.15

Już dziś inauguracja jubileuszowego XX Dolnośląskiego Festiwalu Nauki.

Z tej okazji, jak co roku, Nasz wydział przygotował bogatą ofertę. Zapraszamy serdecznie.

Program w PDF:

http://wfa.uni.wroc.pl/wp-content/uploads/2017/09/program.pdf

Edit:

W programie DFN Wydziału zaszły dwie zmiany. Wykłady:

dra Roberta Bryla "Jak powstaje światło? - czyli o fizycznych podsta wach świecenia" (piątek, godz. 16)

oraz

dra Tomasza Greczyło "Co Einstein powiedział swojemu fryzjerowi?" (s obota, godz. 12)

ZOSTAŁY ODWOŁANE. Przepraszamy.

Halo halo, mamy ulotki, dużo ulotek Bulwaru Fizyków, zapraszam do p. 521 - MM

Projekt otwartego i darmowego dla wszystkich Bulwaru Fizyków we Wroc

ławiu wraca! Zapraszamy do głosowania na nasz projekt WBO w 2017 rok u. Namawiajcie znajomych z całego miasta, to m.in. projekt ogólnomie jski. Nasze numery projektów to:

#12 - projekt ogólnomiejski za 1mln PLN
#15 - projekt lokalny za 250tys. PLN
#263 - projekt lokalny za 750tys. PLN

Głosowanie: www.wroclaw.pl/wbo lub w wybranych jednostkach UM Wrocła wia.

Jak widać z ulotki poniżej plan minimum to doświadczenia podobne do tych ze świetnego parku im. S. Lema w Krakowie. Poza tym chcielibyś my zwierciadła akustyczne. Projekt bardzo rozwojowy, mamy pomysły i plan na dużo więcej - zapraszamy do głosowania!

#wbo2017 #bulwarfizykow WB0BulwarFizykow Wrocławski Budżet Obywatels ki Uniwersytet Wrocławski

mam ciekawe pytanie nad ktorym pracuje.

Fakty:

- -grawitacja to jedno z oddziaływań podstawowych
- -to relatywnie słaba siła zależna od masy
- -poza galaktyką nikoma ilosc materii w przestrzeni
- -brak oporu lub jest bardzo słaby
- -nie można wykryć czarne materii.
- wszechświat się stałe rozszerza

Teza:

brak czarnej materii. Jeśli powrócić do scholastycznej metody analiz y zjawisk przyjmując, że duże obiekty jak słońce wywierają wpływ na tak duże oraz odległe obiekty jak ziemia, mars czy pluton to czarna dziura o masie kilku miliardów słońc powinna mieć poza zasysaniem ta kże możliwość wytyczania stałej orbity gromadom gwiazd. Jeśli jedno cześnie przyjąć, że teoria wielkiego wybuchu jest prawdziwa to sform ułowanie Einsteina, że czarne dziury są aberracja jest bezpodstawne, gdyż to jedyne siły wokół których po wybuchu znów skupiła się materi a. A jeśli isnieje ciemna materia to jest ona w rejonie galaktyk sku tecznie niwelowana przez siły grawitacje ze środka drogi mlecznej. Ktoś mi wytlumaczy jak i gdzie moge sie mylić?

Mamy już pierwsze, jeszcze bardzo koncepcyjne plany jak mógłby wyglą dać bulwar po instalacji urządzeń. To pewnie nie do końca będzie tak wyglądało, ale przedstawiamy, bo dobrze oddaje ducha zmian i kierune k w jakim chcemy iść. Przypominamy, Bulwar Fizyków to trzy projekty:

#12 - projekt ogólnomiejski za 1mln PLN
#15 - projekt lokalny za 250tys. PLN
#263 - projekt lokalny za 750tys. PLN

Głosowanie trwa w internecie: www.wroclaw.pl/wbo lub w wybranych jed nostkach UM Wrocławia.

Zapraszamy do głosowania już teraz!

A za wizualizacje dziękujemy zaprzyjaźnionej architekt - AK :) Koło Naukowe Fizyków "Migacz" Fizyka Fizyk wyjaśnia Fizyka i Matemat yka w Pęgowie Fizyk w podróży Wydział Fizyki Uniwersytetu Warszawski ego Zapytaj Fizyka Fizyka Dla Każdego Fizyka z klasą Wydział Podstaw owych Problemów Techniki (WPPT) Politechnika Wrocławska Politechnika Wrocławska Uniwersytet Wrocławski Wydział Chemii Uniwersytetu Wrocła wskiego Wydział Informatyki i Zarządzania PWr Wydział Matematyki, Fi zyki i Informatyki UMCS Centrum Nauki Kopernik

Żegnaj Cassini

Projekt Bulwar Fizyków wraca! Zapraszamy do głosowania na nasz projekt WBO w 2017 roku. Namawiajcie znajomych z całego miasta, to m.in. projekt ogólnomiejski. Nasze numery projektów to:

#12 - projekt ogólnomiejski za 1mln PLN
#15 - projekt lokalny za 250tys. PLN
#263 - projekt lokalny za 750tys. PLN

Głosowanie: www.wroclaw.pl/wbo lub w wybranych jednostkach UM Wrocławia.

Jak widać z ulotki poniżej plan minimum to doświadczenia podobne do tych ze świetnego parku im. S. Lema w Krakowie. Poza tym chcielibyś my zwierciadła akustyczne. Projekt bardzo rozwojowy, mamy pomysły i plan na dużo więcej - zapraszamy do głosowania!

#wbo2017 #bulwarfizykow
Co tam na niebie?
Ważne - o stypendiach.

RSS feed and feedparser

From Wikipedia:

RSS – a type of web feed which allows users to access updates to online content in a standardized, computer-readable format. These feeds can, for example, allow a user to keep track of many different websites in a single news aggregator. The news aggregator will automatically check the RSS feed for new content, allowing the content to be automatically passed from website to website or from website to user.

Feed aggregators:

- news readers (including email clients)
- · internet services
- web browsers.

Example - TVN24 news

In [22]:

```
import feedparser
from time import strftime

d = feedparser.parse('http://www.tvn24.pl/najnowsze.xml')

print("-"*40)
print(d.feed.title)
print(d.feed.link)
print("-"*40)
print("-"*40)
print()

for e in d.entries:
    print(e.published, e.title)
    print(e.link)
    print()
```

dy mnie nie podrywał. Zawsze będę mu "grateful"

TVN24.pl - Wiadomości - Najnowsze

https://www.tvn24.pl/

Fri, 29 Sep 17 11:26:05 +0200 Joanna Krupa dla TVN24 BiS: Hefner nig

https://www.tvn24.pl/wiadomosci-ze-swiata,2/joanna-krupa-hugh-hefner-nigdy-mnie-nie-podrywal,777015.html

Fri, 29 Sep 17 11:00:48 +0200 "Może być problem, żeby dostać się do lekarzy". Rewolucja na SOR-ach

https://www.tvn24.pl/wiadomosci-z-kraju,3/rewolucja-na-sor-moze-byc-problem-zeby-dostac-sie-do-lekarzy,776982.html

Fri, 29 Sep 17 10:45:42 +0200 Maria w natarciu. Czy zagrozi Wyspom B rytyjskim?

http://tvnmeteo.tvn24.pl/informacje/prognoza,45/maria-w-natarciu-moze-zagrozic-wyspom-brytyjskim,242238,1,0.html

Fri, 29 Sep 17 10:00:13 +0200 Centymetry od tragedii. Wjechał na tor y i machał do maszynisty pendolino

http://tvnwarszawa.tvn24.pl/informacje,news,centymetry-od-tragedii-wjechal-na-tory-i-machal-do-maszynisty-pendolino,242291.html

Fri, 29 Sep 17 08:39:16 +0200 Łabędzi śpiew państwa dżihadu. Zaskaku jąca ofensywa na Rosjan

https://www.tvn24.pl/wiadomosci-ze-swiata,2/ofensywa-dzihadystow-w-syrii-pierwszy-sukces-od-miesiecy,776962.html

Thu, 28 Sep 17 19:09:13 +0200 Książę Harry i dwulatka. Wspólne kibic owanie, wspólny popcorn

https://www.tvn24.pl/ciekawostki-michalki,5/popcorn-ksiecia-harry-eg o-i-2-letnia-emily-zawody-invictus,776852.html

Thu, 28 Sep 17 17:00:55 +0200 Z Wiednia do Warszawy przez Brzeszcze. "A po co? Jak ja do Cieszyna nie mogę się dostać"

https://www.tvn24.pl/katowice,51/z-wiednia-do-warszawy-przez-brzeszc ze-mieszkancy-sceptyczni,776785.html

Thu, 28 Sep 17 15:32:12 +0200 Kopali pod gazociąg, trafili na szkiel ety.

"To masowy grób żołnierzy niemieckich"

https://www.tvn24.pl/poznan,43/robili-wykop-trafili-na-szkielety-to-masowy-grob-zolnierzy,776783.html

Thu, 28 Sep 17 13:33:00 +0200 "Usiłował zabić dwóch policjantów".

Śledztwo przejęła Prokuratura Krajowa

https://www.tvn24.pl/pomorze,42/prokuratura-przejmuje-sprawe-strzela niny-w-inowroclawiu,776738.html

Thu, 28 Sep 17 11:36:00 +0200 Google chce uniknąć gigantycznych kar. Szykuje zmiany w wyszukiwarce

http://tvn24bis.pl/redirect/google-ugina-sie-pod-presja-ke-i-wprowad za-zmiany-aby-uniknac-kar,776697.html

Wed, 27 Sep 17 13:23:00 +0200 Motorniczy, który zabił trzy osoby wra ca za kraty. Sąd nie złagodził wyroku

https://www.tvn24.pl/lodz,69/sad-apelacyjny-podtrzymal-wyrok-motorniczemu-13-5-roku-wiezienia,776390.html

Sat, 23 Sep 17 22:20:35 +0200 Andrzej Godlewski i prof. Krzysztof Mi szczak

https://www.tvn24.pl/horyzont,46,m/wybory-w-niemczech-andrzej-godlewski-i-krzysztof-miszczak-horyzont,775514.html

Wed, 20 Sep 17 12:08:02 +0200 Tomasz Fiałkowski (przewodniczący) https://www.tvn24.pl/nagroda-literacka-nike-2017,159,m/jury,366/tomasz-fialkowski,774466.html

Sat, 26 Aug 17 17:00:19 +0200 Nie żyje Grzegorz Miecugow https://www.tvn24.pl/wiadomosci-z-kraju,3/grzegorz-miecugow-nie-zyje,767507.html

Tue, 07 Mar 17 19:25:59 +0100 "Kwota nieprawdopodobnie kosmiczna". N owy rekord Wielkiej Orkiestry Świątecznej Pomocy https://www.tvn24.pl/wosp-2017/25-final-wosp-zebrano-ponad-105-mln-zl,721403.html

Example - NBP exchange rates

In [23]:

```
import feedparser
from time import strftime

d = feedparser.parse('http://rss.nbp.pl/kursy/TabelaA.xml')

print("-"*40)
print(d.feed.title)
print(d.feed.link)
print("-"*40)
print("-"*40)
print()

for e in d.entries:
    print(e.published, e.title)
    print(e.link)
    print()
```

NBP - Tabela A kursów średnich walut obcych

http://www.nbp.pl/home.aspx?f=/Statystyka/kursy.html

Fri, 29 Sep 2017 11:46:10 +0100 Tabela nr 189/A/NBP/2017 z dnia 2017 -09-29

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a189

Thu, 28 Sep 2017 11:48:14 +0100 Tabela nr 188/A/NBP/2017 z dnia 2017 -09-28

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a188

Wed, 27 Sep 2017 11:46:47 +0100 Tabela nr 187/A/NBP/2017 z dnia 2017 -09-27

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a187

Tue, 26 Sep 2017 11:46:05 +0100 Tabela nr 186/A/NBP/2017 z dnia 2017 -09-26

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a186

Mon, 25 Sep 2017 11:46:04 +0100 Tabela nr 185/A/NBP/2017 z dnia 2017 -09-25

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a185

Fri, 22 Sep 2017 11:46:10 +0100 Tabela nr 184/A/NBP/2017 z dnia 2017 -09-22

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a184

Thu, 21 Sep 2017 11:46:13 +0100 Tabela nr 183/A/NBP/2017 z dnia 2017 -09-21

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a183

Wed, 20 Sep 2017 11:47:12 +0100 Tabela nr 182/A/NBP/2017 z dnia 2017 -09-20

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a182

Tue, 19 Sep 2017 11:45:52 +0100 Tabela nr 181/A/NBP/2017 z dnia 2017 -09-19

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a181

Mon, 18 Sep 2017 11:46:28 +0100 Tabela nr 180/A/NBP/2017 z dnia 2017 -09-18

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a180

Fri, 15 Sep 2017 11:46:32 +0100 Tabela nr 179/A/NBP/2017 z dnia 2017 -09-15

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a179

Thu, 14 Sep 2017 11:45:57 +0100 Tabela nr 178/A/NBP/2017 z dnia 2017 -09-14

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a178

Wed, 13 Sep 2017 11:46:12 +0100 Tabela nr 177/A/NBP/2017 z dnia 2017 -09-13

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a177

Tue, 12 Sep 2017 11:46:10 +0100 Tabela nr 176/A/NBP/2017 z dnia 2017 -09-12

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a176

Mon, 11 Sep 2017 11:46:02 +0100 Tabela nr 175/A/NBP/2017 z dnia 2017 -09-11

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a175

Fri, 8 Sep 2017 11:46:04 +0100 Tabela nr 174/A/NBP/2017 z dnia 2017-09-08

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a174

Thu, 7 Sep 2017 11:46:03 +0100 Tabela nr 173/A/NBP/2017 z dnia 2017-09-07

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a173

Wed, 6 Sep 2017 11:46:39 +0100 Tabela nr 172/A/NBP/2017 z dnia 2017-09-06

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a172

Tue, 5 Sep 2017 11:45:39 +0100 Tabela nr 171/A/NBP/2017 z dnia 2017-09-05

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a171

Mon, 4 Sep 2017 11:45:47 +0100 Tabela nr 170/A/NBP/2017 z dnia 2017-09-04

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a170

Fri, 1 Sep 2017 11:45:57 +0100 Tabela nr 169/A/NBP/2017 z dnia 2017-09-01

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a169

Thu, 31 Aug 2017 11:45:51 +0100 Tabela nr 168/A/NBP/2017 z dnia 2017 -08-31

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a168

Wed, 30 Aug 2017 11:46:14 +0100 Tabela nr 167/A/NBP/2017 z dnia 2017 -08-30

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a167

Tue, 29 Aug 2017 11:45:35 +0100 Tabela nr 166/A/NBP/2017 z dnia 2017 -08-29

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a166

Mon, 28 Aug 2017 11:46:05 +0100 Tabela nr 165/A/NBP/2017 z dnia 2017 -08-28

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a165

In [24]:

link = d.entries[0].link
print(link)

http://rss.nbp.pl/kursy/TabRss.aspx?n=2017/a/17a189

In [25]:

import requests r = requests.get(link) print(r.content)

b'\r\n<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//E N">\r\n<html>\r\n<head>\r\n<meta http-equiv="Content-Type" content ="text/html; charset=iso-8859-2">\r\n<link rel="stylesheet" href="cs s/kw rss.css" type="text/css">\r\n<meta http-equiv="expires" content ="0">\r\n<meta http-equiv="cache-control" content="no-cache, must-re validate">\r\n<meta http-equiv="pragma" content="no-cache">\r\n<meta</pre> http-equiv="window-target" content=" top">\r\n<style type="text/cs"</pre> s">\r\n</style>\r\n</title>Archiwum kurs\xf3w walut</title>\r\n</head >\r\n<body>\r\n<div align="center" width="100%">\r\n 0" cellpadding="0" cellspacing="0" border="0">\r\n\r\n\r\n\r \n\r\n
\r\n ng="0" border="0" bgcolor="#ffffff">\r\n\r\n< br />\r\n 0">\r\nTabela A kurs\xf3w \xb6rednich wal ut obcych\r\nTabela nr 189/A/NB P/2017 z dnia $2017-09-29 \r\n\r\n
\r\n$ ellpadding="4" cellspacing="0" border="0" width="400" class="tabel a">\r\n<colgroup>\r\n<col width="190" />\r\n<col width="90" />\r\n<c ol width="80" />\r\n</colgroup>\r\n\r\n<th class="tabela ar0 lef t pad5 bold">Nazwa waluty\r\n<th class="tabela ar0 right pad5 b old">Kod waluty\r\nKurs \xb6redni\r\n\r\n\r\nbat (Tajlandia)\r\n1 THB\r \n0,1095\r\n\r\n\r\n<td c lass="tabela ar2 left">dolar ameryka\xf1ski\r\n a ar2 right">1 USD\r\n3,6519 \r\n\r\n\r\ndolar australi jski\r\n1 AUD\r\n<td cla ss="tabela ar1 right">2,8625\r\n\r\n\r\n<td class="tab ela ar2 left">dolar Hongkongu\r\n1 HKD\r\n0,4676\r\n \r\n\r\ndolar kanadyjski\r\n<td class="tabela ar1 right">1 CAD\r\n<td class="tabela ar1 ri ght">2,9383\r\n\r\n\r\ndol ar nowozelandzki\r\n1 NZD</td >\r\n2,6338\r\n\r\n\r\n<t</pre> d class="tabela ar1 left">dolar singapurski\r\n a arl right">1 SGD\r\n2,6888 \r\n\r\neuro\r\n< td class="tabela ar2 right">1 EUR\r\n<td class="tabela ar2 right">4,3091\r\n\r\nf orint (W\xeagry)\r\n100 HUF</ td>\r\n1,3846\r\n\r\n\r\n frank szwajcarski\r\n<td class="tab ela ar2 right">1 CHF\r\n3,761 9\r\n\r\n\r\nfunt szterlin g\r\n1 GBP\r\n ="tabela arl right">4,8842\r\n\r\n\r\n<td class="tabel a ar2 left">hrywna (Ukraina)\r\n1& nbsp;UAH\r\n0,1373\r\n\r \n\r\njen (Japonia)\r\n s="tabela arl right">100 JPY\r\n<td class="tabela arl right" t">3,2439\r\n\r\n\r\nkoron a czeska\r\n1 CZK\r\n<td class="tabela ar2 right">0,1655\r\n\r\n\r\nclass ="tabela ar1 left">korona du\xf1ska\r\n<td class="tabela ar1 ri ght">1 DKK\r\n0,5790\r\n \r\n\r\nkorona islandzka\r \n100 ISK\r\n a ar2 right">3,4597\r\n\r\n\r\n<td class="tabela ar1 l

eft">korona norweska\r\n1 NOK \r\n0,4594\r\n\r\n\r \nkorona szwedzka\r\n ela ar2 right">1 SEK\r\n0,449 2\r\n\r\n\r\nkuna (Chorwac ja)\r\n1 HRK\r\n $s="tabela ar1 right">0,5749\r\n\r\n\r\n$ la ar2 left">lej rumu\xf1ski\r\n1& nbsp;RON\r\n0,9368\r\n\r \n\r\nlew (Bu\xb3garia)\r\n<td</pre> class="tabela ar1 right">1 BGN\r\n ight">2,2032\r\n\r\n\r\nli ra turecka\r\n1 TRY\r\n< td class="tabela ar2 right">1,0269\r\n\r\n\r\n<td clas s="tabela ar1 left">nowy izraelski szekel\r\n<td class="tabela arl right">1 ILS\r\n1,0353</ td>\r\n\r\n\r\npeso chilijskie \r\n100 CLP\r\n ="tabela ar2 right">0,5745\r\n\r\n\r\n<td class="tabel a arl left">peso filipi\xf1skie\r\n<td class="tabela arl righ t">1 PHP\r\n0,0718\r\n</ tr>\r\n\r\npeso meksyka\xf1skie \r\n1 MXN\r\n<td class="tabel a ar2 right">0,2010\r\n\r\n\r\n<td class="tabela ar1 l eft">rand (Republika Po\xb3udniowej Afryki)\r\n<td class="tabel" a arl right">1 ZAR\r\n0,2708 \r\n\r\n<tt>\r\nreal (Brazyli a)\r\n1 BRL\r\n<td class ="tabela ar2 right">1,1476\r\n\r\n\r\n<td class="tabel a arl left">ringgit (Malezja)\r\n1 MYR\r\n0,8650\r\n \r\n\r\nrubel rosyjski\r\n lass="tabela ar2 right">1 RUB\r\n ht">0,0631\r\n\r\n\r\nrupi a indonezyjska
\r\n10000 IDR td>\r\n2,7115\r\n\r\n\r\n rupia indyjska\r\n ar2 right">100 INR\r\n5,5825 \r\n\r\nclass="tabela ar1 left">won po\xb3udni owokorea\xf1ski\r\n100 KRW</t d>\r\n0,3192\r\n\r\n\r\n< td class="tabela ar2 left">yuan renminbi (Chiny)\r\n<td class ="tabela ar2 right">1 CNY\r\n $0,5495\r\n\r\n\r\nSDR (MF)$ W)\r\n1 XDR\r\n<td class ="tabela arl right">5,1649\r\n\r\n\r\n

 \r\n\r\n\r\n<table border="0" cellspacing="0" cell padding="10" width="400">\r\n<sp</pre> an class="copy">© 2017 NBP\r\n\r\n</di v>\r\n\r\n</body>\r\n</html>'

In [36]:

```
from bs4 import BeautifulSoup
import pandas as pd

soup = BeautifulSoup(r.content, "lxml")
cells = soup.find_all("td", class_="tabela")
curr = [cell.text for cell in cells]
curr_reshaped = list(zip(*[iter(curr)]*3))
ex = pd.DataFrame(curr_reshaped,columns=["Waluta", "Skrót", "Kurs"])
ex
```

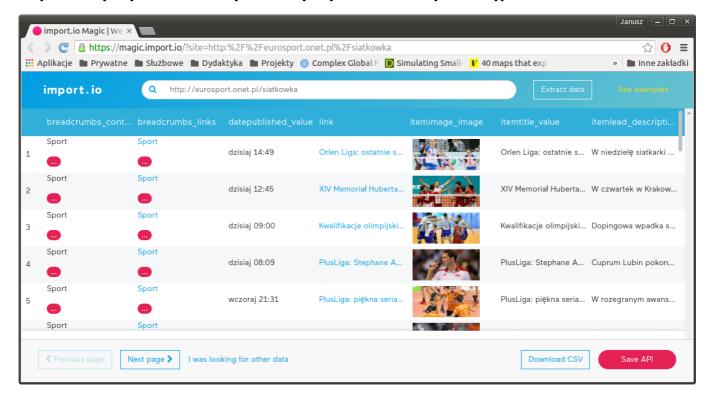
Out[36]:

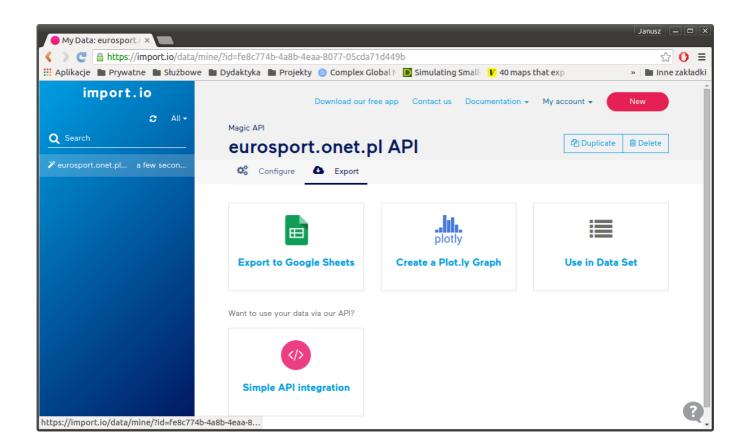
	Waluta	Skrót	Kurs
0	bat (Tajlandia)	1 THB	0,1095
1	dolar amerykański	1 USD	3,6519
2	dolar australijski	1 AUD	2,8625
3	dolar Hongkongu	1 HKD	0,4676
4	dolar kanadyjski	1 CAD	2,9383
5	dolar nowozelandzki	1 NZD	2,6338
6	dolar singapurski	1 SGD	2,6888
7	euro	1 EUR	4,3091
8	forint (Węgry)	100 HUF	1,3846
9	frank szwajcarski	1 CHF	3,7619
10	funt szterling	1 GBP	4,8842
11	hrywna (Ukraina)	1 UAH	0,1373
12	jen (Japonia)	100 JPY	3,2439
13	korona czeska	1 CZK	0,1655
14	korona duńska	1 DKK	0,5790
15	korona islandzka	100 ISK	3,4597
16	korona norweska	1 NOK	0,4594
17	korona szwedzka	1 SEK	0,4492
18	kuna (Chorwacja)	1 HRK	0,5749
19	lej rumuński	1 RON	0,9368
20	lew (Bułgaria)	1 BGN	2,2032
21	lira turecka	1 TRY	1,0269
22	nowy izraelski szekel	1 ILS	1,0353
23	peso chilijskie	100 CLP	0,5745
24	peso filipińskie	1 PHP	0,0718
25	peso meksykańskie	1 MXN	0,2010
26	rand (Republika Południowej Afryki)	1 ZAR	0,2708
27	real (Brazylia)	1 BRL	1,1476
28	ringgit (Malezja)	1 MYR	0,8650
29	rubel rosyjski	1 RUB	0,0631
30	rupia indonezyjska	10000 IDR	2,7115
31	rupia indyjska	100 INR	5,5825
32	won południowokoreański	100 KRW	0,3192
33	yuan renminbi (Chiny)	1 CNY	0,5495

	Waluta	Skrót	Kurs
34	SDR (MFW)	1 XDR	5,1649

Platforms for extracting data

import.io (https://www.import.io/ (https://www.import.io/))





In [37]:

```
import requests
import json

api = "https://api.import.io/store/connector/XXXXX/_query?input=webpage/url:htt
p%3A%2F%2Feurosport.onet.pl%2Fsiatkowka&&_apikey=XXXX"

r = requests.get(api)
results = r.json()
for i in results["results"]:
    print(i["link/_title"])
    print()
```

PlusLiga: PGE Skra Bełchatów zagra z uśmiechem o najwyższe cele

PlusLiga: Indykpol AZS Olsztyn chce zagrać w play off

Jastrzębski Węgiel: półfinał pierwszym celem

PlusLiga: MKS Będzin ma poprawić pozycję

PlusLiga: ZAKSA Kędzierzyn-Koźle z nowym trenerem i odmłodzonym zesp

ołem

PlusLiga: Dafi Społem Kielce ma cel minimum, czyli utrzymanie

PlusLiga: BBTS Bielsko-Biała powalczy o czołową ósemkę

PlusLiga: Cuprum Lubin spróbuje trafić w "szóstke"

PlusLiga: w Cerrad Czarnych Radom nastał czas młodzieży

PlusLiga: celem Asseco Resovii Rzeszów występ w finale

PlusLiga: Piotra Gruszkę i GKS Katowice czeka sezon ciekawszy i trud

niejszy

Jan Such przed startem PlusLigi: będzie ciekawsza

PlusLiga: mniej gwiazd, więcej zagranicznych trenerów

Wraca PlusLiga, osiemnasty sezon w elicie całkiem inny od poprzednic h

Lukas Tichacek: wszyscy patrzą na nas nieco inaczej

Mistrzostwa Europy w siatkówce (Gruzja i Azerbejdżan 2017) - termina rz, wyniki i tabele

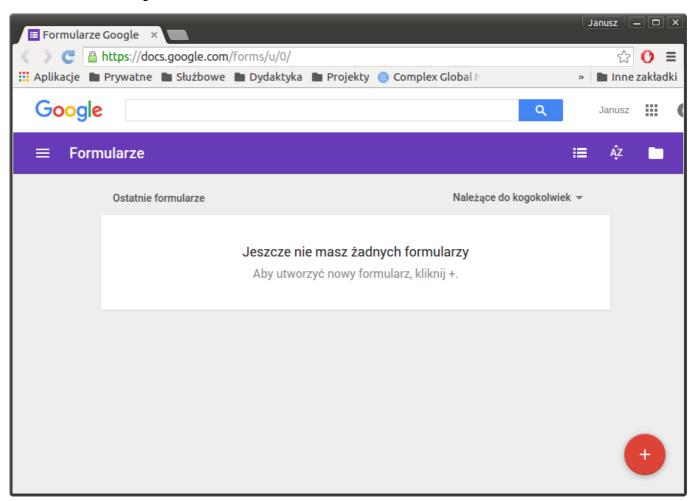
ME siatkarek 2017: Serbia idzie jak burza, Białoruś rozbita w ćwierć finale

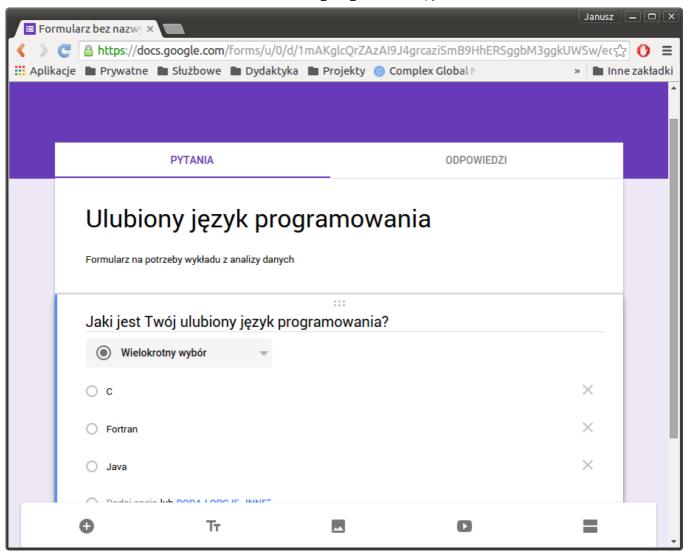
Tokio 2020: FIVB czeka na akceptację systemu kwalifikacji olimpijski ch w siatkówce

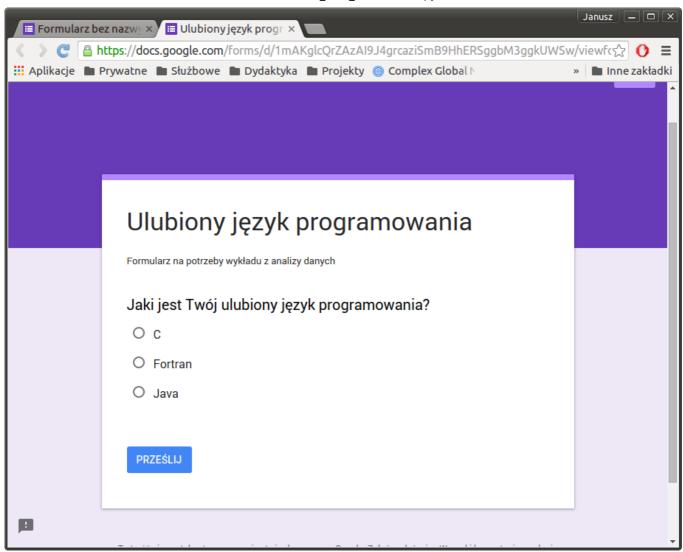
ME siatkarek 2017: Holandia rozbiła Włochy i wzięła półfinał rozgryw ek

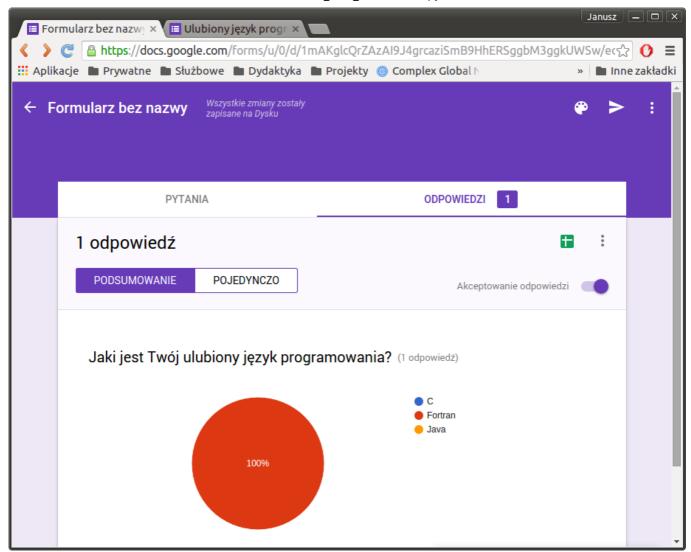
Malwina Smarzek: jest nam przykro, kiedy kibice stawiają na kadrze k rzyżyk

Online surveys









In []: