

```
In [1]: !pip install bs4
```

```
Requirement already satisfied: bs4 in c:\users\asharm\anaconda3\lib\site-packages (0.0.1)
Requirement already satisfied: beautifulsoup4 in c:\users\asharm\anaconda3\lib\site-packages (from bs4) (4.11.1)
Requirement already satisfied: soupsieve>1.2 in c:\users\asharm\anaconda3\lib\site-packages (from beautifulsoup4->bs4) (2.3.1)
```

```
In [2]: !pip install requests
```

```
Requirement already satisfied: requests in c:\users\asharm\anaconda3\lib\site-packages (2.27.1)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\asharm\anaconda3\lib\site-packages (from requests) (1.26.9)
Requirement already satisfied: idna<4,>=2.5 in c:\users\asharm\anaconda3\lib\site-packages (from requests) (3.3)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\asharm\anaconda3\lib\site-packages (from requests) (2.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\asharm\anaconda3\lib\site-packages (from requests) (2021.10.8)
```

```
In [3]: from bs4 import BeautifulSoup
```

```
In [4]: url = "https://scholar.google.com/citations?view_op=top_venues&hl=en"
```

```
In [5]: import requests
r = requests.get(url)
r
```

```
Out[5]: <Response [200]>
```

```
In [6]: soup = BeautifulSoup(r.content)
```

Rank

```
In [7]: Rank = []
```

```
In [8]: for i in soup.find_all("td", class_="gsc_mvt_p"):
Rank.append(i.text)
```

```
In [9]: Rank
```

```
Out[9]: ['1.',  
         '2.',  
         '3.',  
         '4.',  
         '5.',  
         '6.',  
         '7.',  
         '8.',  
         '9.',  
         '10.',  
         '11.',  
         '12.',  
         '13.',  
         '14.',  
         '15.',  
         '16.',  
         '17.',  
         '18.',  
         '19.',  
         '20.',  
         '21.',  
         '22.',  
         '23.',  
         '24.',  
         '25.',  
         '26.',  
         '27.',  
         '28.',  
         '29.',  
         '30.',  
         '31.',  
         '32.',  
         '33.',  
         '34.',  
         '35.',  
         '36.',  
         '37.',  
         '38.',  
         '39.',  
         '40.',  
         '41.',  
         '42.',  
         '43.',  
         '44.',  
         '45.',  
         '46.',  
         '47.',  
         '48.',  
         '49.',  
         '50.',  
         '51.',  
         '52.',  
         '53.',  
         '54.',  
         '55.',  
         '56.',  
         '57.',  
         '58.',  
         '59.',  
         '60.',
```

```
'61.',  
'62.',  
'63.',  
'64.',  
'65.',  
'66.',  
'67.',  
'68.',  
'69.',  
'70.',  
'71.',  
'72.',  
'73.',  
'74.',  
'75.',  
'76.',  
'77.',  
'78.',  
'79.',  
'80.',  
'81.',  
'82.',  
'83.',  
'84.',  
'85.',  
'86.',  
'87.',  
'88.',  
'89.',  
'90.',  
'91.',  
'92.',  
'93.',  
'94.',  
'95.',  
'96.',  
'97.',  
'98.',  
'99.',  
'100.']
```

```
In [10]: # publication
```

```
In [11]: Publication = []
```

```
In [12]: for i in soup.find_all("td",class_="gsc_mvt_t"):  
          Publication.append(i.text)
```

```
In [13]: Publication
```

```
Out[13]: ['Nature',
          'The New England Journal of Medicine',
          'Science',
          'IEEE/CVF Conference on Computer Vision and Pattern Recognition',
          'The Lancet',
          'Advanced Materials',
          'Nature Communications',
          'Cell',
          'International Conference on Learning Representations',
          'Neural Information Processing Systems',
          'JAMA',
          'Chemical Reviews',
          'Proceedings of the National Academy of Sciences',
          'Angewandte Chemie',
          'Chemical Society Reviews',
          'Journal of the American Chemical Society',
          'IEEE/CVF International Conference on Computer Vision',
          'Nucleic Acids Research',
          'International Conference on Machine Learning',
          'Nature Medicine',
          'Renewable and Sustainable Energy Reviews',
          'Science of The Total Environment',
          'Advanced Energy Materials',
          'Journal of Clinical Oncology',
          'ACS Nano',
          'Journal of Cleaner Production',
          'Advanced Functional Materials',
          'Physical Review Letters',
          'Scientific Reports',
          'The Lancet Oncology',
          'Energy & Environmental Science',
          'IEEE Access',
          'PLOS ONE',
          'Science Advances',
          'Journal of the American College of Cardiology',
          'Applied Catalysis B: Environmental',
          'Nature Genetics',
          'BMJ',
          'Circulation',
          'European Conference on Computer Vision',
          'International Journal of Molecular Sciences',
          'Nature Materials',
          'Chemical engineering journal',
          'AAAI Conference on Artificial Intelligence',
          'Journal of Materials Chemistry A',
          'ACS Applied Materials & Interfaces',
          'Nature Biotechnology',
          'The Lancet Infectious Diseases',
          'Frontiers in Immunology',
          'Applied Energy',
          'Nano Energy',
          'Nature Energy',
          'Meeting of the Association for Computational Linguistics (ACL)',
          'The Astrophysical Journal',
          'Gastroenterology',
          'Nature Methods',
          'IEEE Transactions on Pattern Analysis and Machine Intelligence',
          'Cochrane Database of Systematic Reviews',
          'Blood',
          'Neuron',
```

```

'Nano Letters',
'Morbidity and Mortality Weekly Report',
'European Heart Journal',
'Nature Nanotechnology',
'ACS Catalysis',
'Nature Neuroscience',
'American Economic Review',
'Journal of High Energy Physics',
'IEEE Communications Surveys & Tutorials',
'Annals of Oncology',
'Nutrients',
'Accounts of Chemical Research',
'Immunity',
'Environmental Science & Technology',
'Nature Reviews. Molecular Cell Biology',
'Gut',
'Physical Review D',
'ACS Energy Letters',
'Monthly Notices of the Royal Astronomical Society',
'Conference on Empirical Methods in Natural Language Processing (EMNLP)',
'Clinical Infectious Diseases',
'Cell Metabolism',
'Nature Reviews Immunology',
'Joule',
'Nature Photonics',
'International Journal of Environmental Research and Public Health',
'Environmental Pollution',
'Computers in Human Behavior',
'Frontiers in Microbiology',
'Nature Physics',
'Small',
'Cell Reports',
'Molecular Cell',
'Clinical Cancer Research',
'Bioresource Technology',
'Journal of Business Research',
'Molecular Cancer',
'Sensors',
'Nature Climate Change',
'IEEE Internet of Things Journal']

```

```
In [14]: # h5-index
```

```
In [35]: h5_index = []
```

```
In [36]: for i in soup.find_all("a",class_ = "gs_ibl gsc_mp_anchor"):
          h5_index.append(i.text)
```

```
In [37]: h5_index
```

```
Out[37]: ['444',  
          '432',  
          '401',  
          '389',  
          '354',  
          '312',  
          '307',  
          '300',  
          '286',  
          '278',  
          '267',  
          '265',  
          '256',  
          '245',  
          '244',  
          '242',  
          '239',  
          '238',  
          '237',  
          '235',  
          '227',  
          '225',  
          '220',  
          '213',  
          '211',  
          '211',  
          '210',  
          '207',  
          '206',  
          '202',  
          '202',  
          '200',  
          '198',  
          '197',  
          '195',  
          '192',  
          '191',  
          '190',  
          '189',  
          '186',  
          '183',  
          '181',  
          '181',  
          '180',  
          '178',  
          '177',  
          '175',  
          '173',  
          '173',  
          '173',  
          '172',  
          '170',  
          '169',  
          '167',  
          '166',  
          '165',  
          '165',  
          '165',  
          '165',  
          '164',
```

```
'164',  
'163',  
'163',  
'163',  
'163',  
'162',  
'160',  
'160',  
'159',  
'159',  
'159',  
'159',  
'158',  
'158',  
'155',  
'155',  
'155',  
'155',  
'155',  
'154',  
'153',  
'153',  
'152',  
'152',  
'152',  
'152',  
'152',  
'152',  
'151',  
'151',  
'150',  
'149',  
'149',  
'146',  
'146',  
'145',  
'145',  
'145',  
'144',  
'144']
```

```
In [39]: # h5-median
```

```
In [41]: h5_median = []
```

```
In [43]: for i in soup.find_all("span", class_="gs_ibl gsc_mp_anchor"):  
         h5_median.append(i.text)
```

```
In [44]: h5_median
```

```
Out[44]: ['667',  
          '780',  
          '614',  
          '627',  
          '635',  
          '418',  
          '428',  
          '505',  
          '533',  
          '436',  
          '425',  
          '444',  
          '364',  
          '332',  
          '386',  
          '344',  
          '415',  
          '550',  
          '421',  
          '389',  
          '324',  
          '311',  
          '300',  
          '315',  
          '277',  
          '273',  
          '280',  
          '294',  
          '274',  
          '329',  
          '290',  
          '303',  
          '278',  
          '294',  
          '276',  
          '246',  
          '297',  
          '307',  
          '301',  
          '321',  
          '253',  
          '265',  
          '224',  
          '296',  
          '220',  
          '223',  
          '315',  
          '296',  
          '228',  
          '217',  
          '232',  
          '314',  
          '304',  
          '234',  
          '254',  
          '296',  
          '293',  
          '243',  
          '229',  
          '231',
```



```
'207',  
'302',  
'265',  
'264',  
'220',  
'248',  
'263',  
'220',  
'304',  
'243',  
'214',  
'211',  
'242',  
'214',  
'340',  
'235',  
'217',  
'212',  
'194',  
'249',  
'278',  
'211',  
'292',  
'233',  
'228',  
'225',  
'222',  
'214',  
'225',  
'222',  
'196',  
'205',  
'202',  
'201',  
'190',  
'233',  
'209',  
'201',  
'228',  
'212']
```

```
In [45]: len(Rank)
```

```
Out[45]: 100
```

```
In [46]: len(Publication)
```

```
Out[46]: 100
```

```
In [47]: len(h5_index)
```

```
Out[47]: 100
```

```
In [48]: len(h5_median)
```

```
Out[48]: 100
```

```
In [ ]:
```

