In [1]:

```
# import sys
# !{sys.executable} -m pip install beautifulsoup4 pandas requests tqdm
```

In [20]:

```
from bs4 import BeautifulSoup
from tqdm import tqdm
import pandas as pd
import os, requests, time, urllib
```

In [21]:

```
# import books info and links
df = pd.read_excel('5_6134188041515827412-converted.xlsx', sheet_name='Table 1')
df.head()
```

Out[21]:

	S.No.	Book Title	Author	Edition	OpenURL	Downloaded	Remarks
0	1	Fundamentals of Power Electronics	Robert W. Erickson, Dragan\nMaksimovic	2nd ed.\n2001	http://link.springer.com/openurl?genre=book&is...	1	NaN
1	2	Handbook of the Life Course	Jeylan T. Mortimer, Michael J.\nShanahan	2003	http://link.springer.com/openurl?genre=book&is...	1	NaN
2	3	All of Statistics	Larry Wasserman	2004	http://link.springer.com/openurl?genre=book&is...	1	NaN
3	4	Social Anxiety and Social Phobia in Youth	Christopher Kearney	2005	http://link.springer.com/openurl?genre=book&is...	1	NaN
4	5	Discrete Mathematics	László Lovász, József Pelikán, Katalin\nVeszte...	2003	http://link.springer.com/openurl?genre=book&is...	1	NaN

In [23]:

```
# Set headers
headers = requests.utils.default_headers()

# create local download dir
if not os.path.exists('Springer_Ebooks'):
    os.makedirs('Springer_Ebooks')

# define dl function
def download_file(item, df):

    url = df.loc[item, 'OpenURL']      # get server url
    req = requests.get(url, headers)  # request server url using get method
    soup = BeautifulSoup(req.content, 'html.parser') # parsing requested html

    # define necessary server file url and local filename
    server_dir = soup.find("a", title="Download this book in PDF format").get('href')
    server_domain = 'https://link.springer.com'
    local_filename = '{}_{}_{}.pdf'.format(df.loc[item, 'Book Title'],
                                           df.loc[item, 'Author'].split(',')[0],
                                           str(df.loc[item, 'Edition']).replace("\n", ""))
    # NOTE the stream=True parameter below
    with requests.get(server_domain + server_dir, stream=True) as r:
        r.raise_for_status()
        with open('Springer_Ebooks/'+ local_filename, 'wb') as f:
            for chunk in r.iter_content(chunk_size=8192):
                if chunk: # filter out keep-alive new chunks
                    f.write(chunk)
                    # f.flush()

    return local_filename
```

In [24]:

```
for item in tqdm(df.index, desc='Downloading ebook pdf...', unit='pdf', position=0):
    download_file(item, df)
```

```
Downloading ebook pdf...: 100%|██████████| 1/1 [00:09<00:00,  9.77s/pdf]
```

In [26]:

```
print('Total books to download: {}'.format(len(df.index)))
```

```
Total books to download: 408
```

117, 407 not free...