```
In [17]:   import requests as req
           from bs4 import BeautifulSoup
           import pandas as pd
```

```
In [18]:   def get_topic_link(base_url):
               response = req.get(url=base_url)
               page_content = response.text
               print(response.status_code)

               soup = BeautifulSoup(page_content, "html.parser")
               tags = soup.find_all("div", class_ = "py-4 border-bottom d-flex flex-justify-betwe
               topic_links = []
               for tag in tags:
                   url_end = tag.find("a")["href"]
                   topic_link = f"https://www.github.com{url_end}"
                   topic_links.append(topic_link)
               return topic_links
```

```
In [19]:   def get_info_topic(topic_link):
               response1 = req.get(topic_link)
               topic_soup = BeautifulSoup(response1.text, "html.parser")
               topic_tag = topic_soup.find("h1").text.strip() #3D
               topic_desc = topic_soup.find("p").text
               info_topic = {
                   "title" : topic_tag,
                   "desc" : topic_desc
               }
               return info_topic


           def get_info_tags(topic_link):
               response = req.get(topic_link)
               info_soup = BeautifulSoup(response.text, "html.parser")
               repo_tags = info_soup.find_all("div", class_ = "d-flex flex-justify-between flex-i
               return repo_tags


           def get_info(tag):
               repo_username = tag.find_all("a")[0]
               repo_name = tag.find_all("a")[1]

               url_end = tag.find("a")["href"]
               repo_url = f"https://www.github.com{url_end}"

               repo_star = tag.find("span",{"id":"repo-stars-counter-star"}).text.strip()
               repo_value = int(float(repo_star[:-1]) * 1000)

               topics_data = {
                   "repo_name" : "repo_name",
                   "repo_username" : repo_username.text.strip(),
                   "repo_name" : repo_name.text.strip(),
                   "repo_url" : repo_url,
                   "repo_star" : repo_value,
                   }
               return topics_data
```

```
In [20]:   def save_CSV(results):
               df = pd.DataFrame(results)
               df.to_csv("github.csv", index=False)

In [21]:   def save_XLX(results):
               df = pd.DataFrame(results)
               df.to_excel("github.xlsx", index=False)

In [23]:   def main():
               base_url = "https://github.com/topics"
               topic_links = get_topic_link(base_url) # list of url ex: https://github.com/topics
               result2 = []
               for topic_link in topic_links: # https://github.com/topics/3d
                   print(f"getting info {topic_link}")
                   topic_tags = get_info_topic(topic_link) #title, desc
                   repo_tags = get_info_tags(topic_link) # some repo tags, so we can use for loop
                   result1 = []
                   for tag in repo_tags:
                       repo_info = get_info(tag)
                       result1.append(repo_info)
                   for x in result1:
                       gabungan = topic_tags | x
                       result2.append(gabungan)
                   save_CSV(result2)
                   save_XLX(result2)


           if __name__ == "__main__":
               main()
```

```
200
getting info https://www.github.com/topics/3d
getting info https://www.github.com/topics/ajax
getting info https://www.github.com/topics/algorithm
getting info https://www.github.com/topics/amphp
getting info https://www.github.com/topics/android
getting info https://www.github.com/topics/angular
getting info https://www.github.com/topics/ansible
getting info https://www.github.com/topics/api
getting info https://www.github.com/topics/arduino
getting info https://www.github.com/topics/aspnet
getting info https://www.github.com/topics/atom
getting info https://www.github.com/topics/awesome
getting info https://www.github.com/topics/aws
getting info https://www.github.com/topics/azure
getting info https://www.github.com/topics/babel
getting info https://www.github.com/topics/bash
getting info https://www.github.com/topics/bitcoin
getting info https://www.github.com/topics/bootstrap
getting info https://www.github.com/topics/bot
getting info https://www.github.com/topics/c
getting info https://www.github.com/topics/chrome
getting info https://www.github.com/topics/chrome-extension
getting info https://www.github.com/topics/cli
getting info https://www.github.com/topics/clojure
getting info https://www.github.com/topics/code-quality
getting info https://www.github.com/topics/code-review
getting info https://www.github.com/topics/compiler
getting info https://www.github.com/topics/continuous-integration
getting info https://www.github.com/topics/covid-19
getting info https://www.github.com/topics/cpp
```

In [ ]:

In [ ]: