```
In [1]:
                                                                                                   H
from urllib.request import urlopen
import argparse
import requests as req
from bs4 import BeautifulSoup
In [6]:
                                                                                                   Ы
search_info = "https://www.google.co.kr/search?"
params = {
    "q": "송편",
    "tbm": "isch"
}
html_object = req.get(search_info, params)
In [7]:
                                                                                                   M
html object.status code
Out[7]:
200
                                                                                                   H
In [9]:
bs object = BeautifulSoup(html object.text, "html.parser")
img data = bs object.find all("img")
In [10]:
img data
4Yz" width="130"/>,
 <img alt="송편에 대한 이미지 검색결과" height="90" src="https://encrypted-tbn0.gsta</pre>
tic.com/images?q=tbn:ANd9GcTSZ9nytJ2Hs9V8Of1TYreb-Eo8fLBvmfrH6rHPkNVrzPOK7avjHeFDm
kI" width="135"/>,
 <img alt="송편에 대한 이미지 검색결과" height="89" src="https://encrypted-tbn0.gsta</pre>
tic.com/images?q=tbn:ANd9GcRY-rzqkxw00fDR8Rv4VdlmSb8NVET39Uufpy46oW-kvNWRQkl58LUau
uI" width="133"/>,
 <img alt="송편에 대한 이미지 검색결과" height="105" src="https://encrypted-tbn0.gst</pre>
atic.com/images?q=tbn:ANd9GcSCF58xtPPMKEXeavA7I1lOtEFfl8kBCfDte5ovMMdCGJgYXrTfkyqE
9yQ" width="140"/>,
 <img alt="송편에 대한 이미지 검색결과" height="84" src="https://encrypted-tbn0.gsta</pre>
tic.com/images?q=tbn:ANd9GcRBjb379FXJl4olAiNorWf8R VEJ0V1w3D3 pX7zuco8HMwqku7K0YzL
xZv'' width=^{\circ}150''/>,
 <img alt="송편에 대한 이미지 검색결과" height="97" src="https://encrypted-tbn0.gsta</pre>
tic.com/images?q=tbn:ANd9GcRNwnumKuAI-73UlzVgVCUwam8D6A53h0hfVl26nk7MlaRZ6iTDI4M20
NY" width="145"/>,
 <img alt="송편에 대한 이미지 검색결과" height="105" src="https://encrypted-tbn0.gst</pre>
atic.com/images?q=tbn:ANd9GcRg2STceJBZNpFq9E3leCvsKgcohRy5AASBHzPl-NRn9SDoMxMiwsBu
sWcX" width="140"/>,
 <img alt="솔펴에 대한 이미지 검색결과" height="84" src="httns://encrynted-thn0 gsta
```

In [13]:

```
for i in enumerate(img_data[1:]):
    t = urlopen(i[1].attrs['src']).read()
    filename = "img/"+str(i[0]+1)+".jpg"

with open(filename, "wb") as f:
    f.write(t)
print(str(i)+" image success")
```

- (0, <img alt="송편에 대한 이미지 검색결과" height="124" src="https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcT1NY6Rv7N2qZniqyVAH76qyrL5own6Dx0iHPe1\_gu92fcP0cmMiH8d2Huz" width="149"/>) image success
- (1, <img alt="송편에 대한 이미지 검색결과" height="113" src="https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcTKyIddS\_FR7bBCiiGc69BNQszhDU1oGfGUGw0\_6geDNzV-y8458Z6ZKN8-" width="150"/>) image success
- (2, <img alt="송편에 대한 이미지 검색결과" height="100" src="https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcR-mIgpYWbME\_f-2AiR6kgIduRdQEbKZicqn0UL9XSxdtFWbY7\_1 Ahajng" width="133"/>) image success
- (3, <img alt="송편에 대한 이미지 검색결과" height="80" src="https://encrypted-tbn0.g static.com/images?q=tbn:ANd9GcQHZ9u9Ayz1BwFRLyoTfzoKdGmZlewHl2Y30Nsx03aiBARxsiqJiGibUJnV" width="143"/>) image success
- (4, <img alt="송편에 대한 이미지 검색결과" height="111" src="https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQ7\_s2p0CBQmjKbpGiQs2tkYs8qB\_fd1-7v7FYClGKep072oMH3PRIvmWw" width="137"/>) image success
- (5, <img alt="송편에 대한 이미지 검색결과" height="80" src="https://encrypted-tbn0.g static.com/images?q=tbn:ANd9GcRsqoM641jvu2MZFpsFg69XL8hA0yQUPeNxu\_4xTYSx07hZhfICYw 7j3Fi1" width="143"/>) image success
- (6. <img alt="송편에 대한 이미지 검색결과" height="86" src="https://encrypted-tbn0.g
  - 젠장 20개밖에 다운이 안되네

In [21]:

```
from google_images_download import google_images_download
import ssl
ssl._create_default_https_context = ssl._create_unverified_context

def imageCrawLing(keyword, dir):
    response = google_images_download.googleimagesdownload()

    arguments = {
        "keywords" : keyword,
        "limit": 10000,
        "print_urls":False,
        "no_directory":True,
        "output_directory":dir,
        "chromedriver":'/home/lmu/tools/chromedriver'
}

    paths = response.download(arguments)

    print(paths)
```

```
H
In [22]:
imageCrawling('송편','img')
Item no.: 1 --> Item name = \ucla1\ud3b8
Evaluating...
Getting you a lot of images. This may take a few moments...
Reached end of Page.
Starting Download...
Completed Image ====> 1.KOCIS_songpyeon_%284994899563%29.jpg
Completed Image ====> 2.965cc82059734e0f8d3159e99b4af981 %E1%84%81%E1%85%A9%E1%86%
BE%E1%84%89%E1%85%A9%E1%86%BC%E1%84%91%E1%85%A7%E1%86%AB-1030x773.jpg
Completed Image ====> 3.150767748476_20171012.JPG
Completed Image ====> 4.%EC%86%A1%ED%8E%B8-%EC%95%8C%EA%B3%A0-%EB%93%9C%EC%85%A8%E
B%82%98%EC%9A%94.png
Completed Image ====> 5.215cc73649b240d2b230fb5699fd056f.JPG
Completed Image ====> 6.%EC%86%A1%ED%8E%B8.jpg
Invalid or missing image format. Skipping...
Completed Image ====> 7.14731304788431[1].PNG
Completed Image ===> 8.20180923000009 0.jpg
Completed Image ===> 9.%EA%BD%83%EC%86%A1%ED%8E%B8%EC%9B%90%EB%8D%B0%EC%9D%B41809
05-000.ipg
```

In []: