

In [163]:

```
import requests
from bs4 import BeautifulSoup as bsp
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm #그래프 폰트설정
import pyautogui
import re
```

In [2]:

```
resonse=requests.get("http://www.yjjob.or.kr/p/?j=23")
page=bsp(resonse.content,'html.parser')#두줄로 page 파싱 가능.
page
(($ (document).context.documentElement.clientHeight - div_h) / 2) > 10) ?
$(window).scrollTop() + (($ (document).context.documentElement.clientHeight - div_
h) / 2) : 10;
        $('#div-full-wrap').css({"position":"absolute","to
p":"0","left":"0","width":maskWidth,"height":$(document).height(),"opacity":"0.
3","background":"#111","display":"block","z-index":"9990"});
        $('#div-full-wrap').fadeIn(800);
        $('#div-full-wrap').fadeTo("slow",0.75);
        $('#div-full').css({"position":"absolute","top":t
+"px","left":l+"px","width":div_w+"px","height":div_h+"px","margin":"0 auto","disp
lay":"block","z-index":"9993","background":"#fff"});
        $('#full-contents').css({"position":"relative","to
p":"0","left":"0","width":"100%","height":"100%","margin":"0 auto","display":"bloc
k","background":"#fff","z-index":"9995"});
        t_str = (t_str) ? t_str : '이미지 미리보기
('+data.w+'px*'+data.h+'px)';
        $('#full-contents').html('<div id="ej-full-header-
wrap"><h1>'+t_str+'<span><a href="javascript:hideFull();" class="img_ovr" title
="닫기"></a></span></h1>
```

In [3]:

```
#각 훈련에대한 내용이 들어있는 div.r-eview태그 검색
curriculums=page.select('tr > td > div .r-eview')
curriculums
```

Out [3]:

```
[<div class="r-eview">
  <a class="p100" href="..p/?j=58&edu_code=VmtaYVUxTnJOVlpPVkU1UIZrUkJPUT09K00=" title="(기계설계제작) CAD/CAM 3D기계설계가공 (7개월)"><em class="font11 gray">
구직자훈련(국기) &gt; 기계설계가공</em><br/>
  <span class="font14 bold" style="color:#1a7ac3;">(기계설계제작) CAD/CAM 3D기계설
계가공 (7개월)</span></a>
  <p class="tpad3">교육기간 : <em class="blue">2022.02.07</em> ~ <em class="red">20
22.09.07</em></p> <!-- //////////////////추가 시작 -->
  <p class="tpad5">교육시간 · 요일 : <span class="bold">
09:10~16:50
(7교시)
[월, 화, 수, 목, 금]</span></p>
  <p class="tpad5">수강료 : <span class="bold">전액무료</span></p>
  <p class="tpad5">훈련수당 : <span class="bold purple">116,000</span>원 (<span cla
ss="green bold">매달차등지급</span></p>
  <!-- //////////////////추가 -->
</div>,
<div class="r-eview">
```

In [4]:

```
curriculums[0]
```

Out [4]:

```
<div class="r-eview">
<a class="p100" href="..p/?j=58&edu_code=VmtaYVUxTnJOVlpPVkU1UIZrUkJPUT09K00="
title="(기계설계제작) CAD/CAM 3D기계설계가공 (7개월)"><em class="font11 gray">구직자
훈련(국기) &gt; 기계설계가공</em><br/>
<span class="font14 bold" style="color:#1a7ac3;">(기계설계제작) CAD/CAM 3D기계설 계가
공 (7개월)</span></a>
<p class="tpad3">교육기간 : <em class="blue">2022.02.07</em> ~ <em class="red">2022.
09.07</em></p> <!-- //////////////////추가 시작 -->
<p class="tpad5">교육시간 · 요일 : <span class="bold">
09:10~16:50
(7교시)
[월, 화, 수, 목, 금]</span></p>
<p class="tpad5">수강료 : <span class="bold">전액무료</span></p>
<p class="tpad5">훈련수당 : <span class="bold purple">116,000</span>원 (<span class
="green bold">매달차등지급</span></p>
<!-- //////////////////추가 -->
</div>
```

In [5]:

```
curriculums[0].em.text
```

Out [5]:

```
'구직자훈련(국기) > 기계설계가공'
```

In [6]:

```

for curriculum in curriculums:#전체 curriculums에서 하나씩 뽑아서
    for i in curriculum:#뽑은 curriculum에서 다시 반복문을 돌면서
        if(i.text=="\n"):#줄바꿈은 패스하고
            pass
        string=i.text.strip()#아니라면 공백제거
        string2=string.replace(" ", "")#탭키 제거
        print(string2)#텍스트 검색
    print('=====')

```

구직자훈련(국기) > 기계설계가공
(기계설계제작) CAD/CAM 3D기계설계가공 (7개월)

교육기간 : 2022.02.07 ~ 2022.09.07

교육시간 · 요일 : 09:10~16:50(7교시) [월, 화, 수, 목, 금]

수강료 : 전액무료

훈련수당 : 116,000원 (매달 차등지급)

=====

구직자훈련(국기) > 용접/공조
(특수요원) 용접/공조 파견인근요원사 양성 (6개월)

In [12]:

```

list_all=[]
for i in range(len(curriculums)):
    csp=curriculums[i].select('p')
    #과정 구분과 과정명은 모든 curriculum에 있으므로 따로 검색
    #print(f'과정구분 : {curriculums[i].em.text}')
    #print(f'과정명 : {curriculums[i].span.text}')
    dic_item={'과정구분':curriculums[i].em.text,
              '과정명':curriculums[i].span.text}#dictionary형태로 저장
    for j in csp:#과정 구분과 과정명 외에 존재하는 각 태그들 검색
        string=j.text.replace(" ", "")
        #print(string)
        newKey=string.split(':')#기준으로 나누어
        if(newKey[0]=='교육시간 · 요일'):#시간을 표기하기위해 ':'이 사용됐으므로
            newKey[1]=newKey[1]+'_'+newKey[2]+'_'+newKey[3]#value가 될 값 지정
        dic_item[newKey[0].strip()]=newKey[1].strip()#dictionary에 저장
    list_all.append(dic_item)#만들어진 dictionary를 list_all에 저장.
    #print('=====')

```

In [13]:

```
for lecture in list_all: #list_all에서 한 객체씩 뽑아
    for key in lecture.keys(): #키와 밸류를 반복
        print(key, ': ', lecture[key]) #출력
    print('=====')
```

과정구분 : 구직자훈련(국기) > 기계설계가공
 과정명 : (기계설계제작) CAD/CAM 3D기계설계가공 (7개월)
 교육기간 : 2022.02.07 ~ 2022.09.07
 교육시간·요일 : 09:10~16:50(7교시) [월,화,수,목,금]
 수강료 : 전액무료
 훈련수당 : 116,000원 (매달차등지급)

=====

과정구분 : 구직자훈련(국기) > 용접/공조
 과정명 : (특수용접) CO2,TIG,피복아크용접사 양성 (6개월)
 훈련코드 : AIG20210000330093
 교육기간 : 2022.02.07 ~ 2022.07.15
 교육시간·요일 : 09:10~16:50(7교시) [월,화,수,목,금]
 수강료 : 전액무료
 훈련수당 : 116,000원 (매달차등지급)

=====

과정구분 : 구직자훈련(국기) > 전기PLC
 과정명 : (내선공사) 전기공사 및 전기관리자 양성
 훈련코드 : AIG20210000334293
 교육기간 : 2022.02.08 ~ 2022.08.16
 교육시간·요일 : 09:10~16:50(7교시) [월,화,수,목,금]

In [14]:

```
print(list_all)
```

```
{'과정구분': '구직자훈련(국기) > 기계설계가공', '과정명': '(기계설계제작) CAD/CAM 3D기계설계가공 (7개월)', '교육기간': '2022.02.07 ~ 2022.09.07', '교육시간·요일': '09:10~16:50(7교시) [월,화,수,목,금]', '수강료': '전액무료', '훈련수당': '116,000원 (매달차등지급)'}, {'과정구분': '구직자훈련(국기) > 용접/공조', '과정명': '(특수용접) CO2,TIG,피복아크용접사 양성 (6개월)', '훈련코드': 'AIG20210000330093', '교육기간': '2022.02.07 ~ 2022.07.15', '교육시간·요일': '09:10~16:50(7교시) [월,화,수,목,금]', '수강료': '전액무료', '훈련수당': '116,000원 (매달차등지급)'}, {'과정구분': '구직자훈련(국기) > 전기PLC', '과정명': '(내선공사) 전기공사 및 전기관리자 양성', '훈련코드': 'AIG20210000334293', '교육기간': '2022.02.08 ~ 2022.08.16', '교육시간·요일': '09:10~16:50(7교시) [월,화,수,목,금]', '수강료': '전액무료', '훈련수당': '116,000원 (매달차등지급)'}, {'과정구분': '구직자훈련(국기) > 용접/공조', '과정명': '(열냉동설비) 열원운영 전문가 양성-냉동공조유지보수관리 (6개월)', '훈련코드': 'AIG20200000296254', '교육기간': '2022.02.08 ~ 2022.07.26', '교육시간·요일': '09:10~16:50(7교시) [월,화,수,목,금]', '수강료': '전액무료', '훈련수당': '116,000원 (매달차등지급)'}, {'과정구분': '구직자훈련(국기) > 디자인', '과정명': '(디지털디자인) 웹퍼블리셔,UI/UX,웹디자인 (5개월) - 스마트융합훈련', '훈련코드': 'AIG20210000342967', '교육기간': '2022.02.10 ~ 2022.07.18', '교육시간·요일': '09:10~16:50(7교시) [월,화,수,목,금]', '훈련수당': '116,000원 (매달차등지급)'}, {'과정구분': '구직자훈련(국기) > IT·SW', '과정명': '(정보시스템구축) IT시스템구축 및 정보보안 전문가 양성 (5개월) - 스마트융합훈련', '훈련코드': 'AIG20210000346469', '교육기간': '2022.02.10 ~ 2022.07.18', '교육시간·요일': '09:10~16:50(7교시) [월,화,수,목,금]', '수강료': '전액무료', '훈련수당': '116,000원 (매달차등지급)'}
```

In [19]:

```
for i in list_all:
    if '훈련수당' not in i:##딕셔너리에 키가 존재하는지 체크
        print(i)
        print('=====')
```

```
{'과정구분': '일반고특화 > 디자인및설계', '과정명': '[일반고특화] 컴퓨터그래픽운용기
능사 & 디지털 웹디자인 ', '교육기간': '2022.03.02 ~ 2022.12.30', '교육시간·요일':
'09:10~16:50(7교시) [월,화,수,목,금]'}
```

```
=====
{'과정구분': '일반고특화 > 디자인및설계', '과정명': '[일반고특화] 컴퓨터그래픽운용기
능사 & 디지털 웹디자인 ', '교육기간': '2022.03.02 ~ 2022.12.30', '교육시간·요일':
'09:10~16:50(7교시) [월,화,수,목,금]'}
```

In [21]:

```
key_list=[]
for i in list_all:
    for key in i:
        key_list.append(key)
```

In [25]:

```
key_list=set(key_list)
```

In [26]:

```
key_list
```

Out[26]:

```
{'과정구분', '과정명', '교육기간', '교육시간·요일', '수강료', '훈련수당', '훈련코
드'}
```

In [27]:

```
import pyautogui
text=pyautogui.prompt("검색어를 입력하세요")
resonse=requests.get(f"https://stdict.korean.go.kr/search/searchResult.do?pageSize=10&searchKeyword={text}&page=1")
page=bspp(resonse.content, 'html.parser')#두줄로 page 파싱 가능.
```

Out [27]:

```
<!DOCTYPE html>

<html lang="ko">
<title>국립국어원 표준국어대사전</title>
<!-- HEADER -->
<meta charset="utf-8"/>
<meta content="width=device-width, initial-scale=1.0, maximum-scale=1.0, minimum-scale=1.0, user-scalable=no, target-densityDpi=medium-dpi" name="viewport"/>
<meta content="telephone=no, address=no, email=no" name="format-detection"/>
<meta content="본 누리집의 저작권은 국립국어원에 있습니다." name="copyright"/>
<meta content="표준국어대사전" name="keywords"/>
<meta content="[국립국어원 표준국어대사전]" name="description"/>
<meta content="yes" name="apple-mobile-web-app-capable"/>
<meta content="IE=edge" http-equiv="X-UA-Compatible"/>
<link href="/css/jquery.mCustomScrollbar.css" rel="stylesheet" type="text/css"/>
<link href="/css/style.css" rel="stylesheet" type="text/css"/>
<link href="/css/sliderkit-core.css" media="screen" rel="stylesheet" type="text/css"/>
```

In [65]:

```
title=page.select('.t_blue2')
title=title[0].text.strip()
print(title)
dts=page.find_all('dt')
dts
```

```
학원
[<dt>
<a class="t_blue1" href="/search/searchView.do?word_no=362856&searchKeywordTo=3">
학원<sup>1</sup></a><span
class="t_gray hanja_font">(學員)</span> 「명사」
<span style="letter-spacing:1px;line-height:25px;margin-top:5px;"><font class="dataLine"> 공부하는 인원.</font><a class="btn_allmost" href="#none" onclick="javascript:fnSearchView('362856')" style="white-space:nowrap;">전체 보기</a></span>
</dt>, <dt>
<a class="t_blue1" href="/search/searchView.do?word_no=503671&searchKeywordTo=3">
학원<sup>2</sup></a><span
class="t_gray hanja_font">(學院)</span> 「명사」
<span style="letter-spacing:1px;line-height:25px;margin-top:5px;"><span class="t_red1">「1」</span><font class="dataLine">『교육』 일정한 목적·교과 과정·설비·제도 및 법규에 의하여 계속적으로 학생에게 교육을 실시하는 기관.=학교.</font><a class="btn_allmost" href="#none" onclick="javascript:fnSearchView('503671')" style="white-space:nowrap;">전체 보기</a></span>
```

In [169]:

```

text=pyautogui.prompt("검색어를 입력하세요")
resonse=requests.get(f"https://stdict.korean.go.kr/search/searchResult.do?pageSize=10&searchKeyword={text}&page=1")#두줄로 page 파싱 가능.
page=bs4.BeautifulSoup(resonse.content, 'html.parser')
title=page.select('.t_blue2')
title=title[0].text.strip()
print(title)
dts=page.find_all('dt')
base_link='https://stdict.korean.go.kr/'
print(f'검색어 : {title}')
for dt in dts:
    number=dt.select('sup')[0].text
    print(f'순번 : {number}')
    hanja=dt.select('span.hanja_font')[0].text
    print(f'한자 : {hanja}')
    pumsa=dt.text.strip().split(' ')[0].split('「')[1].split('」')[0]
    print(f'품사 : {pumsa}')
    comment=dt.select('font.dataLine')[0].text
    print(f'설명 : {comment}')
    a=dt.select_one('a')
    link=base_link+a['href']
    print(f'링크 : {link}')
    print('=====')

```

사과
 검색어 : 사과
 순번 : 1
 한자 :
 품사 : 명사
 설명 : 『식물』 살이 연하고 달며 물이 많은 참외.=사과참외.
 링크 : https://stdict.korean.go.kr//search/searchView.do?word_no=169066&searchKeywordTo=3 (https://stdict.korean.go.kr//search/searchView.do?word_no=169066&searchKeywordTo=3)

=====

순번 : 2
 한자 : (司果)
 품사 : 명사
 설명 : 『역사』 조선 시대에, 오위(五衛)에 둔 정육품의 군직(軍職). 현직에 종사하고 있지 않은 문관, 무관 및 음관(蔭官)이 맡았다.
 링크 : https://stdict.korean.go.kr//search/searchView.do?word_no=169067&searchKeywordTo=3 (https://stdict.korean.go.kr//search/searchView.do?word_no=169067&searchKeywordTo=3)

=====

순번 : 3

사용자로부터 검색어를 받아 표준국어 대사전 페이지를 불러오는 함수

In [220]:

```

def searchWords():
    text=pyautogui.prompt("검색어를 입력하세요")
    resonse=requests.get(f"https://stdict.korean.go.kr/search/searchResult.do?pageSize=10&searchKeyword={text}&page=1")#두줄로 page 파싱 가능.
    return page

```

page를 매개변수로 받아 dt태그를 찾아 리스트를 반환시켜주는 함수

In [128]:

```
def findDt(page):
    dts=page.find_all('dt')#각 내용이 포함된 dt 검색
    return dts
```

dt에서 검색어를 찾고 출력해주는 함수

In [223]:

```
def findString(dt):
    string = dt.select_one('a').text.strip()
    newstring = re.sub(r'[0-9]+', '', string)
    print(f'검색 : {newstring}')
    return newstring
```

dt에서 Number를 찾고 출력해주는 함수

In [232]:

```
def findNumber(dt):
    if dt.select('sup'):#순번이 있는 경우만 작동
        number=dt.select('sup')[0].text
        print(f'순번 : {number}')
        return number
```

dt에서 한자를 찾고 출력해주는 함수

In [225]:

```
def findHanja(dt):
    hanja=dt.select('span.hanja_font')[0].text #한자 찾기
    print(f'한자 : {hanja}')
    return hanja
```

dt에서 품사를 찾고 출력해주는 함수

In [226]:

```
def findPumsa(dt):
    if '「' in dt.text.strip():
        pumsa=dt.text.strip().split('「')[1].split('」')[0]#품사 찾기
        print(f'품사 : 「{pumsa}」')
        return pumsa
    else:
        print(f'품사 : 없음')
        return '없음'
```

dt에서 설명을 찾고 출력해주는 함수


```
def findComment(dt):
    comment=dt.select('font.dataLine')[0].text #설명 찾기
    print(f'설명 : {comment}')
    return comment
```

In [228]:

```
def findLink(dt):
    base_link='https://stdict.korean.go.kr/'
    a=dt.select_one('a')#링크 찾기
    link=base_link+a['href']
    print(f'링크 : {link}')
    return link
```

In [261]:

```
def forDts(dts):
    dic_dts=[]
    for dt in dts:
        dic_dt={}
        dic_dt['검색']=findString(dt)
        dic_dt['순번']=findNumber(dt)
        dic_dt['한자']=findHanja(dt)
        dic_dt['품사']=findPumsa(dt)
        dic_dt['설명']=findComment(dt)
        dic_dt['링크']=findLink(dt)
        dic_dts.append(dic_dt)
    print('=====')
#print(dic_dts)
return dic_dts
```

In [262]:

[illegible]

