

## 평가방법: 필답형 – 파이썬 활용 빅데이터수집&처리

평가일시	2019년 10월 14일		
과정명	SBA 빅데이터 사이언스 실무자 양성과정		
교과목	파이썬 활용 빅데이터 수집 & 처리	수강생	
단원명	파이썬 활용 빅데이터 수집 & 처리	확 인	

구분 (능력단위요소)	문 항	점 수
	<p>(문항 1~ 문항 10) 8점, 문항 (11), (12) 10점</p> <p>1. 빈칸을 채워 주세요.(각 4점)</p> <p>=====</p> <pre> &lt;!doctype html&gt; &lt;html&gt;   &lt;__(A)____&gt;     &lt;meta charset="utf-8"&gt;     &lt;title&gt;서울산업진흥원 교육과정&lt;/title&gt;   &lt;/__(A)____&gt;   &lt;__(B)____&gt;     &lt;h1&gt;빅데이터 사이언스 실무자 양성과정&lt;/h1&gt;     &lt;p id="subject"&gt;파이썬기반 빅데이터분석&lt;/p&gt;     &lt;p id="teacher"&gt;홍길동&lt;/p&gt;     &lt;p id="school"&gt;서울IT&lt;/p&gt;     &lt;p id="semester"&gt;2019&lt;/p&gt;   &lt;/body&gt; &lt;/html&gt; ===== </pre> <p>(정답)</p> <p>(A)_____</p> <p>(B)_____</p>	100/ 100 점

(2) 파이썬 사용하는 라이브러리 중의 하나로서, URL 정보를 이용하여 웹페이지의 HTML 소스를 가져오기 위해 사용하는 라이브러리는 무엇인가요?

```
=====
import _____
url = "https://www.naver.com"
html = _____.get(url)
html
=====
```

- (A) requests
- (B) geturl
- (C) gethtml
- (D) htmlget

(정답) \_\_\_\_\_

(3) HTML 소스를 파싱하여 웹 HTML에서 정보를 쉽게 가져오려고 합니다. 아래 빈칸에 알맞은 라이브러리는 무엇일까요?  
빈칸에 알맞은 것을 채워주세요.

```
=====
# _____를 이용해 HTML 소스를 파싱
soup = _____ (html, 'xml')
soup
=====
```

- (A) requests
- (B) matplotlib
- (C) BeautifulSoup
- (D) parse\_html

(정답) \_\_\_\_\_

(4) a태그 중에 class가 'f1'인 요소를 하나를 가져오려고 합니다. 빈칸에 알맞은 명령을 선택해 주세요.

=====

```
soup._____ ('a', class="f1")
```

=====

(A) find\_all

(B) get\_all

(C) find\_one

(D) find

(정답) \_\_\_\_\_

빈칸을 채워 주세요. (5) ~ (7)

=====

```
url ="https://www.alexa.com/topsites/countries/KR"
```

```
html =requests.get(url).text
```

```
soup = BeautifulSoup(html , "__ (5) __") # html.parser - 가장 속도 빠름.
```

```
soup.title
```

```
# p 태그의 요소 안에서 a 태그의 요소를 찾음
```

```
website_ranking =soup.select ('__ (6) __')
```

```
top_sites =soup.__(7)__('div') # div 태그 요소를 전체를 찾아, 저장.
```

```
len (top_sites )
```

=====

(정답) (5)\_\_\_\_\_

(정답) (6)\_\_\_\_\_

(정답) (7)\_\_\_\_\_

(8) 정규 표현식을 이용하여 원하는 패턴에 맞는 전화번호를 검색하여 가져오려고 합니다. 아래 빈칸에 적절한 명령을 보기에서 골라주세요.

=====

```
# 전화번호
```

```
cafe_tel = ____ (8) ____ ('\d{3}[-]\d{3}[-]\d{4}', taginfo)
```

```
if cafe_tel is not None:
```

```
    cafe_tel = cafe_tel.group()
```

```
else:
```

```
    cafe_tel = 'No Contact'
```

```
cafe_tel
```

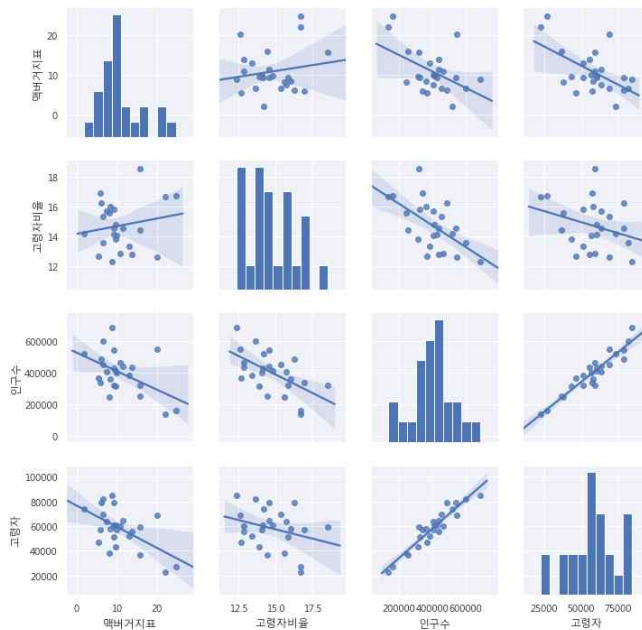
=====

- (A) re.search()
- (B) reg.search()
- (C) reg.pattern()
- (D) re.pattern()

(정답) (8) \_\_\_\_\_

(9) 아래와 같은 그래프를 표현해 주는 seaborn 의 함수는 무엇인가요?

sns.\_\_(9)\_\_(population, vars=["맥버거지표", "고령자비율", "인구수", "고령자"], kind="reg")



- (A) countplot()
- (B) scatterplot()
- (C) barplot()
- (D) pairplot()

(정답) \_\_\_\_\_

(10) O, X 퀴즈

listEle = soup2.find\_all('p') p태그의 요소 전체를 검색하여 이를 딕셔너리 형태의 자료로 listEle에 넣게 된다. ( \_\_\_\_ (A) \_\_\_\_ )

```
=====
rank_list = list()
```

```
menu_list = list()
cafe_list = list()
link_list = list()
```

=====

위의 소스 코드는 빈 리스트를 네 개 선언한 것이다.( \_\_\_\_ (B) \_\_\_\_ )

(정답) (A) \_\_\_\_\_

(정답) (B) \_\_\_\_\_

(11) pip 명령을 이용하여 우리는 프로그램을 설치할 수 있다.

selenium을 설치하려고 한다면 아래 빈칸에 어떤 명령을 이용하여 설치가 가능할까요? 빈칸을 채워주세요.

pip \_\_\_\_ (A) \_\_\_\_ \_\_\_\_ (B) \_\_\_\_

(정답) (A) \_\_\_\_\_

(정답) (B) \_\_\_\_\_

(12) 아래 프로그램은 Selenium를 이용한 웹 데이터를 가지고 오는 부분의 일부이다. 주어진 주석을 확인 후, 해당되는 부분의 빈칸을 채워 보세요.

=====

....

# 영화제목 검색키 입력

movie\_title = '결투'

movie\_title = '알라딘 외전'

driver.find\_element\_by\_css\_selector('#ipt\_tx\_xxx').\_\_\_\_\_(1)\_\_\_\_\_

# 검색 클릭

driver.find\_element\_by\_css\_selector('#jSearchArea > div > button').click()

# 검색 첫번째 페이지로 이동

driver.find\_element\_by\_css\_selector('#old\_content > ul:nth-child(4) > li:nth-child(1) > dl > dt > a').click()

# 페이지 객체의 HTML코드를 html에 저장 후,

# BeautifulSoup 객체로 가져온다

html = \_\_\_\_\_(2)\_\_\_\_\_

soup = \_\_\_\_\_(3)\_\_\_\_\_

soup.title # 가져온 html 정보의 title 객체를 확인한다.

	<pre> cur_url = driver.current_url  # 현재 페이지의 url를 cur_url에 저장. cur_url review_cnt = soup.find("span","cnt") review_cnt ## review_cnt 결과 : &lt;span class="cnt"&gt;총&lt;em&gt;635&lt;/em&gt;건&lt;/span&gt; review_cnt2 = review_cnt.____(4)____  # 635 값을 얻기(find 이용) review_cnt2 ===== (정답) (1)_____ (정답) (2)_____ (정답) (3)_____ (정답) (4)_____ </pre>	
	<b>총점</b>	