## [평생] 멀티미디어 일반형 양식

## [원고] 빅데이터 수집 시스템 개발

12회차: 문자열 처리와 그룹연산

내용전문가	백현숫	교수설계(한기대)	홍길동
협력업체	업체명	교수설계(협력업체)	홍길동

NCS 분류 정보	20-01-02-09	정보통신 - 정보기술 - 정보기술개발 - 빅데이터플랫폼구축
능력단위 정보	03	빅데이터 수집시스템 개발
	2001020903_17v1.1	빅데이터 수집시스템 설계하기
능력단위 요소 정보		
9454 TT 9T		

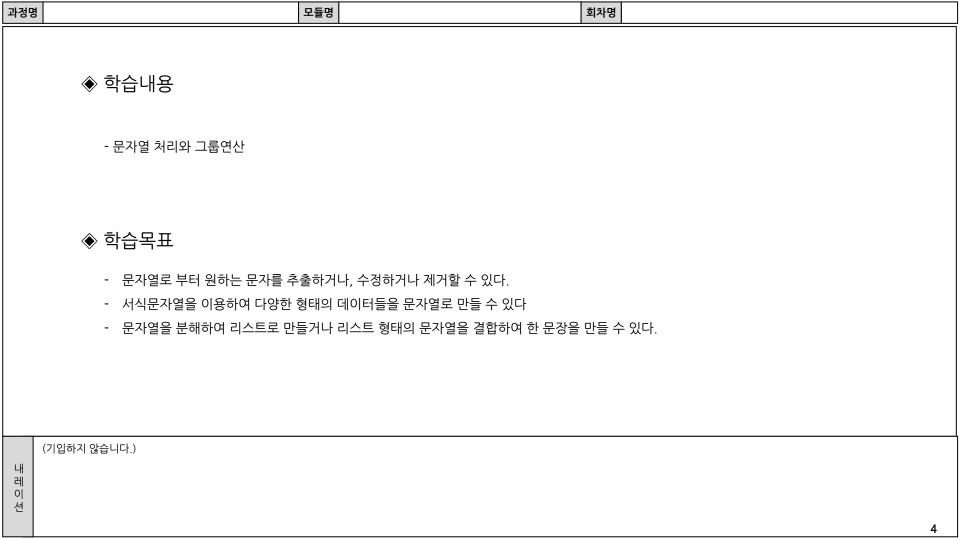
업무	작성자	버전	작성일	특이사항
원고 작성	백현숙	v.1.0	2019/09/11	2회차원고
한기대 피드백		v.1.1		
원고 보완		v.2.0		
자문 진행		v.2.1		
원고 보완				

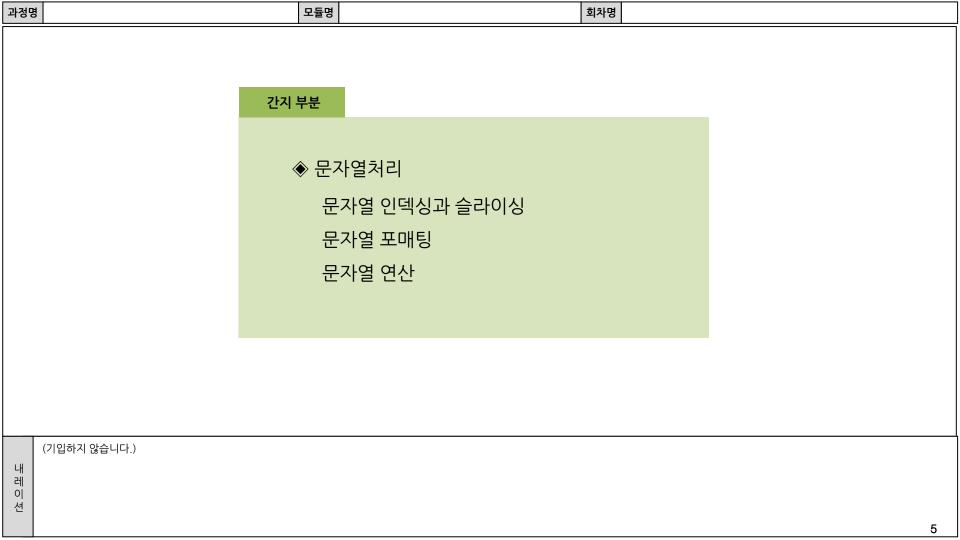


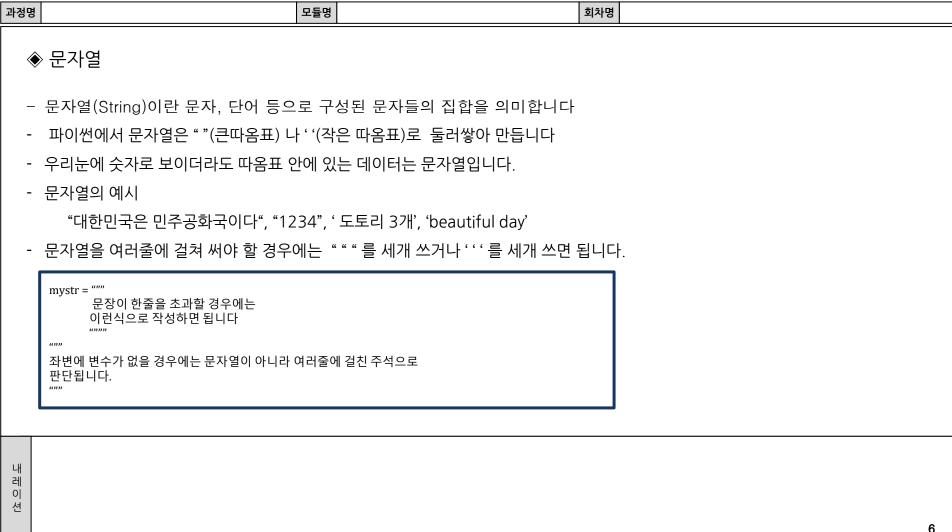
◆ 학습 목차 - 학습 목차를 작성해 주세요. (NCS 비적용 과정일 경우에는 NCS 및 능력단위 정보를 기입하지 않아도 됩니다.)

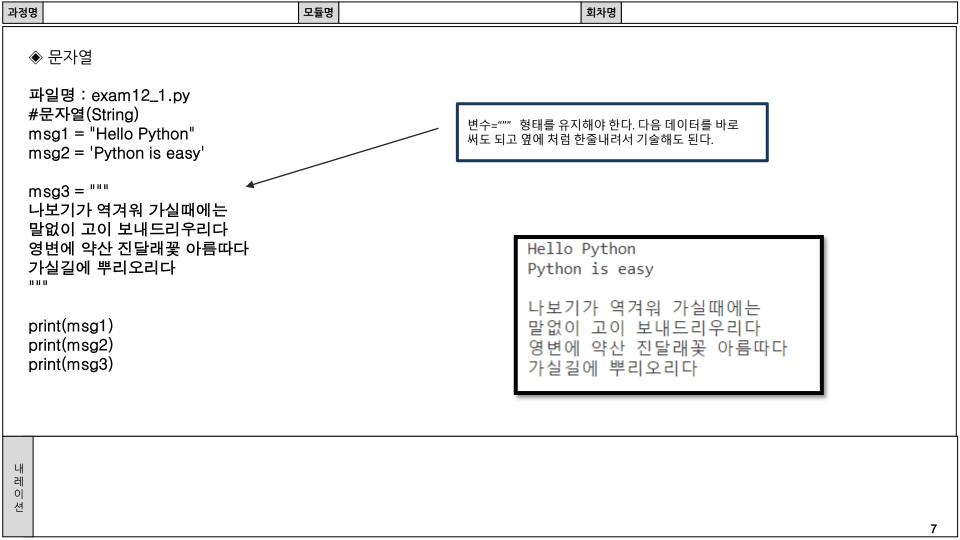
NCS 분류 정보	20-01-02-09	정보통신 - 정보기술 - 정보기술개발 - 빅데이터플랫폼구축
능력단위 정보	능력단위코드 기입	빅데이터 수집시스템 개발

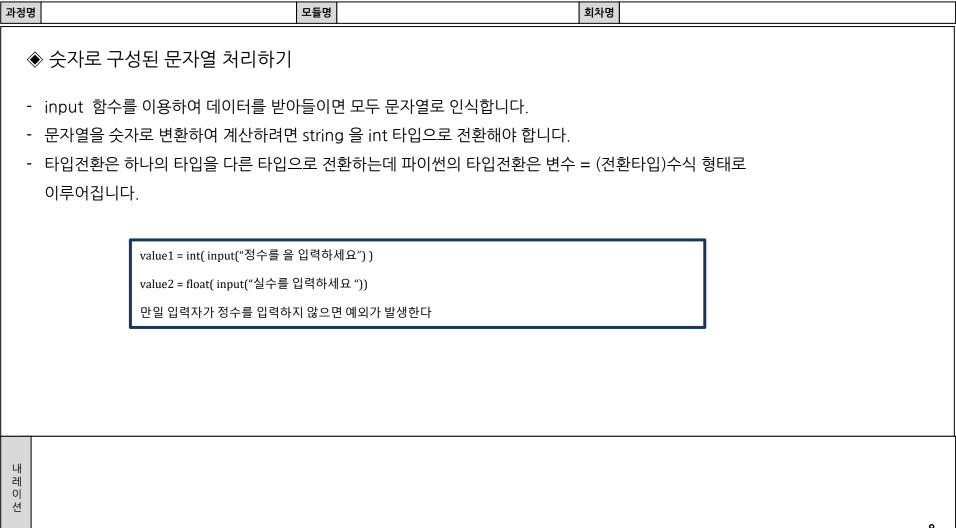
모듈명	회차명		능력단위 요소		수행준거
1. 모듈명 기입	01. 회차명 기입	1	능력단위요소명 기입	1.1	수행준거 기입
	02. 회차명 기입	1	능력단위요소명 기입	1.2	수행준거 기입
	03. 회차명 기입	1	능력단위요소명 기입	1.3 1.4	수행준거 기입 수행준거 기입
	04. 회차명 기입	1	능력단위요소명 기입	1.4	수행준거 기입
2. 모듈명 기입	05. 회차명 기입	2	능력단위요소명 기입	2.1 2.2	수행준거 기입 수행준거 기입
					2



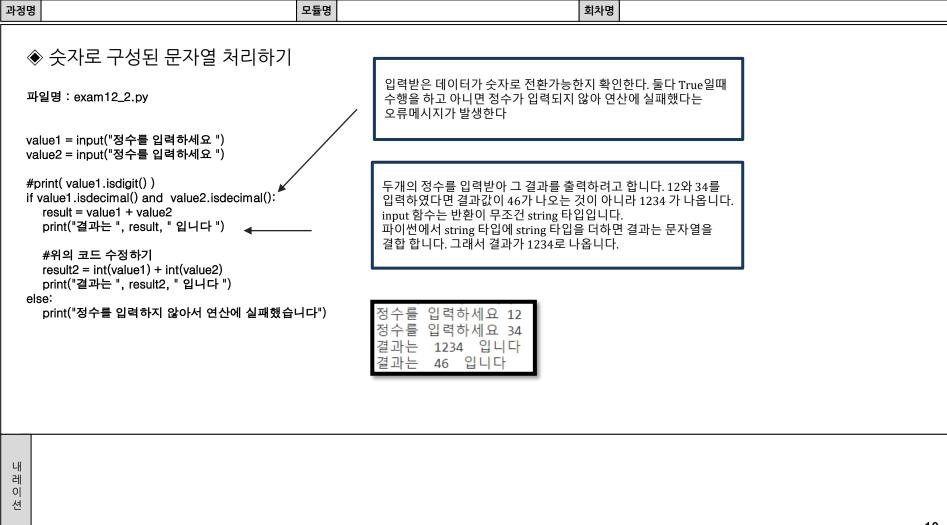




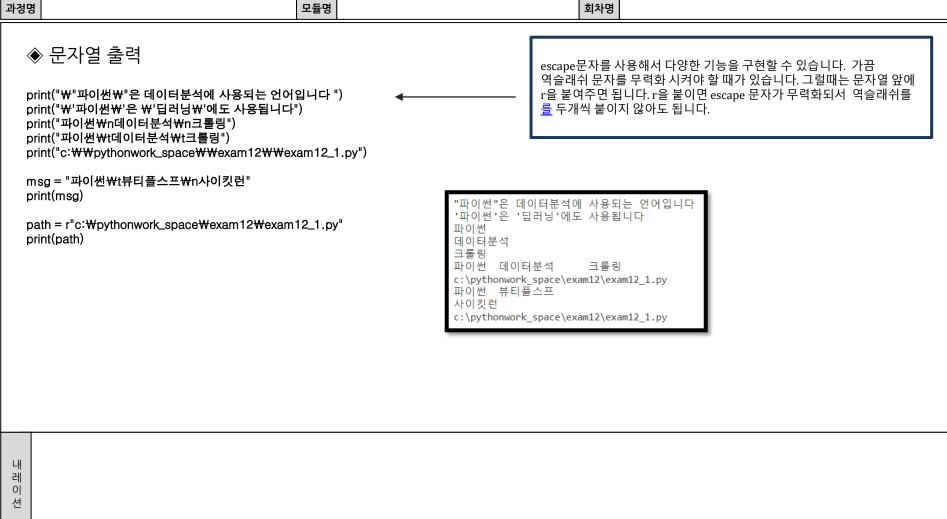


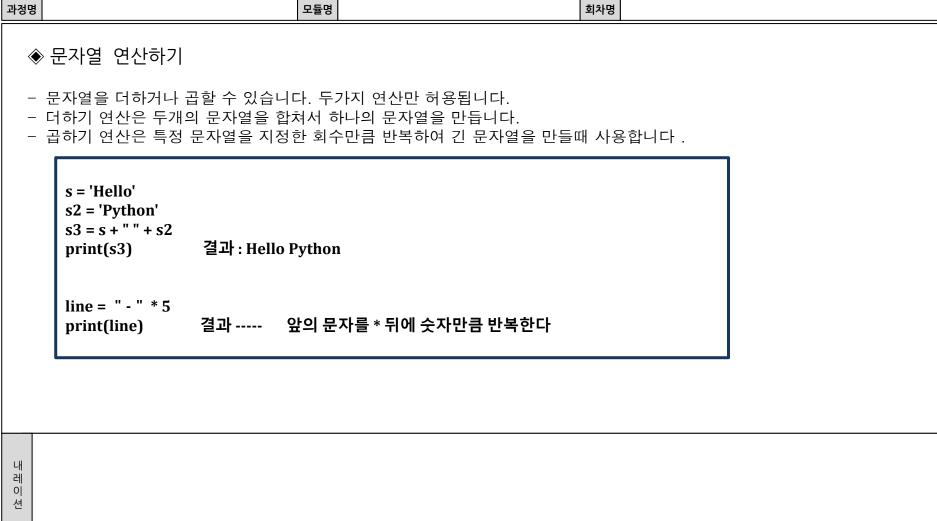


과정명	!	모듈명 회차명					
4	췣 입략	된 문자열이 숫자로 이루어졌는지 확인하는 함수					
-	- input 함수를 이용하여 받아들인 데이터가 만일 숫자가 아닌 문자가 포함되어 있다면 파이썬은 예외를 발생시킵니다.						
-	예외	를 발생시키지 않고 사전에 오류를 점검하고 싶다면 우선 타입전환이 가능한지 확인을 해야 합니다	ł.				
-	문자일	멸이 숫자로 구성되었는지 확인하기 위한 함수가 isdigit, isdecimal, isnumeric 세개가 있습니다.					
	1		]				
		isdecimal : 문자열이 숫자로만 구성되었으면 True, 아니면 False반환					
		isdigit : 특수 기호 중에서 어깨위에 제곱이나 세제곱을 표시하는 문자처럼 '숫자처럼 생긴' 모든 글자를 다 숫자로 판단합니다.					
		idnumeric : 숫자값 표현에 해당하는 텍스트까지 숫자로 인정해줍니다 . 예) "½"					
내 레 이 션			9				

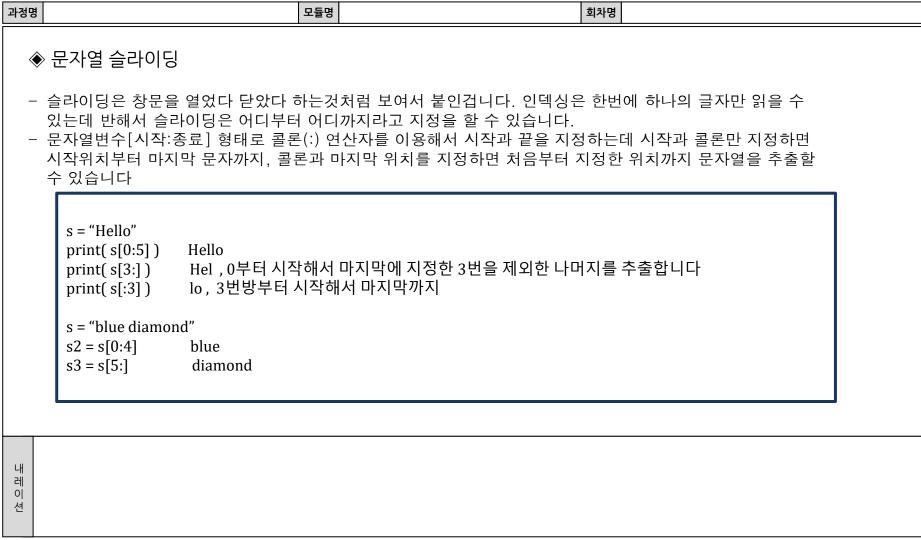


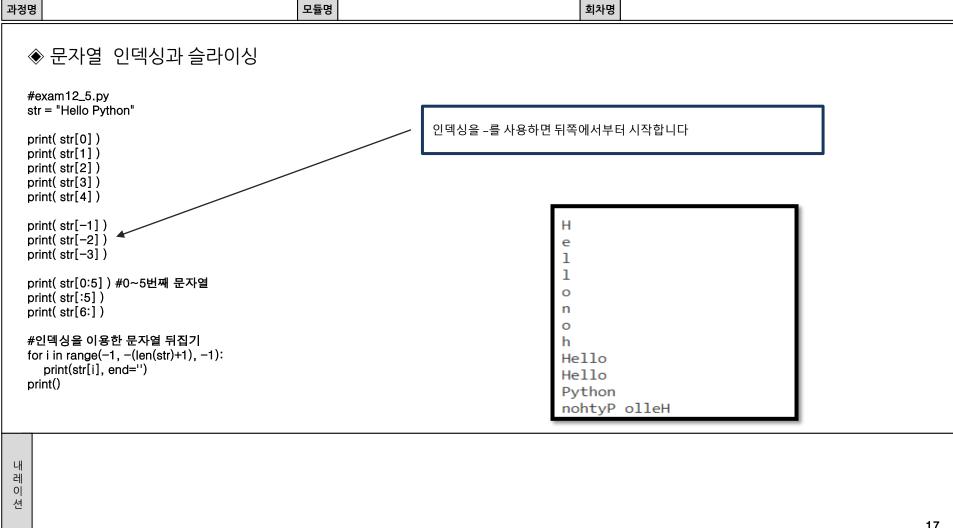
과정당	<b>B</b>	모듈명	회차명
-	이스케이프 코드런 출력을 보기 좋게 파이썬에서는 \(역	을바꿈이나 특별한 문자등을 출력하기 위해 이스케이 한 프로그래밍할 때 사용할 수 있도록 미리 정의해 둔 정렬하는 용도로 사용 합니다. 문자열 안에 함께 사용 슬래쉬) 와 함께 사용하는데 키보드의 우축 상단에 로 보이고 한글폰트에서는 \ 로 보입니다. 다음 표에	"문자 조합 " 입니다. 용합니다 백스페이스 아래에 있는 문자입니다.
	문자	의미	
	\n	문자열에서 줄바꿈을 할 때 사용합니다.	
	\t	문자열에서 탭키를 적용할 때 사용합니다.	
	\\	역슬래쉬 문자 자체를 출력하고자 할 때 사용합니	·   다
	\'	작은 따옴표를 출력하고자 할 때 사용합니다.	
	\"	큰 따옴표를 출력하고자 할 때 사용합니다	
내 레 이 션			11

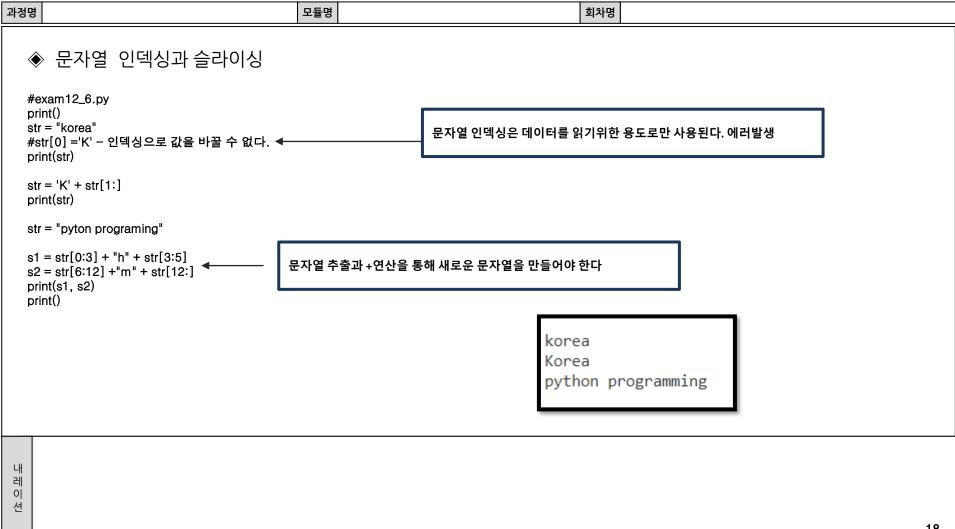


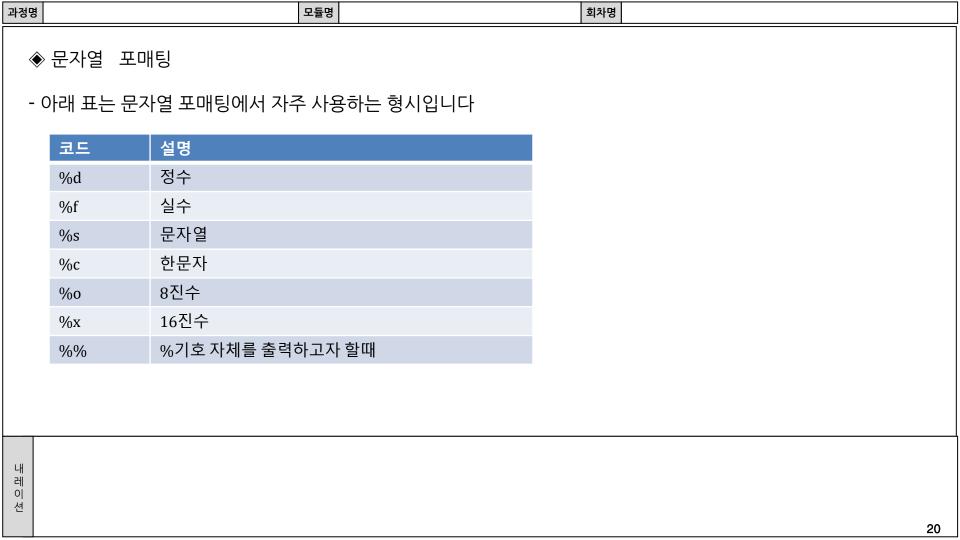


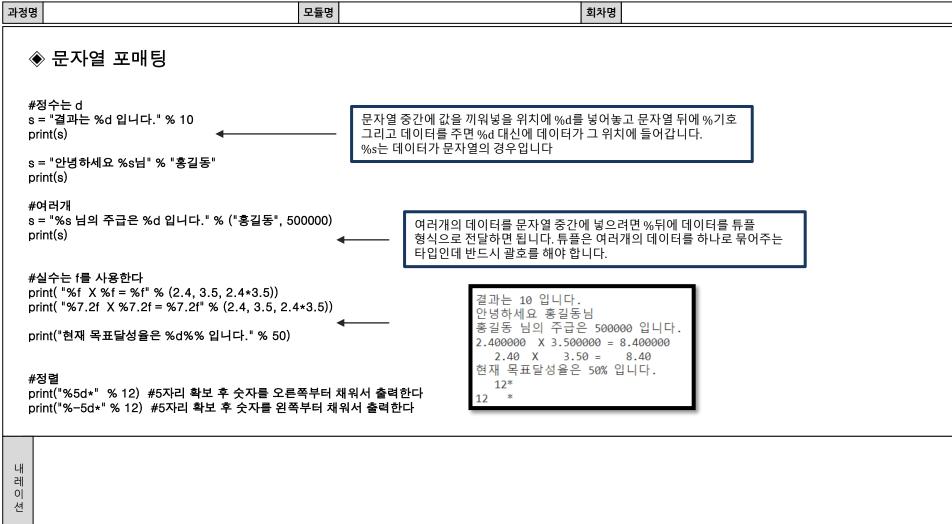


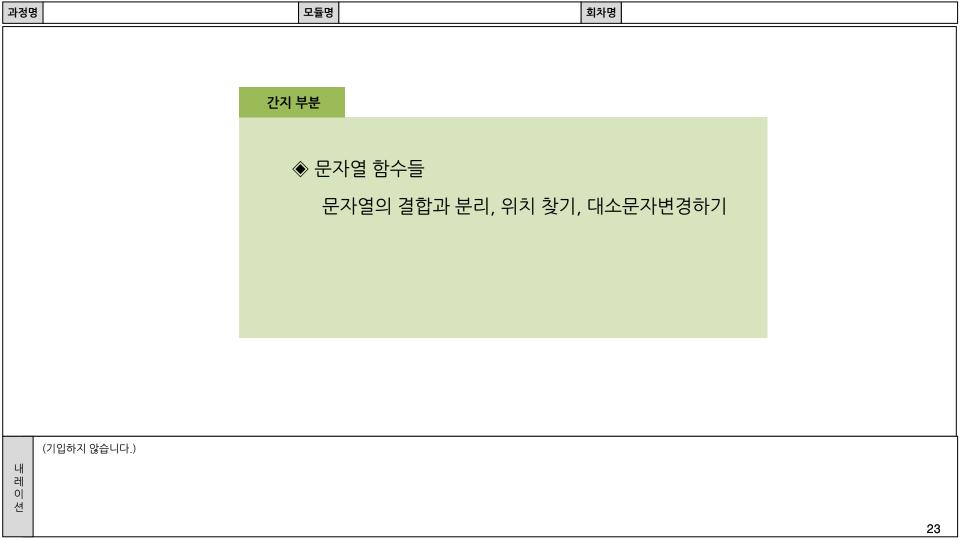












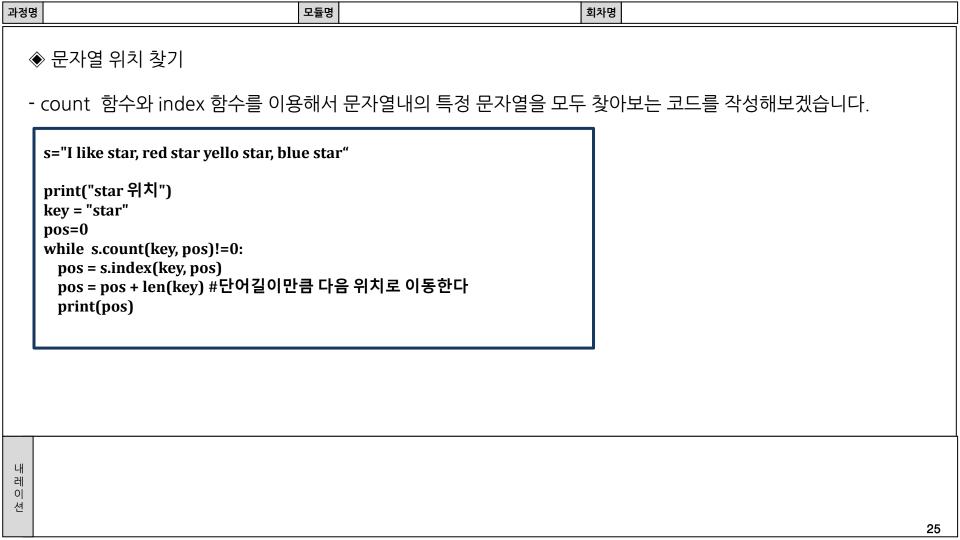
회차명

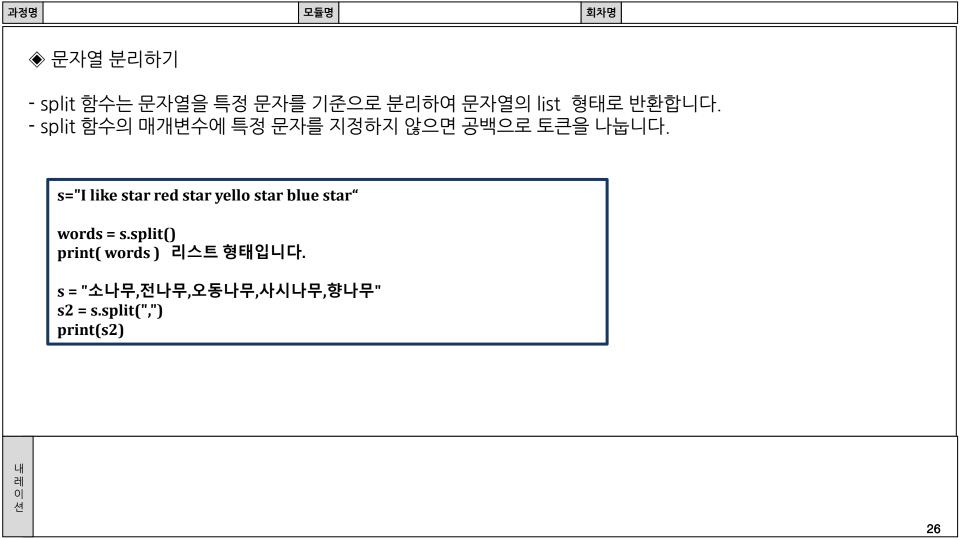
모듈명

문자열에서 특정 문자열이 몇번 존재하는지 찿기 위해서는 count 함수를 사용합니다.

과정명

◈ 문자열 위치 찾기





	과정명		모듈명	회차명	
	#e s= wo	≫ 문자열 관련 함수들 exam12_9.py "I like star, red star yello star, blue star" ords = s.split() int( words )			
	W	= "소나무,전나무,오동나무,사시나무,향나무" pods = s1.split(",") int(woods)			
	pr	int( "문자열 길이 : ", len(s) )			
		특정 문자의 개수 int ("star 개수 : ", s.count("star"))			
	pr	star 의 위치 int("star 첫번째 위치:", s.index("star")) 고든 star 찾기			
	ke po	int("star 위치") y = "star" ps=0 nile s.count(key, pos)!=0: pos = s.index(key, pos) pos = pos + len(key) #단어길이만큼 다음 위치를 print(pos)	로 이동한다		
	내 레 이 션				
١					30

과정명	<b>5</b>	모	듈명	회차명		
	<b>*</b>	적용하기				
	1.	data 폴더내의 data.txt 파일을 읽어서 토큰을 분리하고 토큰의 개수는 몇개인지 출력하시오				
		파일읽기 예 file = open("./data/data.txt", enco data = file.read() print(data) file.close()	oding="utf8")			
	·					
	2.	Too 와 Time 단어가 각각 몇번 ·	등장했는지 횟수를 출력하시오			
	3.	who 등장 위치를 리스트에 추기	ㅏ하여 어떤 위치에 등장했는지 위 <i>₹</i>	니값 리스트 <del>를</del> 출력하시오		
	4.	모든 토큰을 대문자로 바꾸어서	토큰의 번호, 대문자로 전환한 토	큰을 출력하시오		
내 레 이 션	(기입	]하지 않습니다.)				

과정명		모듈명	회차명	
	◈ 적용하기- 풀이			
	file = open("./data/data.txt", encod data = file.read() print(data) file.close()	ing="utf8")		
	#1)토콘분리하기 tokens = data.split() print( len(tokens))			
	#2)Too 와 Time 단어가 각각 몇번 등 print("Too 개수 : ", data.count("Too print("Time 개수 : ", data.count("Tir	)"))		
내레	(기입하지 않습니다.)			
이 션				33
				50

과정명		모듈명	회차명
•	문제풀기		
	문제 1) 다음 연산의 수행 결과는 ?		
	x1 = "123"		
	x2 = "456"		
	result = x1 + x2		
	<ol> <li>123456</li> <li>579</li> <li>예외발생</li> <li>142536</li> <li>정답) 1</li> <li>해설) 문자열간의 + 연산은 문자열을 5</li> <li>난이도) 5(1:아주어려움, 2:어려움 3: 관련페이지번호) 8</li> </ol>		
내 레 이 션	(기입하지 않습니다.)		35

번호	문제	정답	난이 도	해설	관련학습 보기	
2	디음처럼 문장이 있을때 수행 결과를 작성하시오 s = "DeepLearning" print(s[4:])	Learnin g	3	4번째 데이터부터 슬라이딩이므로 Learning가 된다.	16	
3	다중 중 에러가 발생하는 부분은 ? s = "machineLearning"  ① s[0]="M" ② s2 = "M" + s2[:1] ③ s2 = s[0] + s[1] + s[2] + s[3[ + s[4] ④ s2 = "machine" * 2	1	3	인덱스를 이용해서 데이터를 수정할 수 없습니다	15	
4	다음 프로그램의 수행 결과는 s = "{1} {3} {0} {2}".format(1,2,3,4) print(s) 1) 1 2 3 4 2) 3 2 1 4 3) 2 4 1 3 4) 4 3 2 1	3	3	{1} {3} {0} {2} 는 format 함수에 전달되는 인자들의 위치값입니다. 1 -0, 2 -1, 3-2, 4-3 에 해당됩니다. 그래서 2 4 1 3 의 결과가 나옵니다	22	<b>3</b> 6

번호	문제	정답	난이도	해설	관련학습보기
5	토큰을 분리하려고 합니다. 다음 중 적당한 함수는?	4	4	토큰을 분리하기 위해서는 split 함수를 사용합니다.	26
	① index				
	② isdecimal				
	3 strip				
	④ split				
6	다음처럼 토큰 리스트가 있을때 모든 토큰을 결합사여 하나의 문자열로 만들 수 있는 함수는? msg = ['l'', 'like'', ''star'']	1	5	join 함수를 이용해서 토큰들을 하나로 묶을 수 있습니다	29
	① join				
	② append				
	③ concatenation				
	④ add				
	다음 단어를 대문자로 출력하고자 합니다. 적절한 함수는 ?	2	4	대문자로 변환하는 함수는 upper 입니다 .	27
7	① capital				
	② upper				
	③ lower				
	④ capitalize				
					<u>l</u>

번호	문제	정답	난	해설	관련학습보기
8	다음 프로그램의 수행 결과는 ? s = " My name is Tom " print ("*" + s.lstrip() + "*") 1) *My name is Tom * 2) * My name is Tom * 3) *My name is Tom* 4) *MynameisTom*	1	5	lstrip 함수는 문자열 왼쪽의 공백을 삭제해줍니다	28
9	s1 = "35" s2 = "120" result = ( 일때 연산의 결과가 155가 나오려면 어떻게 해야 하는지 비어있는 공간의 코드를 완성하시오	int(s1) + int(s2)	3	문자열의 경우 먼저 정수형태로 전환을 해야 한다 정수 전환 연산자는 int 이다	8
10	다음 처럼 데이터가 출력되길 원할때 올바른 format 방법은 00001 00002 00003 00004 { 이부분에 포맷팅을 완성하세요	:05d {:05d}	1	{ :05d} 정수 출력을 위한 5개의 자리를 잡고 나머지 빈자리를 지정한 문자 0으로 채우시오	22





- 주요학습내용: 해당 회차 또는 레슨의 주요학습내용을 자세히 기입해 주세요.
- 검색 키워드: 학습자가 검색창에 어떤 검색어를 입력하면 본 회차 또는 본 레슨이 검색될 수 있을지 검색 키워드를 5개 기입해 주세요.

제목	주요학습내용	검색 키워드1	검색 키워드2	검색 키워드3	검색 키워드4	검색 키워드5
<sub>12.</sub> 문자열 처리와 그룹 연산	문자열과 문자열 처리함수	join	splite	strip	index	count