DS II sugara Università

SNS빅데이터분석

기말고사 대체 문제 (Q1 ~ Q5)

동서대학교 소프트웨어학과 빅데이터 융합연계전공 박승민 교수



Q1.

나라장터 – 대한민국 조달청에서 운영하는 나라장터 사이트에서 공고명과 기간을 입력하여 결과를 검색한 후 다양한 형식의 파일(txt, xls)로 저장하는 크롤러를 만드

세요





업무	공고번호-차수	분류	공고명	공고기관	수요기관	계약방법	입력일시 (입찰마감일시)	공동 수급	투찰
용역	20190426611- 00	일반	2019학년도 글로벌비전 진로캠 프(미국서부)해외연수 위탁용역 2단계입찰(규격가격동시) 공고	광주시 호남 삼육중학교	광주시 호남 삼육중학교	제한(총액)규 격가격동시	2019/04/19 15:50 (2019/04/30 10:0 0)		지문투찰
용역	20190427413- 00	일반	2019학년도 호남삼육중「글로 벌 비전 진로캠프(미국 동부)」 위탁용역 2단계입찰공고	광주시 호남 삼육중학교	광주시 호남 삼육중학교	제한(총액)규 격가격동시	2019/04/19 15:46 (2019/04/30 10:0 0)		지문투찰
기타	유한산단 입찰 공고 제2019-0 1호-00	일반	취업캠프 위탁운영 건	유한대학 산 학협력단	유한대학교 산학협력단		2019/04/19 15:42 (2019/05/07 10:0 0)		
용역	20190433223- 01	변경	수성대학교 전문대학 혁신지원 사업 HC진로설정캠프 위탁운영 업체선정 공고	수성대학교	수성대학교	일반(총액)협 상에의한계 약	2019/04/19 13:59		
용역	20190432021- 00	일반	2019년도 우체국과 함께 하는 청소년 경제금융캠프	과학기술정 보통신부 우 정사업본부 우정공무원 교육원	과학기술정 보통신부 우 정사업본부 우정공무원 교육원	제한(총액)협 상에의한계 약	2019/04/19 09:53 (2019/04/30 11:0 0)		지문투찰
용역	20190432270- 00	일반	2019학년도 목포고등학교 드림 체험캠프(수학여행) 차량임차용 역 수의계약 안내공고	전라남도교 육청 목포고 등학교	전라남도교 육청 목포고 등학교	수의(총액)소 액-견적입찰 (2인 이상 견 적 제출)	2019/04/19 09:28 (2019/04/25 10:0 0)		지문투찰
용역	20190433286- 01	변경	2019 국제디자인융합캠프 운영 대행 용역	한국디자인 진흥원	한국디자인 진흥원	제한(총액)협 상에의한계 약	2019/04/18 19:16 (2019/04/29 10:0 0)	협정	지문투찰





1.공고명으로 검색할 키워드는 무엇입니까?: 캠프

2.조회 시작일자 입력(예:2019/01/01):2019/01/01

3.조회 종료일자 입력(예:2019/03/31):2019/04/20

4.파일로 저장할 폴더 이름을 쓰세요(예:c:\data\): c:\data\

1번째 공고내용을 추출합니다~~~~

1.업무: 용역

2.공고번호-차수: 20190426611-00

3.분류: 일반

4.공고명: 2019학년도 글로벌비전 진로캠프(미국서부)해외연수 위탁용역 2단계입찰(규격가격동시) 공고

5.URL 주소: http://www.g2b.go.kr:8081/ep/invitation/publish/bidInfoDtl.do?bidno=20190426611&bidseq=00&releaseYn=Y&taskClCd=5

6.공고기관: 광주시 호남삼육중학교 7.수요기관: 광주시 호남삼육중학교 8.계약방법: 제한(총액)규격가격동시

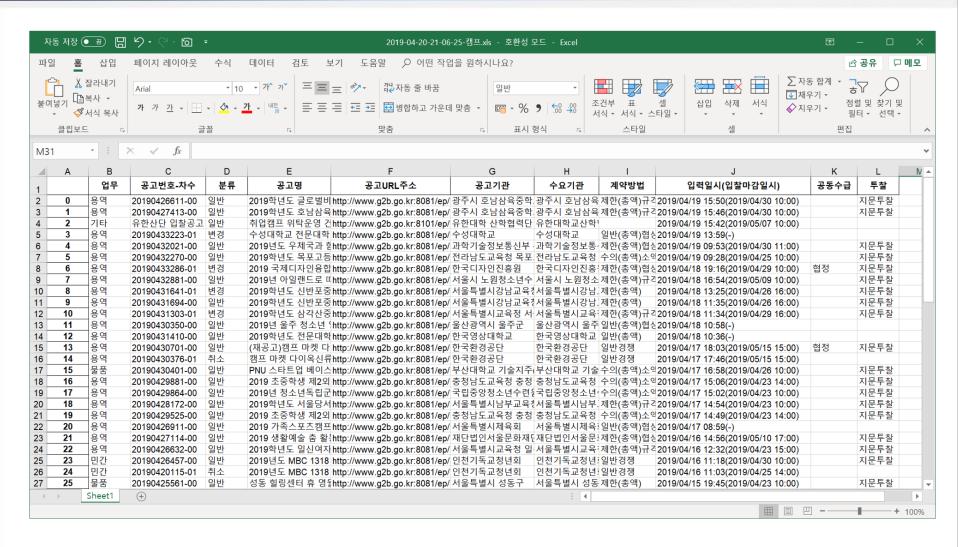
9.입력일시(입찰마감일시): 2019/04/19 15:50(2019/04/30 10:00)

11.공동수급:

12.투찰여부: 지문투찰











Q2. 네이버 블로그 - 검색어와 날짜와 검색 건수를 입력 받은 후 네이버 블로그의 상세 내용을 수집하도록 크롤러를 만드세요~

연습문제 6-5: 블로그 크롤러 : 여러건의 네이버 블로그 정보 추출하여 저장하기

1.크롤링할 키워드는 무엇입니까?(예:여행): 수문딸기체험

- 2.결과에서 반드시 포함하는 단어를 입력하세요(예:국내,바닷가) (여러개일 경우 , 로 구분해서 입력하고 없으면 엔터 입력하세요):
- 3.결과에서 제외할 단어를 입력하세요(예:분양권,해외) (여러개일 경우 , 로 구분해서 입력하고 없으면 엔터 입력하세요): 안경,린드버그
- 4.조회 시작일자 입력(예:2017-01-01):2019-01-01
- 5.조회 종료일자 입력(예:2017-12-31):2019-04-20
- 6.크롤링 할 건수는 몇건입니까?(최대 1000건): 20
- 7.파일을 저장할 폴더명만 쓰세요(예:c:\temp\text{\psi}):c:\text{\psi}data\text{\psi}







- 1.블로그 주소:https://blog.naver.com/nazirite81?Redirect=Log&logNo=221512233931
- 2.작성자 닉네임:둥과 함께

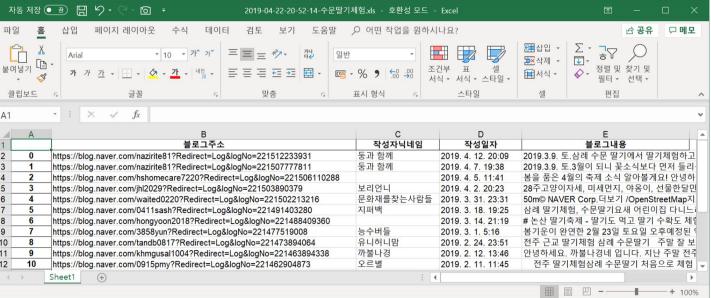
■ 2019-04-22-20-52-14-수문딸기체형.txt - 메모징 파일(F) 편집(E) 서식(O) 보기(V) 도움말

- 3.작성 일자:2019. 4. 12. 20:09
- 4.블로그 내용:2019.3.9. 토.삼례 수문 딸기에서 딸기체험하고 도시락도 먹었는데 집에 가기엔 너무 이른 시간 라까지 들고 왔네. 누군가 그려놓은 하트에 숟가락 얹는 둥이. ㅋㅋ 날 좋다~~날이 좋아지면 나들이 열심

총 12 건 중 2 번째 블로그 데이터를 수집합니다==========

- 1.블로그 주소:https://blog.naver.com/nazirite81?Redirect=Log&logNo=221507777811
- 2.작성자 닉네임:둥과 함께
- 3.작성 일자:2019. 4. 7. 19:38
- 4.블로그 내용:2019.3.9. 토.3월이 되니 꽃소식보다 먼저 들리는 딸기 소식 작년에 처음으로 둥이랑 딸기체험을

며 기분 좋은 둥이. ^^ 한참 딸기 좋고 이벤트도 있어 좋았는데 체험? 수집 예시 화면들







Q3. pixabay 사이트의 이미지 다운로드하기

무료로 이미지를 다운로드 할 수 있는 멋진 사이트 중에서 많이 이용되는 곳이 바로 pixabay입니다. 이 사이트에서 특정 키워드를 입력하여 이미지를 검색한 후 컴퓨터의 지정된 폴더로 다운로드 받는 크롤러를 작성하세요.

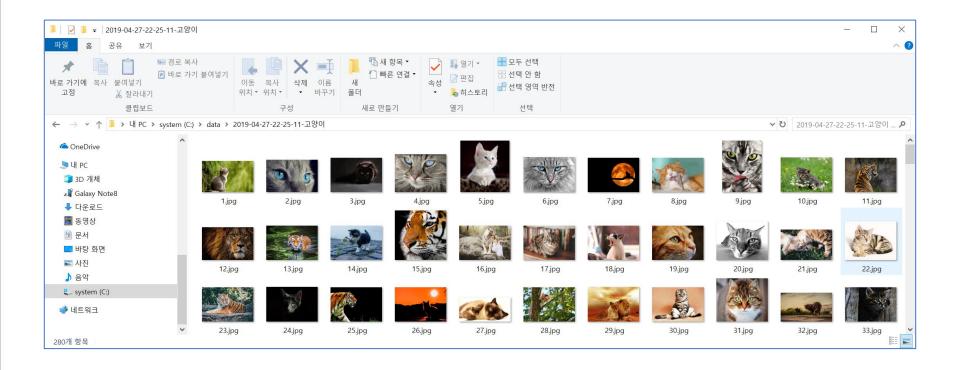
- 사이트 주소 : <u>https://pixabay.com/ko/</u>

pixabay 사이트에서 이미지를 검색하여 수집하는 크롤러 입니다

- 1.크롤링할 이미지의 키워드는 무엇입니까?: 고양이
- 2.크롤링 할 건수는 몇건입니까?: 280
- 3.파일이 저장될 경로만 쓰세요(예: c:\temp\): c:\data\





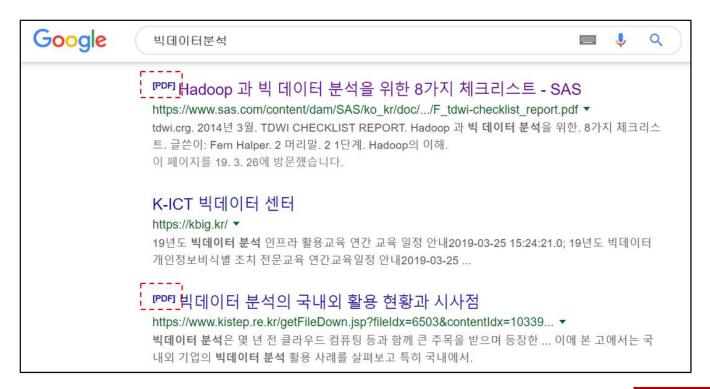






Q4. pdf 형식의 파일 다운로드 받기

인터넷에 있는 수많은 자료 중에는 pdf 형식의 파일로 제공되는 경우가 아주 많습니다. 그래서 이번에는 google 사이트에서 특정 키워드로 검색(키워드는 자유)한후 pdf 파일로 제공되는 파일을 지정된 폴더로 다운로드 받는 크롤러를 만들겠습니다.



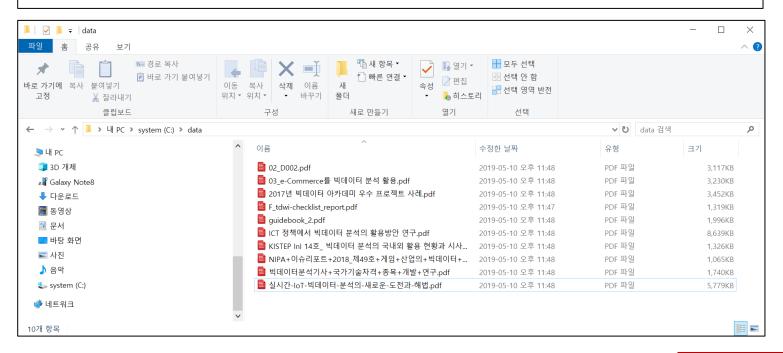




연습문제 7-4.구글 사이트에서 pdf 파일을 검색하여 수집하는 크롤러

- 1.크롤링할 키워드는 무엇입니까?: 빅데이터분석
- 2.크롤링 할 건수는 몇건입니까?: 10
- 3.파일이 저장될 경로만 쓰세요(예: c:₩temp₩) : c:₩data₩

입력하신 경로가 존재하여 c:₩data₩ 폴더에 저장하겠습니다







Q5. 유튜브 영상의 댓글 수집하기

유투브(www.youtube.com)에서 특정 키워드로 영상을 검색한 후 사용자가 요청한 영상의 건수만큼 각 영상 별 댓글을 수집하여 아래의 예시 화면과 같이 txt, csv, xls 형식으로 저장하는 크롤러를 만드세요.

연습문제 8-3 유투브 영상의 댓글 수집하기

- 1.유튜브에서 검색할 주제 키워드를 입력하세요(예:롯데마트):올리브영
- 2.위 주제로 댓글을 크롤링할 유튜브 영상은 몇건입니까?:3
- 3.각 동영상에서 추출할 댓글은 몇건입니까?: 10
- 4.크롤링 결과를 저장할 폴더명만 쓰세요(예:c:\temp\):c:\data\





1 번째 영상의 1 번째 댓글

1.URL 주소: https://www.youtube.com//watch?v=K2i20U bgHw

2.댓글 작성자명: Mona C

3.댓글 작성일자: 2시간 전

4.댓글 내용: Youtuber Taengoo unnie you should apply for the youtube play button awards!!! The silver and the gold (and eventually the diamond one!!) ��

1 번째 영상의 2 번째 댓글

1.URL 주소: https://www.youtube.com//watch?v=K2j20U_bgHw

2.댓글 작성자명: K4CE

3.댓글 작성일자: 2시간 전

4.댓글 내용: Thank you again with sharing these little moments with us ♥ you're so cute as always. Love yo u 3900

(중간 내용은 생략합니다)

3 번째 영상의 9 번째 댓글

1.URL 주소: https://www.youtube.com//watch?v=wgP1INNUZzg

2.댓글 작성자명: •Oryxs Chan•

3.댓글 작성일자: 1개월 전

4.댓글 내용: 너는 매우 귀엽고 멋지다. 나는 너의 비디오와 너의 채널을 좋아한다.��

3 번째 영상의 10 번째 댓글

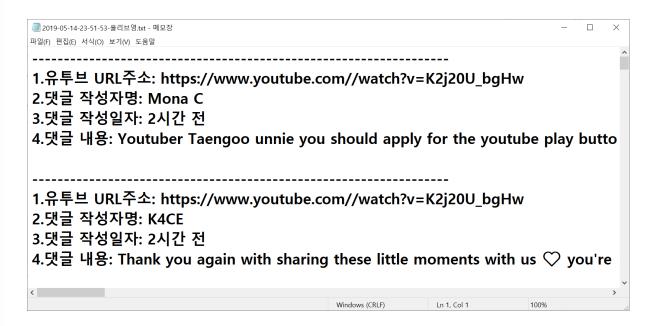
1.URL 주소: https://www.youtube.com//watch?v=wgP1INNUZzg

2.댓글 작성자명: 박소희 3.댓글 작성일자: 1개월 전

4.댓글 내용: 두콩언니 ◈♀◈◈◈◈������������������ㅠㅠ 넘 최고 ,,,, 추천템 영상 제일 조아요 ㅠㅠ 매일 하나씩 영업당해갑니다 ...저번 더샘 더마수딩트리트먼트 ?ㅠㅠㅠ그고 영업당해서 샀는데 피부 너무 조아져뜹니다 ...◈ 연두콩언니최공



[txt 파일 저장 예시 화면]







[csv , xls 파일 저장 예시 화면]

			2019-05-14-23-51	53-올리브영.xks - 호환성 모드 - Excel 조		
파일 를	<u>혹</u> 삽입 페이지 레이아웃 수식 데이터 검	토 보기 도움말	♪ 검색		🖒 공유	- 메모
불여넣기	■ 복사 ·		다동 줄 바꿈 병합하고 가운데 맞춤	일반 *** *** *** *** *** *** *** *** *** *		
클립보	보드 5 글꼴 5	맞춤		ҕ 표시 형식 ҕ 스타일 셀 편집		
A1	· : × ✓ fx					
			_	<u>-</u>		_
1 A	B URL 주소	C 댓글작성자명	D 댓글작성일자	E 댓글 내용		F
2 0	https://www.youtube.com//watch?v=K2j20U_bgHw	Mona C	2시간 전	Youtuber Taengoo unnie you should apply for the youtube play button awards!!! The silver and the gold (and e	wontually the	o diamond
3 1	https://www.youtube.com//watch?v=K2j20U_bgHw	K4CE	2시간 전	Thank you again with sharing these little moments with us \bigcirc you're so cute as always. Love you 3900	ventually the	s diamond
4 2	https://www.youtube.com//watch?v=K2j20U_bgHw	Yumi Cassandra	3시간 전	Thank you taeveon for keep uploading vlog You are the best idol coz you know what your fans like		
5 3	https://www.youtube.com//watch?v=K2j20U_bgHw	탱탱볼	3시간 전	헐 태연누나 진짜 짱슈스다 베라에도 저런 싸인용 숟가락 있구나 누나 우리 집에도 와서 싸인해줘요 숟기	라 널리고 날	클렸어
6 4	https://www.youtube.com//watch?v=K2j20U_bgHw	Millicent Wong	2시간 전	I always wonder who is filming	1 2 12 2	
7 5	https://www.youtube.com//watch?v=K2j20U_bgHw		2시간 전	탱구tv 특징 : 아직 못본 사람은 있어도 한번만 본 사람은 없다.		
8 6	https://www.youtube.com//watch?v=K2j20U_bgHw		2시간 전	ㅋㅋㅋㅋㅋㅋㅋ분명히 태연언니 막 이곳저곳 많이많이 다니시는 것 같은데 정작 한번도 본적이 없을 무	착한 사루	박에게만 보
9 7	https://www.youtube.com//watch?v=K2j20U_bgHw	김가을	2시간 전	언니 정말 아무것도 안 나오고 김태연 씨만 나오는데 이렇게 재밌을 수가 있나요		
10 8	https://www.youtube.com//watch?v=K2j20U_bgHw	ss45678912	2시간 전	I don't know what you are talking about but i still watch the video till the end		
11 9	https://www.youtube.com//watch?v=K2j20U_bgHw	SugarxMochi	2시간 전	TAEYEON! I LOVE YOU SO MUCH!! ��One comment = One Taeyeon fanPff, who am I kidding, we're a	ll a fan of he	r
12 10	https://www.youtube.com//watch?v=HwXPIX0yWqY	e R	5일 전	와이지 막해라램ㅋㅋㅋㅋㅋ		
13 11	https://www.youtube.com//watch?v=HwXPIX0yWqY	서지혜	5일 전	나도 찍을래 브이로구✓		
14 12	https://www.youtube.com//watch?v=HwXPIX0yWqY	mang9 han	2일 전	머스테브블러셔는 어디서 살수있나요?		
15 13	https://www.youtube.com//watch?v=HwXPIX0yWqY	최지원	2일 전	보타닉힐보쿠션 색 추천해주세영 ㅜㅜㅜ데일리로 지금 쓰고있는 색!!! ㅎㅎㅎ♦♡		
16 14	https://www.youtube.com//watch?v=HwXPIX0yWqY	ty고대오징어	3일 전	파우치 영상 여기저기서 많이 봤었는데 역시 직원분들은 다르시군요 영업력이 정말 만땅ㅋㅋㅋ 캡처하느리	· 정신없었습	늘니다
17 15	https://www.youtube.com//watch?v=HwXPIX0yWqY	김다혜	3일 전	==== 우리 최고당♠♠	I O L L M L	
18 16	https://www.youtube.com//watch?v=HwXPIX0yWqY	Lanimouse 라니마우스		이런 파우치 탈기 넘 좋아요^^ 남의 파우치 구경 꿀잼 *		
19 17	https://www.youtube.com//watch?v=HwXPIX0yWqY	권롯국	1일 전	퀄리티조타링		
20 18	https://www.youtube.com//watch?v=HwXPIX0yWqY	톱유튭	3일 전	삐아는 어디서 살수있을까요? 사고싶은데 올리브영은 비싼것맘 입점해유 ㅋㅋ		
21 19	https://www.youtube.com//watch?v=HwXPIX0yWqY	김은경	3일 전	수고많았다요~ 나 일빠 !! 그나저나 저 말되게 많네여 (당황:)		
22 20	https://www.youtube.com//watch?v=wgP1lNNUZzg		1개월 전	♦나중에 내가 보기 편할려고 쓴 것♦00:35 청귤 비타C 잡티세럼 (구달) ₩2400002:22 청귤 비타C 잡티 톤입	(크림 (구달)	₩22000
23 21	https://www.youtube.com//watch?v=wgP1lNNUZzg		1개월 전	눈화장지우고 항상 찝찝하고 눈에 화장품 들어간 것 같고 인공눈물쓰면 화장품이 더 들어가는 기분이여서 고		
24 22	https://www.youtube.com//watch?v=wgP1lNNUZzg	도라에몽	1개월 전	올리브영두콩 컴온이네요☆☆☆☆언니 덕분에 판테놀크림 너무너무×999999잘쓰고있어요! (구달 청귤세럼		
25 23	https://www.youtube.com//watch?v=wgP1lNNUZzg		1개월 전	언니 추천템 영상이 제일 좋아요! 영상 올라오면 늘 메모해놓고 쇼핑할 때 찾아보게 돼요 ♡ 언니가 추천했던		
26 24	https://www.youtube.com//watch?v=wgP1lNNUZzg		1개월 전	ㅠㅜㅜㅜ치카치카 머야 ㅜㅜㅜㅜㅜ 언니 29 개월이죠 ㅠㅠㅠ흑흑♡�		
27 25	https://www.youtube.com//watch?v=wgP1lNNUZzg	뽈뽈이	1개월 전	아앗 루시도엘 주위에 말 안하고 혼자 조용히 쓰고 있었는데 두콩언니가 소개해서 더더욱 인기가 많아지겠너	용 ㅠㅠㅋㅋㅋ	
28 26	https://www.youtube.com//watch?v=wgP1lNNUZzg	인채	1개월 전	두콩언니 영상은 내 삶의 활력소!!! 다른 채널은 보다안보다 반복하고 몇개월동안 안 보는 채널도 많은데 이상	하게 두콩언	너니 영상은
29 27	https://www.youtube.com//watch?v=wgP1lNNUZzg	김현경	1개월 전	구달세럼조심하세요전따갑고트러블올라왓다능		
30 28	https://www.youtube.com//watch?v=wgP1lNNUZzg	•Oryxs Chan•	1개월 전	너는 매우 귀엽고 멋지다. 나는 너의 비디오와 너의 채널을 좋아한다.�◆		
31 29	https://www.youtube.com//watch?v=wgP1INNUZzg	박소희	1개월 전	두콩언니 ♦♀♦♦♦♦♦। 이 주는 하는 사람들이 하는 사람들이 되었다면 하는 사람들이 하는 다음	□ 매일 하나	·씩 영업당
32						
→	Sheet1 (+)			1 (>
						+ 1009

