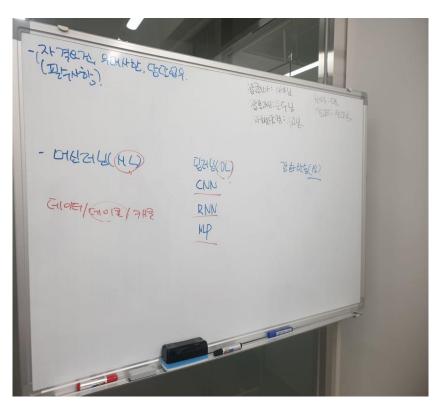
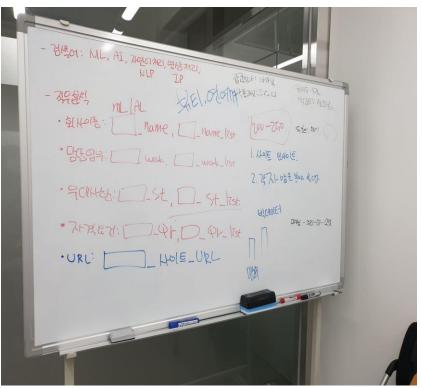
빅데이터 취업과 관련한 기술 동향

이소정, 장마가, 하주리, 석은수

초기구상





주제 선정및 배경설명

주제 선정 이유: 빅데이터 관련 취업을 위해 회사에서 요구하는 기술동향을 조사

크롤링한 사이트: ICT 학점연계 프로젝트 인턴십

크롤링 실패한 사이트: 잡코리아, 원티드, 알리오, 잡플래닛, 사람인인

ICT 선정이유: 회사정보, 기술, 직무내용 등이 상세하게 기술

출처: https://www.ictintern.or.kr/main.do

ICT 학점연계 프로젝트 인턴십이란?



66

국내·외 기업에서 제안한 ICT분야 직무 중심의 프로젝트, 서비스 개발 등에 대학생이 참여하여 인턴십을 수행하고, 이를 학점으로 인정받도록 하는 이론과 실무 역량을 겸비한 문제해결형 창의인재양성 프로그램입니다.

"

사업특성

- ✓ ICT 직무 중심의 프로젝트 인턴십 수행(국내•글로벌)
- 기업 및 대학(학생)의 인턴십 역량 제고를 위한 현실적인 지원
- 운영기관을 통한 체계적인 전주기적 관리•운영



관련 법령 및 지침

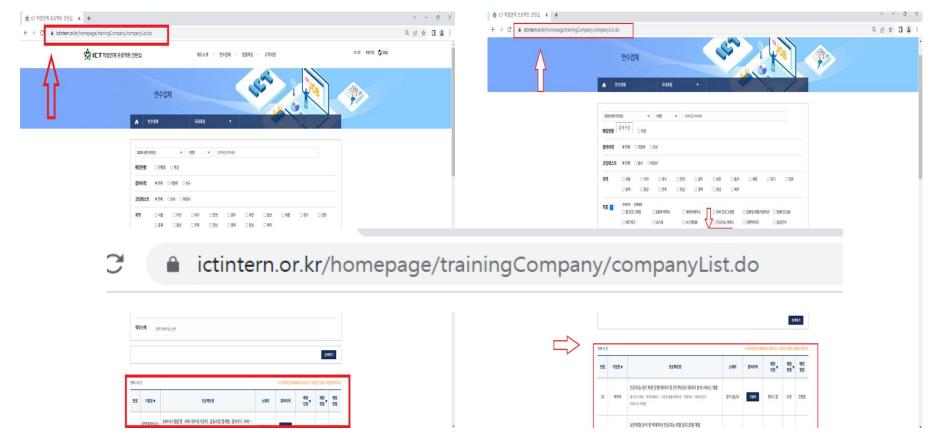
- 정보통신 진흥 및 융합 활성화 등에 관한 특별법
 (정보통신융합법) 제12조(학점이수 인턴제도),
 동법 시행령 제12조∼제14조, 동법 시행규칙 제2조
- 정보통신 방송 연구개발 관리 규정 및 관련 지침
- 정보통신 방송 기반조성 사업 수행 관리지침 등



전체시스템구조

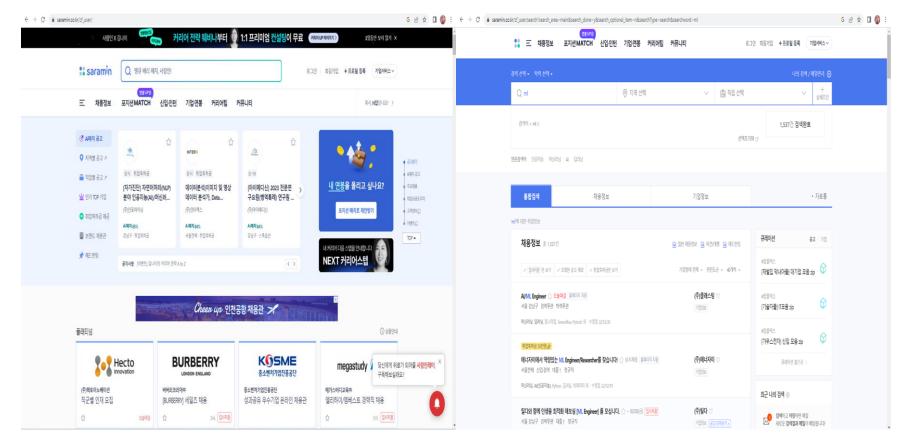
- 상세조건을 선택하고 검색한 경우에도 주소는 변하지 않음=> 동적 크롤링 필요
- 은수: 2022년 상반기
- 주리: 2022년 하반기
- 마가: 2023년 상반기
- 소정: 2023년 상반기 빅데이터

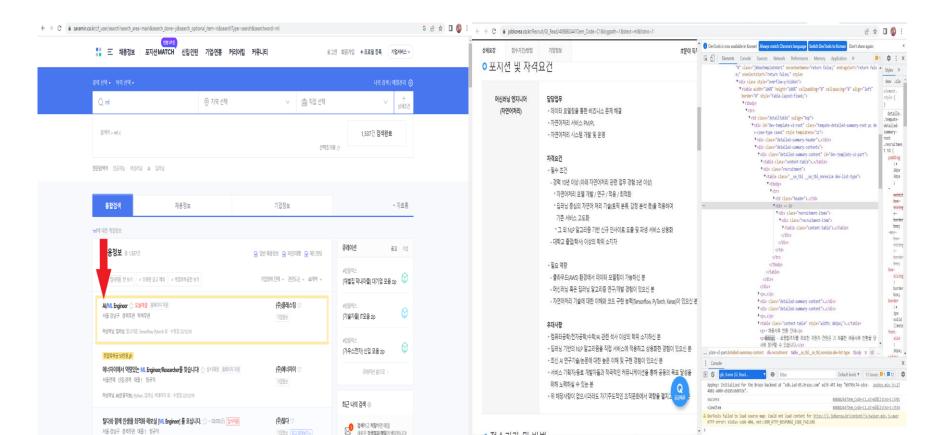
주소를 통해 알아보는 동적크롤링



여기서부터 개인발표 ICT

석은수





코드 구현

```
29
         driver.find_element_by_xpath(f'//*[@id="pagingNavi"]/div[2]/a[{j}]').click()
30
         time.sleep(0.5)
31
32
         # ======기술스택
33
34
35
             for m in range(1.21):
36
                 driver.find_element_by_xpath(f'//*[@id="container"]/form/div/div[3]/div[2]/table/tbody/tr[{m}]/td[3]/a').click()
37
38
                html=driver.page source
39
                 soup = BeautifulSoup(html, 'html.parser')
40
41
42
                 std wokr = soup.find all('tbody')[13]
                 std wokr tr = std wokr.find all('tr')
43
44
45
                 for stack j in std wokr tr[1].select('td div div div.item'):
                    stacks.append(stack j.text)
46
47
48
                driver.find element by xpath(f'//*[@id="contents area"]/div/div[3]/a').click()
49
50
51
         except: continue
52
         # ======기술스택
53
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

□ Python + ~ □ · A ×

지털콘텐츠', 4), ('AR/VR', 4), ('플랫폼 서비스', 4), ('데이터 분석', 4), ('IoT', 3), ('임베디드SW', 3), ('스마트팩토리', 2), ('블록체인', 2), ('시스템SW', 2), ('플라우드 컴퓨팅', 2), ('보안', 2), ('내/X', 2), ('지능형 로봇', 2), ('IT영업/관리', 1), ('IT데스트/마케팅', 1), ('자율주행차', 1), ('웨어러를 디바이스', 1), ('필스커팅', 1), ('스마트홈', 1), ('자원어처리', 1), ('웹스케어', 1)]
['- 필수 요구 사항 - Git', 'C#', 'C++ 등을 사용해본 경험 - Ne', 'CSS에 대한 이해 - MySQL등의', 'Github를 이용한 개발 협업을 수행해본 경', 'Goog le Cloud Platform등의 를', 'Memcached등의 인메모리 데이터베이스를 ', 'React Testing Library등을 이', 'Git', 'JavaScript', 'React', 'REST API', 'spring-framewor k', 'Amazon AWS', 'Amazon EC2', 'Android', 'Android Studio', 'Back-end', 'CSS', 'CSS 3', 'CSS3', 'figma', 'Firebase', 'Flutter', 'Front-end', 'Git', 'GitHub', 'Google Tag Manager', 'ios', 'Java', 'JavaScript', 'Python', 'React Native', 'React Native', 'TypeScript', 'Anazon AWS', 'Amazon AWS', 'JavaScript', 'Jouery', 'Anazon AWS', 'JavaScript', 'Jouery', 'Adobe Illustrator', 'Adobe Photoshop', 'embery's', 'Less', 'Amazon AWS', 'Docker', 'Gatt', 'React', 'React', 'Saas', 'Java', 'Ilnux', 'Amazon AWS', 'Docker', 'Git', 'Python', 'React', 'Front-end', 'AwS', 'Docker', 'Git', 'Python', 'React', 'Tensorflow', 'Front-end', 'AwS', 'Docker', 'Git', 'Python', 'React', 'Tensorflow', 'Back-end', 'Front-end', 'AwS', 'Docker', 'Git', 'Python', 'React', 'Tensorflow', 'Amazon AWS', 'CSS', 'Python', 'React', 'Tensorflow', 'Back-end', 'Front-end', 'AwS', 'Docker', 'Git', 'Python', 'React', 'Tensorflow', 'Back-end', 'Front-end', 'AwS', 'Docker', 'Git', 'Python', 'Manzon AWS', 'Tensorflow', 'Manzon AWS', 'Python', 'React', 'Tensorflow', 'Manzon AWS', 'Docker', 'Git', 'Python', 'Rest', 'Tensorflow', 'Back-end', 'Front-end', 'AwS', 'Docker', 'Git', 'Python', 'Manzon AWS', 'Tensorflow', 'Manzon AWS', 'Docker', 'Git', 'Python', 'Rest', 'Tensorflow', 'Back-end', 'Front-end', 'AwS', 'Docker', 'Git', 'Python', 'Manzon AWS', 'Tensorflow', 'Manzon AWS', 'Docker', 'Git', 'Python', 'REST API', 'Tensorflow', 'Back-end', 'Front-end', 'AwS', 'Docker', 'Git', 'Python', 'Manzon AWS', 'Docker', 'Git', 'Pyth

2022 상반기

직무



2022 상반기

기술스택



워크넷데이터

□ 모집요강

식무내용	모집즥종
	DBA-데이터베이스,ERP-시스템분석-설계,비데이타-AI(인공지능)

경력조건	확력	고용형태	모집인원	근무예정지
무관	대졸(4년)	정규직	○명	서울 영등포구, 서울

전공	자격면허	외국어능력	병역대체 복무자채용	고용허가제
전산학·컴퓨터공학 (학 과 : 컴퓨터공학전공) 응용 소프트웨어 공학 (학과 : IT융합응용공학 과)	박데이터분석기사(우대) , 리눅스마스터(우대)		비희망	

우대조건	컴퓨터 활용능력
고용촉진장려금대상자	문서작성 (워드프로세스 활용),표계산 (스프레드시트 활용) ,프레젠테 이션 프로그램 활용
기타우대사항	작업환경
Python을 활용한 머신러닝/딥러닝 개발 경험이 있으신 분 프로젝트 관리 (PM) 경험 S. 백데이터 언어 (R, Python) 개발 가능 리눅스 환경에서 쉘 스크립트 개발 가능 관공서 사업 업무 경험이 있으신분	

```
from bs4 import BeautifulSoup
from selenium import webdriver
import time
from konlpy.tag import Okt
from collections import Counter
import platform
from wordcloud import WordCloud
import matplotlib.pyplot as plt
import csv
import pandas as pd

driver=webdriver.Chrome('chrome\chromedriver.exe')

driver.get('https://www.ictintern.or.kr/homepage/trainingCompany/companyList.do')

driver.find_element_by_xpath('//*[@id="s_PLAN_NO"]/option[3]').click()
driver.find_element_by_xpath('//*[@id="container"]/form/div/div[2]/div/a').click()

data_list=[]
stack=[]
```

▼ 학점연계 프로젝트 인턴십	д	도소개 * 연수업체	* 알림마당	* 고객지원	
연수업체		105	S C	JOB	
♠ 연수업체	국내과정	•		M	
2022년 하반기 인턴십	V 기업명 V 검색	어플입력하세요			
- 전체 2023년 상반기 인단실 2022년 하반기 인단실 2022년 하반기 인단실 2022년 상반기 인단실 국내과정) 2021년 하반기 인탄실 국내과정)	신규				
2023년 상반기 인단십 2022년 생각기 인단실 2022년 생각기 인단실 2021년 해박기 인단실 국내주병 2021년 해박기 인단실 국내주병 2021년 해박기 인단실 국내주병 2021년 생박기 인단실 국내주경 (CT 인단실 대소프라정 2020년 해박기 인단실 국내주경)	실시	ONE OFF	0.91	DIIX C	77 C 761
2022년 성반기 인단성 2022년 독립가 입단된 2022년 학원기 인단성 2021년 학원기 인단성 국내과정 2021년 학원기 인단성 교로발표정 2021년 성반기 인단성 교로발표정 2021년 성반기 인단성 교로발표정 (CT 인단성 작은 필요점		□ 경주 □ 대전 □ 경북 □ 경남	□ 울산 □ 제주	□ 세종 · □	경기 □ 강원

2022년 하반기인	턴십	~	기업명	✔ 검색(어를 입력하세요.					
매칭현황	□ 진행중	□ 마감								
참여이력	전체	○기참여	○ 신규							
코딩테스트	전체	○실시 ○	미실시							
지역	□ 서울	□ 부산	□ 대구	□ 인천	□ 광주	□ 대전	□ 울산	□ 세종	□ 경기	□ 강원
	□ 충북	□ 충남	□ 전북	□ 전남	□ 경북	□ 경남	□ 제주			
직종 📴	전체선택 김	전체해제								
	□ 웹 프로그	그래밍	□ SW아키텍처		이터베이스	□ 서버 프로	2그래밍	□ 모바일 애플리케이	션 🗆 임타	베디드SW
	□ 네트워크	1	□ UI/UX	□ 시:	스템SW	□ 인공지능	서비스	□ 자연어처리	□ 음성	성인식
	□ 데이터 /	시각화	□ 데이터 분석	□ 플	뱃폼 서비스	□ 빅데이터		□ 핀테크	□ 클리	라우드 컴퓨팅
	□ 디지털콘	린텐츠	□ 소셜미디어	□ AR	/VR	□ 스마트팩	토리	□ 스마트홈	□ 지	능형 로봇
	□ 웨어러블	를 디바이스	□ 3D 프린터	□ 반5	도체	□ 자율주행	차	□ 스마트그리드	□ 신기	재생에너지
	□ loT		☐ ICT 융합	□ 헬:	스케어	□ 블록체인		□ 보안	□ 전1	파·위성
	□ 통신서비	스	□ 정보기술전략	□ 교 ፥	육서비스	□ IT영업/	반리	□ IT테스트/마케팅		
직무스택	입력후엔터	또는선택								

합계						266 명 배정/ 175 명 선발		
141	복용 (마감)	스마트모빌리타 장착용 리튬이온배타리팩 인쇄회로기판(Printed Cir··· ICT 용합	인천 연수구		최대 1명	0명	마감	
142	딜리버리랩 (마감)	◦ 프로젝트 명 1 : 외식업 식자제 통합 중개 플랫폼 고도화 ◦ 프로젝트 … 웹 프로그레잉	서울성동구	기참여	최대 2 명	2명	마감	
143	아이코어 (<mark>마감)</mark>	머신비잔용 핵심 제품 개발 스마트릭토리	경기 안양시		최대 1명	0명	마감	
144	유클리드소프트 (마감)	프로젝트 #1 : 인공지능(AI) 기반의 쓰레기 무단투기 행위 탐지 시스템 … 물죗품 서비스	대전 동구	기참여	최대 3 명	2명	마감	
145	플랜아이 <mark>(마감)</mark>	SaaS 기반 웹 사용자 데이터 분석 서비스 데이터 분석	대전 유성구	기참여	최대 3 명	2명	마감	

○ 실습생 직무

직무구분	UI/UX
기술스택	figma
직무내용	1. 앱 내 콘텐츠 디자인 머터리얼(아이콘, UXUI 디자인 보조) 제작 2. 국내/외 외부 채널에 노출되는 디지털 광고 콘텐츠 제작 3. 조직문화 머터리얼 제작(포스터, 현수막, 명함 등)
요구기술	《우대사항》 - 영어, 일본어, 프랑스어, 아랍어 중 하나가 언어 소통에 어려움이 없으신 분 - 피그마, 어도비, 스케치 등의 디자인툴을 다루어 보신 분 - 프리미어, 파이널 컷 등의 영상 제작 툴을 다루어 보신 분 - Google Analytics4 등의 광고 분석툴을 다루어 보신 분 - 영상 및 광고 기획을 경험해보신 분 - 기업의 SNS채널 운영 경험을 통해 오가닉 사용자를 확보해보신 분



```
html=driver.page_source
   all_div=soup.find_all('div',{'class':'selectize-input items not-full has-options has-items disabled locked'})
   for i in all div:
       data_list.append(i.text)
   print(len(data list))
driver.quit()
print(stack)
   with open('bigdata.csv', 'w', encoding='utf-8', newline='') as f:
       wr = csv.writer(f)
       wr.writerow(data_list)
```

```
# 워드 클라우드 (big_list)

counts = Counter(big_list)

tags = counts.most_common(40)

print(tags)

if platform.system() == 'Windows':
    path = r'c:\Windows\Fonts\malgun.ttf'
elif platform.system() == 'Darwin': # Mac OS
    path = r'/System/Library/Fonts/AppleGothic'
else:
    path = r'/system/Library/Fonts/AppleGothic'
else:
    path = r'/usr/share/fonts/truetype/name/NanumMyeongjo.ttf'

wc = WordCloud(font_path=path, background_color="white", max_font_size = 100, width=600, height=400)
cloud = wc.generate_from_frequencies(dict(tags))

plt.figure(figsize=(10, 8))
plt.ais('off')
plt.imshow(cloud)
plt.show()
```

```
[('웹 프로그래밍', 33), ('인공지능 서비스', 17), ('모바일 애플리케이션', 16), ('플랫폼 서비스', 15), ('빅데이터', 7), ('시스템SW', 7), ('디지털콘텐츠', 6), ('ICT
융합', 6), ('서버 프로그래밍', 6), ('임베디드SW ', 6), ('UI/UX', 5), ('보안', 5), ('AR/VR', 4), ('IoT', 3), ('데이터 분석', 3), ('스마트팩토리', 3), ('자율주행차'
, 3), ('클라우드 컴퓨팅', 3), ('지능형 로봇', 2), ('SW아키텍처', 2), ('반도체', 2), ('헬스케어', 1), ('웨어러블 디바이스', 1), ('네트워크 ', 1), ('IT테스트/마케팅
', 1), ('자연어처리', 1), ('교육서비스', 1)]
Counter({'Python': 61, 'JavaScript': 59, 'Java': 47, 'C++': 35, 'React': 32, 'Node.js': 29, 'C#': 26, 'Back-end': 23, 'Front-end': 23, 'MySOL': 21, 'CSS': 21, 'HT
ML': 19, 'Amazon AWS': 18, 'linux': 17, 'Android': 16, 'Tensorflow': 13, 'TypeScript': 12, 'Django': 12, 'Flutter': 11, 'Git': 11, 'GitHub': 10, 'REST API': 10, '
HTML5': 10, 'figma': 9, 'React Native': 9, 'Ubuntu': 9, 'Spring Boot': 9, 'ios': 9, 'Vue.js': 8, 'windows': 8, 'Docker': 8, 'Swift': 8, 'MongoDB': 7, 'jQuery': 7,
 'PostgreSQL': 7, 'Kotlin': 6, 'PHP': 6, 'Spring': 6, 'Next.js': 5, 'nginx': 5, 'CSS3': 5, 'Express': 4, 'Microsoft Azure': 4, 'ASP.NET': 4, 'zeplin': 4, 'MariaDB
': 4, 'CentOS': 4, 'NoSQL': 4, 'spring-framework': 4, 'GraphQL': 3, 'JIRA': 3, 'SaaS': 3, 'Angular': 3, 'Elasticsearch': 3, 'Kafka': 3, 'Jenkins': 3, 'Amazon RDS'
: 3, '인공지능': 3, 'Firebase': 3, 'React-native': 3, 'restful': 3, 'Redis': 2, 'Redux': 2, 'Socket.IO': 2, 'PyCharm': 2, 'MSSql': 2, 'Ionic': 2, 'Visual Studio':
2, 'Android Studio': 2, 'Xcode': 2, 'Kubernetes': 2, 'pytorch': 2, 'unity': 2, 'Amazon S3': 2, 'Go': 2, 'Laravel': 2, 'Kubernetes ': 2, 'Illustrator': 2, 'Azure'
: 2, 'Verilog': 2, '딥러닝': 2, 'Microsoft SQL Server': 2, 'Ruby': 2, 'C': 2, 'AR': 2, 'VR': 2, 'illustrator': 2, 'Grafana': 1, 'npm': 1, 'Postman': 1, 'Adobe Aft
er Effects': 1, 'Adobe Photoshop': 1, 'Adobe Premire': 1, 'Canva': 1, 'Final Cut Pro': 1, 'PaaS': 1, 'Bootstrap': 1, 'OrCAD': 1, 'PADS': 1, 'apache': 1, 'Embedded
': 1, 'firmware': 1, '임베디드': 1, '펌웨어': 1, 'keras': 1, 'AioHTTP': 1, 'RabbitMQ': 1, 'Travis CI': 1, 'postgresql': 1, 'tailwindcss': 1, 'vuejs': 1, 'mfc': 1,
'Amazon CloudFront': 1, 'Apache Spark': 1, 'AWS ELB': 1, 'AWS Lambda': 1, 'Docker Compose': 1, 'CSS 3': 1, 'Photoshop': 1, 'Sketch': 1, '멀티쓰레드 프로그래밍':
1, '윈도우/윈도우 네트워크 프로그래밍': 1, 'YOLO': 1, 'Computer Architecture': 1, '논리회로': 1, '자료구조': 1, 'PyTorch': 1, 'DATA BASE': 1, 'Bitbucket': 1, 'NLP
': 1, 'VirtualBox': 1, '범용 DBMS 지식 보유': 1, '비식별 조치관한 기초': 1, '빅데이터 분산 처리 관한 기초': 1, 'EDA': 1, 'Pvtorch': 1, 'Oueue': 1, 'Recoil': 1, 'S
OL': 1, 'Tensorflow2': 1, 'restful API': 1, 'Eclipse': 1, 'Ruby On Rails': 1, 'Unreal': 1, 'CNN': 1, '데이터베이스': 1, '파이썬': 1, 'Debian': 1, 'GitLab': 1, 'Go
ogle Compute Engine': 1, 'Oracle': 1, 'Sentry': 1, 'SourceTree': 1, 'Unity': 1, 'JSP': 1, 'Amazon EC2': 1, 'mac': 1, 'Microsoft IIS': 1, 'embed': 1, 'micro python
': 1, 'node-red': 1, 'amazone ec2': 1, 'bootstrap': 1, 'centos': 1, 'php': 1, 'postman': 1, 'react': 1, 'saas': 1, 'AngularJS': 1, 'JPA': 1, 'kotlin': 1, 'SQLite'
: 1, 'Data 처리기술': 1, '웹 개발 프레임워크': 1, 'NumPy': 1, 'UML': 1, 'embedded': 1, 'Apache HTTP Server': 1, 'Backbone.js': 1, 'cocos2d': 1, '유니티엔진': 1,
'PhpStorm': 1, 'opencv': 1, 'robot': 1, 'Animate.css': 1, 'HWP': 1, 'InDesign': 1, 'Media Encoder': 1, 'PowerPoint': 1, 'Premiere': 1, 'photoshop': 1, 'UI UX': 1,
'unrealengine': 1, 'FPGA': 1, 'Linear Algebra': 1, 'Linux/Ubuntu': 1, 'Number Theory': 1, 'Auto Layout': 1, 'Constraint Layout': 1, 'Jetpack compose': 1, 'MVVM':
1, 'RESTful API': 1, 'UIKit': 1, 'Objective-C': 1, 'vscode': 1, 'wordscript': 1, 'RESTFul API': 1, 'JSON': 1, 'Google Tag Manager': 1, 'GTM': 1, 'Google Analytic
s': 1, 'adobe photoshop': 1, '미리캔버스': 1})
```

빅데이터: 7개

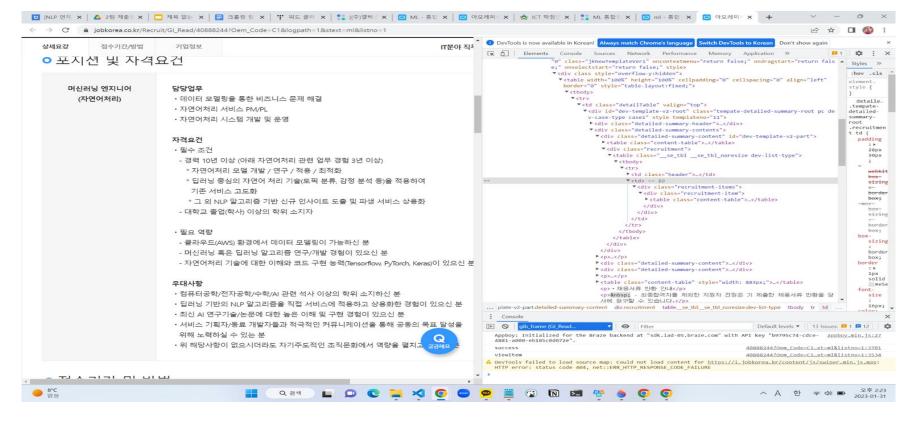
cloud



cloud



장마가 잡코리아 실패



ICT로 전격 이적.

2023년 상반기인	J턴십 ·	기업명	✔ 검색이	어를 입력하세요.					
매칭현황	□ 진행중 □ 마감								
참여이력	◉ 전체 ○ 기참여	○ 신규							
코딩테스트	◉ 전체 ○실시	○ 미실시							
지역	□ 서울 □ 부신	□ 대구	□ 인천	□ 광주	□대전	□ 울산	□ 세종	경기	□ 강원
	□ 충북 □ 충남	□ 전북	□ 전남	□ 경북	□ 경남	□제주			
직종 -	전체선택 전체해제 □ 웹 프로그래밍	□ SW아키텍처		기터베이스	□ 서버 프롤	2 7 24 01	□ 모바일 애플리케이션	1 □ 임베□	TIESW/
	□ 네트워크	□ UI/UX		스템SW	□ 인공지능		□ 자연어처리	□ 음성(
	□ 데이터 시각화	□ 데이터 분석		뱃폼 서비스	□ 백데이트	1	□ 핀테크		우드 컴퓨팅
	□ 디지털콘텐츠	□ 소셜미디어	□ AR	/VR	□ 스마트팩	부토리	□ 스마트홈	□ 지능	형 로봇
	□ 웨어러블 디바이스	□ 3D 프린터	□ 반5	근체	□ 자율주현	성사	□ 스마트그리드	□ 신재성	생에너지
	□ IoT	□ ICT 융합	□ 헬스	스케어	□ 블록체인	Í	□ 보안	□ 전파	·위성
	□ 통신서비스	□ 정보기술전략	<u></u>	육서비스	□ IT영업/	관리	□ IT테스트/마케팅		
직무스택	입력 후 엔터 또는 선택								

전체 170 건

※ 클릭하면 상세내용이 보입니다. 마감된 기업은 저원불가합니다

번호	기업명 ▼	프로젝트명	소재지	참여이력	배정 인원 ▼	채용 현황 ▼	매칭 현황
10	테크랩스	빅데이터 개인화 추천 서비스의 상품기반 성과 분석 결과 화면 개발 웹 프로그래밍 데이터 시각화 UI/UX 빅데이터 클라우드 컴퓨팅	서울 강남구	기참여	최대 1 명	0 명	진행중
9	셈웨어	Unity 게임엔진 기반의 메타버스 서비스(교육, 제조, 환경플랜트 등… 웹 프로그래밍 플랫폼 서비스 디지털콘텐츠 IT테스트/마케팅 데이터베이스 시스템SW 연공지능서비스 교육서비스 네트워크	서울 관악구	기참여	최대 2 명	0 명	진행중

2023 상반기

cloud



코드

```
html = driver.page source
soup = BeautifulSoup(html, 'html.parser')
result = []
for i in range(1,10):
    driver.find element by xpath(f'//*[@id="pagingNavi"]/div[2]/a[{i}]').click()
    time.sleep(1)
    item = soup.select('tbody>tr>td>a>p>div>div>div')
    for x in item:
        result.append(x.text)
print(result)
it table = pd.DataFrame(result)
print(it table.head())
it_table.to_csv('it_table.csv', encoding='utf-8', mode='w', index=True)
counts = Counter(result)
tags = counts.most_common(40)[1:]
print(tags)
```

2023 상반기

그래프



```
html = driver.page_source
soup = BeautifulSoup(html, 'html.parser')
result = []
stack=[]
for m in range(1,10):
    driver.find_element_by_xpath(f'//*[@id="pagingNavi"]/div[2]/a[{m}]').click()
        for m in range(1,20):
            driver.find_element_by_xpath(f'//*[@id="container"]/form/div/div[3]/div[2]/table/tbody/tr[{m}]/td[3]/a').click()
            html=driver.page_source
            soup = BeautifulSoup(html, 'html.parser')
            std wokr = soup.find_all('tbody')[13]
            std_wokr_tr = std_wokr.find_all('tr')
            for stack_j in std_wokr_tr[1].select('td div div div.item'):
                stack.append(stack j.text)
            driver.find_element_by_xpath(f'//*[@id="contents_area"]/div/div[3]/a').click()
        continue
print(stack)
```

```
0,JavaScript
     1,Python
     2,Java
     3,JavaScript
     4,C#
     5,C++
     6,Java
     7, Python
     8,C#
10
11
     9, Python
12
     10,C++
13
     11,CSS
14
     12,HTML
15
     13, JavaScript
16
     14,C++
17
     15, MySQL
18
     16, Python
19
     17,Kotlin
20
     18, Python
21
     19, React
22
     20,Swift
23
     21,R
24
     22, Amazon AWS
25
     23, JavaScript
26
     24,linux
     25,MSSql
```

,0

4번째 - 원티드드



소스코드에는 있으나 출력되는 것이 없음.

```
# options = webdriver.ChromeOptions()
     options = webdriver.ChromeOptions()
     options.add experimental option('excludeSwitches',
     ['enable-logging'])
     driver = webdriver.Chrome(options=options)
11
12
13
     html = r'https://www.wanted.co.kr/wd/143115'
14
     driver.get(html)
15
16
     html = requests.get(html)
     bs = BeautifulSoup(html.text, 'html.parser')
17
     # print(bs)
18
19
     for i in bs.select('h6'):
20
21
         print(i)
22
```

```
ROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PROJECT\job_ja> conda activate base base) PS C:\Users\user\KNU\PROJECT\job_ja> & C:\ProgramData/Anacond 3/python.exe c:\Users\user\KNU\PROJECT\job_ja> & C:\Users\user\KNU\PROJECT\job_ja> & C:\Users\user\KNU\PROJECT\job_ja> & C:\Users\user\KNU\PROJECT\job_ja>
```

이소정

빅데이터 살펴 보기

데이터불러오기

번호	기업명▼	프로젝트명	소재지	참여이력	배정 인원 ▼	채용 현황 ▼	매칭 현황	
23	아이카	인공지능 IOT 차량 운행 데이터 및 전기차(EV) 데이터 분석 서비스 개발 웹 프로그래밍 데이터베이스 모바일 애플리케이션 빅데이터 데이터 분석 IT테스트/마케팅	경기 성남시	기참여	최대 1 명			t('202203','202203_기업 'C')
22	포테이토넷	보안위협 분석 및 빅데이터 인공지능 위협 탐지 모델 개발 인공지능 서비스 자연어처리 데이터 시각화 데이터 분석 빅데이터 보안 데이터베이스 웹 프로그래밍	서울 광진구		최대 1 명	0명	진행중	
21	화이트스캔	답러닝 기반의 의사결정 지원 플랫폼 개발(자연어 처리, 예측 모델링) 인공지능서비스 빅데이터 자연어처리 플랫폼서비스	서울 용산구	기참여	최대 2 명			t('202203','202203_기업 'C')
20	대성에너지	통합 데이터 분석 시스템 구축(시범사업) 데이터 분석 빅데이터 데이터 시각화 데이터베이스	대구 중구		최대 1 명	0명	진행중	
		P #1. 딥러닝 기반의 공사 현장 및 다양한 작업공간에서의 이상 행동 ···						

데이터불러오기

```
data_list, com, work, stack = [], [], []
for _ in range(2):
    html = driver.page source
    bs = BeautifulSoup(html, 'html.parser')
    href_c = bs.select('.left a')
   for i in href c:
        try:
            script url = i['href'].split(':')[1]
            print('script url: ', script url)
            driver.execute script(script url)
            time.sleep(2)
            html = driver.page source
            bs = BeautifulSoup(html, 'html.parser')
            ht = bs.find all('tbody')[4]
            com.append(ht.find_all('tr')[7].text)
```

자연어처리 및 wordcloud 함수 만들기

```
def remove stopword(text):
                                                def make wordcloud(text):
   tokens = okt.pos(text)
                                                    wc = WordCloud(font path = path, background color
   print(tokens)
                                                = 'white', max font size = 100, width=600, height=400)
   token list = []
   for word, tag in tokens:
                                                    counts = Counter(text)
       if tag in ['Noun', 'Adjective',
                                                    counts = counts.most common(50)
'Alpha'l:
                                                    print(counts)
           token list.append(word)
   stopwords = ['및', '서비스', '위', '데이터',
                                                    cloud = wc.generate from frequencies(dict(counts))
'플랫폼', '개발', '계', '등', '분석', '기반', '사
업', '기여', '공간', '비대', '통합', '통해', '관
                                                    plt.figure(figsize = (10, 8))
련', '체제', '0', 'B', '분야', '제공', '기술', '
                                                    plt.axis('off')
국가', '헬', '스케', '장비', '영역', '소재']
                                                    plt.imshow(cloud)
   token list = [t for t in token list if t
                                                    plt.show()
not in stopwords]
   print(token list)
                                                make wordcloud(info noun)
   return token list
                                                make wordcloud(work)
                                                make wordcloud(stack)
```

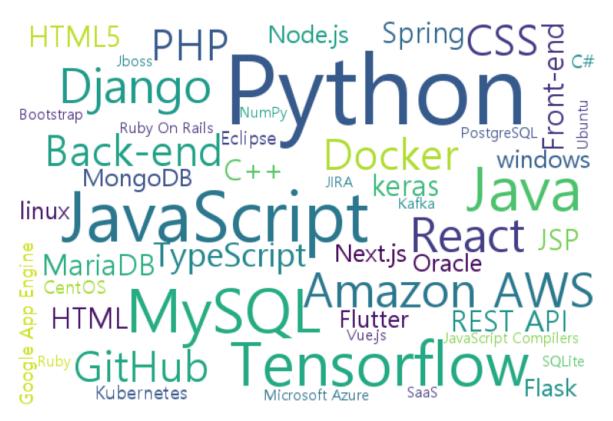
빅데이터 기업들의 사업분야



빅데이터 직무



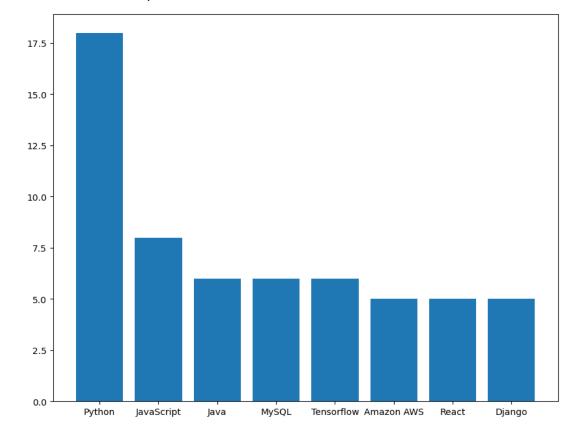
빅데이터 기술스택



빅데이터 기술스택

```
count = Counter(stack)
stack_count = count.most_common(8)
stack_count = dict(stack_count)
plt.figure(figsize = (10, 8))
plt.bar(stack_count.keys(),
stack_count.values())
plt.show()
```

파이썬, JavaScript, JAVA, SQL, Tensorflow 중요!!!!



결론

- 빅데이터에 대한 기업의 관심과 수요가 시간이 지남에 따라 커짐
- 빅데이터 직무와 함께 볼 수 있는 직무는 데이터 시각화, 데이터 분석 등 이 있음
- 빅데이터 직무 스텍은 파이썬, JavaScript, JAVA, SQL, Tensorflow 임