**중간 보고서**

**T-SA: Twitter keyword Search API based tweet Analysis**

**(트위터 키워드 검색 API기반 트윗 분석)**



과 목: 산학캡스톤디자인1

분 반: 01

담 당 교 수: 정현숙 교수님

팀 명: 브이아이(VI)

조 원: 이석준(20165072)

이윤혁(20165062)

배인규(20165073)

서재익(20144773)

**- 목 차 -**

**제1장 서 론**

제1절 개발 동기 및 필요성 --------------------------------------------------------- 1

제2절 기존 연구 ------------------------------------------------------------------- 1

제3절 개발 목표 ------------------------------------------------------------------- 1

제4절 시스템 명세 ----------------------------------------------------------------- 2

1.4.1. 시스템 동작 과정 ----------------------------------------------------------- 2

1.4.2. 원본 데이터 SET ----------------------------------------------------------- 2

**제2장 본 론**

제1절 개발 환경

2.1.1. Ubuntu

2.1.2. Python

2.1.3. MariaDB

2.1.4. Eclipse, JDK

2.1.5 Hadoop Echo System

제2절 TwitterAPI

제3절 T-SA 흐름도

제4절 구현(Python)

2.2.1. TwitterAPI.py

2.2.2. DBModule.py

2.2.3. Visual.py

제5절 구현(DB)

2.3.1. DB/Table 정의서

제6절 구현(Hadoop)

2.4.1. Sqoop(DB/HDFS 연동)

2.4.2. Map/Reduce

**제3장 결 론**

제1절 실행 화면

제2절 참고문헌 및 사이트

**-부 록-**

1. 부 록 1 (깃허브)

제1장 서 론

**제1절 개발 동기 및 필요성**

SNS(Social Networking Service)는 사용자 간의 자유로운 의사소통, 정보공유 그리고 인맥 확대 등을 통하여 사회적 관계를 생성하고 강화해주는 온라인 플랫폼을 말한다. 대표적인 SNS로는 페이스북, 인스타그램, 블로그, 트위터 등이 존재한다. 그중에서도 기업, 공공기관, 정부 기관과 같은 단체뿐만 아니라 연예인, 국회의원, 대통령 등 공인이 다른 SNS보다 상대적으로 많이 사용하는 트위터를 선택하게 되었다.

단체가 작성한 트윗, 팔로잉 등을 분석하여 그 단체의 슬로건, 비전, 성향, 인재상 등 다양한 정보를 예상하거나 수집하는데 이용할 수 있을 거라 생각한다.



**제2절 기존 연구**

|  |  |
| --- | --- |
|  |  |
| URL: <http://tweetrend.com/>  Input: 기간, 단어  Ouput: 기간별 언급 횟수 그래프, 트윗정보 | URL: <https://foller.me/>  Input: 사용자 아이디  Output: 사용자의 정보 |

**제3절 개발 목표**

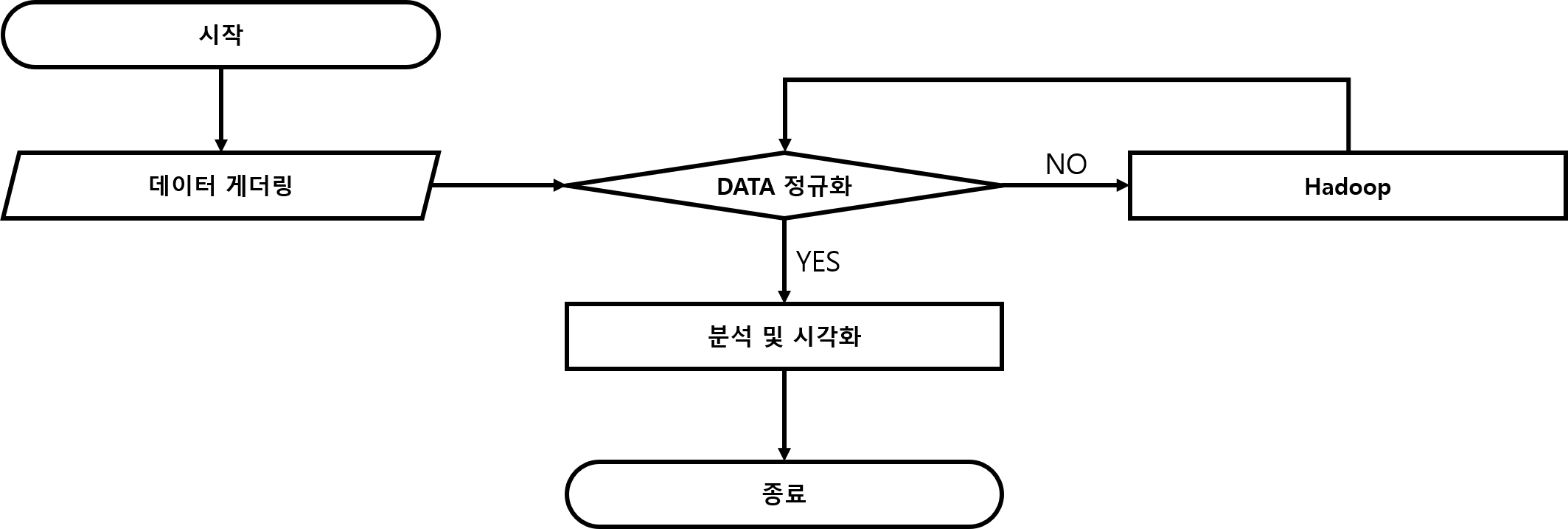
이 프로젝트에서는 개발 동기 및 필요성과 기존 연구를 토대로 2가지의 목표를 정하였다.

첫 번째로는, 키워드와 기간을 설정하여 해당 기간동안 작성된 트윗을 가져와서 포함된 단어와 해시태그의 빈도수를 분석하고 시각화할 것이다.

두 번째로는, 사용자의 아이디혹은 이름을 통해 해당 사용자의 정보를 가져와서 분석하고 시각화할 것이다.

**제4절 시스템 명세**

**1.4.1. 시스템 동작 과정**



**1.4.2. 원본 데이터 SET**

|  |
| --- |
| Status(\_api=<tweepy.api.API object at 0x7f6fd80f87b8>,\_json={'created\_at':'Sun Apr 28 08:12:14 +0000 2019','id':1122413198013911040,'id\_str': '1122413198013911040','text':'aaaaaaaaa 이윤혁','truncated':False,'entities':{'hashtags':[],'symbols':[],'user\_mentions':[],'urls':[]},'metadata': {'result\_type':'recent','iso\_language\_code':'ko'},'source':'<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>','in\_reply\_to\_status \_id':None,'in\_reply\_to\_status\_id\_str':None,'in\_reply\_to\_user\_id':None,'in\_reply\_to\_user\_id\_str':None,'in\_reply\_to\_screen\_name':None,'user':{'id':1107113829123555328,'id\_str':'1107113829123555328','name':'Lee SeokJune','screen\_name':'LSeokjune','location':'','description':'','url':None,'entiti es':{'description':{'urls':[]}},'protected':False,'followers\_count':0,'friends\_count':0,'listed\_count':0,'created\_at':'Sun Mar 17 02:58:00 +0000 2019 ','favourites\_count':0,'utc\_offset':None,'time\_zone':None,'geo\_enabled':False,'verified':False,'statuses\_count':8,'lang':'ko','contributors\_enabled':False,'is\_translator':False,'is\_translation\_enabled':False,'profile\_background\_color':'F5F8FA','profile\_background\_image\_url':None,'profile\_background\_image\_url\_https':None,'profile\_background\_tile':False,'profile\_image\_url':'http://abs.twimg.com/sticky/default\_profile\_images/default\_profile\_normal.png','profile\_image\_url\_https':'https://abs.twimg.com/sticky/default\_profile\_images/default\_profile\_normal.png','profile\_link\_color':'1DA1F2','profile\_sidebar\_border\_color':'C0DEED','profile\_sidebar\_fill\_color':'DDEEF6','profile\_text\_color':'333333','profile\_use\_background\_image':True,'has\_extended\_profile':False,'default\_profile':True,'default\_profile\_image':True,'following':False,'follow\_request\_sent':False,'notifications':False,'translator\_type':'none'},'geo':None,'coordinates':None,'place':None,'contributors':None,'is\_quote\_status':False,'retweet\_count':0,'favorite\_count':0,'favorited':False,'retweeted':False,'lang':'ko'},created\_at=datetime.datetime(2019,4,28,8,12,14),id=1122413198013911040,id\_str='1122413198013911040',text='aaaaaaaaa 이윤혁',truncated=False,entities={'hashtags':[],'symbols':[],'user\_mentions':[],'urls':[]},metadata={'result\_type':'recent','is o\_language\_code':'ko'},source='Twitter Web Client',source\_url='http://twitter.com',in\_reply\_to\_status\_id=None,in\_reply\_to\_status\_id\_str=None,In\_ reply\_to\_user\_id=None,in\_reply\_to\_user\_id\_str=None,in\_reply\_to\_screen\_name=None,author=User(\_api=<tweepy.api.API object at 0x7f6fd80f87b8 >,\_json={'id':1107113829123555328,'id\_str':'1107113829123555328','name':'Lee SeokJune','screen\_name':'LSeokjune','location':'','description':'','url ':None,'entities':{'description':{'urls':[]}},'protected':False,'followers\_count':0,'friends\_count':0,'listed\_count':0,'created\_at':'Sun Mar 17 02:58:00 +0 000 2019','favourites\_count':0,'utc\_offset':None,'time\_zone':None,'geo\_enabled':False,'verified':False,'statuses\_cou nt':8,'lang':'ko','contributors\_ enabled':False,'is\_translator':False,'is\_translation\_enabled':False,'profile\_background\_color':'F5F8FA','profile\_background\_image\_url':None,'profile\_background\_image\_url\_https':None,'profile\_background\_tile':False,'profile\_image\_url':'http://abs.twimg.com/sticky/default\_profile\_images/default\_profile\_normal.png','profile\_image\_url\_https':'https://abs.twimg.com/sticky/default\_profile\_images/default\_profile\_normal.png','profile\_link\_color':'1DA1F2','profile\_sidebar\_border\_color':'C0DEED','profile\_sidebar\_fill\_color':'DDEEF6','profile\_text\_color':'333333','profile\_use\_background\_image':True,'has\_extended\_profile':False,'default\_profile':True,'default\_profile\_image':True,'following':False,'follow\_request\_sent':False,'notifications':False,'translator\_type':'none'},id=1107113829123555328,id\_str='1107113829123555328',name='Lee SeokJune',screen\_name='LSeokjune',locati on='',description='',url=None,entities={'description':{'urls':[]}},protected=False,followers\_count=0,friends\_count=0,listed\_count=0,created\_at=datetime.datetime(2019,3,17,2,58),favourites\_count=0,utc\_offset=None,time\_zone=None,geo\_enabled=False,verified=False,statuses\_count=8,lang='ko',contributors\_enabled=False,is\_translator=False,is\_translation\_enabled=False,profile\_background\_color='F5F8FA',profile\_background\_image\_url=None,profile\_background\_image\_url\_https=None,profile\_background\_tile=False,profile\_image\_url='http://abs.twimg.com/sticky/default\_profile\_images/default\_profile\_normal.png',profile\_image\_url\_https='https://abs.twimg.com/sticky/default\_profile\_images/default\_profile\_normal.png',profile\_link\_color='1DA1F2',profile\_sidebar\_border\_color='C0DEED',profile\_sidebar\_fill\_color='DDEEF6',profile\_text\_color='333333',profile\_use\_background\_image=True,has\_extended\_profile=False,default\_profile=True,default\_profile\_image=True,following=False,follow\_request\_sent=False,notifications=False,translator\_type='none'),user=User(\_api=<tweepy.api.API object at 0x7f6fd80f87b8>,\_json={'id':1107113829123555328,'id\_str':'110711382912 3555328','name':'Lee SeokJune','screen\_name':'LSeokjune','location':'','description':'','url':None,'entities':{'description':{'urls':[]}},'protected':Fals e,'followers\_count':0,'friends\_count':0,'listed\_count':0,'created\_at':'Sun Mar 17 02:58:00 +0000 2019','favourites\_count':0,'utc\_offset ':None,'time\_zone':None,'geo\_enabled':False,'verified':False,'statuses\_count':8,'lang':'ko','contributors\_enabled':False,'is\_translator':False,'is\_translation\_enabled':False,'profile\_background\_color':'F5F8FA','profile\_background\_image\_url':None,'profile\_background\_image\_url\_https':None,'profile\_background\_tile':False,'profile\_image\_url':'http://abs.twimg.com/sticky/default\_profile\_images/default\_profile\_normal.png','profile\_image\_url\_https':'https://abs.twimg.com/sticky/default\_profile\_images/default\_profile\_normal.png','profile\_link\_color':'1DA1F2','profile\_sidebar\_border\_color':'C0DEED','profile\_sidebar\_fill\_color':'DDEEF6','profile\_text\_color':'333333','profile\_use\_background\_image':True,'has\_extended\_profile':False,'default\_profile':True,'default\_profile\_image':True,'following':False,'follow\_request\_sent':False,'notifications':False,'translator\_type':'none'},id=1107113829123555328,id\_str='1107113829123555328',name='Lee SeokJune',screen\_name='LSeokjune',location='',description='',url=None,entities={'des cription':{'urls':[]}},protected=False,followers\_count=0,friends\_count=0,listed\_count=0,created\_at=datetime.datetime(2019,3,17,2,58),favourites\_count=0,utc\_offset=None,time\_zone=None,geo\_enabled=False,verified=False,statuses\_count=8,lang='ko',contributors\_enabled=False,is\_translator=False,is\_translation\_enabled=False,profile\_background\_color='F5F8FA',profile\_background\_image\_url=None,profile\_background\_image\_url\_https=None,profile\_background\_tile=False,profile\_image\_url='http://abs.twimg.com/sticky/default\_profile\_images/default\_profile\_normal.png',profile\_image\_url\_https='https://abs.twimg.com/sticky/default\_profile\_images/default\_profile\_normal.png',profile\_link\_color='1DA1F2',profile\_sidebar\_border\_color='C0DEED',profile\_sidebar\_fill\_color='DDEEF6',profile\_text\_color='333333',profile\_use\_background\_image=True,has\_extended\_profile=False,default\_profile=True,default\_profile\_image=True,following=False,follow\_request\_sent=False,notifications=False,translator\_type='none'),geo=None,coordinates=None,place=None,contributors=None,is\_quote\_status=False,retweet\_count=0,favorite\_count=0,favorited=False,retweeted=False,lang='ko') |

- TwitterAPI 중 Search메서드의 결과값 중 하나의 Row에 해당하는 값

제2장 본 론

**제1절 개발 환경**

|  |  |
| --- | --- |
| Ubuntu | 18.04.2 LTS |
| Python | 3.6.8 |
| MariaDB | 10.1.38 |
| Eclipse | 2019-03 (4.11) |
| JDK | 1.8.0\_191 |
| Hadoop  Sqoop | 3.2.0  1.4.7 |

**1.2.1 Ubuntu (18.04.2 LTS)**

제3장 결 론

1

-부 록-

1