

2018년도 연구계획

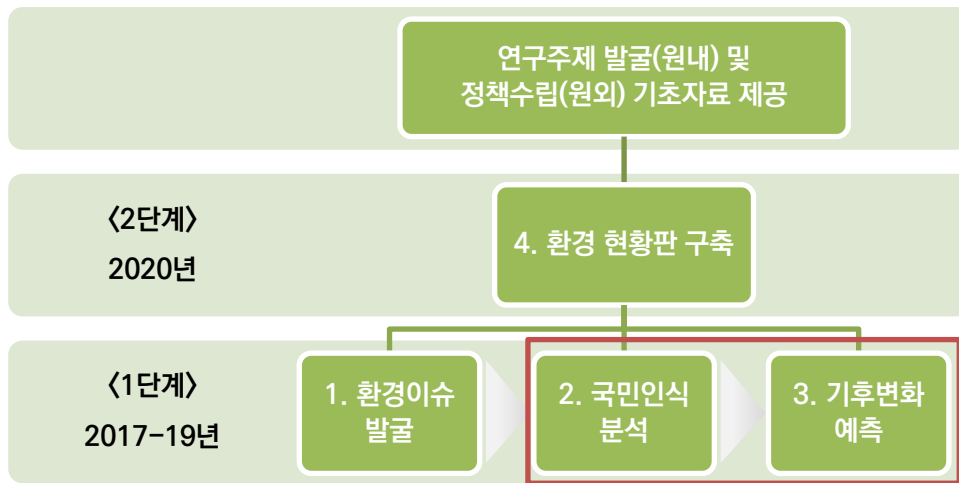
2018.01.04. 김도연

2018-20년도 연구계획

□ 향후 계획

○ 환경이슈 현황판 구축

- 실시간 온라인 텍스트 데이터 '수집 > 전처리 > 분석 > 시각화' 자동화 구현
- 분석은 '환경이슈 발굴(Text Mining) > 국민인식 분석(Opinion Mining) > 이슈 예측'순으로 수행하고, 분석결과를 실시간 대시보드 형태로 구현
- 실시간 환경이슈 현황판을 토대로, 주기적으로 분석결과를 원내외 제공

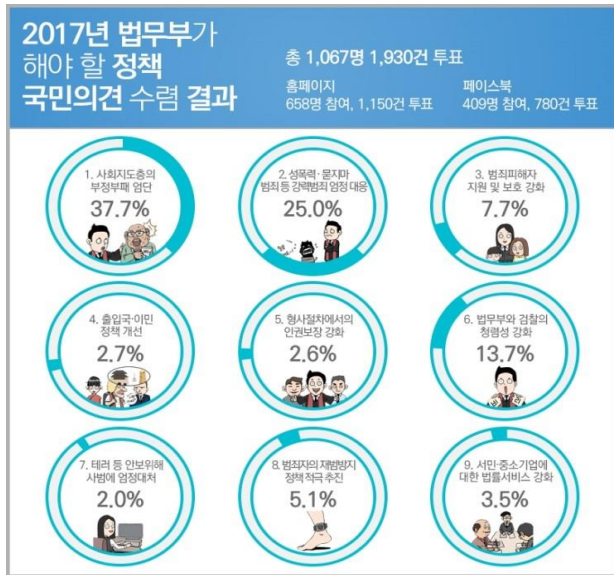


〈3차년도 계획 (2018-20)〉

연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석 및 탐지

□ 연구의 필요성 및 목적

- 환경이슈에 대한 정책이 국민의 의견을 수렴하고 있는 지 여부에 대한 회의 존재
 - 국민의견 수렴은 설문지, 전문가 의견, 콜센터 등의 전통적인 방법으로 진행되고 있음
 - 기존의 전통적인 방법은 표본의 수가 적고 시간과 비용이 많이 소비됨



〈법무부〉

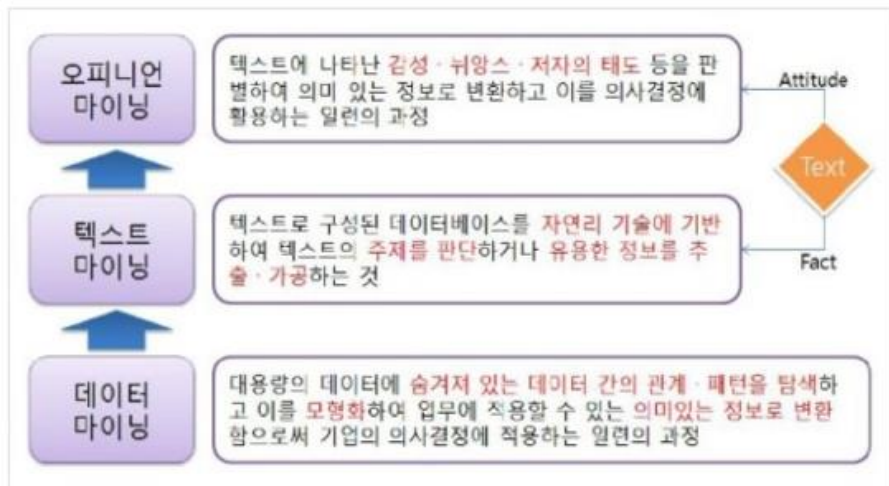


〈교육부〉

연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석

□ 연구의 필요성 및 목적

- 실시간으로 생산되는 소셜 빅데이터의 오피니언 마이닝(Opinion Mining)을 통해 환경이슈에 대한 국민인식 파악 가능
 - 오피니언 마이닝은 텍스트에 나타난 감성, 태도 등을 판별하여 의미 있는 정보로 변환하고 이를 의사결정에 활용
 - 환경이슈 관련 국민인식이 다량으로 표출되는 온라인 채널의 텍스트를 기반으로 오피니언 마이닝을 수행하여 국민의 니즈를 발굴하는 작업 가능



<그림 1> 데이터 마이닝 분류

연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석

□ 연구의 필요성 및 목적

- 소셜 미디어에 표출되는 환경이슈에 대한 국민인식을 분석하여, 향후 환경정책 수립을 위한 기초자료 제공에 기여하고자 함
 - 성공적인 정책운영을 위한 선제적 대응에 본 연구 결과를 활용할 수 있음
 - 기존의 전통적인 방식을 보완하는 방법으로 ‘오피니언 마이닝을 이용한 국민인식 분석 방법’의 예를 제공

연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석

□ 선행연구 현황 및 선행연구와 본 연구의 차별성

구 분		선행연구와의 차별성			
		연구목적	연구방법	주요 연구내용	
주요 선행 연구	분석 방법	1	<ul style="list-style-type: none"> - 과제명: 2016년 소셜 빅데이터 기반 보건복지 이슈 동향 분석 - 연구자(년도): 송태민 외(한국보건사회연구원, 2016) - 연구목적: 소셜 빅데이터 기반으로 우리나라 주요 보건복지 정책 수요를 예측 	<ul style="list-style-type: none"> - 연관분석 - 의사결정나무 - 감성분석 	<ul style="list-style-type: none"> - 5개 주제(보건복지 정책 미래신호 예측, 담배, 통일, 어린이 예방접종 온톨로지 개발, 지카바이러스 온톨로지 개발)에 대해 소셜 빅데이터를 수집하여 동향 분석과 수요 예측 및 위험 예측 모형 제시
		2	<ul style="list-style-type: none"> - 과제명: 소셜미디어 빅데이터 분석을 통한 캠핑에 대한 정부 정책과 국민들의 인식 변화 연구 - 연구자(년도): 장미화 외(관광연구 31(1), 2016) - 연구목적: 캠핑에 대한 정부 정책과 국민들의 인식을 분석 	<ul style="list-style-type: none"> - 의미연결망분석 	<ul style="list-style-type: none"> - 2012년 부터 2015년 10월까지 캠핑에 대한 정부 정책과 국민들의 인식을 연도별 분석
		3	<ul style="list-style-type: none"> - 과제명: Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll - 연구자(년도): CODY, Emily M, et al (2015.08) - 연구목적: 해도노 미터(Hedonometer)를 이용하여 '기후' 포함된 트윗의 일별, 주별, 월별 행복도 측정 	<ul style="list-style-type: none"> - 감성분석 	<ul style="list-style-type: none"> - '기후' 키워드가 포함된 트윗(20089~20147) 데이터를 기반으로 행복도 측정 - 트위터가 기후변화 인식 확산의 중요한 요소임을 발견함

연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석

□ 선행연구 현황 및 선행연구와 본 연구의 차별성

구 분			선행연구와의 차별성		
			연구목적	연구방법	주요 연구내용
주요 선행 연구	분석 대상	1			
		2			
		3			
본 연구			<ul style="list-style-type: none"> - 기존의 연구들은 분석 매체와 범위가 한정된 소셜 데이터를 기반으로 사회현상을 분석하여 분석결과가 제한적임, 새로운 이슈 발굴이 어려움 - 일정기간의 사회현상을 분석하는데 그쳐 새로운 이슈 발굴을 위한 실시간 분석에 한계가 존재함 - 자연현상에 대한 오피니언 마이닝과 예측연구는 초기단계에 있다고 판단함 	<ul style="list-style-type: none"> - 오피니언 마이닝 - 예측 분석 - 실시간 자동화 구현 	<ul style="list-style-type: none"> - 소셜 데이터의 특성을 고려하여 실시간 자동화 분석 연구가 필요 - 텍스트 마이닝 연구 대상이 사회현상이 아닌, 자연현상을 대상으로 예측하는 선도적 연구가 필요

연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석

□ 연구 내용

- ① 오피니언마이닝을 활용한 이상기온에 대한 국민인식 분석
- ② 이상기온 예측 연구 가능성 분석

연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석

□ 연구 추진 방법

① 환경이슈 선정

- 기후변화 세부 현상 키워드 : 이상기온(폭염, 한파, 이상온난 등)

· 텍스트 마이닝을 이용한 환경이슈 분석결과 기후변화 키워드가 주요키워드로 나타났으며, 향후 기후변화의 세부현상에 대한 연구가 유망한 연구 주제로 파악 (환경 빅데이터 분석 및 서비스 개발, 2017)

· SNS 데이터는 지역별 수집이 어려움으로 지리적 특성을 덜 받는 이상기온 현상을 선정

② 이상기온 현상 키워드 사전 구축

- 텍스트 데이터 수집(웹 크롤링)을 진행하기에 앞서 불필요한 데이터의 수집을 제거하기 위해 이상기온 주제에 적합한 키워드 사전 구축이 필요함

· 키워드 사전은 향후 도출되는 분석 결과에 직접적으로 영향을 미치며, 결과의 신뢰성을 확보하기 위해 가장 중요한 작업이라 할 수 있음

- 이상기온 현상 별(폭염, 한파, 이상온난 등) 키워드 사전 구축

· 환경관련 생산문서에서 머신러닝 기법(Word2Vec)을 이용하여 이상기온 현상과 근접한 순으로 키워드 추출

· 추출한 키워드를 환경 전문가 집단의 의견을 반영하여 키워드 선정 및 이상기온 현상 별 키워드 사전 DB 구축

연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석

□ 연구 추진 방법

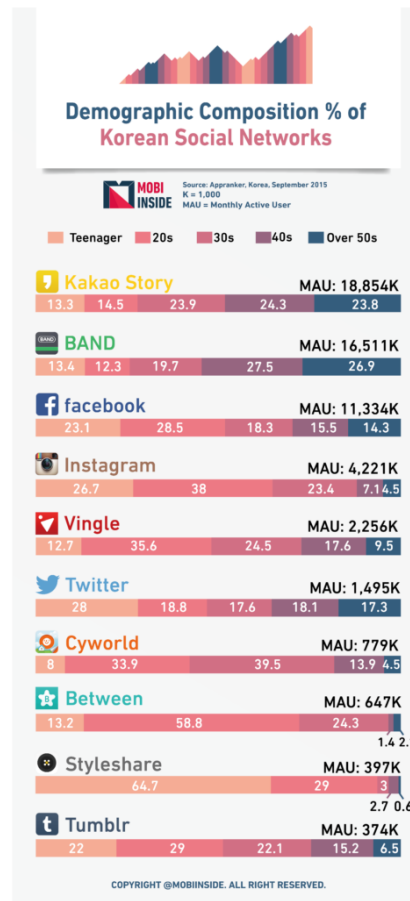
③ 데이터 수집

- 환경이슈 관련 국민인식이 표출되는 온라인 채널 확인
 - SNS(Social Network Service)*와 온라인 댓글**에서 발생하는 구전 데이터 수집

* SNS: Twitter, Facebook, Instagram

** 온라인 댓글: 뉴스 댓글, 블로그 댓글

- 온라인 채널별 이용자 연령층 분포 확인 (우측 그림 참조 : 2015 기준)
- 이상기온 현상 키워드 사전을 기반으로 온라인 텍스트 데이터 수집
 - 빅데이터팀 서버에 웹 크롤링 자동화 구현 (2018.2.3~ 2018.3.31[8회] 파이썬을 활용한 실전 웹크롤링 CAMP)
- 소셜데이터 수집 시 위치 데이터 수집 고민



연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석

□ 연구 추진 방법

- ④ 오피니언 마이닝

- ⑤ 예측 연구 가능성 확인

연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석

□ 향후 계획

○ 환경이슈 예측 분석

- 환경이슈에 영향을 미치는 버즈 데이터를 기반으로 예측 알고리즘 구현

1. 구글 : 독감예보 시스템

구글은 사용자들이 입력한 검색어를 토대로 독감에 걸리면 나타나는 증상들에 대한 검색어가 얼마나 자주 검색되었는지 분석하여 독감 확산을 예측했습니다. 그 결과 미국 질병통제 예방센터의 공표보다 1주일~10일 정도 독감의 창궐을 탐지해냈습니다.



미국: 미국 질병 통제 센터에서 공개한 인플루엔자 의사환자(LI) 데이터

기후변화 세부현상 예측 분석

예)

1. 폭염 예측
2. 한파 예측
3. 미세먼지 예측

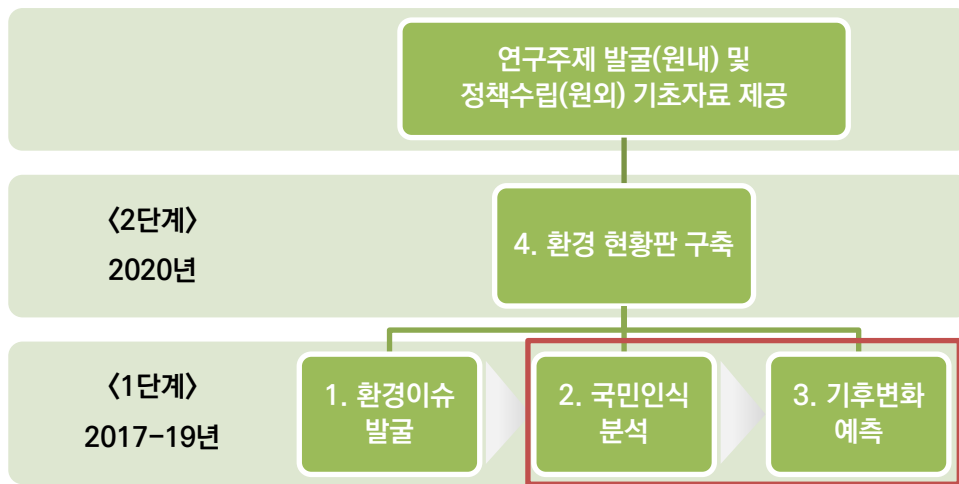
- 이상현상 전 날 SNS에 언급된 이상현상 관련 버즈량을 통해 예측

연구주제 : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석

□ 향후 계획

○ 환경이슈 현황판 구축

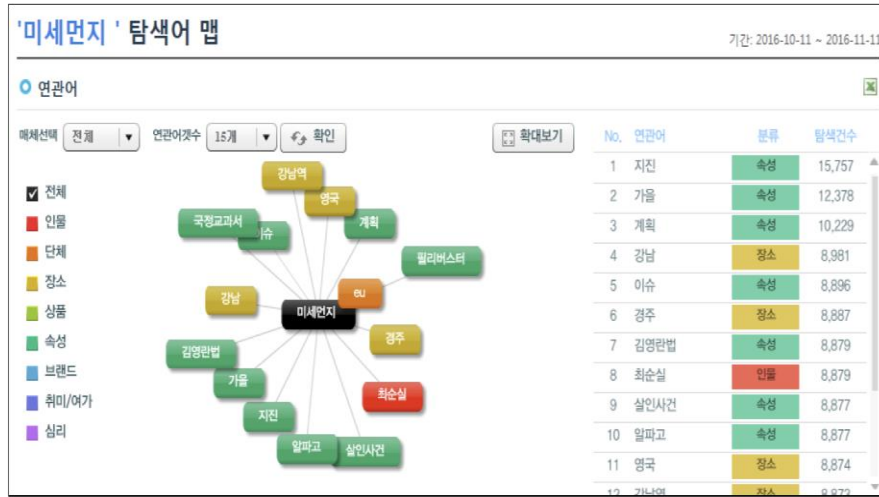
- 실시간 온라인 텍스트 데이터 '수집 > 전처리 > 분석 > 시각화' 자동화 구현
- 분석은 '환경이슈 발굴(Text Mining) > 국민인식 분석(Opinion Mining) > 이슈 예측'순으로 수행하고, 분석결과를 실시간 대시보드 형태로 구현
- 실시간 환경이슈 현황판을 토대로, 주기적으로 분석결과를 원내외 제공



〈3차년도 계획 (2018-20)〉

소셜 빅데이터 분석 서비스 사례(1)

- 소셜메트릭스(다음소프트) <http://www.socialmetrics.co.kr/campaign.html>
 - 검색한 키워드에 대해 소셜미디어(트위터, 블로그, 인스타그램 등)에서 노출된 빈도 추이와 관련 연관어맵 제공
 - 키워드에 대한 긍부정 감정정보 및 날짜별로 가장 많이 확산된 메시지 제공

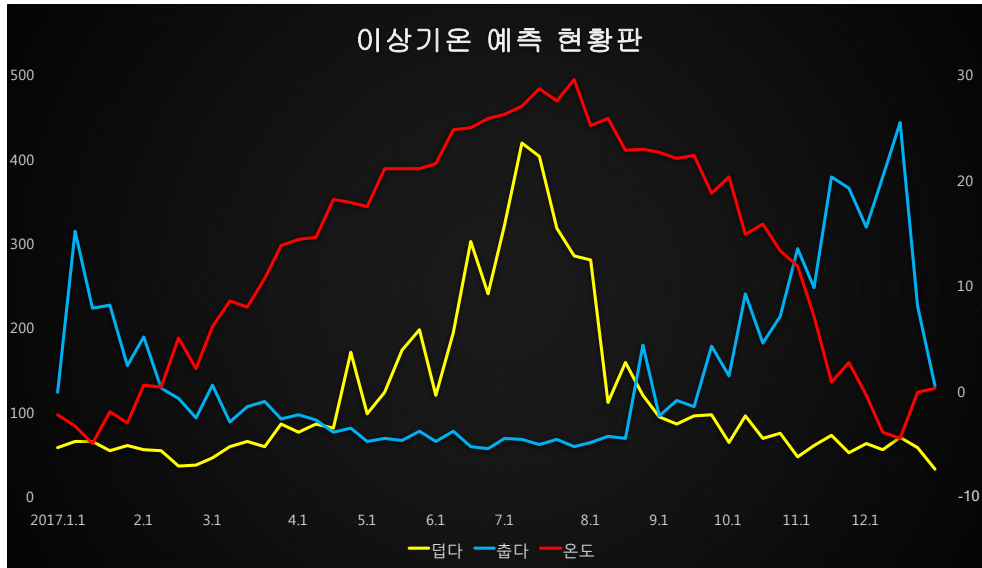


소셜 빅데이터 분석 서비스 예(3)

- 실시간 소셜 데이터를 이용한 이상기온 예측 현황판 구축

- 기간: 2017.01.01~ 2017.12.31

- 대시보드에 이용한 데이터 : 주별 평균 우리나라 기온, 답다 검색량(구글, 네이버), 춥다 검색량(구글, 네이버)



2018년도 연구관리

• 연구(가제) : 소셜 빅데이터를 통한 이상기온에 대한 국민인식 분석 및 예측

구분	'17. 12	1	2	3	4	5	6	7	8	9	10	11	12
논문 작업	■												
연구주제 선정		■											
1. 사전구축			■										
2. 데이터 수집 자동화 + 논문 Full paper 제출				■	■								
3. 오피니언 마이닝						■	■	■					
4. 예측 분석									■	■			
결론 및 시사점 도출											■	■	
향후 계획 수립													■