# 2018년도 연구계획

2017.11.28. 김도연

# 2018년도 연구주제

연도	2017년
연구주제	텍스트 마이닝을 활용한 KEI 연구동향 분석
분석 분야	환경 전 분야
분석 매체	1. 뉴스 • NAVER 뉴스 2. 전문사이트 • KEI 연구보고서
분석 방법	텍스트 마이닝

2018년				
소셜 오피니언 마이닝을 활용한 일반국민의 기후변화 인식 분석				
기후변화				
1. 뉴스 • NAVER 뉴스 2. 전문사이트 • KEI 연구보고서, 학회지, 환경영향평가보고서 3. SNS • Facebook, Twitter 4. 댓글 • 뉴스 댓글, 블로그 댓글				
텍스트 마이닝 + 소셜 오피니언 마이닝 + 예측분석				

## 2018년도 연구관리

• 연구(가제): 소셜 오피니언 마이닝(Social Opinion Mining)을 활용한 일반국민의 기후변화 인식 분석

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

1. 데이터 수집

2. 텍스트 마이닝

3. 오피니언 마이닝

4. 예측분석

### '기후변화' 사전 구축

#### Word2Vec 분석

• 환경 DB에서 '기후변화'와 근접한 순으로 단어 출력



### 데이터 수집

### 1. 뉴스

- NAVER 뉴스
- 2. 전문사이트
- KEI 연구보고서, 학회지, 환경 영향평가보고서

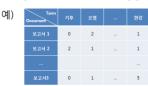
#### **3. SNS**

- Facebook, Twitter
- 4. 댓글
- 뉴스 댓글, 블로그 댓글

### 전처리

#### 텍스트 데이터 전처리 과정

- 1. 형태소분석기 수행
  - R에서 제공하는 형태소분석기 패키지 (KoNLP)를 사용하여 명사 추출
- 2. 특수문자, 특정단어 등 불용어 삭제
- 3. 단어길이 한글자 삭제
- 4. 출현빈도가 매우 낮은 단어(Sparse Terms) 삭제
- 5. Low TF-IDF 단어 삭제
- 6. DTM(Document Term Matrix)형태로 변형



1. 데이터 수집

2. 텍스트 마이닝

3. 오피니언 마이닝

4. 예측분석

### 매체별 키워드 분석

### 1. 형태소 분석

• 명사, 형용사 빈도 분석

#### 2. Word2Vec

- Skip-gram
- CBOW

### 매체별 이슈 키워드 비교

#### 키워드 시계열 비교 분석

• 연도별, 월별 키워드 빈도 빈도 분석

3. 오피니언 마이닝 1. 데이터 수집 2. 텍스트 마이닝 4. 예측분석 텍스트 마이닝 Fact 텍스트로 구성된 데이터를 자연 분석하여 주제를 판단하거나 유용한 정보를 추출, 가공 텍스트 기반의 데이터 오피니언 마이닝 텍스트에 나타난 감성, 느낌, 저자의 태도 등을 판별하여 의미 있는 정보로 변환

### 1. 데이터 수집

### 2. 텍스트 마이닝

### 3. 오피니언 마이닝

### 4. 예측분석

### Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll

Published: 2015.08.20 https://doi.org/10.1371/journal.pone.0136092 Authors: CODY, Emily M., et a

- 데이터: '기후' 포함된 트윗(2008.9-2014.7)
- 내용: 해도노 미터(Hedonometer)를 이용하여 '기후'포함된 트윗의 일별, 주별, 월별 행복도 측정
- 결과 : 트위터가 기후 변화 인식 확산의 중요한 요소임을 발견

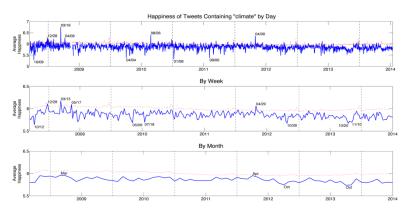


Fig2. Average happiness of tweets containing the word "climate" from September 2008 to July 2014 by day (top), by week (middle), and by month (bottom).

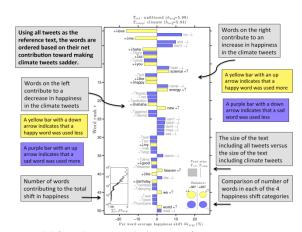


Fig 3. A word shift graph comparing the happiness of tweets containing the word "climate" to unfiltered tweets.

1. 데이터 수집

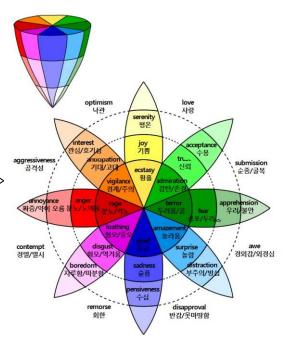
2. 텍스트 마이닝

3. 오피니언 마이닝

4. 예측분석

### 감성 분석(Sentimental Analysis)

- 1. 감성 사전
  - 해도노 미터(Hedonometer): http://hedonometer.org/words.html
  - SentiWordNet(SWN): <a href="http://sentiwordnet.isti.cnr.it/">http://sentiwordnet.isti.cnr.it/</a>
- 2. 로버트 푸루칙의 감정 수레바퀴 (Robert Plutchik's Wheel of Emotions) -> 64가지 감정 중 분석에 적합한 몇가지 감정 선택
- 3. Tensorflow를 활용한 '기후변화 한국어 감성분석기' 구현
- 4. 전문사이트, 전통 미디어에서 추출한 키워드 시계열 분석결과와 비교분석



1. 데이터 수집

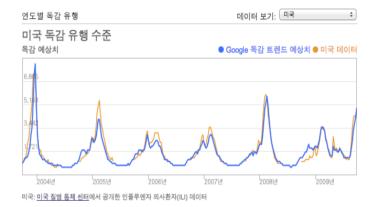
2. 텍스트 마이닝

3. 오피니언 마이닝

4. 예측 분석

#### 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 참궐을 탐지해냈습니다.



### 기후변화 세부현상 예측 분석

#### 예) 폭염 예측

- 폭염 주의보: 일 최고기온이 33도 이상인 상태가 2일 이상 지속될 것으로 예상될 때
- 폭염 경보: 일 최고기온이 35도 이상인 상태가 2일 이상 지속될 것으로 예상될 때



• SNS에 언급된 폭염 관련 버즈량을 통해 폭염 예측

## 2017년도 연구성과 논문작업

ICIC(International Conference on Industrial Convergence) in 2018

Site: <a href="http://icic.dsim.or.kr">http://icic.dsim.or.kr</a>

	Important dates
Full Paper Submission	December 8, 2017
Acceptance Notification	December 11, 2017
Registration	December 13, 2017
Conference Date	January 17~20, 2018