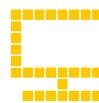
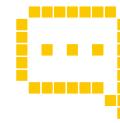




# 뉴스 빅데이터 해커톤 수상작 모음집

2019



한국언론진흥재단

04520 서울특별시 중구 세종대로 124 프레스센터 13층

Tel 02. 2001. 7574 Fax 02. 2001. 7560

[www.kpf.or.kr](http://www.kpf.or.kr)

뉴스 2019  
빅데이터  
해커톤  
수상작  
모음집



## 2019 뉴스 빅데이터 해커톤

한국언론진흥재단은 뉴스빅데이터를 활용한 새로운 사업 모델 및 공공서비스 발굴을 위해 '2019 뉴스빅데이터 해커톤'을 개최했습니다.

3~5월 동안 진행된 대회에는 95개팀, 340명이 참가해 뉴스데이터에 대한 높은 관심을 보여주셨습니다.

이 책은 수상작을 모은 것으로, 한국언론진흥재단 홈페이지 [www.kpf.or.kr](http://www.kpf.or.kr)와  
빅카인즈 [www.bigkinds.or.kr](http://www.bigkinds.or.kr)에서 무료로 보실 수 있습니다.

2019 뉴스빅데이터 해커톤에 참가해주신 모든 분들께  
진심으로 감사드립니다.

# CONTENTS

2019  
뉴스 빅데이터 해커톤  
수상작 모음집

Chapter  
01

2019 뉴스빅데이터 해커톤 소개

7

Chapter  
02

수상작 소개

11

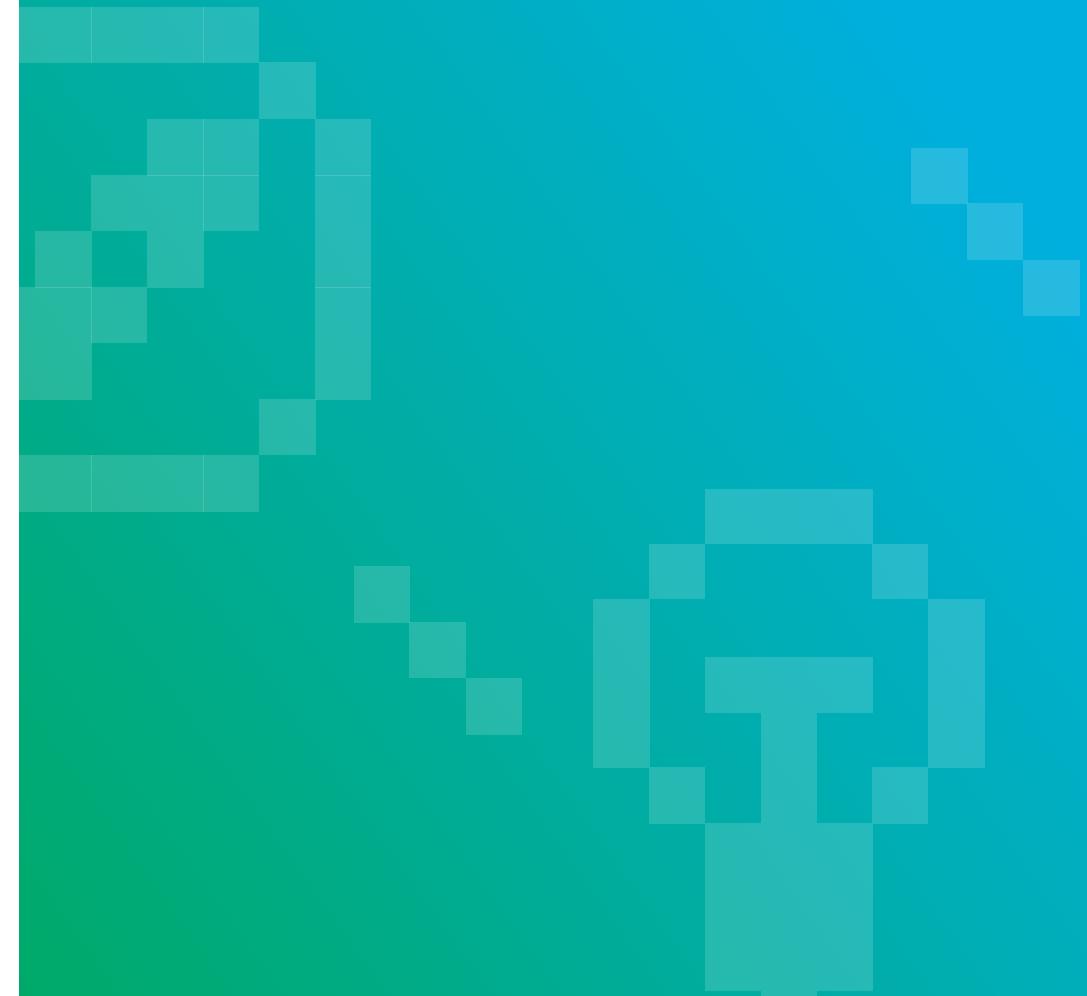
여행자 위치기반 실시간 뉴스서비스 모먼트 Moment	13
의료정보 큐레이팅 서비스	17
중소기업 취업파트너, 뉴스데이터로 중소기업을 보여주다	22
이글스 Easy하게 글 써지는 뉴스	26
브랜드 뉴스 Brand-news	30
뉴스데이터 기반의 건강정보 플랫폼 서비스 '건빵'	33
뉴스빅데이터를 활용한 기사 영향력 다종지표 서비스 플랫폼	38
사용자 참여 뉴스 검증 서비스 'JINJJA'	42
뉴스플로우 News flow	47
기업뉴스 SPY Speed Perfect You	51

2019

# 뉴스빅데이터 해커톤 수상작 모음집

2019  
뉴스빅데이터  
해커톤 수상작  
모음집

Chapter  
**01**



# 01 대회 목적

- 뉴스데이터를 활용한 IT 서비스 아이디어를 발굴함으로써, 우수 스타트업을 지원하고 국민 편의 제고에 기여함

# 02 대회 주제

- 뉴스데이터를 활용하여 사업화할 수 있는 유료 서비스 개발
  - 뉴스데이터를 활용하여 사회문제를 예방·해결할 수 있는 공공서비스 개발



03

## 활용 데이터

- 빅카인즈 [www.bigkinds.or.kr](http://www.bigkinds.or.kr)에서 제공하는 모든 데이터
- 공공데이터 포털 등에서 공개된 오픈 데이터
- 기타 활용하는 데 저작권 등의 문제가 없는 데이터

04

## 빅카인즈

[www.bigkinds.or.kr](http://www.bigkinds.or.kr)

- 종합일간지, 경제지, 지역일간지, 방송사 등을 포함한 6,000만 건의 뉴스 DB에 빅데이터 분석 기술을 접목해 만든 뉴스 분석 서비스
- 관계도 분석, 키워드 트렌드, 연관어 분석 등 뉴스빅데이터를 활용한 다양한 시각화 분석 제공



10

2019 뉴스빅데이터 해커톤 소개

2019  
뉴스빅데이터  
해커톤 수상작  
모음집

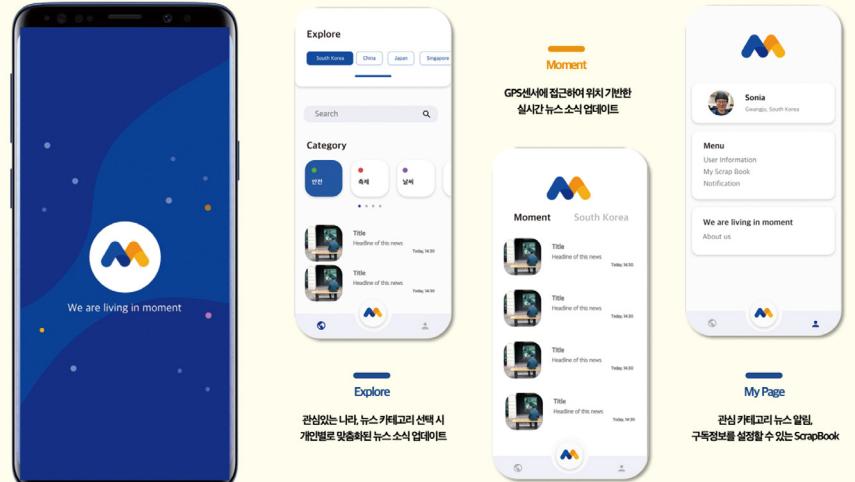
Chapter

02

# 수상작 소개



# 여행자 위치기반 실시간 뉴스서비스 모먼트 Moment



해외 여행 시 예상치 못한  
안전문제를 해결하기 위해  
실시간 위치기반 뉴스데이터를  
활용하여 여행객에게  
안전 정보를 제공

**팀명**  
모먼트(Moment)

**연락처**  
seoinha4964@gmail.com

**구성원**  
**서인하** 전남대 사회학과 4학년  
**김기표** 전남대 산업공학과 4학년  
**배성희** 전남대 컴퓨터정보통신공학과  
4학년

01

## 발상 동기

여행 중 예상하지 못한 사건과 마주친다면 어떻게 대처해야 할까? 국내라면 포털사이트 검색을 통해 무슨 일인지 파악할 수 있겠지만, 국외, 특히 비영어권 국가를 여행 중이라면 어떻게 정보를 얻을지 막막할 것이다.

모먼트 Moment 서비스는 팀원 중 한 명이 터키 여행 중 쿠데타 시위대를 만났던 경험에서 출발했다. 시위대로 인해 도로교통이 마비됐지만 언제쯤 정상화될지, 우회로는 있는지, 정보를 알 수 없어 난감했던 것이다.

해마다 여행자는 늘어나고 있지만 안전에 대한 정보는 미비하다. 대부분 여행자는 자주 사용하는 포털사이트를 통해 정보를 찾는다. 하지만 정확한 키워드를 모른다면 정보에 대한 접근 자체가 어렵다.

모먼트는 여행자의 이동경로에 따라 실시간 위치 기반 현지 뉴스를 제공하고, 관심사에 따라 카테고리별 뉴스를 구독할 수 있도록 기획되었다.

02

## 아이디어 내용

### 지역별 실시간 뉴스 제공

이용자의 GPS 센서에 접근하여 위치 기반의 실시간 뉴스를 업데이트한다.

### 구독 카테고리별 뉴스 제공

이용자가 구독한 관심국가, 카테고리별로 뉴스데이터를 제공한다. 카테고리는 여행자에게 필요한 안전, 행사, 맛집, 날씨 등으로 구성한다.

### 오지 여행자를 위한 알림 서비스

오지 여행 시에는 와이파이를 오랫동안 사용하지 못한다. 여행자에게 와이파이 신호가 작아지면 알림을 띄우고, 확인 버튼을 누르게 되면 보호자에게 와이파이 존을 벗어났음을 알린다. 오랫동안 연락이 가지 않아 걱정할 보호자를 안심시킬 수 있다.

14

03

## 서비스 구현

### 뉴스 메모·공유 기능

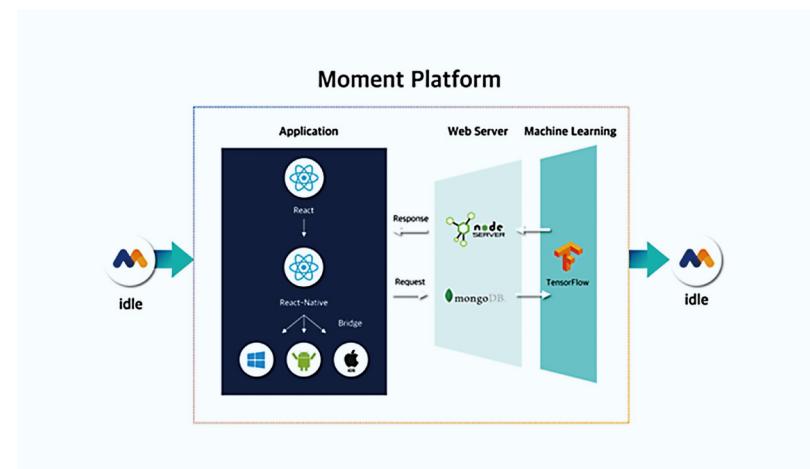
뉴스를 보다가 떠오른 아이디어나 메모를 기록하고 스크랩하는 기능이다. 마이페이지에서 관리, 공유할 수 있다.

### 활용 데이터

- 빅데이터
- 구글 뉴스 API
- 전 세계 뉴스 사이트 크롤링

### 내부 기술 구성도

- 전세계 뉴스 사이트에서 beautiful soup, selenium 이용해서 크롤링을 진행
- 자연어처리로 데이터 전 처리 후, 품사를 선별하여 머신러닝 training data
- 이 데이터들을 Text CNN 알고리즘을 사용하여 모델 생성
- 뉴스 API들을 모델을 통해서 지역별 카테고리별로 분류 진행
- Google translation API를 이용하여 번역 후 사용자에게 제공



15

## 04 기대 효과

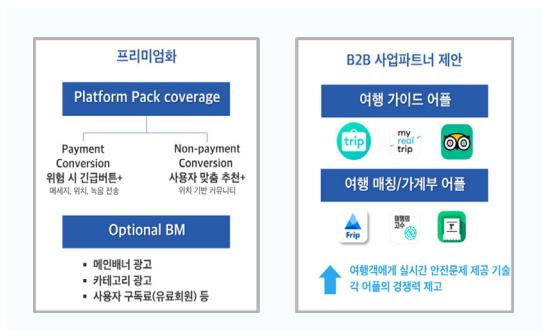
### 사용자 기술 구성도

- React기반 Library View를 생성
- React Native를 활용하여 iOS, Android, Windows 다양한 플랫폼 출시
- Node.js server 사용하여 MongoDB NoSQL에 저장
- 각 data들은 tensorflow를 사용한 MachineLearning 실행
- Category된 뉴스를 사용자에게 제공

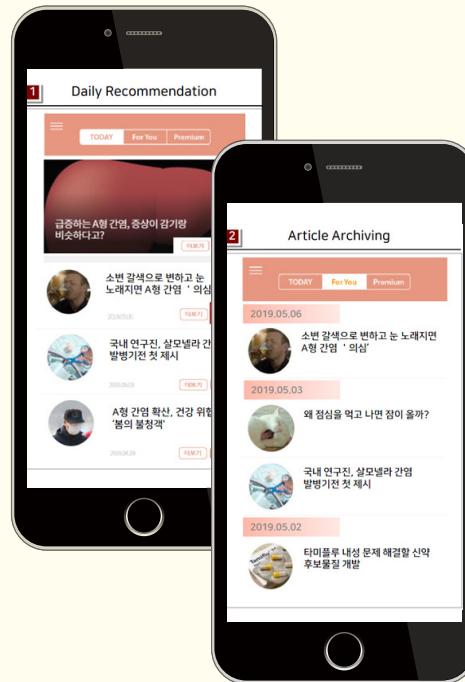
기존 뉴스 서비스는 일반적인 뉴스 이용자를 위한 것으로, 여행자를 위해 특화되지 않았다. 우리나라의 출국자 비율은 매년 증가 추세이며, '이동'은 라이프 스타일에 중요한 요소가 되고 있다.

모먼트는 여행객의 현지 정보 접근성을 높여 자유롭게 그 나라를 여행할 수 있는 진정한 의미의 로컬라이제이션 Localization을 꿈꾼다. 여행자에 특화된 뉴스 서비스를 통해 여행지에서 일어날 수 있는 안전 사고를 예방하고, 여행지에서 즐길 수 있는 다양한 정보를 제공할 수 있을 것이다.

더불어 B2B 사업 파트너 제안도 가능하다. 현재 플레이스토어 상위에 있는 여행 가이드 어플은 실시간 위치 기반이기는 하나, 여행자의 안전 문제에 대한 솔루션은 제공하고 있지 않다. 우리는 이를 여행플랫폼 시장의 블루오션이라 판단하며, 실시간 안전 정보 제공을 통해 각 어플의 경쟁력을 제고 할 수 있을 것이라고 생각한다.



## 의료정보 큐레이팅 서비스



텍스트 마이닝 기법을 활용한  
의료뉴스추천 시스템으로  
환자의 니즈에 따른  
맞춤형 의료정보 제공

### 팀명

MBT(Medical Bigdata Team)

### 연락처

010-2782-4626

### 구성원

서석윤 연세대 디지털애널리틱스  
융합협동과정 석사 1학년

김호현 연세대 디지털애널리틱스  
융합협동과정 석사 1학년

김미소 연세대 디지털애널리틱스  
융합협동과정 석사 1학년

권지혜 연세대 디지털애널리틱스  
융합협동과정 석사 1학년

## 01 발상 동기

국민 건강을 위해서는 질병의 조기 인지와 예방, 관리가 중요하다. 하지만 현 의료 시스템에서는 환자가 병원을 통해 의료 정보를 얻기 힘들다. 우리나라의 인구 1,000명당 의사 수는 2.3명으로 OECD 국가 중 최하 수준이고, 대학병원의 진료시간은 3분 내외로 매우 짧아 환자가 질병을 관리하는 데 필요한 정보를 충분히 얻기 어렵다.

의료 정보를 찾는 확실한 방법은 전문 의학서적을 보는 것이다. 그러나 용어의 전문성 등으로 환자나 보호자가 원하는 정보를 얻기는 쉽지 않다. 반면 접근성이 높은 유튜브, 블로그, 카페 등에 나오는 의료 정보는 신뢰성이 떨어지고, 정보의 깊이가 얕으며 범주 또한 제한적이다.

이러한 상황에서 의료 관련 뉴스콘텐츠는 좋은 대안이 될 수 있다. 의료뉴스는 의학 전문기자나 교수 등의 전문가가 대중의 눈높이에 맞게 전달하므로 신뢰성과 접근성이 높다. 의료뉴스는 다양한 미디어에서 생산하고 있지만, 뉴스데이터를 한 곳에 모아 주제별로 분류·정제하고 환자에게 맞춤형 의료정보를 큐레이션 하는 서비스는 아직 없다.

## 02 아이디어 내용

### 핵심기능

#### • 기사필터링

의료뉴스 중에는 환자의 관점에서 유의미하지 않은 것들이 있다. 병원, 의사, 의료산업 종사자 및 기업 주주들을 위한 뉴스가 그 예다. 이러한 뉴스를 일차적으로 토픽 모델링을 통해 걸러낸다. 이를 통해 환자가 의료 정보를 얻는 데 소요되는 탐색 비용을 줄일 수 있다.

#### • 큐레이션

두 가지 방향으로 큐레이션 서비스를 제공한다. 첫째, 환자가 질병명을 입력할 경우 관련된 맞춤 뉴스를 제공한다. 둘째, 환자가 본인의 증상이 정확히 어떤 질환에 해당하는지 모를 경우, 증상을 검색하면 관련 질환과 뉴스 목록을 제공한다.

## • 요약서비스

뉴스를 검색하면 많은 양의 뉴스가 나오지만, 그 중 무엇이 나에게 필요한 뉴스인지 제목으로는 파악되지 않는 경우가 많다. 이런 과정에서 오는 탐색 비용을 줄이기 위하여, 뉴스 요약서비스를 제공한다.

### タ깃이용자

- 의학 정보에 대한 니즈가 강한 일반 대중
- 의료정보를 통해 질병의 관리 및 개선, 합병증의 예방을 원하는 환자 및 보호자
- 의학 정보 소외 계층 60대 이상 인구 등

## 03 서비스 구현

토픽 모델링을 이용한 의학 기사 필터: 빅카인즈를 통해 수집한 전체 질병 관련 의료기사에 대해서 토픽 모델링을 진행해, 각 기사에 높은 확률의 토픽을 태그 Annotation 한다. 또한 생활습관, 식이 요법 등 환자가 필요로 하는 주제를 사전 형태로 만들어, 도출된 주제 중 뉴스의 토픽이 사전 내에 있는 경우만을 필터한다.

### 활용 데이터

빅카인즈를 통해 뉴스를 자체 DB에 저장 및 관리

### 내용 기반 추천 시스템

의료뉴스를 학습시킨 Word2Vec을 문서를 기준으로 임베딩 시킨 후, 뉴스의 임베딩 값과 다른 뉴스의 임베딩 값의 코사인 유사도를 계산한다. 이후 문서-문서의 매트릭스를 만들어, 가장 높은 유사도를 가진 기사를 추천한다.

### 인공신경망(RNN)을 이용한 문서요약

기사 요약은 Chang 등이 2015년에 AP뉴스 100만 건을 훈련 시킨 LSTM을 이용하고, 학습된 모델을 통해 기사 요약을 수행한다. 요약된 기사는 3줄 내외로 간략히 제공되며, 전문을 읽을 수 있는 URL도 함께 제공한다.

**B2C 모델**

큐레이션 서비스에서 질환과 연관성 높은 광고를 노출하는 방법이다. 예를 들어 만성질환 및 4대 중증질환 환자가 광고 타깃이 될 수 있다.

**B2B 모델**

병원은 환자를 대상으로 대중의 눈높이에 맞는 의료정보를 제공하기 위해 노력하고 있다. 예방, 관리에 대한 의료정보는 병원에서 경쟁사와의 차별화를 위해 니즈가 큰 부분이다. 여기에 초점을 맞추어 병원의 요구에 맞춘 B2B 서비스가 가능하다.

**B2G 모델**

각종 부처에서 운영하는 취약계층 진료비 지원사업, 질병관리본부에서 시행하는 특정 질환의 예방·관리를 위한 사업 등과 연계할 수 있다. 의료정보 취약계층을 위한 홍보자료를 주기적으로 공급하여 보건소에 비치하는 서비스도 가능하다.

**질병에 대한 인지 제고**

질병에 대한 정보 부족으로 증상을 정확히 인지하지 못하는 경우가 많다. 예를 들어 손 떨림 증상을 겪는 대부분의 사람들은 정형외과를 찾아 물리치료만 받는 경우가 많은데, 해당 증상은 목디스크, 파킨슨병의 초기 증상인 경우가 많아 빠른 진단 및 인지가 매우 중요하다. 본 서비스는 특정 증상에 대한 심각성 및 질병 유무 인지에 도움을 줌으로써 질병과 관련된 탐색 비용을 줄여준다.

**질병의 관리**

증상의 개선 및 치료를 위해서는 식단, 생활 습관, 운동 등 다양한 정보가 종합적으로 제공돼야 한다. 본 서비스는 질환별 키워드 검색을 통해 해당 질병에 대한 다양한 관리 요법들을 제공함으로써 질병의 재발병률을 낮추고 증상 개선에 기여할 수 있다.

**라이프 로그 데이터와의 융합**

최근 웨어러블 디바이스가 보급되면서, 걸음 수, 수면시간, 식이, 활동량 등의 데이터가 축적되고 있다. 실시간으로 측정되는 데이터와 연계하여, 환자에게 필요한 질병 관리 정보를 푸시 알람 등의 서비스로 제공할 수 있다. 그 외 이용자 중심으로 수집되는 각종 데이터와 결합하여 의료정보 대중화에 기여할 수 있다.

B2C 모델		
모델 설명	B2C 모델	B2B 모델
수익모델	웹/앱을 통해 기사 구독	CRM 니즈가 있는 병원이 고객들에게 증상에 맞는 의료기사 제공
소비자 접점	연관 광고	월 구독(사용료)
Customer	웹/앱	온라인 책자, 팜플렛 자료 배포
Consumer	카카오플러스친구 등 병원의 SNS	공익 모델
	1/2차 병원	보건소 및 지자체
	병원 등록/방문 고객(환자)	보건복지부 산하 보건소, 지자체
	노인 등 질병취약계층	환자/보호자 및 일반 대중
	환자/보호자 및 일반 대중	환자/보호자 및 일반 대중

# 중소기업 취업파트너, 뉴스데이터로 중소기업을 보여주다



중소기업을 목표로 취업을  
준비하는 청년들에게  
우수한 중소기업에 대한  
정보를 뉴스데이터를  
활용하여 제공

## 팀명

뭣이 중순디

## 연락처

010-3082-6748

## 구성원

체원진 성균관대 문헌정보학과 2학년  
김무성 성균관대 문헌정보학과 4학년  
김혜원 상명대 컴퓨터과학과 4학년  
오하늬 상명대 컴퓨터과학과 4학년

## 01 발상 동기

2017년 4월 통계청의 경제활동 인구조사에 따르면, 대한민국 청년 실업률은 11.2%로 외환위기 때와 거의 맞먹고 있다. 이에 우리는 구직자의 다양한 기호를 고려하면서 정보접근성을 높이는 방안을 고민하다가, 유독 중소기업 관련 취업정보가 부족함을 파악했다.

취업정보 사이트들은 대기업 위주로 정보를 제공하고 있었다. 소개하는 중소기업의 수가 적을 뿐더러 대기업만큼 자세하지도 않았다. 구직자의 약 70%가 중소기업을 목표로 하거나 중소기업도 고려하는 데 비해 정보습득이 용이하지 않은 것이다. 준비하고자 하는 분야에 어떤 중소기업들이 있는지, 업계에서 해당 중소기업의 위치는 어떤지 파악할 수 없으니 어느 중소기업에 지원할지 판단하기 어려울 수밖에 없다.

이러한 문제의식을 안고, 중소기업에 취업하고자 하는 청년을 위해 우수한 중소기업에 대한 정보를 제공하는 차별화된 취업정보 제공 시스템을 만들기로 했다.

## 02 아이디어 내용

- 뉴스데이터 분석 자료와 중소기업 현황 정보 시스템이 제공하는 우수 중견기업 DB를 활용
- 뉴스데이터를 활용해 중소기업의 투자 현황과 규모, 시연하고 있는 산업, 언론에 보도된 기업 이미지, 다른 기업들과 준비하는 컨퍼런스, 투자받은 사람이나 단체, 주식현황, 해당 중소기업을 검색할 때 자주 등장하는 키워드를 제공
- 서비스의 효용성
  - 모든 정보를 일일이 찾아보지 않고도 한눈에 볼 수 있어 효율적임
  - 특정 기업에 대해 가졌던 막연한 이미지를 데이터를 활용하여 구체화
  - 중소기업에 대한 이해를 높이고, 지원 기업 선택 시 판단 자료가 됨

03

# 서비스 구현

- 정부로부터 검증받은 우수중소기업 데이터를 수집하기 위해 '중소기업현황정보시스템'을 크롤링하여 기업명, 업종명, 주소 등 필요한 정보를 추출한다. 이를 토대로, 중소기업에 대한 기사를 분석하기 위해 빅카인즈 API 뉴스 검색을 사용한다. 서비스에 필요한 데이터를 수집하고 DB에 저장한다.
  - 기업의 성장성, 안전성, 활동성 등 구직자에게 유용한 항목을 지표로 삼아 우수 중소기업에 대한 뉴스데이터를 분석한다. 빅카인즈 뉴스검색을 통해서 얻은 결과에서 '키워드 트렌드' 인터페이스를 이용하여 기업의 기간별 기사 건수 언급량의 척도로 고려로 기업의 활동성을 파악한다. 추가적으로 관계도 분석과 연관어 분석 API를 이용해서 해당 버즈 언급들이 어떤 측면에서 언급되었는지 분석한다. 긍정적인 측면의 단어 투자, 수주, 상장과 부정적인 측면의 단어 부채, 실형, 폐지의 맥락 정보를 파악한다. 또한, 관계도·연관어 분석을 통해서 해당 기업의 주력사업 정보와 언론에 비친 대외 이미지 또한 제공할 수 있다.

- 투자액, 사업규모 등의 추가 정보 데이터를 기사로 추출하기 위해서 빅카인즈 분석 API 내 정보 추출 기능을 우선 활용한다. API 외에 추가적 기능에 따라 저장했던 로데이터를 파일 선과 같은 프로그래밍 언어로 통제한다.
- 최종적으로 각 중소기업에 대한 정제된 뉴스데이터, 중소기업 현황정보시스템의 데이터를 통합하여 웹형태로 시각화된 서비스를 구현한다. 각 기업에 따른 기본적인 소개와 뉴스분석을 통해 투자현황, 활동성, 성장성, 주력산업, 기업의 키워드 등을 보여준다. 또, 기업의 채용공고, 가장 최근 언급된 기사의 헤드라인 등을 묶어서 제공한다.

## 04 기대 효과

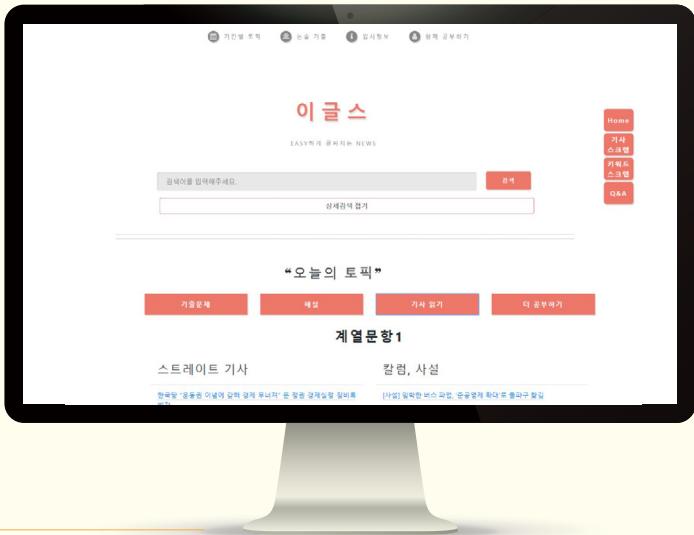
- 중소기업 취업을 준비하는 청년들에게 중소기업의 사업현황, 기술력, 성장가능성 등 다양한 측면에 대한 입체적인 정보를 제공할 수 있다.
  - 구직난에 시달리는 중소기업이 청년들을 대상으로 자사를 홍보하는 플랫폼으로 활용할 수 있다.
  - 중소기업 뉴스를 활용한 서비스를 확산함으로써, 중소기업 보도에 대한 언론의 관심을 제고할 수 있다.



# 이글스

Easy하게 글 써지는 뉴스

뉴스빅데이터와  
함께 하는  
논술 공부



팀명  
한화

연락처

ksh971002@naver.com

구성원

김상희 숙명여대 미디어학부 3학년

안선빈 숙명여대 컴퓨터과학부 3학년

현정민 숙명여대 컴퓨터과학부 4학년

01

## 발상 동기

2019년도 논술 입시에서 연세대학교는 2년 연속 교육과정에서 벗어난 시험 문제를 출제해 교육청으로부터 모집정지 처분을 받았다. 고등학교 정규 교과과정에서 가르치는 수준의 제시문을 출제하라고 요구하는 정부와 변별력 확보를 원하는 대학가가 수년 째 갈등을 빚은 와중 수면 위로 터진 사건이다. 이처럼 수험생이 대학에서 요구하는 수준에 맞추려면 사교육의 힘을 빌릴 수밖에 없는 것이 현 입시 상황이다.

우리는 기존 논술교육 시스템에서 아쉬운 점을 발견했다. 논술 입시학원의 주요전략은 답안 구성 스킬과 기출문제 분석 정도이다. 지문을 이해하고 글을 쓰는 능력은 결국 개인의 역량개발에서 나오는데, 풍부한 내용을 채우기 위한 배경지식을 갖추는 것은 일방향 강의로 해결하기 어렵다. 하지만 수험생에게는 별도의 상식을 갖출 만한 시간적 여유도 없다.

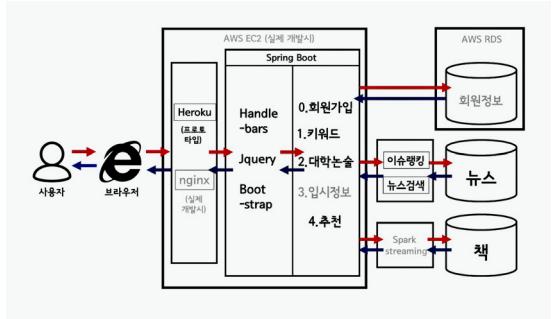
이 문제의 답은 뉴스에서 찾을 수 있다. 인문논술 준비에서 뉴스데이터는 필수불가결한 정보를 제공한다. 시의성 있는 소식과 역사적 지식, 분석적 시각을 갖춘 자료들로 배경지식을 쌓을 수 있다. 하지만 뉴스 포털에 들어가면 하루에도 수백 개씩 쏟아지는 기사들 중 꼭 필요한 내용만 선별하는 것은 수험생에게 어려운 일이다. 이러한 문제를 해결하기 위해 뉴스빅데이터 분석을 활용한 수험생 맞춤형 논술대비 웹사이트를 만들기로 했다.

02

## 아이디어 내용

- 첫째, 기간별 핵심 키워드 추천이다. 통상적으로 그 해의 이슈가 논술주제로 선정될 가능성이 높기 때문에, 학생들에게 시의성 높은 키워드를 뽑아주는 것이다. 가장 많이 기사화된 키워드들을 5개씩 추려서 보여준다. 방대한 뉴스자료를 직접 분류하고 선별할 여유가 없는 이용자는 시간을 아낄 수 있다.
- 둘째, 추천된 키워드와 관련된 뉴스를 자율학습에 용이하도록 배치한다. 학생들이 이슈를 한눈에 확인한 후 선택적으로 이슈를 클릭하면 그 이슈와 관련된 헤드라인이 나열된다. 그 중 원하는 뉴스를 읽고, 배경지식을 쌓는다.

## 03 서비스 구현



### 키워드 추천

이용자가 원하는 기간을 전송하면 빅카인즈 이슈랭킹 API를 통하여 해당 기간의 토픽을, 뉴스검색 API를 사용하여 관련 뉴스를 가져온다. 로그인 시 오늘의 이슈랭킹 키워드들과 관련된 뉴스를 메인 화면에 보여준다.

### 뉴스 출력

뉴스검색 API를 사용하여 가져온 뉴스들을 제목이나 작성자가 '칼럼', '사설' 등의 단어가 포함되는 경우 분석 뉴스로, 그렇지 않은 경우 단신 뉴스로 분류한다. 왼쪽에는 단신 뉴스, 오른쪽에는 분석 뉴스로 나눠서 화면에 출력한다.

### 논술 기출문제

각 대학 입학처에 올라온 기출문제 답안의 핵심 키워드를 추출하여 내부 데이터를 만든다. 이를 기반으로 뉴스검색 API를 통해 해당 키워드와 관련된 뉴스를 가져온다.

### 입시 정보

연도별·대학별 입시 정보를 뉴스검색 API로 실시간으로 가져온다. 입시 정보는 교육청, 대학별 보도자료, 학원에서 제공하는 정보로 나눠서 보여준다.

### 수험생 맞춤형 학습자료

원하는 키워드를 검색하면 기출문제, 키워드 관련 책, 키워드 관련 인강 추천이 출력된다. 추천 알고리즘은 스파크 스트리밍을 이용 한다.

### 질의 응답 서비스

외부 업체와의 제휴로 튜터를 협력 받는다. 튜터에게 전용 계정을 제공하고, 제공받은 계정으로 로그인하면 학생들에게 받은 질문들을 볼 수 있다.

## 04 기대 효과

- 에듀테크 시장의 성장세에 발맞춰 온라인 교육업체의 강점을 극대화
- 기존 학원과의 제휴로 고정회원을 확보하고, 도서추천 등 타 데이터와의 결합을 통해 부가수익 창출



## 현명한 소비를 도와주는 뉴스 검색 솔루션

### 팀명 뉴서스(Newser's)

연락처  
kso4013@gmail.com

### 구성원

김시온 송실대 컴퓨터학부 4학년  
박진수 줌인터넷  
박윤성 송실대 컴퓨터학부 4학년  
서민 건국대 응용통계학과 3학년  
조찬욱 송실대 컴퓨터학부 4학년

## 01 발상 동기

착한 기업, 윤리적 기업 제품을 선호하는 경향이 커지는 추세다. 기업의 사회공헌 활동이 기업 이미지와 선호도, 신뢰도뿐 아니라 소비자 구매에까지 영향을 미치고 있는 것이다.

하지만 소비자가 기업에 대한 정보를 알지 못하는 경우가 많다. 예를 들어, 일본 대기업 '미쓰비시'는 전범 기업으로, 전투기, 군함 등 일제강점기 군수 물품을 조달하여 팽창했다. 이 외에도 '캐논', '아사히' 등의 전범 기업이 있지만, 소비자는 이를 알지 못한 채 제품을 구매하는 경우가 많다.

소비자가 기업의 윤리성에 대한 정보에 쉽게 접근할 수 있고, 이러한 정보를 바탕으로 구매 여부를 판단할 수 있도록 지원하는 서비스가 필요하다.

## 02 아이디어 내용

- 쇼핑몰 상품 구매 전, '착한 소비'를 원하는 소비자들에게 상품을 만든 기업에 대한 정보를 제공하는 서비스를 만든다.
- 검색 시, 선택된 기업의 뉴스데이터를 분석하여, 기업의 선행과 여론의 질타를 받았던 주요 뉴스 및 키워드 등의 분석자료를 제공한다.
- 워드 클라우드, 뉴스 목록 등의 기능이 있는 뉴스 검색 샘플 앱 및 코드를 제공한다.

### 주요タ겟

- 해마다 성장하는 소셜·E커머스 유통기업: 치열한 경쟁 속에서 차별화된 전략을 필요로 함. 현재 78조 원인 시장규모는 올해 100조 원으로 커질 것이라는 전망
- 기업에 대한 정보를 제공하는 아르바이트, 취업 및 기업 정보 제공업체

## 서비스 구현

### 데이터 활용

빅데이터 API를 1시간 간격으로 주기적으로 호출하여 정제 과정을 거친다. 한 눈에 볼 수 있는 이슈 차트, 브랜드 연관 키워드, 상세 뉴스 리스트 등의 정보를 제공한다.

### 맞춤형 기능

브랜드마다 원하는 대로 직접 커스터마이징 할 수 있는 검색 API를 제공하고, 정제된 자료를 시각화한 '통합 대시 보드' 기능을 통해 브랜드에 대한 정보를 한눈에 확인할 수 있게 한다.

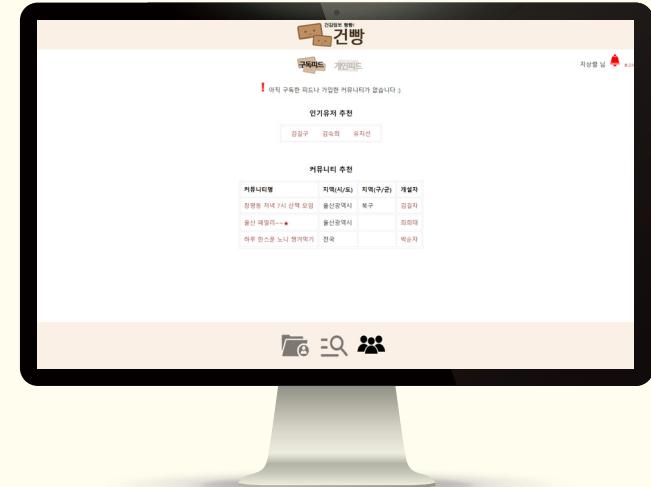


통합 대시 보드 프로토타입

## 기대 효과

- 내가 소비하는 기업에 대해 정확히 하는 스마트 컨슈머 확대에 기여한다.
- 기업의 윤리경영 실천을 유도할 수 있다.
- 온라인 쇼핑몰에서 서비스 도입 시, 소비자의 브랜드 정보에 대한 니즈를 충족시킴으로써, 타 쇼핑몰과 차별화할 수 있다.

## 뉴스데이터 기반의 건강정보 플랫폼 서비스 '건빵'



건강 관련 뉴스데이터를  
기반으로 이용자 맞춤형  
건강정보를 제공하고  
다양한 헬스케어 기업과의  
제휴를 통해 광고 및  
판매 중개를 하는 플랫폼

**팀명**  
더조은  
**연락처**  
skdudenzl11@naver.com  
**구성원**  
최나영  
**최나영**  
이수정 인하대 경영학과 4학년

### 1. '건강정보'에 대한 중·장년층의 높은 관심

중·장년층을 타깃으로 한 TV 건강 프로그램은 방송국의 오랜 협력 프로그램이고, 미디어에 소개된 건강식품은 순식간에 포털사이트 검색 순위에 오른다. 이처럼 중·장년층에게 '건강'은 주된 관심사이다.

### 2. 그러나 신뢰할 수 있고 필요한 정보를 얻기는 힘든 현 상황

넘쳐나는 정보 속에는 가짜 뉴스나 과대 광고성 글이 너무 많고, 그 속에서 필요한 정보를 찾기 위해서는 시간을 들여 검색하는 수고를 감수해야 한다.

### 3. 해결책은 '뉴스데이터 기반의 맞춤형 건강 정보 플랫폼'

뉴스가 유료화될 경우, 가장 구매의사가 높은 분야는 '건강·의료·복지'라고 한다. 뉴스의 장점은 신뢰할 수 있으면서 읽기 쉽다는 것이다. 건강 정보에 대한 수요와 뉴스의 장점을 볼 때, 뉴스데이터를 기반으로 개인 사용자 맞춤형 건강 정보를 제공하고, 사용자끼리 건강에 관한 정보를 나누는 플랫폼은 현 상황의 해결책이 될 수 있다.

건빵 플랫폼의 참여주체는 크게 서비스 이용자와 헬스케어 분야 기업 [건강식품](#), [건강용품](#), [병원](#), [보험사](#) 등으로 나뉜다. 건빵은 이용자에게 건강 관련 뉴스데이터를 기반으로 실생활에 유용한 맞춤형 건강 정보를 제공하고, 건강을 중심으로 교류하는 대화의장을 열어준다.

또한, 다양한 헬스케어 분야 기업들과 제휴를 맺어 광고를 진행하고, 이용자의 건강정보나 소비습관 분석을 바탕으로 제휴 기업의 건강상품을 추천한다. 이용자는 건빵을 통해 일상에 맞닿아 있으면서 믿을 수 있는 건강 정보를 얻을 것이고, 헬스케어 기업은 효과적인 광고 채널을 통해 매출을 올릴 수 있을 것이다.

### 1. 개인 맞춤형 건강 정보 추천

이용자가 입력한 프로필 [나이](#), [거주 지역](#), [질병](#), [관심사](#) 등과 플랫폼 내에서의 실제 건강정보 이용 습관을 바탕으로, 이용자의 세부적인 관심사에 맞는 건강정보를 날마다 추천하는 기능이 핵심이다. 이용자는 원하는 정보를 일일이 검색할 필요 없이 한 화면에서 다양한 카테고리의 건강 정보를 모아볼 수 있다.

- **정책정보** 의료·건강에 관한 정부 정책 신설 및 변경사항

예) 진료비 감액 혜택 연령 기준 65세-> 70세 상향 조정

- **지역정보** 이용자 거주 지역 내 건강 관련 소식

예) 충주시 '치매진료소' 개관

- **건강정보** 건강 상식, 건강식품 정보, 건강에 관한 최신 연구 결과 등

예) 저혈당리라고 먹나니…… 'G지수' 모르세요?

모임가 vs 노니, 임청난 효능에도 부작용 조심해야 하는 이유

운동(동영상) : 건강 정보TV 프로그램의 방송 영상 클립

### 2. 이용자 간 소셜 네트워킹 기능

모든 이용자는 플랫폼 내에서 '개인 페이지'를 갖는다. 이용자는 이곳에 관심 정보를 스크랩해 두거나, 글을 올릴 수 있다. 글쓰기 양식에는 건강 제품 정보를 간편하게 불러오는 기능이 있고, 상품후기 작성 및 공유가 가능하다.

또한 다른 이용자의 글은 '구독피드'에서 열람할 수 있다. 상품 후기를 작성한 이용자는 플랫폼 내 부가 서비스나 제휴 업체의 상품을 구매할 때 사용 가능한 포인트를 적립받는다. 또한 비슷한 관심사를 가진 이용자끼리 커뮤니티를 형성할 수 있어 교류를 통한 건강 증진을 돋는다.

### 3. 건강상품 추천시스템

이용자 프로필을 기반으로 헬스케어 제휴업체의 상품·서비스를 추천한다. 추천은 알람 형태로 띄우며, 알림 글을 누르면 플랫폼 내 구매 페이지로 연결되도록 한다. 기본적으로 업체가 제공하는 정보를 이용하되, 건강식품의 경우 '식품의약품안전처 식품안전정보포털'의 '건강 기능 식품 데이터'를 추가적으로 제공한다.

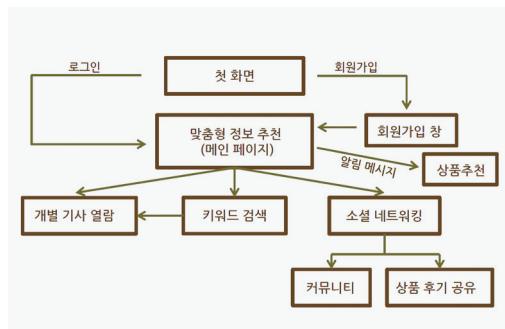
## 1. Rstudio를 활용한 데이터 전 처리 및 DB 적재

- 빅카인즈 뉴스검색 API로 ‘사회>의료\_건강’ 카테고리의 뉴스데이터 불러오기
- 데이터 전 처리 | 건강 정보와 관련 없는 뉴스데이터 제거, 날짜 형식 변경 등
- 건강 정보 카테고리 분류 | 정책정보, 지역정보, 건강정보
- 개별 기사의 지역을 분류 | 시/도, 구/군
- 빅카인즈 키워드 추출 API를 이용하여 건강정보 카테고리 데이터의 토픽 생성
- Oracle과 연동하여 완성된 데이터 프레임을 Oracle DBMS에 적재

## 2. Oracle + Eclipse를 활용하여 테이블 구성

건강뉴스, 유저 프로필, 유저 토픽, 유저 커뮤니티, 유저 글, 제휴상품, 식품안전정보 DB 등으로 구성한다.

## 3. Eclipse에서 jsp로 웹 구현



## 이용자 | 높은 접근성, 편리한 이용, 믿을 수 있는 정보

현재 건강정보를 얻는 방법에는 포털 검색이나 공공기관이 운영하는 앱 이용 등이 있다. 전자는 신뢰성 있는 정보를 찾기 힘들고, 후자는 나를 위한 맞춤형 정보를 찾기 어렵다. 건빵은 기존 서비스의 문제를 해결하여 ‘접근이 용이하고, 이용이 편리하며, 믿을 수 있는 차별화된 건강 정보 소비 경험을 가져다 줄 것이다.

## 제휴기업 | 효과적인 광고 채널 확득

건빵 플랫폼에서는 건강에 관심이 많은 소비자 집단을 대상으로 광고가 전달되므로, 광고 수용도가 높을 것으로 예상한다. 또한 ‘뉴스데이터 기반 정보 플랫폼’이라는 서비스 특성 덕분에 이용자들이 서비스 내에서 접하는 정보에 대한 신뢰도가 높고, 이는 플랫폼 내부의 광고에 대한 신뢰도에도 영향을 미칠 가능성이 높다.

## 정부기관 | 효과적인 정책 홍보 가능

정부기관은 의료·건강에 대한 대국민 정책 홍보가 필요할 때, 건빵 플랫폼을 통해 해당자에게 효과적으로 정책에 관한 내용을 전달 할 수 있을 것이다.

# 뉴스빅데이터를 활용한 기사 영향력 다중지표 서비스 플랫폼



뉴스의 영향력을  
시각화·수치화 정보로  
제공하여 뉴스 이용자의  
선택에 도움을 주는  
서비스

**팀명**  
퇴근후 강남역  
**연락처**  
010-7101-6540  
**구성원**  
조수현 와이즈넛  
민경현 서울대 건설환경공학부 석사과정 2학년  
고형권 한양대 수학과 4학년  
양세원 중앙대 역사학과 4학년  
이선미 국방부

## 01 발상 동기

우리는 뉴스를 통해 사건 수사의 진행 과정, 정부의 재난재해에 대한 대응 활동 등을 접하게 된다. 뉴스를 접한 후 댓글이나 SNS를 통해 자신의 생각을 표출하기도 한다. 뉴스의 방향에 따라 긍정적이었던 평가가 부정적으로 바뀌는 경우도 쉽게 볼 수 있다.

시간과 뉴스기사에 따라 급격하게 변하는 사람들의 반응을 보면서 뉴스기사가 미치는 파급력이 대단하다는 것을 느꼈다. 뉴스 기사 하나로 긍정적인 평가가 부정적으로 바뀔 수도 있고, 이슈가 되지 않았던 사안에 대해 사람들이 관심 갖고 평가하기도 한다.

이에 뉴스빅데이터와 SNS 데이터를 융합하여 뉴스기사의 영향력을 객관적으로 확인할 수 있는 다중지표를 제공하는 서비스 플랫폼을 착안하게 되었다.

## 02 아이디어 내용

이슈 혹은 키워드 [인물](#), [장소](#), [사건 등](#)에 대한 SNS 상의 긍정·부정 반응과 뉴스기사의 상관관계를 시각적 그래프와 수치화된 지표를 통해 나타낸다. 서비스 기능은 다음과 같다.

- 특정 키워드에 대한 뉴스 전후 기간별 긍정·부정 감성지수 비교 제공
- 감성지수 변화정도를 변화지수로 제공
- 특정 키워드에 대해 세부 이벤트별 클러스터링을 통해 군집화된 기사 제공
- 시계열적 반응 곡선 클릭 시 뉴스와 연계 정보 제공
- 뉴스에 대한 영향력 평가 지표 제공
- 긍정·부정 평가지수를 통해 해당 키워드에 대한 평균 반응 지표 제공
- 각 기사의 카운트 등 자체 개발할 지표를 통하여 관심도 지표 제공
- 기사의 벡터화 [article2vec](#)·시각화 정보 제공

다양한 필터 기능을 제공하여 사용자가 원하는 정보를 쉽게 볼 수 있도록 한다.

## 03 서비스 구현

데이터는 빅카인즈에서 제공하는 뉴스빅데이터와 SNS 대표 플랫폼 중 하나인 '트위터' 데이터를 활용한다. 빅카인즈는 1990년부터 주요 언론사를 포함한 54개 매체의 약 6천만 건의 뉴스빅데이터가 유형별로 정형화되어 있어 분석 및 활용에 용이하다. 140자 글자수 제한이 있는 트위터는 이슈가 발생했을 때 사용자의 반응이 즉각 올라오고, 리트윗 등의 기능으로 확산이 빨라 이슈의 화제성을 나타낼 수 있는 대표적인 SNS이기 때문에 반응 데이터로 적합하다고 판단했다.

키워드 몇 개를 선정하고, 해당 키워드를 포함한 트위터와 뉴스데이터를 분석한다. 시계열분석은 트위터 데이터의 'created\_at' 컬럼을 활용하며, 뉴스데이터의 '일자' 항목을 사용한다. 뉴스 및 트윗에서 키워드를 추출할 때는 제공되는 '키워드 추출 API'를 활용한다.



긍정·부정 시계열 분석 프로토타입

## 04 기대 효과

관심 있는 키워드에 대한 감성-시간 추이, 관련 기사, 다양한 지표들을 통해 다양한 서비스에서 마케팅, 서비스기획, 상품개발 등에 활용할 수 있다. 여론의 반응을 검색량 추이, 개별 뉴스 또는 한정된 SNS 및 댓글 내용으로만 파악하던 기존 서비스의 한계를 넘어 '왜 이런 여론이 생성되었으며, 왜 여론의 감정 반응이 변화했을까'에 초점을 맞추어 여론 동향의 원인을 자동으로 분석할 수 있다.

기업이나 공공기관에서는 정책이나 서비스에 대한 여론 반응 동향과 변화 추이를 보도된 뉴스와 연계하여 쉽게 파악하는 효과를 거둘 수 있다. 또한 제공되는 확산지수, 변화지수 등을 참고하여 계획수립에 활용할 수 있고, 재난발생 등 특이한 변동사항을 빠르게 인지하고 대처하여 피해를 최소화할 수 있다.

# 사용자 참여 뉴스 검증 서비스 'JINJJA'



독자 스스로 뉴스의  
사실 여부를 평가하는  
참여형 뉴스 검증 모델



## 팀명

On the Bridge

## 연락처

010-9129-4712

## 구성원

최태민 고려대 일반대학원 기계공학과 석사과정

김형석 글로벌네트웍스

이혜연 아주대 전자공학과 3학년

유준영 서울과학기술대 산업공학과 4학년

한정은

## 01 발상 동기

최근 미디어 환경의 급속한 변화와 함께 가짜뉴스가 큰 화두로 떠 오르고 있다. 가짜뉴스는 명예훼손이나 사생활 침해 등 인격권 침해나, 정치 이슈의 사실 여부를 판단하는 과정에서 사회적 혼란과 분열을 조장하며 우리 사회에 전반적으로 부정적 영향을 끼친다. 뉴스를 접하는 독자들이 자극적인 가짜뉴스에 현혹될수록 사실만을 전달하는 언론사는 독자들의 관심에서 벗어나게 되며 결과적으로 신뢰성을 잃게 되는 악순환으로 이어지게 된다.

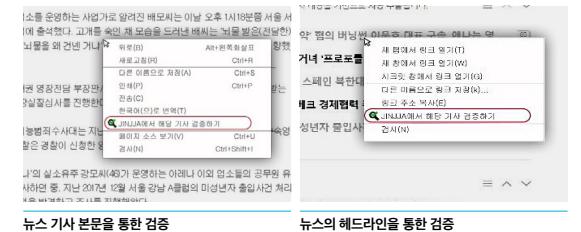
프랑스와 독일에서는 가짜뉴스의 생산과 유통을 제도와 법으로 제재하기 위해 정부와 국회에서 목소리를 내고 있다. 국내 또한 미디어 교육 지원법, 가짜정보 유통 방지에 관한 법률안 발의를 하는 등 가짜뉴스에 대한 대응 방안을 강구 중이다.

뉴스에 대한 선별된 규제는 표현의 자유를 침해하거나, 언론을 탄압하는 수단으로 악용될 수 있다. 국가의 법에 따라 규제된 뉴스를 일방적으로 제공하는 것보다 독자 개개인이 뉴스의 진실성 여부를 스스로 판단할 수 있는 여지를 주는 것이 바람직하다. 따라서 우리는 사용자에게 뉴스의 사실 정도를 확인할 수 있는 'JINJJA' 서비스를 제안한다.

## 02 아이디어 내용

### 1. 'JINJJA' 접근 방법

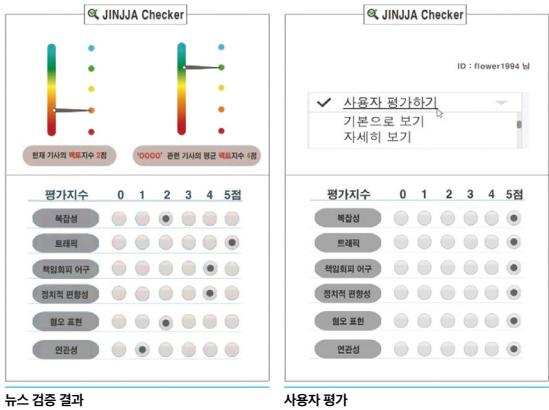
사용자가 웹에서 뉴스를 보다 사실 여부를 검증하고 싶을 때 다음 그림과 같이 크롬 확장프로그램 형식의 'JINJJA'를 사용할 수 있다.



## 03 서비스 구현

### 2. 'JINJJA' 검증 결과

사용자는 해당 뉴스의 팩트 지수를 확인할 수 있으며, '자세히 보기'를 클릭하면 뉴스의 험오 표현 정도, 정치적 편향성 등 세부 항목 별 분석 결과와 다른 사용자들의 평가 통계치를 확인할 수 있다.



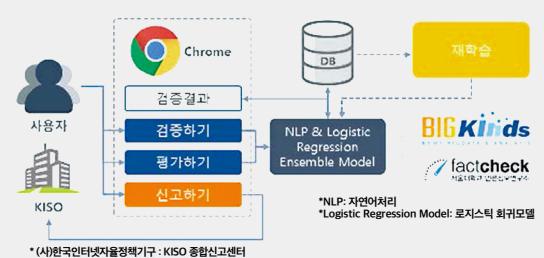
### 3. 'JINJJA' 결과에 대한 사용자 평가 및 모델 재학습

사용자는 검증 결과에 대해 평가를 할 수 있으며, 평가 참여를 독려하기 위해 캐시슬라이드의 포인트와 같은 보상을 제공할 계획이다. 또한 사용자의 평가를 바탕으로 검증 모델을 주기적으로 재학습시킨다.

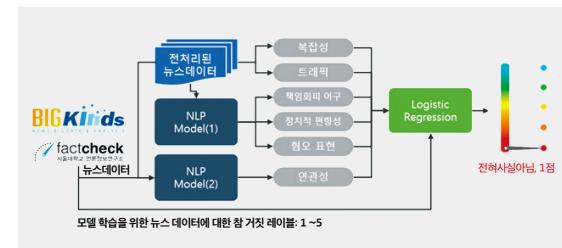
### 4. 예상 고객

예상 고객은 뉴스 구독자, 뉴스로 2차 뉴스를 제작하는 기자, 뉴스 데이터를 활용하는 업체 및 연구자, 언론중재위원회 심의위원 등이 있다.

### 1. 서비스 아키텍처



### 2. 양상별 모델 구조



- 빅카인즈와 SNU 팩트체크 데이터를 바탕으로 NLP 모델과 로지스틱 회귀모델을 학습
- NLP & Logistic Regression Ensemble Model을 통해 뉴스를 검증
- 사용자의 평가를 바탕으로 Ensemble Model을 재학습
  - 팩트 지수에 대한 평가 → Logistic Regression 재학습
  - 세부 항목별 결과에 대한 평가 → NLP Model 재학습

### 3. 독립변수

- 페이지 소스의 복잡성: 가짜뉴스일수록 페이지 소스가 복잡
- 기사의 책임회피 어구 사용 빈도: 가짜뉴스일수록 책임회피 어구가 자주 나타남

## 04 기대 효과

- 정치적 편향성: 가짜뉴스일수록 정치색이 짙음
- 혐오 표현 정도: 가짜뉴스일수록 특정 인물이나 조직을 겨냥하여 혐오 표현을 사용
- 페이지 트래픽 분석: 가짜뉴스일수록 시간에 대한 트래픽 분산이 큼
- 연관성 검사: DB [빅데이터&서울대 팩트체크](#) 와 해당 뉴스 기사와의 연관성을 분석

### 언론 신뢰 향상

가짜뉴스에 의해 언론의 신뢰도가 지속해서 하락하고 있다. ‘JINJJA’를 언론사 및 기타 업체에서 이용한다면, 불필요한 자원을 투입할 필요가 없이 빠른 사실 검증이 가능해질 것이다.

### 미디어 리터러시 능력 향상

사회가 급변하며 쏟아지는 정보 속에서 미디어를 비판적으로 수용하고 해석하는 능력은 선택이 아닌 필수가 되었다. ‘JINJJA’를 통해 뉴스 검증 결과의 근거들과 다른 사용자들의 평가 정보를 바탕으로 사용자의 미디어 리터러시 능력을 함양할 수 있다.

### 언론중재위원회 지원

가짜뉴스에 대한 명확한 판단 기준이 없어 피해자가 있더라도 언론중재위원회를 통해 구제받기 힘들다는 것이 현 상황이다. JINJJA는 팩트지수를 통해 가짜뉴스 여부를 판단하는 기준을 만들어 언론중재위원회와 관련 피해자들을 도와줄 수 있다.

## 뉴스플로우 News flow



사건의 발단부터 해결까지  
뉴스를 통해 알아보는  
타임라인형 뉴스 플랫폼

### 팀명

사연남(사건 연대기 만들어주는 남자들)

### 연락처

kcs00217@naver.com

### 구성원

김창섭 경희대 국제학과 4학년  
고재원 국민대 법학과 3학년  
노승곤 서울시립대 경영학과 4학년  
한동훈 경희대 국제학과 3학년

단편적인 뉴스는 사건의 핵심을 명확하게 전달해준다는 장점이 있다. 하지만 너무 핵심만을 전달다보니 복잡한 사건을 단편적으로 알려주게 되고 사안의 전체적인 흐름을 말해주지 못한다. 사람들은 하나의 사안을 알기 위해 관련 기사들을 일일이 모두 찾아봐야 한다. 사건의 전후 맥락 없는 보도는 사안에 대한 사람들의 흥미와 관심을 떨어뜨린다.

그러나 사람들은 사건의 진행 과정을 알고자 하는 욕구가 크다. 한국언론진흥재단에서 발간한 ‘한국 언론 신뢰도: 진단과 처방’에 따르면, 일반인이 원하는 기사 중 ‘사건의 진행, 배경정보 등을 상세하게 담은 기사’<sup>5.27점</sup>는 ‘광고 없는 기사’<sup>5.58점</sup>, ‘논란에 대한 사실 확인 기사’<sup>5.40점</sup>에 이어 3번째였다. 현재 사람들이 보는 뉴스와 원하는 뉴스 사이에는 괴리가 있다.

이와 같은 문제의식은 현재 독자뿐만 아니라 학계와 언론계 전체에 퍼져있다. 우리는 이 문제를 해결하기 위해 ‘뉴스플로우’를 개발하게 되었다.

## 핵심 기능

‘뉴스플로우’는 주요 이슈에 대한 뉴스를 모아 타임라인형으로 시각화해준다. 각 타임라인에는 한 이슈의 세부 사건들이 표시된다. 따라서 사람들이 타임라인을 보면 하나의 이슈가 어떻게 훌려왔는지 사건의 흐름을 알 수 있다.

## 타깃층

‘뉴스플로우’는 시간 제약으로 단발성 뉴스밖에 보지 못하지만, 사건의 전후를 파악하고자 하는 사람들을 위한 프로그램이다. 구체적으로는 특정 사건의 흐름에 대한 체계적인 정리를 원하는 사람, 사건 국면이 바뀌는 양상을 보고 싶은 사람 등을 대상으로 한다.

## 강점 및 차별점

- 꾸준한 업데이트:** ‘뉴스플로우’는 자동으로 업데이트된다. 따라서 과거의 자료들을 한눈에 볼 수 있을 뿐더러, 현재의 뉴스도 볼 수 있다.

- 방대한 데이터양:** 빅카인즈 웹사이트를 통해 많은 양의 기사를 자료로 쓴다. 따라서 어느 플랫폼보다 더 많은 자료로 이슈를 정리한다.
- 객관성:** 특정 매체의 기사만 사용하는 것이 아니라 모든 뉴스를 이용해 콘텐츠를 만든다. 따라서 편집자의 시각이 들어가지 않아, 다른 매체보다 객관적일 수 있다.

## 1. 사건 시기별 뉴스 모으고 구분 짓기

- 이슈의 시작점과 끝점 찾기**  
빅카인즈 ‘타임라인 API’를 통해서 해당 주제가 이슈화되기 시작한 기점을 찾아낸다. 현재는 관련 주제에 대해 하루에 특정 횟수 이상으로 기사가 보도되면 이슈의 시작점으로 인식한다.
- 날짜별 핵심 키워드 검색**  
빅카인즈 ‘워드 클라우드 API’를 통해 이슈화되기 시작한 시간부터 현재 혹은 이슈가 끝난 시점까지 날짜별로 해당 주제와 관련된 키워드를 뽑아낸다.
- 단계 구분**  
찾아낸 핵심 키워드를 바탕으로, 각 날짜의 상위에 있는 키워드가 전날의 키워드와 얼마나 유사한지를 검사한다. 만약 당일의 상위 키워드가 전날 키워드와 비교하여 달라져 있다면 사건이 새로운 단계로 접어든 것으로 인식하고, 해당 일을 기점으로 날짜를 나눈다.
- 단계별 핵심 키워드 뽑기**  
모든 단계의 키워드를 종합한다. 해당 기간에 가장 높은 순위에 있는 키워드를 골라낸다. ‘워드 클라우드 API’에서 키워드와 함께 제공된 가중치를 사용해서, 특정 기간의 키워드 가중치들을 누적한다. 그리고 이를 가장 높은 순으로 정렬해서 상단에 있는 키워드들을 추출한다.
- 단계별 핵심 키워드로 키워드 재검색**  
자의적 이슈 시작점 설정으로 놓쳤던 키워드를 찾아낸다.

- 대표 뉴스

빅데이터 '뉴스 검색 API'를 사용하여 각 단계의 기간별로 뽑아낸 키워드를 사용하여 뉴스를 검색한다. 그중 가장 대표적인 뉴스가 타임라인 안에 해당 키워드와 함께 뜨게 한다. 나머지 뉴스들은 키워드 클릭 시에 정확도 순으로 뜨게 한다.

## 2. 추가 기능

- 제목 달기

중요 뉴스 안에 있는 키워드 중 가장 상위 랭크의 키워드로 타임라인의 제목을 만든다.

- 사건 요약문 작성

한 사건에 대한 요약문이 만들어진다.

- 사건 전환 색깔 시각화

한 사건이 어떻게 변화됐는지, 긍정·부정에 따라 타임라인에 올려진 키워드를 다르게 표시한다.

- 뉴스 중요도 분석

TF-IDF 알고리즘을 사용하여 과거 사건의 반전이 발생했을 때 댓글 반응 변화를 유사도 분석을 통하여 찾아낸다. 이에 따라 해당 뉴스의 중요도 여부를 판단한다. 부정적 댓글이 많을수록 뉴스의 중요도를 낮춘다.

- 타임라인형으로 이슈를 정리해놓았기에 독자들은 주요 이슈를 빠르게 파악할 수 있다. 또한 사건의 전후 맥락을 파악할 수 있게 된다.
- 독자들은 사건을 알기 위해 수많은 기사를 검색하지 않아도 되므로 정보 검색 시간을 절약할 수 있다.
- 과거에는 뉴스의 조각만을 보았기에 무엇이 새로운 뉴스인지 몰랐다. 하지만, 이제는 뉴스 플로우를 통해 사건의 전후 맥락을 파악했기에 무엇이 새로운 뉴스인지 알 수 있다.

## 기업뉴스

SPY Speed Perfect You

취업준비생의  
정보검색 시간을  
줄여주는  
뉴스빅데이터 활용  
기업분석 서비스

**팀명**  
뉴스 이슈(Issue)?  
  
**구성원**  
신예슬 서울시립대 대학원  
김담희  
김재순 인천대



01

## 발상 동기

취업 경쟁이 높아지면서 취업준비생이 준비해야 할 것도 많아지고 있다. 하지만 방대한 정보가 오히려 시간을 허비하게 만드는 경우도 있다. 어떤 자료를 어떻게 찾아볼지 모르거나, 검색방법을 알더라도 일일이 찾아보기에는 시간이 많이 걸린다.

이러한 문제를 해결하기 위해 뉴스데이터를 활용하여 취업에 필요한 정보를 정리해서 제공하는 서비스를 기획했다. 희망기업에 대한 뉴스, 면접준비에 필요한 이슈에 대한 뉴스 등을 제공하여 불필요한 검색시간을 줄이고 효율적인 취업준비를 할 수 있도록 돋는다.

02

## 아이디어 내용

### 기업 분석

- 관심 있는 기업을 검색하면, 특정기간 동안 기업 관련 뉴스의 상위 키워드 10개를 뽑아준다.
- 뽑아낸 상위 키워드를 클릭하면 원래 검색한 키워드 [관심기업](#)과 뽑아낸 키워드가 같이 들어간 기사가 최근순서로 표출된다.

### 이슈 분석

- 이슈가 되는 뉴스 2-3개 정도를 이용자에게 보여주고, 이용자끼리 자유롭게 의견을 달고 논쟁할 수 있는 공간을 제공한다.
- 잡이룸의 '이달의 시사주제'에서 아이디어를 얻어, 시사 주제를 순위권 뉴스에서 가져와서 제공한다.

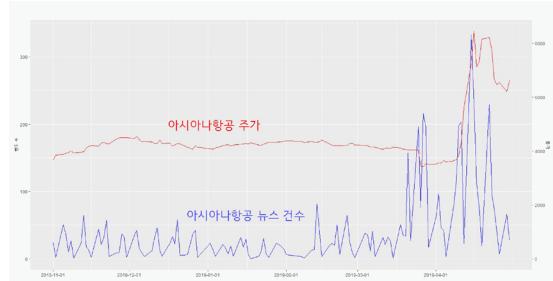
### 기업이슈 분석

- 관심 기업의 이슈를 한눈에 볼 수 있도록 검색한 기업의 최근 1년 동안의 주가와 그 기간 내에 기업의 검색 횟수를 그래프로 제공한다.
- 3월에 관심기업의 주가에 변화가 있다면, 해당월 동안 관심기업 뉴스를 '기업 분석' 메뉴와 같은 방식으로 보여준다.

03

## 서비스 구현

- 빅카인즈 키워드 추출 API를 활용하여 빈도순 상위 10개 키워드를 추출한다.
- 추출한 키워드에 대해 워드 클라우드 API를 활용하여 연관성이 높은 키워드 10개를 추출한다.
- 뉴스검색 API 활용 시 입력 키워드에 해당하는 산업군에 대한 상위 10개 키워드를 추출한다.
- 입력 키워드에 대한 주식데이터를 불러와, 뉴스타임라인 API, 인기검색어 API를 활용하여 주식데이터와 입력 키워드에 대한 뉴스 건수를 비교한다.
- 이슈랭킹 API를 활용하여 시사 토론 주제를 선정한다.
- 뉴스데이터를 통해 얻은 분석내용은 전체적으로 시각화한다.



기업 주가와 뉴스건수 비교 예시

04

## 기대 효과

- 취업준비생이 관심 있는 기업의 정보를 손쉽게 찾고 비교할 수 있게 함으로써, 정보검색 및 자료분석에 소요되는 시간을 단축시킨다.
- 키워드 중심의 기업 분석으로 관심 기업과 해당 기업이 속한 산업에 대한 취업준비생들의 이해도를 높인다.

# 2019 뉴스빅데이터 해커톤 본선대회

일시 2019. 5. 15.(수) 13시 30분

장소 마루180 이벤트홀(B1)

주최 한국언론진흥재단  
Korea Press Foundation



# 2019 뉴스 빅데이터 해커톤 수상작 모음집

발행인	민병욱
편집인	임영섭
기획	한국언론진흥재단 뉴스빅데이터팀
발행일	2019. 7. 19
발행처	한국언론진흥재단 04520 서울특별시 종구 세종대로 124 프레스센터 13층
	Tel 02. 2001. 7574
	Fax 02. 2001. 7560
	<a href="http://www.kpf.or.kr">www.kpf.or.kr</a>
편집	써니아일랜드