

2019 뉴스빅데이터 해커톤 과 제 설 명 서

2019. 3

한국언론진흥재단



1

과제 설명

I

대회 취지

- ☐ 한국언론진흥재단은 언론사 뉴스데이터 5,000만여 건을 축적하여, 검색 및 분석 서비스(www.bigkinds.or.kr)를 제공하고 있음
- ☐ 빅데이터가 미래기술로 떠오르면서 뉴스빅데이터도 주목받고 있지만, 뉴스를 활용한 IT 서비스의 실제 성공사례는 아직 부족한 상황임
- ☐ 뉴스데이터를 활용한 수익 모델 및 공공서비스를 발굴하여 산업 활성화 및 대국민 서비스 개선에 기여하고자 함

II

대회 목적

- ☐ 뉴스데이터를 활용한 IT 서비스 아이디어를 가진 우수 스타트업 발굴 및 지원
- ☐ 수상작에 대한 언론보도 · 수상집 공유 등을 통해 뉴스데이터 활용 아이디어 확산

III

스타트업 사례

- ☐ 시지온(Cizion)
 - 서로 다른 언론사 뉴스의 댓글을 모아 서비스하는 소셜 댓글 솔루션 ‘LiveRe’
 - 뉴스에서 표제어를 추출하여 대표 태그 선정, 유사한 주제의 뉴스를 클러스터링
 - 독자는 SNS로 로그인하여 간편하게 댓글을 달고, 언론사는 풍부한 댓글을 확보
 - 관련 사이트: www.livere.com
- ☐ 엠로보(M-Robo)
 - 기업공시, 금융지표 등에서 이벤트를 선별하여 뉴스를 자동 생성하는 로봇 뉴스 서비스
 - 특허정보를 뉴스, 공시정보 데이터와 매칭하여 검증하는 팩트체크 서비스
- ☐ 머니스테이션(Money Station)
 - 뉴스데이터, 금융정보 등을 분석하여 개인투자자를 위한 투자정보 분석 서비스
 - 주식 종목에 대한 투자정보와 함께 연관 뉴스 링크 제공
 - 관련 사이트: www.moneystation.net

□. 빅카인즈에서 제공하는 모든 데이터(www.bigkinds.or.kr)

○ 뉴스 메타정보 항목

항목	내용
뉴스식별자	해당 뉴스의 ID값
일자	발행 일자
언론사	발행 언론사
기고자	뉴스 작성자
제목	뉴스 제목
통합 분류	정치, 경제, 문화 등 카테고리 정보
사건 / 사고 분류	범죄, 사고 등 사건 / 사고 관련 카테고리 정보
인물 / 위치 / 기관	해당 뉴스에서 언급된 인물, 위치, 기관 추출 정보
키워드	해당 뉴스에서 추출된 모든 명사
특성추출	검색키워드와 한 문장 내에서 동시에 출현하는 빈도를 기준으로 추출한 중요 키워드 (텍스트랭크 알고리즘 사용)
본문	뉴스 본문 첫 200자
URL	원본 뉴스 주소
분석제외 여부	검색결과에 있으나 분석에서 제외되는 뉴스의 제외 사유를 기재 - 중복: 비슷한 기사가 반복적으로 수집되면 최신 기사 1건만 남기고 제외 - 예외: 인사, 부고, 동정, 사진은 의미 있는 분석결과를 기대하기 어려우므로 제외

○ 빅카인즈 API 제공

- 대회 참가자는 기획 및 개발을 위해 신청 시 빅카인즈 API를 이용할 수 있음
 - (API 기능) 뉴스검색, 뉴스조회, 이슈랭킹, 워드클라우드, 뉴스타임라인, 인기 검색어, 뉴스 인용문 검색, 분야별 키워드, 특성 추출, 키워드 추출
 - (이용가능기간) 발급일(신청일로부터 1~2일 소요) ~ 19.4.26 * 본선 진출자는 연장 가능
 - (신청방법) 이메일(sea@wplanet.co.kr) * 이름, 휴대전화번호 기재
- ※ API는 해커톤 대회 참가 목적으로만 이용 가능하며, 그 외 목적으로 활용 금지

□. 그 외 오픈 데이터 안내

- 우리말 샘 사전(opendict.korean.go.kr)
- 공공데이터 포털(www.data.go.kr)

□. 기타 활용하는 데 저작권 등의 문제가 없는 데이터