

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 왕경문	작성일 : 2023-06-21	버전 : 1.49	

content

본문 텍스트에 대한 키워드 추출 결과를 반환합니다.

13. 코드 테이블

13.1 언론사

코드값	언론사
01100101	경향신문
01100201	국민일보
01100301	내일신문
01100401	동아일보
01100501	문화일보
01100611	서울신문
01100701	세계일보
01100751	아시아투데이
01100801	조선일보
01100901	중앙일보
01101001	한겨레
01101101	한국일보
01101202	일요신문
01200101	경기일보
01200201	경인일보
01200301	기호일보
01200401	인천일보
01200501	경기신문
01300101	강원도민일보
01300201	강원일보
01400101	충청타임즈
01400201	대전일보
01400301	동양일보
01400351	중도일보
01400401	중부매일
01400501	중부일보
01400551	충북일보
01400601	충청일보
01400701	충청투데이
01400801	금강일보
01500051	경남신문
01500101	경남일보
01500151	경남도민일보
01500201	경북일보
01500301	경상일보
01500401	국제신문
01500501	대구일보
01500601	매일신문
01500701	부산일보
01500801	영남일보
01500901	울산매일
01501001	대구신문

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 왕경문	작성일 : 2023-06-21	버전 : 1.49	

코드값	언론사
01501101	경북도민일보
01501201	경북매일신문
01501301	울산신문
01600201	광주매일신문
01600301	광주일보
01600451	남도일보
01600501	무등일보
01600601	새전북신문
01600801	전남일보
01600901	전라일보
01601001	전북도민일보
01601101	전북일보
01620001	광남일보
01700101	제민일보
01700201	한라일보
01700401	제주일보
02100051	대한경제
02100101	매일경제
02100201	머니투데이
02100311	서울경제
02100351	이투데이
02100401	메트로경제
02100501	파이낸셜뉴스
02100601	한국경제
02100701	헤럴드경제
02100801	아시아경제
02100851	아주경제
02100861	브릿지경제
04100058	노컷뉴스
04100078	뉴스핌
04100158	데일리안
04100608	브레이크뉴스
04100958	EBN
04101008	이데일리
04101808	비즈워치
04102008	쿠키뉴스
04104008	프레시안
04400108	헬로디디
05202752	평택시민신문
05300152	설악신문
05420702	홍성신문
05510752	영주시민신문
05520352	당진시대
06100252	시사IN
06101202	주간한국
06101302	한겨레21
07100001	뉴스핌권
07100251	미디어오늘
07100501	전자신문
07100502	환경일보
07100602	PD-저널
07100702	기자협회보
07101201	디지털타임스



사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 왕경문	작성일 : 2023-06-21	버전 : 1.49	

코드값	언론사
07200502	소년한국일보
08100101	KBS
08100201	MBC
08100301	SBS
08100401	YTN
08200101	OBS
10100101	스포츠서울
10100301	스포츠한국
10100401	스포츠월드

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 왕경문	작성일 : 2023-06-21	버전 : 1.49	

13.2 뉴스 통합 분류체계

LEVEL	분류체계	코드값
1레벨	정치	001000000
2레벨	정치>북한	001001000
2레벨	정치>선거	001002000
2레벨	정치>외교	001003000
2레벨	정치>국회_정당	001004000
2레벨	정치>행정_자치	001005000
2레벨	정치>청와대	001006000
2레벨	정치>정치일반	001007000
1레벨	경제	002000000
2레벨	경제>무역	002001000
2레벨	경제>외환	002002000
2레벨	경제>유통	002003000
2레벨	경제>자원	002004000
2레벨	경제>산업_기업	002005000
2레벨	경제>증권_증시	002006000
2레벨	경제>취업_창업	002007000
2레벨	경제>금융_재테크	002008000
2레벨	경제>반도체	002009000
2레벨	경제>부동산	002010000
2레벨	경제>자동차	002011000
2레벨	경제>서비스_쇼핑	002012000
2레벨	경제>국제경제	002013000
2레벨	경제>경제일반	002014000
1레벨	사회	003000000
2레벨	사회>날씨	003001000
2레벨	사회>여성	003002000
2레벨	사회>환경	003003000
2레벨	사회>교육_시험	003004000
2레벨	사회>노동_복지	003005000
2레벨	사회>사건_사고	003006000
2레벨	사회>의료_건강	003007000
2레벨	사회>미디어	003008000
2레벨	사회>장애인	003009000
2레벨	사회>사회일반	003010000
1레벨	문화	004000000
2레벨	문화>생활	004001000
2레벨	문화>영화	004002000
2레벨	문화>음악	004003000
2레벨	문화>종교	004004000
2레벨	문화>출판	004005000
2레벨	문화>미술_건축	004006000
2레벨	문화>방송_연예	004007000
2레벨	문화>요리_여행	004008000
2레벨	문화>전시_공연	004009000
2레벨	문화>학술_문화재	004010000
2레벨	문화>문화일반	004011000
1레벨	국제	005000000
2레벨	국제>일본	005001000
2레벨	국제>중국	005002000
2레벨	국제>미국_북미	005003000

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 왕경문	작성일 : 2023-06-21	버전 : 1.49	

LEVEL	분류체계	코드값
2레벨	국제>중동_아프리카	005004000
2레벨	국제>러시아	005005000
2레벨	국제>아시아	005006000
2레벨	국제>중남미	005007000
2레벨	국제>유럽_EU	005008000
2레벨	국제>국제일반	005009000
1레벨	지역	006000000
2레벨	지역>강원	006001000
2레벨	지역>경기	006002000
2레벨	지역>경남	006003000
2레벨	지역>경북	006004000
2레벨	지역>광주	006005000
2레벨	지역>대구	006006000
2레벨	지역>대전	006007000
2레벨	지역>부산	006008000
2레벨	지역>울산	006009000
2레벨	지역>전남	006010000
2레벨	지역>전북	006011000
2레벨	지역>제주	006012000
2레벨	지역>충남	006013000
2레벨	지역>충북	006014000
2레벨	지역>지역일반	006015000
1레벨	스포츠	007000000
2레벨	스포츠>골프	007001000
2레벨	스포츠>야구	007002000
3레벨	스포츠>야구>메이저리그	007002001
3레벨	스포츠>야구>일본프로야구	007002002
3레벨	스포츠>야구>한국프로야구	007002003
2레벨	스포츠>축구	007003000
3레벨	스포츠>축구>해외축구	007003001
3레벨	스포츠>축구>국가대표팀	007003002
3레벨	스포츠>축구>한국프로축구	007003003
2레벨	스포츠>농구_배구	007005000
2레벨	스포츠>월드컵	007006000
2레벨	스포츠>올림픽_아시안게임	007010000
2레벨	스포츠>스포츠일반	007013000
1레벨	IT_과학	008000000
2레벨	IT_과학>과학	008001000
2레벨	IT_과학>보안	008002000
2레벨	IT_과학>모바일	008003000
2레벨	IT_과학>콘텐츠	008004000
2레벨	IT_과학>인터넷_SNS	008005000
2레벨	IT_과학>IT_과학일반	008006000

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 왕경문	작성일 : 2023-06-21	버전 : 1.49	

13.3 뉴스 사건/사고 분류체계

분류체계	코드값
범죄	1
범죄>성범죄	1,5
범죄>성범죄>성폭행	1,5,14
범죄>성범죄>성추행	1,5,15
범죄>성범죄>성희롱	1,5,16
범죄>성범죄>성매매	1,5,17
범죄>성범죄>음란물	1,5,18
범죄>기업범죄	1,6
범죄>기업범죄>반독점범죄	1,6,19
범죄>기업범죄>계약위반	1,6,20
범죄>기업범죄>횡령	1,6,21
범죄>기업범죄>내부자거래	1,6,22
범죄>기업범죄>거래제한	1,6,23
범죄>정치	1,7
범죄>정치>뇌물수수	1,7,24
범죄>범죄일반	1,8
범죄>범죄일반>방화	1,8,25
범죄>범죄일반>폭행	1,8,26
범죄>범죄일반>절도	1,8,27
범죄>범죄일반>유괴/납치	1,8,28
범죄>범죄일반>살인	1,8,29
범죄>범죄일반>사기	1,8,30
범죄>범죄일반>마약	1,8,31
사고	2
사고>산업사고	2,9
사고>산업사고>붕괴	2,9,32
사고>산업사고>폭발	2,9,33
사고>산업사고>화재	2,9,34
사고>산업사고>원자력사고	2,9,35
사고>교통사고	2,10
사고>교통사고>항공사고	2,10,36
사고>교통사고>우주사고	2,10,37
사고>교통사고>해상사고	2,10,38
사고>교통사고>철도사고	2,10,39
사고>교통사고>노상사고	2,10,40
재해	3
재해>자연재해	3,11
재해>자연재해>가뭄	3,11,41
재해>자연재해>지진	3,11,42
재해>자연재해>홍수	3,11,43
재해>자연재해>눈사태_산사태	3,11,44
재해>자연재해>태풍	3,11,45
재해>자연재해>폭염	3,11,46
재해>자연재해>해일	3,11,47
재해>자연재해>화산폭발	3,11,48
사회	4
사회>사회문제	4,12
사회>사회문제>학대	4,12,49
사회>사회문제>중독	4,12,50
사회>사회문제>미성년범죄	4,12,51

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 왕경문	작성일 : 2023-06-21	버전 : 1.49	

사회>사회문제>노예	4,12,52
사회>사회문제>자살	4,12,53
사회>사회갈등	4,13
사회>사회갈등>전쟁	4,13,54
사회>사회갈등>테러행위	4,13,55
사회>사회갈등>시위	4,13,56
사회>사회갈등>반란_혁명_폭동	4,13,57
사회>사회갈등>대량학살	4,13,58

13.4 토픽랭크 (TOPIC RANK) 알고리즘

알고리즘	상세
토픽랭크	<p>문서 내 중요 키워드를 통계적으로 뽑는 방식에는 대표적으로 TF/IDF 방식과 TEXTRANK 방식이 있습니다.</p> <p>TF/IDF 방식은 문서 내 해당 키워드의 발생 빈도수인 TF(Term Frequency)와 전체 분석대상 문서 군에서 발생빈도를 분모로 하는 IDF(Inverse Document Frequency)의 조합방식으로 분석 문서에서 많이 발생하면서 다른 문서에는 적게 발생하는 경우에 해당 문서의 대표 키워드로 선정합니다.</p> <p>계산식이 단순하기 때문에 정보검색에서 많이 활용되나 단순히 키워드의 영향력을 빈도수로만 측정하기 때문에 분석품질의 한계점이 존재합니다.</p> <p>Text Rank 분석은 문서 내의 키워드가 다른 키워드와 얼마나 빈번하게 연결되는가를 측정해 키워드의 문서내의 영향력을 측정하는 방식입니다.</p> <p>IDF값을 측정하지 않고 개별 문서만을 분석해 중요 키워드를 뽑을 수 있으며 영향력에 기반한 키워드의 가중치(weight)를 계산할 수 있기 때문에 텍스트마이닝에서 많이 활용되고 있습니다.</p> <p>토픽랭크(Topic Rank)는 Text Rank 알고리즘에서 토픽을 식별하는 기능을 추가 보완한 알고리즘입니다.</p>