
OpenAPI 사용자 지침서

V 1.45

뉴스 빅데이터 분석 시스템 구축

- 인수 -

2016-01-26 (제정)

2016-08-10 (개정)

2017-12-20 (개정)

2018-08-29 (개정)

2019-03-18 (개정)

2019-04-01 (개정)

2019-04-17 (개정)

2019-04-23 (개정)

2019-08-21 (개정)

2019-09-10 (개정)

2020-03-10 (개정)

2021-06-07 (개정)

2021-10-28 (개정)

관 리 본 문 서	Controlled Copy
문서 관리 번호	K-2015-D-101
배 포 번 호	/
작 성 자 (PM)	김진우

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

제.개정 이력서

버전	제,개정 일자	제,개정 페이지 및 내용	작성자	승인자
1.0	2016-01-26	제정	김진우	김진우
1.1	2016-02-17	Request 재정의	김진우	김진우
1.2	2016-04-25	Request 추가, 코드테이블 추가	정재훈	김진우
1.3	2016-04-29	키워드검색 Response 추가	정재훈	김진우
1.31	2016-05-17	언론사분류코드 분리, publisher_code 추가	정재훈	김진우
1.32	2017-04-24	언론사 코드테이블 추가, 접속URL 최신화, 뉴스검색 return필드 추가	유동주	정현철
1.33	2017-08-10	인용구 검색 추가	정현철	정현철
1.34	2017-12-20	분야별 키워드 추가	정현철	정현철
1.35	2018-08-29	특성 추출, 키워드 추출 추가	이승진	김은아
1.36	2019-03-18	서비스 언론사 목록 현행화	이승진	김은아
1.37	2019-04-01	이슈랭킹API 호출 결과에 topic_keyword 항목 설명 추가	이승진	김은아
1.38	2019-04-17	API 샘플 내용 수정	이승진	김은아
1.39	2019-04-23	API 설명 수정	이승진	김은아
1.4	2019-08-21	TopN 키워드 API 추가, 일부 API 명칭 변경 (이슈랭킹 API, 워드 클라우드 API, 뉴스 타임라인 API, 분야별 키워드 API)	이승진	김은아
1.41	2019-09-10	API KEY 수정	이승진	김은아
1.43	2020-03-10	뉴스 조회 API highlight 최대값 수정(300자→200자)	김은아	김은아
1.44	2021-06-07	뉴스 검색 API return_from 수정 최대값 20000	김민수	김민수
1.45	2021-10-28	뉴스검색,조회 tms_raw_stream(키워드) 컬럼 등록	김민수	김민수

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

<제목 차례>

1. 개요	5
1.1 연계 방식	5
1.2 인증 방식	5
2. 뉴스 검색 API	6
2.1 요청 URL	6
2.2 요청 형식	6
2.3 출력 결과 형식	10
3. 뉴스 조회 API	12
3.1 요청 URL	12
3.2 요청 형식	12
3.3 출력 결과 형식	13
4. 오늘의 이슈 API(이슈 랭킹 API)	13
4.1 요청 URL	13
4.2 요청 형식	13
4.3 출력 결과 형식	15
5. 연관어 분석 API(워드 클라우드 API)	16
5.1 요청 URL	16
5.2 요청 형식	16
5.3 출력 결과 형식	17
6. 키워드 트렌드 API(뉴스 타임라인 API)	18
6.1 요청 URL	18
6.2 요청 형식	18
6.3 출력 결과 형식	20
7. 인기검색어 API	21
7.1 요청 URL	21
7.2 요청 형식	21
7.3 출력 결과 형식	21
8. 뉴스 인용문 검색 API	22
8.1 요청 URL	22
8.2 요청 형식	22

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

8.3 출력 결과 형식	26
9. 오늘의 키워드 API(분야별 키워드 API)	28
9.1 요청 URL	28
9.2 요청 형식	28
9.3 출력 결과 형식	28
10. 특성 추출 API	31
10.1 요청 URL	31
10.2 요청 형식	31
10.3 출력 결과 형식	31
11. 키워드 추출 API	32
11.1 요청 URL	32
11.2 요청 형식	32
11.3 출력 결과 형식	32
12. TopN 키워드 API	33
12.1 요청 URL	33
12.2 요청 형식	33
12.3 출력 결과 형식	35
13. 코드 테이블	36
13.1 언론사	36
13.2 뉴스 통합 분류체계	38
13.3 뉴스 사건/사고 분류체계	40
13.4 토픽랭크 (TOPIC RANK) 알고리즘	41

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

1. 개요

Open API는 외부 시스템에서, "뉴스 빅데이터 분석 시스템"에서 제공하는 서비스들을 접근하여 사용할 수 있도록 기능을 제공합니다.

1.1 연계 방식

- HTTP 프로토콜의 POST 방식을 사용하여 연계할 수 있습니다.
- 입력/출력 데이터 형식은 UTF-8로 인코딩된 JSON 구조를 사용합니다.
- 입력 데이터는 JSON 객체로 전달해야 합니다.(JSON파일은 지원되지 않습니다.)

1.2 인증 방식

인증 방식은 인증키를 발급받은 후, 해당 인증키가 정상적으로 등록된 인증키여야만 사용 가능합니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :		작업명 :
작성자 : 김민수	작성일 : 2021-10-28		버전 : 1.45

2. 뉴스 검색 API

키워드와 다양한 조건을 지정하여 뉴스 데이터를 검색할 수 있습니다.

2.1 요청 URL

<http://tools.kinds.or.kr:8888/search/news>

2.2 요청 형식

```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "query": "서비스 AND 출시",
    "published_at": {
      "from": "2019-01-01",
      "until": "2019-03-31"
    },
    "provider": [
      "경향신문",
    ],
    "category": [
      "정치>정치일반",
      "IT_과학"
    ],
    "category_incident": [
      "범죄",
      "교통사고",
      "재해>자연재해"
    ],
    "byline": "",
    "provider_subject": [
      "경제", "부동산"
    ],
    "subject_info": [
      ""
    ],
    "subject_info1": [
      ""
    ],
    "subject_info2": [
      ""
    ],
    "subject_info3": [
      ""
    ],
    "subject_info4": [

```

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

```

    " ",
    ],
    "sort": {"date": "desc"},
    "highlight": 200,
    "return_from": 0,
    "return_size": 5
    "fields": [
        "byline",
        "category",
        "category_incident",
        "provider_news_id"
    ]
  }
}

```

query

검색 질의어를 지정합니다.(검색어 없이 검색 가능)

o 검색 연산자

구분	연산자	사용 예
부울린 검색	AND, OR, NOT	사과 AND 배 NOT 과일
구문 검색	" "	"아담과 이브"
연산자 우선 순위	()	(사과 OR 배) AND 과일

o 검색 영역 지정

설명	사용 예
기본 값은 제목과 본문에서 검색합니다.	"query": "사과 AND 배 NOT 과일"
제목에서 검색	"query": { "title": "사과 AND 배 NOT 과일" }
본문에서 검색	"query": { "content": "사과 AND 배 NOT 과일" }

published_at

검색 기간을 지정합니다.

- 날짜 형식은 한국 시간을 기준으로 YYYY-MM-DD 형식으로 지정합니다.

예약어	설명
from	조회 시작일을 지정합니다. 지정한 일자를 포함합니다.
until	조회 종료일을 지정합니다. 지정한 일자를 제외합니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

provider

언론사를 지정합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
- 여러 언론사들을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

category

한국언론진흥재단에서 지정한 뉴스 통합 분류체계 값을 조건으로 지정합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
- 여러 분류 값을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

category_incident

한국언론진흥재단에서 지정한 뉴스 사건/사고 분류체계 값을 조건으로 지정합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
- 여러 분류 값을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

byline

기자 이름을 조건으로 지정합니다.

- 기자 이름을 여러명 지정하기 위해서 OR 연산자를 사용 가능합니다.
- ※ 언론사에서 기자 이름을 제공한 News인 경우에만 검색이 가능합니다.

provider_subject

언론사에서 제공한 News의 주제 분류 값(코드)을 조건으로 지정 합니다.

☞ **subject_info** 또는 **subject_info1 ~ subject_info4** 를 이용한 검색을 권장합니다.

설명	사용 예
SubjectInfo = “경제, 부동산” SubjectInfo1 = “ent” SubjectInfo2 = “신문, 방송” SubjectInfo3 = “10000123” SubjectInfo4 = “others”	
SubjectInfo 에 값이 경제 또는 부동산이고, SubjectInfo3 은 “10000123” 인 기사 검색	"provider_subject": [“경제, 부동산”, “10000123”]
"provider_subject": []	배열의 항목은 AND로 검색됩니다. 배열 안의 “,” 구분자는 OR로 검색됩니다.

※ NewsML에서 정의된 SubjectInfo 와 SubjectInfo1 ~ 4 의 값을 검색합니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

※ 해당 주제 분류 값(코드)는 기사를 제공한 언론사와 시점에 따라 다르므로 통일된 값을 제공하지 않습니다.

subject_info / subject_info1 ~ subject_info4

언론사에서 NewML에 정의한 프로퍼티 SubjectInfo, SubjectInfo1~4의 내용을 검색합니다.

※ 기존 provider_subject의 모호성부분을 해결하기 위해 각 프로퍼티 값을 분리했습니다.

※ comma 또는 파이프(|) 구분자로 값을 지정한 경우는 각 해당하는 값을 배열로 선언해서 사용합니다.

sort

검색된 뉴스기사의 정렬 방식을 지정합니다.

설명	사용 예
지정하지 않으면, 기본 값은 검색 정확도 순으로 조회됩니다.	-
최신 기사 순서대로 조회	"sort":{ "date":"desc" }
과거 기사 순서대로 조회	"sort":{ "date":"asc" }
제목으로 정렬 (제목을 오름차순만 정렬됩니다.)	"sort":{ "title":"asc" }
정확도 정렬 (정확도는 내림차순으로만 정렬됩니다.)	"sort":{ "_score":"desc" }
최신순, 정확도순으로 다중정렬	"sort":[{"date":"desc"}, {"_score":"desc"}]
제목순, 날짜순으로 다중정렬	"sort":[{"title":"asc"}, {"date":"desc"}]
정렬로 지정가능 한 필드와 정렬값	date : 날짜, desc / asc title : 제목, asc _score : 정확도, desc byline : 기고자, asc provider_code : 매체코드, desc / asc

highlight

기사에서 검색어가 포함된 영역을 찾아 지정한 범위(최대 200자)를 출력 결과로 제공합니다.

return_from

검색된 결과 목록에서 반환할 시작 위치를 지정합니다.

return_from 필드는 최대 20000까지만 가능합니다.

return_size

검색된 결과 목록에서 반환할 시작 위치부터 반환할 개수를 지정합니다.

최대 10000까지 가능합니다. 특정 언론사는 최대 15000까지 제한을 두고 있습니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

fields

출력 결과에서 반환할 필드 값을 지정합니다.

필요하지 않은 경우라도 비어있는 배열([])을 입력해야 합니다.

※ 불필요한 데이터 전송을 줄이기 위해 지정하지 않은 필드의 값은 반환 하지 않습니다.

※ 지정할 수 있는 값은 2.3 출력 결과 형식의 documents 값을 참조하세요.

2.3 출력 결과 형식

```
{
  "result": 0,
  "return_object": {
    "total_hits": 155040,
    "documents": [
      {
        "news_id": "02100601.20211027093629001",
        "title": "여의도 하나금융투 담은 리츠, '코람코더원'으로 이름 변경",
        "highlight": "여의도 하나금융투 담은 리츠, ..'코람코더원'으로 이름 변경 이 기사는 10월 27일 05:56 .. '마켓인사이트'에 게재된 기사입니다... ..여의도 하나금융투자 빌딩을 담은 ..'코크렙제30호위탁관리부동산투자회사'(코크렙제30호리츠)가 회사명을 ..'코람코더원'으..25일 국토<b>교통</b>부는 코크렙제30호리츠의 변경인가를 승인한다고 밝혔다...",
        "published_at": "2021-10-27T00:00:00.000+09:00",
        "enveloped_at": "2021-10-28T09:36:29.000+09:00",
        "dateline": "2021-10-28T09:36:29.000+09:00",
        "provider": "한국경제",
        "category": [
          "경제>증권_증시",
          "경제>산업_기업",
          "경제>부동산"
        ],
        "category_incident": [],
        "byline": "윤아영",
        "images": "/02100601/2021/10/27/02100601.20211027093629001.01.jpg",
        "provider_news_id": "202110263143r",
        "publisher_code": "202110263143r",
        "tms_raw_stream": "여의도리츠코람코더원이름 변경기사05:56 마켓인사이트Wn게재기사Wn여의도금융 투자 빌딩Wn코크렙Wn30호Wn위탁관리부동산투자회사Wn코크렙Wn30호Wn리츠Wn회사명Wn코람코더원Wn변경Wn내년Wn기업공개WnIPOWn절차Wn진행Wn국토 교통부Wn코크렙제30호Wn리츠Wn변경Wn승인Wn상호Wn코람코더원Wn정관Wn변경Wn주식 일반 청약Wn유상증자Wn자금조달계획Wn변경Wn코람코더원리츠 운영Wn코람코자산신탁Wn내년Wn리츠 상장Wn추진Wn금융투자Wn주관사 리츠 상장Wn선정Wn보유 자산Wn여의도금융투자빌딩Wn준공Wn지하Wn지상Wn규모Wn연면적Wn지하철Wn환승역Wn여의도역Wn이용Wn여의도 공원Wn여의도환승센터"
      }
    ]
  }
}
```

total_hits

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

검색된 전체 문서 건수를 반환합니다.

documents

검색된 문서들에서 요청한 범위에 해당하는 문서를 JSON 객체의 배열로 반환합니다.

※ fields 항목에 지정한 필드만 값이 반환됩니다.

필드명	설명
news_id	언론진흥재단내에서 관리하는 뉴스 식별자입니다.
title	뉴스기사 제목
content	뉴스기사 본문
published_at	뉴스기사 발행일을 세계표준시간 기준으로 반환합니다. ex) "2019-03-14T00:00:00.000+09:00"
enveloped_at	언론사로부터 뉴스 기사를 수집한 시간
dateline	언론사로부터 전달받은 언론사의 뉴스 출고시간
provider	언론사
category	뉴스통합분류체계
category_incident	뉴스사건/사고 분류체계
highlight	검색어가 포함된 영역입니다. 요청파라미터에서 지정한 경우에만 반환합니다.
byline	기자 이름
images	이미지가 포함된 뉴스 기사의 경우 이미지 파일 이름을 반환 ex) /01100701/2016/02/17/P160217124532014021.jpg
provider_subject	언론사에서 제공한 뉴스 분류 값(코드) ex> "provider_subject" : [{ "subject_info" : "사회" , "subject_info1" : "1234" , ... }]
subject_info	NewsML에서 정의한 프로퍼티 SubjectInfo 의 값 comma 또는 파이프() 구분자는 배열로 변환되어 반환
subject_info1	NewsML에서 정의한 프로퍼티 SubjectInfo1 의 값 comma 또는 파이프() 구분자는 배열로 변환되어 반환
subject_info2	NewsML에서 정의한 프로퍼티 SubjectInfo2 의 값 comma 또는 파이프() 구분자는 배열로 변환되어 반환
subject_info3	NewsML에서 정의한 프로퍼티 SubjectInfo3 의 값 comma 또는 파이프() 구분자는 배열로 변환되어 반환
subject_info4	NewsML에서 정의한 프로퍼티 SubjectInfo4 의 값 comma 또는 파이프() 구분자는 배열로 변환되어 반환
provider_news_id	언론사에서 제공한 뉴스 고유 식별값 ※ 언론사 마다 서로 다른 형식을 사용합니다.
publisher_code	언론사에서 제공한 뉴스 고유 식별값 ※ 언론사 마다 서로 다른 형식을 사용합니다.
provider_link_page	언론사 원본 링크페이지(없는 경우 존재)
printing_page	지면번호(없는 경우 존재)
tms_raw_stream	title, content에 포함되어 있는 키워드(명사)를 추출하여 String 값으로 반환

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

3. 뉴스 상세 정보 조회 API

검색한 뉴스에 대해서 뉴스 상세 정보를 제공하기 위해 지정한 뉴스 식별자에 해당하는 뉴스만을 제공합니다.

해당 뉴스 상세 정보는 fields 하위 Element 필드에 의해 상세 정보가 반환됩니다.

3.1 요청 URL

<http://tools.kinds.or.kr:8888/search/news>

3.2 요청 형식

```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "news_ids": [
      "01500701.2015083110018412570",
      "01100701.20150826100000152"
    ],
    "fields": [
      "content",
      "byline",
      "category",
      "category_incident",
      "images",
      "provider_subject",
      "provider_news_id",
      "publisher_code"
    ]
  }
}
```

news_ids

조회할 뉴스 식별자들을 지정합니다.

※ provider_news_id 의 값을 지정해도 같은 결과가 반환됩니다.

※ publisher_code 의 값을 지정해도 같은 결과가 반환됩니다.

※ fields 하위 element - content : 본문내용, byline : 기고자, category : 언론진흥재단 통합 분류, category_incident : 언론진흥재단 사건사고 분류, images : 이미지 정보, provider_subject : 언론사 주제 분류코드, publisher_code : 언론사 고유 기사코드

fields

출력 결과에서 반환할 필드 값을 지정합니다.

※ 불필요한 데이터 전송을 줄이기 위해 지정하지 않은 필드의 값은 반환 하지 않습니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

3.3 출력 결과 형식

출력 결과 형식은 뉴스 검색 API의 2.3 출력 결과 형식을 참조하세요.

4. 오늘의 이슈 API

※ (기존) 이슈랭킹 → (변경) 오늘의 이슈

최근 게재된 뉴스를 분석하여 주요 이슈와 관련된 뉴스를 목록(랭킹)으로 제공합니다.

4.1 요청 URL

http://tools.kinds.or.kr:8888/issue_ranking

4.2 요청 형식

```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "date": "2019-01-01"
    "provider": []
  }
}
```

date

조회 일자를 지정합니다. 검색 일자를 지정합니다.

- 제공되는 조회일자는 1990-01-01부터 지정할 수 있습니다.

provider

지정한 언론사에서 제공한 뉴스가 포함된 토픽만을 반환합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
- 여러 언론사들을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.

※ 코드 값은 하단 코드 테이블 참조

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

4.3 출력 결과 형식

```
{
  "result": 0,
  "return_object": {
    "date": "2019-01-01",
    "topics": [
      {
        "topic": "임종석 · 조국 국회 운영위 출석 여야 난타전",
        "topic_rank": 270,
        "topic_keyword": "조국,임종석,종석,사찰,국회,김태우,출석,운영,정치,운영위,민간인,대통령,한국,수석,나경원,의혹,한국당,정부,민정,공세,민정수석,리스트,블랙리스트,vs,삼인성호,정치적,정치공세,비리,특감반,홍영표,여야,논란,국민,개탄,파면,청와대,농간,민간사찰,민주,문재인,책임,비위,공방,수사,개인,희대,시비,사람,일탈,전희경",
        "news_cluster": [
          "01101101.20181231050408001",
          "01101101.20181231050400002",
          "01101101.20181231050354002",
          "01101101.20181231050259002",
          "02100201.20181231085859001"
        ]
      },
      {
        "topic": "신재민 청와대 적자 국채 발행 강요 폭로",
        "topic_rank": 85,
        "topic_keyword": "기재부,신재민,사무관,폭로,국채,발행,주장,사장,적자,교체,법적,대응,지시,사실,청와대,강요,서울신문,서울,적자국채,정부,압력,국채발행,신뢰성,압박,의심,추가,개입,인사,김병준,파악,조치,위치,국가,유튜브,차관,친서,타락,4조,취소,양심,유감,전여옥,권력,의혹,5급,전직,메가스터디,김정은,토론,두절",
        "news_cluster": [
          "01101101.20181231050252001",
          "01100101.20181231213652002",
          "01100801.20181231213032001",
          "08200101.20181231205951001"
        ]
      }
    ]
  }
}
```

date

조회된 일자를 반환합니다.

topics

분석된 이슈 키워드와 관련된 뉴스 식별자들을 중요도 순서로 10개를 반환합니다.

필드명	설명
topic	뉴스에서 추출한 주요 이슈 키워드들
news_cluster	이슈와 관련된 뉴스들의 고유 식별자를 배열로 반환합니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

topic_rank	내부적으로 순위를 부여하기 위한 일련 번호입니다. 숫자가 작을수록 중요한 이슈입니다.
topic_keyword	토픽랭크 알고리즘으로 뉴스에서 추출한 키워드

5. 연관어 분석 API

※ (기존) 워드 클라우드 → (변경) 연관어 분석

입력 키워드와 연관이 높은 키워드들을 지정한 검색 조건에 해당하는 문서들에서 실시간으로 추출하여 제공합니다.

※ 연관어 분석은 문서 내에서 사용자가 입력한 질의어를 기반으로 연관 단어를 추출하여 문서의 특징을 직관적으로 확인할 수 있도록 하는 분석기법입니다. (TOPIC RANK 알고리즘을 사용하며 하단 TOPIC RANK 알고리즘 참조)

5.1 요청 URL

http://tools.kinds.or.kr:8888/word_cloud

5.2 요청 형식

```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "query": "청와대",
    "published_at": {
      "from": "2019-01-01",
      "until": "2019-01-31"
    },
    "provider": [
      "경향신문",
    ],
    "category": [
      "정치>정치일반",
      "스포츠일반",
    ],
    "category_incident": [
      "범죄",
      "교통사고",
    ],
    "byline": "",
    "provider_subject": [ "사회" ]
  }
}
```

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

query(필수값)

관련된 키워드를 추출하기 위한 키워드를 지정합니다.

published_at

연관 키워드를 추출하기 위한 검색 기간을 지정합니다.

- 날짜 형식은 한국 시간을 기준으로 YYYY-MM-DD 형식으로 지정합니다.

예약어	설명
from	조회 시작일을 지정합니다. 지정한 일자를 포함합니다.
until	조회 종료일을 지정합니다. 지정한 일자를 제외합니다.

provider

지정한 언론사에서 게재한 뉴스만을 분석합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
 - 여러 언론사들을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

category

한국언론진흥재단에서 지정한 뉴스 통합 분류체계 값을 조건으로 지정합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
 - 여러 분류 값을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

category_incident

한국언론진흥재단에서 지정한 뉴스 사건/사고 분류체계 값을 조건으로 지정합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
 - 여러 분류 값을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

byline

기자 이름을 조건으로 지정합니다.

- 기자 이름을 여러명 지정하기 위해서 OR 연산자를 사용 가능합니다.
- ※ 언론사에서 기자 이름을 제공한 News인 경우에만 검색이 가능합니다.

provider_subject

언론사에서 제공한 News의 주제 분류 값(코드)을 조건으로 지정 합니다.

- ※ 해당 주제 분류 값(코드)는 기사를 제공한 언론사와 시점에 따라 다르므로 통일된 값을 제공하지 않습니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

※ 뉴스검색 API 2.2. 요청형식을 참고하세요.

5.3 출력 결과 형식

```
{
  "result": 0,
  "return_object": {
    "nodes": [
      {
        "id": 2,
        "name": "청원",
        "level": 3,
        "weight": 1.33
      },
      {
        "id": 3,
        "name": "유씨",
        "level": 3,
        "weight": 0.86
      },
      {
        "id": 4,
        "name": "누리꾼",
        "level": 3,
        "weight": 0.67
      },
      {
        "id": 5,
        "name": "명예훼손 혐의",
        "level": 3,
        "weight": 0.67
      }
    ]
  }
}
```

nodes

지정한 검색어와 연관된 키워드들을 반환합니다.

필드명	설명
name	토픽명칭을 반환합니다.
weight	토픽의 중요도를 수치로 반환합니다.(하단 TOPIC RANK 알고리즘 참고)
id	토픽들을 구분하기 위한 식별자입니다.
level	지정한 검색어와 토픽간의 연결 단계를 나타냅니다. 그래프 분석을 위한 예약어로 워드 클라우드에서는 무시하셔도 됩니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

6. 키워드 트렌드 API

※ (기존) 뉴스 타임라인 → (변경) 키워드 트렌드

입력 키워드와 언론사, 분류 조건을 입력하여 기간별 게재된 뉴스 데이터의 건수를 집계하여 제공합니다.

6.1 요청 URL

http://tools.kinds.or.kr:8888/time_line

6.2 요청 형식

```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "query": "정부",
    "published_at": {
      "from": "2018-01-01",
      "until": "2018-12-31"
    },
    "provider": [
      "경향신문",
    ],
    "category": [
      "정치>정치일반",
      "스포츠일반",
    ],
    "category_incident": [
      "범죄",
      "교통사고",
    ],
    "byline": "",
    "provider_subject": [ "사회" ],
    "interval": "month",
    "normalize": false
  }
}
```

query(필수값)

관련된 키워드를 추출하기 위한 키워드를 지정합니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :		작업명 :
작성자 : 김민수	작성일 : 2021-10-28		버전 : 1.45

published_at

연관 키워드를 추출하기 위한 검색 기간을 지정합니다.

- 날짜 형식은 한국 시간을 기준으로 YYYY-MM-DD 형식으로 지정합니다.

예약어	설명
from	조회 시작일을 지정합니다. 지정한 일자를 포함합니다.
until	조회 종료일을 지정합니다. 지정한 일자를 제외합니다.

provider

지정한 언론사에서 게재한 뉴스만을 분석합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
 - 여러 언론사들을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

category

한국언론진흥재단에서 지정한 뉴스 통합 분류체계 값을 조건으로 지정합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
 - 여러 분류 값을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

category_incident

한국언론진흥재단에서 지정한 뉴스 사건/사고 분류체계 값을 조건으로 지정합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
 - 여러 분류 값을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

byline

기자 이름을 조건으로 지정합니다.

- 기자 이름을 여러명 지정하기 위해서 OR 연산자를 사용 가능합니다.
- ※ 언론사에서 기자 이름을 제공한 News인 경우에만 검색이 가능합니다.

provider_subject

언론사에서 제공한 News의 주제 분류 값(코드)을 조건으로 지정 합니다.

- ※ 해당 주제 분류 값(코드)는 기사를 제공한 언론사와 시점에 따라 다르므로 통일된 값을 제공하지 않습니다.
- ※ 뉴스검색 API 2.2. 요청형식을 참고하세요.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

interval

집계할 기간을 지정합니다.

예약어	설명
day	일자 별로 검색된 뉴스 건수를 집계합니다.
month	월 별로 검색된 뉴스 건수를 집계합니다.
year	년도 별로 검색된 뉴스 건수를 집계합니다.

normalize

boolean 값을 지정하여 정규화 수행여부 선택.

true이면 색인된 문서건수에 맞추어 검색 된 문서 건수를 정규화 수행.

정규화 공식 : 기간 평균 수집 문서양*(해당 기간에 검색된 문서 건수/해당 기간에 수집된 문서 건수)

6.3 출력 결과 형식

```
{
  "result": 0,
  "return_object": {
    "total_hits": 27,
    "time_line": [
      {
        "label": "201801",
        "hits": 2
      },
      {
        "label": "201802",
        "hits": 2
      },
      {
        "label": "201803",
        "hits": 2
      },
      {
        "label": "201804",
        "hits": 1
      },
      {
        "label": "201805",
        "hits": 6
      }
    ]
  }
}
```

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :		작업명 :
작성자 : 김민수	작성일 : 2021-10-28		버전 : 1.45

total_hits

검색된 전체 문서 건수를 반환합니다.

time_line

기간별 조회된 뉴스 현황을 목록으로 반환합니다.

필드명	설명
label	집계기간으로, interval 값에 따라 형식이 달라집니다. day=yyyyMMdd, month=yyyyMM, year=yyyy
hits	해당 기간에 조회조건으로 검색된 문서 건수

7. 인기검색어 API

사용자가 가장 많이 검색한 키워드를 추출하여 제공합니다.

7.1 요청 URL

http://tools.kinds.or.kr:8888/query_rank

7.2 요청 형식

```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "from": "2018-01-01",
    "until": "2018-01-01",
    "offset": 5,
    "target_access_key": ""
  }
}
```

from ~ until

검색할 날짜의 범위를 지정합니다.

offset

결과값으로 전달받을 개수를 지정합니다.

target_access_key

발급받은 access_key 값을 조건에 넣을 경우 해당업체에서 뉴스검색에 사용한 키워드에 대해서만 인기검색어를 추출합니다.(값을 지정하지 않을 경우 기본적으로 제공되는 인기검색어는 빅카인즈 웹서비스 + OpenAPI를 사용하는 모든업체의 뉴스검색 키워드를 기준으로 추출한 데이터입니다.)

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

7.3 출력 결과 형식

```
{
  "result": 0,
  "return_object": {
    "queries": [
      {
        "date": "20180101",
        "query": "부시",
        "count": 2105
      },
      {
        "date": "20180101",
        "query": "조지",
        "count": 1405
      },
      {
        "date": "20180101",
        "query": "오바마",
        "count": 1336
      },
      {
        "date": "20180101",
        "query": "칼럼",
        "count": 880
      },
      {
        "date": "20180101",
        "query": "남북",
        "count": 842
      }
    ]
  }
}
```

8. 뉴스 인용문 검색 API

키워드와 다양한 조건을 지정하여 뉴스 인용문 데이터를 검색할 수 있습니다.

8.1 요청 URL

<http://tools.kinds.or.kr:8888/search/quotation>

8.2 요청 형식

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "query": "빅데이터",
    "published_at": {
      "from": "2019-01-01",
      "until": "2019-03-30"
    },
    "provider": [
      "경향신문",
      "국민일보",
      "내일신문"
    ],
    "category": [
      "정치>정치일반",
      "스포츠일반",
      "IT_과학"
    ],
    "category_incident": [
      "범죄",
      "교통사고",
      "재해>자연재해"
    ],
    "byline": "",
    "provider_subject": [
      ""
    ],
    "subject_info": [
      ""
    ],
    "subject_info1": [
      ""
    ],
    "subject_info2": [
      ""
    ],
    "subject_info3": [
      ""
    ],
    "subject_info4": [
      ""
    ],
    "sort": {
      "date": "desc"
    },
    "highlight": 200,
    "return_from": 0,
    "return_size": 5,
    "fields": [
```

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

```

        "highlight",
        "byline",
        "category",
        "category_incident",
        "images",
        "provider_subject",
        "subject_info",
        "provider_news_id",
        "publisher_code"
    ]
}
}

```

query

검색 질의어를 지정합니다. (검색어 없이 검색가능)

o 검색 연산자

구분	연산자	사용 예
부울린 검색	AND, OR, NOT	사과 AND 배 NOT 과일
구문 검색	" "	"아담과 이브"
연산자 우선 순위	()	(사과 OR 배) AND 과일

published_at

검색 기간을 지정합니다.

- 날짜 형식은 한국 시간을 기준으로 YYYY-MM-DD 형식으로 지정합니다.

예약어	설명
from	조회 시작일을 지정합니다. 지정한 일자를 포함합니다.
until	조회 종료일을 지정합니다. 지정한 일자를 제외합니다.

provider

언론사를 지정합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
 - 여러 언론사들을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

category

한국언론진흥재단에서 지정한 뉴스 통합 분류체계 값을 조건으로 지정합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
 - 여러 분류 값을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

category_incident

한국언론진흥재단에서 지정한 뉴스 사건/사고 분류체계 값을 조건으로 지정합니다.

- 명칭 또는 코드 값으로 지정할 수 있습니다.
 - 여러 분류 값을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.
- ※ 코드 값은 하단 코드 테이블 참조

byline

기자 이름을 조건으로 지정합니다.

- 기자 이름을 여러명 지정하기 위해서 OR 연산자를 사용 가능합니다.
- ※ 언론사에서 기자 이름을 제공한 News인 경우에만 검색이 가능합니다.

provider_subject

언론사에서 제공한 News의 주제 분류 값(코드)을 조건으로 지정 합니다.

 **subject_info 또는 subject_info1~ subject_info4 를 이용한 검색을 권장합니다.**

설명	사용 예
SubjectInfo = “경제, 부동산” SubjectInfo1 = “ent” SubjectInfo2 = “신문, 방송” SubjectInfo3 = “10000123” SubjectInfo4 = “others”	
SubjectInfo 에 값이 경제 또는 부동산이고, SubjectInfo3 은 “10000123” 인 기사 검색	"provider_subject": [“경제, 부동산”, “10000123”]
"provider_subject": []	배열의 항목은 AND로 검색됩니다. 배열 안의 “,” 구분자는 OR로 검색됩니다.

※ NewsML에서 정의된 SubjectInfo 와 SubjectInfo1 ~ 4 의 값을 검색합니다.

※ 해당 주제 분류 값(코드)는 기사를 제공한 언론사와 시점에 따라 다르므로 통일된 값을 제공하지 않습니다.

subject_info / subject_info1 ~ subject_info4

언론사에서 NewML에 정의한 프로퍼티 SubjectInfo, SubjectInfo1~4의 내용을 검색합니다.

- ※ 기존 provider_subject의 모호성부분을 해결하기 위해 각 프로퍼티 값을 분리했습니다.
- ※ comma 또는 파이프(|) 구분자로 값을 지정한 경우는 각 해당하는 값을 배열로 선언해서 사용합니다.

sort

검색된 뉴스기사의 정렬 방식을 지정합니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

설명	사용 예
지정하지 않으면, 기본 값은 검색 정확도 순으로 조회됩니다.	-
최신 기사 순서대로 조회	"sort":{ "date":"desc" }
과거 기사 순서대로 조회	"sort":{ "date":"asc" }
제목으로 정렬 (제목을 오름차순만 정렬됩니다.)	"sort":{ "title":"asc" }
정확도 정렬 (정확도는 내림차순으로만 정렬됩니다.)	"sort":{ "_score":"desc" }
최신순, 정확도순으로 다중정렬	"sort":[{ "date":"desc" }, { "_score":"desc" }]
제목순, 날짜순으로 다중정렬	"sort":[{ "title":"asc" }, { "date":"desc" }]
정렬로 지정가능 한 필드와 정렬값	date : 날짜, desc / asc title : 제목, asc _score : 정확도, desc byline : 기고자, asc provider_code : 매체코드, desc / asc

highlight

기사에서 검색어가 포함된 영역을 찾아 지정한 범위(최대 300자)를 출력 결과로 제공합니다.

return_from

검색된 결과 목록에서 반환할 시작 위치를 지정합니다.

최대 1000까지 가능합니다.

return_size

검색된 결과 목록에서 반환할 시작 위치부터 반환할 개수를 지정합니다.

최대 100까지 가능합니다.

fields

출력 결과에서 반환할 필드 값을 지정합니다.

필요하지 않은 경우라도 비어있는 배열([])을 입력해야 합니다.

※ 불필요한 데이터 전송을 줄이기 위해 지정하지 않은 필드의 값은 반환 하지 않습니다.

※ 지정할 수 있는 값은 2.3 출력 결과 형식의 documents 값을 참조하세요.

8.3 출력 결과 형식

```
{
  "result": 0,
  "return_object": {
```

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

```

"total_hits": 72,
"documents": [
  {
    "news_id": "01100201.20190111193338001",
    "published_at": "2019-01-11T00:00:00.000+09:00",
    "date": "20190111",
    "provider": "국민일보",
    "category": [
      "IT_과학>인터넷_SNS",
      "IT_과학>과학"
    ],
    "category_incident": [
      "범죄>기업범죄>거래제한"
    ],
    "provider_subject": [
      {
        "subject_info1": "경제일반",
        "subject_info2": "11151100",
        "subject_info3": "",
        "subject_info4": "",
        "subject_info": "경제"
      }
    ],
    "publisher_code": "",
    "subject_info": [
      "경제"
    ],
    "source": "문재인 대통령",
    "quotation": "한국형 규제 샌드박스 시행은 신기술, 신제품의 빠른 시장성 점검과 출시를 도울 것"
  }
]

```

total_hits

검색된 전체 문서 건수를 반환합니다.

documents

검색된 문서들에서 요청한 범위에 해당하는 문서를 JSON 객체의 배열로 반환합니다.

※ fields 항목에 지정한 필드만 값이 반환됩니다.

필드명	설명
news_id	언론진흥재단내에서 관리하는 뉴스 식별자입니다.
published_at	뉴스기사 발행일을 세계표준시간 기준으로 반환합니다. ex) "2015-12-22T00:00:00.000+09:00"
date	발행일
provider	언론사
category	뉴스통합분류체계

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

category_incident	뉴스사건/사고 분류체계
provider_subject	언론사에서 제공한 뉴스 분류 값(코드) ex> “provider_subject” : [{ “subject_info” : “사회” , “subject_info1” : “1234” , ... }]
subject_info	NewsML에서 정의한 프로퍼티 SubjectInfo 의 값 comma 또는 파이프() 구분자는 배열로 변환되어 반환
subject_info1	NewsML에서 정의한 프로퍼티 SubjectInfo1 의 값 comma 또는 파이프() 구분자는 배열로 변환되어 반환
subject_info2	NewsML에서 정의한 프로퍼티 SubjectInfo2 의 값 comma 또는 파이프() 구분자는 배열로 변환되어 반환
subject_info3	NewsML에서 정의한 프로퍼티 SubjectInfo3 의 값 comma 또는 파이프() 구분자는 배열로 변환되어 반환
subject_info4	NewsML에서 정의한 프로퍼티 SubjectInfo4 의 값 comma 또는 파이프() 구분자는 배열로 변환되어 반환
publisher_code	언론사에서 제공한 뉴스 고유 식별값 ※ 언론사 마다 서로 다른 형식을 사용합니다.

9. 오늘의 키워드 API

※ (기존) 분야별 키워드 → (변경) 오늘의 키워드

빅카인즈 메인화면에서 사용하는 오늘의 키워드 정보를 볼 수 있습니다.

9.1 요청 URL

http://tools.kinds.or.kr:8888/today_category_keyword

9.2 요청 형식

```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {}
}
```

9.3 출력 결과 형식

```
{
  "result": 0,
  "return_object": {
    "date": "20190423",
    "cate_ratio": [
      {
        "category_percent": 100,

```

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

```

    "category_name": "전체",
    "category_code": "000000000",
    "category_count": 733
  },
  {
    "category_percent": 8.9,
    "category_name": "정치",
    "category_code": "001000000",
    "category_count": 65
  },
  {
    "category_percent": 13.6,
    "category_name": "경제",
    "category_code": "002000000",
    "category_count": 100
  },
  {
    "category_percent": 13.4,
    "category_name": "사회",
    "category_code": "003000000",
    "category_count": 98
  },
  {
    "category_percent": 12.3,
    "category_name": "문화",
    "category_code": "004000000",
    "category_count": 90
  },
  {
    "category_percent": 11.1,
    "category_name": "국제",
    "category_code": "005000000",
    "category_count": 81
  }
],
"cate_keyword": [
  {
    "category_name": "전체",
    "category_code": "000000000",
    "named_entity_count": 681,
    "named_entity": "바른미래당",
    "entity_step": "step4",
    "named_entity_type": "OG"
  },
  {
    "category_name": "전체",
    "category_code": "000000000",
    "named_entity_count": 582,
    "named_entity": "삼성전자",

```

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

```

        "entity_step": "step4",
        "named_entity_type": "OG"
    },
    {
        "category_name": "전체",
        "category_code": "000000000",
        "named_entity_count": 526,
        "named_entity": "한국당",
        "entity_step": "step4",
        "named_entity_type": "OG"
    }
]
}

```

date

오늘 날짜 정보를 반환합니다.

documents (cate_ratio)

오늘의 키워드 카테고리 정보를 JSON 객체의 배열로 반환합니다.

필드명	설명
category_percent	카테고리 전체 퍼센테이지 정보를 반환합니다.
category_name	카테고리명 정보를 반환합니다.
category_code	카테고리 내부 고유코드값을 반환합니다.
category_count	카테고리 개수를 반환합니다.

documents (cate_keyword)

오늘의 키워드 카테고리 정보를 JSON 객체의 배열로 반환합니다.

필드명	설명
category_name	카테고리명 정보를 반환합니다.
category_code	카테고리 내부 고유코드값을 반환합니다.
named_entity	오늘의 키워드 명을 반환합니다.
named_entity_count	오늘의 키워드 개수를 반환합니다.
named_entity_type	오늘의 키워드 개체타입을 반환합니다. (PS : 인물, LC : 장소, OG:기관)
entity_step	step1~step5을 반환합니다. 해당 정보를 이용해서 UI에서 글자크기로 표시하기 위한 목적. (step1 : 10개 미만, step2 : 20개 미만, step3 : 50개 미만, step4 : 100개 미만, step5 100개 이상)

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

10. 특성 추출 API

입력한 텍스트의 특성 추출 결과를 제공합니다.

JSON 형태로 사용 시 텍스트 내용에 쌍따옴표("")가 포함되어있는 경우, 해당 문자 앞에 역슬래시(\\) 문자 입력이 필요합니다.

예) 국지성 호우..."피해 대비 철저히" -> 국지성 호우..."\\피해 대비 철저히\\"
 "주의가 필요하다"고 말했다. -> \"주의가 필요하다\"고 말했다.

10.1 요청 URL

<http://tools.kinds.or.kr:8888/feature>

10.2 요청 형식

```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "title": "'태풍 솔릭'에 항공기 대규모 결항...오후 10시까지 총 532편 결항",
    "sub_title": "일자리 창출 · 사회안전망 확충...내년 사상 최대 '나랏돈' 분다",
    "content": "2018아시안게임 남자 축구에서 박항서 감독이 이끄는 베트남 국가대표팀의 돌풍이 심상치 않다."
  }
}
```

10.3 출력 결과 형식

```
{
  "result": 0,
  "return_object": {
    "result": {
      "title": "항공기|0.1714048555587672 항공기_대규모_결항|0.17083741949082015  
대규모|0.1702700199700766 항공기_대규모|0.15692294958078706 532편_결항  
|0.14910529827199845 솔릭|0.1396095846626425 태풍_솔릭|0.10714353565037664 결항  
|0.0020622143949391588 태풍|0.001001650421856614",
      "sub_title": "사상_최대|0.09947331867942348 사회안전망|0.09871378441773095  
사상_최대_나랏돈|0.09792899284307392 안전망|0.09784893867007033 나랏돈  
|0.09602439002004316 일자리_창출|0.07639166062533455 일자리|0.07091165132567441 사상  
|0.0012416507498564445 창출|0.0010053642701117122 사회|9.812542367098709E-4 확충  
|9.747380712681627E-4 내년|9.643652168532188E-4 최대|9.601827804673364E-4",
      "content": "아시안|0.10055281584337647 국가대표팀|0.10054541885073295 베트  
남|0.09873194315687864 박항서|0.09801727923793947 베트남_국가대표팀|0.09151655431344134  
박항서_감독|0.0901096877211987 남자_축구|0.09010562295113558 2018아시안  
|0.07092712146925532 게임|9.873683661305109E-4 감독|9.818052150006968E-4 남자  
|9.817490185732548E-4 축구|9.801404465515614E-4 돌풍|7.092032554073097E-4"
    }
  }
}
```

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

```

    }
  }
}

```

title

제목 텍스트에 대한 특성 추출 결과(키워드|중요도)를 중요도순으로 정렬하여 반환합니다.

sub_title

부제목 텍스트에 대한 특성 추출 결과(키워드|중요도)를 중요도순으로 정렬하여 반환합니다.

content

본문 텍스트에 대한 특성 추출 결과(키워드|중요도)를 중요도순으로 정렬하여 반환합니다.

11. 키워드 추출 API

입력한 텍스트의 키워드 추출 결과를 제공합니다.

JSON 형태로 사용 시 텍스트 내용에 쌍따옴표(")가 포함되어있는 경우, 해당 문자 앞에 역슬래시(\) 문자 입력이 필요합니다.

예) 국지성 호우..."피해 대비 철저히" -> 국지성 호우...\\"피해 대비 철저히\"/>

"주의가 필요하다"고 말했다. -> \\"주의가 필요하다\"/>

11.1 요청 URL

<http://tools.kinds.or.kr:8888/keyword>

11.2 요청 형식

```

{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "title": "'태풍 솔릭'에 항공기 대규모 결항...오후 10시까지 총 532편 결항",
    "sub_title": "일자리 창출·사회안전망 확충...내년 사상 최대 '나랏돈' 쏟아",
    "content": "2018아시안게임 남자 축구에서 박항서 감독이 이끄는 베트남 국가대표팀의 돌풍이 심상치 않다."
  }
}

```

11.3 출력 결과 형식

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

```
{
  "result": 0,
  "return_object": {
    "result": {
      "title": "태풍 솔릭 항공기 대규모 결항 10시 532편 결항 ",
      "sub_title": "일자리 창출 사회 안전망 확충 내년 사상 최대 나랏돈 ",
      "content": "2018아시안 게임 남자 축구 박항서 감독 베트남 국가대표팀 돌풍 "
    }
  }
}
```

title

제목 텍스트에 대한 키워드 추출 결과를 반환합니다.

sub_title

부제목 텍스트에 대한 키워드 추출 결과를 반환합니다.

content

본문 텍스트에 대한 키워드 추출 결과를 반환합니다.

12. TopN 키워드 API

다양한 조건을 지정하여 검색된 뉴스결과에서 가장 많이 언급된 키워드를 제공합니다.

12.1 요청 URL

http://tools.kinds.or.kr:8888/topn_keyword

12.2 요청 형식

- 유형 1. 특정조건(입력날짜시간의 24시간 전부터 입력날짜시간까지)의 전체 뉴스 대상


```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "date_hour": "2019080100"
  }
}
```
- 유형 2. 특정조건(기간, 키워드, 분야)의 전체 뉴스 대상


```
{
  "access_key": "d3a80ae9-482c-41d0-9c39-146fe526e09d",
  "argument": {
    "published_at": {
```

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

```

        "from": "2019-08-01",
        "until": "2019-08-05"
    },
    "query": "청와대",
    "category": [
        "001000000",
        "002000000",
        "003000000",
        "004000000",
        "005000000",
        "006000000",
        "007000000",
        "008000000"
    ]
}

```

date_hour

yyyyMMddHH 형식으로 조회할 날짜시간을 지정합니다.
지정한 날짜시간으로부터 24시간전까지 조회합니다.

query

검색 질의어를 지정합니다.

o 검색 연산자

구분	연산자	사용 예
부울린 검색	AND, OR, NOT	사과 AND 배 NOT 과일
구문 검색	" "	"아담과 이브"
연산자 우선 순위	()	(사과 OR 배) AND 과일

published_at

검색 기간을 지정합니다.

• 날짜 형식은 한국 시간을 기준으로 YYYY-MM-DD 형식으로 지정합니다.

예약어	설명
from	조회 시작일을 지정합니다. 지정한 일자를 포함합니다.
until	조회 종료일을 지정합니다. 지정한 일자를 제외합니다.

category

뉴스 통합 분류체계(정치, 경제, 사회, 문화, 국제, 지역, 스포츠, IT 과학) 코드 값을 조건으로 지정합니다.

- 분류체계 코드 값 미입력 시 전체 뉴스에 대해 조회합니다.
- 여러 분류 값을 지정한 경우 OR 조건으로 같이 검색하여 결과를 제공합니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

분류체계	코드 값
정치	001000000
경제	002000000
사회	003000000
문화	004000000
국제	005000000
지역	006000000
스포츠	007000000
IT_과학	008000000

12.3 출력 결과 형식

```
{
  "result": 0,
  "return_object": {
    "result": {
      {
        "rank": "1",
        "name": "일본"
      },
      {
        "rank": "2",
        "name": "한국"
      },
      {
        "rank": "3",
        "name": "미국"
      },
      {
        "rank": "4",
        "name": "서울"
      }
    }
  }
}
```

rank

키워드 언급량 순위입니다.

name

언급된 키워드입니다.

content

본문 텍스트에 대한 키워드 추출 결과를 반환합니다.

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :		작업명 :
작성자 : 김민수	작성일 : 2021-10-28		버전 : 1.45

13. 코드 테이블

13.1 언론사

언론사	코드값
01100101	경향신문
01100201	국민일보
01100301	내일신문
01100401	동아일보
01100501	문화일보
01100611	서울신문
01100701	세계일보
01100801	조선일보
01100901	중앙일보
01101001	한겨레
01101101	한국일보
01200101	경기일보
01200201	경인일보
01300101	강원도민일보
01300201	강원일보
01400201	대전일보
01400351	충도일보
01400401	충부매일
01400501	충부일보
01400551	충북일보
01400601	충청일보
01400701	충청투데이
01500051	경남신문
01500151	경남도민일보
01500301	경상일보
01500401	국제신문
01500501	대구일보
01500601	매일신문
01500701	부산일보
01500801	영남일보
01500901	울산매일
01600201	광주매일신문
01600301	광주일보
01600501	무등일보
01600801	전남일보
01601001	전북도민일보
01601101	전북일보
01700101	제민일보
01700201	한라일보
02100101	매일경제
02100201	머니투데이
02100311	서울경제
02100501	파이낸셜뉴스
02100601	한국경제
02100701	헤럴드경제
02100801	아시아경제
02100851	아주경제
07100501	전자신문

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

07101201	디지털타임스
08100101	KBS
08100201	MBC
08100301	SBS
08100401	YTN
08200101	OBS

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

13.2 뉴스 통합 분류체계

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

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축
		OpenAPI 서비스
단계명 : 운영	활동명 :	작업명 :
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45

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

사용자 매뉴얼		프로젝트 명 : 뉴스 빅데이터 분석 시스템 구축	
		OpenAPI 서비스	
단계명 : 운영	활동명 :	작업명 :	
작성자 : 김민수	작성일 : 2021-10-28	버전 : 1.45	

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 : 김민수	작성일 : 2021-10-28	버전 : 1.45

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>