

NEWS CLOUD

: 네이버 뉴스 수집을 통해 뉴스의 핵심을 빠르게 이해하자

- 중간 발표 -

발표날짜 : 2019.05.02

과목명 : 산학 캡스톤 디자인1

교수명 : 정현숙

팀 : 5조

팀원 : 전연지 양진이 김소원

발표자 : 양진이

1

News Cloud 개요

- News Cloud 소개
- 유사 제품인 빅카인즈 소개
- 빅카인즈와 News Cloud의 차별성

2

News Cloud 구성도

- News Cloud 전체 구성도
- News Cloud 화면 구성도

3

News Cloud 기능

- News Cloud 기능 흐름도
- S/W(Web 개발) 주요 기능
- S/W(Data 분석) 주요 기능
- N2H4 설명

4

News Cloud 개발 환경

- 개발 환경 및 설치 순서

5

News Cloud 수행 내용

- 팀원 소개 및 담당 업무
- News Cloud 수행 일정
- News Cloud 현재 상황

6

News Cloud 참고 문헌

- 참고 서적
- 참고 사이트

NEWS CLOUD

실시간으로 특정 주제의 네이버 뉴스를 분석 및 수집하여
전체 뉴스를 읽지 않고도 Word Cloud 기법으로 시각화 하여
뉴스 내용을 **빠르게 이해**하고 언론사에 따라 시사점이 어떻게 다른지
비교할 수 있다.

Word Cloud란,
문서의 키워드, 개념 등을 직관적으로 파악할 수 있도록
핵심단어를 시각적으로 돋보이게 하는 기법



BIG Kinds

- 한국언론진흥재단이 1990년부터 운영하고 있는 뉴스 데이터베이스 카인즈(KINDS)에 빅데이터 분석을 가미한 서비스
- 키워드 검색을 기반으로 키워드 언급 추이, 뉴스 히스토리, 매체별 언급 빈도 등 다양한 뉴스를 파악할 수 있는 기능을 제공

News Cloud 개요

+

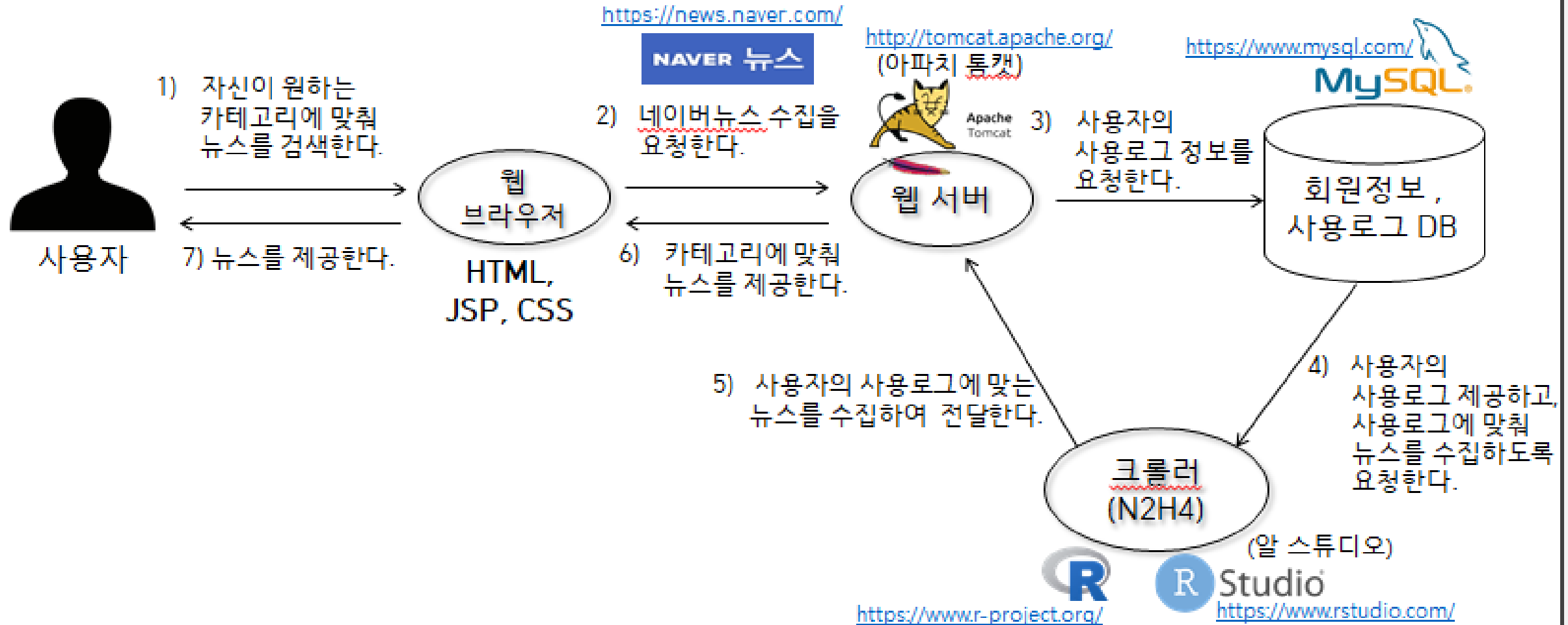
News Cloud와 BIG Kinds의 차별성

비교 구분	BIG Kinds	News Cloud
카테고리 중심 서비스	<u>키워드 중심</u> 의 검색 기반 서비스를 제공한다.	<u>카테고리와 서브 카테고리를 선택</u> 한 후 뉴스 데이터를 실시간으로 가져와서 서비스를 제공한다.
분야 키워드	<u>전체뉴스를 기반</u> 으로 분석된 키워드와 뉴스를 제공한다.	카테고리, 서브 카테고리 설정 후 추가로 두 언론사를 선택함으로써 각각의 <u>언론사가 중점으로 두는 키워드를 쉽게 파악</u> 할 수 있다.
주요 타겟층	<u>(전문적인 직업을 가진)</u> 직장인들	<u>(시간이 부족한 바쁜)</u> 직장인, 학생

News Cloud 구성도

+

News Cloud 전체구성도

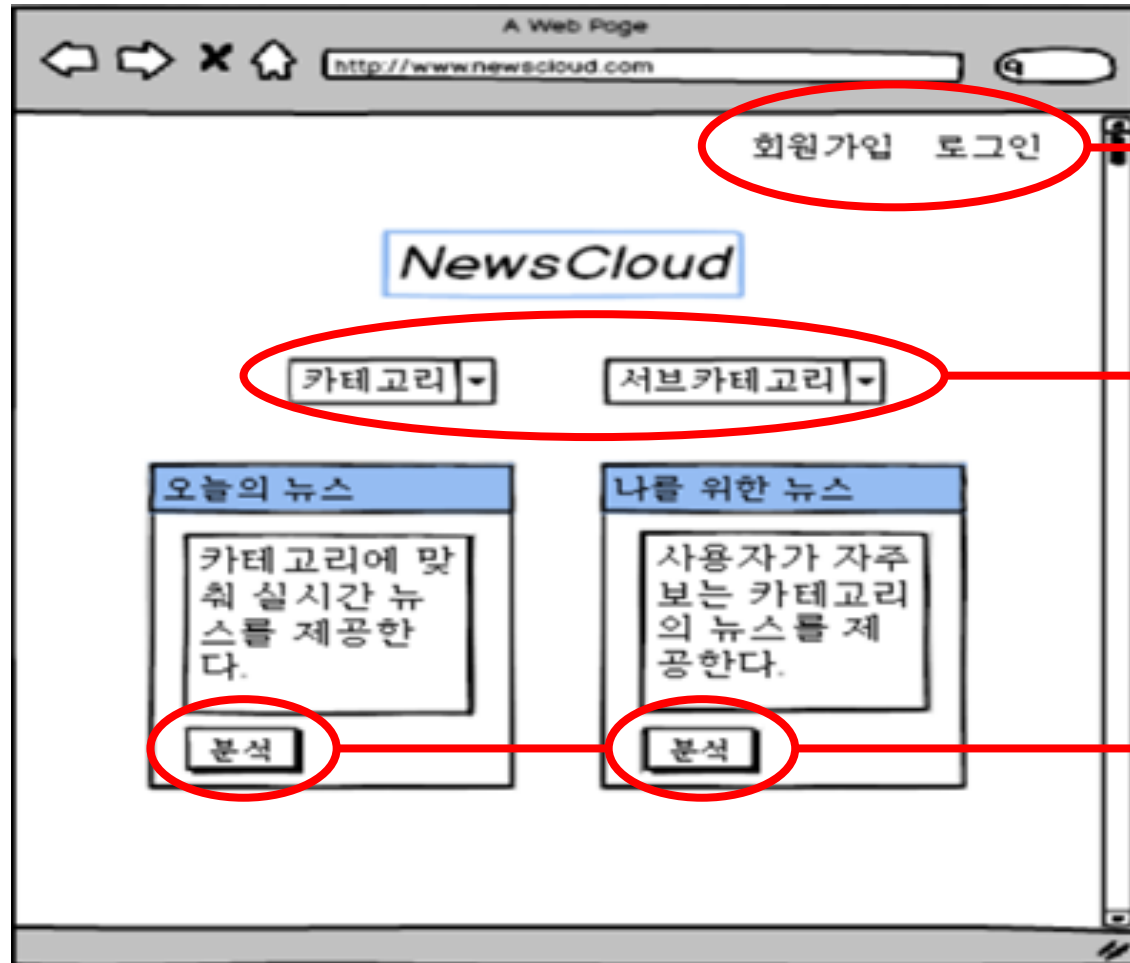


News Cloud 구성도

+

News Cloud 화면 구성도

<메인화면>



맨 위 우측 화면에 회원가입, 로그인 기능을 넣어 사용자가 로그인 기능을 사용할 수 있도록 한다.

사용자는 자신이 원하는 카테고리 and 서브카테고리를 선택한 후

<오늘의 뉴스>기능이나, <나를 위한 뉴스>기능 분석을 선택할 수 있다.

News Cloud 구성도

+

News Cloud 화면 구성도

<오늘의 뉴스>

The screenshot shows a web browser window titled 'A Web Page' with the URL 'http://www.newscloud.com'. The page content includes a header 'NewsCloud' and a section '오늘의 뉴스'. Below this, there are several filter sections: '홈/오늘의 뉴스' (a dropdown menu), '카테고리' (a section header), '카테고리' (a dropdown menu), '서브카테고리' (a dropdown menu), '기간' (a section header), '2000.00.00 ~ 2019.00.00' (a date range input), '언론사1' (a dropdown menu), '언론사2' (a dropdown menu), and a '분석' button. At the bottom, there is a '워드 클라우드' section with two large gray rectangular boxes representing the word cloud.

메인 화면에서 오늘의 뉴스 분석 기능을 선택하면,

사용자가 선택한 카테고리 and 서브카테고리에 관련된

기간, 언론사, 언론사2 목록이 생성된다.
언론사 2개를 선택해 분석 버튼을 누르게 되면

해당 기사들의 중요 핵심 키워드 빈도수에 따른
워드 클라우드 이미지를 볼 수 있게 된다.
이미지를 선택하게 되면
해당되는 실시간 뉴스를 볼 수 있다.

News Cloud 구성도

+

News Cloud 화면 구성도

<나를 위한 뉴스>

NewsCloud

나를 위한 뉴스

종/나를 위한 기사

카테고리

카테고리 서브카테고리

기간

2000.00.00 ~ 2019.00.00

언론사1 언론사2

분석

워드 클라우드

메인 화면에서 나를 위한 뉴스 분석 기능을 선택하면,

사용자가 선택한 카테고리 and 서브카테고리에 관련된

기간, 언론사, 언론사2 목록이 생성된다.
언론사 2개를 선택해 분석 버튼을 누르게 되면

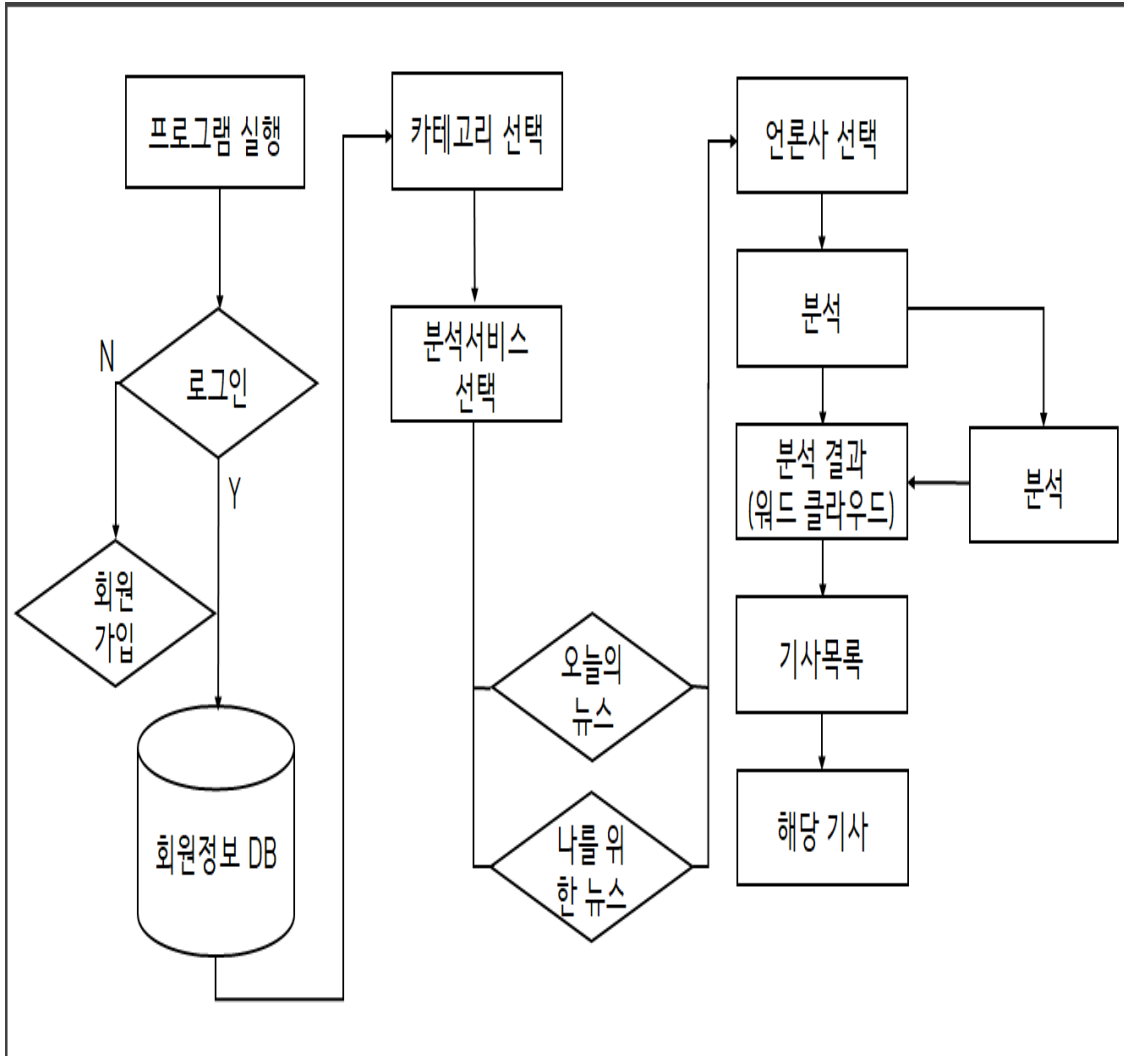
해당 기사들의 중요 핵심 키워드 빈도수에 따른
워드 클라우드 이미지를 볼 수 있게 된다.
이미지를 선택하게 되면
해당되는 개인 맞춤 뉴스를 볼 수 있다.

News Cloud 기능

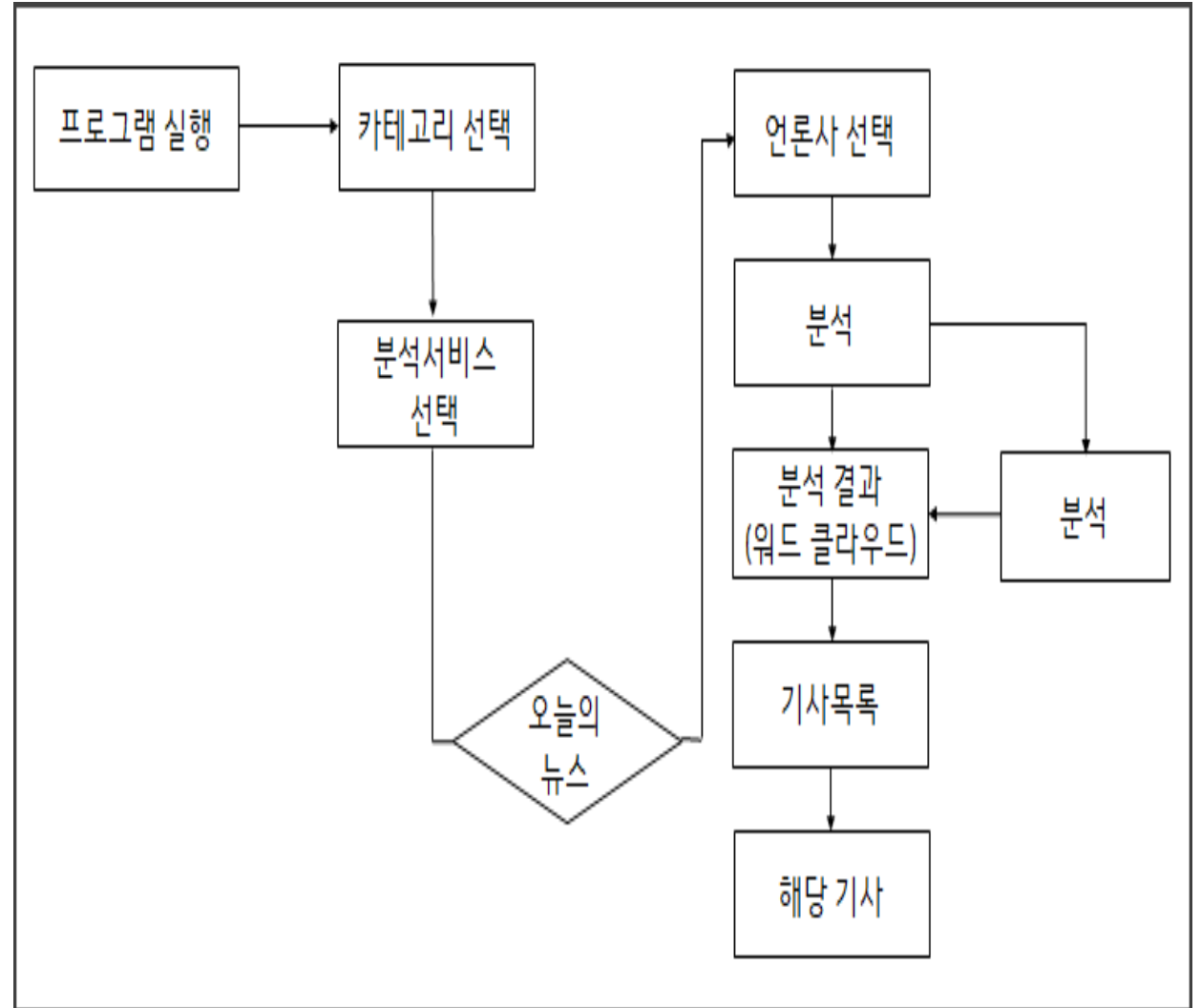
+

News Cloud 기능 흐름도

<로그인 기능을 사용할 때>



<로그인 기능을 사용하지 않을 때>



News Cloud 기능

+

S/W(Web 개발) 주요 기능

기능	설명
로그인	데이터베이스에 저장된 회원 정보를 이용한 로그인 기능 제공
메인 화면	회원가입, 로그인, <오늘의 뉴스>, <나를 위한 뉴스> 분석 기능 제공
오늘의 뉴스	선택한 기간, 언론사의 카테고리에 관한 워드 클라우드 이미지 제공
나를 위한 뉴스	사용자의 사용로그를 기준으로 분석하여 사용자가 선택한 기간, 언론사의 카테고리 에 관한 워드 클라우드 이미지 제공
기사 제공	선택한 언론사의 카테고리에 관해 워드 클라우드 이미지를 선택하면 해당 기사 목록을 보여주고 링크를 걸어 기사를 볼 수 있도록 해줌

News Cloud 기능

+

S/W(Data 분석) 주요 기능

기능	설명
뉴스 수집기	<u>N2H4</u> 를 사용하여 네이버 뉴스에서 제공하는 6개의 카테고리 (정치,경제,사회,생활/문화,세계,IT/과학)과 그에 따른 서브카테고리에 대한 뉴스를 수집
워드 클라우드(빈도분석-시각화)	뉴스에서 사용된 단어들의 빈도수를 파악하여 빈도수에 따라 크기와 색을 다르게 하여 쉽게 뉴스의 키워드를 파악할 수 있도록 이미지 제공

N2H4란, 네이버 뉴스 크롤링을 위한 도구

- R로 네이버 뉴스 데이터를 가져오는 방법
(R의 패키지 중 하나)

크롤링이란,
실시간으로 웹 데이터를
가져오는 방법

News Cloud 기능

+

N2H4 설명

네이버 오픈 API 목록

N 트위터 드림 공유하기 8개

네이버 오픈API 목록 및 안내입니다.

API명	설명	호출제한
검색	네이버 블로그, 이미지, 웹, 뉴스, 백과사전, 책, 카페, 지식iN 등 검색	25,000회/일
지도 (Web, Mobile)	네이버 지도 표시 및 주소 좌표 변환	20만/일
네이버 아이디로 로그인	외부 사이트에서 네이버 아이디로 로그인 가능 구현	없음
네이버 회원 프로필 조회	네이버 회원 이름, 닉네임, 이메일, 성별, 연령대, 프로필 조회	없음
Papago NMT 번역	인공신경망 기반 기계 번역 (영,중)	10,000글자/일
Papago SMT 번역	통계 기반 기계 번역 (영,일,중)	10,000글자/일
Clova Face Recognition	입력된 사진을 입력받아 얼굴윤곽/부위/표정/유명인 닮음도를 리턴	1,000건/일
데이터랩(검색어트렌드)	통합검색어 트렌드 조회	1,000회/일
데이터랩(쇼핑인사이트)	쇼핑인사이트 분야별 트렌드 조회	1,000회/일
캡차(이미지)	자동 입력 방지용 보안 이미지 생성 및 입력값 비교	1,000회/일
캘린더	로그인한 사용자 캘린더에 일정 추가 가능	5,000회/일
카페	특정 네이버 카페 가입하기	50회/일
카페	네이버 회원이 가입한 카페 게시판에 글 쓰기	200회/일
블로그	네이버 회원의 블로그에 글쓰기	200회/일
단축URL	입력된 URL을 me2.do 형태의 짧은 URL로 변환	25,000회/일
공유하기	네이버 블로그, 카페, 풀라 공유하기	없음

네이버에서
제공하는 API 중
뉴스 API는 없음

N2H4 개발자 :



박찬엽

- (현)코빗 재무팀 데이터 담당자
 - 재무DB 구축/관리 및 자동화
- (전)서울도시가스 선행연구팀 연구원
 - 챗봇 엔진 개발 및 서버 구축
- (전)2017년 패스트 캠퍼스 데이터 분석 R 강의
 - 데이터 분석을 위한 중급 R 프로그래밍
- N사 뉴스 크롤러 N2H4, D사 뉴스 크롤러 DNH4 관리자
 - ForkonLP 프로젝트
- FACEBOOK@mrchypark
- GITHUB@mrchypark

한글 분석 및 생성 연구를 위한 기반 데이터를 확보하고 연구하는 모임

getComment : 네이버 뉴스 페이지의 관련 댓글 정보를 가져오는 기능

getContent : 네이버 뉴스 페이지 내에 url, 기사입력시간, 수정시간, 신문사, 제목, 내용 정보를 가져오는 기능

getMainCategory : 네이버 뉴스의 메인 카테고리를 가져오는 기능(정치, 경제, 사회 등등)

getMaxPageNum : 메인 카테고리의, 서브 카테고리 페이지에서 마지막 페이지 수를 가져오는 기능

getNewsTrend : 네이버 뉴스에서 검색시 검색 결과에 나오는 총 검색량을 가져오는 기능

getSubCategory : 네이버 뉴스의 서브 카테고리를 가져오는 기능

getUrlListByCategory : 뉴스 페이지의 제목과 url을 가져오는 기능

getUrlListQuery : 네이버 뉴스가 있는 기사들의 url을 가져오는 기능

News Cloud 기능



N2H4 기능 설명

forkonlp / N2H4

Watch ▾

16

Unstar

129

Fork

63

Code

Issues 5

Pull requests 0

Projects 1

Wiki

Insights

Branch: master ▾

N2H4 / R /

Create new file

Upload files

Find file

History



mrchypark fix datetime parse

Latest commit 23ef3c9 3 days ago

..

.searchNews.R	for poll	2 years ago
N2H4-package.R	update tibble	2 months ago
getComment.R	update tibble	2 months ago
getContent.R	fix datetime parse	3 days ago
getMainCategory.R	update tibble	2 months ago
getMaxPageNum.R	refactor code remove dep pack	a year ago
getNewsTrend.R	update tibble	2 months ago
getQueryUrl.R	fix getQueryUrl	2 months ago
getSubCategory.R	update tibble	2 months ago
getUrlListByCategory.R	update tibble	2 months ago
getUrlListByQuery.R	update tibble	2 months ago
getVideoUrl.R	refactor code remove dep pack	a year ago
setUrls.R	refactor code remove dep pack	a year ago

<https://github.com/forkonlp/N2H4>

N2H4 기능 예시

getContent 명령어 사용

```
> library(N2H4)
> url<-"http://news.naver.com/main/read.nhn?mode=LS2D&mid=shm&sid1=105&sid2=731&oid=031&aid=0000393444"
> getContent(url)
# A tibble: 1 x 6
  url          datetime      edittime      press title      body
  <chr>        <dtm>        <dtm>        <chr>  <chr>    <chr>
1 http://ne~ 2016-11-17 14:39:00 2016-11-17 14:39:00 아이뉴스2~ 네이버-카카~ "[성상훈기자]~
> |
```

아이뉴스24

네이버-카카오, 新 성장동력 키워드는 '콘텐츠'

기사입력 2016.11.17. 오후 2:39 기사원문 스크랩 본문듣기 · 설정



2



댓글

요약봇 ^{Beta}

가



<아이뉴스24>

[성상훈기자] 네이버와 카카오가 '변신'을 선언했다. 네이버는 포털 서비스를 넘어 '콘텐츠 플랫폼'으로 옷을 갈아 입고 있는 가운데 카카오도 카카오 페이지를 선봉에 내세

News Cloud 개발 환경

+

개발환경 및 설치 순서

구분	상세내용
운영체제	Windows 10 64bit
개발도구	②Eclipse 2019-03 ③Tomcat 8 ⑥Rstudio 1.1.463 ④MySQL 8.0.25
사용언어	⑤R-3.5.3 ①Java(jdk-12) HTML JSP CSS

<설치 순서>

- ① Java를 사용하기 위해 **jdk를 설치**
- ② **Eclipse 설치**
- ③ 웹 프로그램은 구동시키기 위해선 네트워크 통신이 필요하게 됨 이것을 해결할 수 있는 **웹 서버 Tomcat을 설치**
- ④ 웹 서버안에서 데이터베이스를 사용하기 위해 **MySQL을 설치**
- ⑤ 데이터 분석을 위한 **R 설치**
- ⑥ R을 사용하여 데이터분석을 할 수도 있지만 R은 명령창 하나만 뜨기에 사용하는데에 조금 불편함 감이 있음.
반면, **Rstudio**는 네 분할(코딩 창, 명령창, 변수에 대한 정보 창, 파일/패키지/뷰어/도움말)을 실행시켜주기에 **사용하는데 더욱 편리하여 설치**

News Cloud 수행 내용

+

팀원소개 및 담당 업무

구분	전연지	양진이	김소원
학번	20164244	20164316	20164226
연락처	010-3746-7697	010-3763-9216	010-7167-5324
이메일	jeonyeonji1028@gmail.com	zcxsad@naver.com	kksw5324@naver.com
깃허브	https://github.com/szduswldz	https://github.com/zcxsad	https://github.com/kswon
역할	팀장	팀원	팀원
담당업무	프로젝트 진행상황 관리, 데이터 분석/통합 개발 담당	데이터 분석 개발 담당	웹 개발 담당

+

[illegible]

News Cloud 수행 내용

+

News Cloud 수행 순서

빅데이터
시용환경 구축

크롤러 설치

빅데이터 수집

크롤러를 사용하여
네이버 뉴스 R로 구현

빅데이터 저장

네이버 뉴스 데이터
JAVA와 연동

빅데이터 분석

R을 이용해 카테고리별,
언론사별 분석

JAVA와 R을 Eclipse에 연
동 구현

빅데이터 처리

워드 클라우드 기법을
이용해 시각화

빅데이터 시각화

News Cloud 수행 내용

+

News Cloud 현재 상황

동영상 설명

+

Q&A

+

감사합니다