

한이음 공모전 2017

개 발 보 고 서

프로젝트명	국문	R을 활용한 뉴스 실시간 분석 웹 개발
	영문	Web Services of Real-time Article Analysis using R
작 품 명	R찬뉴스	
신 청 자	아주대학교 / 임동희	

요 약 본

팀 정보

팀 명	R찬뉴스			
팀 원	이 름	소 속	부서/학과	직위/학년
멘 토	장경애	조은소프트		
지도교수				
멘티 1(팀장)	임동희	아주대학교	산업공학과	4학년
멘티 2	이재호	경기대학교	응용정보통계학과	4학년
멘티 3	정승완	전남대학교	소프트웨어공학과	3학년
멘티 4	김도현	전남대학교	소프트웨어공학과	3학년
멘티 5	신현중	차의과대학교	식품생명공학과	2학년



작품 정보														
프로젝트명	국문	R을 활용한 뉴스 실시간 분석 웹 개발												
	영문	Web Services of Real-time Article Analysis using R												
작품명	R찬뉴스													
작품 소개	<p>실시간으로 특정 주제의 뉴스를 분석하여 전체 뉴스를 읽지 않고도 시각화하여 그 주요 키워드를 파악할 수 있도록 하는 R 기반의 뉴스수집과, 분석을 웹으로 제공하는 실시간 뉴스 분석 웹서비스입니다.</p> <ul style="list-style-type: none">○ 뉴스 수집 (웹 크롤링): 네이버뉴스에서 제공하는 6개의 카테고리의 뉴스 수집○ 자연어처리: R의 자연어 처리 패키지와 자체개발 자연어 정제함수 적용○ 워드클라우드 시각화: 뉴스의 사용단어를 기반으로 한 빈도행렬 워드클라우드 분석○ 네트워크분석 시각화: 뉴스에 동시에 출현하는 단어의 연관 네트워크분석													
작품 구성도	<pre>graph TD subgraph Frontend WUI[Web UI: JSP, HTML5, CSS] WSV[Web Server: Java] end subgraph Backend subgraph Data R[R] RStudio[RStudio] RMarkdown[R Markdown] end subgraph Algorithm TextMining[Text Mining] WordCloud[Word Cloud] NetworkAnalysis[Network Analysis] end end subgraph Infra CS[client-server Internet] AT[Aeroche Tancet] end WUI --> Data WSV --> Data Data --> Algorithm Algorithm --> CS CS --> AT</pre>													
작품의 개발배경 및 필요성	<ul style="list-style-type: none">○ 빅데이터시대 대량의 데이터 생산에 따른 자연어 분석, 머신러닝, 인공지능연구 필요○ 언론사의 개수가 70여개가 넘는 다양한 뉴스에 대한 이해 어려움, 시간절감 필요○ 데이터 분석결과 서비스의 제약에 따른 기술적인 해결필요 (R, JAVA)													
작품의 특징점	<ul style="list-style-type: none">○ R을 사용하여 뉴스데이터 수집과 R과 JAVA 언어 연동을 통한 웹서비스 제공○ 네이버에서 제공되는 주제별로 다양한 언론사의 뉴스를 실시간으로 수집○ 텍스트마이닝 분석을 통해 키워드를 이해하기 쉽도록 시각화 기법 제공													
작품 기능	<table><tr><td>로그인/메인</td><td>회원가입/로그인 기능, 로그인을 통하여 메인 화면으로 이동</td></tr><tr><td>오늘의기사</td><td>카탈로그 기반의 오늘의 주요 기사에 대한 분석결과 제공</td></tr><tr><td>워드클라우드</td><td>특정 뉴스에 대한 빈도수가 높은 단어의 시각화 기능 제공</td></tr><tr><td>키워드연관분석</td><td>특정 키워드의 네트워크 연관분석을 통해 분석결과 시각화 기능 제공</td></tr><tr><td>개인 맞춤기사</td><td>개인의 선택 기사를 수집하여 개인의 성향에 맞춤 기사 제공</td></tr><tr><td>통계 분석</td><td>기사, 언론사, 주요 키워드 등 뉴스 분석결과에 대한 통계정보 제공</td></tr></table>		로그인/메인	회원가입/로그인 기능, 로그인을 통하여 메인 화면으로 이동	오늘의기사	카탈로그 기반의 오늘의 주요 기사에 대한 분석결과 제공	워드클라우드	특정 뉴스에 대한 빈도수가 높은 단어의 시각화 기능 제공	키워드연관분석	특정 키워드의 네트워크 연관분석을 통해 분석결과 시각화 기능 제공	개인 맞춤기사	개인의 선택 기사를 수집하여 개인의 성향에 맞춤 기사 제공	통계 분석	기사, 언론사, 주요 키워드 등 뉴스 분석결과에 대한 통계정보 제공
로그인/메인	회원가입/로그인 기능, 로그인을 통하여 메인 화면으로 이동													
오늘의기사	카탈로그 기반의 오늘의 주요 기사에 대한 분석결과 제공													
워드클라우드	특정 뉴스에 대한 빈도수가 높은 단어의 시각화 기능 제공													
키워드연관분석	특정 키워드의 네트워크 연관분석을 통해 분석결과 시각화 기능 제공													
개인 맞춤기사	개인의 선택 기사를 수집하여 개인의 성향에 맞춤 기사 제공													
통계 분석	기사, 언론사, 주요 키워드 등 뉴스 분석결과에 대한 통계정보 제공													
작품의 기대효과 및 활용분야	<ul style="list-style-type: none">○ 뉴스를 읽을 시간이 없는 직장인, 학생 / 뉴스를 생산하는 언론사에 빠른 정보제공○ 철자를 정확히 읽지 못하는 난독증 환자를 위한 뉴스 요약서비스 제공○ 수집된 데이터와 자연어처리를 활용한 머신러닝 서비스 개발에 활용													

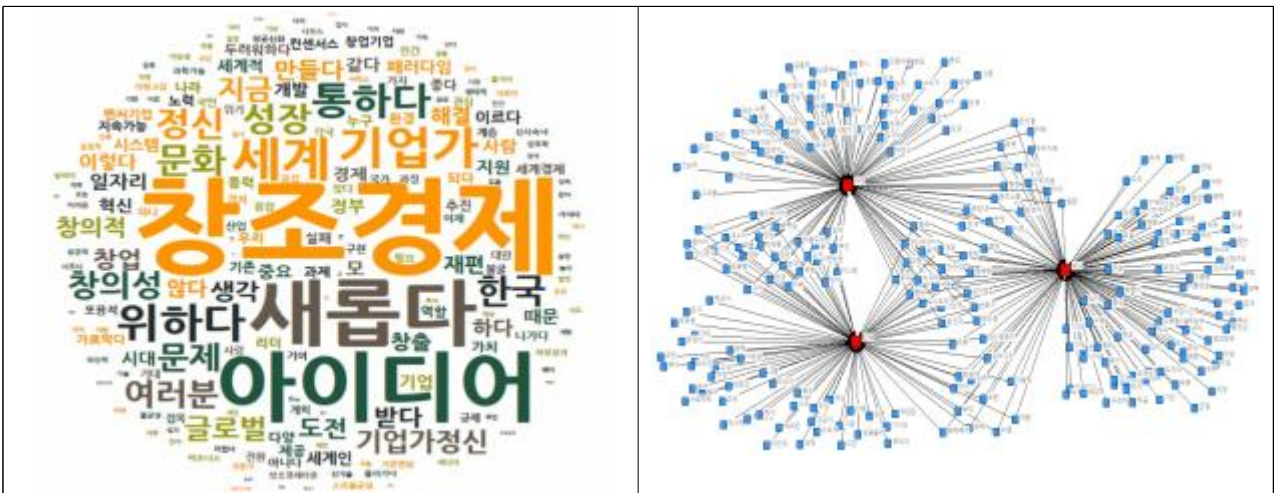
본 문

I. 작품 개요

1. 작품 소개

실시간으로 특정 주제의 뉴스를 분석하여 전체 뉴스를 읽지 않고도 시각화하여 그 주요 키워드를 파악할 수 있도록 하는 웹 크롤링과 R 기반의 분석을 웹으로 제공하는 실시간 뉴스 분석 웹서비스

- 뉴스 수집 (웹 크롤링): 네이버뉴스에서 제공하는 6개의 카테고리의 뉴스 수집
- 자연어처리: R의 자연어 처리 패키지와 자체개발 자연어 정제함수 적용
- 워드클라우드 시각화: 뉴스의 사용단어를 기반으로 한 빈도행렬 워드클라우드 분석
- 네트워크분석 시각화: 뉴스에 동시에 출현하는 단어의 연관 네트워크분석



2. 작품의 개발 배경 및 필요성

2.1 작품의 개발 배경

- 자연어 분석, 머신러닝, 인공지능연구 필요
 - 빅데이터 시대의 도래에 따라 대량의 데이터들을 수집 가공, 분석할 수 있는 가능성 제시, 활용 방안에 대한 연구는 요즘 시대의 큰 이슈 중 하나임
 - 인터넷 대부분의 데이터가 비정형 데이터로 구성, 이에 따라 사람의 언어를 이해, 분석하고자 하는 연구들 존재 함
 - 분석 결과를 인공지능 기술과 연결 시켜 언어 번역, 질문 응답과 같은 서비스에

접목시키고자 하는 움직임이 발생되고 있음

○ 데이터 분석 결과의 서비스화 제약

- R은 데이터 분석을 위해서는 효과적 이지만 웹이나 어플리케이션 상에서 실시간으로 구동시키고 그 결과물을 주고받는 데에는 어려움이 존재
- R의 활용성을 극대화하기 위해서는 JAVA 등의 프로그래밍 언어와 원활한 연동 필요

2.2 작품의 개발 필요성

○ 다양한 언론사 및 뉴스 이해를 위한 시간절감 필요

- 과거에는 다양한 뉴스를 접하기 위해서는 추가적인 비용 지불, 현재에는 포털사이트를 통하여 대부분의 언론매체의 뉴스 접근 가능
- 사용자들은 시간 부족, 편의성 추구 등을 이유로 뉴스의 기사들을 모두 읽는 것을 기피함
- 기사의 내용을 빠르게 이해하고, 언론 매체에 따라서 시사점이 어떻게 다른지 손쉽게 비교해보고자 하는 필요성 대두

3. 작품의 특징 및 장점

3.1 유사 제품 분석

해외 유사서비스로는 다음과 같은 서비스가 존재한다.

- 섬리(Summlly) 기사를 약 200자 정도로 축약해서 제공
- 펄스(Pulse): 개인화 뉴스 서비스
- 서카(Circa): 하나의 토픽에 관한 다양한 목소리를 보여주는 앱
- 프리즈매틱(Prismatic): 페이스북과 트위터 글을 분석해 관련 뉴스를 찾아 보여줌

국내 유사 서비스로는 zziz.net(개인), 똑똑한 뉴스 필터링 앱(개인), 빅카인즈가 존재하며, 현재 실제로 서비스를 제공하는 것은 빅카인즈만 존재

3.1.1 기능적 분석

○ 빅카인즈

한국언론진흥재단이 1990년부터 운영하고 있는 뉴스 데이터베이스 카인즈(KINDS)에 빅데이터 분석을 가미한 서비스로 키워드 검색을 기반으로 하며, 키워드 언급 추이, 뉴스 히스토리, 매체별



언급 빈도 등 다양하고 입체적으로 뉴스를 파악할 수 있는 기능을 제공하고 있다.

빅카인즈의 모태가 되는 '뉴스소스 베타' 시스템에 자연어 처리와 의미 연결망 분석 기능이 강화된 것으로 분석 시스템은 솔트룩스라는 자연어 처리 전문 업체가 제작 일반용/ 전문가용 서비스를 나눠 제공하고 있으며 일반용의 주요서비스는 다음과 같다.

- 오늘 이슈 - 매일 수집된 뉴스 중 선택한 지역에서 가장 많이 다뤄진 이슈 분석
- 오늘인물 - 매일 수집된 뉴스 내 정보원과 인용문을 분석하여 인용이 많이 된 정보원 보여주는 기능
- 말말말 - 특정 이슈에 대한 주요 인물의 발언 내용 제공
- 분야키워드 - 매일 수집된 뉴스 내 주요 키워드 분석
- 네트워크 - 오늘이슈, 인물, 분야키워드에서 선택한 항목과 관계된 개체를 네트워크로 보여주는 기능

이외에도 뉴스심층분석, 뉴스분석사례, 뉴스라이브러리, 이슈모음 등의 서비스를 제공 또한 로그인을 하여, 관심키워드 관리, 관심 뉴스관리/스크랩, 공유사전, 나의사전 기능으로 검색식을 관리할 수 있다.

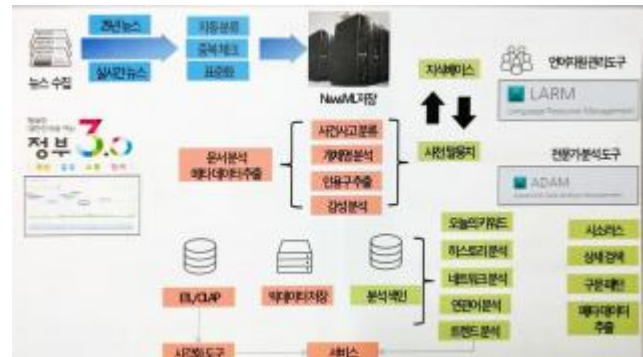
3.1.2 기술적 분석

빅카인즈 분석의 주요 기술은 다음과 같다.

○ 뉴스 클러스터링

분석 대상이 되는 뉴스 1건씩 형태소를 분석하여 명사 상당 어구를 추출
명사 상당 어구에서 토픽랭크(TopicRank) 알고리즘을 사용하여 뉴스 내 중요

키워드를 선정한 후 키워드를 벡터(Vector)로 구성 문서 벡터 간 코사인 거리(Cosine Distance) 계산하여 문서 간의 유사도(Similarity)가 임계치(0.2) 이상인 뉴스를 그룹화 모든 문서가 클러스터링 될 때까지 문서 간 그룹화 반복 수행



○ 개체명분석

뉴스 텍스트 내에서 인물, 기관, 장소 키워드를 SSVM(Structured Support Vector Machine) 알고리즘을 사용하여 추출 개체명은 고유명사인 경우가 많으며, 항상 새롭게 만들어지고 때로는 같은 단어라도 사용되는 문장에 따라 상이하하며, 고유명사 하나가 하나의 개체명을 이룰 수도 있지만 2개 이상의 고유명사나 일반명사와 결합하여 복합명사 형식의 개체명을 이룰 수 있음

○ 텍사노미

텍사노미는 맥락 속에서 배치될 수 있는 단어, 텍사노미를 연산자(AND, OR)로 연결해서 검색에 활용

○ 시소러스

원하는 키워드에 대한 뉴스를 효율적으로 검색하고 확인할 수 있음
시소러스 표제어를 선택할 경우, 해당 표제어의 동의어, 관련어 등이 자동으로 검색식에 입력

○ 레인보우

빅카인즈 자체 시각화 도구, D3기반으로 레인보우로 만든 인터랙티브차트는 기사에 임베드해서 활용할 수 있다.

3.2 작품의 유사제품과의 차별성

3.2.1 기능적 차별성

비교 구분	빅카인즈	R찬뉴스
카테고리 중심 서비스	키워드 중심의 검색서비스를 제공	6개의 카테고리를 선택하고, 세부카테고리를 까지 설정한 후 데이터를 실시간으로 끌어와서 분석이 진행됨
분야 키워드	분야키워드가 전체뉴스를 기반으로 분석된 키워드와 네트워크를 제공	카테고리를 먼저 설정하고, 수집된 데이터에서 비교할 언론사를 추가로 선택하여 분석이 진행됨, 따라서 전체적인 맥락보다 두 언론사의 주요 키워드를 비교·대조 하여 파악할 수 있으며 해당 언론사가 중점적으로 두는 키워드를 파악할 수 있음
개인 맞춤뉴스	관심뉴스, 스크랩뉴스, 최근 본 뉴스, 나의 뉴스분석 등의 개인 로그 관련된 서비스를 제공	사용로그를 기반으로 개인 맞춤 뉴스를 제공
주요 타겟층	언론학자, 정치학자, 경영학자, 변호사, 컨설턴트, 회계사 등 전문가를 타겟으로 하여 기자들에겐 데이터 저널리즘의 재료, 변호사의 변론문, 컨설턴트의 보고서 등에 활용될 것으로 타겟	바쁜 학생, 직장인, 난독증 환자들이 주요 타겟

3.2.2 기술적 차별성

○ 세부카테고리

빅카인즈의 분야키워드 서비스가 정치, 경제, 사회, 문화, 국제, 지역, 스포츠, IT과학으로 분류하여 전체 뉴스를 가지고 분석을 제공하는 반면, R찬뉴스는 정치, 경제, 사회, 생활·문화, 세계, IT·과학 6가지 카테고리 외 적어도 3개 이상의 세부카테고리를 선택하고, 언론사까지 선택하여 분석을 진행

○ 실시간성

빅카인즈의 분석에 사용되는 데이터는 전일 오전 5시부터 금일 오후 8시까지 수집된 뉴스데이터를 기반으로 하지만, R찬 뉴스는 해당 일을 기준으로 실시간으로 데이터를 수집하여 분석을 진행

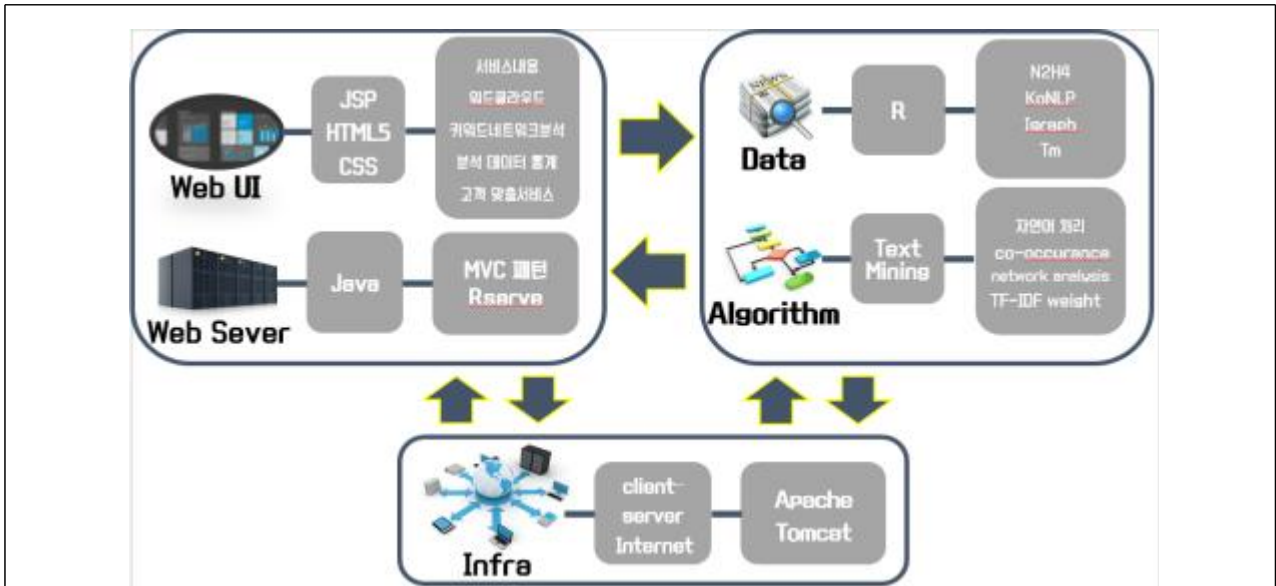
○ 오픈소스 언어 R And Java 기반

빅카인즈가 서울대학교 대학원의 뉴스소스 베타 시스템과 데이터 전문기업 솔트룩스가 제작한 분석 시스템을 결합한 것과 다르게 ‘R찬뉴스’는 오픈소스 기반의 R언어와 Java를 연동시켜 구동됨

II. 작품 내용

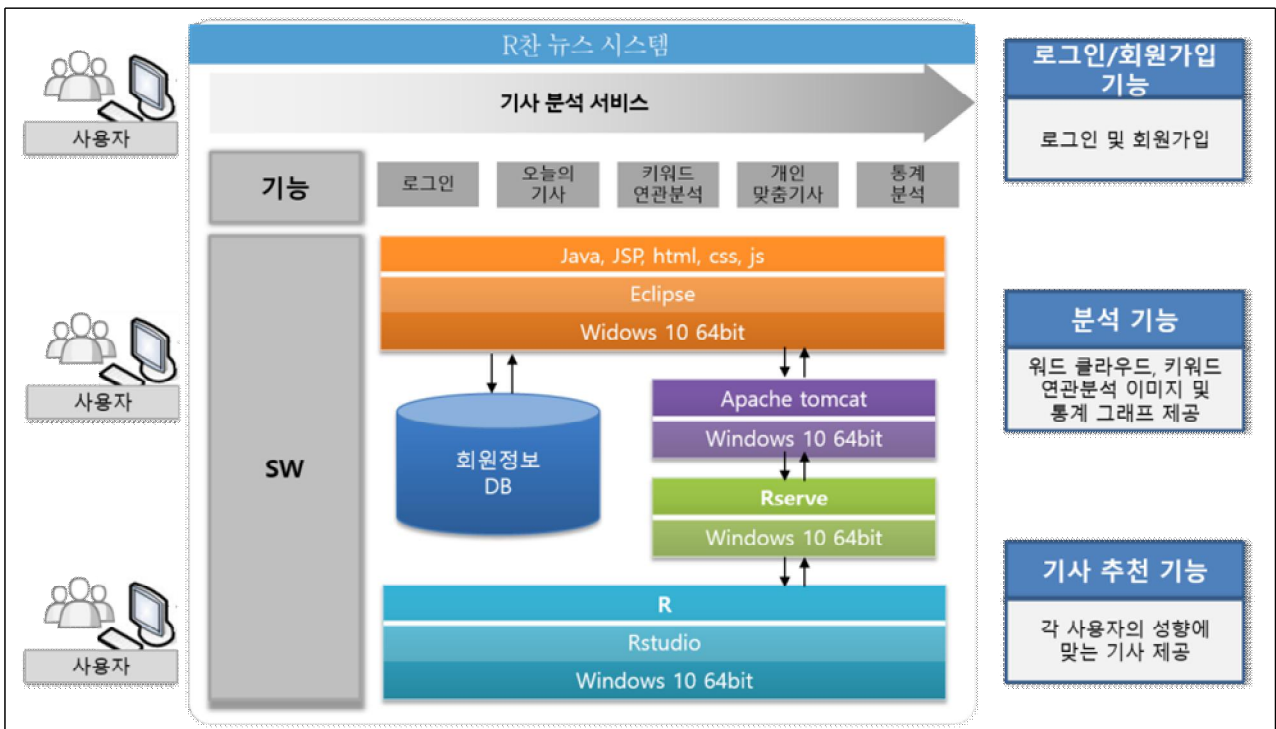
1. 작품 구성도

○ 전체 구성도



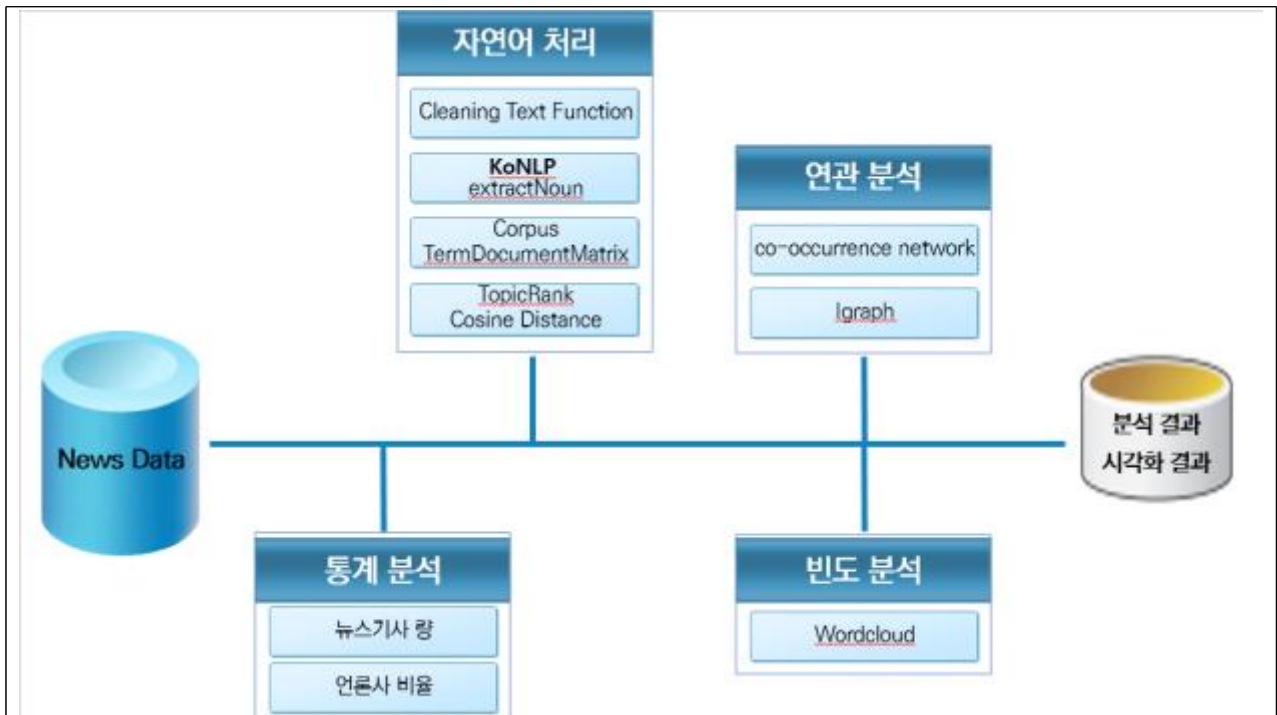
- 작품은 Web UI, Web Server, Data분석, 분석 알고리즘, 인프라영역으로 구성

○ S/W 구성도



- 소프트웨어 영역은 Window환경 기반에서 R/R Studio, Rserver연동 모듈, Java/JSP/html/css/JSP로 구성

○ Data 분석 구성도



- Data 영역은 자연어처리, 연관분석, 통계분석, 빈도분석을 통한 시각화로 구성



2. 작품 기능

2-1. 전체 기능 목록

구분	기능	설명	현재진척도(%)
S/W (Web 개발)	로그인/회원가입	데이터베이스에 저장된 회원 정보를 이용한 로그인 기능 제공	100%
	메인	오늘의 기사, 키워드 연관분석, 통계 분석, 개인 맞춤 기사 등의 연결 기능 제공	100%
	오늘의 기사	선택한 언론사의 카테고리에 관한 워드클라우드 이미지 제공	100%
	키워드 연관분석	선택한 언론사의 카테고리에 관한 키워드 연관분석 제공	100%
	개인 맞춤기사	해당 기사에서 가장 높은 빈도수의 키워드를 기준으로 다시 검색하여 기사 제공	100%

	통계 분석	각 언론사의 카테고리에 관한 기사 건수 그래프 제공	100%
S/W (Data 분석)	뉴스 수집기	네이버뉴스에서 제공하는 6개의 정치/ 경제/ 사회/ 생활·문화/ 세계/ IT·과학과 그 세부 카테고리의 뉴스 수집	100%
	자연어 처리	R의 자연어 처리 패키지와 자체제작한 자연어 정제함수, TF-IDF 방식의 가중치를 통하여 주요 단어 추출	100%
	빈도분석-시각화 (Word Cloud)	뉴스에서 사용된 단어의 빈도수를 파악하여 행렬을 만들고 빈도수에 따라 크기와 색을 달리 하여 쉽게 뉴스의 키워드를 파악할 수 있도록 이미지 제공	100%
	연관성분석-시각화 (Word Network)	뉴스 간에 등장하는 단어의 동시성을 파악하여 특정 단어가 들어간 뉴스에서 어떠한 단어가 동시에 출현하며, 연관성이 있는지 네트워크 분석결과 제공	100%

2-2. S/W(Web 개발) 주요 기능

기능	설명	작품실물사진
로그인	데이터베이스에 저장된 회원 정보를 이용한 로그인 기능 제공	
메인	오늘의 기사, 키워드 연관분석, 통계 분석, 개인 맞춤 기사 등의 연결 기능 제공	 카테고리 선택->분석기능 선택
오늘의 기사	선택한 언론사의 카테고리에 관한 워드 클라우드 이미지 제공	언론사 선택->분석->분석 결과 이미지

		
키워드 연관분석	선택한 언론사의 카테고리에 관한 키워드 연관분석 제공	 <p>언론사 선택->분석->분석 결과 이미지</p>
개인 추천기사	사용자의 로그 중 최빈값(카테고리와 언론사)을 기준으로 분석하여 워드 클라우드 이미지 제공	
통계 분석	각 언론사의 카테고리에 관한 기사 건수 그래프 제공	 <p>분석 결과 그래프</p>
기사 제공	선택한 언론사의 카테고리에 관한 기	

3. 주요 적용 기술

3.1 Web 개발 기술

○ CSS(cascading style sheets)

- 기술의 정의

기존의 HTML은 웹 문서를 다양하게 설계하고 수시로 변경하는데 많은 제약이 따르는데, 이를 보완하기 위해 만들어진 것이 스타일 시트이고 스타일 시트의 표준안이 바로 CSS기술 임.

- 기술의 특징

추가적인 sw나 플로그인 필요 없이 html기능의 확장가능

설정하기 쉬우며 단순한 규칙으로 이루어진 텍스트

흐름이 같은 문서양식으로 전체 구성 가능

디자인과 웹구조의 완벽한 분리로 하나의 html문서에 다양한 디자인 적용 가능

- 본 작품에서 화면의 구성 전 영역에서 사용함.

○ Jsp

- 기술의 정의

Java Server Page의 준말로, HTML내에 자바 코드를 삽입하여 웹 서버에서 동적으로 웹 페이지를 생성하여 웹 브라우저에 돌려주는 언어임.

- 기술의 특징

수행 성능과 확장성

다중 요청에 의해 멀티스레드로 동작

여러 스레드간의 리소스 공유가 쉬우므로 성능향상

java언어의 장점을 그대로 수용할 수 있고 web서버 독립적 사용가능

○ javabeans 컴포넌트

- 기술의 정의

주요 운영 체제 플랫폼의 네트워크 환경에서 재사용 가능 애플리케이션 또는 프로그램 컴포넌트들을 시각적인 환경 아래 개발할 수 있게 해주는 구조.

- 기술의 특징

jsp는 java언어를 활용하여 동적 컨텐츠 구현 가능

어떠한 환경에서든 쉽게 적용할 수 있는 javabeans컴포넌트를 활용하여

컴포넌트의 내부구조를 모르고도 쉽게 필요한 기능을 구현 가능

○ HTML

- 기술의 정의

인터넷 서비스의 하나인 월드 와이드 웹을 통해 볼 수 있는 문서를 만들 때

하이퍼텍스트를 작성하기 위해 개발되어 대부분의 웹 페이지 작성에 사용

- 기술의 특징

플래시나 ActiveX 없이 HTML5를 이용하여 인터랙티브한 사이트 제작 가능

HTML5를 이용해 모바일용 어플리케이션으로 제작 가능

- 본 작품의 웹페이지 전반적 제작에 사용함

- Rserve

- 기술의 정의

R과 JSP를 연동해주는 TCP/IP 통신 소켓 서버

- 기술의 특징

R과 Jsp를 연동해주어서 Jsp 상에서 R을 쉽게 사용할 수 있게 함.

다양한 개발 환경에서 초기화 등의 설정 없이 통신이 가능.

각각의 연결은 클라이언트에 대해 독립적이어서 쓰레드(Thread)문제가 발생하지 않음.

- 웹 실행 전 Rserve와 R의 사용할 패키지를 설치 및 실행하고 미리 작성해놓은 R스크립트를 웹에서 이용하여 각 웹페이지별로 필요로 하는 분석 기능을 사용.

3.2 Data 분석 기술

○ 웹 크롤링

- 기술의 정의

조직적, 자동화된 방법으로 월드 와이드 웹을
탐색하는 컴퓨터 프로그램

- 기술의 특징

빠르게 웹 페이지의 내용을 수집하여, 차후 활용을 목적으로함

- 기술의 활용

R의 N2H4 패키지를 활용하여

네이버뉴스에서 서비스되는 카테고리별
기준으로 하루 동안의 기사들을 수집
수집된 기사들은 URL, 날짜, 언론사,
기사명, 기사본문의 형태로 data.frame
및 csv 파일로 저장, 활용 됨

 N_2H_4

license	MIT License	build	passing	coverage	50%	CRAN	not published
---------	-------------	-------	---------	----------	-----	------	---------------

네이버 뉴스 크롤링을 위한 도구

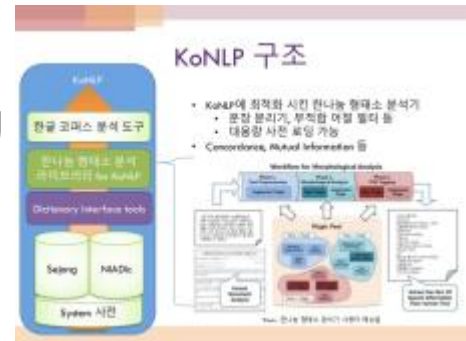
[illegible]

○ 자연어 처리

- 정의 : 인간이 발화하는 언어 현상을 기계적으로 분석해서 컴퓨터가 이해할 수 있는 형태로 만드는 자연 언어 이해, 혹은 그러한 형태를 다시 인간이 이해할 수 있는

언어로 표현하는 제반 기술을 의미

- KoNLP : Korean NLP Package 의 약자로, R 기반에서 한글의 자연어 처리를 위한 패키지



KoNLP 특징
R 패키지로는 유일한 한글 텍스트 전처리 도구
라이브러리로 제공하는 기존 형태소 분석기와는 다르게 한글 텍스트 분석에 자주 쓰이는 기능 위주 인터페이스 제공
한글 분석에 대해 특별한 배경 지식 없이 간단한 함수로 사용가능
형태소 분석 및 POS Tagging 결과를 튜닝할 수 있는 사용자 사전 입력 인터페이스 제공
구 분야별 사전 적용 가능, 연구자들의 활용도 높음
현재 시스템사전(28만) , 세종사전(8만), NIADic(93만) 사전을 제공

- KoNLP 단어-문서행렬

본 작품에서는 KoNLP를 활용하여 단어-문서행렬을 구성

문서를 분류하려면 문서를 기술하는 표현을 문서로부터 추출하고, 이로부터 분류를 예측하는 알고리즘을 만들어야 한다. 문서의 행렬 표현 방식은 이러한 목적으로 가장 많이 사용되는 기법

- TF-IDF 가중치

TF는 Term Frequency로 단어 빈도를 의미하며 IDF는 Inverse Document Frequency의 약어로, 단어가 출현한 문서 수의 역수를 의미

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

로그를 활용하여 TF를 정규화하고 IDF를 활용하여 중요한 단어의 가중치를 주는 방법은 전체 자연어를 정제하지 않아도 중요한 단어를 선별하는데 효과적

tf_{ij} = number of occurrences of i in j
 df_i = number of documents containing i
 N = total number of documents

o 텍스트마이닝

- 기술의 정의

비/반정형 텍스트 데이터에서 자연어처리 기술에 기반하여 유용한 정보를 추출, 가공하는 것을 목적으로 하는 기술

- 기술의 특징

컴퓨터가 이해할 수 있는 TermDocumentMatrix 형태로 텍스트를 가공,

- 기술의 활용

단어의 빈도수를 파악하고, 단어 행렬 구성, 워드 클라우드, 워드 네트워크 분석을

위한 전처리 작업으로 사용

○ 계층적 클러스터링

- 기술의 정의

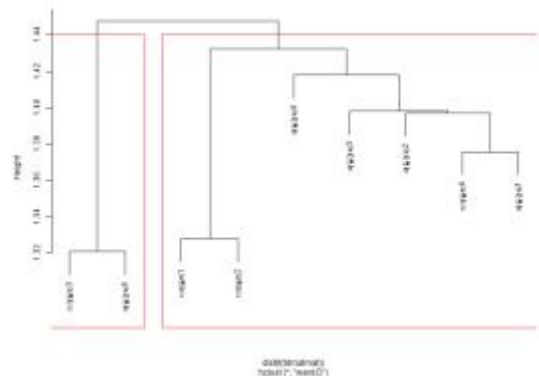
객체들의 유사성을 측정하여 유사성이 높은 집단을 분류하고 군집에 속한 객체들의 유사성과 서로 다른 군집에 속한 객체간의 상이성을 규명하는 방법을 클러스터링이라 함

- 기술의 특징

서로 다른 n개의 군집으로 시작해 유사한 군집끼리 합쳐 나가는 방식

최종적으로 1개의 군집으로 모든 집단이 만남

- 본 작품에서 단어의 빈도수에 따라 유사한 뉴스 문서들을 군집화 하기 위하여 사용함



4. 작품 개발 환경

구분		상세내용
S/W 개발환경	OS	Windows 10
	개발환경(IDE)	Rstudio ver 3.3.3 / eclipse / Tomcat
	개발도구	Client / Server computer
	개발언어	R / Jsp/ Java / HTML5 / CSS / JS
	기타사항	-
프로젝트 관리환경	형상관리	구글 팀드라이브, Github
	이슈관리	메신저, 구글 팀드라이브
	의사소통관리	메신저, 구글 팀드라이브
	기타사항	-

III. 프로젝트 수행 내용

1. 멘티(참여학생) 업무분장

번호	이름	대학	학과	학년	역할	담당업무
1	임동희	아주대학교	산업공학과	4	팀장	- 프로젝트 진행상황 관리, 데이터 분석/통합개발 담당
2	이재호	경기대학교	응용정보통계학과	4	팀원	- 데이터 분석 개발 담당, 웹 개발/통합개발 지원
3	정승완	전남대학교	소프트웨어공학과	3	팀원	- 웹 개발 담당, 데이터 분석/통합개발 지원
4	김도현	전남대학교	소프트웨어공학과	3	팀원	- 웹 개발 담당, 데이터 분석/통합개발 지원
5	신현중	차의과대학교	식품생명공학과	2	팀원	- 데이터 분석 지원, 공모전 준비 담당

2. 프로젝트 수행일정

프로젝트 기간 (한이음 사이트 기준)		2017.00.00. ~ 2017.00.00.											
구분	추진내용	프로젝트 기간											
		1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
계획	프로젝트 계획수립, 필요능력 학습												
분석	분석요구사항 도출, 분석												
설계	데이터 분석, 기능개발 설계												
	통합개발 설계												
개발	데이터 분석 개발												
	기능 개발												
	통합 개발												
테스트	정상 구현 테스트 및 디버깅												
공모전	한이음 공모전 출품 준비												
종료	프로젝트 종료												

3. 프로젝트 추진 과정에서의 문제점 및 해결방안

3-1. 프로젝트 관리 측면

○ 프로젝트는 크게 디자인 부분과 데이터 분석 부분으로 나뉘어져 있었습니다. 팀원 별로 자신의 역량을 최대한 발휘 할 수 있도록 자신의 영역에서 맡은 역할에 집중하였고, 디자인 파트와 데이터 분석 파트간에 서로 구현 방향과 구현 가능성에

대한 의견이 다르고 작업해온 환경 및 사용하는 언어에 관한 이해도가 달랐기에 의사소통에 어려움을 겪기도 하였지만, 매 달 멘토님의 주관 하에 오프라인 모임을 가지고, 이 외에는 온라인 상에서 가능한 개발 방향에 관하여 지속적인 의견 교환을 통해 서로 개발 방향을 합의 할 수 있었습니다.

3-2. 작품 개발 측면

○ Web 개발

○ Web 기능 :

사용자가 사용하기에 편리한 방향으로 설계를 하려고 했는데, 직접 디자인 하는 것보다 부트스트랩을 사용하는 것이 더 효율적이라고 판단을 해서 부트스트랩을 사용하여 디자인을 구현하였습니다. 그러나 기존의 부트스트랩 템플릿에 없는 디자인과 동적 기능을 추가하는 것이 생각보다 어려워서 인터넷을 참고하여서 저희가 원하는 방향으로 구현하였습니다. 그리고 페이지 간에 데이터 전달을 할 때, Jsp와 Java Script의 실행 시간이 달라서 Java Script의 변수와 Jsp의 변수를 혼용하여 사용하는 것이 저희가 생각했던 대로 되지 않았습니다. 그래서 그 부분을 고려하여 다시 코드를 작성하여 해결하였습니다.

○ R과 Jsp 연동 :

R언어와 Jsp를 연동하는 과정에서 Rserve와 라이브러리가 필요하여서 설치하였는데, 라이브러리를 사용하는 과정에서 특정 문자(한글, 특수문자)가 들어간 경우 호환이 되지 않아서 Rserve Error가 발생하였습니다. 그래서 R과 Jsp를 연결 시켜주는 Rserve 명령어에 인코딩(UTF-8) 기능을 추가시켜 실행하여 그 문제를 해결하였습니다. 그리고 R의 데이터 형식과 Java 데이터 형식이 달라서 그 두 개를 매칭시키는 것이 어려웠습니다. 하지만, 여러 번의 시행착오를 거쳐서 R에서의 데이터 반환 원리를 정확히 알게 되어서 R의 데이터를 Java에 잘 연결시킬 수 있게 되었습니다.

○ Data 분석 개발

○ 웹 크롤링 :

R언어의 N2H4 패키지를 활용하여 네이버뉴스에서 서비스되는 언론사들의 기사를 검색어를 이용하여 검색된 기사들을 수집하고 수집된 기사들을 URL, 날짜, 언론사, 기사명, 기사본문의 형태로 저장하는데 성공하였고 이후 분석단계를 개발하고 있었으나, 네이버뉴스의 웹서비스가 각 언론사들의 기사를 네이버뉴스를 통해서가 아닌 각 언론사의 기사 본문으로 이동하게 변경되면서 뉴스 수집권한을 획득하지 못하여 검색어를

활용한 뉴스 수집기를 사용하지 못하게 되었습니다.

개발 방향을 네이버뉴스의 주제별로 내용을 수집하는 것으로 개발 목표를 변경하여 뉴스 수집권한 문제를 해결하였습니다.

○ R을 이용한 자연어 처리 :

수집된 기사는 인간이 사용하는 자연어로, 수학, 통계학, 컴퓨터공학에서 분석에 활용하는 숫자와 , 기계어와는 거리가 있는 언어입니다. 따라서 분석을 위해서 자연어를 빈도를 통해서 수치화 하고 그것을 수리적으로 분석해야 했습니다.

자연어 처리를 위한 R에서 사용할 수 있는 다양한 패키지들이 존재하지만, 뉴스 기사를 처리해야 하는 목적에 맞게, 온라인 기사에 자주 나타나는 단어(기자 , 재배포, 무단전재, 바로가기, ...)와 특수문자(▶ ,© ...) 같은 자연어의 처리가 필요하였고, 분석결과 의 해석상에 어려움이 있는 숫자와 영문자 등을 사전에 제거하는 R 함수를 만들어 사용하였습니다. 하지만 그 과정에서도 모든 뉴스기사의 패턴에 맞춰 정제할 수 없었기에 가중치의 개념을 활용하였습니다.

○ 텍스트마이닝 방법론 :

텍스트마이닝에 활용되는 빈도분석 기반의 워드클라우드, 단어 네트워크 분석, 감성 분석 외에도 군집과 머신러닝들을 활용하는 다양한 방법론이 후보로 있었으나, 감성분석에 필요한 한국어 감성분류 사전이 현재 제공되는 곳이 전무하고, 제공되는 곳도 개발하는 서비스의 의도대로 연동하여, 활용하기 어려운 측면이 존재하였습니다. 이외에도 군집분석과 머신러닝 또한, 현재 개발하는 서비스의 목적과 형태에 사용하기 어려운 측면이 존재하였고, 따라서 빈도분석 기반의 워드클라우드와 단어행렬을 구성하고, 행렬곱을 이용하여 동시성을 구하여 네트워크 분석을 이용하였습니다.

4. 프로젝트를 통해 배우거나 느낀 점

○ 프로젝트를 진행함에 있어 각 개발 파트별로 개발 목표를 통일 시키는 것이 중요함을 깨닫는 좋은 시간 이였습니다. 타 개발 파트의 개발 현황을 정확히 파악하기 어렵기 때문에, 서로 기능 구현 목표나, 사용자의 요구사항들을 서로 이해한 바가 같은 지 면밀히 확인하는 과정이 필요하다는 것을 알 수 있었습니다.

또한 개발 파트 별로 연동이 중요함을 알 수 있었는데, R 개발 파트에서의 결과물이 Java환경에서 구현 될 수 있는지, 결과물을 어떠한 형태로 변형 시켜야 되는지 등을 사전에 파악해야 함을 알 수 있었습니다.

○ 분석 팀은 데이터마이닝 분야의 지식 습득, 기술에 대한 이해도 상승의 효과가 있었습니다. R을 활용한 크롤러개발 및 자연어 처리에 관한 이해도가 상승하였으며, 데이터 분석 능력 및 데이터 마이닝 알고리즘을 이해하고 적용함으로써 실전능력을 배

양할 수 있었습니다. 또한 다른 언어와 R을 연동시키기 위해서 코드에서 패키지를 사용을 줄이기 위해 노력하였고, 따라서 코딩능력 또한 향상 되었습니다.

IV. 작품의 기대효과 및 활용분야

1. 작품의 기대효과

○ 언론사별 시사점 파악

R찬 뉴스는 두 개의 언론사를 선택하여 해당 카테고리 내에서 언론사별 주요 키워드 분석 진행하게 된다. 언론사별로 사용된 키워드 및 키워드 빈도 파악으로 중심 키워드를 파악 가능하고 언론사 간에 키워드 차이를 비교할 수 있게 된다. 전체적인 뉴스를 기반으로 주요 키워드와 네트워크만을 제공 하는 빅카인즈와 차이가 있으며, 사용자는 본 작품을 통해 언론사간 서로 다른 시사점을 파악 할 수 있다.

○ 개인 맞춤 뉴스 제공

R찬 뉴스 사용자들의 서비스 이용 로그를 수집하고, 로그를 기반으로 선호하는 카테고리나 키워드를 파악하게 된다. 파악한 결과를 바탕으로 개인별 성향에 따라 개인 맞춤 뉴스를 제공하게 된다.

○ 세부 카테고리별 뉴스 분석

빅카인즈의 분야키워드 서비스는 정치, 경제, 사회, 문화 등 8개로 분류하여 뉴스 분석 진행하지만 R찬 뉴스는 사회, 생활·문화, 세계, IT·과학 6가지 카테고리 3개 이상의 추가적 세부 카테고리 나뉘어 뉴스 분석을 진행하게 된다. 이를 통하여 조금 더 세밀한 주제에 따른 분석 결과를 제공 할 수 있다.

○ 실시간성

빅카인즈의 뉴스 콘텐츠 수집은 전일 오전 5시부터 금일 오후 8시까지로 한정되어 있다. 하지만 R찬 뉴스는 해당 일을 기준으로 실시간으로 데이터를 수집 및 분석 진행하게 된다. 본 서비스를 통해서 시간에 한정 되지 않고 뉴스를 수집할 수 있으며, 사용자는 최신의 뉴스 기사에 관한 분석 결과를 서비스로 받을 수 있다.

2. 작품의 활용분야

○ 사용자 측면

- 뉴스를 읽을 시간이 없는 직장인, 학생

국내 인터넷 사용자들은 다양한 종류의 뉴스 콘텐츠들을 포털사이트를 통해서 접할 수 있지만, 시간상의 이유로 접할 여력이 없는 경우가 많다. 본 작품은 이러한 불편함

을 겪는 직장인, 학생들을 대상으로 뉴스의 콘텐츠를 빠르게 전달하여 시간적 편의성의 문제를 해결 할 수 있다.

- 언론사에게 빠른 정보 제공

언론사들은 자신들이 제공한 뉴스 콘텐츠가 자신들이 원하는 시사점대로 전달이 되었는지 확인할 필요가 있다. 본 작품을 활용하면 각 기사 별로 주요 키워드를 파악하고, 언론사가 제공한 기사들 간의 연관성을 파악함으로써 언론사에게 기사에 대한 피드백을 빠르게 전달 할 수 있다.

- 난독증 환자를 위한 뉴스 요약서비스 제공

단어나 철자를 정확히 읽고 이해하는 데 어려움을 가지고 있는 난독증 환자들은 국내 5% 정도라고 알려져 있다. 말소리를 가장 작은 단위까지 인지하고 처리하는 능력인 음운처리능력의 이상은 난독증의 대표적 증상으로 환자들은 글의 내용을 의미 단위로 쪼개어 인지하는데 어려움을 겪는다. 본 작품은 뉴스 콘텐츠의 내용을 의미 단위의 단어로 보여주어 차후 난독증 환자의 뉴스 서비스 이용에 도움을 줄 수 있다.

○ 확장성 측면

- 뉴스 감성분석

감성분석이란 텍스트의 정보를 추출하는 텍스트마이닝과는 다르게 어떤 주제에 대한 주관적인 인상, 감정, 태도, 개인의 의견들을 텍스트로부터 뽑아내는 분석이다. 일반적으로 찬성/반대와 같은 2진 형식으로 표현된다.

수집된 뉴스 데이터가 분류될 수 있는 한국어 감성사전이 구축된다면 latent semantic analysis/ support vector machines 같은 머신러닝을 사용하여 주제에 관한 기사의 긍정/부정도를 분류 할 수 있고, 개인별 로그데이터가 결합된다면, 새로운 뉴스들을 개인의 성향에 맞춰서 예측하고 추천하는 서비스로 확장 가능하다.

- 언론사 성향 분류

비지도학습의 머신러닝의 하나인 클러스터링을 통해 앞서 계산된 뉴스 데이터들의 코사인 거리 데이터를 활용하여 뉴스 기사를 학습 시킬 수 있다. 모든 언론사들의 기사를 통합하여 뉴스기사 데이터들을 가지고 학습시키고 언론사들을 기존에 정해진 가치관이 아닌 기사 유사도를 통해 언론사들의 분류가 가능하다.

- 개인 서비스 개발

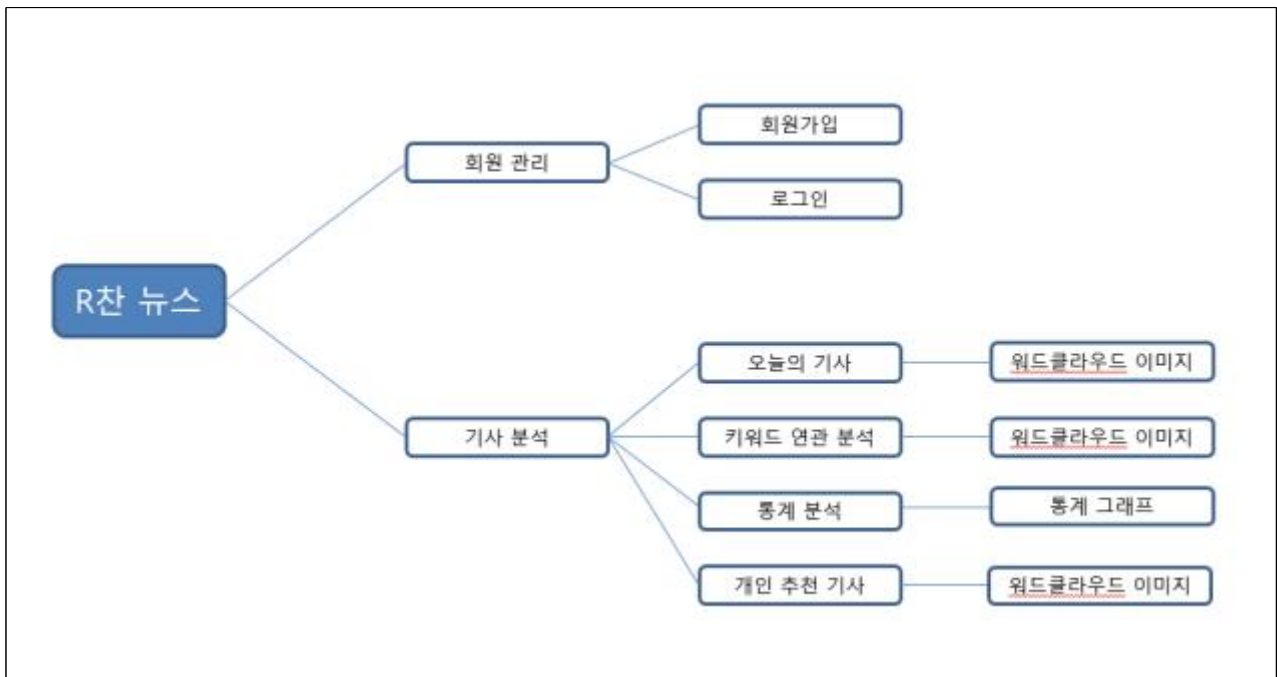
수집된 데이터에서 서비스 사용 유저들의 행동 특성 데이터가 모이게 된다면 그 데이터를 활용하여 새로운 개인 맞춤 서비스를 개발할 수 있을 것으로 기대 된다. 또한 정

제된 텍스트 데이터는 이후 음성, 영상데이터 분석기법의 발전에 따라 수집된 텍스트 데이터와 결합하여 새로운 분석 서비스로 확장 가능하다.

V. 개발산출물

1. 메뉴 구성도

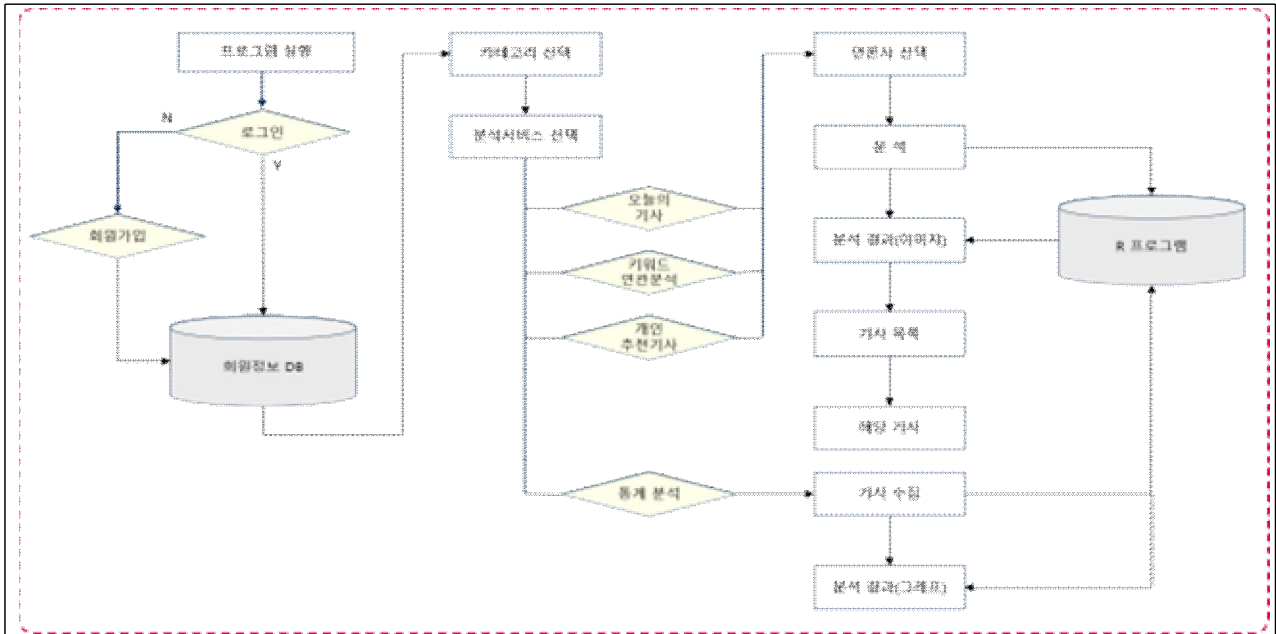
- 크게 회원 관리 기능과 기사 분석 기능으로 나뉩니다. 회원 관리 기능은 회원가입 및 로그인 기능이 있고, 기사 분석에는 오늘의 기사, 카테고리 연관 분석, 통계 분석, 개인 추천 기사 등의 기능이 있습니다. 그리고 카테고리 연관 분석을 통해 분석을 하면 워드 클라우드 이미지를 볼 수 있고, 통계 분석을 하면 해당 언론사들의 수에 대한 결과를 그래프로 볼 수 있습니다.



2. 프로그램 목록

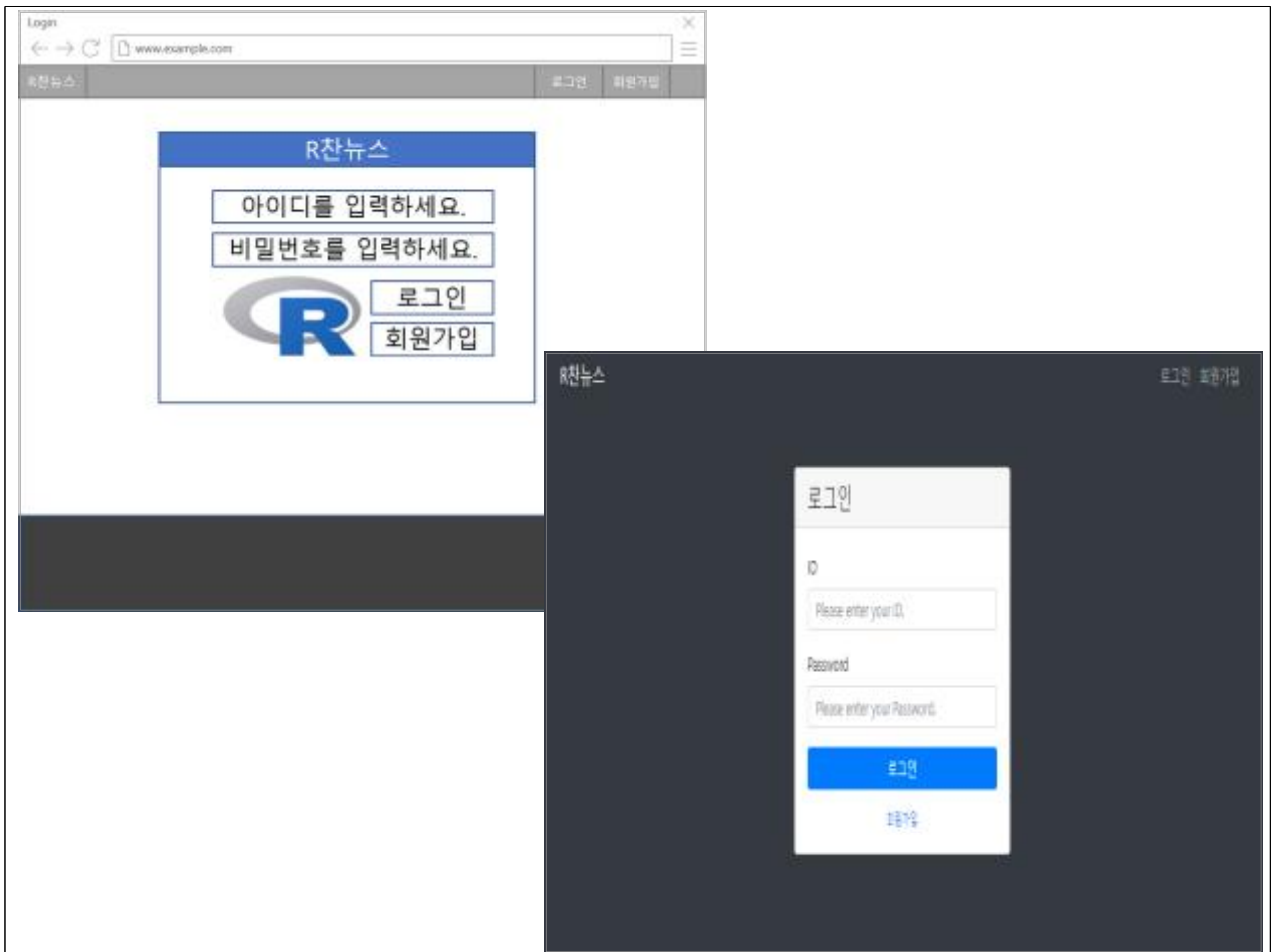
LOG-01-01	로그인	
MAIN-02	메인	
NEWS-02-01	오늘의기사	
NEWS-02-02	워드클라우드	
NEWS-02-03	키워드연관분석	
NEWS-02-04	개인 맞춤기사	

3. 기능 흐름도

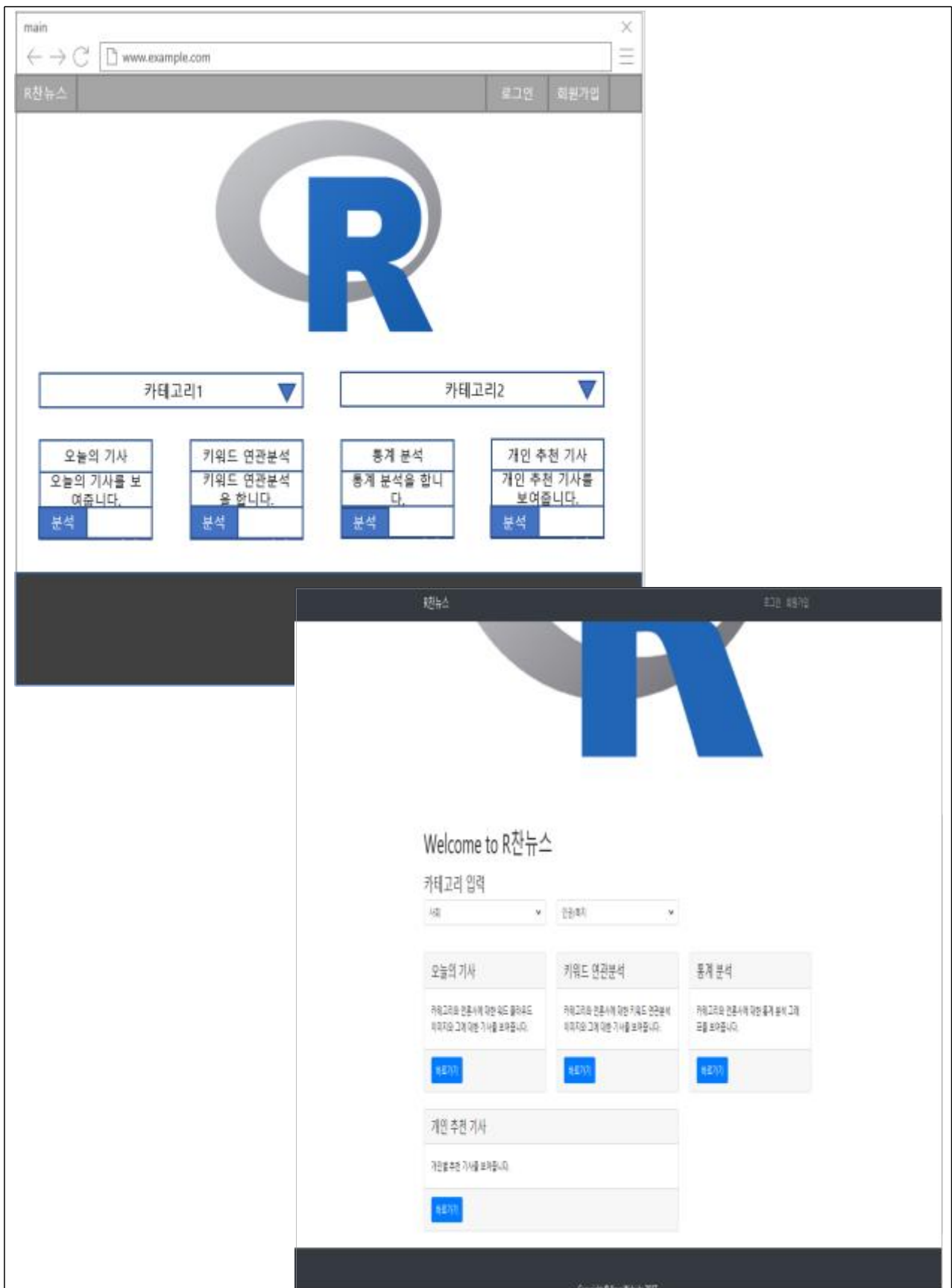


4. 화면 설계서

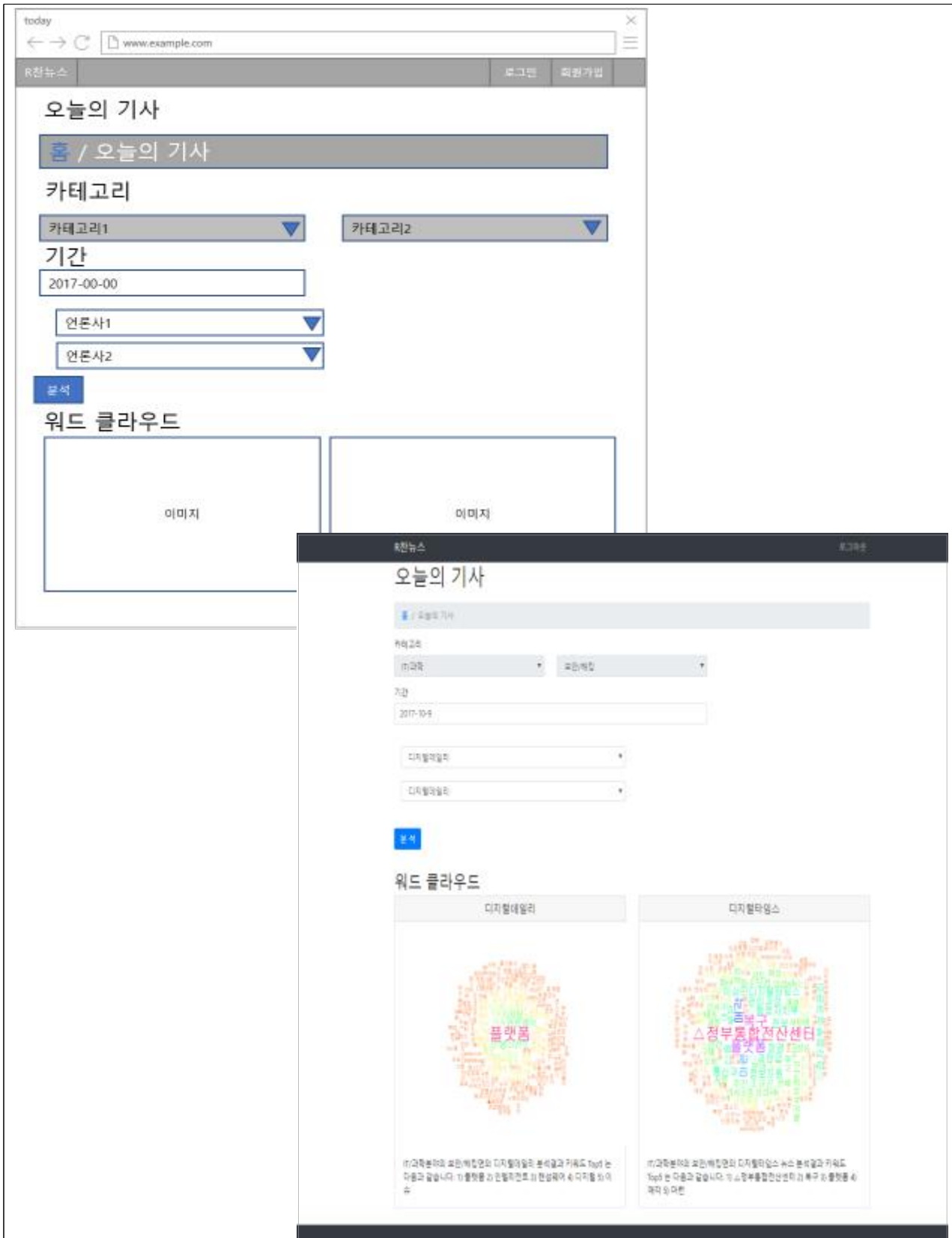
- 로그인 화면 : 회원가입 및 로그인을 할 수 있는 화면입니다.



- 메인 화면 : 카테고리1, 카테고리2를 선택하고 자신이 원하는 분석 기능을 선택합니다.



- 오늘의 기사 : 오늘의 기사 분석 기능을 선택 하면, 해당 카테고리의 오늘의 기사가 있는 언론사1, 언론사2 목록이 생성됩니다. 언론사 2개를 선택 후 분석 버튼을 누르면 해당 기사들의 키워드 빈도수에 따른 워드클라우드 이미지를 보여줍니다. 그리고 이미지를 클릭하면 해당 기사들의 목록을 볼 수 있습니다.



- 키워드 연관분석 : 키워드 연관분석 기능을 선택하면, 해당 카테고리의 키워드 연관분석 이미지를 볼 수 있습니다. 언론사 2개를 선택 후 분석 버튼을 누르면 해당 기사들의 키워드와 연관된 단어들이 연결된 이미지를 보여줍니다. 그리고 이미지를 클릭하면 해당 기사들의 목록을 볼 수 있습니다.

The screenshot displays the '키워드 연관분석' (Keyword Association Analysis) interface. At the top, there's a navigation bar with 'R찬뉴스' and '로그인' (Login) links. The main header shows the current path: '홈 / 키워드 연관분석'. Below this, users can select two categories ('카테고리1', '카테고리2'), a date ('기간'), and two newspapers ('언론사1', '언론사2'). A blue '분석' (Analyze) button is located below the newspaper selection. The main content area is divided into two sections, each labeled '키워드 연관분석' and '이미지' (Image). The bottom section shows two network graphs: '아이뉴스24' and '전자신문'. Each graph consists of a central node connected to several peripheral nodes, representing keyword associations. A legend at the bottom explains that the size of the nodes represents the degree of association and the thickness of the lines represents the frequency of co-occurrence.

- 통계 분석 : 통계 분석 기능을 선택하면, 해당 카테고리에 대한 기사가 있는 언론사 통계 그래프를 볼 수 있습니다. 해당 카테고리에 존재하는 모든 언론사에 대한 여러 가지 통계 자료를 볼 수 있습니다.



- 개인 추천 기사 : 개인 추천 기사 기능을 선택하면, 회원의 기록을 기반으로 분석한 내용에 대한 워드 클라우드 이미지를 볼 수 있습니다. 언론사 2개를 선택 후 분석 버튼을 누르면 회원 기록을 기반으로 분석한 워드 클라우드 이미지를 보여줍니다. 그리고 이미지를 클릭하면 해당 기사들의 목록을 볼 수 있습니다.

The image displays a web application interface for 'Personal Recommendation Article' (개인 추천 기사). The interface includes a navigation bar with 'R News' and '로그인' (Login) and '회원가입' (Sign Up) buttons. The main content area has a breadcrumb '홈 / 개인 추천 기사' and a '카테고리' (Category) section with two dropdown menus for '카테고리1' and '카테고리2'. Below this is a '기간' (Period) section with a date input '2017-00-00' and two dropdown menus for '언론사1' and '언론사2'. A blue '분석' (Analyze) button is present. The '워드 클라우드' (Word Cloud) section shows two empty boxes labeled '이미지' (Image). An inset window shows the same interface after analysis, with the '워드 클라우드' section displaying two word clouds. The left word cloud is titled '디지털대일리' and the right one is titled '디지털타임스'. Both word clouds feature the word '플랫폼' (Platform) prominently. Below each word cloud is a list of top keywords and a brief description of the analysis.

- 해당 기사 목록 : 각 위드클라우드, 키워드 연관분석 등의 내용에 해당하는 기사들의 목록을 보여줍니다. 그리고 링크를 클릭하면 그 해당 기사로 페이지가 바뀝니다.

today
www.example.com
R전뉴스
로그인
회원가입

해당 기사 목록

[홈](#) / [오늘의 기사](#) / [해당 기사 목록](#)

번호	제목	기사 링크	등록일
1	초·한도미사일 1발 발사...북 "강력 대응"	http://innnews.inbc.com/inslay/2017...	2017/06/29
2	'천원강' 도시 일부 요가매트서 유해물질 검출	http://innnews.inbc.com/inslay/2017...	2017/06/29
3	韓, NSC 상임위 소집...北 강핵 규탄	http://innnews.inbc.com/inslay/2017...	2017/06/29
4	이재동 부회장 1심 판결문, 박근혜 재판 증거 채택	http://innnews.inbc.com/inslay/20...	2017/06/29

< 1 >

R전뉴스
로그인
회원가입

기사 목록

[홈](#) / [오늘의 기사](#) / [기사 목록](#)

번호	제목	등록일
1	공정 임국정 면세점 허용될때..내년 정기국회 '즉각'	2017-06-30 16:01:00
2	탄신음료 '사이다' 유행타고 '유희노란' 넘었다	2017-06-30 15:22:00
3	[주목 이사람]환상 속의 언어공주기 나타났다	2017-06-30 15:20:00

< 1 >

Copyright © Your Website 2017

5. 테이블 설계서

<유저 테이블> : 회원가입 시 회원 정보가 쌓이는 테이블

항목명	Type	필수/선택	값 예시	설명
<u>userID</u>	VARCHAR (20)	필수	ex) ssupo11	▪ ID(기본키)
userPassword	VARCHAR (20)	필수	ex) 1234	▪ 비밀번호
userName	VARCHAR (20)	필수	ex) 한이름	▪ 이름
userPhoneNumber	VARCHAR (20)	필수	ex) 01045681234	▪ 핸드폰 번호(숫자만 가능)
userEmail	VARCHAR (50)	필수	ex) adb@naver.com	▪ 이메일(@ 필수)

<로그 테이블> : R찬뉴스에서 제공하는 분석 서비스 이용 시 로그가 쌓이는 테이블

항목명	Type	필수/선택	값 예시	설명
<u>userID</u>	VARCHAR (20)	필수	ex) ssupo11	▪ 회원 아이디(기본키 및 외래키)
<u>press</u>	VARCHAR (20)	필수	ex) 한국일보	▪ 언론사 이름(기본키)
category	VARCHAR (20)	필수	ex) 경제	▪ 주 카테고리
subcategory	VARCHAR (20)	필수	ex) 생활경제	▪ 서브 카테고리
<u>dateTime</u>	DATETIME	필수	ex) 2017-10-09 15:31:05	▪ 날짜 및 시간(기본키)

6 / 7. 프로그램 상세 로직 & 개발소스

○ 웹 개발

- 로그인화면 / html 부분

```
<div class="container">
    <!-- Login Section -->
    <div class="row">
        <div class="col-lg-4 mb-4 login">
            <div class="card h-100">
                <h4 class="card-header">로그인</h4>
                <div class="card-body">
                    <form method="post" action="loginAction.jsp"
name="login" id="contactForm" novalidate>
                        <div class="control-group form-group">
                            <div class="controls">
                                <label>ID</label> <input
type="text" class="form-control"
                                id="name"
placeholder="Please enter your ID." name="userID"
maxlength="20">

```

```

class="help-block"></p>
</div>
</div>
<div class="control-group form-group">
    <div class="controls">
        <label>Password</label>
        <input type="password" class="form-control"
            id="password"
            placeholder="Please enter your Password."
            name="userPassword" maxlength="20">
    </div>
</div>
<div id="success"></div>
<!-- For success/fail messages -->
<button type="submit" class="col-lg-12
    btn btn-primary">로그인</button>
    <div class="text-center">
        <a class="d-block small mt-3"
            href="join.jsp">회원가입</a>
    </div>
</form>
</div>
</div>
</div>
</div>
</div>
</div>

```

- 로그인화면 / loginAction 부분

```

<%
    String userID =null;
    if(session.getAttribute("userID") != null) {
        userID = (String) session.getAttribute("userID");
    }
    if (userID != null) { // 세션이 이미 존재, 중복 로그인 방지
        PrintWriter script = response.getWriter();
        script.println("<script>");
        script.println("alert('이미 로그인이 되어있습니다.');"");
        script.println("location.href = 'main.jsp'"");
        script.println("</script>");
    }
}

```

```

        UserDao userDao = new UserDao();
        int result = userDao.login(user.getUserID(), user.getUserPassword());
        if (result == 1) {
            session.setAttribute("userID", user.getUserID());
            PrintWriter script = response.getWriter();
            script.println("<script>");
            script.println("location.href = 'main.jsp'");
            script.println("</script>");
        }
        else if (result == 0) {
            PrintWriter script = response.getWriter();
            script.println("<script>");
            script.println("alert('비밀번호가 틀립니다.');"");
            script.println("history.back()");
            script.println("</script>");
        }
        else if (result == -1) {
            PrintWriter script = response.getWriter();
            script.println("<script>");
            script.println("alert('존재하지 않는 아이디입니다.');"");
            script.println("history.back()");
            script.println("</script>");
        }
        else if (result == -2) {
            PrintWriter script = response.getWriter();
            script.println("<script>");
            script.println("alert('데이터베이스 오류가 발생했습니다.');"");
            script.println("history.back()");
            script.println("</script>");
        }
    }
}
%>

```

- 로그인화면 / 유저 관련 DB연동 부분

```

public int login(String userID, String userPassword) {
    String SQL = "select userPassword from user where userID = ?";

    try {
        pstmt = conn.prepareStatement(SQL);
        pstmt.setString(1, userID);
        rs = pstmt.executeQuery();
        if (rs.next()) {
            if(rs.getString(1).equals(userPassword)) {

```

```

        return 1; // 로그인 성공
    }
    else return 0; // 비밀번호 불일치
}
return -1; // 아이디가 없음
} catch(Exception e) {
    e.printStackTrace();
}
return -2; // 데이터베이스 오류
}

public int join(User user) {
    String SQL = "insert into user VALUES (?, ?, ?, ?, ?)";

    try {
        pstmt = conn.prepareStatement(SQL);
        pstmt.setString(1, user.getUserID());
        pstmt.setString(2, user.getUserPassword());
        pstmt.setString(3, user.getUserName());
        pstmt.setString(4, user.getUserPhoneNumber());
        pstmt.setString(5, user.getUserEmail());
        return pstmt.executeUpdate(); // 0이상의 숫자가 반환

    } catch(Exception e) {
        e.printStackTrace();
    }
    return -1; // 데이터베이스 오류
}

public User getUser(String userID) {
    String SQL = "select * from user where userID = ?";
    User user = new User();
    try {
        PreparedStatement pstmt = conn.prepareStatement(SQL);
        pstmt.setString(1, userID);
        rs = pstmt.executeQuery();
        while(rs.next()) {
            user.setUserID(rs.getString(1));
            user.setUserPassword(rs.getString(2));
            user.setUserName(rs.getString(3));

```



```

        user.setUserPhoneNumber(rs.getString(4));
        user.setUserEmail(rs.getString(6));
    }

    } catch (Exception e) {
        e.printStackTrace();
    }
    return user;
}

```

- 메인화면 / 카테고리 부분

```
<h1 class="my-4">Welcome to R찬뉴스</h1>
```

```

<%
    Rserve_C rc = new Rserve_C();
    rc.init();
    rc.assign("setwd(\"C:/RPrj\")");

    String[] category = rc.category();
    String[] subCategory = {" "};
    int j;

    int flagCate = 0;
    if (request.getParameter("flagCate") != null) {
        flagCate = Integer.parseInt(request.getParameter("flagCate"));
    }

    int subFlagCate = 0;
    if (request.getParameter("subFlagCate") != null) {
        subFlagCate = Integer.parseInt(request.getParameter("subFlagCate"));
    }

    if (flagCate != 0)
        rc.assignCategory(flagCate);

    if (subFlagCate != 0)
        rc.assignSubCategory(subFlagCate);
%>

<!-- Form Section -->
<div class="row">

```

```

<div class="col-lg-12">
  <h3>카테고리 입력</h3>
  <div class="control-group form-group">
    <div class="controls">
      <div class="row">
        <div class="col-lg-4 mb-4">
          <select name="category" class="form-control"
            onChange="selectCate(this)">
            <option value=0>카테고리</option>
            <%
              for (int i = 0; i < category.length; i++) {
                if (flagCate == i + 1) {
                  %>
                  <option value=<%i + 1%> selected><%=category[i]%></option>
                  <%
                    } else {
                      %>
                      <option value=<%i + 1%>><%=category[i]%></option>
                      <%
                        }
                      }
                    %>
                  </select>
                </div>
                <div class="col-lg-4 mb-4">
                  <select name="subcategory" class="form-control"
                    onChange="selectSubCate(this)">
                    <option value=0>서브카테고리</option>
                    <%
                      if (flagCate != 0) {
                        subCategory = rc.subCategory();
                        for (int i = 0; i < subCategory.length; i++) {
                          if (subFlagCate == i + 1) {
                            %>
                            <option value=<%i + 1%> selected><%=subCategory[i]%></option>
                            <%
                              } else {
                                %>
                                <option value=<%i + 1%>><%=subCategory[i]%></option>
                                <%
                                  }
                                }
                              %>
                            </select>
                          }
                        }
                      }
                    }
  </div>
</div>

```

```

        }
        %>
    </select>
</div>
</div>
<div>
    <p class="help-block"></p>
</div>
</div>
</div>
</div>

```

- 메인화면 / 오늘의 기사 버튼 부분

```

<div class="col-lg-4 mb-4">
    <div class="card h-100">
        <h4 class="card-header">오늘의 기사</h4>
        <div class="card-body">
            <p class="card-text">카테고리와 언론사에 대한 워드 클라우드 이미지와
            그에 대한 기사를 보여줍니다.</p>
        </div>
        <div class="card-footer">
            <button type="submit" class="btn btn-primary"
            onClick="moveToday();">바로가기</button>
        </div>
    </div>
</div>

```

- 메인화면 / 자바 스크립트 부분

```

<script>
    function moveToday() {

        if (
        <%=flagCate != 0%>
        && document.getElementsByName('subcategory')[0].value != "0") {
            var cate = document.getElementsByName('category')[0];
            var subcate = document.getElementsByName('subcategory')[0];

            location.href = 'today.jsp';

        } else
            alert("카테고리를 택하세요.");
        }
    }

```

```

function moveKeyword() {

    if (
<%=flagCate != 0%>
    && document.getElementsByName('subcategory')[0].value != "0") {

        location.href = 'keyword.jsp';

    } else
        alert("카테고리를 선택하세요.");
    }

function moveStatistics() {

    if (
<%=flagCate != 0%>
    && document.getElementsByName('subcategory')[0].value != "0") {

        location.href = 'statistics.jsp';

    } else
        alert("카테고리를 선택하세요.");
    }

function selectCate(selectObj) {
    var keyword = document.getElementsByName('category');

    location.href = 'main.jsp?flagCate=' + selectObj.value;

    alert(selectObj.value + "가 선택 되었습니다.");
}

function selectSubCate(selectObj) {

    location.href = 'main.jsp?flagCate=' +
<%=flagCate%>
    + '&subFlagCate=' + selectObj.value;

    alert(selectObj.value + "가 선택 되었습니다.");
}
</script>

```

- 오늘의 기사 화면 / 언론사 선택 부분 + 로깅

<%

```
Rserve_C rc = new Rserve_CO;

Calendar cal = Calendar.getInstance();

String[] press_names = rc.crawling();

String cate = rc.getCategory();
String subCate = rc.getSubCategory();

rc.getDescription("summary1 = \"준비중\"");
rc.getDescription("summary2 = \"준비중\"");

int press1 = -1;
if (request.getParameter("press1") != null) {
    press1 = Integer.parseInt(request.getParameter("press1"));
}

int press2 = -1;
if (request.getParameter("press2") != null) {
    press2 = Integer.parseInt(request.getParameter("press2"));
}

int pressFlag = 0;
if (request.getParameter("pressFlag") != null) {
    pressFlag = Integer.parseInt(request.getParameter("pressFlag"));
}

String summary1 = "준비중";
String summary2 = "준비중";

LogDAO logDAO = new LogDAO();
if (pressFlag == 1) {
    System.out.println(press_names[press1] + " " +
press_names[press2]);

    rc.assignPress(press_names[press1], press_names[press2]);
    rc.execute(rc.getCategory(), rc.getSubCategory());
    summary1 = rc.getDescription("summary1");
    summary2 = rc.getDescription("summary2");
    logDAO.logging(userID, press_names[press1], cate, subCate,
        cal.get(Calendar.YEAR) + "-" +
```

```

(cal.get(Calendar.MONTH) + 1) + "-" + cal.get(Calendar.DATE)
                                + " " +
cal.get(Calendar.HOUR_OF_DAY) + ":" + cal.get(Calendar.MINUTE) + ":"
                                + cal.get(Calendar.SECOND));
                                logDAO.logging(userID, press_names[press2], cate, subCate,
                                cal.get(Calendar.YEAR) + "-" +
(cal.get(Calendar.MONTH) + 1) + "-" + cal.get(Calendar.DATE)
                                + " " +
cal.get(Calendar.HOUR_OF_DAY) + ":" + cal.get(Calendar.MINUTE) + ":"
                                + cal.get(Calendar.SECOND));
                                }
                                %>

<!-- Form Section -->
<div class="row">
    <div class="col-lg-12">
        <label>카테고리</label>
        <div class="control-group form-group">
            <div class="controls">
                <div class="row">
                    <div class="col-lg-4">
                        <select name="category"
class="form-control" disabled>
                            <option
value=0><%=cate%></option>
                        </select>
                    </div>
                    <div class="col-lg-4">
                        <select name="subcategory"
class="form-control" disabled>
                            <option
value=0><%=subCate%></option>
                        </select>
                    </div>
                </div>
                <p class="help-block"></p>
            </div>
        </div>
    </div>
</div>
<div class="row">
    <div class="col-lg-8 mb-4">

```

```

        <div class="control-group form-group">
            <div class="controls">
                <label>기간</label> <input type="text"
class="form-control"
                id="period"
                value=<%=cal.get(Calendar.YEAR) + "-"
+ (cal.get(Calendar.MONTH) + 1) + "-" + cal.get(Calendar.DATE)%>>
                <p class="help-block"></p>
            </div>
        </div>
    </div>
</div>

<%

%>

<!-- 언론사 Section -->
<div class="row">
    <div class="col-lg-12 mb-4">
        <div class="col-lg-6 mb-4">
            <select name="press1" class="form-control">
                <%
                    for (int i = 0; i < press_names.length;
i++) {
                        if
(lrc.getArticleSum()[i].equals("1")) {
                            %>
                            <option
value=<%=i%>><%=press_names[i]%></option>
                            <%
                                }
                            }
                        %>
                    </select>
                </div>
                <div class="col-lg-6 mb-4">
                    <select name="press2" class="form-control">
                        <%
                            for (int i = 0; i < press_names.length;
i++) {

```

```

                                                                    if
(lrc.getArticleSum()[i].equals("1")) {
                                                                    %>
                                                                    <option
value=<%=i%>><%=press_names[i]%></option>
                                                                    <%
                                                                    }
                                                                    }
                                                                    %>
                                                                    </select>
                                                                    </div>
                                                                    </div>
</div>

```

- 오늘의 기사 화면 / 워드 클라우드 부분

```

<div class="col-lg-6 col-sm-6 portfolio-item">
  <div class="card h-100">
    <%
      if (pressFlag == 1) {
    %>
    <a href="Board.jsp?press=press_idx1"></a>
    <div class="card-body">
      <p class="card-text"><%=summary1%></p>
    </div>
    <%
      } else {
    %>
    <a href="#"></a>
    <div class="card-body">
      <p class="card-text">분석 전 상태</p>
    </div>
    <%
      }
    %>
  </div>
</div>

```

- 오늘의 기사 화면 / 자바 스크립트 부분

```

<script>

```



```

function selectPress() {
    var press1 = document.getElementsByName('press1')[0].value;
    var press2 = document.getElementsByName('press2')[0].value;

    if (press1 != press2) {

        location.href = 'today.jsp?press1=' + press1 + '&press2='
            + press2 + '&pressFlag=1';

    } else {
        alert("언론사가 중복 선택 되었습니다.");
    }
}
</script>

```

- 개인 추천 기사 화면 / 최빈값(카테고리, 언론사) 추출 부분

```

<%
    Rserve_C rc = new Rserve_CO;

    Calendar cal = Calendar.getInstance();
    LogDAO logDAO = new LogDAO0;

    String[] press_names = rc.crawling();

    String cate = logDAO.getCategory(userID);
    String subCate = logDAO.getSubCategory(userID, cate);

    ArrayList<String> press1_List = logDAO.getPress(userID, "0");
    ArrayList<String> press2_List = logDAO.getPress(userID, "1");
    String press1 = null;
    String press2 = null;

    outer1 : for (int i = 0; i < press1_List.size(); i++) {
        for (int j = 0; j < press_names.length; j++) {
            if (press1_List.get(i).equals(press_names[j]) &&
!rc.getArticleSum0[j].equals("1")) {
                press1 = press_names[j];
                break outer1;
            }
        }
    }
}

```

```

        if (press1 == null) {
            for (int i = 0; i < press_names.length; i++) {
                if (lrc.getArticleSum0[i].equals("1")) {
                    press1 = press_names[i];
                    break;
                }
            }
        }

        outer2 : for (int i = 0; i < press2_List.size(); i++) {
            for (int j = 0; j < press_names.length; j++) {
                if (press2_List.get(i).equals(press_names[j]) &&
!lrc.getArticleSum0[j].equals("1")
                                && !press_names[j].equals(press1)) {
                    press2 = press_names[j];
                    break outer2;
                }
            }
        }

        if (press2 == null) {
            for (int i = 0; i < press_names.length; i++) {
                if (lrc.getArticleSum0[i].equals("1") &&
!press_names[i].equals(press1)) {
                    press2 = press_names[i];
                    break;
                }
            }
        }

        rc.getDescription("summary1 = \"준비중\"");
        rc.getDescription("summary2 = \"준비중\"");

        int pressFlag = 0;
        if (request.getParameter("pressFlag") != null) {
            pressFlag = Integer.parseInt(request.getParameter("pressFlag"));
        }

        String summary1 = "준비중";
        String summary2 = "준비중";

        System.out.println(press1 + "" + press2);

```

```

        rc.assignPress(press1, press2);
        rc.execute(cate, subCate);
        summary1 = rc.getDescription("summary1");
        summary2 = rc.getDescription("summary2");
    %>

```

- 개인 추천 기사 화면 / 로그관련 DB연동 부분

```

    public int logging(String userID, String press, String category, String subCategory,
String dateTime) {
        String SQL = "insert into log VALUES (?, ?, ?, ?, ?)";

        try {
            pstmt = conn.prepareStatement(SQL);
            pstmt.setString(1, userID);
            pstmt.setString(2, press);
            pstmt.setString(3, category);
            pstmt.setString(4, subCategory);
            pstmt.setString(5, dateTime);
            return pstmt.executeUpdate(); // 0이상의 숫자가 반환

        } catch(Exception e) {
            e.printStackTrace();
        }
        return -1; // 데이터베이스 오류
    }

    public ArrayList<String> getPress(String userID, String order) {
        String SQL = "select l.press, count(*) from rprj.log l where userID = ? group
by l.press having count(*)="
            + "(SELECT max(count) FROM "
            + "(select l.press, count(*) as count from rprj.log l where
userID = ? group by l.press) as temp) - " + order + ";;";
        ArrayList<String> press = new ArrayList();
        try {
            PreparedStatement pstmt = conn.prepareStatement(SQL);
            pstmt.setString(1, userID);
            pstmt.setString(2, userID);
            rs = pstmt.executeQuery();
            while(rs.next()) {
                press.add(rs.getString(1));
            }
        }
    }

```

```

        } catch (Exception e) {
            e.printStackTrace();
        }
        return press;
    }

    public String getCategory(String userID) {
        String SQL = "select l.category, count(*) from rprj.log l where userID = ?
group by l.press having count(*)="
            + "(SELECT max(count) FROM "
            + "(select l.press, count(*) as count from rprj.log l where
userID = ? group by l.press) as temp);";
        ArrayList<String> cate = new ArrayList();
        try {
            PreparedStatement pstmt = conn.prepareStatement(SQL);
            pstmt.setString(1, userID);
            pstmt.setString(2, userID);
            rs = pstmt.executeQuery();
            while(rs.next()) {
                cate.add(rs.getString(1));
            }

        } catch (Exception e) {
            e.printStackTrace();
        }
        return cate.get((int) (Math.random()*cate.size()));
    }

    public String getSubCategory(String userID, String cate) {
        String SQL = "select l.subCategory, count(*) from rprj.log l where userID = ?
and category = ? group by l.press having count(*)="
            + "(SELECT max(count) FROM "
            + "(select l.press, count(*) as count from rprj.log l where
userID = ? and category = ? group by l.press) as temp);";
        ArrayList<String> subCate = new ArrayList();
        try {
            PreparedStatement pstmt = conn.prepareStatement(SQL);
            pstmt.setString(1, userID);
            pstmt.setString(2, cate);
            pstmt.setString(3, userID);
            pstmt.setString(4, cate);
            rs = pstmt.executeQuery();

```

```

        while(rs.next()) {
            subCate.add(rs.getString(1));
        }

    } catch (Exception e) {
        e.printStackTrace();
    }
    return subCate.get((int) (Math.random()*subCate.size()));
}

public int getLogCount(String userID) {
    String SQL = "select count(*) as count from rprj.log l where userID = ?";
    int count = 0;
    try {
        PreparedStatement pstmt = conn.prepareStatement(SQL);
        pstmt.setString(1, userID);
        rs = pstmt.executeQuery();
        while(rs.next()) {
            count = rs.getInt(1);
        }

    } catch (Exception e) {
        e.printStackTrace();
    }
    return count;
}

```

- 기사 목록 화면 / 기사 목록 부분

<table summary="공지사항 목록입니다.">

<caption>공지사항</caption>

<colgroup>

<col width="10%" />

<col width="*" />

<col width="14%" />

</colgroup>

<thead>

<tr>

<th style="text-align: center;">번호</th>

```

        <th style="text-align: center;">제목</th>
        <th style="text-align: center;">등록일</th>
    </tr>
</thead>
<%
    Rserve_C rc = new Rserve_CO();

    String press = "";
    if (request.getParameter("press") != null) {
        press = request.getParameter("press");
    }

    ArrayList<String> URL = new ArrayList<>();
    Collections.addAll(URL, rc.getURL(press));
    ArrayList<String> title = new ArrayList<>();
    Collections.addAll(title, rc.getTitle(press));
    ArrayList<String> date = new ArrayList<>();
    Collections.addAll(date, rc.getDate(press));

    int maxPageNumber = URL.size() / 10 + 1; //최대페이지수

    int pageNumber = 1; //현재페이지

    int pageSize = 0; // 현재 페이지 사이즈
    PrintWriter script = response.getWriter();

    if (request.getParameter("pageNumber") != null) {
        pageNumber = Integer.parseInt(request.getParameter("pageNumber"));
        if (pageNumber == 0)
            pageNumber = 1;
        else if (pageNumber == maxPageNumber + 1) //확실히 하기 위해.
            pageNumber = maxPageNumber;
    }

    if (maxPageNumber == pageNumber) {
        pageSize = URL.size() % 10 == 0 ? 10 : URL.size() % 10;
    } else if (maxPageNumber > pageNumber)
        pageSize = 10;
%>

<tbody>
<%

```

```

        if (URL.size() != 0) {
            for (int i = (pageNumber - 1) * 10; i < (pageNumber - 1) * 10 +
pageSize; i++) {
                %>
                <tr>
                    <td><%=i + 1%></td>
                    <td class="article"><strong><a
                        href="<%=URL.get(i)%>"><%=title.get(i)%></a></strong></td>
                    <td><%=date.get(i)%></td>
                </tr>
                <%
                    }
                } else {
                %>
                <tr>
                    <td colspan="3">등록된 뉴스가 없습니다</td>
                </tr>
                <%
                    }
                %>
            </tbody>

            <tfoot>

            </tfoot>
        </table>

    </div>
</div>
<!-- /.row -->
<ul class="pagination justify-content-center">
    <li class="page-item"><a class="page-link"
        href="Board.jsp?pageNumber=<%=pageNumber - 1%>" aria-label="Previous">
        <span aria-hidden="true">&laquo;</span> <span class="sr-only">Previous</span>
    </a></li>
    <%
        for (int i = (pageNumber - 1) / 10 * 10; i < (maxPageNumber / 10 == pageNumber
/ 10 ? maxPageNumber
            : (pageNumber - 1) / 10 * 10 + 10); i++) {
            if (pageNumber == i + 1) {
                %>

```

```

<li class="page-item active"><a class="page-link"
  href="Board.jsp?pageNumber=<%=i + 1%>"><%=i + 1%></a></li>
<%
  } else {
%>
<li class="page-item"><a class="page-link"
  href="Board.jsp?pageNumber=<%=i + 1%>"><%=i + 1%></a></li>
<%
  }
}
%>
<li class="page-item"><a class="page-link"
  href="Board.jsp?pageNumber=<%=pageNumber + 1%>" aria-label="Next">
  <span aria-hidden="true">&raquo;</span> <span class="sr-only">Next</span>
</a></li>

</ul>
</div>

```

○ Data 개발

프로그램 ID	CT-1	프로그램 명	클렌징함수	작성일	2017. 07.	Page	1
개요	뉴스 텍스트를 정제하는 함수					작성자	이재호
상세 로직							

제작한 추가적 클렌징 함수

```

cleaning_text <- function(dat){
  char <- gsub("[[:cntrl:]]", "", dat)
  char <- gsub("@[[:graph:]]*", "", char)
  char <- gsub("[A-z]", "", char)
  char <- gsub("WW▶", "", char)
  char <- gsub("WW@", "", char)
  char <- gsub("무단전재 및 재배포 금지", "", char)
  char <- gsub("금지", "", char)
  char <- gsub("재배포", "", char)
  char <- gsub("년", "", char)
  char <- gsub("무단", "", char)
  char <- gsub("전재", "", char)
  char <- gsub("바로그가기", "", char)
  char <- gsub("기자", "", char)
  char <- gsub(press_idx[1], "", char)
  char <- gsub(press_idx[2], "", char)
  char <- gsub("[^WuAC00-WuD7A3xfe a-zA-Zㄱ-ㅎ 가-힣Wws]", "", char, perl = FALSE)
  char <- gsub("@", "", char)
}

```


프로그램 ID	TM-1	프로그램 명	TDM	작성일	2017. 07.	Page	2
개요	뉴스 텍스트를 정제하여, TermDocumentMatrix를 생성					작성자	이재호

상세 로직

```

press_body1 <- cleaning_text(tmp1$body) %>% sapply(extractNoun, USE.NAMES = FALSE)
press_body2 <- cleaning_text(tmp2$body) %>% sapply(extractNoun, USE.NAMES = FALSE)

cps1 <- Corpus(VectorSource(press_body1))
cps2 <- Corpus(VectorSource(press_body2))

tdm1 <- TermDocumentMatrix(cps1,
  control=list(tokenize=ko.words,
    removePunctuation=T,
    stopwords = T,
    removeNumbers=T,
    wordLengths=c(4, Inf),
    weighting=weightTfIdf))
tdm2 <- TermDocumentMatrix(cps2,
  control=list(tokenize=ko.words,
    removePunctuation=T,
    stopwords = T,
    removeNumbers=T,
    wordLengths=c(4, Inf),
    weighting=weightTfIdf))

```

프로그램 ID	NC	프로그램 명	네이버뉴스크롤러	작성일	2017. 07.	Page	3
개요	뉴스 텍스트를 정제하는 함수					작성자	이재호

상세 로직

```

naver_news_search.cl <- function(url, locale){
  id <- Sys.time()
  t1 <- Sys.time()

  startDate <- paste0(year(), "-", month(), day(), sep="")
  endDate <- paste0(year(), "-", month(), day(), "-1", sep="")

  pageUrl <- paste0("http://news.naver.com/main/list.naver?cid=", locale, "&sid1=", locale, "&mid=a4&endNodes=LSO&dates=", startDate)

  max <- getMaxPageNum(pageUrl)
  print(paste0("Maximum page : ", max, " crawling start."))
  final_data <- c()

  for (pageNum in 1:max){
    print(paste0("pageNum : ", pageNum, " / start Time : ", round(Sys.time()-t1, 2), "s / spent Time at first : ", round(Sys.time()-t1, 2), "s"))
    t1 <- Sys.time()
    newsData <- c()
    newsUrl <- getURLbyCategory(pageUrl)
    for (newsLink in newsUrl$links){
      try <- 0
      tem <- tryGetContent(newsLink, silent = TRUE)

      while(try <- 360 && length(tem) < "try-error"){
        tem <- tryGetContent(newsLink, silent = TRUE)
        try <- try+1
        print(paste0("try again : ", newsLink))
      }
      if(class(tem)$datetime[1] != "POSIXct"){
        newsData <- cbind(newsData, tem)
      }
      final_data <- rbind(final_data, newsData)
    }
  }
  idx.rm <- which(duplicated(final_data))
  final_data <- final_data[-idx.rm,]

  t2 <- Sys.time()
  time <- t2-t1
  print(paste0("Total time : ", round(time, 2), "min"))
  dx <- create("data", showWarnings=F)
  write.csv(final_data, file=paste0("data/news", locale, "_", locale, "_", startDate, ".csv"), row.names = F)

  print(paste0("저장된 파일명은 : ", "news", locale, "_", locale, "_", startDate, ".csv", " 입니다."))
  return(final_data)
}

```

프로그램 ID	WND	프로그램 명	노드 가중치 함수	작성일	2017. 08. 30	Page	4
개요	단어와 단어간의 빈도수에 따라 노드별(단어) 가중치를 부여하게 된다.					작성자	임동희

상세 로직

```
co.dataframe1 <- as.data.frame(co.matrix1)
co.dataframe1$child = names(co.dataframe1)
co.dataframe1 <- co.dataframe1 %>% gather(parent,degree,1:(ncol(co.dataframe1)-1))
co.dataframe1 <- left_join(co.dataframe1,co.dataframe1 %>% group_by(child) %>% summarise( sum = sum(degree)),by="child")
co.dataframe1 <- co.dataframe1[-which(co.dataframe1$degree==0),]
co.dataframe1$sum <- co.dataframe1$sum*30
co.dataframe1 <- co.dataframe1[-which(co.dataframe1$child==co.dataframe1$parent),]

row.names(co.dataframe1) = NULL
duple = vector()
for(i in 1:nrow(co.dataframe1)){
  for(j in i:nrow(co.dataframe1)){
    if(i!=j){
      if(co.dataframe1$child[i]==co.dataframe1$parent[j]&co.dataframe1$parent[i]==co.dataframe1$child[j]){
        duple = append(duple, j)
      }
    }
  }
}
co.dataframe1 <- co.dataframe1[-duple,]
co.dataframe1$degree=NULL
```

프로그램 ID	WNG	프로그램 명	워드 네트워크 생성 함수	작성일	2017. 08. 30	Page	5
개요	단어별로 주어진 가중치에 따라 워드 네트워크 그래프를 생성, 저장하게 된다.					작성자	임동희

상세 로직

```
png("./data/as_image.png",
     width=800,
     height=600
)
#par(mfcol= c(1,1))

p<-
co.dataframe1%>%
  graph_from_data_frame() %>%
  ggraph(layout = "fr") +
  geom_edge_link(aes(edge_alpha = sum, edge_width = sum)) +
  geom_node_point(color = "darkslategray4", size = 17) +
  geom_node_text(aes(label = name), vjust = 0) +
  theme_void()
plot(p)

dev.off()
print("co-occurrence matrix 네트워크 분석결과 1 저장완료.")
```

프로그램 ID	WCG	프로그램 명	워드 클라우드 함수	작성일	2017. 08. 30	Page	6
개요	단어별 빈도수에 따라 색과 크기를 달리하여 워드 클라우드 그래프를 생성, 저장하게 된다.					작성자	임동희
상세 로직							

```

windowsFonts(nanum = windowsFont("서울남산체 EB"))
pal= brewer.pal(8,"Dark2")#pal= rainbow(14,alpha = 1)

png("./data/wc_image.png",
    width=800,
    height=600
)
#par(mfcol= c(1,1))
wordcloud(words = names(word.count1), freq = word.count1 , min.freq = 2,
    random.order = T,rot.per = 0.10 , scale = c(4,5), max.words = 150 , colors = pal , family = `nanum`)

dev.off()

png("./data/wc_image2.png",
    width=800,
    height=600
)

```