

# 컴퓨터공학 종합 설계 최종 보고서

- 해외축구 어플리케이션 ISPN -

학 과
컴퓨터공학과
팀 명
FM(1조)
참 여 인 원
3명(강혁준, 김민철, 배재경)
개 발 주 제
사용자 맞춤 해외 축구 뉴스 통합 어플리케이션

## 목 차

제1장 개요	3
1.1 목적 및 기대효과	3
1.1.1 목적	3
1.1.2 기대효과	3
1.2 프로그램 개요	4
1.2.1 배경	4
1.2.2 관련 기술 동향	5
1.2.3 사용 대상자	6
1.3 개발 도구	7
1.4 업무분담 및 일정계획표	7
1.4.1 업무분담	7
1.4.2 일정계획표	7
1.5 소요비용	8
제2장 분석	9
2.1 업무 흐름도	9
2.2 프로세스 분석	10
2.2.1 자료 흐름도	10
2.2.2 상세 자료 흐름도	11
2.2.3 기능요구조건 명세서	14
2.2.4 성능요구조건 명세서	19
2.3 인터페이스 분석	13
2.3.1 입.출력 인터페이스 종류와 목적	13
2.3.2 입.출력 인터페이스 내용	13
제3장 설계	20
3.1 DB 설계	20
3.1.1 테이블 구성	20
3.2 모듈 설계	23
3.2.1 크롤링 설계	23
3.2.2 데이터 마이닝 설계	24
3.2.3 중복기사 필터링 설계	24
3.2.3 트위터 클러스터링 설계	25
3.2.4 추천시스템 설계	26
3.3 사용자 인터페이스 설계	28
3.3.1 세부 UI	28
3.3.2 화면흐름도	32

제4장 연구결과	33
4.1 테스트케이스를 통한 요구조건 명세 목표달성 결과	33
4.1.1 기능요구조건 목표달성 결과	33
4.1.2 성능요구조건 목표달성 결과	41
4.2 추천시스템 평가	42
4.2.1 기존 추천시스템과의 비교	42
4.2.2 전체추천	42
4.2.3 그룹추천	43
4.2.4 개인추천	44
4.2.5 기존 추천시스템 성능 평가	44
4.2.6 추천시스템 성능 평가	45
제5장 결론	47
5.1 문제점 및 개선방안	47
5.2 결론	47
5.3 참고 문헌	47

# 1 개요

## 제1장 개요

### 1.1 목적 및 기대효과

#### 1.1.1 목적

해외 축구의 관심 키워드에 따라 언론사들의 기사와 실시간 SNS를 통합하여 추천 및 제공해주는 어플리케이션

- 10개 언론사 통합 기사 제공
  - 통합 기사 최신순 제공 기능
- 관심 키워드 지정
  - 1) 관심 키워드에 맞춰 최신순, 관련도순 기사 제공
  - 2) 키워드기반 기사 검색 기능 제공
  - 3) 날짜별 기사 제공
  - 4) 관심 키워드 수정 기능
- 사용자 맞춤 기사 추천
  - 추천시스템을 적용한 사용자 맞춤 추천 기사 제공
- 트위터를 통한 팀별 실시간 반응 확인 가능
- 사용자의 시간절약과 관심 있는 해외 축구 기사 접근성 높임
- 심플한 UI를 통한 사용자 편의성 증대

#### 1.1.2 기대효과

첫째, 10개의 언론사를 통합한 해외 축구 기사와 SNS를 어플리케이션 내에서 제공하여 해외 축구 정보에 대한 다양성, 편리성, 접근성을 증대시킬 수 있다.

둘째, 관심 키워드 지정

사용자가 관심 있는 팀, 선수 또는 감독으로 3개의 키워드를 지정하면 각 키워드별 기사를 최신순, 관련도순으로 정렬하여 볼 수 있다. 또한 검색 기능을 통해 관련 키워드에 해당하는 키워드를 선택하면 해당 키워드 기사들을 볼 수 있다. 날짜를 선택하여 해당 날짜의 기사를 볼 수도 있다. 관심 키워드는 언제든지 수정할 수 있다.

셋째, 사용자 맞춤 기사 추천

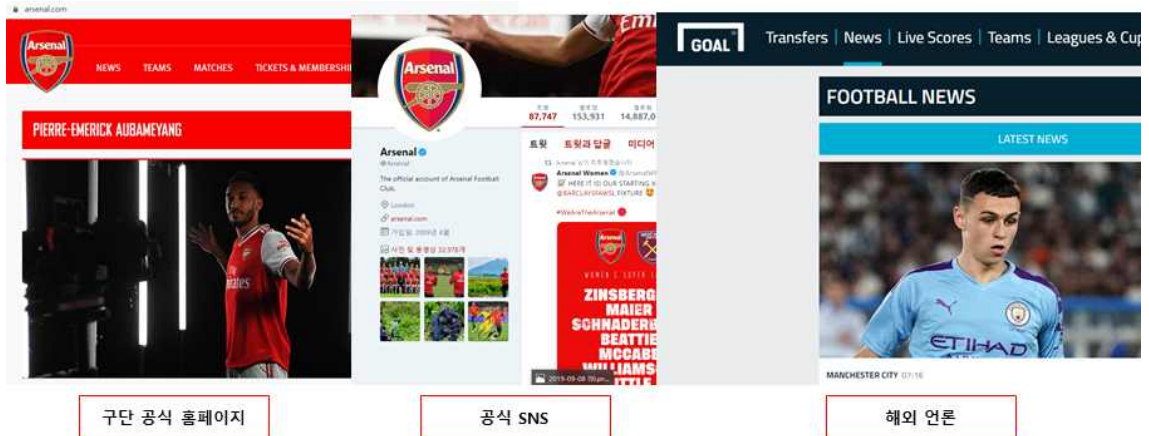
사용자가 어플리케이션을 사용하면 사용자의 성향과 행동 데이터를 분석하여 기사를 추천해주는 추천시스템을 적용하여 개개인의 사용자에게 맞춤 추천 기사를 제공할 수 있다. 추천시스템은 전체 추천, 그룹 추천, 개인 추천으로 나뉘어 사용자가 처음 어플리케이션을 사용할 때도 전체 추천과 그룹 추천 기사를 확인할 수 있다.

넷째, 트위터를 통한 팀별 실시간 반응 확인

실시간 SNS인 트위터를 어플리케이션 내에서 제공하여 20개의 팀별 실시간 반응을 빠르고 쉽게 확인할 수 있다.

## 1.2 프로그램 개요

### 1.2.1 배경



분산되어있는 해외 축구관련 정보들이 한곳에 통합되어있으면서, 원하는 팀과 선수 관련 기사만 볼 수 있는 어플리케이션 개발의 필요성을 느껴 제작하게 되었다.

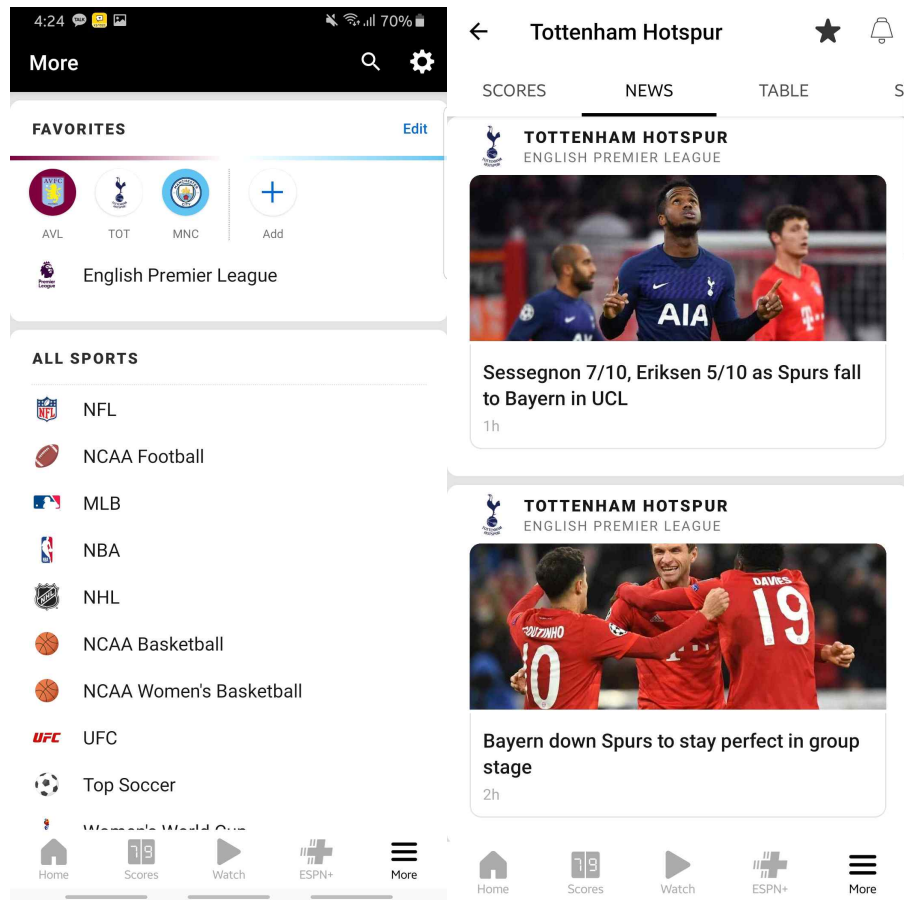
해외 축구를 좋아하는 사람들은 해외 축구 기사와 축구 팬들의 실시간 반응을 한곳에서 빠르고 쉽게 확인하고 싶다는 바람이 있을 것이다. 하지만 국내 포털 사이트들의 해외 축구 기사는 해외 축구 언론사의 기사를 번역해오기 때문에 실시간으로 올라오는 기사는 해외 언론사 사이트에 직접 들어가 확인하거나 해당 언론사 어플리케이션을 사용할 수밖에 없었다. 해외 언론사는 그 수가 다양하기에 특정 언론사를 자주 확인하고, 실시간 해외 팬들의 반응은 트위터에서 해당 팀 관련 검색을 통해 반응을 확인해야만 했다. 또한 자신이 좋아하는 선수 또는 감독 관련 기사를 찾아보기 위해서는 직접 검색하거나 ESPN에 로그인하여 미리 지정해야하는 불편함이 있었다.

스포츠 기사에서 중요한 것은 빠르게 올라오는 기사와, 실시간 팬들의 반응을 확인하는 것이라고 생각하였다. 한 눈에 원하는 선수의 기사를 실시간으로 확인하고 싶다는 바람으로 어플리케이션을 제작하게 되었다. 이름은 Inha+ESPN=ISPEN으로 정하였다.

현재 포털 사이트 외에는 추천 시스템을 도입한 해외 축구 언론사 또는 어플리케이션이 없다는 점이 아쉬웠었다. 따라서 추천 시스템을 개발하여 사용자가 지정한 관심 키워드 관련 기사 외에도 충분히 관심을 가질만한 기사를 추천해주는 '기사 추천' 기능을 도입하였다. 이를 통해 사용자가 관심 키워드 외에도 흥미로운 기사들을 접할 수 있게 함으로써 사용자에게 제공하는 정보의 다양성을 향상시키고자 하였다.

## 1.2.2 관련 기술 동향

- ESPN

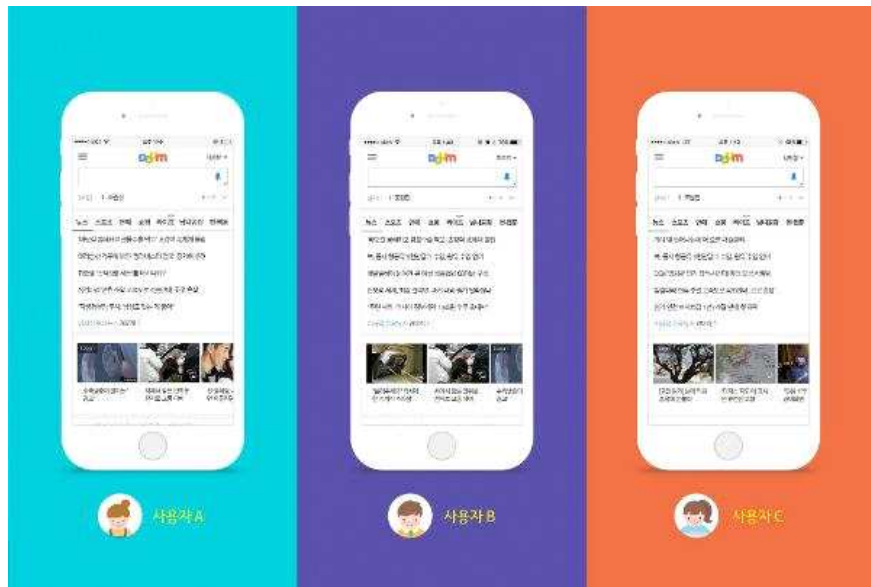


<ESPN 팀 키워드 지정 화면과 관련 기사 제공 화면>

ESPN은 미국 최대 스포츠 방송사로, 현재 5,000만 이상 다운로드된 통합 스포츠 어플리케이션이다. More 탭에서 지정한 관심 팀을 누르면 해당 팀 관련 기사를 최신 순으로 확인할 수 있다.

관심 키워드를 리그와 팀 단위로만 지정할 수 있다는 점, 팀 관련 기사를 최신 순으로만 제공한다는 점, 실시간 SNS 반응을 확인할 수 없다는 점, 그리고 추천시스템이 없다는 점에서 ESPN과 차이가 있다. ESPN은 해외 축구에서 EPL(English Premier League)의 20개의 팀과 선수, 감독을 키워드로 지정할 수 있으며, 키워드마다 관련도순, 최신순으로 기사를 정렬하여 볼 수 있다. 트위터 반응을 각 팀별로 실시간으로 확인할 수도 있다. 또한 추천시스템을 도입하여 기사를 추천해주는 기능이 탑재되어있다. EPL 해외 축구 기사를 자주 접하는 사용자에게 편리성과 특수성이 강조되는 어플리케이션이다.

- 카카오 RUBICS(루빅스)



<사용자마다 다른 메인 뉴스를 추천해주는 비교화면(제공: 카카오 RUBICS)>

카카오 RUBICS는 다음 포털 사이트에서 사용자가 선호하는 뉴스에 따른 기사를 추천해주는 AI기반 추천 시스템이다. 협력 필터링과 내용기반 필터링을 주로 사용하는 추천시스템에 맞춤형 멀티암드밴딩을 추가로 적용하여 사용자의 행동 데이터를 분석하고 관심 기사를 추천해 준다. 일반적인 추천시스템에서 사용자의 성향을 모르는 초반에는 추천을 해주기 어렵다는 Cold Start(콜드 스타트) 문제를 맞춤형 멀티암드밴딩으로 해결하였다.

ISPN은 루빅스의 추천 시스템에서 아이디어를 착안하여 해외 축구 기사 추천 시스템을 전체 추천, 그룹 추천, 개인 추천으로 나누어 설계하였다.

전체추천에서는 텍스트 마이닝, 그룹추천과 개인추천에서는 트위터 클러스터링을 적용하였다. 트위터 클러스터링은 트위터 내용을 추천에 반영한 것으로, 이를 통해 키워드 관련 주제에 대한 사람들의 트렌드를 반영하고자 하였다.

이용자 행동 정보가 거의 없는 콜드 스타트 이용자들의 기사 소비 패턴을 분석해 기사를 제공하는 루빅스와 달리 ISPN은 10개 언론사의 해외 축구 뉴스를 통합 제공하여 해당 서비스를 자주 사용하는 해비 유저를 대상으로 하며, 선호하는 키워드를 설정하면 관련 기사와 추천 기사를 제공한다.

이를 통해 사용자의 소비 패턴을 보다 자세히 분석할 수 있어 사용자가 관심 있을만한 기사, 또는 사용자가 원하는 기사를 볼 수 있게 하여 사용자의 편리성과 개별성을 높일 수 있다는 장점이 있다.

### 1.2.3 사용 대상자

본 어플리케이션의 기능은 대부분의 스마트폰과 태블릿 PC 이용자들이 사용할 수 있다. 해외 축구 기사를 편하고 빠르게 접하고 싶거나, 해외 축구팀별 트위터를 실시간으로 확인하고 싶은 해외 축구 팬들을 주요 대상으로 한다.

### 1.3 개발 도구

- 1) Django  
서버와 클라이언트 연동하여 데이터 전송
- 2) MySQL  
DBMS 개발 도구
- 3) RStudio  
R 언어 기반 개발 도구
- 4) PyCharm  
Python 언어 기반 개발 도구
- 5) Java JRE 1.8.0  
Java 언어 기반 개발 도구
- 6) Android Studio 3.5  
3.5 버전의 안드로이드 형식의 어플리케이션 개발 도구
- 7) Android SDK 9.0 (Pie)  
Android SDK 9.0 버전의 개발 도구
- 8) Ubuntu  
Linux OS

### 1.4 업무분담 및 일정계획표

#### 1.4.1 업무 분담

팀 원 명	역할
김민철 (팀장)	어플리케이션 설계, 크롤링 설계, 추천 시스템 설계 및 구현
강혁준 (팀원)	서버 / DB 구축 및 구현, 크롤링 구현, 어플리케이션 구현, 추천 시스템 설계
배재경 (팀원)	데이터 마이닝, 중복 필터링, 트위터 클러스터링, 추천 시스템 설계



### 1.4.2 일정 계획표

세 부 과 제		일 정							
		1주	2주	3주	4주	5주	6주	7주	8주
분석	설계주제 선정								
	기초 자료조사							중	중
	제안서 작성							간	간
설계	시스템 설계							고	고
	사용자인터페이스 설계							사	사
	구조 설계								
	데이터베이스 설계								

세 부 과 제		일 정							
		9주	10주	11주	12주	13주	14주	15주	16주
구현	서버 구축								
	데이터베이스 구축								
	서버 세부기능 구현								
	추천 시스템 구현								
	어플리케이션 개발								
시험	통합 시험 실시								
	최종 발표								

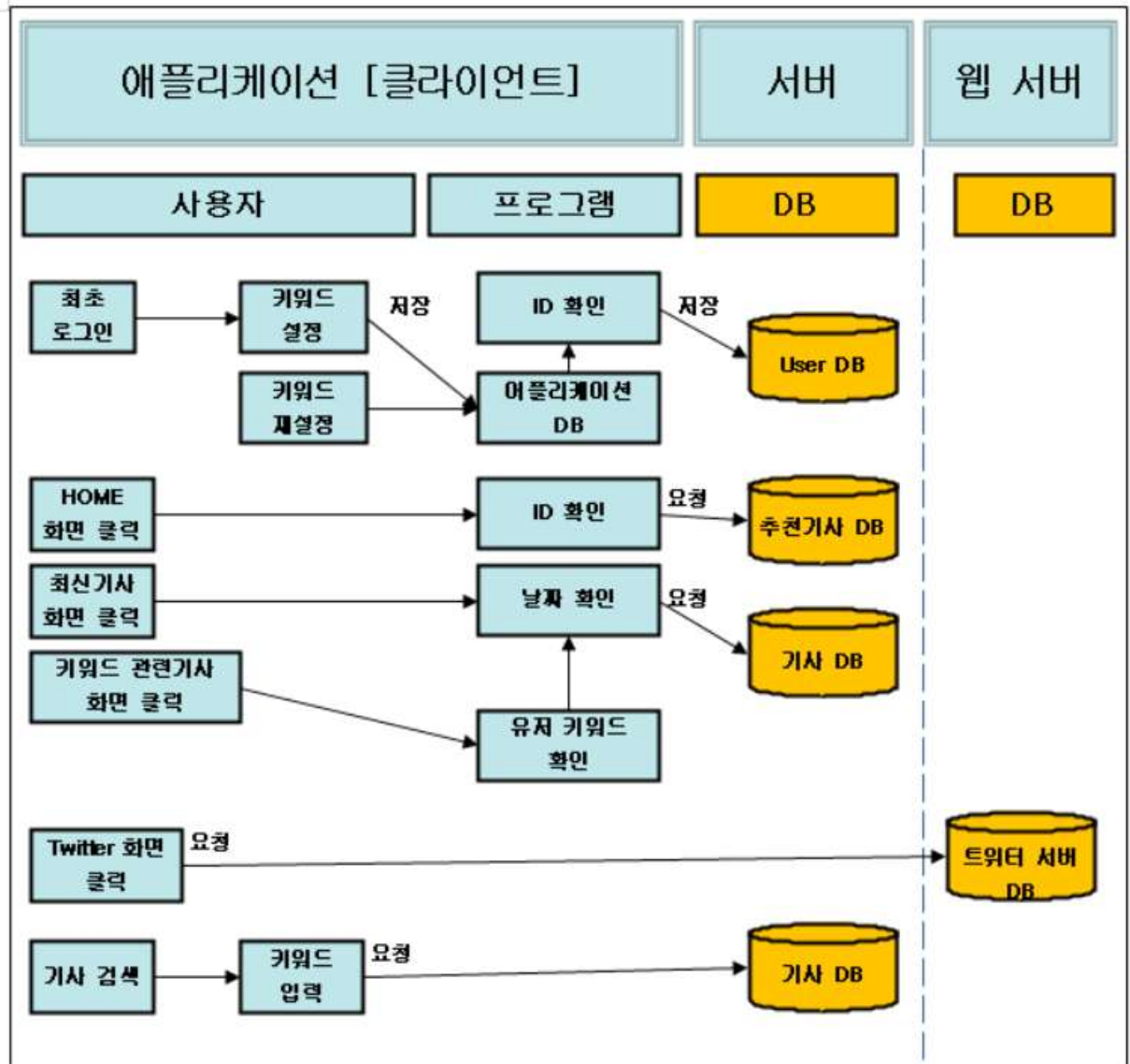
### 1.5 소요 비용

항목	관련정보
컴퓨터 SSD 추가	153,360원
안드로이드 UI 구매	24,000원
기타 자료	20,000원
총 비용	197,360원

장비는 개인 노트북을 각각 사용하며, 어플리케이션 테스트는 개인 스마트폰에서 하였다. 서버 및 프로그램 구현은 Team Viewer를 통해 SSD를 추가한 컴퓨터에서 작업하였다.

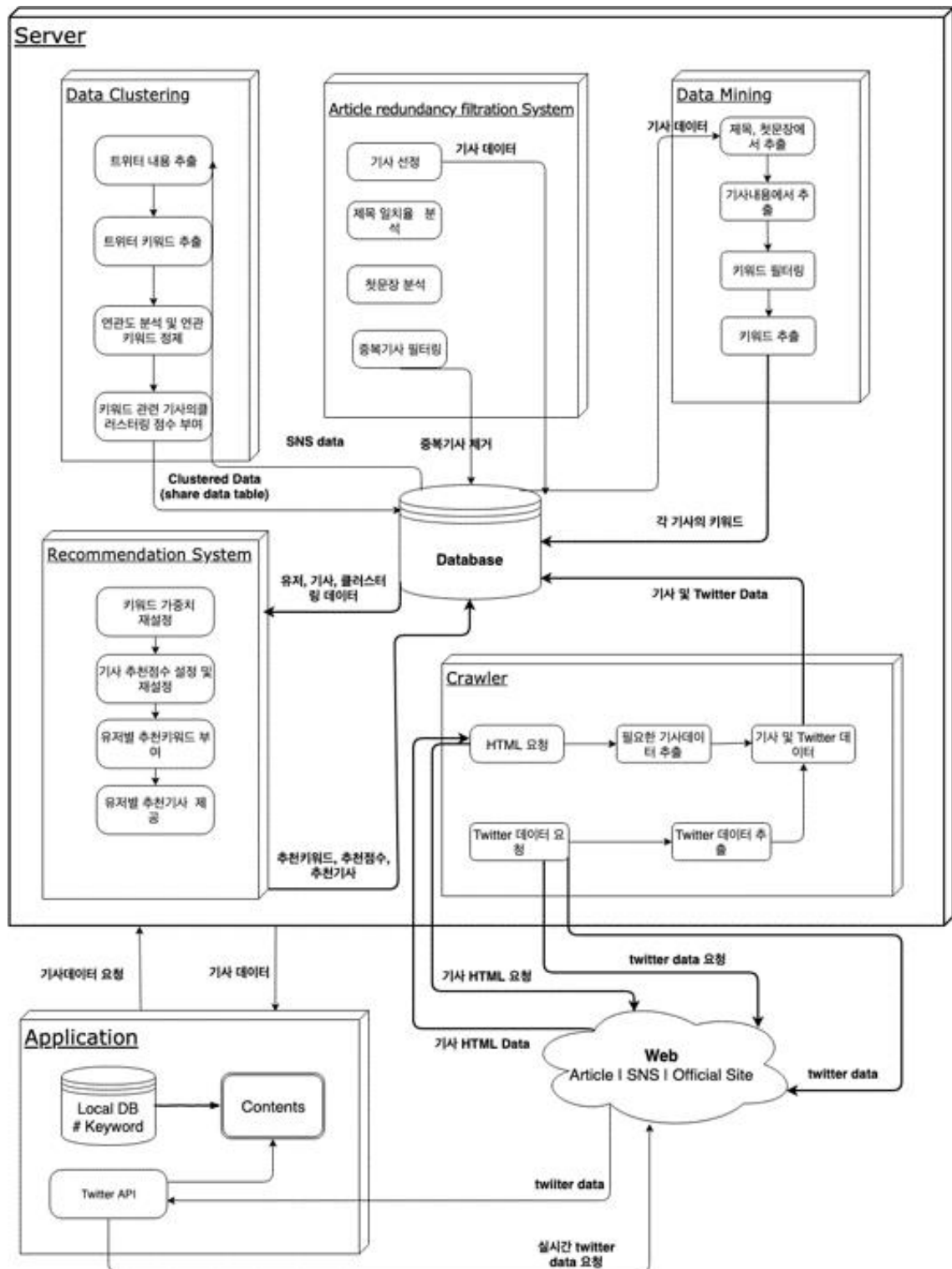
## 2 분석

### 2.1 업무흐름도



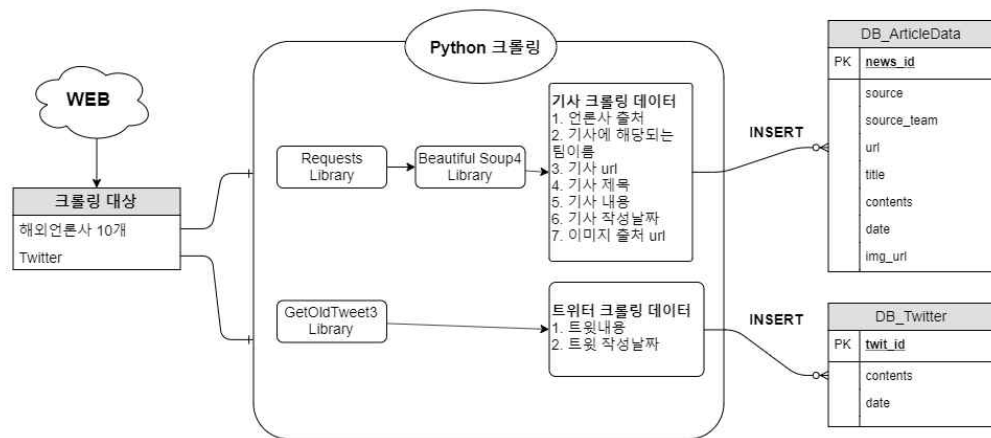
## 2.2 프로세스 분석

### 2.2.1 자료흐름도

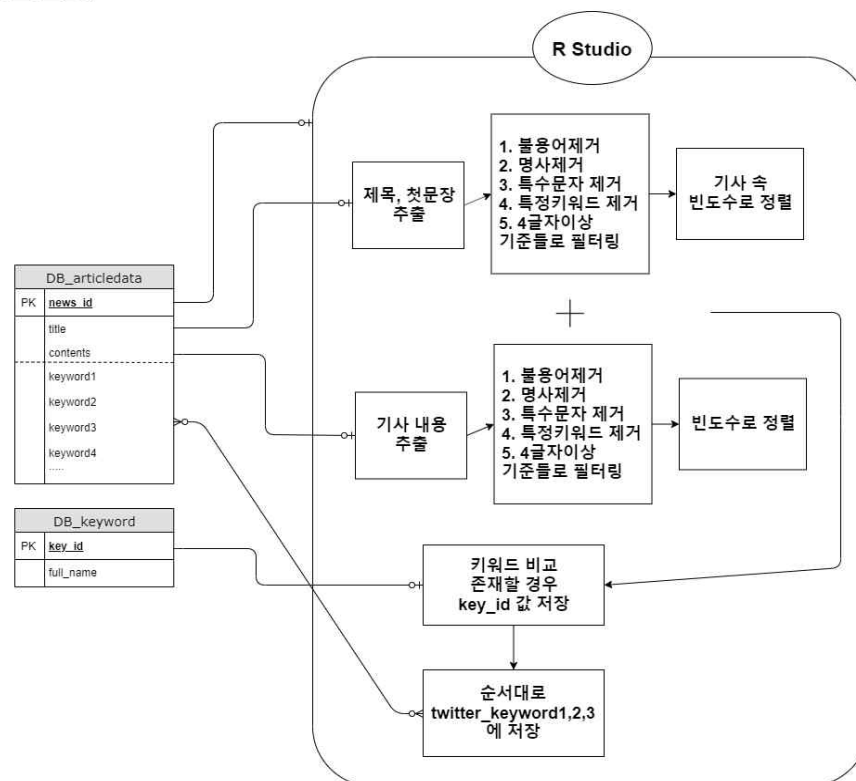


## 2.2.2 상세 자료흐름도

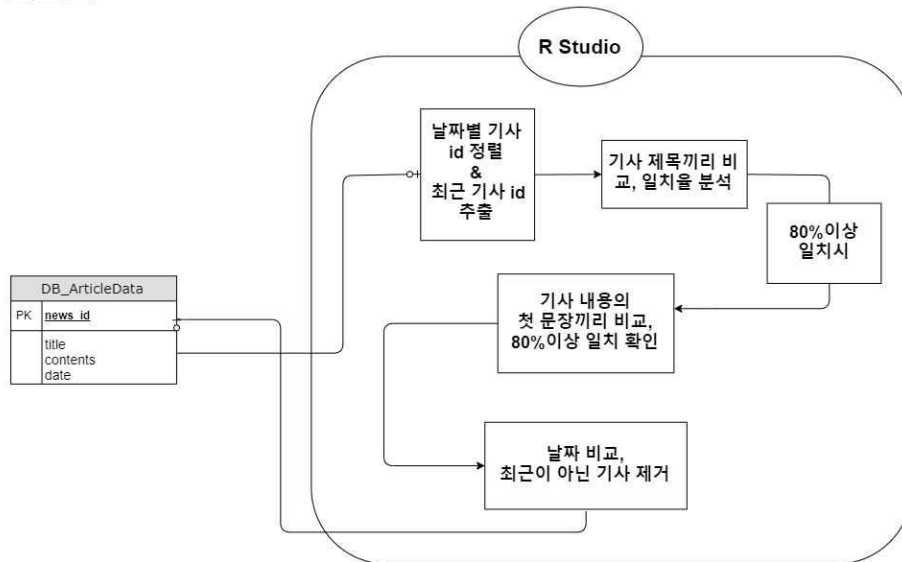
### 크롤링



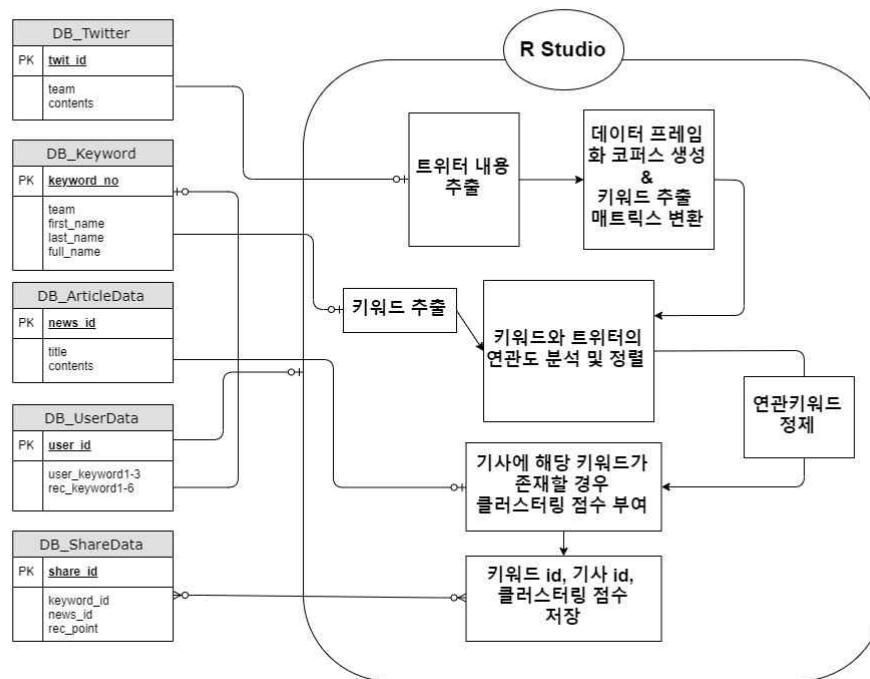
### 데이터 마이닝

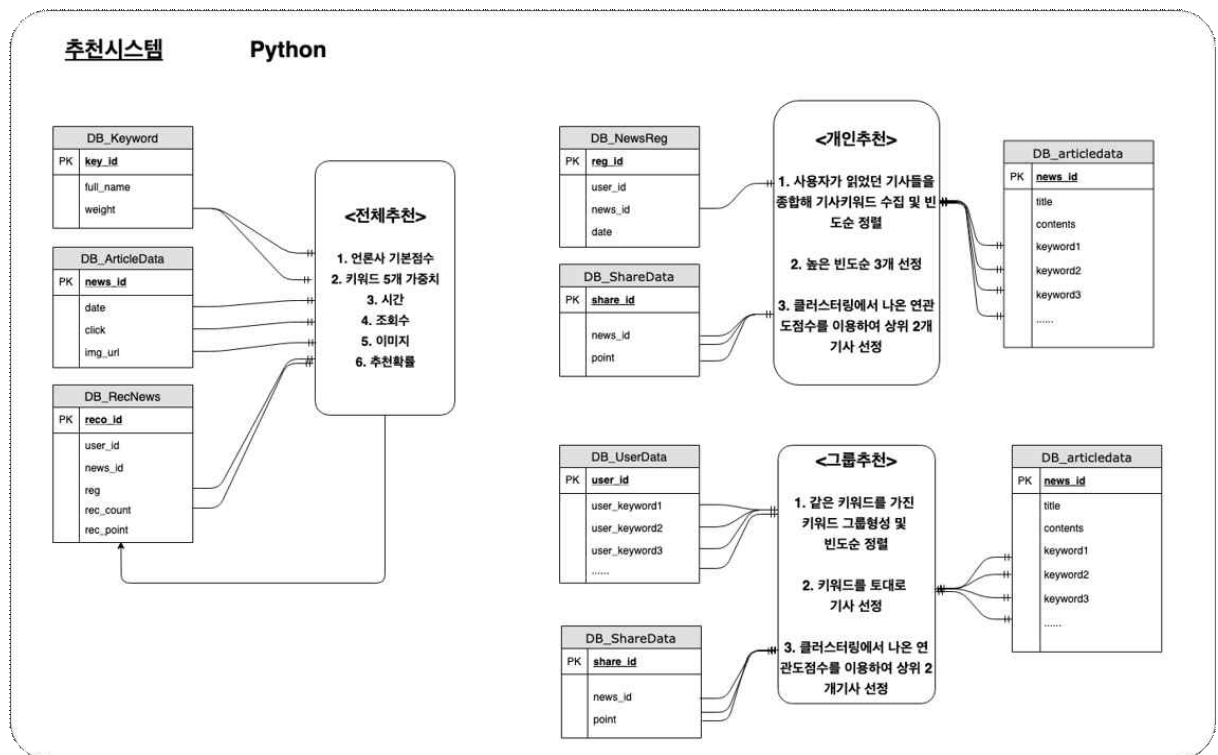


## 중복 기사 제거



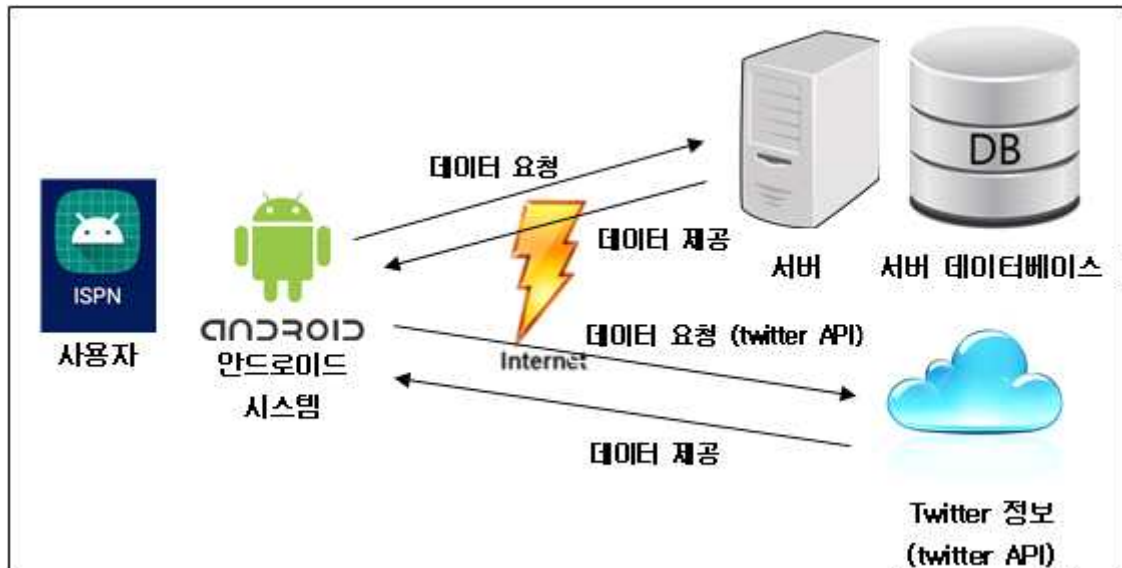
## 트위터 클러스터링





## 2.3 인터페이스 분석

### 2.3.1 입·출력 인터페이스 종류와 목적



프로그램 사용자 : 안드로이드 시스템을 이용하여 서비스를 요청하고 제공 받는다.  
서비스 제공자 : 서버 데이터베이스. Twitter API를 통해 요청된 정보를 제공한다.

### 2.3.2 입·출력 인터페이스 내용

프로그램 사용자 : 화면 - Home(추천기사리스트), 최신기사 리스트, 날짜별 관련기사 리스트 (최신순, 관련도순), 각 구단 트위터정보, 키워드 검색, 키워드 검색 결과(기사 리스트), 키워드 선택

### 2.2.3 기능요구조건 명세서

기능 번호	기능 설명
1.1	해외언론사 10개 페이지에서 언론사출처, 기사제목, 기사내용, URL, 이미지 URL, 기사 작성 날짜, 팀 이름 추출
1.2	트위터 페이지에서 각 구단 공식 트위터 및 구단관련 팬 페이지로부터 트윗 내용 및 트윗 작성 날짜 추출
1.3	기사와 트윗에서 추출한 내용을 데이터베이스에 저장
1.4	위의 3개의 과정을 30분마다 자동으로 수행
2.1	데이터베이스에서 추출한 기사들로부터 기사제목과 기사내용 가져오는 기능
2.2	기사제목과 기사내용들로부터 특정기준들로 필터링하여 키워드 선별
2.3	선별된 키워드를 데이터베이스에 저장
3.1	EPL 내의 20개 팀의 구단이름, 감독이름, 선수이름을 지정하여 키워드 형태로 저장
3.2	분석된 기사를 키워드별로 분류하여 데이터베이스에 저장
3.3	트위터에서 추출한 트윗 내용들을 저장
3.4	어플 내의 사용자 아이디 및 사용자 지정키워드 저장
3.5	어플 내의 사용자가 조회한 기사들을 저장
3.6	추천시스템을 통해 각 사용자별로 추천되는 기사를 저장
3.7	트위터에서 크롤링한 내용들로 클러스터링 과정을 거친 연관된 키워드 데이터 정보를 저장
4.1	데이터마이닝을 이용해 중복되는 기사를 필터링하여 제외하는 기능
5.1	키워드 조회수에 따른 키워드 가중치 재설정.
5.2	조회수가 많은 기사에서 키워드들을 종합한 다음, 많이 중복된 키워드 순으로 키워드 선정하여 선정된 키워드를 토대로 기사를 추천하는 전체추천기능
5.3	사용자내에서 같은 키워드를 가진 사용자의 그룹을 형성하여, 그룹 내에서 사용자들이 가진 다른 키워드들을 종합해 통계를 낸 다음, 그 중 가장 많이 중복된 키워드를 이용하여 기사를 추천하는 그룹추천 기능
5.4	사용자가 조회한 기사들의 키워드들 중에서 사용자가 지정한 키워드를 제외한 가장 많이 중복된 키워드를 이용하여 기사를 추천하는 개인추천 기능
6.1	크롤링 된 트위터 내용을 키워드 별로 클러스터링하여 키워드 별로 연관도를 나타내는 기능
7.1	Django를 통해 서버와 클라이언트를 연동하여 서버와 클라이언트 사이의 데이터 전송을 담당하는 기능
8.1	어플리케이션 내에서 사용자가 원하는 키워드(팀/감독/선수) 지정할 수 있게 하는 기능
8.2	사용자가 지정한 키워드 및 다양한 정보를 이용하여 사용자가 흥미를 느낄만한 뉴스를 추천뉴스로 선정하여 제공
8.3	사용자가 지정한 키워드가 제목 및 본문에 포함된 기사들을 최신순 및 관련도 순으로 기사로 제공
8.4	기사들이 최초 업로드된 시간을 기준으로 정렬하여 제공
8.5	각 팀별 최신 트위터 검색결과 제공
8.6	키워드 검색을 통해 검색된 기사목록을 제공

시스템	Crawling Automation System (크롤링)
요구조건	지정한 페이지의 소스를 Python내의 Requests, BeautifulSoup, GetOldTweet3 라이브러리를 이용하여 기사내용을 추출하고, 리눅스의 crontab을 이용해 크롤링 자동화
관련 기능번호	1.1 , 1.2 , 1.3 , 1.4
관련 DB	ArticleData
입력 데이터	해외언론사 10개, 각 구단 공식트위터 및 구단관련 팬페이지 트위터
출력 데이터	해외언론사 10개 페이지에서 언론사출처, 기사제목, 기사내용, URL, 이미지 URL, 기사작성날짜, 팀 이름. 트위터에서 트윗내용 및 트윗작성날짜
상세 설명	
Python의 Requests, BeautifulSoup4 라이브러리를 이용. 각 언론사 페이지마다 HTML 태그 구성이 다르기 때문에, 각 웹페이지마다 BeautifulSoup4 라이브러리의 크롤링 알고리즘을 다르게 적용해 클래스별로 적용. 트위터 크롤링은 Python의 GetOldTweet3 라이브러리를 이용해 수집. 수집한 데이터는 데이터베이스에 저장. 이 과정을 리눅스의 crontab을 이용해 스케줄러를 설정하여 30분마다 자동으로 수행.	

언론사 이름	URL 주소
BBC Sport	<a href="https://www.bbc.com/sport/football">https://www.bbc.com/sport/football</a>
DailyMail	<a href="https://www.dailymail.co.uk/sport/football">https://www.dailymail.co.uk/sport/football</a>
Telegraph	<a href="https://www.telegraph.co.uk/football">https://www.telegraph.co.uk/football</a>
The Sun	<a href="https://www.thesun.co.uk/sport/football">https://www.thesun.co.uk/sport/football</a>
Guardian	<a href="https://www.theguardian.com/football">https://www.theguardian.com/football</a>
Goal	<a href="https://www.goal.com/en-gb/news">https://www.goal.com/en-gb/news</a>
Skysports	<a href="https://www.skysports.com/football">https://www.skysports.com/football</a>
Express	<a href="https://www.express.co.uk/news/">https://www.express.co.uk/news/</a>
ESPN	<a href="https://www.espn.com/soccer/">https://www.espn.com/soccer/</a>
Mirror	<a href="https://www.mirror.co.uk/sport/football">https://www.mirror.co.uk/sport/football</a>

< 크롤링 가능한 해외 언론 사이트 URL >



종류	트위터 아이디
구단공식 트위터	@Arsenal, @AVFCOfficial, @AFCBournemouth, @OfficialBHAFc, @BurnleyOfficial, @ChelseaFC, @CPFC, @Everton, @LCFC, @LFC, @ManUtd, @ManCity, @NUFC, @NorwichCityFC, @SheffieldUnited, @SouthamptonFC, @SpursOfficial, @WatfordFC, @WestHam, @Wolves
아스날	@gunnerforever78, @ThrowbackAFC, @afcstuff, @GurjitAFC, @AFTVMedia, @VipArsenal, @GunnersNews2019
아스톤빌라	@avfcnews2019, @AVFC_News, @astonforza, @AVFCchats, @VillaTil1Die
본머스	@AFCB_Fanly, @FWPBournemouth
브라이튼	@FWPBrighton, @AlbionAnalytics, @BHaseagulls_com
번리	@Burnleysbest, @BurnleyFC_Com, @TheClaretsChat, @BXSport
첼시	@BestOfChelsea1, @CFCNewsReport, @BoysCfc, @ChelseaFarleh, @TheBlues_---
크리스탈 팰리스	@PalaceEaglesc0m, @AdvertiserCPFC, @CPFCBC
에버튼	@ReadEverton, @vitaleverton @EvertonNewsFeed
레스터시티	@lfcawaydays_, @FIRST4LCFC, @FoxesofLCFC
리버풀	@Liverpool_zy, @Liverpoolcom_, @AnfieldWatch, @iLiverpoolApp, @TheAnfieldWrap
맨시티	@City_Xtra, @ManCityMEN, @ManCityzenscom
맨유	@MUFCscoop, @UnitedStandMUFC, @FullTimeDEVILS
뉴캐슬	@FWPNewcastle, @NUFCCollective, @nufcnews2019
노리치	@TalkNorwichCity, @NCFC_--News
세필드	@The_Bladesman, @SheffieldUNews, @sufcdevelopment
사우스햄튼	@iSouthamptonApp, @ReadSouthampton
토트넘	@thespursweb, @News2019Thfc, @iSpursApp, @Spurs_Centre
왓포드	@watfordfclivehq, @FWPWatford, @WatfordFNH, @watfordsz
웨스트햄	@Sam_InkersoleTM, @WestHam_fl, @westham_gossip
울버햄튼	@Wolves_TT1, @FWPWolves, @LiveWolvesNews, @WolvesTalk
EPL	@premierleague, @OfficialFPL, @Premiereleague7

< 크롤링 가능한 트위터 아이디 목록 >

시스템	Article Analysis System (데이터 마이닝)
요구조건	R에서 DB를 연결하여 DB의 기사로부터 필터링을 통한 키워드 추출 후 빈도순으로 정렬하여 키워드 아이디를 DB의 기사 키워드에 저장
관련 기능번호	2.1 , 2.2 , 2.3
관련 DB	ArticleData, Keyword
입력 데이터	크롤링된 기사내용
출력 데이터	각 기사마다 선별된 키워드(최대 10개)
상세 설명	
<p>각 기사마다 데이터 마이닝을 통해 선별된 키워드의 키워드 아이디 최대 10개까지 저장에 사용. 기사의 제목과 내용 가져옴.</p> <p>(1) 4글자 이상 대명사 (2) 불용어 제거 (3) 특수문자 제거 (4) 특정 키워드 제거의 4가지 키워드 필터링 기준으로 키워드 추출. 제목+첫 문장/제목+내용 필터링한 키워드들을 빈도수로 정렬하고 제목+첫 문장 키워드의 우선 순위가 높도록 리스트로 병합. Keyword 테이블에 있는 키워드들로 선별하여 키워드 아이디를 기사 키워드에 순서대로 저장.</p>	

시스템	Data Management System (데이터베이스)
요구조건	Crawling System, Article Analysis System, Recommendation System, Application User Interface System에서 사용되는 데이터를 저장.
관련 기능번호	3.1 , 3.2 , 3.3 , 3.4 , 3.5 , 3.6 , 3.7
관련 DB	모든 데이터베이스 테이블
입력 데이터	크롤링 시스템, 데이터 마이닝, 중복기사 제거, 트위터 클러스터링, 사용자 인터페이스
출력 데이터	사용자 인터페이스
상세 설명	
1. 지정키워드 : 팀이름 20개, 감독이름 20개, 선수이름 506개, 언론사이름 10개로 총 556개 키워드 저장. 2. Article Analysis System에서 추출한 키워드 저장. 3. 어플내의 사용자 고유ID값 및 사용자 지정키워드 저장. 4. 어플내의 사용자가 조회한 기사들을 저장. 5. 추천시스템을 통해 각 사용자별로 추천되는 기사를 저장. 6. 트위터에서 크롤링한 내용들로 클러스터링 과정을 거친 연관된 키워드 데이터 정보를 저장.	

시스템	Article Redundancy Filtration System (중복기사 제거)
요구조건	R 언어로 기사들의 제목과 첫 문장을 비교하여 높은 일치율(중복도)를 나타내는 기사 필터링.
관련 기능번호	4.1
관련 DB	ArticleData
입력 데이터	크롤링된 기사 제목과 내용
출력 데이터	중복된 기사 제거
상세 설명	
여러 언론 사이트에서 크롤링을 하는 과정에서 중복되는 내용의 기사를 필터링하기 위해 사용. 날짜별 기사 정렬 후 기사 제목끼리 비교하여 일치율이 80% 이상 일치 시 기사 본문의 첫 문장끼리 비교하고, 80% 이상 일치하는지 확인 후 두 기사 중 최근이 아닌 기사를 제거	

```
Q<-"Southampton 2 Watford 1 James Ward Prowse secures
fightback win for Saints after stunning free kick"
sQ<-"Southampton 2 1 Watford James Ward Prowse free kick fires
Saints to precious win"
```

```
> per
[1] 0.8
```

<80%의 일치율을 보이는 두 기사의 제목(예시)>

시스템	User Recommendation System (추천시스템)
요구조건	데이터베이스내의 사용자 정보를 이용하여 키워드를 선정하고 선정된 키워드의 기사 정보를 이용하여 사용자에게 가장 흥미가 있을만한 기사를 선택.
관련 기능번호	5.1 , 5.2 , 5.3 , 5.4
관련 DB	UserData, ArticleData, RecNews, NewsReg, ShareData
입력 데이터	사용자 데이터, 기사 데이터
출력 데이터	각 사용자마다 추천키워드 및 선정된 추천기사
상세 설명	
<p>추천방식은 전체추천, 그룹추천, 개인추천이 있다. 전체추천은 키워드 구분 없이 축구 팬들이라면 누구나 관심 있을만한 기사를 추천. 그룹추천은 키워드별 그룹데이터를, 개인 추천은 개인 데이터를 이용하여 기사를 추천. 첫 사용자의 추천 방식은 개인추천을 제외한 전체추천과 그룹추천을 이용하여 기사를 추천.</p>	

시스템	Twitter Clustering System (트위터 클러스터링)
요구조건	크롤링된 트위터 내용을 유저 키워드 별로 클러스터링하여 연관도와 연관 키워드를 추출, 연관 키워드가 포함된 기사에 클러스터링 점수 부여.
관련 기능번호	6.1
관련 DB	Twitter, Keyword, ArticleData, UserData, ShareData
입력 데이터	크롤링된 트위터 내용, 유저 키워드
출력 데이터	키워드별 연관 키워드에 따른 추가점수
상세 설명	
<p>크롤링된 트위터를 기반으로 기사에 클러스터링 점수 부여에 사용. 트위터 내용 매트릭스화하여 유저 키워드와 클러스터링. 연관도순으로 정렬하고 정제과정을 거쳐 5개의 연관 키워드 추출. 순서대로 기사 내용에 해당 연관 키워드 존재 시 클러스터링 점수(5점부터 1점)를 저장.</p>	

시스템	Server to Client Connect System
요구조건	Django를 통해 서버와 클라이언트를 연동하여 서버와 클라이언트 사이의 데이터 전송을 담당.
관련 기능번호	7.1
관련 DB	-
입력 데이터	서버로의 요청
출력 데이터	요청에 대한 결과
상세 설명	
<p>서버: Django의 REST Api. 클라이언트: Android의 Retrofit2 라이브러리 이용. HTTP 요청 및 응답을 이용.</p>	

시스템	Application User Interface System
요구조건	Android Studio를 이용해 구현. 안드로이드 어플리케이션 UI 형태로 제공.
관련 기능번호	8.1 , 8.2 , 8.3 , 8.4 , 8.5 , 8.6
관련 DB	UserData, Rec_News, Keyword, News_Reg, ArticleData
입력 데이터	사용자 입력
출력 데이터	사용자 입력에 따른 결과
상세 설명	
Server to Client Connect System를 이용해 추천기사 목록, 최신기사 목록 및 키워드관련기사 목록, 사용자 지정키워드 저장 기능, 팀별 트위터 화면, 키워드를 통한 기사 검색기능 제공.	

#### 2.2.4 성능요구조건 명세서

시스템	요구조건
Database Management System	최대 1000명의 User 데이터를 저장할 수 있는 저장 공간이 필요하고 사진을 포함한 기사 1개당 1MB일 때, 10개의 해외 언론, 하루에 100개의 기사가 나올 경우 하루에 대략 1GB 가 필요하므로 2주 ~ 1 달 동안의 기사를 저장하기 위해서는 약 50GB 의 데이터 저장 공간이 필요함.
Crawling Update Cycle	일반적으로 기사와 기사간의 업데이트 시간차이가 최소 10분이기 때문에 서버 데이터베이스에 기사를 업데이트하는 주기를 30분으로 예상. 트위터 크롤링 업데이트 주기 24시간으로 예상.
Article Analysis System	R프로그래밍에서 DB에서 기사 속 키워드 추출 후 빈도수로 정렬한 키워드 id 다시 DB에 저장하는 데 걸리는 시간 3초 이내.
Recommendation System	기사의 추천 점수에 따른 가중치부여 및 각 사용자별 추천키워드 및 추천기사 선정 주기는 12시간. 추가로 크롤링 업데이트 주기(30분) 내에 추천시스템이 원활하게 돌아가면 됨.
Application User Interface System	LTE기준 어플리케이션과 서버와의 통신 속도는 2~5초이고 ESPN 어플리케이션 접근 시 평균적으로 6~8초이기 때문에 일반적인 핸드폰 기기인 G7 안드로이드 9.0 파이, 4GB RAM, 옥타코어(2.7GHz + 1.7GHz) 기준으로 최초 업데이트 혹은 새로고침 시 5 sec 이내 업데이트를 목표로 함.
Server to Client Connect System	동시 접속자 수 100명에 대해 최대 응답시간 5초 이내로 서비스 제공.

## 2 설계

### 3.1 DB 설계

#### 3.1.1 테이블 구성

테이블명	UserData: 사용자데이터 정보					
이름	자료형	자료길이	필수	키	실례	비고
user_id	VARCHAR	100	Y	primary key	9f2de55245824a0	사용자 아이디
user_keyword (1~3)	INTEGER	-	-	foreign key	506/514/414	사용자 지정 키워드 3개
rec_keyword (1~6)	INTEGER	-	-	foreign key	522/537 ...	사용자 추천 키워드 6개
rec_percent	FLOAT	-	-	-	0.65	추천확률

테이블명	Rec_News: 추천뉴스					
이름	자료형	자료길이	필수	키	실례	비고
reco_id	INTEGER	-	Y	primary key	2034	추천뉴스 고유번호
user_id	VARCHAR	100	-	foreign key	9f2de55245824a0	사용자 아이디
news_id	INTEGER	-	-	foreign key	6840	뉴스 아이디
rec_reg	INTEGER	-	-	-	0 아니면 1	추천뉴스 안읽었으면 0, 읽었으면 1
rec_point	INTEGER	-	-	-	6	추천점수
date	DATETIME	-	-	-	2019-12-03 17:16:23	추천날짜

테이블명	Keyword: 키워드					
이름	자료형	자료길이	필수	키	실례	비고
keyword_no	INTEGER	-	Y	primary key	9	팀/선수/감독 키워드 556개
team	VARCHAR	45	-	-	Arsenal	팀 이름
first_name	VARCHAR	45	-	-	Kieran	선수/감독의 성
last_name	VARCHAR	45	-	-	Tierney	선수/감독의 이름
player_no	INTEGER	-	-	-	3	선수 등번호
view_no	INTEGER	-	-	-	97	키워드 조회수
weight	FLOAT	-	-	-	11.4962	키워드 가중치
full_name	VARCHAR	100	-	-	Kieran Tierney	선수/감독의 전체이름

테이블명	News_Reg: 사용자가 조회한 기사					
이름	자료형	자료길이	필수	키	실례	비고
reg_id	INTEGER	-	Y	primary key	10	고유번호
user_id	VARCHAR	100	-	foreign key	9f2de55245824a0	사용자 아이디
news_id	INTEGER	-	-	foreign key	9484	뉴스 아이디
date	DATETIME	-	-	-	2019-11-30 17:49:59	조회한 날짜

테이블명	ShareData: 추천뉴스 추가점수 부여					
이름	자료형	자료길이	필수	키	실례	비고
shared_id	INTEGER	-	Y	primary key	67651	고유번호
keyword_id	VARCHAR	45	-	-	514	키워드 아이디
news_id	VARCHAR	45	-	-	7788	뉴스 아이디
rec_point	INTEGER	-	-	-	4	추가점수

테이블명	ArticleData: 기사정보					
이름	자료형	자료길이	필수	키	실례	비고
news_id	INTEGER	-	Y	primary key	4303	기사 고유번호
title	VARCHAR	400	-	-	기사제목	기사제목
contents	LONGTEXT	-	-	-	기사내용	기사내용
source	VARCHAR	30	-	-	Skysports	언론사출처
date	DATETIME	-	-	-	2019-11-18 10:06:47	기사 작성날짜
image_URL	VARCHAR	400	-	-	http:// ~	기사의 이미지 URL
keyword(1~10 )	INTEGER	-	-	foreign_key	13/506 ...	선별된 키워드 10개
source_team	VARCHAR	30	-	-	Chelsea	기사 관련 팀이름
URL	VARCHAR	200	-	-	http:// ~	기사 URL
click_no	INTEGER	-	-	-	10	기사 클릭수(조회수)
rec_score	FLOAT	-	-	-	48.083	추천점수
rec_percent	FLOAT	-	-	-	32.56	추천확률
rec_count	INTEGER	-	-	-	52	추천횟수

테이블명	Twitter: 크롤링된 트위터 정보					
이름	자료형	자료길이	필수	키	실례	비고
twit_id	INTEGER	-	Y	primary key	1	고유번호
team	VARCHAR	45	-	-	Arsenal	트윗 출처 팀이름
contents	VARCHAR	45	-	-	트윗 내용	트윗 내용
date	DATETIME	-	-	-	2019-12-01 13:05:23	트윗 작성날짜

## 3.2 모듈 설계

### 3.2.1 크롤링 설계

모듈명	기사 크롤링	
설 명	해외언론사 10개 페이지에서 언론사출처, 기사제목, 기사내용, URL, 이미지URL, 기사작성 날짜, 팀 이름 추출 후 데이터베이스에 저장.	
관련 라이브러리	Python의 'Requests' , 'Beautiful Soup4'	
입력 데이터	언론사 URL	
출력 데이터	각 기사의 언론사이름, 기사제목, 기사작성날짜, 기사내용, URL, 이미지URL, 팀 이름	
단 계	처 리 내 역	
1	해외언론사 10개 사이트의 각 팀별 기사를 정렬해놓은 URL 입력.	
2	팀별 기사목록의 HTML코드에서 각 기사의 URL 추출.	
3	첫 기사부터 데이터베이스에 저장되어 있지 않은 기사 선별(URL로 구분)	
4	각 언론사에서 중복된 기사일 경우 제외 (팀별 기사를 추출하므로 같은 기사가 서로 다른 팀에서 나타남)	
5	각 기사의 HTML코드에서 언론사이름, 기사제목, 기사작성날짜, 기사내용, URL, 이미지URL, 팀 이름을 추출 후 리스트로 객체화	
6	가장 최신 크롤링된 기사 URL 텍스트파일로 저장(3단계에서 활용)	
7	각 기사의 정보를 담은 객체를 팀별로 리스트화한 후 DB에 저장	

모듈명	트위터 크롤링	
설 명	트위터 페이지에서 각 구단 공식 트위터 및 구단관련 팬 페이지로부터 트윗내용 및 트윗 작성날짜 추출	
관련 라이브러리	Python의 'GetOldTweet3', 'pymysql', 'multiprocessingpool'	
입력 데이터	각 구단 팀페이지 및 팬페이지 트위터 아이디	
출력 데이터	트윗내용 및 트윗 작성날짜	
단 계	처 리 내 역	
1	현재시간 및 크롤링 기간 3주로 설정	
2	각 팀 페이지, 팬 페이지 아이디를 Twitter API(GetOldTweets3)로 실행	
3	3주간의 한 아이디의 Tweet 정보 중 내용이 10글자 미만일 경우 제외	
4	Tweet 정보의 트윗 내용, 트윗 날짜 추출한 후 Date 객체를 정제	
5	추출된 Tweet 정보를 리스트화하여 최종 데이터에 리스트로 저장	
6	최종 데이터를 DB에 저장, 각 ID에 대한 API 실행은 멀티프로세싱을 활용	



### 3.2.2 데이터 마이닝 설계

모듈명	데이터 마이닝
설 명	R에서 DB를 연결하여 DB의 기사로부터 필터링을 통한 키워드 추출 후 빈도순으로 정렬하여 키워드 아이디를 DB의 기사 키워드에 저장
관련 라이브러리	R의 'KoNLP', 'tm', 'dplyr', 'stringr', 'DBI', 'RMySQL', 'dbConnect' Library
입력 데이터	기사 아이디, 제목, 내용, 키워드 아이디, full name
출력 데이터	키워드 아이디를 순서대로 기사 키워드 1-10에 저장하는 쿼리문
단 계	처 리 내 역
1	dbConnect와 dbGetQuery를 통해 MySQL DB의 articledata 테이블에서 기사를 가져옴
2	기사 아이디로 각 기사의 제목과 내용을 불러옴(아래 단계는 루프로 수행)
3	각 기사의 제목과 기사 본문의 첫 문장을 추출함
4	불용어, 특수문자, 특정 키워드를 제거한 4글자 이상의 대명사 키워드 추출(필터링)
5	추출한 키워드를 기사 속 빈도수로 정렬하여 리스트화
6	각 기사의 본문으로 4, 5단계 수행 후 리스트 밑으로 합침(제목과 첫 문장의 키워드 우선순위가 더 높게 하기 위함)
7	DB의 keyword 테이블의 키워드 이름과 리스트 속 키워드를 비교, 존재할 경우 키워드 아이디를 리스트에 저장
8	dbSendQuery로 키워드 아이디를 DB의 기사 키워드 1부터 10까지 리스트에 정렬된 순서대로 저장

### 3.2.3 중복기사 필터링 설계

모듈명	중복 기사 필터링
설 명	R에서 DB의 기사를 불러와 기사의 제목과 첫 문장의 일치율을 서로 비교하여 높은 일치율을 보일 경우(중복도가 높을 경우) 최근이 아닌 기사를 DB에서 제거
관련 라이브러리	R의 'dplyr', 'stringr', 'DBI', 'RMySQL', 'dbConnect' Library
입력 데이터	기사 아이디, 제목, 내용, 날짜
출력 데이터	해당 기사 DB에 제거 요청 쿼리문
단 계	처 리 내 역
1	dbConnect와 dbGetQuery를 통해 MySQL DB의 ArticleData 테이블에서 기사를 가져옴
2	특정 날짜 이후의 날짜별 기사 추출(최근에 크롤링된 기사 추출)
3	두 기사의 제목과 기사 본문의 첫 문장을 각각 추출함
4	두 기사의 제목끼리 일치율을 분석하여 80%이상 일치하는지 확인(단어로 끊어서 해당 단어가 비교 대상인 문장에 있는 횟수/문장의 단어 개수*100)
5	일치 시, 첫 문장끼리 비교하여 80%이상 일치하는지 확인
6	일치 시, 해당 기사의 날짜를 비교하여 상대적으로 최근이 아닌 기사의 아이디 추출
7	DB의 ArticleData에 해당 기사 제거를 요청하는 쿼리문을 보냄

### 3.2.4 트위터 클러스터링 설계

모듈명	트위터 클러스터링	
설 명	R에서 DB의 트위터와 키워드의 연관도 분석하여 연관 키워드가 기사에 존재할 경우 해당 기사에 클러스터링 점수 부여	
관련 라이브러리	R의 'dplyr', 'stringr', 'tm', 'NLP', 'openNLP', 'reticulate', 'DBI', 'RMySQL', 'dbConnect' Library	
입력 데이터	트위터 아이디, 팀, 내용, 키워드 아이디, 팀, first name, last name, 기사 아이디, 제목, 내용, 유저 아이디, 유저 키워드 1-3	
출력 데이터	키워드 아이디, 기사 아이디, 클러스터링 점수 ShareData 테이블에 저장	
단 계	처 리 내 역	
1	dbConnect와 dbGetQuery를 통해 MySQL DB의 ArticleData, Keyword, Twitter, UserData 테이블에서 기사, 키워드, 트위터, 유저 정보를 가져옴	
2	트위터 아이디로 트위터 내용과 팀 추출	
3	tm 라이브러리로 트위터 내용을 데이터 프레임화하고 코퍼스를 생성하여 공백, 특수 문자, 숫자, 불용어를 제거하여 매트릭스 변환	
4	userdata 테이블에서 각 유저 아이디로 유저 키워드 3개 추출	
5	각각 유저 키워드내 키워드 아이디로 keyword 테이블의 팀과 first name, last name 추출	
6	findAssocs로 키워드의 last name과 트위터 매트릭스의 연관도 분석 및 연관 키워드 추출	
7	연관도가 높은 순으로 연관 키워드 정렬	
8	연관 키워드 명사화 후 first name, 팀 키워드 제외, 5글자이상 10글자 이하의 5개 연관 키워드 추출	
9	각 기사에 연관 키워드가 존재하는지 순서대로 확인	
10	해당 연관 키워드 존재 시 클러스터링 점수 5점 저장(5개 연관 키워드의 순서대로 5점부터 1점까지 저장)	
11	DB의 ShareData 테이블에 연관도를 분석했던 키워드의 아이디, 연관키워드가 존재하는 기사 아이디, 클러스터링 점수를 저장하는 쿼리문을 보냄	

### 3.2.5 추천시스템 설계

모듈명	추천 시스템 - 키워드 가중치 설정	
설 명	각 기사의 조회수를 반영하여 기사가 포함하는 키워드의 조회수를 증가시키고 키워드의 조회수를 활용하여 키워드 가중치 재설정	
관련 라이브러리	Python의 'pymysql'	
입력 데이터	기사데이터의 조회수	
출력 데이터	키워드의 가중치 값	
단 계	처 리 내 역	
1	DB로부터 기사데이터를 가져온 후 모든 기사로부터 조회수를 키워드 조회수에 반영	
2	키워드의 종류인 팀, 선수, 감독, 언론사 별로 조회수 합계를 구함	
3	각 키워드에 대해 ((키워드 조회수 / 조회수 합계) * 키워드 종류의 가중치 합)으로 가중치를 설정 후 키워드 테이블 가중치에 반영(키워드 종류의 가중치 합 = 팀 : 120, 선수 : 2530, 감독 : 140, 언론사 : 30)	

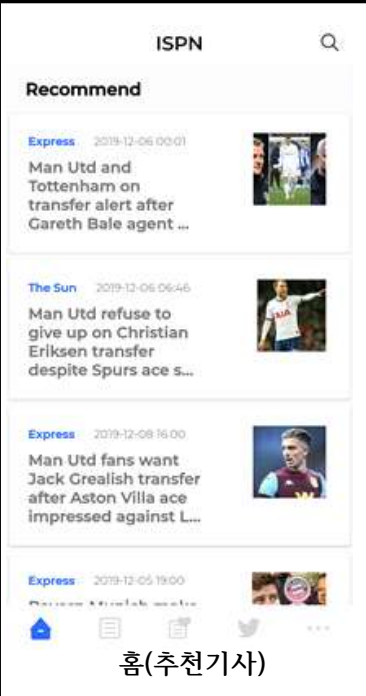
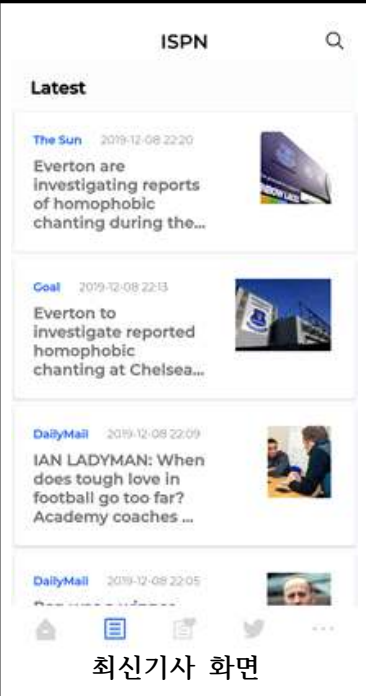
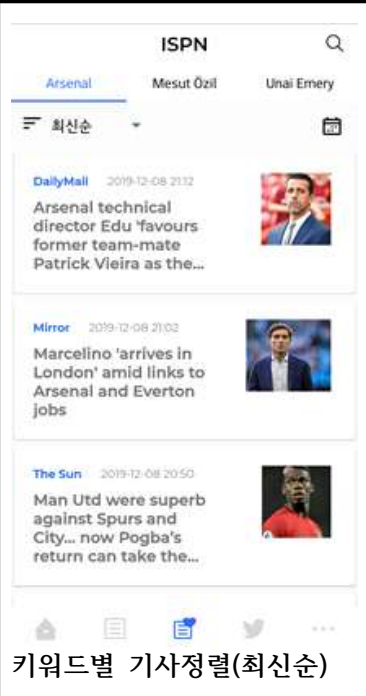
모듈명	추천 시스템 - 추천 점수 설정	
설 명	각 기사들의 데이터에서 클릭 수, 언론사, 날짜, 키워드, 이미지 유/무, 추천 확률을 이용하여 기사의 추천점수를 설정 및 재설정함	
관련 라이브러리	Python의 'pymysql'	
입력 데이터	기사데이터의 조회수	
출력 데이터	키워드의 가중치 값	
단 계	처 리 내 역	
1	DB로부터 기사데이터와 키워드데이터를 가져옴	
2	기사데이터의 클릭수를 이용하여 기사점수에 반영함( 3회의 조회수당 2점 추가)	
3	기사데이터의 언론사, 키워드5개를 이용하여 키워드 테이블의 언론사, 키워드가중치로부터 기사점수에 합산함	
4	기사데이터의 날짜 데이터를 반영하여 기사점수에 가감함. (6시간당 -1점이지만 3시간일 경우 -0.5점 1시간은 -0.166점으로 동점수가 나오지 않도록 설정)	
5	기사데이터에 이미지가 있을 경우 기사점수에 20점을 합산함	
6	5회 이상의 추천횟수일 경우 추천확률을 반영하여 점수를 가감함 (30%미만 -3점, 50%미만 -1점, 70% 미만 +1점, 100% 이하 +3점) 이후 기사데이터에 추천점수 반영	

모듈명	추천 시스템 - 추천 키워드 설정	
설 명	유저데이터로부터 그룹추천, 개인추천을 위한 추천 키워드를 설정함.	
관련 라이브러리	Python의 'pymysql'	
입력 데이터	유저테이블 데이터	
출력 데이터	각 유저의 추천키워드 6개	
단 계	처 리 내 역	
1	DB로부터 유저 테이블 데이터를 가져옴	
2	한 유저에 대해 같은 키워드를 가진 유저들을 종합하여 다른 키워드를 수집	
3	각 키워드를 정렬하여 높은 값을 가진 키워드 3개를 DB 추천키워드1,2,3에 저장	
4	DB로부터 기사기록 테이블에서 한 유저의 아이디 값을 가진 데이터를 가져옴	
5	기사기록의 테이블에서 얻은 news_id로 기사테이블에서 그 기사의 키워드들을 가져온 후 모든 데이터를 종합함	
6	종합한 키워드를 빈도순으로 정렬하고 상위 3개의 키워드를 DB 추천키워드 4,5,6에 저장	



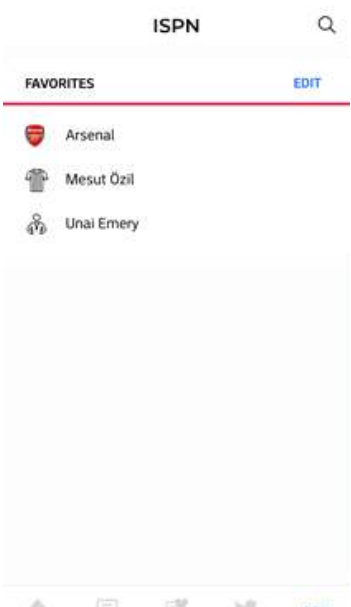
모듈명	추천 시스템 - 추천 기사 설정	
설 명	설정된 추천키워드를 이용하여 클러스터링 후 나온 결과의 ShareData 테이블을 이용하여 각 기사에 추가점수를 부여하고 15개의 추천 기사를 선정, 추천키워드하나당 2개의 기사를 선정하고 전체추천에서 3개의 기사를 선정함	
관련 라이브러리	Python의 'pymysql'	
입력 데이터	추천키워드 6개	
출력 데이터	추천기사 15개	
단 계	처 리 내 역	
1	DB로부터 유저데이터의 추천키워드와 ShareData테이블, 기사테이블을 가져옴	
2	유저에게 필요한 기사테이블과 ShareData 테이블을 유저 추천키워드에 맞게 선별	
3	각 ShareData의 news_id와 기사테이블의 news_id를 매칭 시키고 sharedata의 추가점수를 기사테이블의 추천점수에 추가로 부여함(이 때 시간복잡도를 최소화하기 위해 각 테이블을 news_id로 정렬한 후 서치시간을 최소화함)	
4	종합된 기사데이터를 추천점수 순으로 정렬하고 키워드 당 상위 2개의 기사를 선정함, 이 때 키워드가 없거나 기사가 2개 이하일 경우 전체추천기사에서 선정함	
5	각 추천키워드에 대해 추천기사를 설정한 후 전체추천기반으로 추천기사를 선정함 (이 때 3개의 전체추천 기사뿐 아니라 키워드에서 부족했던 기사도 할당하여 추천함)	
6	선정된 기사15개를 Rec_news table에 저장함	

### 3.3 사용자 인터페이스 설계

#### 3.3.1 세부 UI

사용자 인터페이스 세부 UI <1>		
 <p>ISP News App Screenshot: Recommend section. It displays a list of football news items with headlines like 'Man Utd and Tottenham on transfer alert after Gareth Bale agent...' and 'Man Utd refuse to give up on Christian Eriksen transfer despite Spurs ace s...'. The bottom navigation bar includes icons for home, list, and search.</p> <p>홈(추천기사)</p>	 <p>ISP News App Screenshot: Latest section. It displays a list of football news items with headlines like 'Everton are investigating reports of homophobic chanting during the...' and 'Everton to investigate reported homophobic chanting at Chelsea...'. The bottom navigation bar includes icons for home, list, and search.</p> <p>최신기사 화면</p>	 <p>ISP News App Screenshot: Keyword-based article sorting (Latest) section. It displays a list of football news items with headlines like 'Arsenal technical director Edu 'favours former team-mate Patrick Vieira as the...'' and 'Marcelino 'arrives in London' amid links to Arsenal and Everton jobs'. The bottom navigation bar includes icons for home, list, and search.</p> <p>키워드별 기사정렬(최신순)</p>
1. 전체추천, 그룹추천, 개인 추천 기사를 혼합한 추천 기사를 보여줌	1. 최신기사 화면	1. 지정 키워드별 기사들 보여줌
2. 각 기사별 기사제목, 이미지, 연관키워드 보여줌	2. 가장 최근 업데이트 된 순서대로 정렬	2. 원하는 날짜 선택 후 그 날짜의 기사들을 최신순으로 정렬
3. 아래로 스와이핑시 새로 고침	3. 기사 클릭 시 기사화면으로 이동	3. 좌우 스와이핑을 통해 이동
4. 기사 클릭 시 기사화면으로 이동	4. 검색 버튼 클릭 시 검색 화면으로 넘어감	4. 기사 클릭 시 기사화면으로 이동
5. 검색 버튼 클릭 시 검색 화면으로 넘어감	-	5. 검색 버튼 클릭 시 검색 화면으로 넘어감




사용자 인터페이스 세부 UI <2>

 <p>키워드별 기사정렬(관련도순)</p>	 <p>각 팀별 트위터</p>	 <p>설정</p>
1. 지정 키워드별 기사들 보여줌	1. 각 팀별 트윗을 최신순으로 제공	1. 사용자가 지정한 키워드를 출력
2. 원하는 날짜 선택 후 그 날짜의 기사들을 관련도순으로 정렬	2. 팀 선택 탭을 좌우 스와이핑을 통해 이동가능	2. Edit 버튼 클릭 시 키워드 선택화면으로 이동
3. 좌우 스와이핑을 통해 이동	3. 트위터 클릭 시 트위터 어플로 이동.(동영상 재생가능)	3. 검색 버튼 클릭 시 검색 화면으로 넘어감
4. 기사 클릭 시 기사화면으로 이동	4. 검색 버튼 클릭 시 검색 화면으로 넘어감	-
5. 검색 버튼 클릭 시 검색 화면으로 넘어감	-	-

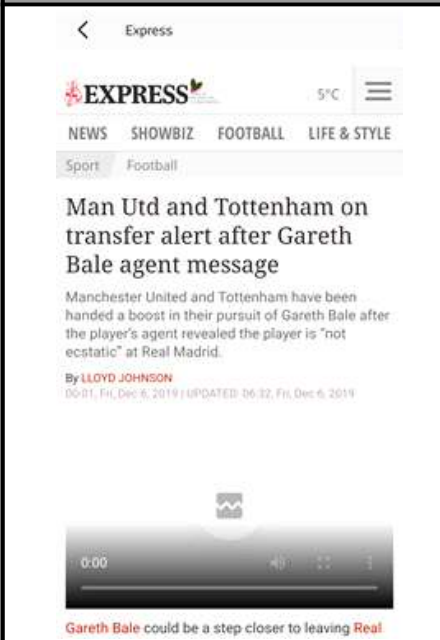
사용자 인터페이스 세부 UI <3>

 <p>키워드 선택&lt;1&gt;</p>	 <p>키워드 선택&lt;2&gt;</p>	 <p>키워드 선택&lt;3&gt;</p>
1. EPL 20개 팀 선택가능	1.각 팀 소속 선수 선택가능	1.각 팀 소속 감독 선택가능
2. 팀 선택 시 파랗게 표시됨	2.선수 선택시 FOLLOWING으로 버튼이 바뀜	2.감독 선택시 FOLLOWING으로 버튼이 바뀜
3. 상하 스와이핑을 통해 팀/선수/감독 키워드 선택화면으로 이동가능	3. 상하 스와이핑을 통해 팀/선수/감독 키워드 선택화면으로 이동가능	3. 상하 스와이핑을 통해 팀/선수/감독 키워드 선택화면으로 이동가능
4. FINISH 버튼 클릭 시 키워드 저장 후 설정화면으로 이동	4. FINISH 버튼 클릭시 키워드 저장 후 설정화면으로 이동	4. FINISH 버튼 클릭시 키워드 저장 후 설정화면으로 이동
5. 뒤로가기 버튼 클릭시 저장되지 않고 설정화면으로 이동	5. 뒤로가기 버튼 클릭시 저장되지 않고 설정화면으로 이동	5. 뒤로가기 버튼 클릭시 저장되지 않고 설정화면으로 이동

사용자 인터페이스 세부 UI <4>

 <p>검색</p>	 <p>검색결과</p>	 <p>기사내용</p>
1. 팀/선수/감독이름으로 검색가능	1. 검색된 결과 클릭시 보여주는 화면	1. 세부 기사내용 확인가능
2. 검색된 결과 클릭시 검색결과 화면으로 이동	2. 기사 클릭시 기사화면으로 이동	2. Read Article 버튼 클릭시 기사원문화면으로 이동
3. 검색결과 자동완성 제공	3. 뒤로가기 버튼 클릭시 홈 화면으로 이동	3. 뒤로가기 버튼 클릭시 기사클릭 전 화면으로 이동
4. X 버튼 클릭시 검색 초기화	-	4. 상하 스와이핑으로 통해 기사내용 확인



사용자 인터페이스 세부 UI <5>
 <p>기사원문</p>
1. 기사내용 원문으로 확인 가능
2. 뒤로가기 버튼 클릭시 기사내용 화면으로 이동
3. 상하 스와이핑으로 통해 기사내용 확인

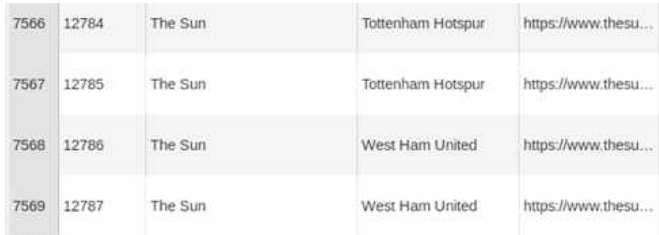
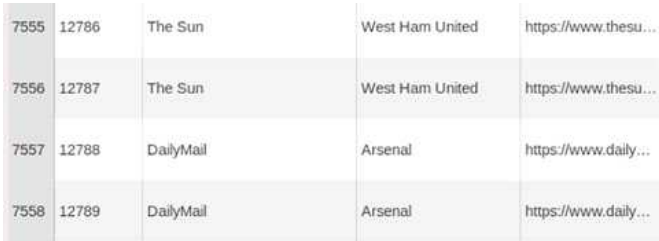

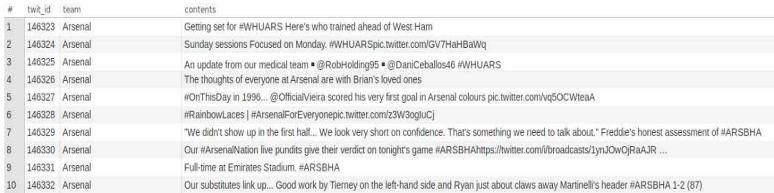
### 3.3.2 화면흐름도



## 4 연구결과

### 4.1 테스트케이스를 통한 요구조건 명세 목표달성 결과

#### 4.1.1 기능요구조건 목표달성 결과

시스템	Crawling Automation System	기능번호	1.1 , 1.3
기능	해외언론사에서 기사 추출 후 DB 에 저장하는 기능		
테스트항목	10 개의 해외 언론사		
테스트방법	해외기사 크롤링 모듈 실행		
기대결과	데이터베이스 기사테이블에 새로운 기사가 추가됨		
실제결과	 <p>&lt;크롤링 전 마지막기사&gt;</p>  <p>&lt;크롤링 완료 후 DB&gt;</p>		
시스템	Crawling Automation System	기능번호	1.2 , 1.3
기능	각 구단 공식 트위터 및 구단관련 팬페이지로부터 트윗 내용 및 트윗 작성 날짜 추출 후 DB 에 저장		
테스트항목	공식 트위터 및 구단관련 팬페이지의 3 주동안의 트윗내용		
테스트방법	트위터 크롤링 모듈 실행		
기대결과	데이터베이스 트위터테이블에 트위터 내용이 추가됨		
실제결과	 <p>&lt;트위터 크롤링 전 clear 된 DB&gt;</p>  <p>&lt;트위터 크롤링 완료 후 DB&gt;</p>		

시스템	Article Analysis System	기능번호	2.1 , 2.2 , 2.3																																															
기능	기사제목과 기사내용들로부터 특정기준들로 필터링하여 키워드 선별																																																	
테스트항목	키워드가 선정되지 않는 기사																																																	
테스트방법	키워드가 없는 기사에 대해 데이터마이닝 모듈 실행																																																	
기대결과	모든 기사에 대해 각각의 키워드가 설정된 상태로 기사테이블에 저장됨																																																	
실제결과	<table><tr><td>https://i.dailymail.co.uk/1s/2019/12/...</td><td>2019-12-08 22:05:25.0...</td><td>1</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr><tr><td>https://i.dailymail.co.uk/1s/2019/12/...</td><td>2019-12-08 22:09:48.0...</td><td>1</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr><tr><td>https://images.daznservices.com/di/...</td><td>2019-12-08 22:13:50.0...</td><td>1</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr></table>												https://i.dailymail.co.uk/1s/2019/12/...	2019-12-08 22:05:25.0...	1	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	https://i.dailymail.co.uk/1s/2019/12/...	2019-12-08 22:09:48.0...	1	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	https://images.daznservices.com/di/...	2019-12-08 22:13:50.0...	1	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL				
	https://i.dailymail.co.uk/1s/2019/12/...	2019-12-08 22:05:25.0...	1	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL																																						
	https://i.dailymail.co.uk/1s/2019/12/...	2019-12-08 22:09:48.0...	1	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL																																							
	https://images.daznservices.com/di/...	2019-12-08 22:13:50.0...	1	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL																																							
	<데이터 마이닝 전 DB>																																																	
<table><tr><td>https://i.dailymail.co.uk/1s/2019/12/...</td><td>2019-12-08 22:05:25.0...</td><td>1</td><td>507</td><td>516</td><td>519</td><td>527</td><td>371</td><td>355</td><td>513</td><td>514</td><td>357</td><td>413</td></tr><tr><td>https://i.dailymail.co.uk/1s/2019/12/...</td><td>2019-12-08 22:09:48.0...</td><td>5</td><td>240</td><td>515</td><td>511</td><td>517</td><td>514</td><td>423</td><td>510</td><td>531</td><td>159</td><td>404</td></tr><tr><td>https://images.daznservices.com/di/...</td><td>2019-12-08 22:13:50.0...</td><td>9</td><td>511</td><td>513</td><td>517</td><td>516</td><td>299</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr></table>												https://i.dailymail.co.uk/1s/2019/12/...	2019-12-08 22:05:25.0...	1	507	516	519	527	371	355	513	514	357	413	https://i.dailymail.co.uk/1s/2019/12/...	2019-12-08 22:09:48.0...	5	240	515	511	517	514	423	510	531	159	404	https://images.daznservices.com/di/...	2019-12-08 22:13:50.0...	9	511	513	517	516	299	NULL	NULL	NULL	NULL	NULL
https://i.dailymail.co.uk/1s/2019/12/...	2019-12-08 22:05:25.0...	1	507	516	519	527	371	355	513	514	357	413																																						
https://i.dailymail.co.uk/1s/2019/12/...	2019-12-08 22:09:48.0...	5	240	515	511	517	514	423	510	531	159	404																																						
https://images.daznservices.com/di/...	2019-12-08 22:13:50.0...	9	511	513	517	516	299	NULL	NULL	NULL	NULL	NULL																																						
<데이터 마이닝 완료후 DB>																																																		

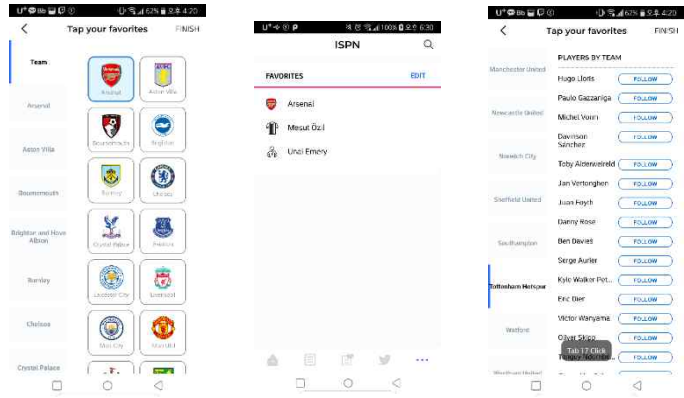
시스템	Article Redundancy Filtration System	기능번호	4.1																								
기능	데이터마이닝을 이용해 중복되는 기사를 필터링하여 제외하는 기능																										
테스트항목	임의의 기사를 중복하여 삽입																										
테스트방법	중복된 기사를 포함한 기사테이블에 대해 중복 제거 모듈 실행																										
기대결과	중복된 기사가 삭제됨																										
실제결과	<table><tr><td>7803</td><td>13164</td><td>The Sun</td><td>Chelsea</td><td>https://www.thesu...</td><td>Everton are investigating reports of homophobic chanting during their victory over Chelsea</td></tr><tr><td>7804</td><td>13165</td><td>The Sun</td><td>Chelsea</td><td>https://www.thesu...</td><td>Everton are investigating reports of homophobic chanting during their victory over Chelsea</td></tr><tr><td>7805</td><td>13166</td><td>The Sun</td><td>Chelsea</td><td>https://www.thesu...</td><td>Everton are investigating reports of homophobic chanting during their victory over Chelsea</td></tr><tr><td>*</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr></table>			7803	13164	The Sun	Chelsea	https://www.thesu...	Everton are investigating reports of homophobic chanting during their victory over Chelsea	7804	13165	The Sun	Chelsea	https://www.thesu...	Everton are investigating reports of homophobic chanting during their victory over Chelsea	7805	13166	The Sun	Chelsea	https://www.thesu...	Everton are investigating reports of homophobic chanting during their victory over Chelsea	*	NULL	NULL	NULL	NULL	NULL
	7803	13164	The Sun	Chelsea	https://www.thesu...	Everton are investigating reports of homophobic chanting during their victory over Chelsea																					
	7804	13165	The Sun	Chelsea	https://www.thesu...	Everton are investigating reports of homophobic chanting during their victory over Chelsea																					
7805	13166	The Sun	Chelsea	https://www.thesu...	Everton are investigating reports of homophobic chanting during their victory over Chelsea																						
*	NULL	NULL	NULL	NULL	NULL																						
	<p>&lt;중복된 기사&gt;</p> <table><tr><td>7803</td><td>13164</td><td>The Sun</td><td>Chelsea</td><td>https://www.thesu...</td><td>Everton are investigating reports of homophobic chanting during their victory over Chelsea</td></tr><tr><td>*</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr></table>			7803	13164	The Sun	Chelsea	https://www.thesu...	Everton are investigating reports of homophobic chanting during their victory over Chelsea	*	NULL	NULL	NULL	NULL	NULL												
7803	13164	The Sun	Chelsea	https://www.thesu...	Everton are investigating reports of homophobic chanting during their victory over Chelsea																						
*	NULL	NULL	NULL	NULL	NULL																						
	<p>&lt;중복 제거 완료&gt;</p>																										

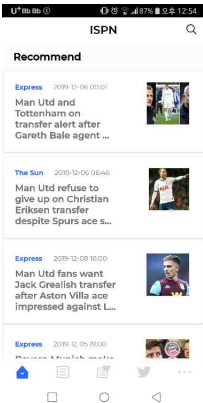
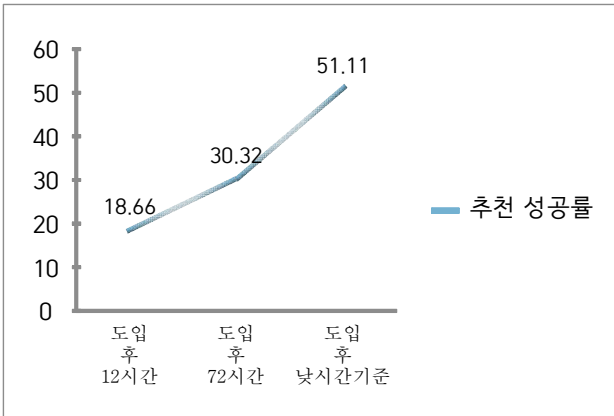
시스템	Twitter Clustering System	기능번호	6.1
기능	크롤링 된 트위터 내용을 키워드 별로 클러스터링 하여 키워드 별 연관도를 나타내는 기능		
테스트항목	크롤링한 트위터의 매트릭스와 “Arsenal” 또는 “Emeri” 키워드 클러스터링 후, 연관도와 연관 키워드가 적절한지 확인		
테스트방법	트위터 테이블을 이용하여 클러스터링 모듈 실행, 그래프 생성		
기대결과	Arsenal 과 Emeri 가 서로 관련 있으면서 의미 있는 키워드가 추출됨		
실제결과	<div></div> <p>&lt;Arsenal(붉은색) , Emeri(파란색) 의 연관도 그래프&gt;</p>		

시스템	User Recommendation System	기능번호	5.1, 5.2																																																
기능	각각의 기사에 추천점수를 부여하여 높은 추천점수의 기사를 추천하는 전체추천 기능																																																		
테스트항목	추천점수가 부여되지 않은 기사																																																		
테스트방법	추천점수부여 모듈 실행																																																		
기대결과	모든 기사에 대해 추천점수가 부여됨																																																		
실제결과	<div><div><p>&lt;추천점수 부여 전&gt;</p><table><thead><tr><th>keyword9</th><th>keyword10</th><th>rec_score</th></tr></thead><tbody><tr><td>NULL</td><td>NULL</td><td>65.2927</td></tr><tr><td>357</td><td>413</td><td>0</td></tr><tr><td>159</td><td>404</td><td>0</td></tr><tr><td>NULL</td><td>NULL</td><td>0</td></tr><tr><td>NULL</td><td>NULL</td><td>0</td></tr><tr><td>NULL</td><td>NULL</td><td>0</td></tr><tr><td>NULL</td><td>NULL</td><td>NULL</td></tr></tbody></table></div><div>➡</div><div><table><thead><tr><th>keyword9</th><th>keyword10</th><th>rec_score</th></tr></thead><tbody><tr><td>NULL</td><td>NULL</td><td>65.0716</td></tr><tr><td>357</td><td>413</td><td>64.1939</td></tr><tr><td>159</td><td>404</td><td>85.1096</td></tr><tr><td>NULL</td><td>NULL</td><td>66.5935</td></tr><tr><td>NULL</td><td>NULL</td><td>58.1271</td></tr><tr><td>NULL</td><td>NULL</td><td>66.6132</td></tr><tr><td>NULL</td><td>NULL</td><td>NULL</td></tr></tbody></table><p>&lt;추천점수 부여 완료&gt;</p></div></div>			keyword9	keyword10	rec_score	NULL	NULL	65.2927	357	413	0	159	404	0	NULL	NULL	0	NULL	NULL	0	NULL	NULL	0	NULL	NULL	NULL	keyword9	keyword10	rec_score	NULL	NULL	65.0716	357	413	64.1939	159	404	85.1096	NULL	NULL	66.5935	NULL	NULL	58.1271	NULL	NULL	66.6132	NULL	NULL	NULL
keyword9	keyword10	rec_score																																																	
NULL	NULL	65.2927																																																	
357	413	0																																																	
159	404	0																																																	
NULL	NULL	0																																																	
NULL	NULL	0																																																	
NULL	NULL	0																																																	
NULL	NULL	NULL																																																	
keyword9	keyword10	rec_score																																																	
NULL	NULL	65.0716																																																	
357	413	64.1939																																																	
159	404	85.1096																																																	
NULL	NULL	66.5935																																																	
NULL	NULL	58.1271																																																	
NULL	NULL	66.6132																																																	
NULL	NULL	NULL																																																	

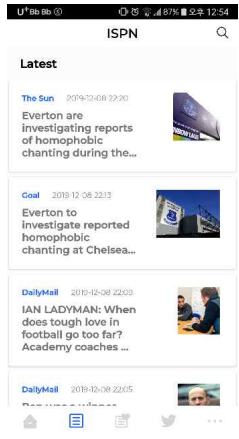
시스템	User Recommendation System	기능번호	5.3																																																	
기능	그룹 및 개인별로 추천 키워드 생성 및 기사 추천																																																			
테스트항목	추천 키워드와 추천기사가 없는 사용자																																																			
테스트방법	추천시스템 모듈 실행																																																			
기대결과	유저별로 추천 키워드 및 기사가 추천됨																																																			
실제결과	<table><tr><th>#</th><th>user_id</th><th>user_keyword1</th><th>user_keyword2</th><th>user_keyword3</th><th>rec_keyword1</th><th>rec_keyword2</th><th>rec_keyword3</th><th>rec_keyword4</th><th>rec_keyword5</th><th>rec_keyword6</th></tr><tr><td>1</td><td>707d066df22be3f</td><td>506</td><td>512</td><td>514</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>2</td><td>5a3c2dd2bc9a1e5c</td><td>522</td><td>424</td><td>509</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>3</td><td>818a1a5b554c7310</td><td>506</td><td>522</td><td>424</td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table>			#	user_id	user_keyword1	user_keyword2	user_keyword3	rec_keyword1	rec_keyword2	rec_keyword3	rec_keyword4	rec_keyword5	rec_keyword6	1	707d066df22be3f	506	512	514							2	5a3c2dd2bc9a1e5c	522	424	509							3	818a1a5b554c7310	506	522	424											
	#	user_id	user_keyword1	user_keyword2	user_keyword3	rec_keyword1	rec_keyword2	rec_keyword3	rec_keyword4	rec_keyword5	rec_keyword6																																									
	1	707d066df22be3f	506	512	514																																															
	2	5a3c2dd2bc9a1e5c	522	424	509																																															
3	818a1a5b554c7310	506	522	424																																																
	<p>&lt;추천키워드 선정 전&gt;</p> <table><tr><th>#</th><th>user_id</th><th>user_keyword1</th><th>user_keyword2</th><th>user_keyword3</th><th>rec_keyword1</th><th>rec_keyword2</th><th>rec_keyword3</th><th>rec_keyword4</th><th>rec_keyword5</th><th>rec_keyword6</th></tr><tr><td>1</td><td>707d066df22be3f</td><td>506</td><td>512</td><td>514</td><td>424</td><td>522</td><td>146</td><td>526</td><td>13</td><td>542</td></tr><tr><td>2</td><td>5a3c2dd2bc9a1e5c</td><td>522</td><td>424</td><td>509</td><td>506</td><td>87</td><td>376</td><td>517</td><td>515</td><td>514</td></tr><tr><td>3</td><td>818a1a5b554c7310</td><td>506</td><td>522</td><td>424</td><td>514</td><td>524</td><td>146</td><td>43</td><td>517</td><td>542</td></tr></table>			#	user_id	user_keyword1	user_keyword2	user_keyword3	rec_keyword1	rec_keyword2	rec_keyword3	rec_keyword4	rec_keyword5	rec_keyword6	1	707d066df22be3f	506	512	514	424	522	146	526	13	542	2	5a3c2dd2bc9a1e5c	522	424	509	506	87	376	517	515	514	3	818a1a5b554c7310	506	522	424	514	524	146	43	517	542					
#	user_id	user_keyword1	user_keyword2	user_keyword3	rec_keyword1	rec_keyword2	rec_keyword3	rec_keyword4	rec_keyword5	rec_keyword6																																										
1	707d066df22be3f	506	512	514	424	522	146	526	13	542																																										
2	5a3c2dd2bc9a1e5c	522	424	509	506	87	376	517	515	514																																										
3	818a1a5b554c7310	506	522	424	514	524	146	43	517	542																																										
	<p>&lt;추천키워드 선정 완료&gt;</p> <hr/>																																																			
	<table><tr><th>#</th><th>reco_id</th><th>user_id</th><th>news_id</th><th>rec_point</th><th>rec_reg</th><th>date</th></tr><tr><td>★</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr></table>			#	reco_id	user_id	news_id	rec_point	rec_reg	date	★	NULL	NULL	NULL	NULL	NULL	NULL																																			
#	reco_id	user_id	news_id	rec_point	rec_reg	date																																														
★	NULL	NULL	NULL	NULL	NULL	NULL																																														
	<p>&lt;추천기사 선정 전&gt;</p> <table><tr><th>#</th><th>reco_id</th><th>user_id</th><th>news_id</th><th>rec_point</th><th>rec_reg</th><th>date</th></tr><tr><td>1...</td><td>79475</td><td>9f2de55245824a08</td><td>10939</td><td>144</td><td>0</td><td>2019-12-09 08:30:44</td></tr><tr><td>1...</td><td>79476</td><td>9f2de55245824a08</td><td>13021</td><td>135</td><td>0</td><td>2019-12-09 08:30:44</td></tr><tr><td>1...</td><td>79477</td><td>9f2de55245824a08</td><td>11599</td><td>132</td><td>0</td><td>2019-12-09 08:30:44</td></tr><tr><td>1...</td><td>79478</td><td>9f2de55245824a08</td><td>6840</td><td>80</td><td>0</td><td>2019-12-09 08:30:44</td></tr><tr><td>1...</td><td>79479</td><td>9f2de55245824a08</td><td>11257</td><td>70</td><td>0</td><td>2019-12-09 08:30:44</td></tr><tr><td>1...</td><td>79480</td><td>9f2de55245824a08</td><td>12874</td><td>140</td><td>0</td><td>2019-12-09 08:30:44</td></tr></table>			#	reco_id	user_id	news_id	rec_point	rec_reg	date	1...	79475	9f2de55245824a08	10939	144	0	2019-12-09 08:30:44	1...	79476	9f2de55245824a08	13021	135	0	2019-12-09 08:30:44	1...	79477	9f2de55245824a08	11599	132	0	2019-12-09 08:30:44	1...	79478	9f2de55245824a08	6840	80	0	2019-12-09 08:30:44	1...	79479	9f2de55245824a08	11257	70	0	2019-12-09 08:30:44	1...	79480	9f2de55245824a08	12874	140	0	2019-12-09 08:30:44
#	reco_id	user_id	news_id	rec_point	rec_reg	date																																														
1...	79475	9f2de55245824a08	10939	144	0	2019-12-09 08:30:44																																														
1...	79476	9f2de55245824a08	13021	135	0	2019-12-09 08:30:44																																														
1...	79477	9f2de55245824a08	11599	132	0	2019-12-09 08:30:44																																														
1...	79478	9f2de55245824a08	6840	80	0	2019-12-09 08:30:44																																														
1...	79479	9f2de55245824a08	11257	70	0	2019-12-09 08:30:44																																														
1...	79480	9f2de55245824a08	12874	140	0	2019-12-09 08:30:44																																														
	<p>&lt;추천기사 선정 후&gt;</p>																																																			

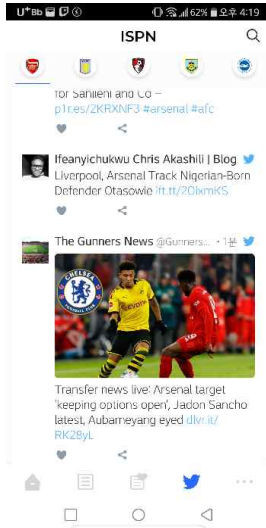
시스템	Server to Client Connect System	기능번호	5.4
기능	Django 를 통해 서버와 클라이언트를 연동하여 서버와 클라이언트 사이의 데이터 전송을 담당하는 기능		
테스트항목	클라이언트에서 서버에 대한 기사 데이터 요청		
테스트방법	어플리케이션에서 기사클릭		
기대결과	클릭한 기사에 대한 정보를 어플리케이션으로 볼 수 있음		
실제결과	<div>  <p>&lt;기사 목록&gt;</p>  <p>&lt;기사 클릭 후&gt;</p> </div>		

시스템	Application User Interface System	기능번호	8.1																											
기능	앱 내에서 사용자가 원하는 키워드 지정할 수 있게 하는 기능																													
테스트항목	팀, 선수, 감독 키워드 각각 하나씩 선정. 예) Arsenal, Mesut Ozil, Unai Emery																													
테스트방법	앱 내의 키워드 선택화면에서 키워드 선택																													
기대결과	선택된 키워드가 데이터베이스에 저장되고 클라이언트에 적용됨																													
실제결과																														
	<div>&lt;팀선택 화면&gt;   &lt;선수/감독 선택화면&gt;   &lt;앱 내의 키워드 저장결과&gt;</div> <table><thead><tr><th>#</th><th>user_id</th><th>user_keyword1</th><th>user_keyword2</th><th>user_keyword3</th><th>rec_keyword1</th><th>rec</th></tr></thead><tbody><tr><td>1</td><td>707d066df272be3f</td><td>506</td><td>512</td><td>514</td><td></td><td></td></tr><tr><td>2</td><td>5a3b2dd2bc9a1e5c</td><td>522</td><td>424</td><td>509</td><td></td><td></td></tr><tr><td>3</td><td>816a1a5b554c7310</td><td>506</td><td>522</td><td>424</td><td></td><td></td></tr></tbody></table> <div>&lt;사용자가 선택한 키워드 DB 저장 결과&gt;</div>			#	user_id	user_keyword1	user_keyword2	user_keyword3	rec_keyword1	rec	1	707d066df272be3f	506	512	514			2	5a3b2dd2bc9a1e5c	522	424	509			3	816a1a5b554c7310	506	522	424	
#	user_id	user_keyword1	user_keyword2	user_keyword3	rec_keyword1	rec																								
1	707d066df272be3f	506	512	514																										
2	5a3b2dd2bc9a1e5c	522	424	509																										
3	816a1a5b554c7310	506	522	424																										

시스템	Application User Interface System	기능번호	8.2								
기능	사용자가 지정한 키워드 및 다양한 정보를 이용하여 사용자가 흥미를 느낄만한 뉴스를 추천뉴스로 선정하여 제공										
테스트항목	각 사용자마다 12 시간마다 업데이트되는 추천뉴스 15 개										
테스트방법	사용자 5 명을 3 일동안 유저테스트 진행										
기대결과	추천기사 화면 제공, 추천 기사에 대한 추천성공률이 50%를 넘음										
실제결과											
	<p>&lt;추천기사 화면&gt;</p>										
	 <table border="1"><caption>추천 성공률 그래프 데이터</caption><thead><tr><th>도입 후 시간</th><th>추천 성공률 (%)</th></tr></thead><tbody><tr><td>도입 후 12시간</td><td>18.66</td></tr><tr><td>도입 후 72시간</td><td>30.32</td></tr><tr><td>도입 후 낮시간기준</td><td>51.11</td></tr></tbody></table> <p>— 추천 성공률</p> <p>&lt;추천성공률&gt;</p>			도입 후 시간	추천 성공률 (%)	도입 후 12시간	18.66	도입 후 72시간	30.32	도입 후 낮시간기준	51.11
도입 후 시간	추천 성공률 (%)										
도입 후 12시간	18.66										
도입 후 72시간	30.32										
도입 후 낮시간기준	51.11										

시스템	Application User Interface System	기능번호	8.3
기능	사용자가 지정한 키워드가 제목 및 본문에 포함된 기사들을 최신순 및 관련도순으로 기사로 제공		
테스트항목	지정한 키워드를 토대로 제공되는 기사		
테스트방법	키워드 설정 후 키워드별 기사 정렬 화면 확인		
기대결과	키워드별 기사 정렬 화면에서 키워드 관련 기사를 날짜별 최신순, 관련도순으로 출력		
실제결과	 <p>&lt;최신순 정렬화면&gt;                      &lt;관련도순 정렬화면&gt;                      &lt;날짜선택&gt;</p>		

시스템	Application User Interface System	기능번호	8.4
기능	기사들이 최초 업로드된 시간을 기준으로 정렬하여 제공		
테스트항목	날짜순으로 제공되는 최신기사		
테스트방법	최신기사 화면 확인		
기대결과	최신 기사 화면에서 최신순으로 기사가 출력		
실제결과	 <p>&lt;최신기사 화면&gt;</p>		

시스템	Application User Interface System	기능번호	8.5
기능	팀별 트위터 검색결과 제공		
테스트항목	팀별 트위터 검색결과		
테스트방법	트위터 액티비티 확인		
기대결과	트위터 액티비티에서 팀별 트위터 검색결과 출력		
실제결과	 <p>&lt;팀별 트위터 검색결과 화면&gt;</p>		

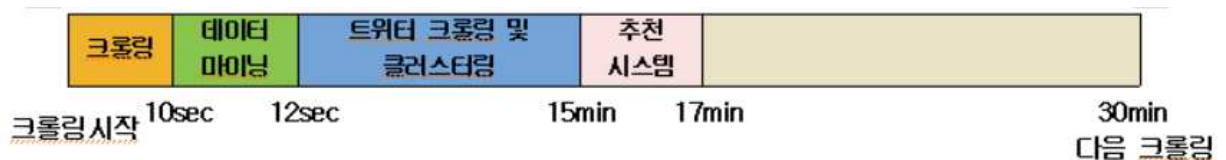
시스템	Application User Interface System	기능번호	8.6
기능	키워드 검색을 통해 검색된 기사목록을 제공		
테스트항목	키워드 검색결과 예) Jamie Vardy 키워드 검색		
테스트방법	검색화면 및 검색결과 액티비티 확인		
기대결과	키워드 검색화면에서 검색한 키워드로 기사목록 제공		
실제결과	 <p>&lt;키워드 검색화면&gt;      &lt;검색결과 화면&gt;</p>		



시스템	성능목표치	달성 여부	목표달성방법
Database Management	최대 1000명의 User 데이터를 저장할 수 있는 저장 공간이 필요하고 사진을 포함한 기사 1개당 1MB일 때, 10개의 해외 언론, 하루에 100개의 기사가 나올 경우 하루에 대략 1GB가 필요하므로 2주 ~ 1 달 동안의 기사를 저장하기 위해서는 약 50GB의 데이터 저장 공간이 필요함.	O	기사 데이터의 용량을 다수 차지하는 사진 데이터는 URL을 통해 받아오므로 상당수의 용량 절약이 가능함. 가동하는 서버컴퓨터의 CPU는 Intel i5-8400 6코어 프로세서, 6 쓰레드이며, RAM은 16GB, SSD 500GB이다. 현재 SSD 500GB중 OS포함 프로그램 및 데이터 전체 용량 70GB 즉 14%의 용량을 차지하고 있다. 또한 DB에 하루에 최소 185개, 최대 413개로 평균 302개의 기사가 저장되어 약 4300개의 기사(2주차분량)와 약 4만개의 트위터(3주차분량), 556개의 키워드로 총 80MB가 저장되어 있으며, 따라서 2주 ~ 1달 이상의 기사를 저장하기 충분함.
Crawling Update Cycle	일반적으로 기사와 기사간의 업데이트 시간차이가 최소 10분이기 때문에 서버 데이터베이스에 기사를 업데이트 하는 주기를 30분으로 예상. 트위터 업데이트 주기 24시간으로 예상.	O	크롤링 데이터는 3주전까지의 기사 데이터를 크롤링하여 DB에 저장함. 트위터 3주차 크롤링시 10분 내외의 시간이 소요되기 때문에 24시간 주기로 트위터 크롤링시 28sec 소요. 하루에 최소 185개, 최대 413개로 평균 302개의 기사나오고, 이를 크롤링하는데 최소 22sec ~ 최대 242sec 이고, 한시간분량의 기사를 크롤링하는데 최소 0.9sec ~ 최대 10sec로 30분 마다 기사를 문제없이 크롤링 가능.
Article Analysis System	R프로그래밍에서 DB에서 기사 속 키워드 추출 후 빈도수로 정렬한 키워드 id 다시 DB에 저장하는 데 걸리는 시간 3초이내.	O	한 개의 기사를 데이터마이닝 하는데 0.12 sec가 걸리고, 따라서 하루동안 평균 302개의 기사를 데이터마이닝 하는데 38 sec, 한 시간동안의 기사를 데이터마이닝 하는데 1.59 sec가 소요되기 때문에 문제없이 데이터마이닝 가능.
Recommendation System	기사의 추천 점수에 따른 가중치부여 및 각 사용자별 추천키워드 및 추천기사 선정 주기는 12시간. 추가로 크롤링 업데이트 주기(30분) 내에 추천시스템이 원활하게 돌아가면됨.	O	기사 추천점수 및 가중치 부여 시간: 기사 7000개에 대해 44.47 sec소요 사용자 1명에 대해 추천키워드 선정 및 추천기사 선정에 걸리는 시간: 0.28 sec소요 사용자 100명에 대해 추천기사 선정에 걸리는 시간: 74.83 sec소요

#### 4.1.2 성능요구조건 목표달성 결과

시스템	성능목표치	달성 여부	목표달성방법
Application User Interface System	LTE기준 어플리케이션과 서버와의 통신 속도는 2~5초이고 ESPN 어플리케이션 접근시 평균적으로 6~8초이기 때문에 일반적인 핸드폰 기기인 G7 안드로이드 9.0 파이, 4GB RAM, 옥타코어(2.7GHz + 1.7GHz) 기준으로 최초 업데이트 혹은 새로고침 시 5 sec 이내 업데이트를 목표로 함.	O	어플리케이션 실행시 약 2초내외, 추천기사와 최신기사 업데이트 약 1초, 키워드 변경 후 최신순, 관련도순으로 정렬된 기사 업데이트에 약 1초, 트위터 새로고침시 약 2초소요.
Server to Client Connect System	동시 접속자 수 100명에 대해 최대 응답시간 5초이내로 서비스제공.	O	가상사용자를 만들어 테스트 하기위해 jmeter 프로그램 이용. 1초 내에 100명의 사용자가 각각 추천기사와 최신기사 총 45개의 기사를 가져오는 작업을 50번 반복하는 충분한 테스트환경 적용. 평균 응답시간: 461(ms) = 0.46초, 최소 응답시간: 12(ms) = 0.012초, 최대응답시간: 1930(ms) = 1.93초, 에러확률 0%의 결과도출. 동시접속자 100명에 대해 문제없이 서비스제공 가능.



※ 사용자 100명의 100번의 평균

<기사 업데이트 주기(30분)내에 시스템 동작 순서>

## 4.2 추천시스템 평가

### 4.2.1 기존 추천시스템과의 비교

ISPN의 추천 시스템은 크게 전체 추천, 그룹 추천, 개인 추천으로 나뉘며, 각 추천에 따른 기사를 모아 사용자에게 추천 기사로 제공한다.

카카오 루빅스 추천 시스템은 내용기반 필터링과 협업 필터링, 멀티암드밴딧을 혼합하여 콘텐츠를 추천하고, Netflix, Spotify 등의 기업에서는 협업 필터링 중 하나인 모델 기반 협력 필터링으로 사용자가 소비한 콘텐츠를 기반으로 선호하는 콘텐츠를 유추한다.

ISPN은 멀티암드밴딧을 전체추천에, 협업 필터링을 그룹추천에, 내용기반 필터링을 개인추천에 반영하여 추천시스템을 개발하였다.

ISPN은 루빅스의 멀티암드밴딧을 이용하기 때문에 이용자 행동 정보가 거의 없는 콜드 스타트 이용자들의 기사 소비 패턴을 분석해 기사를 제공할 수 있다. 또한 내용기반 필터링 및 협업기반 필터링 개념을 이용해 유저에게 선호하는 키워드를 설정하게끔 함으로써 더욱 맞춤화 된 추천 기사를 제공할 수 있다.

또한 기존의 뉴스 추천시스템들은 스포츠, 정치, 경제, 연예 등 전반적인 분야에 걸쳐 추천 뉴스를 제공하는 반면, ISPN은 해외축구에 초점을 맞춰서 추천 뉴스를 제공하기 때문에 유저에게 더욱 특화된 서비스를 제공할 수 있는 장점이 있다.

추가적으로 전체추천에서는 텍스트 마이닝, 그룹추천과 개인추천에서는 트위터 클러스터링을 적용하였다. 트위터 클러스터링은 트위터 내용을 추천시스템에 반영한 것으로, 이를 통해 키워드 관련 주제에 대한 사람들의 트렌드를 반영하고자 하였다.

따라서 ISPN은 사용자의 뉴스 소비 패턴을 보다 자세히 분석하여 사용자의 뉴스 소비패턴이 쌓임에 따라 더욱 맞춤화된 기사를 제공하고, 사용자가 필요한 기사만을 볼 수 있게 하여 사용자의 편리성과 개인성을 높일 수 있다.

전반적인 뉴스 추천 어플리케이션을 만드는 것이 아닌, 해외 축구에 한정된 사용자들을 대상으로 뉴스 추천 어플리케이션을 만들면서, 사용자들이 전반적으로 가지고 있는 성향과 특성화된 소비 성향을 적절히 조합하여 유의미한 결과를 만드는데 있어서 많은 고민과 논의를 거쳐 개발하였다.

### 4.2.2 전체추천

확인됐다. 즉, 추천 서비스를 제공할 대부분의 이용자들은 뉴스 서비스를 자주 사용하는 열성적인 이용자(Heavy User)가 아니라 이용자 행동 정보가 거의 없는 '콜드 스타트 이용자'인 것이다.

이러한 이용자에 대해서는 개인화 추천의 대표적인 방법인 협업 필터링 혹은 내용 기반 필터링으로 개인화 추천을 제공하기 어렵다. 이에 따라, 카카오는 전술한 두 방법 대신에 통계적 기계학습 기법인 '맞춤형 멀티암드밴딧'(Customized Multi-Armed Bandit) 알고리즘을 추천 시스템에 적용했다. 여기서 '맞춤형'이란 말은 카카오의 뉴스 서비스에 맞게 알고리즘이 수정됐다는 뜻이다.

<"기계 학습 기반의 뉴스추천 서비스 구조와 그 효과에 대한 고찰: 카카오의 루빅스를 중심으로" 사이버 커뮤니케이션학보, 통권 pw 34권 1호, 18p>

- 콜드 스타터인 초기사용자를 위해 멀티암드밴딩과 유사한 추천시스템으로 추천 기사를 제공한다.  
멀티암드밴딩은 사용자의 행동데이터가 충분히 모이지 않아 추천하기 힘든 상황일 때 추천할 수 있고 사용자에게 신선했(Serendipity)을 줄 수 있는 추천방식이다.  
멀티암드밴딩과 유사한 추천시스템을 통해 사용자의 데이터가 거의 없음에도 기사에 자체적으로 부여된 점수를 통해 기사를 추천해준다.  
사용자가 ISPN을 처음 사용했을 때, 또는 사용한지 얼마 되지 않았을 때도 사용자에게 기사를 추천해줄 수 있게 하기 위해서 멀티암드밴딩 기반 추천시스템을 반영하였고 이를 전체 추천이라고 부른다.

속성	예시	점수	비고
기본점수	-	20점	기본점수
언론사	ESPN	3.5점	언론사 가중치점수
조회수	20	+6.66점	3회 / +1점
시간	3일전	-6점	12시간 / -1점
팀, 감독, 선수 키워드 5개	Arsenal, Ozil, heung-min Son, rose, Mourinho	$8.6 + 18.6 + 5.4 + 13.4 + 22.7 = 68.7$	각 키워드의 가중치 합산점수
사진	O	+20점	사진있을시 +20점
추천 확률	45% / 13회	-1점	추천 횟수 10회 이상 30% 미만 -3점 30 ~ 50% -1점 50 ~ 70% +1점 70% 이상 +3점
최종 점수	-	112.86	최종 기사 추천 점수

<추천점수부여 표>

전체추천은 위 표에 따른 방식으로 각 기사에 점수가 부여되고, 점수가 높은 기사순으로 추천 기사에 제공된다.

#### 4.2.3 그룹추천

협업 필터링 시스템은 이용자와 콘텐츠 소비성향이 비슷한 유사 이용자들을 찾아내고, 이용자가 소비하지 않았지만 유사 이용자군에서 많이 소비된 콘텐츠를 추천해 주는 체계이다. 협업 필터링 시스템은 1990년대 후반 이후 추천에서 가장 많이 사용되고 있다.

<"기계 학습 기반의 뉴스추천 서비스 구조와 그 효과에 대한 고찰: 카카오의 루빅스를 중심으로" 사이버 커뮤니케이션학보, 통권 pw 34권 1호, 9p>

협업 필터링은 사용자데이터를 기반으로 형성된 그룹의 데이터를 이용하기 때문에 사용자가 선호할 만한 기사를 추천할 수 있다.

즉 ISPN을 많은 사용자들이 이용할수록 공통된 키워드를 가지는 사용자들이 생기고, 한 키워드를 중심으로 여러 사용자들을 그룹화할 수 있다. 이는 사용자가 지정하지 않은 주제지만 그룹 내의 다른 사용자들이 많이 지정한 키워드의 기사를 추천해줄 수 있다.

#### 4.2.4 개인추천

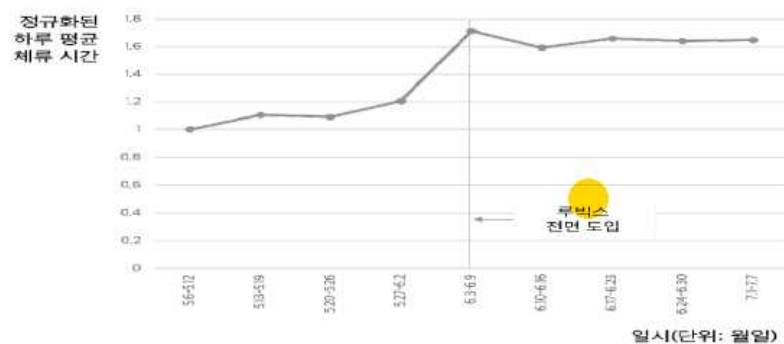
내용기반 필터링은 새로운 콘텐츠를 추천할 때 특히 유용하다. 새로운 콘텐츠들에 대한 다른 이용자들의 행동 데이터 유무에 상관없이 과거에 이용자가 선호했던 특성들을 많이 가지고 있으면 이용자 대상의 추천이 가능하기 때문이다. 이러한 장점 때문에, 내용기반 필터링과 협업 필터링을 결합한 다양한 혼합(Hybrid) 알고리즘들이 고안됐다(Balabanovi

<"기계 학습 기반의 뉴스추천 서비스 구조와 그 효과에 대한 고찰: 카카오의 루빅스를 중심으로" 사이버 커뮤니케이션학보, 통권 pw 34권 1호, 12p>

개인추천은 협업 필터링 중 하나인 모델기반 협력 필터링으로 사용자가 읽었던 기사를 분석하여 키워드를 찾아내는 방식으로 이루어져 있다. 이를 통해 그룹추천처럼 사용자가 선호할 만한 기사를 추천해 줄 수 있다.

#### 4.2.5 기존 추천시스템 성능 평가

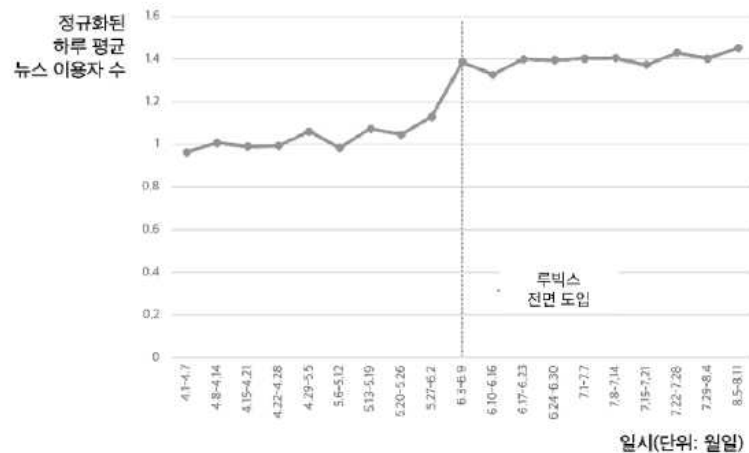
기존 추천시스템 도입 결과(루빅스)



<그림 8> 루빅스 전면 도입 전후의 하루 평균 총 체류 시간 추이(자료: 카카오)

<"기계 학습 기반의 뉴스추천 서비스 구조와 그 효과에 대한 고찰: 카카오의 루빅스를 중심으로" 사이버 커뮤니케이션학보, 통권 pw 34권 1호, 35p>

뉴스 소비의 질로 평가될 수 있는 기준중 하나인 체류 시간에서 루빅스를 도입하고 효과를 확인할 수 있었다고 한다. <그림 8>은 2015년 5월 이후부터 7월 초까지 다음 모바일 뉴스 서비스에서 이용자 총 체류 시간의 주간 단위 하루 평균의 변화를 나타낸 그래프이다. 7월 첫째 주 다음 모바일 뉴스 서비스에서의 이용자 체류 시간은 루빅스 도입 이전인 5월 둘째 주보다 65% 상승했다. 이 결과는 루빅스의 도입에 따른 클릭 수 상승이 실제 이용자가 더 많은 뉴스를 보기 위한 소비 행동의 결과임을 의미한다.



〈그림 9〉 루빅스 전면 도입 전후의 하루 평균 뉴스 이용자 수 추이(자료: 카카오)

<"기계 학습 기반의 뉴스추천 서비스 구조와 그 효과에 대한 고찰: 카카오의 루빅스를 중심으로" 사이버 커뮤니케이션학보, 통권 pw 34권 1호, 36p>

<그림 9>는 2015년 4월부터 8월 둘째 주까지의 다음 모바일 뉴스 서비스에서 주별 하루 평균 뉴스 이용자 수를 나타낸 그래프로, 루빅스 도입 이후 8월 둘째 주의 4월 대비 뉴스 이용자 수가 45% 증가했다. 이는 루빅스 도입 이후 다음 모바일 뉴스 서비스의 이용자의 규모가 확장했음을 의미한다.

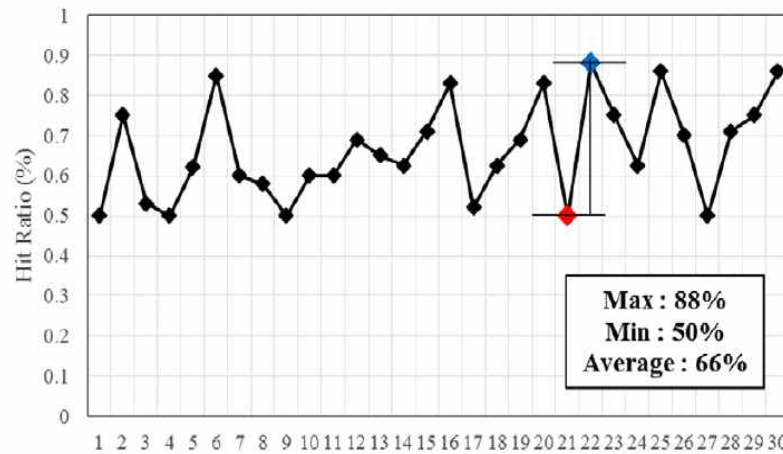
루빅스 도입 이후 다음 모바일 서비스의 뉴스 이용자와 체류시간의 증가를 통해 루빅스 추천시스템의 긍정적 효과를 확인할 수 있었다. ISPN은 루빅스 추천시스템을 기반으로 설계하여 비슷한 맥락에서 ISPN 사용자들에게서 이러한 효과를 기대해볼 수 있다.

#### 4.2.6 추천시스템 성능 평가

추천 시스템의 평가 기준은 추천 성공률이다. 추천 성공률이 높아질수록 사용자에게 추천한 기사가 흥미를 끌었다는 의미로, 추천 시스템이 효과가 있었다고 평가할 수 있다.

$$\text{추천성공률} = \frac{\text{사용자가 읽은 추천기사의 횟수}}{\text{사용자에게 추천한 기사횟수}} \times 100(\%)$$

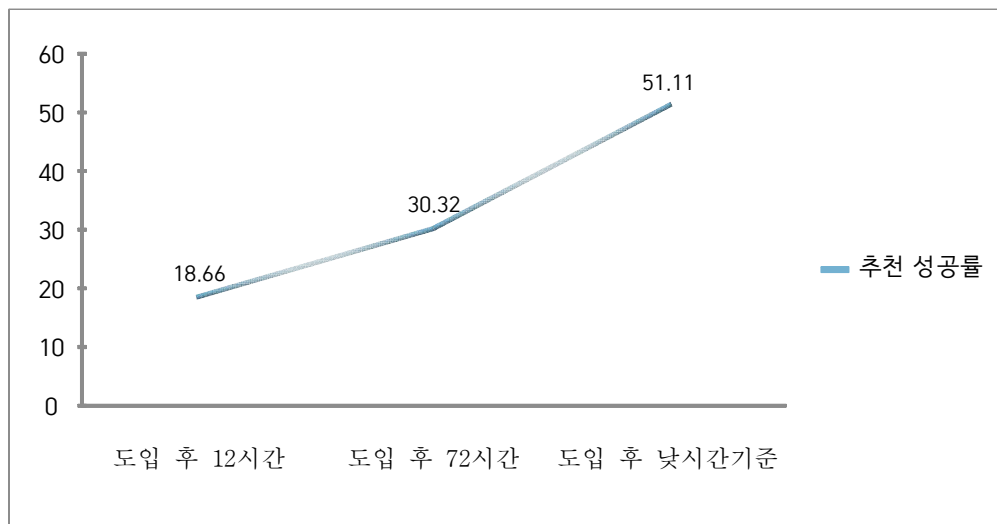
사용자가 뉴스를 읽었다고 확인하기 위해서 최소 체류 시간을 정하고, 최소 체류 시간 이상 머무는지 확인하였다. 뉴스를 확인하고 관심이 없을 경우 나가는 시간까지 약 6초가 걸리고 기사에 이미지가 있을 경우 +3초를 하여 6초~ 9초 이상의 시간동안 기사에 체류 시 기사를 읽었다고 판단했다.



<“정보과학회논문지 제42권 제5호 사용자의 소셜 카테고리를 이용한 유튜브 동영상 추천 알고리즘”>

해당 실험은 유튜브 동영상 추천 알고리즘에 따른 실험 결과의 Hit Ratio 수치를 분석한 그래프이다. 실험 결과가 Min 50%, Max 88%, Average 66%이다. 유의미한 클릭만을 클릭수로 모았을 때 제공한 기사의 추천 성공률을 유튜브 추천 알고리즘의 Minimum인 50%과 같게 잡아, 사용자가 사용자에게 제공된 추천 기사들 중에서 절반 이상의 기사를 클릭할 수 있도록 하는 것을 목표로 하였다.

ISPIN 추천시스템 도입 결과(추천 성공률)



해외축구 뉴스에 대한 추천이라는 특수성을 반영한 사용자 관심 키워드 지정 기능을 제공하고 트위터 클러스터링을 적용함으로써 추천시스템의 추천 성공률(Hit ratio)을 목표치 이상 달성하려하였다.

테스트 유저 10명에 대해 3일간의 테스트를 진행하였으며, 위의 도표는 테스트의 결과를 보여준다. 대부분 전체추천으로만 제공되었을 때(사용자들의 뉴스 소비패턴이 쌓이지 않았을 때, 즉 초기 사용자) 추천 성공률이 18.66% 이었다. 하지만 사용자의 소비패턴이 쌓임에 따라(3일이 지난 후 - 테스트 진행 72시간 후) 30.32%로 추천 성공률이 증가하는 것을 볼 수 있었다. 사용자들이 뉴스를 가장 활발히 소비했던 낮 시간 기준(오후 12시부터 오전 12시, 총 36시간)으로 통계를 낸 결과, 51.11%로 목표치였던 추천 성공률 50%를 넘긴 것을 확인

할 수 있었다.

해당 테스트를 통해 작은 표본이지만 기대하던 목표치를 일정 수준 달성할 수 있음과 추천 시스템의 효과를 긍정적으로 평가할 수 있다.

## 5 결론

### 5.1 문제점 및 개선방안

어플리케이션을 개발하면서 모듈을 세분화하여 관리하고 통합 테스트를 반복하는 과정에서 자잘하게 발생하던 에러들을 수정해왔고, 현재로서는 어플리케이션 내부 시스템에서 큰 문제를 발생시키는 에러가 없는 상태이다.

추천시스템으로 기사를 추천하는 과정에서 오버헤드를 줄이기 위해 12시간마다 추천시스템을 실행하여 유저별 추천 기사를 제공하도록 시스템을 설계 했다. 허나 동시접속자 100명에 대해서 테스트 해본 결과 예상보다 추천시스템 작동에 시간이 많이 소요되지 않았다. 따라서 사용자가 클라이언트에 접속 시, 그리고 새로고침 시 실시간으로 추천 기사를 제공해주는 방식으로 시스템을 개선해볼 수 있을 것이다. 추천 기사를 새로 받을 수 있는 버튼을 따로 설정하여 버튼을 누르면 추천시스템이 작동하고, 기사가 새로 추천되기 전까지는 기존에 있던 추천 기사를 유지하는 방법을 생각해볼 수도 있다. 비록 오버헤드의 가능성과 추천기사가 제공되는데 걸리는 시간이 길어지지만 사용자가 새로운 추천 기사를 자주 제공 받을 수 있기 때문에 사용자들의 어플리케이션 체류 시간을 늘릴 수 있을 것이다.

### 5.2 결론

수많은 사람들이 스마트폰을 사용하며, 어플리케이션을 통해 정보를 편리하게 얻고 있다. 스포츠 뉴스, 그 중에서도 해외 축구 뉴스는 실시간으로 올라오는 기사와 관련 반응이 중요하다. 그럼에도 불구하고 언론사와 SNS가 통합되어 한 눈에 볼 수 있는 해외 축구 기사 어플리케이션과 해외 축구 뉴스에 추천 시스템을 적용하여 사용자에게 맞춤 기사를 추천해주는 어플리케이션이 부재하였다. 이 둘의 융합을 고려하면서 ISPN 개발의 필요성을 깨달았고, 결국 목표치를 성공적으로 달성하면서 어플리케이션을 개발해낼 수 있었다. 해외 축구 기사를 자주 확인하는 사용자들에게 맞춰진, 편리한 기능들로 구성되어있어 원하는 키워드에 맞게 기사를 읽고, 기사를 추천 받는 편리한 맞춤 서비스를 지원하는 것이 본 어플리케이션의 최대 장점이다. 향후 꾸준한 버그 수정과 테스트를 통해 안정성을 높이면서, 영국 포함 유럽 5대 리그인 스페인, 이탈리아, 독일, 프랑스 리그로 서비스를 확장하여 제공하고 싶은 바람이 있다.

### 5.3 참고 문헌

- [1] 박승택, 성인재, 서상원, 황지수, 노지성, 김대원, “기계학습 기반의 뉴스 추천 서비스 구조와 그 효과에 대한 고찰 : 카카오의 루빅스를 중심으로” 사이버커뮤니케이션학보, 통권 제 34 권 1호, 2017, 3김
- [2] 서봉원, “콘텐츠 추천알고리즘의 진화”, Broadcasting trend & insight 2016.04
- [3] 유소엽, 정옥란, “사용자의 소셜 카테고리를 이용한 유튜브 동영상 추천 알고리즘”, Journal of KIISE, Vol. 42, N. 5. Pp 664-670, 2015, 5



- [4] 김수연, 이상훈, 황현석, “AHP와 하이브리드 필터링을 이용한 개인화된 추천 시스템 설계 및 구현”, J Korea Industr Inf Syst Res Volume 17 Number 7, 2012
- [5] 김석훈 “Django로 배우는 쉽고 빠른 웹개발 파이썬 웹프로그래밍“ 한빛미디어 2018 개정판 1쇄
- [6] 천인국 “그림으로 쉽게 설명하는 안드로이드 프로그래밍” 생능출판 2018 제4판 1쇄
- [7] <https://pypi.org/project/GetOldTweets3/>, 트위터 API, 오픈소스
- [8] <https://developer.android.com/?hl=ko>, 안드로이드 공식 가이드
- [9] <https://www.uplabs.com/>, 안드로이드 UI 참고 사이트