

프로젝트 개발 계획서

1. 프로젝트 이름

A. 프로젝트 이름 : VINE-DING

B. 프로젝트 이름의 의미 : 덩굴이라는 의미의 Vine 과 묶는다는 의미의 binding 의 음을 따와서, 여러 정보들을 합쳐서 보여준다는 의미로 선정

2. 프로젝트 개요

A. VINE-DING 은 젊은 세대를 대상으로 하는 정보 분석 서비스 입니다. 여러 커뮤니티 사이트와 SNS 등, 많은 데이터들을 기반으로 하여 실시간으로 지금 사람들이 어떤 것에 관심을 가지고 있는지, 사람들이 지금 어떤 감정을 많이 가지고 있는지를 이해하기 쉽게 시각화 하여 고객들에게 제공합니다.
여러분의 세상을 넓혀보세요.
저희는 그 시간을 제공해 드리겠습니다.

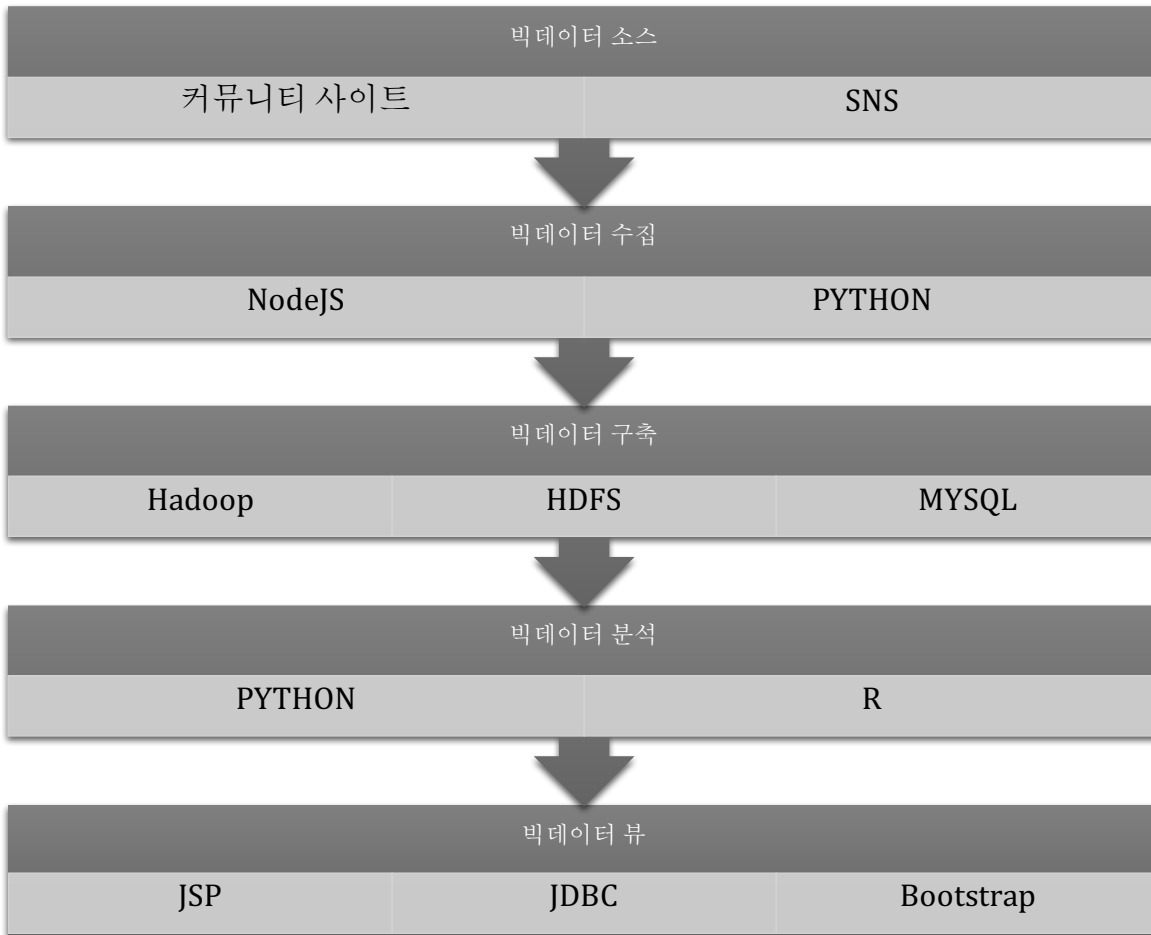
3. 유사 서비스 벤치마킹

- A. tweettrend : <http://tweettrend.com>
- B. pulsek : <http://www.pulsek.com/pulsek/>
- C. 서울 열린 데이터 광장 : <http://data.seoul.go.kr/>
- D. 네이버 데이터 랩 : <http://datalab.naver.com/>
- E. 구글 트렌드 : <https://trends.google.com/trends/>
- F. 오피디아 : <https://www.odpia.org/main.odpia>
- G. 소셜매트릭스 : <http://insight.some.co.kr/campaign.html>

4. 프로젝트 주요기능

- A. SNS 와 각종 커뮤니티에서 데이터를 모아 이를 베이스로 한다.
- B. 관심사/카테고리 별 많이 언급된 항목 순위 제공
 - i. 관심사의 항목은 영화, 음악, 여행지, 뉴스, 드라마 등이 있다.
- C. SNS 트렌드 실시간 분석
- D. 어떤 이슈에 대한 감정 분석(긍정, 부정의 퍼센트)
- E. 실시간 정보 밀집도 조사를 통해서 새로운 이슈 탐지

5. 시스템 구성 & 사용 기술



6. 작업 명세서

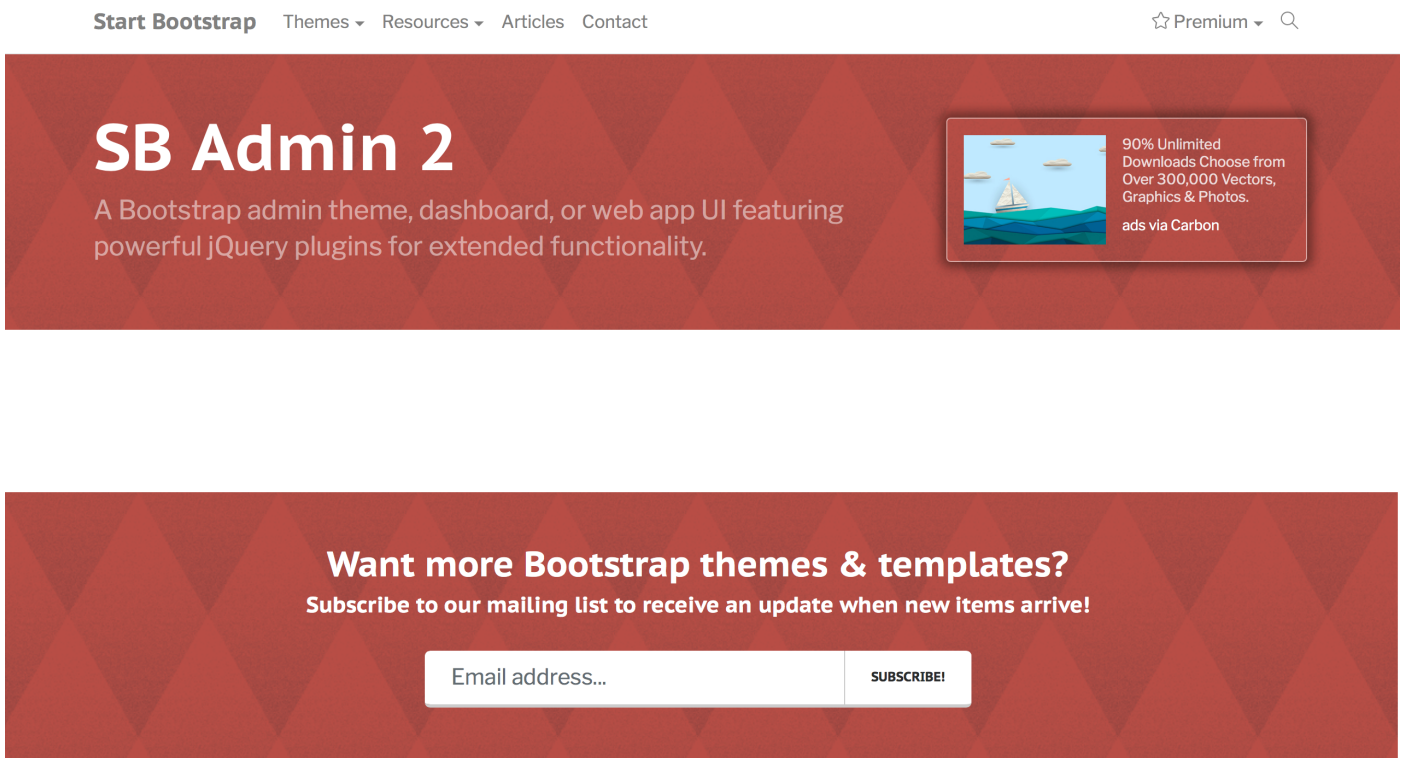
| No | 작업명 | 담당자 | 예상소요시간 |
|----|---|-----|--------|
| 01 | 크롤링 대상 사이트 정하기 | 공통 | 3 시간 |
| 02 | 개발 환경 구축하기(가상머신, Node.js 설치, 에디터) | 최봉근 | 2 시간 |
| 03 | 웹 페이지 다운로드 | 최봉근 | 2 시간 |
| 04 | HTML 해석(링크와 이미지 추출) | 최봉근 | 3 시간 |
| 05 | 사이트 통째로 다운로드 | 최봉근 | 2 시간 |
| 06 | XML/RSS 해석 | 최봉근 | 2 시간 |
| 07 | 정기적으로 다운로드(정기적인 처리를 수행) | 최봉근 | 3 시간 |
| 08 | 로그인이 필요한 웹사이트 크롤링 데이터 다운로드(로그인 동반 데이터 수집) | 최봉근 | 5 시간 |
| 09 | 문자 코드와 호환 | 최봉근 | 2 시간 |
| 10 | 정규 표현식을 사용해 데이터 변환 | 최봉근 | 5 시간 |

| | | | |
|----|---|-----|------|
| 11 | 각각의 데이터 형식을 node.js 로 다루기 | 최봉근 | 5 시간 |
| 12 | 데이터를 DB 에 저장 | 최봉근 | 3 시간 |
| 13 | 형태소 분석 | 최봉근 | 5 시간 |
| 14 | 특정 품사의 단어들을 추출 | 최봉근 | 3 시간 |
| 15 | 단어의 출현 빈도 조사 | 최봉근 | 3 시간 |
| 16 | 데이터 시각화(구글 차트를 이용한 차트 작성) | 최봉근 | 6 시간 |
| 17 | 관심사 도출과 관련한 핵심 문제점 인식 및 가치 발견 | 공통 | 2 시간 |
| 18 | 관심사 목록 도출 (음악, 영화, 여행, 음식 등) | 이성진 | 1 시간 |
| 19 | 수집된 데이터의 범위 파악(필요한 테이블, 데이터 기간, 데이터 량 등) | 이성진 | 3 시간 |
| 20 | 수집된 데이터 기반의 Output 설정 (유저들이 어떠한 정보들을 원할지에 대해 생각) | 이성진 | 3 시간 |
| 21 | 결과물을 이용한 가설 설정 | 이성진 | 2 시간 |
| 22 | 형태소 라이브러리(NIADic 등)을 통한 형태소 추출 | 이성진 | 6 시간 |
| 23 | 추출한 형태소로 관심사 키워드 분석 | 이성진 | 3 시간 |
| 24 | 관심사에 관한 연관어 추출 | 이성진 | 6 시간 |
| 25 | 각 연관어에 대한 빈도 분석 | 이성진 | 6 시간 |
| 26 | 각 연관어의 빈도에 따른 중요도 설정 | 이성진 | 3 시간 |
| 27 | 관심사에 관한 키워드를 연관어로 분석 | 이성진 | 3 시간 |
| 28 | 관심사 도출을 위한 빅데이터 알고리즘 선정 | 이성진 | 2 시간 |
| 29 | 선정된 알고리즘 바탕의 분석 모델 작성 | 이성진 | 3 시간 |
| 30 | 분석 모델 코드화 | 이성진 | 6 시간 |
| 31 | 작성된 모델을 이용하여 샘플 데이터의 관심사 키워드 추출 테스트 | 이성진 | 6 시간 |
| 32 | 가설을 샘플데이터로 테스트하여 신뢰성 검증 | 이성진 | 3 시간 |
| 33 | 실제 크롤링된 데이터 기반의 관심사 추출 테스트 | 이성진 | 6 시간 |
| 34 | 테스트 결과물 수정 및 보완 | 이성진 | 3 시간 |
| 35 | 게시물 등록 빈도수 쿼리문 작성 (연령대별, 날짜별, 관심사별 등) | 최봉근 | 2 시간 |
| 36 | 쿼리 결과물의 불필요한 부분을 없애고 유저의 기호에 맞춰 데이터 정제 | 최봉근 | 3 시간 |
| 37 | 데이터 필터링 최적화 (유해, 스팸 및 광고성 글 필터링) | 최봉근 | 6 시간 |
| 38 | 데이터 가치가 있는 SNS 판별 | 이동기 | 3 시간 |
| 39 | SNS API 분석 및 레퍼런스 정리 | 이동기 | 3 시간 |
| 40 | API 사용법 정리 | 이동기 | 4 시간 |

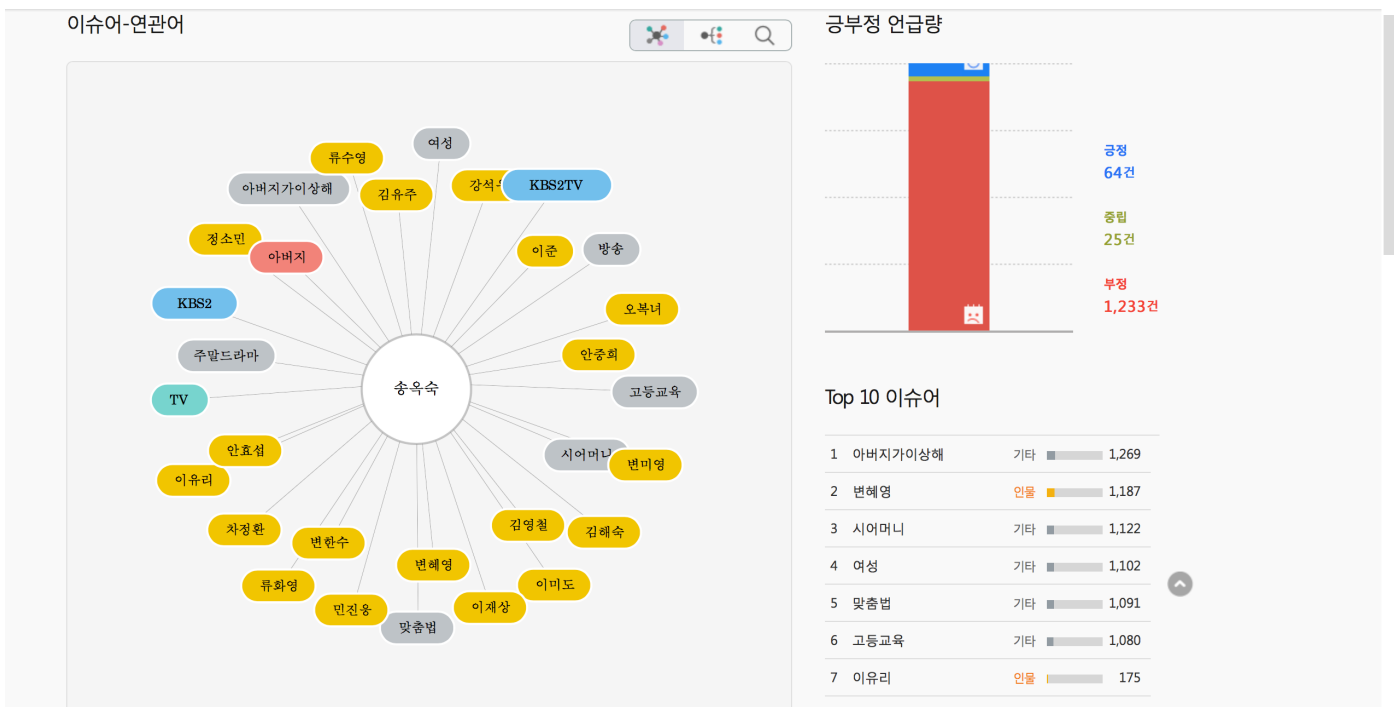
| | | | |
|----|--------------------------|-----|---------|
| 41 | 실제 사용 가능한 API 정리 | 이동기 | 2 시간 |
| 42 | API 를 통해 받아 올 정보저장 방법 선정 | 이동기 | 2 시간 |
| 43 | 샘플데이터를 선정한 DB 에 저장 해보기 | 이동기 | 2 시간 |
| 44 | SNS API-DB 연동 | 이동기 | 3 시간 |
| 45 | 감정 분석에 필요한 정보 추론 | 이동기 | 2 시간 |
| 46 | 감정 분석에 불필요한 정보 필터링 | 이동기 | 3 시간 |
| 47 | 단어-감정 분석 모델 설계 | 이동기 | 6 시간 |
| 48 | 각 단어와 감정상태를 매핑 | 이동기 | 6 시간 |
| 49 | 모델 구현을 위한 모듈 설계 | 이동기 | 4 시간 |
| 50 | 단어-감정 모듈 1 구현 | 이동기 | 2 시간 |
| 51 | 단어-감정 모듈 2 구현 | 이동기 | 2 시간 |
| 52 | 단어-감정 모듈 3 구현 | 이동기 | 2 시간 |
| 53 | 실제 데이터를 이용한 테스트 | 이동기 | 2 시간 |
| 54 | 결과 데이터를 표시할 방법 선택 | 이동기 | 1 시간 |
| 55 | 선택한 방법의 API 예시 제작 | 이동기 | 3 시간 |
| 56 | 실제 결과와 데이터 표시 API 연동 | 이동기 | 3 시간 |
| 57 | 웹 앱 UI 설계 | | 공통 4 시간 |
| 58 | 설계된 UI 를 기반으로 반응형 템플릿 제작 | 최봉근 | 4 시간 |
| 59 | 반응형 뷰를 위한 기술 학습 | | 공통 3 시간 |
| | 총 합 | 이동기 | 55 시간 |
| | 총 합 | 최봉근 | 62 시간 |
| | 총 합 | 이성진 | 65 시간 |

7. 화면 설계서

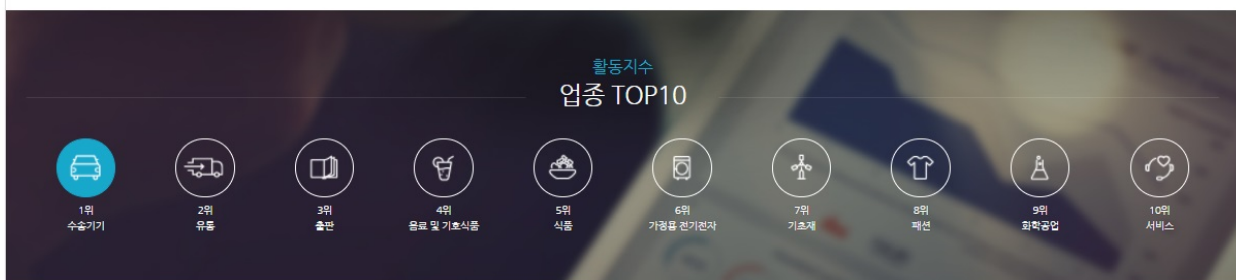
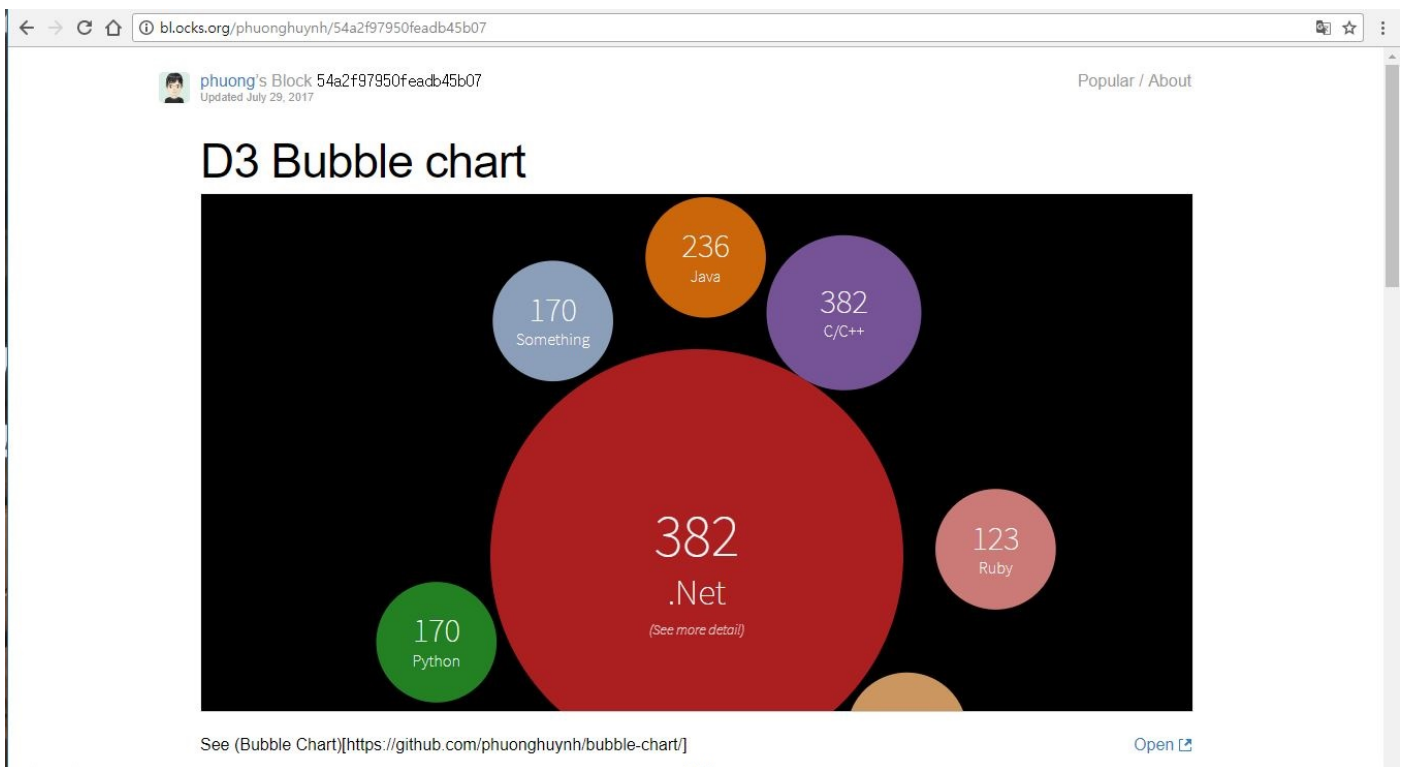
A. 템플릿



B. Content View(감정분석)



C. Content View(관심사 분석)



2017-07-24 ~ 2017-07-30

| 랭킹 | | 활동지수 📈 | 팬층감수 📉 | 보유팬수 📈 | 포스트수 📈 | 반응수 📈 |
|----|--|-------------|--------|------------|--------|---------|
| 1 |  하이마트(Hi-mart) 유통 | 86.42 ★★★★★ | 10,760 | 6,631,409 | 30 | 45,126 |
| 2 |  오늘 뭐 먹지? 서비스 | 86.03 ★★★★★ | 852 | 4,317,144 | 49 | 865,890 |
| 3 |  패션정보 서비스 | 84.50 ★★★★★ | 193 | 423,025 | 430 | 59,568 |
| 4 |  Big Hit Entertainment 서비스 | 84.45 ★★★★★ | 9,148 | 1,207,258 | 4 | 120,134 |
| 5 |  Google 컴퓨터 및 정보통신 | 84.27 ★★★★★ | 912 | 10,221,048 | 3 | 1,931 |
| 6 |  BIGBANG 서비스 | 84.26 ★★★★★ | 290 | 10,490,395 | 6 | 23,279 |
| 7 |  YTN 컴퓨터 및 정보통신 | 84.00 ★★★★★ | 416 | 587,801 | 284 | 216,019 |

D. Content View(SNS 트렌드 분석) - Google Chart(Line chart)

1 미네소타 트윈스, 하이메 가르시아, 오كل랜드 애슬레틱스, 뉴욕 양키스



2 다니엘 리카르도, 헝가리 그랑프리, 레드불 레이싱, 헝가리, 루이스 해...



3 Pocono Raceway, 스프린트 컵 시리즈, Camping World Truck S...



4 토미 래런, 첼시 핸들러, 도널드 트럼프, 환자보호 및 부담적정보험법



5 타이론 우들리, 데미안 마이아, UFC 214, 얼티밋 파이팅 챔피언십, ...



6 보스턴 레드삭스, 캔자스시티 로열스, 에두아르도 누녜스, 마이크 무...



7 보스턴 패트리어츠, 롬 닌코비치, 내셔널 풋볼 리그, Bill Belichick



8. 프로젝트 스케줄

| 작업 명 | 1 주 | 2 주 | 3 주 | 4 주 | 5 주 | 6 주 | 7 주 |
|-----------------|-----|-----|-----|-----|-----|-----|-----|
| 데이터 설계 시스템 완성 | | | | | | | |
| 기능별 데이터 사용방법 설계 | | | | | | | |
| 기능 구현 및 테스트 | | | | | | | |
| 뷰 제작 | | | | | | | |
| 문서작업 | | | | | | | |

9. 프로젝트 법률문제 분석

A. SNS 정보수집시에 개인정보 침해

- i. 자신의 영업을 위해 무단으로 사용함으로써 타인의 경제적 이익을 침해하는 행위를 부정경쟁행위로 규정한다.

B. 분석결과

- i. 이 프로젝트에서 수집하는 정보는 타인의 경제적 이익을 침해하는 행위가 아니라서 문제가 없다고 판단된다.