작업명세서

(놈놈놈 팀 이동기, 이성진, 최봉근)

|  |  |  |  |
| --- | --- | --- | --- |
| No | 작업명 | 담당자 | 예상소요시간 |
| 01 | 크롤링 대상 사이트 정하기 | 공통 | 3시간 |
| 02 | 개발 환경 구축하기(가상머신, Node.js설치, 에디터) | 최봉근 | 2시간 |
| 03 | 웹 페이지 다운로드 | 최봉근 | 2시간 |
| 04 | HTML 해석(링크와 이미지 추출) | 최봉근 | 3시간 |
| 05 | 사이트 통째로 다운로드 | 최봉근 | 2시간 |
| 06 | XML/RSS 해석 | 최봉근 | 2시간 |
| 07 | 정기적으로 다운로드(정기적인 처리를 수행) | 최봉근 | 3시간 |
| 08 | 로그인이 필요한 웹사이트 크롤링 데이터 다운로드(로그인 동반 데이터 수집) | 최봉근 | 5시간 |
| 09 | 문자 코드와 호환 | 최봉근 | 2시간 |
| 10 | 정규 표현식을 사용해 데이터 변환 | 최봉근 | 5시간 |
| 11 | 각각의 데이터 형식을 node.js로 다루기 | 최봉근 | 5시간 |
| 12 | 데이터를 DB에 저장 | 최봉근 | 3시간 |
| 13 | 형태소 분석 | 최봉근 | 5시간 |
| 14 | 특정 품사의 단어들을 추출 | 최봉근 | 3시간 |
| 15 | 단어의 출현 빈도 조사 | 최봉근 | 3시간 |
| 16 | 데이터 시각화(구글 차트를 이용한 차트 작성) | 최봉근 | 6시간 |
| 17 | 관심사 도출과 관련한 핵심 문제점 인식 및 가치 발견 | 공통 | 2시간 |
| 18 | 관심사 목록 도출 (음악, 영화, 여행, 음식 등) | 이성진 | 1시간 |
| 19 | 수집된 데이터의 범위 파악(필요한 테이블, 데이터 기간, 데이터 량 등) | 이성진 | 3시간 |
| 20 | 수집된 데이터 기반의 Output 설정  (유저들이 어떠한 정보들을 원할지에 대해 생각) | 이성진 | 3시간 |
| 21 | 결과물을 이용한 가설 설정 | 이성진 | 2시간 |
| 22 | 형태소 라이브러리(NIADic 등)을 통한 형태소 추출 | 이성진 | 6시간 |
| 23 | 추출한 형태소로 관심사 키워드 분석 | 이성진 | 3시간 |
| 24 | 관심사에 관한 연관어 추출 | 이성진 | 6시간 |
| 25 | 각 연관어에 대한 빈도 분석 | 이성진 | 6시간 |
| 26 | 각 연관어의 빈도에 따른 중요도 설정 | 이성진 | 3시간 |
| 27 | 관심사에 관한 키워드를 연관어로 분석 | 이성진 | 3시간 |
| 28 | 관심사 도출을 위한 빅데이터 알고리즘 선정 | 이성진 | 2시간 |
| 29 | 선정된 알고리즘 바탕의 분석 모델 작성 | 이성진 | 3시간 |
| 30 | 분석 모델 코드화 | 이성진 | 6시간 |
| 31 | 작성된 모델을 이용하여 샘플 데이터의 관심사 키워드 추출 테스트 | 이성진 | 6시간 |
| 32 | 가설을 샘플데이터로 테스트하여 신뢰성 검증 | 이성진 | 3시간 |
| 33 | 실제 크롤링된 데이터 기반의 관심사 추출 테스트 | 이성진 | 6시간 |
| 34 | 테스트 결과물 수정 및 보완 | 이성진 | 3시간 |
| 35 | 게시물 등록 빈도수 쿼리문 작성  (연령대별, 날짜별, 관심사별 등) | 최봉근 | 2시간 |
| 36 | 쿼리 결과물의 불필요한 부분을 없애고 유저의 기호에 맞춰 데이터 정제 | 최봉근 | 3시간 |
| 37 | 데이터 필터링 최적화  (유해, 스팸 및 광고성 글 필터링) | 최봉근 | 6시간 |
| 38 | 데이터 가치가 있는 SNS 판별 | 이동기 | 3시간 |
| 39 | SNS API 분석 및 레퍼런스 정리 | 이동기 | 3시간 |
| 40 | API 사용법 정리 | 이동기 | 4시간 |
| 41 | 실제 사용 가능한 API 정리 | 이동기 | 2시간 |
| 42 | API를 통해 받아 올 정보저장 방법 선정 | 이동기 | 2시간 |
| 43 | 샘플데이터를 선정한 DB에 저장 해보기 | 이동기 | 2시간 |
| 44 | SNS API-DB연동 | 이동기 | 3시간 |
| 45 | 감정 분석에 필요한 정보 추론 | 이동기 | 2시간 |
| 46 | 감정 분석에 불필요한 정보 필터링 | 이동기 | 3시간 |
| 47 | 단어-감정 분석 모델 설계 | 이동기 | 6시간 |
| 48 | 각 단어와 감정상태를 매핑 | 이동기 | 6시간 |
| 49 | 모델 구현을 위한 모듈 설계 | 이동기 | 4시간 |
| 50 | 단어-감정 모듈1 구현 | 이동기 | 2시간 |
| 51 | 단어-감정 모듈 2 구현 | 이동기 | 2시간 |
| 52 | 단어-감정 모듈3 구현 | 이동기 | 2시간 |
| 53 | 실제 데이터를 이용한 테스팅 | 이동기 | 2시간 |
| 54 | 결과 데이터를 표시할 방법 선택 | 이동기 | 1시간 |
| 55 | 선택한 방법의 API 예시 제작 | 이동기 | 3시간 |
| 56 | 실제 결과와 데이터 표시 API 연동 | 이동기 | 3시간 |
| 57 | 웹 앱 UI 설계 |  | 공통 4시간 |
| 58 | 설계된 UI를 기반으로 반응형 템플릿 제작 | 최봉근 | 4시간 |
| 59 | 반응형 뷰를 위한 기술 학습 |  | 공통 3시간 |