
올인원 전시회 플랫폼, 아뜰리에(Atelier)

- 2022 1학기 캡스톤 디자인 최종보고서 -

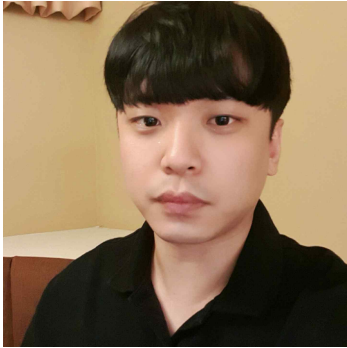


| | |
|-----|-------------|
| 과목명 | 캡스톤디자인(2) |
| 교수명 | 이형태 교수님 |
| 제출일 | 2022.06.03. |
| | 20174266 |
| | 컴퓨터공학부 이지호 |
| 팀원 | 20194863 |
| | 소프트웨어학부 이승연 |
| | 20191181 |
| | 소프트웨어학부 유승욱 |

- 목차 -

| | | |
|----|------------------------|------|
| 1) | 프로젝트 기획 | |
| | 팀원 소개 | - 1 |
| | 프로젝트 의도 | - 1 |
| | 기존 사례 분석 | - 3 |
| 2) | 프로젝트 개발 | |
| | 전체 시나리오 | - 5 |
| | 웹 개발: Frontend | - 6 |
| | 웹 개발: Backend | - 10 |
| | 기술 개발: 크롤링 및 전처리 | - 11 |
| | 기술 개발: 키워드 추출 및 전시회 추천 | - 18 |
| 3) | 프로젝트 스케줄 | |
| | 팀 일정 | - 23 |
| | 교수님 피드백 일정 | - 25 |
| | 멘토님 피드백 일정 | - 26 |

1-1. 팀원 소개



이지호
컴퓨터공학부 4학년



이승연
소프트웨어학부 4학년



유승욱
소프트웨어학부 4학년

React를 통한 UI 생성
Springboot를 통한 웹서비스 생성
Flask를 통한 전시회 추천 API 생성
개발환경 구축 및 DB 모델링

전시회 크롤링 및 image2text
텍스트 전처리 및 키워드 추출
키워드를 통한 전시회 추천
Flask API를 통한 기술 연동

1-2. 프로젝트 의도

온라인 전시 열풍!
코로나로 멈춘 미술관, 안방 찾아간다...

코로나 바이러스의 여파로 박물관과 미술관의 휴관이 장기화되며 온라인 전시가 대안으로 떠오르고 있다. 특히 젊은층들 사이에서 큰 인기를 끌었던 각종 전시회들이 코로나 영향으로 인해 축소되는 분위기에서, 다시 그 활기를 찾아가려는 모양새다.



<그림 1> 코로나에 따른 예술시장 타격

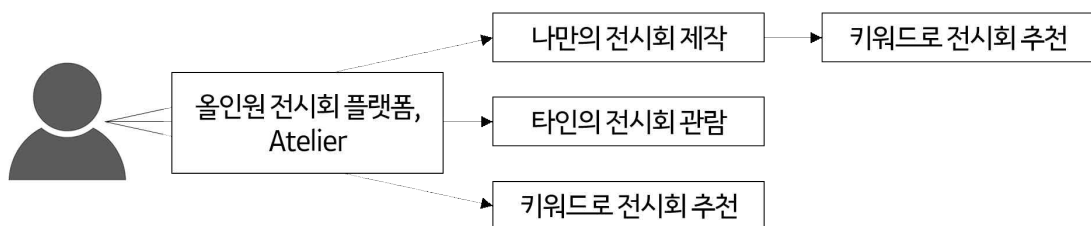
올인원 전시회 플랫폼, 아뜰리에(이하 Atelier)는 본인만의 전시회를 온라인에서 개최하고, 다른 사람들과 공유할 수 있는 커뮤니티를 제공한다. 나만의 스타일대로 온라인 전시를 기획하고 타인과 교류하는 장을 통해 코로나로 인해 지속되는 언택트(untact) 상황에서도 개성

넙치는 문화생활을 *Atelier*에서 온전히 즐길 수 있다. *Atelier*는 ‘예술가가 작업하는 공간이 나 건물’이라는 뜻으로, 서비스를 사용하는 유저라면 그 순간만큼은 예술가가 되어 본인만의 공간을 마음껏 꾸며보라는 의미를 담았다.

예상치 못하게 코로나 바이러스가 오랫동안 지속되며 문화예술계 전반은 큰 타격을 입었다. 전시회를 포함하여 대면으로 진행되었던 행사들은 모두 취소되어야만 했고, 그로 인해 무명 예술인들은 흔히 말하는 ‘배고픈 예술가’로 살아야만 했다. 이렇게 상대적으로 덜 알려진 예술가의 전시회더라도, 사용자가 원하는 취향의 전시회라면 우리 *Atelier*를 통해 추천될 수 있는 시스템을 마련하고 싶었다. 각자의 개성과 그에 따른 실력을 겸비한 예술가라면 누구든 더 많은 사람에게 알려지고 발굴될 수 있도록 전시회가 추천될 수 있는 것이다.

*Atelier*를 가장 많이 이용할 것으로 예상되는 MZ 세대는 ‘자신을 표현하고자 하는 세대’다. SNS 등을 통해 자기 사진을 공유하며, 개개인의 일상을 담은 유튜브 브이로그도 많은 인기를 끌고 있다. 특히 예술 계열 대학생들의 경우 졸업작품을 전시하거나 자신의 이야기를 담은 개인전을 개최하기도 한다.

그러나 대관을 통해 전시를 진행하기에는 청년들에게 경제적인 요건 등의 제약이 따른다. 더군다나 코로나 바이러스로 인해 대면 활동에 대한 수요 자체가 적어져 전시회를 열고픈 수요는 있지만 실제로 대면 전시를 아마추어 입장에서 진행하기는 쉽지 않다. 이러한 측면에서 *Atelier*는 본인만의 개성 있는 온라인 전시를 간단하게 할 수 있도록 도와주는 서비스이며, 본인만의 전시를 토대로 실제 진행되는 다른 미술가들의 전시회를 함께 추천해준다.



<그림2> *Atelier* 프로젝트 기획 의도

우선 ‘나만의 전시회 제작’을 통해 평상시에 찍어두었던 사진들로 본인만의 테마를 정한 전시를 개최할 수 있다. 사진마다 적절한 해시태그 및 설명을 덧붙이고, 전시에 적합한 BGM을 설정하는 등 나만의 전시를 개최하고 꾸미는 성취감을 느낄 수 있다. 본인의 전시회 제작 뿐만 아니라 타인의 온라인 전시에 방문하여 댓글과 같은 리액션을 남기는 커뮤니티 기능 역시 제공된다.

나만의 전시 개최에 끝나지 않고, 본인이 개최한 전시와 가장 어울리는 ‘오프라인 전시 추천’을 돕는다. 해시태그를 통해 본인의 전시회를 잘 설명할 수 있다면, 해당 해시태그를 통

해 본인이 개최한 전시와 가장 유사한 오프라인 전시를 추천받을 수 있다. 전시 기간 및 위치 등 해당 전시회의 대략적인 정보를 함께 제공함으로써 사용자의 편의를 돕는다. 그저 사이트에 사진과 글을 올리는 것 보다, 다른 전시들과 함께 추천이 이루어지기 때문에 기존의 오프라인 전시회처럼 문화생활에 가까운 활동을 기대할 수 있다.

1-3. 기존 사례 분석

Atelier 서비스는 기존 전시회 관련 서비스들과는 달리 확실한 차별점을 가진다.

1) 기존 전시회 관련 사이트들과의 비교

| 서비스명 | 특징 |
|---|---|
| 캔고루  | 테마별로 전시회를 나열해주나 실질적으로 추천 양이 너무 적음. (거의 없다시피 함) |
| 아트맵  | 전시회, 작가, 작품을 안내해주는 큐레이터 서비스. 전시를 감상하고 즐기는 쪽에 초점이 맞춰져 있음. 인기 있는 대중적인 전시 추천의 위주. |
| 아트허브  | 대한민국의 전시자료를 저장하고 서비스하는 디지털 아카이브 역할. 단순 역대 전시회 기록용도. |
| 기타 전시회 예매 관련 (인터파크, 티켓링크, 예술의전당, 문화포털 등) | 날짜에 맞게 진행되는 전시회 예매 가능. 인기 있는 대중적인 전시 추천의 위주. |

<표1> 기존 전시회 관련 사이트들

기존 전시회 관련 사이트들의 경우 아트맵과 인터파크, 티켓링크와 같은 기타 전시회 예매 관련 서비스들에서는 인기 있는 대중적인 전시 추천이 위주가 되었다. 아트허브의 경우는 애초에 아카이브 역할을 목표로 만들었기 때문이었는지 저장된 전시회의 정보가 가장 많았으나, 단순 나열의 용도였을 뿐 사용자 추천 등의 부가 기능은 찾아볼 수 없었다.

그나마 전시회를 세세하게 분류함으로써 추천을 시도한 캔고루의 경우는 #인테리어/건설/건축, #가정용품/선물용품 등 비교적 세세한 카테고리별 분류가 있었으나 그러한 카테고리가 너무 세세했던 탓인지 실제로 어떤 전시회가 포함되었나 확인해보면 추천 양이 거의 없었다.

그에 반해 *Atelier*는 본인이 직접 개최한 온라인 전시를 겸한 커뮤니티 기능에, 직접 개최한 전시를 기반으로 한 오프라인 전시회 추천 기능까지 제공한다. 이는 단순히 온라인상에서 본인의 전시를 개최하고 꾸미는 것에서 사용자의 경험을 제한하지 않고, 본인이 개최한 온라인 전시와 성격이 비슷한 오프라인 전시까지 추천해줌으로써 *Atelier*를 통해 체험한 미술적인 감정을 오프라인으로도 꼭 이어갈 수 있게 도와주는 역할을 한다.

위와 같은 기존 전시회 관련 사이트의 경우, 사람들이 많이 찾는 대중적인 전시회나 비교적 최신 전시회를 중심으로 추천을 해준다. 이는 개개인의 취향을 기반으로 둔 것이 아니므로, 사용자 본인에게 맞는 전시회를 추천받기는 어렵다. *Atelier*에서는 본인이 직접 개최한 온라인 전시회를 기반으로, 사용자 본인의 취향에 더욱 확고한 오프라인 전시까지 추천해준다.

2) 기존 SNS와의 비교



<그림3> 인스타그램의 주요 활용 용도

(사진을 통한 글쓰기, 짧은 동영상을 통한 일상 공유, 이들이 정리된 개인 프로필)

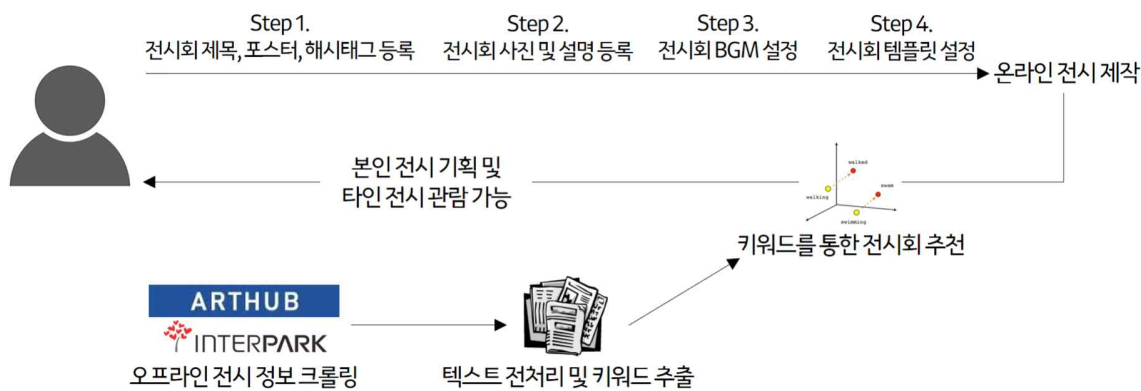
프로젝트 주제를 정하는 초기에 가장 많이 받았던 피드백으로, 기존의 SNS 서비스와 확연한 차이점이 돋보이지 않았다는 점이 있었다. 따라서 단순히 전시회 관람 후기가 올려지는 인스타그램 등과 같은 SNS와는 달리, *Atelier*는 개성이 돋보이는 개인만의 전시회를 개최한다는 의미를 더욱 명확하게 살릴 수 있도록 개발되었다.

*Atelier*를 가장 많이 이용할 것으로 예상하는 MZ 세대가 자주 사용하는, 인스타그램의 경우 개인마다 이용하는 목적이 다르다. 누구는 일기처럼 본인의 일과를 기록하기도 하고, 누구는 자신을 드러내고 과시하는 용으로 사용한다. 이렇듯 사용자마다 다른 목적을 가지고 다양하게 SNS를 이용한다. 그에 반해 *Atelier*는 ‘전시’라는 하나의 뚜렷한 목적을 가진 사용자들이 이용하므로 기존 SNS보다 전시에 특화된 기능 및 커뮤니티가 제공될 수 있다.

UI 등 외적인 부분에서도 실제 오프라인 전시를 감상하는 듯한 경험을 제공할 수 있도록 하였다. 본인이 찍어왔던 사진들을 통해 일정한 테마에 따른 전시를 기획하고, 계획한 전시와 어울리는 배경 음악 등을 설정하고, 사진 배치 및 각 사진에 대한 설명 등을 고민하게 하는 ‘단계별 전시회 제작’을 명시함으로써 실제 전시회를 기획한다는 느낌을 더욱 돋구었다.

2-1. 전체 시나리오

사용자가 *Atelier* 서비스를 사용하기 위한 전체 시나리오이다.



<그림 4> *Atelier* 서비스 전체 시나리오

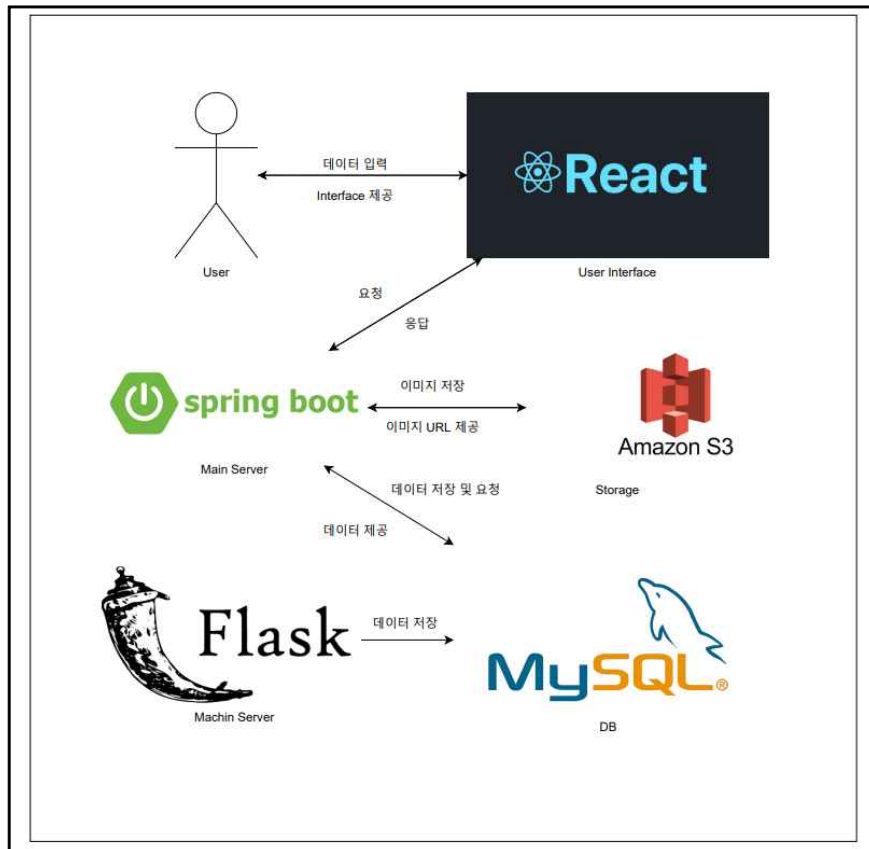
| | |
|----|---|
| 1. | 사용자가 서비스를 이용하기 위한 회원가입을 진행하고, 로그인한다. |
| 2. | 본인만의 온라인 전시회 제작을 시작한다. 전시회 제작은 총 Step 4단계로 이루어진다. |
| 3. | (Step 1.) 전시회 제목, 포스터, 전시회를 잘 나타낼 수 있는 해시태그 3개를 등록한다. |
| | <p>a. 이때 설정되는 해당 온라인 전시회의 해시태그 3개는, 추후 오프라인 전시회를 추천해주는 과정에서 사용된다.</p> <p>b. 추천될 오프라인 전시회와 관련한 정보는 사전에 크롤링된 상태로 텍스트 전처리 및 키워드 추출이 완료되어 DB에 저장되어 있다.</p> <p>c. 사용자가 해시태그를 선정할 때 너무 광범위하게 사용되는 단어보다는 적절한 특징이 담긴 단어를 선택하는 것이 전시회 추천의 성능을 높일 수 있다.</p> |
| 4. | (Step 2.) 전시회에 올릴 사진들 및 각각에 대한 설명을 등록한다. |
| 5. | (Step 3.) 등록된 4개의 테마(귀여운, 잔잔한, 신나는, 밝은)에서 사용자가 원하는 전시회 BGM을 설정한다. |

| | |
|----|--|
| | a. 사용자가 원하는 특정한 BGM을 추가하게 하려 했으나, 경우에 따라 저작권 문제가 발생할 수 있기에 BGM 목록을 미리 설정했다. |
| 6. | (Step 4.) 온라인 전시를 마무리하기 위한 템플릿을 설정한다. |
| 7. | 완성된 본인만의 온라인 전시회를 확인하고, 타인의 온라인 전시회에 방문하여 댓글 및 좋아요 등의 리액션을 남길 수 있다. |

<표2> Atelier 서비스 전체 시나리오

2-2. 웹 개발: Frontend

웹을 통해 사용자와의 서비스가 이루어지는 구조는 다음과 같다.

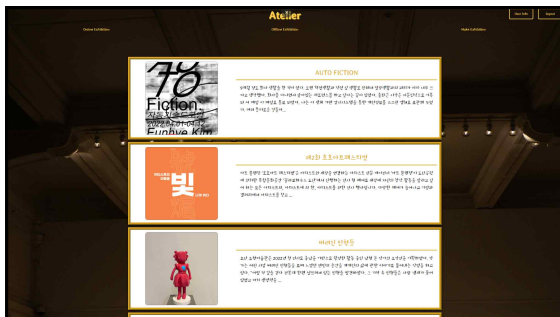


<그림5> Atelier 웹 서비스 구조

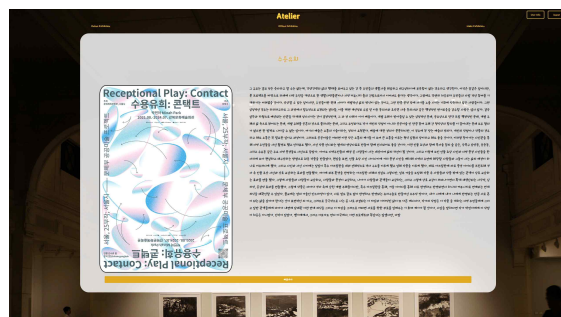
프론트엔드 개발은 전반적으로 React를 활용해 진행하였다. 디자인의 경우 React bootstrap을 참고하여 각종 레이아웃을 만들 때 활용하였다. 세부 페이지로는 메인페이지, 전시회 관련 커뮤니티 등을 구현하였다. 특히 전시회 관람 기능 페이지들의 경우, 실제 전시회를 관람하는 듯한 느낌이 날 수 있도록 프론트엔드 개발을 진행하였다.



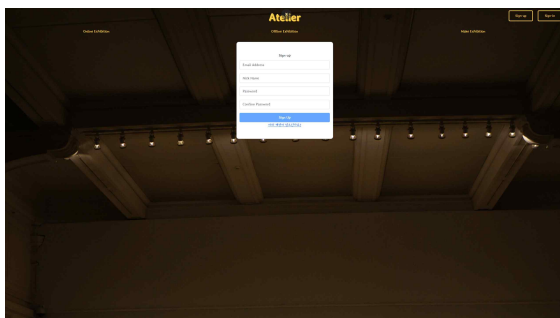
<그림6> Atelier 메인 페이지



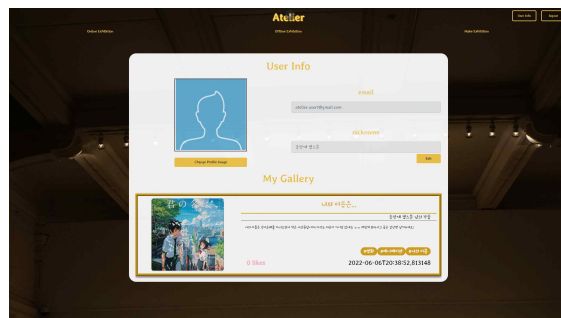
<그림7> Atelier 오프라인 전시 목록 페이지



<그림8> Atelier 특정 전시회 정보



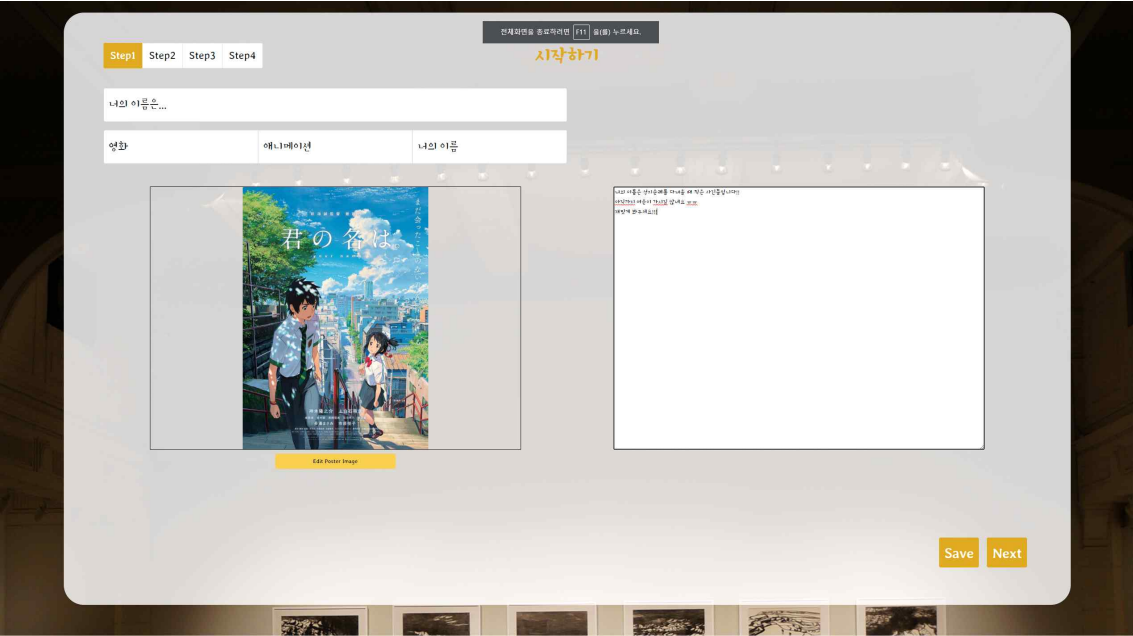
<그림9> Atelier 로그인 페이지



<그림10> Atelier 회원 정보 페이지

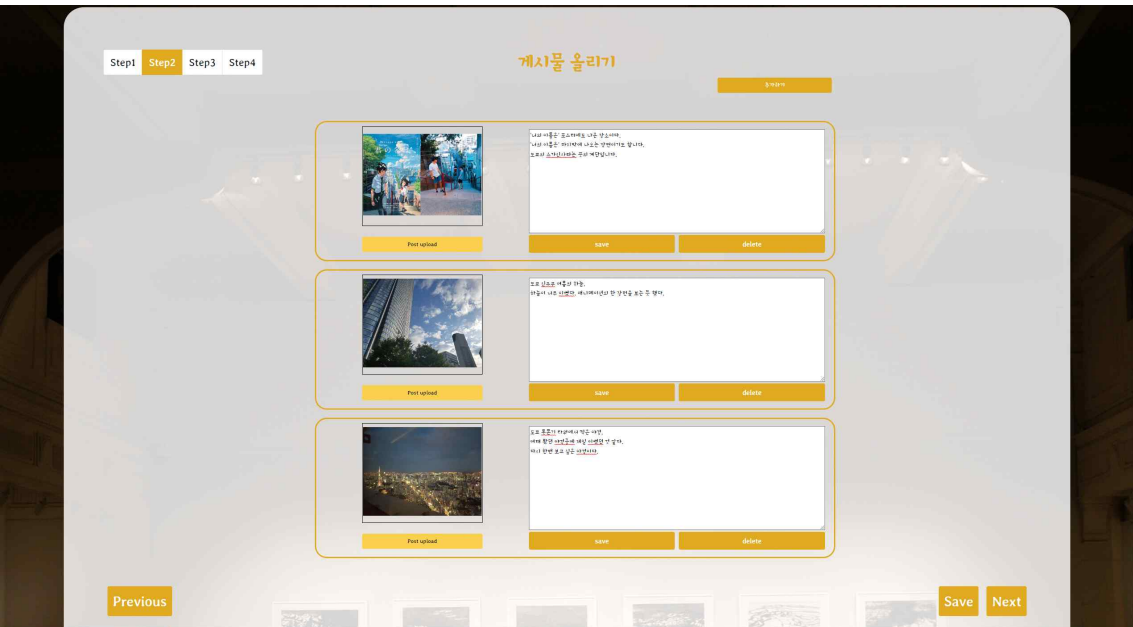
막상 전시회를 제작한다고 해서 자칫 복잡하게 느껴질 수 있는데, *Atelier*는 총 4가지 단계로 사용자만의 온라인 전시를 구성할 수 있다.

‘Step 1’에서는 전시회 제목, 포스터, 해시태그를 등록한다. 사용자가 만들 온라인 전시회의 제목과 포스터 이미지를 등록하고, 해당 전시회를 가장 잘 대표할 수 있는 해시태그 3개를 작성한다. 이때 작성된 해시태그는, 추후 오프라인 전시회를 추천할 때 활용된다.



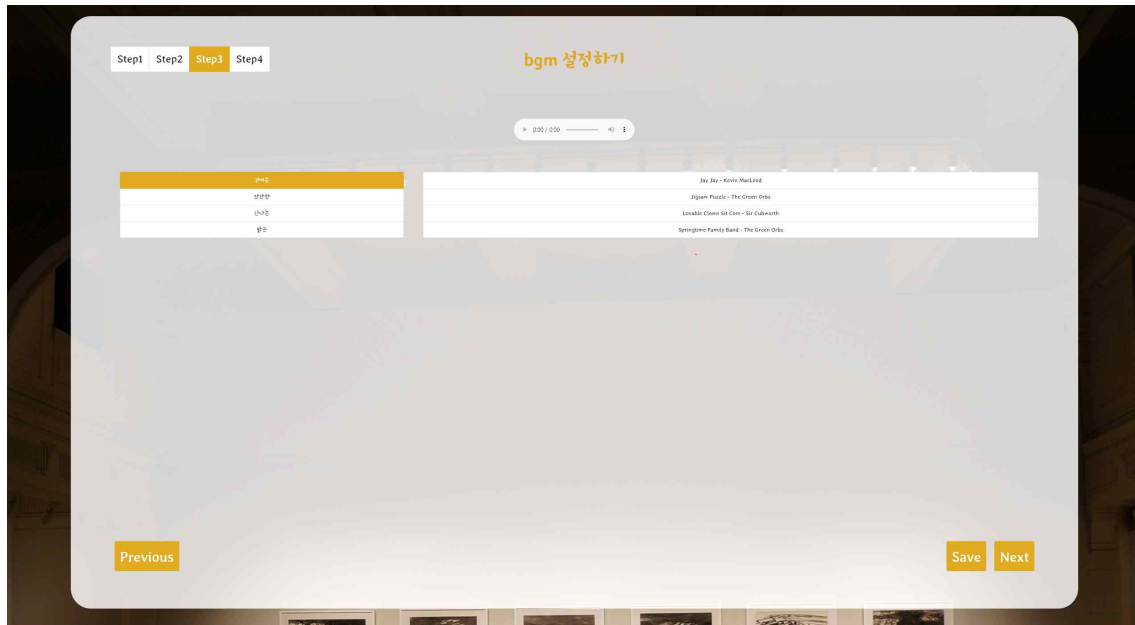
<그림 11> 온라인 전시회 제작 Step 1

‘Step 2’에서는 전시회 사진 및 설명을 등록한다. 사용자의 전시회에 올라갈 사진을 업로드하고, 그에 따른 적절한 설명을 함께 작성한다.



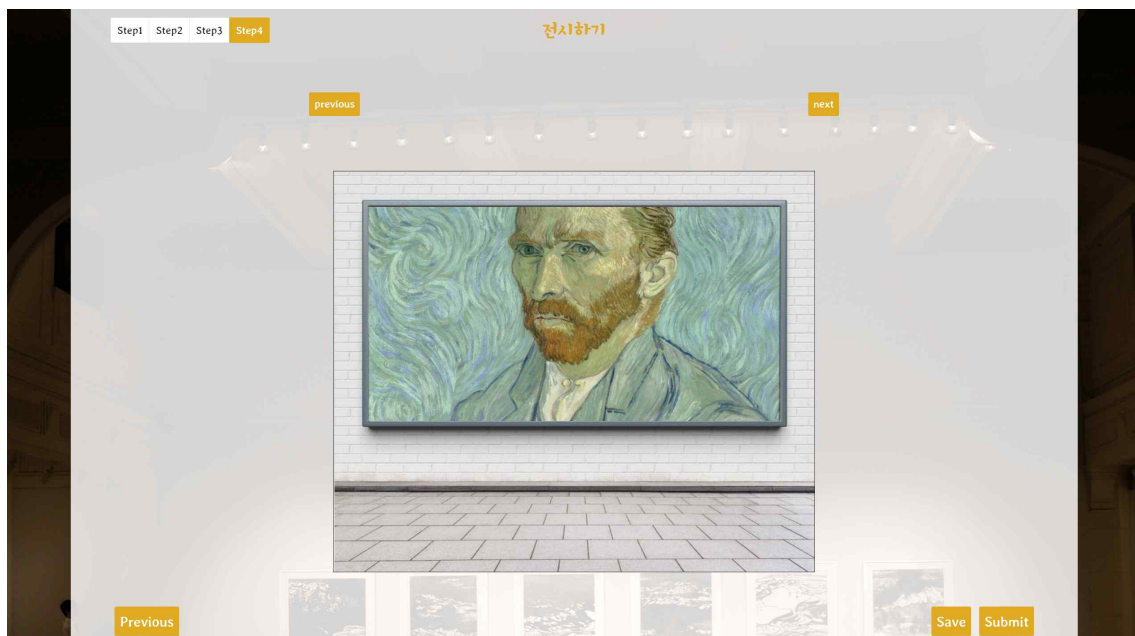
<그림 12> 온라인 전시회 제작 Step 2

‘Step 3’에서는 전시회 BGM을 설정한다. 사용자의 전시회에 어울리는 BGM을 설정하며, 이때 BGM은 총 4가지 테마인 귀여운/잔잔한/신나는/밝은 노래 중 하나를 선택할 수 있다. 사용자가 직접 올리는 BGM은 저작권 문제가 우려되어 허용하지 않았다.



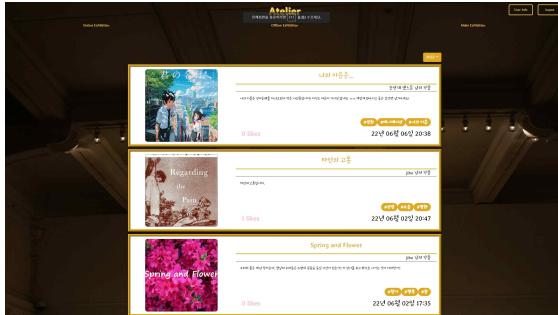
<그림13> 온라인 전시회 제작 Step 3

마지막으로 ‘Step 4’에서는 전시회 템플릿을 설정한다. 사용자가 업로드한 사진이 어떤 프레임에 담기게 할지 설정할 수 있다. 이렇게 총 4단계의 과정을 거침으로써 본인만의 온라인 전시회를 만들 수 있다. 이렇게 본인만의 온라인 전시회를 제작할 수 있으며, 이때 사용한 해시태그에 따라 오프라인 전시회를 추천받을 수 있다.

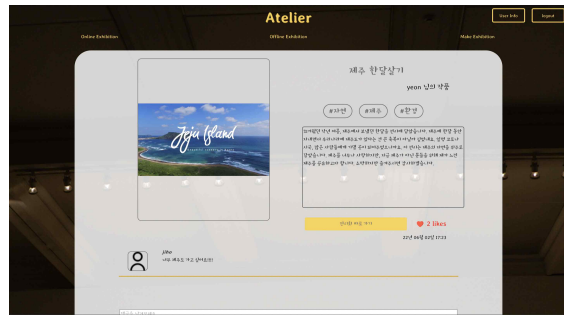


<그림14> 온라인 전시회 제작 Step 4

만들어진 전시회는 온라인 전시 목록 페이지에서 확인할 수 있다. 이때 타인의 온라인 전시 역시 관람할 수 있다. 단순히 본인만의 전시회를 기획하고 업로드하는 것을 넘어, 타인의 전시회에 방문하여 그에 따른 리액션을 남길 수 있다. 특정 전시회에 좋아요를 남겨 호감을 표현하거나, 댓글로 본인만의 감상평을 남김으로써 온라인 전시 커뮤니티를 더욱 알차게 활용할 수 있다.



<그림15> Atelier 온라인 전시 목록 페이지



<그림16> 타인 전시회에 리액션 남기기

2-3. 웹 개발: Backend

백엔드 개발은 전반적으로 Springboot를 활용해 진행하였다. Atelier 서비스 이용을 위한 사용자 회원가입을 구현하였다. 이메일 및 기타 정보들을 입력하면 DB에 새로운 유저로 추가되며, 이때 등록한 이메일로 인증이 진행된다. 이외에도 온라인/오프라인 전시회 DB로부터 불러오기 등의 API를 구현하였다.



아틀리에 이메일 인증 받은편지함 x



caucse.atelier@gmail.com

나에게 ▼

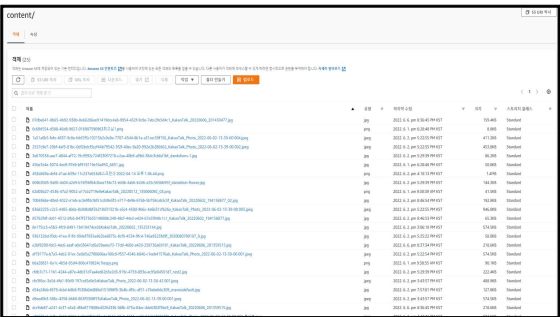
오후 7:57 (0분 전)



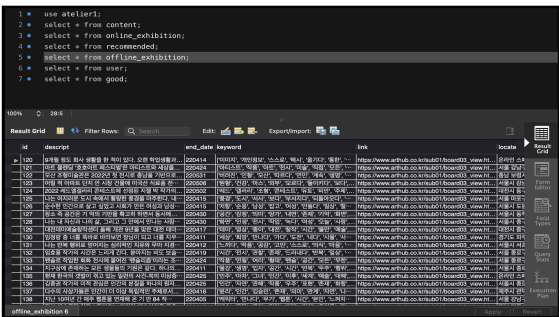
<http://localhost:8080/api/confirm-email?token=00484b55-40b1-45e7-8126-d70ebab32d59>

<그림17> Atelier 회원가입시 이메일 인증

사용자가 온라인 전시회를 만들 때 프로필 사진이나 전시회 포스터 등의 이미지는 AWS S3 클라우드 스토리지에 저장하였다. 데이터들을 이용할 때는 MySQL을 통해 online_exhibition, offline_exhibition, user 등의 테이블을 만들어 사용하였다.

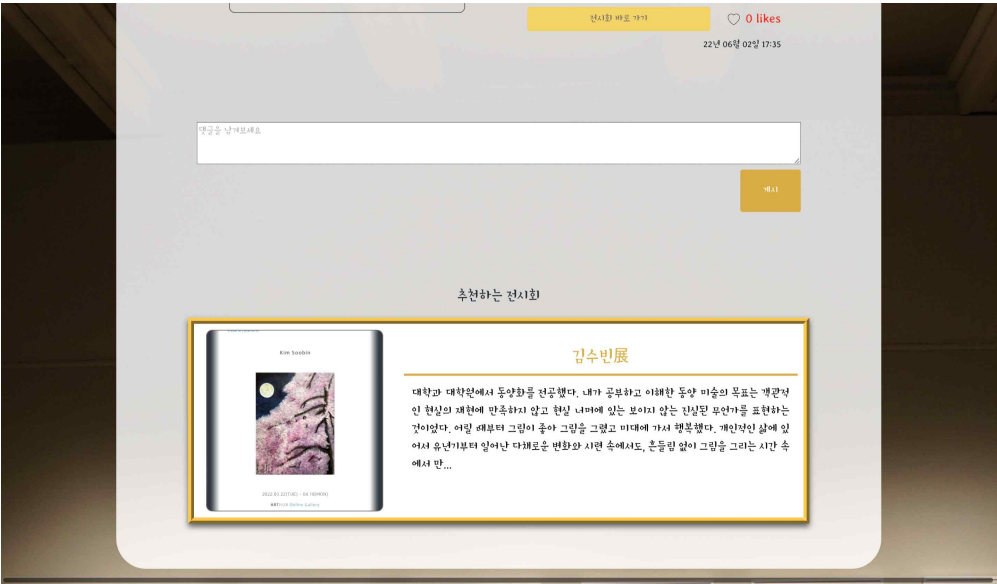


<그림18> AWS S3 클라우드 스토리지 이용



<그림19> MySQL 이용

또한, 온라인 전시회 제작 과정에서의 해시태그를 통해 적절한 오프라인 전시회를 추천하고 자 전시회 추천 API를 구현하였다. 해시태그 3개를 통해 추천된 정보를 확인할 수 있으며, Spring에서 Flask API를 호출하였다.



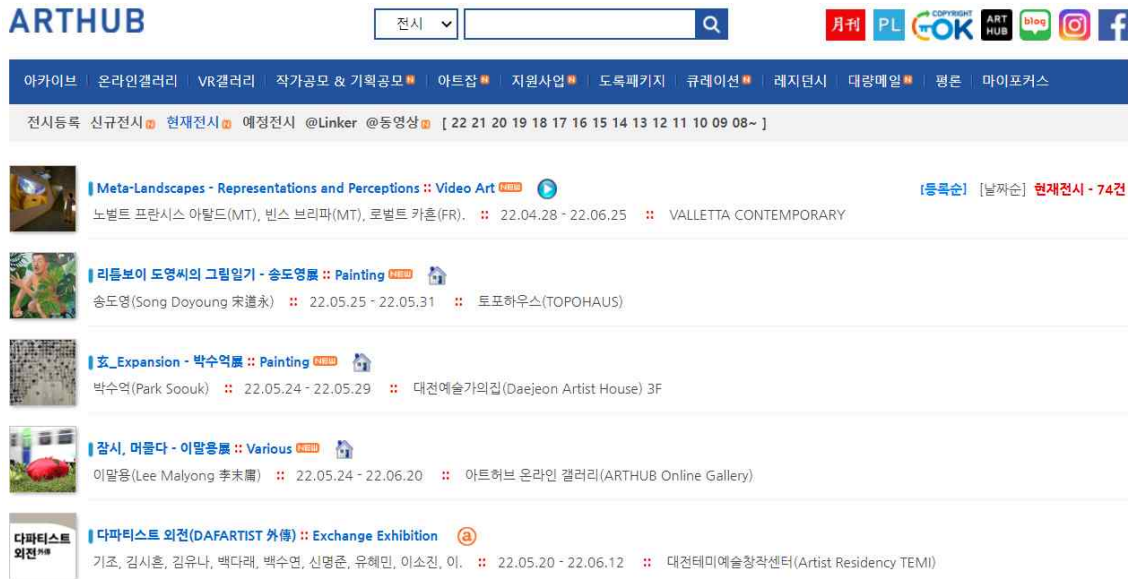
<그림20> Atelier 온라인 전시회를 통해 추천된 오프라인 전시회

2-4. 기술 개발: 크롤링 및 전처리

추후 사용자가 온라인 전시회를 만들었을 때, 해당하는 전시회의 해시태그에 따라 적절한 오프라인 전시회를 추천하기 위하여 웹상의 오프라인 전시회 정보를 크롤링하였다. 어느 웹

사이트에서 크롤링해야 할지 여러 가지 후보들이 있었지만, 디지털 아카이브의 역할로서 수많은 전시회 정보들이 담겨 있는 아트허브와, 많은 사람이 전시회를 예약할 때 찾는 대중적인 사이트인 인터파크 총 두 군데에서 크롤링을 진행하였다.

1) 아트허브 크롤링



<그림21> 아트허브 - 현재전시 페이지

상단의 아트허브 - 현재전시 페이지에 접속해서 (크롤링 진행 일자 기준) 나열된 전시회들을 모두 크롤링하였다. 아트허브와 같은 경우는 한 페이지 내에 모든 전시회 정보들이 있었기에, 위의 페이지에서 전시회 정보로 이어지는 링크들을 크롤링할 수 있었다.

이렇게 한 페이지 안에 있는 모든 전시회의 링크들이 모아지면, 각각의 링크에 따로 들어가서 세부적인 전시회 정보를 크롤링하였다.



<그림22> 아트허브 내 특정 전시회 정보

[illegible]

5. 장로회 작가 개관의 필요성을 극대화하는 방안: 작가 회원이나 자원봉사 사제들 중에 전문성 제로 쓰임에 한하여 대한 반포문 제호의 개관과 그 후속의 극대화 협업을 보여주고 싶은 취지이다. 자원봉사자 전문 역량을 구상해서도 교육적으로도 정당한 대가를 지불하여 대외적으로 제호를 확장하고 개관한 실용과 교육 효과도 있다. 다시 말해, 출판과 제호의 혼동을 막고 글과 제호

〈그림23〉 아트허브 내 특정 전시회 내용

13

2) 인터파크 크롤링

The screenshot shows the Interpark website's '전시/행사' (Exhibition/Event) page. The page has a red header with navigation links like '유지컬', '콘서트', '연극', etc. The main content area is titled '전시/행사 EXHIBITION | 전시' and lists several exhibitions. The first exhibition is '경복궁의 인문학산책' (In-depth Cultural Walk in Gyeongbokgung) at Gyeongbokgung, running from 2018.06.23 to 2022.12.31. The second is '반고흐빌리지展' (Van Gogh Village Exhibition) at Bonda Banchi Museum Gunsan, running from 2021.06.12 to 2022.05.29. The third is '구하우스 미술관 관람권' (Goo House Museum Admission Ticket) at Goo House, running from 2019.01.01 to 2022.12.31.

<그림 24> 인터파크 - 전시/행사 페이지

상단의 인터파크 - 전시/행사 페이지에 접속해서 (크롤링 진행 일자 기준) 나열된 전시회들을 모두 크롤링하였다. 인터파크와 같은 경우도 한 페이지 내에 모든 전시회 정보들이 있었기에, 위의 페이지에서 전시회 정보로 이어지는 링크들을 크롤링할 수 있었다.

이렇게 한 페이지 안에 있는 모든 전시회의 링크들이 모아지면, 각각의 링크에 따로 들어가서 세부적인 전시회 정보를 크롤링하였다.

The screenshot shows the details for the '반고흐빌리지展' (Van Gogh Village Exhibition). The page includes a poster image of Vincent van Gogh, the exhibition title '반고흐빌리지展', the venue '본다빈치뮤지엄 군산' (Bonda Banchi Museum Gunsan), the dates '2021.06.12 ~ 오픈런', and ticket information '전체 관람가' (All Ages) and '입장권 13,500원'. Annotations point to the exhibition title, venue, dates, and poster image.

<그림 25> 인터파크 내 특정 전시회 정보

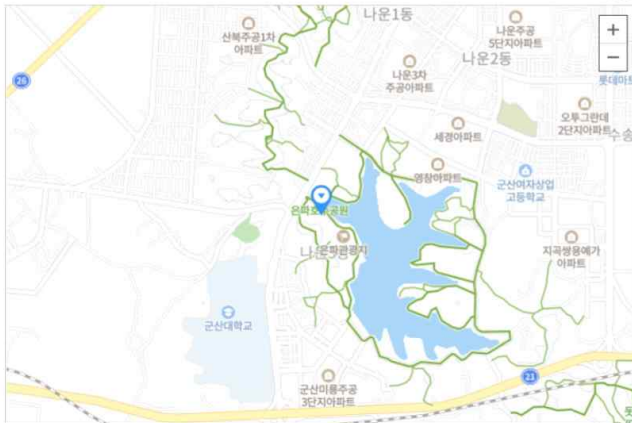
각각의 링크에 따로 들어갔을 때 보이는 화면이다. 인터파크와 같은 경우도 위의 그림과 같이 세부적인 전시회 정보를 확인할 수 있었다. 그중에서도 전시회 제목, 전시회 시작 일자 및 종료 일자, 전시 진행 장소, 포스터 이미지 링크를 크롤링하였다.

공연장 정보

×

본다빈치뮤지엄 군산

주소 전라북도 군산시 은파순환길 20 (나운동) 호텔 은파팰리스 지하 1층



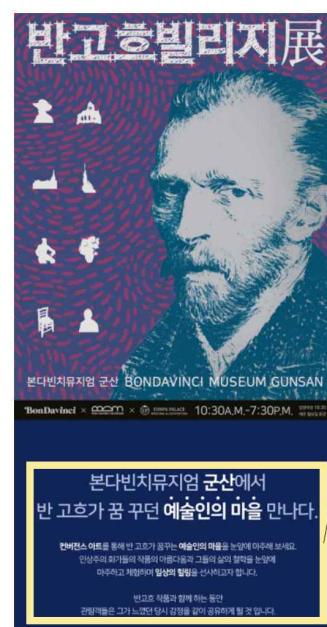
전시 진행 위치

<그림26> 인터파크 내 특정 전시회 정보 중 위치

인터파크의 경우 아트허브와는 달리 전시 진행 위치가 별도의 창에 따로 보이는 형태였다. 따라서 크롤링할 때도 마찬가지로 webdriver가 별도의 창을 띄워 전시 진행 위치를 저장하고, 해당 창을 닫게 함으로써 전시 진행 위치도 모두 크롤링하였다. 이렇게 크롤링된 내용들은 추후 Atelier 웹에서 오프라인 전시회 정보를 보여주는 과정에 사용되었다.

제목 등과 같은 전시회 정보는 전시회 정보 링크 상단에서 바로 추출할 수 있었지만, 전시 내용은 하단의 본문에서 따로 추출해야 했다.

아트허브의 경우는 전시 내용이 텍스트 형태로 웹에 게시되어 있어 크롤링을 통해 긁을 수 있었지만, 인터파크의 경우는 전시 내용이 이미지 내부에 포함되어 있었다. 따라서 인터파크에 게시된 각각의 전시 내용을 가져오기 위해서는 이미지상의 내용을 텍스트로 변환하는 과정(이하 image2text)이 필요했다.



전시내용

<그림27> 인터파크 내 특정 전시회 내용

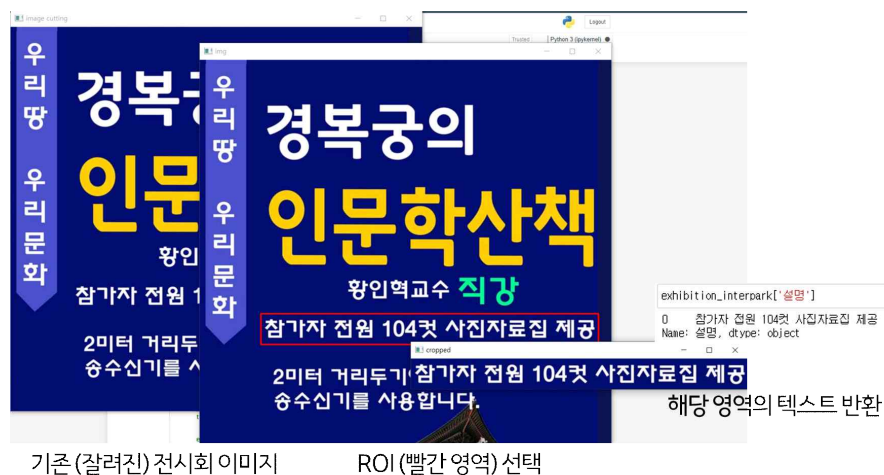
우선 전시 설명이 포함된 이미지가 세로로 매우 긴 형태였기 때문에, image2text를 원활하게 진행하기 위해서는 그림을 적절한 크기로 잘라야 했다. 따라서 파이썬 PIL 라이브러리의 Image를 사용하여 기존의 이미지 하나를 여러 장의 이미지로 잘랐다.



<그림 28> 전시 정보 이미지 텍스트화

그렇게 잘려진 각각의 이미지에서는 파이썬 pytesseract 라이브러리를 사용하여 텍스트화를 진행하였다. Pytesseract는 파이썬용 광학 문자 인식(OCR) 도구로서 이미지에 포함된 텍스트를 인식하고 읽어낸다. 이때 잘려진 이미지에서 모든 내용을 다 인식하는 것이 아니라, 직접 ROI(Research of Interest)를 지정함으로써 해당하는 영역에서만 image2text가 진행되도록 하였다.

아래 그림을 예시로 보면, 잘려진 전시회 이미지 전체에서 ‘참가자 전원 104컷 사진 자료집 제공’에 해당하는 부분만 빨간색 박스로 ROI를 지정하였고, 해당 부분에서만 pytesseract를 통해 image2text가 진행되어 이미지 속 ROI의 내용이 텍스트로 변환됨을 확인할 수 있다.



<그림 29> ROI를 통한 image2text 구현

image2text의 정확도를 더욱 높이하고자 텍스트화를 거치기 전에 파이썬 cv2 라이브러리를 통하여 이미지를 grayscale로 변환하고 thresholding을 추가로 진행하였다. 입력으로 들어오는 전시 정보 이미지들이 각기 다른 색상으로 구성되어 있으며, 그 안에서 글자 크기도 일정하지 않고 모두 달랐기 때문에 위와 같은 적절한 이미지 전처리 과정을 수행해주었다.

대부분의 경우에서 image2text가 기대한 성능을 보여주었지만, 일부분에서 텍스트화가 제대로 되지 못한 모습을 보여주었다. image2text의 성능을 높이기 위해 이미지 전처리를 진행하고 pytesseract가 동작하는 인자도 여러 가지 경우로 실험해봤지만, 그럼에도 성능적인 측면에서 아쉬움이 존재하였다. 이러한 부분을 교정하기 위해 추후 텍스트 전처리 과정에서 기존 다른 태스크에서의 전처리보다 더욱 세심한 전처리를 진행하였다.

이렇게 인터파크 사이트에서는 2022년 4월 18일자로 총 한 번의 크롤링이 이루어졌으며, 그 결과 #건의 전시회 정보들을 축적할 수 있었다.

파이썬 selenium 라이브러리의 webdriver를 통해 인터파크 사이트 내에서의 모든 크롤링이 이루어졌다. 크롬 개발자도구의 element selector를 사용하여 수집 대상의 element를 선택하고, 해당 element의 css selector 혹은 xpath 정보를 가져왔다.

그중에서도 전시 시작 일자 및 종료 일자, 전시 진행 장소를 추출할 때에는 일부 특수 문자를 제거하는 기초 전처리 과정을 거쳤다.

3) 텍스트 전처리

아트허브와 인터파크 사이트 내의 전시회 정보 크롤링을 통해 텍스트 데이터를 축적할 수 있었으며, 이를 추후 활용하기 위한 텍스트 전처리가 진행되었다. 주로 파이썬 re 라이브러리를 통해 다양한 정규표현식을 사용했다. 다른 특수한 라이브러리를 사용하는 경우 하단의 표에 언급하였다.

수월한 전처리를 위해 두 가지 클래스를 구성하여 사용하였는데, 그중 Remover() 클래스는 텍스트 자체에 대한 전처리를 수행하며, Denoiser() 클래스는 제목/날짜/설명 등 특정한 형태의 텍스트가 들어왔을 때 Remover() 클래스의 메서드를 사용한 전처리를 수행한다.

| 전처리 목록 | 특징 |
|------------|---------------|
| 공백 제거 | 정규표현식으로 형태 지정 |
| 괄호 안 내용 제거 | |
| 특수문자 제거 | |
| url 제거 | |
| 한자 제거 | |

| | |
|--------|---|
| 맞춤법 교정 | 파이썬 hanspell 라이브러리 사용, 틀린 맞춤법 옳게 교정. |
|--------|---|

<표3> 전처리 Remover() 클래스 구성

| 전처리 목록 | 특징 |
|------------------|--|
| 아트허브 内 제목 전처리 | 제목에서 불필요한 특수문자 제거 |
| 아트허브/인터파크 날짜 전처리 | 기존의 날짜 형태를 6자리 정수로 변환 |
| 아트허브/인터파크 위치 전처리 | 정규표현식으로 형태 지정 |
| 인터파크 内 장소 전처리 | 텍스트 표현 형태 수정 |
| 아트허브 内 설명 전처리 | Remover() 클래스 사용, 공백/괄호 안 내용/ 특수문자/url/한자 제거. 파이썬 pykospacing 라이브러리 사용, 틀린 띄어쓰기 옳게 교정. |
| 인터파크 内 설명 전처리 | Remover() 클래스 사용, 공백/괄호 안 내용 제거. 및 맞춤법 교정. 파이썬 pykospacing 라이브러리 사용, 틀린 띄어쓰기 옳게 교정. |

<표4> 전처리 Denoiser() 클래스 구성

위 두 개의 클래스를 이용해 텍스트 전처리를 수행하였다. 경우에 따라 특정한 전시회의 설명이 결측값으로 채워진 경우는, 이를 미리 제목으로 대체한 후에 전처리가 진행되었다.

2-5. 기술 개발: 키워드 추출 및 전시회 추천

사용자의 해시태그 3개가 입력됨에 따라 적절한 전시회를 추천하고자 키워드 추출 알고리즘을 활용했으며, 사전 훈련된 단어 임베딩 모델을 선정 후 추출된 키워드의 단어 임베딩 값과 사용자 해시태그의 단어 임베딩 값을 비교하여 사용자 취향에 맞는 전시회를 추천하였다.

1) 키워드 추출

키워드 추출 알고리즘으로는 KR-WordRank¹⁾을 선택하였다. 다른 일반적인 언어와는 달리 한국어는 원래의 단어에 문법적인 기능을 가진 요소가 결합하는 교착어의 특성을 갖는데, 해

1) Kim, H. J., Cho, S., & Kang, P. (2014). KR-WordRank: An Unsupervised Korean Word Extraction Method Based on WordRank. Journal of Korean Institute of Industrial Engineers, 40(1), 18-33

당 알고리즘은 이러한 특성을 살려 주어진 텍스트에서의 단어 후보들을 걸러내었다.

한국어의 명사구는 명사 단독으로 사용되거나 ‘명사+조사’ 혹은 ‘명사+명사’의 형태이고, 동사구나 형용사구는 ‘어근+어미’의 형태이다. 명사나 어근은 띄어쓰기를 기준으로 나눈 토큰의 왼쪽에 위치하고, 어미나 조사는 오른쪽에 위치하게 되는 것인데 KR-WordRank는 이렇게 각 토큰 내에서의 위치 정보를 활용한다.

전시회 설명

KR-WordRank로 단어추출

'나는 어지러운 도시 속에서 황량한 풍경을 마주한다. 내가 발 딛고 선 곳이 그 곳이 기 때문이다. "어디에 설 것인가?" 하는 물음은 좀처럼 발 디딜 틈 없는 내 주변에 대한 답답함과 허탈함으로 되돌아온다. 도시는 어디론가 바쁘게 움직이고 있지만, 알고 보면 그저 바쁘게 가고 있는 시늉만 하고 있는 것이 아닌가 상상해 본다. 지나가는 이에게 어디에 가는지, 왜 가는 지 물어도 아무도 답하지 않는다. 온 도시의 움직임이 마치 목적지 없는 피난 행렬 같다. 나는 주변의 모두가 향하는 방향의 반대편에서 이 풍경을 발견했다. 도시의 모든 것은 설계 당시의 의도대로 작동할 때 유용하다. 하지만 도시의 풍경을 유기적으로 만드는 것은 그 무용함에 있다. 벽들 사이에서 자라난 잡초와 가로수, 무서진 콘크리트와 어지럽게 쌓인 쓰레기 더미의 풍경은 사람의 의도로 만들어진 것이 아니지만, 특별한 목적을 위해 설계되지 않았다는 점에서부터 서사가 부여된다. 사람들은 자기 자신이 선 자리에서 세상을 본다. 나의 발로 밟은 땅, 나의 눈과 귀로 보고 들은 것들을 통해 세상은 점차 확고해진다. 나의 세계도 감각과 기억들로 차츰 쌓여간다. 하지만 작가로서 그렇게 인식한 세상이 쉽사리 확고해지지 않는 이유는, 그 하나하나가 모두 의문으로 쌓아 올린 위태로운 것이기 때문이다. 그 풍경을 가능한 사실적으로 옮겨 보려는 나의 붓질은 그 모든 물음에 나름의 답을 찾기 위한 노력이고, 어느 하나 확신으로 다져지지 어려운 세상에 대한 불안함을 덜어내기 위한 몸부림이다. 작업들은 4년간의 중국 생활과 귀국 후 도시생활에서 발견한 풍경을 그린 것이다. 미처 매듭짓지도 못하고 열쳐버리지도 못했던 미련 같은 생각들이 전염병 사태와 생계 활동이라는 지루한 터널을 지나며 차근차근 작업의 형태로 소화되었다. 이번 전시를 통해 지어진 한 매듭이 내가 발 딛고 선 땅의 의미를 더 깊이 발견하는 매개가 되기를 희망한다.'

['나의',
'풍경을',
'도시',
'위한',
'나는',
'것은',
'대한',
'서사가',
'하지만',
'미처',
'않는',
'모두',
'못하고',
'본다',
'무서진',
'않았다는',
'허탈함으로',

<그림30> KR-WordRank의 단어 추출

기존의 전시회 설명 텍스트에 KR-WordRank의 단어 추출을 적용한 결과는 위 그림과 같다. 중요도에 따른 단어가 잘 추출된 듯 보이지만, 단어를 하나씩 살펴보면 ‘나의’, ‘나는’과 같이 키워드를 선정하는 과정에서 비교적 중요하지 않은 단어 ‘나’와 함께 그에 따른 조사 ‘의’와 ‘는’이 결합된 형태가 추출됨을 확인할 수 있었다.

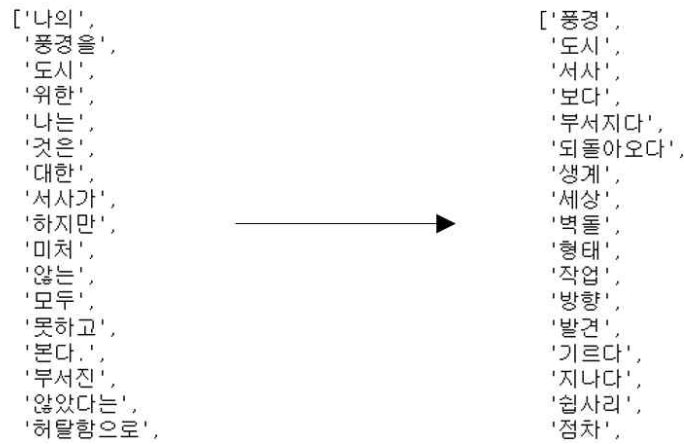
‘위한’과 ‘미처’와 같은 부사 형태 역시 등장하기도 했다. 키워드에 가장 가까워 보이는 단어인 ‘풍경을’이라는 단어의 경우, ‘을’이라는 조사까지는 필요하지 않았으나 해당 조사까지 모두 결합된 형태로 추출됨을 확인할 수 있었다.

이는 KR-WordRank의 추출 과정에 의한 것인데, KR-WordRank는 한글의 문법적인 형태보다는 실제로 사람들이 빈번하게 사용하는 ‘형태’를 단어로 인식하기 위한 알고리즘이기 때문이다. 보통은 ‘풍경을’과 같이 조사까지 결합된 형태를 단어로 사용하지 않지만, KR-WordRank가 ‘을’이라는 조사까지 결합된 ‘풍경을’이라는 하나의 형태를 빈번하게 사용하는 단어 그 자체로 인식한 것이다.

따라서 조금 더 정제된 형태의 키워드를 만들어주고자 추출된 단어 형태에서 조사를 분리하는 과정이 필요했으며, ‘나’와 같이 자주는 사용되나 그 중요도는 낮은 불용어를 제거하는 과정 역시 필요했다. 여기에 만약 동사와 같이 어간 및 어미가 일부 변형될 수 있는 품사의 단어가 등장하는 경우 어간 추출까지 진행하였다.

KR-WordRank로 단어추출

키워드 정제 작업 이후



<그림31> KR-WordRank로 추출된 단어에 정제 작업 진행

파이썬 konlpy 라이브러리의 Okt를 통해 기존의 단어 형태에 숫자 혹은 불용어가 포함되거나, 각 단어의 품사가 ‘형용사’ / ‘부사’ / ‘조사’ / ‘구두점’ / ‘외국어’ 중 하나로 판단되는 경우 해당 단어는 제거해주었다. 그전에는 단어별로 형태소 분석 및 어간 추출을 진행하는 과정이 선행되었다.

KR-WordRank로 추출된 단어에 키워드 정제를 적용한 결과는 위 그림과 같다. 기존과는 달리 ‘나’라는 불용어가 포함된 경우와, ‘위한’과 ‘미처’와 같은 부사 형태 모두 제거되었다. 키워드라고 여겨졌던 형태의 ‘풍경을’이라는 단어는 조사 ‘을’이 제거되고 ‘풍경’이라는 단어만 남았으며, ‘부서진’이라는 단어는 어간 추출이 진행되어 그 원래 형태인 ‘부서지다’로 바뀌었음을 확인할 수 있다. 단순히 KR-WordRank만 사용하는 것이 아니라 필요에 따른 키워드 정제 작업을 진행함으로써 양질의 키워드를 획득할 수 있었다.

2) 전시회 추천

나만의 전시회를 제작할 때 입력되는 사용자 해시태그 3개와, KR-WordRank 및 키워드 정제 작업을 통해 얻은 오프라인 전시회들 각각의 키워드를 비교하기 위해 사전훈련된 단어 임베딩 모델을 선택하였다.

단어 임베딩이란 주어진 단어를 그 의미에 따라 적절한 위치의 벡터 형태로 변환하고, 해당 벡터를 조밀한 차원에 표현하게 되는 개념이다. 임베딩 모델의 종류에 따라 주어진 단어를 벡터 형태로 바꾸는 과정이 다양하며, 그에 따라 같은 단어일지라도 어떤 모델을 선택하였느냐에 따라 각기 다른 벡터로 반환될 수 있다.

그중에서도 대표적으로 사용되는 word2vec, fasttext, GloVe 총 3가지 모델을 선택하였다. 각 모델에서 단어는 밀집 벡터로 표현되며 불필요한 공간상의 낭비가 없고 단어의 의미적 유사성을 잘 파악할 수 있다는 장점이 있다.

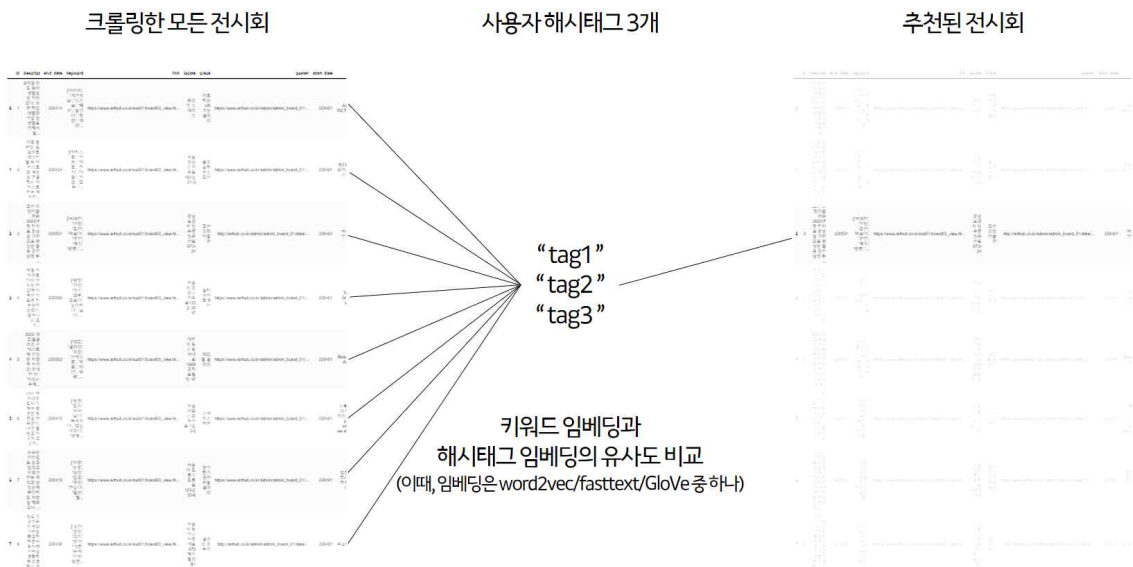
만약 현재 축적해둔 전시회 관련 데이터가 상당히 많았다면 해당 데이터를 이용하여 3가지 모델 각각을 새로 학습해 그에 따른 임베딩 값을 얻었을 수도 있었겠지만, 전시의 특성상 진행되는 기간이 짧지 않기 때문에 크롤링과 image2text를 통해 전시회 관련 텍스트를 모두 모아도 이를 가지고 모델 학습을 진행하기에는 턱없이 부족한 양이었다.

그에 따라 사전학습한 단어 임베딩 결과를 사용해야 했다. 그중에서도 이기창님께서 한국어 위키피디아, 네이버 영화 말뭉치, KorQuAD 데이터셋을 은전한닢으로 형태소 분석한 말뭉치를 가지고 임베딩하신 파일을 사용하였다.

| | word2vec | fasttext | GloVe |
|-----------|-----------------------|----------|--------|
| 학습 방법 | 이기창님의 사전학습된 임베딩 파일 사용 | | |
| 단어 차원 | 100차원 | 100차원 | 100차원 |
| 임베딩 로드 시간 | 1~3초 | 35~38초 | 10~12초 |

<표5> 단어 임베딩 모델

위의 표는 각 임베딩 파일에 대한 세부사항이다. 단어 차원은 모두 100차원으로 같지만, 각각의 임베딩 파일을 불러오는 로드 시간이 너무 제각기로 달랐다. 단순히 모델이 등장한 순서에 따라 word2vec, fasttext, GloVe 순으로 성능이 점차 나아질 것이라는 기대가 있었지만, 그렇다고 일부 모델을 부를 때 10초가 넘어가는 시간을 웹서비스에서 사용하기에는 너무 길다고 생각했다. 따라서 두 단어의 임베딩 유사도에 따른 비교 과정을 일단 구현한 후, 전시회 추천을 가장 잘 수행하는 모델을 선택하기로 하였다. 정리하자면, 모델을 불러오는 시간은 일단 뒤로하고 가장 좋은 성능을 보여주는 모델이 무엇인지 확인하였다.



<그림 32> 임베딩 모델을 통한 전시회 추천 과정

임베딩 유사도에 따른 비교 과정은 위 그림과 같다. 기존에 KR-WordRank를 이용해서 전 사회별 키워드를 상위 30개를 추출하였다. 이렇게 많은 개수를 저장한 이유는 사전훈련된 단어 임베딩에, 전사회별 키워드가 학습되지 않아 OOV(Out of Vocabulary)로 존재할 가능성을 염두했기 때문이다.

30개 중 가장 중요도가 높은 키워드의 임베딩부터 하나씩 차례대로 사용자 해시태그 임베딩과의 코사인 유사도를 비교하였다. 만약 키워드가 OOV였다면, 해당 단어는 건너뛰고 다음 키워드로 계산하였다. 이렇게 총 5번이 카운트되면 비교 과정을 마쳤으며 한번이 카운트 될 때마다 일정한 가중치를 부여한 코사인 유사도를 합산하였다. 그 결과 임베딩 모델에 따른 전시회 추천 결과는 아래와 같다.

| 사용자 해시태그: 청소, 청결, 일상 | | | |
|----------------------|---|---|---|
| | word2vec | fasttext | GloVe |
| 추천된 전시회 제목 | 후지시로 세이지 북촌스페이스 | 후지시로 세이지 북촌스페이스 | Re |
| 추천된 전시회 설명 | 북촌의 계동길에 위치한 옛 중앙탕은 1960년대부터 목욕탕으로 사용되었던... | 북촌의 계동길에 위치한 옛 중앙탕은 1960년대부터 목욕탕으로 사용되었던... | 배태열 작가의 '-'에서는 버려지고, 쓸모 없어진 나무를 그들이 쓰임새... |
| 추천된 전시회 키워드 | 사위실, 북촌, 중앙, 사용... | 사위실, 북촌, 중앙, 사용... | 버려진, 작업, 암흑, 나무... |
| 사용자 해시태그: 향기, 행복, 꽃 | | | |
| | word2vec | fasttext | GloVe |
| 추천된 전시회 제목 | 김수빈展 | 김수빈展 | 김수빈展 |
| 추천된 전시회 설명 | 대학과 대학원에서 동양화를 전공했다. 내가 공부하고 이해한 동양 미술의... | 대학과 대학원에서 동양화를 전공했다. 내가 공부하고 이해한 동양 미술의... | 대학과 대학원에서 동양화를 전공했다. 내가 공부하고 이해한 동양 미술의... |
| 추천된 전시회 키워드 | 벚꽃, 아름, 가득, 가장... | 벚꽃, 아름, 가득, 가장... | 벚꽃, 아름, 가득, 가장... |
| 사용자 해시태그: 휴식, 일상, 자유 | | | |
| | word2vec | fasttext | GloVe |
| 추천된 전시회 제목 | 팀보타: 머스타드블루 '탐의 숲' | Oikos_Kore's Bouquet | Oikos_Kore's Bouquet |
| 추천된 전시회 설명 | 팀보타는 자연물과 현대미술기법 등을 접목하여 자연의 | 모든 생명체의 삶은 화사한 꽃부케보다 거친 풀부케와 더 많이 | 모든 생명체의 삶은 화사한 꽃부케보다 거친 풀부케와 더 많이 |

| | | | |
|-------------|----------------------|----------------------|----------------------|
| | 아름다움을... | 깊아 있다... | 깊아 있다... |
| 추천된 전시회 키워드 | 현대, 자연, 해석, 팀보... | 시간, 모든, 겨울, 대지... | 시간, 모든, 겨울, 대지... |

<표6> 임베딩 모델에 따른 전시회 추천 비교

3-1. 팀 일정

팀 전체 일정

| 내용 | 1주 | 2주 | 3주 | 4주 | 5주 | 6주 | 7주 | 8주 | 9주 | 10주 | 11주 | 12주 | 13주 | 14주 | 15주 |
|-----------------------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|
| 1차 교수님 피드백 및 제출 | | | | | | | | | | | | | | | |
| 1차 CM 피드백 및 제출 | | | | | | | | | | | | | | | |
| 회원가입, 로그인, 유저정보 | | | | | | | | | | | | | | | |
| 전시 커뮤니티 | | | | | | | | | | | | | | | |
| 2차 교수님 피드백 및 제출 | | | | | | | | | | | | | | | |
| 2차 CM 피드백 및 제출 | | | | | | | | | | | | | | | |
| 온라인 전시 | | | | | | | | | | | | | | | |
| 3차 교수님 피드백 및 제출 | | | | | | | | | | | | | | | |
| 3차 CM 피드백 및 제출 | | | | | | | | | | | | | | | |
| 4차 교수님 피드백 및 제출 | | | | | | | | | | | | | | | |
| 4차 CM 피드백 및 제출 | | | | | | | | | | | | | | | |
| 최종 보고서 제출 | | | | | | | | | | | | | | | |

팀원 - 이지호 일정

| 내용 | 1주 | 2주 | 3주 | 4주 | 5주 | 6주 | 7주 | 8주 | 9주 | 10주 | 11주 | 12주 | 13주 | 14주 | 15주 |
|------------------------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|
| 프로젝트 제안서 작성 | | | | | | | | | | | | | | | |
| 개발환경 구축 | | | | | | | | | | | | | | | |
| DB 모델링 | | | | | | | | | | | | | | | |
| 유저 관련 기능 구현 (REST API) | | | | | | | | | | | | | | | |
| 전시 커뮤니티 페이지 | | | | | | | | | | | | | | | |
| 온라인 전시 페이지 | | | | | | | | | | | | | | | |
| 전시 추천 기능 구현 (REST API) | | | | | | | | | | | | | | | |
| 테스트 및 보완 | | | | | | | | | | | | | | | |

팀원 - 이승연 일정

| 내용 | 1주 | 2주 | 3주 | 4주 | 5주 | 6주 | 7주 | 8주 | 9주 | 10주 | 11주 | 12주 | 13주 | 14주 | 15주 |
|---------------------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|
| 프로젝트 제안서 작성 | | | | | | | | | | | | | | | |
| 개발환경 구축 | | | | | | | | | | | | | | | |
| 회원가입, 로그인, 유저정보 페이지 | | | | | | | | | | | | | | | |
| 전시 커뮤니티 관련 기능 구현 | | | | | | | | | | | | | | | |
| 온라인 전시 페이지 | | | | | | | | | | | | | | | |
| 온라인 전시 기능 구현 | | | | | | | | | | | | | | | |
| 전시 추천 페이지 | | | | | | | | | | | | | | | |
| 테스트 및 보완 | | | | | | | | | | | | | | | |

팀원 - 유승욱 일정

| 내용 | 1주 | 2주 | 3주 | 4주 | 5주 | 6주 | 7주 | 8주 | 9주 | 10주 | 11주 | 12주 | 13주 | 14주 | 15주 |
|-------------------------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|
| 프로젝트 제안서 작성 | | | | | | | | | | | | | | | |
| 개발환경 구축 | | | | | | | | | | | | | | | |
| 전시회 정보 이미지 크롤링 | | | | | | | | | | | | | | | |
| 이미지 내의 텍스트 추출 | | | | | | | | | | | | | | | |
| 텍스트 전처리 | | | | | | | | | | | | | | | |
| 키워드 추출 알고리즘 개발 | | | | | | | | | | | | | | | |
| 사용자/전시회 키워드 임베딩 비교 (추천) | | | | | | | | | | | | | | | |
| 기능 연동 | | | | | | | | | | | | | | | |
| API 구축 | | | | | | | | | | | | | | | |
| 테스트 및 보완 | | | | | | | | | | | | | | | |

3-2. 교수님 피드백 일정

| 일시 | 내용 | 비고 |
|---------------|-------------|----------|
| 2022. 03. 11. | 이형태 교수님 피드백 | 1차 보고 내용 |
| 2022. 03. 17. | 이형태 교수님 피드백 | |
| 2022. 03. 25. | 이형태 교수님 피드백 | 2차 보고 내용 |
| 2022. 04. 29. | 이형태 교수님 피드백 | 3차 보고 내용 |
| 2022. 05. 13. | 이형태 교수님 피드백 | 4차 보고 내용 |

3-3. 멘토님 피드백 일정

| 일시 | 내용 | 비고 |
|---------------|-------------|------------------|
| 2022. 03. 18. | 조용덕 멘토님 피드백 | 1차 보고 내용 |
| 2022. 04. 07. | 조용덕 멘토님 피드백 | 2차 보고 내용 |
| 2022. 05. 12. | 조용덕 멘토님 피드백 | 3차 보고 내용 |
| 2022. 05. 27. | 조용덕 멘토님 피드백 | 4차 보고 내용, 대면 멘토링 |