

날짜	25. 09. 03. (수) 25. 09. 04 (목)	회의 형태	대면
참여자	김가희, 김현진, 김예빈, 문수빈, 전수아	시간	16:30 - 18:00 17:30 - 20:00

## 1. 개발 현황

- [태깅 데이터셋] TITLE 업그레이드, DATE 정규화, ARTIST 업그레이드
- 카카오 API 로그인
- 인디 공연 공지 게시물 여부 판별

## 2. 문제점

### 1) 알고리즘

- NER 태깅을 통해 데이터셋을 학습시킨 알고리즘(KoELECTRA + CRF)의 정확도가 높지 않음.
- 특히 '공연 제목'과 '아티스트' 추출에 혼란스러운 모습을 보임.

```

===== 최종 추출 결과 =====
TITLE: 말해 뭐해! / , . / _ / , ! / . / , / | / • 2024년 12월 6일(금) 오후 8시 / • 프리즘홀(
서울 마포구 양화로12길 6 지하) / • 11월 14일(목) 오후 6시 / • 멜론티켓 단독 오픈 / @ruby_records / #모트 #Motte /
#OCEAN #단독공연 #prismhall #프리즘홀 #홍대공연 #인디음악
ARTIST: 모트 연말 / 연말을 맞아
VENUE: 프리즘홀
DATE: 2025.12.06 토요일 오후 8시
PRICE_ON:
PRICE_OFF:
TICKET_OPEN:
ACCOUNT: @ruby_records

```

- 데이터셋을 더 늘려야 하는지, 혹은 알고리즘의 코드를 변경해야 하는지 혼란스러운 상황
- 모델의 정확도가 중요한 서비스임을 고려하면 차선택 필요

### 2) 인스타그램 크롤링

- 새 게시물 텍스트 가져오는 크롤링의 세분화 필요.
- 총 N개의 계정을 탐색하여 N개의 게시물을 가져오는 데 40분 소요
- 인스타그램 봇에 탐지되지 않기 위해 계정 탐색 사이에 불가피한 delay 삽입으로 발생

## 3. 해결 방안

### 1) 알고리즘

- n8n(오픈소스 워크플로우 자동화 툴) + chatGPT api를 통한 자동화

### 2) 인스타그램 크롤링

- 1차 크롤링 → 새로운 게시물인지 판별 → 2차 크롤링 → 3차 크롤링으로 단계 세분화
- 2차 크롤링에서 시간을 대폭 줄이는 방법

**Step 1.** 1차 크롤링 : 게시물 최대 4개에 대한 쇼트코드, 게시시각만 가져오기 (게시물이 고정 게시물이면 4개까지 가지 않고 그 다음 게시물에서 중지)

**Step 2.** 새로운 게시물인지 판별: 쇼트코드, 게시 시각 비교하여 새로 올라온 게시물에 대한 쇼트코드만 저장

**Step 3. 2차 크롤링:** 남은 쇼트코드에 대해서만 원본 게시물 링크에 접속 후, 텍스트만 가져와서 인디 공연 공지 판별 모델에 전달

**Step 4. 3차 크롤링 :** 인디 공연 공지 판별 모델에서 인디 공연 공지라고 판별된 게시물에 한해서만 이미지 가져와서 게시물 추출 모델에 전달

#### 4. 정량 지표 후보

- 1주 합산 웹페이지 방문수 1000회 이상
- 모델 (추출 알고리즘) 정확도 90%
- 사용자 웹 만족도 4점 이상 (5점 만점)
- 1주 / 일일 웹페이지 세션 접속 수 N명 이상
- 홍대 등 공연장 밀집 장소에서 가까운 공연 찾기 버튼을 클릭 수
- 모델(인디공연판별) 정확도 90%
- 사용자의 웹 만족도 N % 이상
- 재방문률 N % 이상

#### 5. 피드백

- 현재 김삼문 팀이 집중해야 하는 것은 ‘사람의 마음’을 이끄는 것
- 머신러닝, 딥러닝 등 기술적인 부분은 향후 개발자의 담당
- 서비스 초기에는 직접 공연 정보를 모으고, 사용자의 유입이 많아지면 자동화로 변경

#### 6. 역할 분담

\*추가된 기능 : 스탬프, 김삼문 큐레이팅, mood별 공연 추천

\*삭제(유보)된 기능 : 인디 공연 공지 판별 모델, 게시물 추출 모델

- ☐ 가희 : 피그마 웹 UI 수정, 웹 프론트 퍼블리싱 완성
- ☐ 현진 : 인스타그램에서 게시물 가져오는 시간 단축하도록 수정
- ☐ 예빈 : 피그마 웹 UI 수정, 웹 프론트 퍼블리싱 완성
- ☐ 수빈 : 추가된 기능 고려하여 DB 및 ERD 수정
- ☐ 수아 : 카카오톡인, 인스타그램에서 게시물 가져오는 시간 단축하도록 수정