

# 온섬표\_감정기반 음악추천봇

20172015 강민서, 20172042 박세원, 20172133 박소정

## I. 이 소프트웨어는 왜 필요한가?

### i. 계기

기존의 음악 사이트에서는 1) 타사용자가 미리 생성한 플레이리스트를 활용해 추천하거나 2) 장르, 아티스트 등 곡의 정보와 그동안 축적한 사용자가 선호하는 곡을 기반으로 유사한 곡을 추천하는 방식을 사용한다. 위와 같이 추천받은 곡은 그날 사용자의 무드와 일치하지 않거나 곡의 분위기만 유사해, 가사 등 곡의 내용이 사용자의 감정과 어긋나는 경우가 많다. 우리는 이러한 기존 방식에서 벗어나 좀 더 사용자의 감정과 유사한 가사의 곡을 추천해 사용자의 만족도를 높이하고자 한다. 챗봇을 통해 사용자에게 수시로 일기를 전달받고 감성사전을 활용해 단어와 문맥에 따라 분석해, 가장 유사한 감정을 가진 곡을 추천한다. 이 프로그램을 구상하게 되었다.

### ii. 도출된 정의

사용자는 챗봇과 대화함으로써 그날의 간단한 일기와 감정을 전달한다. 프로그램 내에서 전달받은 데이터를 분류해 둔 12가지 감정 카테고리과 감성사전을 통해 분석해 “사용자 감정 좌표를 출력한다. 같은 방식으로 이미 도출해둔 “곡의 좌표” 와 거리를 비교하여 작성한 일기 감정에 가장 적합한 음악을 추천한다.

### iii. 기대효과

#### - 음악 감상의 만족도 향상

매시간 바뀌는 감정이나 기분에 따라 듣고 싶은 음악의 유형을 사용자가 직접 작성한 일기를 통해 추천하기 때문에 곡에 더 많이 공감할 수 있다.

#### - 감정에 끼치는 긍정적인 영향

간단한 일기 작성을 통해 매일 자신의 생각을 요약하고 정리할 수 있으며, 과거 일기를 열람하는 일기 조회 기능을 통해 사용자는 당일 혹은 과거의 겪었던 일과 본인의 감정을 성찰할 수 있는 시간을 가진다. 일기를 쓰고 그에 맞는 음악을 들음으로써 자신의 감정을 명확하게 파악할 수 있다

#### - 쉬운 이용성과 편리한 접근성

매일매일 일기를 작성하는 것이 막막하고 낯선 사용자가 가장 친숙한 챗봇을 통해 간단하게 생각날 때마다 일기를 작성할 수 있기 때문에 부담스럽지 않게 이용 가능하다. 또한 접근성이 좋은 카카오톡 챗봇을 활용했기 때문에 출근길, 퇴근길, 산책길 등 일상의 틈새 시간에 서비스를 이용할 수 있다.

#### - 데이터 학습을 통한 정확성 향상과 발전 가능성

챗봇을 이용하는 사용자들을 통해 더 많은 데이터들을 수집하고 분석, 학습할 수 있기 때문에 지속적으로 발전할 수 있다.

#### - 상용화 가능

더 많은 사용자를 유치하기 위해 챗봇에 가장 먼저 적용했지만 기존의 음악 사이트들에 적용해 음악 사이트를 이용하는 사용자들이 더 다양한 방식으로 새로운 곡들을 접할 수 있다.

## II. 무슨 소프트웨어를 개발하였는가?

### i. 플랫폼 및 기술 스택 선정 이유

- 사용자의 접근성 및 편의성
- 해당 기술이 많이 사용되는 기술인지
- 평소에 관심이 있거나, 익숙한 기술인지

#### 1) 플랫폼 - 카카오톡 챗봇 (카카오 i 오픈 빌더)

사용자의 접근성 및 편의성을 고려하여 사용자가 가장 많은 플랫폼인 카카오톡의 ‘카카오 i 오픈 빌더’ 플랫폼을 사용하여 챗봇을 개발하였다.

#### 2) Backend - AWS EC2, AWS RDS, MYSQL

AWS는 용량 추정과 같은 유지관리에 비용 투자가 필요 없으며, 배포가 용이하다. 서버비와 안정성과 보안 측면에서 AWS EC2를 선정하였다. AWS의 RDS는 MYSQL 엔진을 제공하며, 확장성 측면에서 편리하기 때문에 선택하였다.

#### 3) 사용 언어 및 프레임워크 - Python, Flask

프로젝트의 기능을 챗봇에 적용하기 Python 언어를 사용하고, 챗봇에 출력하기 위해 웹 프레임워크인 Flask를 사용하였다.

#### 4) 이외의 플랫폼 및 기술 - 지니 뮤직, Konlpy(Komoran)

- 지니 뮤직은 국내 2위의 음악 스트리밍 플랫폼으로 국내 음악의 음원 수가 많아 선정.
- Konlpy의 Komoran 형태소 분석기는 한국어 형태소 분석기 중 연산 속도가 상위권에 속하며, 자소 분리나 오타자에 대한 품질이 보장된다. 모바일 챗봇의 맞춤법과 같은 오타자가 나올 가능성이 높았기 때문에 오타자 측면에서 형태소 분석기 중 가장 정확한 Komoran을 사용했다.

### ii. 기능 설명

#### - 감정 일기로 노래 추천

메인 기능은 구축한 감정 사전과 노래 추천 알고리즘을 이용하여 사용자가 일기를 작성하면 그날의 감정과 어울리는 가사의 노래를 추천한다. 해당 기능의 상세한 알고리즘은 추후 서술한다.

#### - 간단한 음악 추천

감정 일기를 작성하지 않고, 사용자의 감정만 간단히 입력해도 노래를 추천하는 기능이다.

#### - 일기 조회 및 삭제

일기 조회 기능에서 사용자가 작성한 일기 목록 데이터를 통해 원하는 일자의 일기를 조회할 수 있다. 일기 삭제 기능은 사용자가 현재까지 작성한 모든 일기의 데이터를 삭제하는 기능이다.

#### - 온오프트 차트

지난 한 달간 음악 추천 기능의 추천 곡 중 사용자가 가장 만족한 곡 3곡을 조회하는 기능이다.

### iii. 인용 논문 설명

- 문창배 외 3.(한국산업정보학회). 『음악추천을 위한 분위기 태그 분석』

감정 사전의 아이디어와 감정 사전 구축에 Thayer 모델을 참고하였다. 실제 프로젝트에서는 x축과 y축을 각각 감정의 긍/부정과 감정의 크기로 두어 감정이 분류를 진행하였다. 해당 논문의 내용 중 음악의 12 분위기 카운트 값 계산 및 AV값 계산을 통해 사용자가 입력한 일기 내용과 노래 가사를 좌표로 변환하여 거리 기반 유사도를 기준으로 노래를 추천하였다.

- 신기원 외3.(한국지능정보시스템학회). 『감정 온톨로지를 활용한 가사 기반의 음악 감정 추출』

분석 결과의 정확성을 위한 부정어를 처리하는 방법을 참고하였다. 부사와 용언을 사용하여 부정문이 만들어지는 부정 구문 패턴을 알고리즘에 적용하여 분석 결과의 정확성을 높였다.

### III. 어떤 방법으로 개발하였는가?

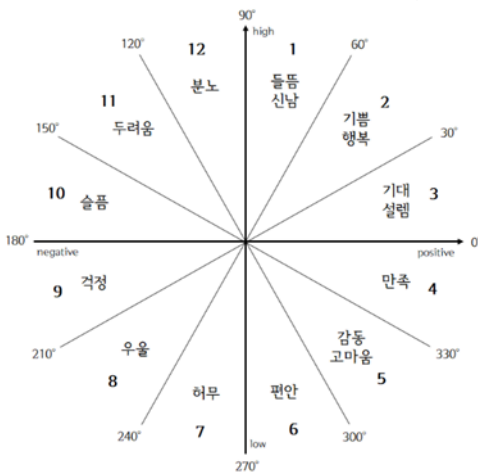
#### i. 감성 분석

감성 분석은 단어와 문맥을 분석하여 텍스트의 감정을 파악하는 기술이다. 이는 크게 2 가지 방식으로 나뉘는데, 사전을 구축하고 규칙을 개발하는 어휘 기반 분석 방식이 있고 기계학습을 이용한 머신러닝 분석 방식이 있다. 이 프로젝트는 사용자의 감정을 파악하여 그 감정에 어울리는 음악을 추천해주므로, 감성 분석 기술의 사용이 필수적이다. 그러나 프로젝트와 적합한 학습 데이터를 찾을 수 없었기 때문에, 어휘 기반의 감성 분석 방식을 적용하여 직접 감성 사전을 구축 후 활용하였다.

#### 1. 감성 사전 구축

##### 1) Thayer 가 제안한 모델을 일부 변형하여 재정의

감성 사전의 구축을 위해서 우선 Thayer 가 제안한 2 차원 감성 모델을 참고하였다. 한국어로 Thayer 모델을 그대로 사용하기에는 단어 편중이 심하게 나타나서, 키워드를 일부 변형하여 재정의 했다.



<그림 1. Thayer 모델 기반으로 변형한 감성 모델>

##### 2) 각종 데이터를 참고하여 단어 수집

사전에 답을 단어를 수집하기 위해서 크게 3 가지 방법을 활용하였다. 첫 번째는 『한국어 감정 단어의 목록 작성과 차원 탐색』 논문에서 분류한 감정어 데이터를 활용하였다. 두 번째로는 프로젝트를 위해 진행한 설문 조사 결과 데이터를 활용하였다. 약 100 명의 참여자가 작성한 텍스트 데이터를 참고하여 감정어를 분류하였다. 세 번째는 SNS 에 작성한 글을 일부 수집하여 감정어를 분류하였다. 이를 통해 일상어까지 반영한 사전을 구축하였다.

##### 3) 수집한 단어를 모델에 분류

앞서 수집한 단어 데이터를 미리 제작한 모델에 맞게 분류하였다. 12 가지 각각의 감정 별로 소분류 키워드를 지정하여 단어 분류를 더욱 용이하도록 했다.

5	감동, 고마움	공감	고맙다	반갑다	백자다	너그럽다	여름
			온정	관기	고마워하다	반가워하다	등등
			복돋다	복돋우다	감영	감사	경덕
			감사하다	감개	감격하다	감명	감책
			동감하다	감개무량하다	감동하다	감복하다	동행
6	편안하다	안심	편하다	친하다	평안하다	밝다	환영
			편안	편차능	편준	따스다	힐링
			편안하다	평온하다	평화롭다	물가분하다	후원
			따뜻하다	부드럽다	포근하다	여유롭다	느그
			덤덤하다	상관없다	아무렇지 않다	무덤덤하다	
		개이지않음	개이지않다	친근하다	상냥하다	배려	
7	허무	황당	황당하다	어이없다	가관이다	가소롭다	기상
			터무니없다	별나다	유별나다	특이하다	이상
		실망	실망하다	대실망	낙담하다	헛되다	독
		질림	질리다	실증나다	따분하다	지루하다	지극
			재미없다	지긋지긋하다	진저리나다	식상하다	신들

<그림 2. 감성 사전 소분류>

##### 4) 단어의 원형과 품사를 기록

분류가 끝난 단어를 형태소 분석기로 분석하여 사전에 원형과 품사를 기록하였다. 기록된 데이터가 적절하지 않은 경우, Komoran 기능 중 하나인 사용자 사전을 이용하여 원형과 품사를 직접 태깅하였다.

## 2. 텍스트 감정 분석

### 1) 데이터를 정제하여 형태소 분석

정규표현식을 활용하여 텍스트 안의 영문을 제거하고 형태소 분석기에 오류가 발생하지 않도록 데이터를 정제했다. 이후 Komoran 형태소 분석기를 이용하여 텍스트 데이터의 형태소를 분석하였다.

### 2) 부정어 처리

감성 분석의 정확도를 높이기 위해 간단한 부정어 처리 과정을 적용했다. 문장 안에 부정어가 포함되어 있을 경우, 문장 전체를 다른 배열에 저장한다. 이후 따로 저장된 문장을 분석하여 연결 어미를 기준으로 끊은 뒤, 부정어와 감정어가 함께 있을 때 기존에 갖고 있던 감정과 반대된다고 판단했다.

### 3) 텍스트 데이터의 감정어 탐색

BoW 방식을 활용하여 텍스트 데이터가 가진 형태소를 분류한 뒤, 미리 제작한 감성 사전을 이용해 가사에 포함된 감정어를 탐색한다. 텍스트에서 사전에 있는 단어를 발견하면 미리 분류한 12 가지 감정 카테고리에 맞게 카운트 값을 증가한다. 예를 들어 사전에 4 번 카테고리로 분류된 '좋다' 라는 단어를 텍스트에서 발견하면, 미리 만들어둔 배열의 4 번째에 1 을 더해준다. ( $v = [0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]$ ) 이러한 방식으로 해당 배열을 채워 나간다.

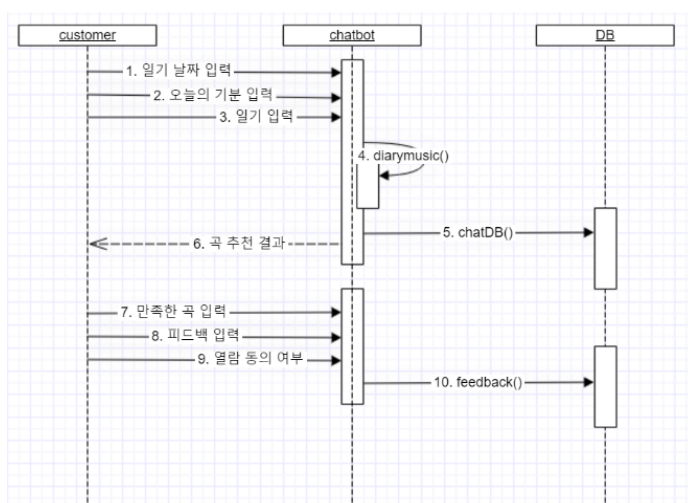
### 4) 데이터 정규화 및 좌표 값 계산

텍스트마다 가지고 있는 감정어의 수가 다르므로, 앞서 채운 배열을 모두 합쳐 1 이 될 수 있도록 소수점으로 정규화한다. 이후 배열 각각의 값에 적절한 각의 코사인 값을 곱해서 12 로 나누어 x 좌표를 구하고, 마찬가지로 사인 값을 곱해 12 로 나누어 y 좌표를 구한다. 이렇게 계산한 좌표를 적절한 카테고리로 분류하여 저장한다.

## ii. 음악 추천

앞서 소개한 감성 분석 방식을 통해 가사 데이터 분석 결과를 미리 저장해둔 후, 사용자 텍스트 데이터가 들어올 때마다 감정을 분석하여 음악을 추천하도록 만들었다. 사용자는 감정을 간단히 적는 오늘의 기분 데이터 (이하 메인 감정이라고 지칭) 와 일기처럼 자세히 감정을 적는 일기 감정 데이터 두 가지를 입력한다. 이렇게 입력받은 사용자 데이터 분석 결과에 따라 음악 추천 방식을 달리했는데, 총 3 가지 패턴으로 분류하여 음악을 추천했다.

1. 메인 감정과 일기 감정이 서로  $\pm 1$  범위 내에 있거나 일기 감정만 분석 가능할 경우 - 일기 감정 데이터의 좌표 값을 기준으로 가장 근접한 가사 데이터 좌표 3 가지를 판별하여 사용자에게 추천한다.
2. 메인 감정만 분석 가능할 경우 - 메인 감정이 가진 카테고리와 같이 분류된 가사 데이터 중 3 가지를 랜덤 추출하여 사용자에게 추천한다.
3. 두 감정 모두 분석이 불가능할 경우 - 이전에 따로 분류해둔 감정값이 5 이하인 가사 데이터 중 3 가지를 랜덤 추출하여 사용자에게 추천한다.



<그림 3. 챗봇 메인 기능 시퀀스 다이어그램>