

추천 알고리즘의 설계흐름을 3 가지 방식으로 생각했다. 이 방식들을 분석 후 가장 적절한 방법을 택하기로 한다.

1. 로그인없이, 감정에 따라 감정 테이블에서 노래를 랜덤으로 선택한다.

2. 사용자 정보를 저장한 뒤, 사용자의 선택특성 분석 혹은 사용자의 노래 추천 서비스 만족도 조사 등의 특성을 기계학습을 통해 학습시킨다. 하지만 개발 초기 단계에서 데이터 확보가 불가능하므로 힘들다고 판단되어 시스템을 확장할 때 이용하기로 한다.

이때 기능이 더 추가될 수 있는데, 사용자 디비 테이블, 사용자 정보를 서버에 활용, 로그인 화면을 만들어서 서버에 데이터를 넘겨주기 등이 해당될 수 있다.

3. 노래 가사에서 나오는 키워드의 특성을 기반으로한 노래 추천 알고리즘. 감정의 최종결과가 노래에서 명시된다고 보장을 못하는 애매함이 있다. 예를 들어, 슬픈 노래에 ‘슬픔’이라는 단어가 등장하지 않을 수 있기 때문이다. 따라서 다음과 같은 방법을 생각해보았다.

TFIDF 알고리즘을 이용해 키워드를 뽑는다. TFIDF는 텍스트 처리와 관련된 것으로 텀에 부여된 가중치 값이다. 한 문서에서 텀이 몇 번 나왔는지(TF:Term Frequency)에 비례하고, 한 단어가 몇 개의 고유한 문서에서 나왔는지(IDF:Inverse Document Frequency)에 반비례한다는 의미이다.

최종 결과인 감정을 워드 클라우드로 나타내어 연관단어들을 뽑는다.

전자와 후자의 연관성 또는 유사도를 이용해 추천 알고리즘을 설계한다.