

닐로 날로잡기 프로젝트
Database Project



인천대학교
INCHEON
NATIONAL UNIVERSITY

Prof. 박재휘 교수님
Project Manager 201401447박재성
201401438남궁찬
201401454방기현
201401500최영훈

목 차(Table of Contents)

1. 개요
 - 1.1 프로젝트 개요 및 필요성
 - 1.2 프로젝트 목적
 - 1.3 기대효과
2. 프로젝트 일정
 - 2.1 프로젝트 일정표
3. 팀 구성 및 역할
 - 3.1 조직 구성
 - 3.2 역할
4. 사용 시나리오
 - 4.1 사용할 데이터
 - 4.2. 사용할 데이터 확보 방법
5. 개발 시나리오
 - 5.1 개발 시나리오
 - 5.2 흐름도
 - 5.3 데이터베이스 모델(ERD)
6. 개발 환경
 - 6.1. 데이터베이스
 - 6.2 데이터 분석
 - 6.3 머신러닝
 - 6.4 데이터 수집
 - 6.5 협업, 버전관리
7. 개발 과정에서의 유의점
 - 7.1. 알고리즘 구현 난이도
 - 7.2. 데이터 확보의 어려움
8. 참고문헌

1. 개요

1.1. 프로젝트 개요 및 필요성

음원 사이트들은 음원 발매 이후 한 시간 단위로 집계되는 ‘실시간 차트’를 발표하고 있다. 최근 음원 문화의 변화에 따라 음원차트의 상위권에 노출되지 못하면 사장되어버린다는 위기의식이 콘텐츠 제작자와 극성팬들 사이에 팽배하면서 일부 이해관계자들의 불안 심리를 이용해 음원차트의 공정성에 악의적인 위해를 가하고자 행해지는 음원 사재기가 문제가 되고 있다. 실제로 2015년 9월 말 국내 최대 음원 사이트인 멜론이 각 음반 기획사를 대상으로 ‘[로엔] 음원사재기 (어뷰징) 행위 근절을 위한 협조 요청의 건’이라는 제목의 공문을 보낸 사실이 드러났으며, 공문에는 음원 사재기를 이용한 순위 차트 조작에 유감을 표하며 외부 업체와의 부당한 사례가 있을 경우 법률적 조치를 취하겠다는 내용이 담겨 있었다. 이에 대응하여 문화체육관광부 주관으로 2016년 3월 3일 ‘음악 산업 진흥에 관한 법률’을 통과시켰다.

위와 같은 이유로 음원 차트 데이터를 분석하여 사재기로 의심되는 음원을 알아내는 프로젝트를 수행하게 되었다.

1.2. 프로젝트의 목적

프로젝트의 목적은 지금까지의 음원차트와 트렌드 분석을 통해 조작으로 의심되는 음원들을 찾아내고 해당 음원의 패턴을 알아내는 것이다. 나아가 알아낸 패턴을 통해 새로운 음원에 대한 조작 의심 정도를 알려주는 머신러닝 모델을 만드는 것이다.

1.3 기대효과

머신러닝 모델을 통해 나온 조작 의심 정도를 음원 차트 순위 알고리즘에 패널티로 적용한다면 좀 더 공정하고 신뢰할 만한 음원 차트를 제공할 수 있을 것이다. 이는 음악 산업의 교란을 근절하고 음악 산업 진흥을 이끌어 낼 수 있을 것이다.

2. 프로젝트 일정

2.1. 프로젝트 일정표

PROJECT TIMELINE

	9월					10월					11월					12월				
	3	10	17	24	31	1	8	15	22	29	5	12	19	26	30	3	10	17	24	
PROJECT WEEK	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
기획서 초안																				
	조직 구성																			
	주제 선정					피드백														
	초안 작성																			
기획서 최종안								피드백 회의												
									최종안 작성											
								발표												
프로젝트 구현							RDS시작													
								DB설계					DB보안 수정							
							구현 및 회의													
								알고리즘 설계				알고리즘 보안 수정								
												인터페이스 구현								
보고서 초안											초안 작성									
											피드백 회의									
보고서 최종안															최종안 작성					
															발표 및 시연					
프로젝트 완료																				

3. 팀 구성 및 역할

3.1 조직 구성

한 학기 동안의 프로젝트 인 것과 프로젝트의 규모를 고려하여, 빠른 의사소통과 결정을 위해서 분산형 팀조직으로 조직을 구성하였다.

- 1)민주주의식 의사결정
- 2)같이 협동, 수행하는 비이기적인 팀
- 3)자신의 역할에 맞는 일을 알아서 수행

3.2 역할

구성원		할당된 작업 및 역할
조원명	*박 재 성	Project Manager, 전반적인 프로젝트 관리, 각종 지원, 자료수집
	남 궁 찬	Question & Answer, 자료수집, 기술 파악 및 수집, 동향분석
	방 기 현	Development Manager, DBA, 코드 기획 및 개발 ,자료 수집
	최 영 훈	Document Manager, 기획-보고서 초안 및 최종안 제작, 발표자료 제작

4. 사용 데이터

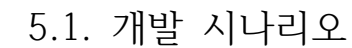
4.1. 사용할 데이터

지니, 엠넷, 벅스 3곳의

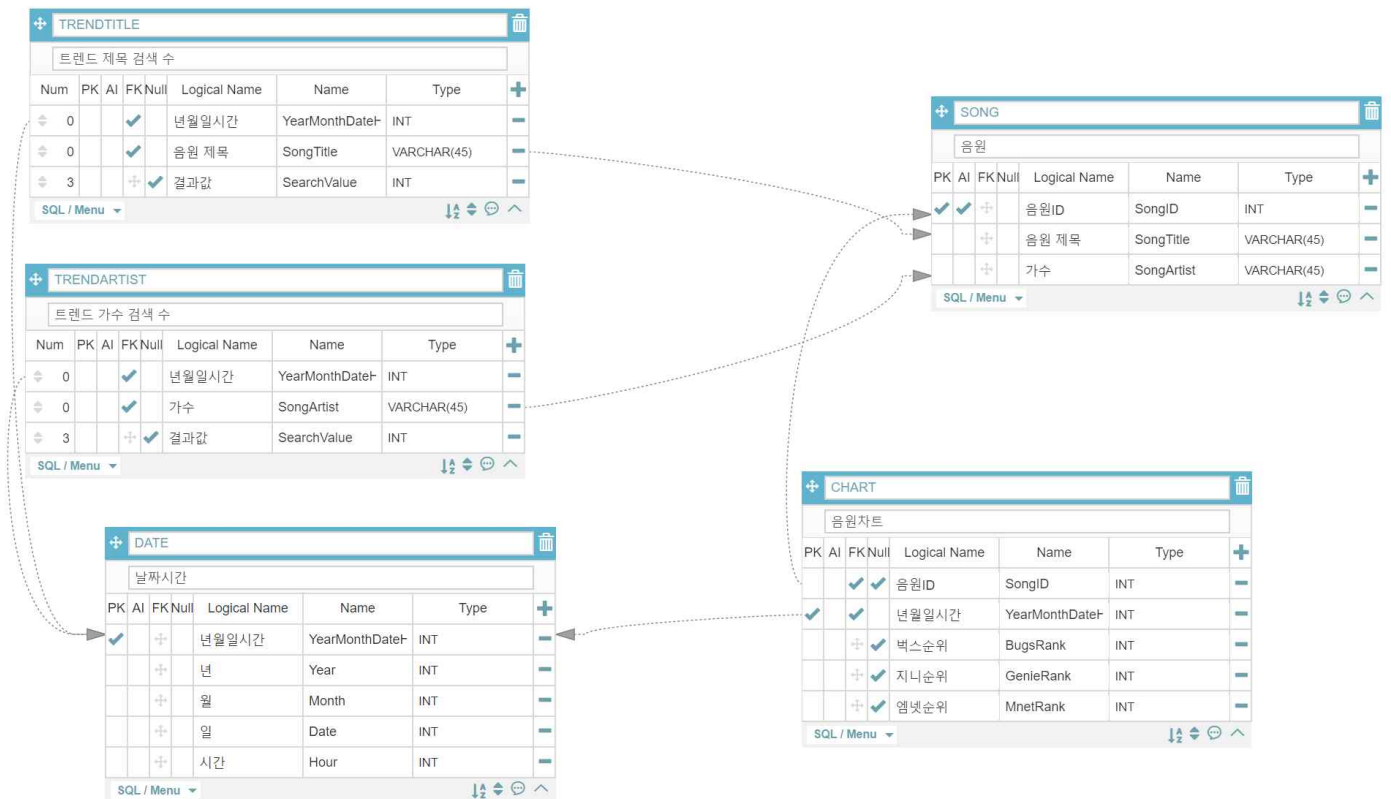
- 시간대별 음원순위 : 각 시간대별 음원순위 데이터를 가져와 변동폭 등을 분석한다.
 - 해당 음원의 화제성 : 음원순위와 해당 음원의 화제성 관계를 분석한다.
 - 아티스트의 유명세 정도 : 음원순위와 해당 음원 아티스트의 유명세 정도 관계를 분석한다.
- 총 5가지의 데이터를 사용하려 합니다.

4.2. 사용할 데이터 확보 방법

- 시간대별 음원순위 : 지니, 엠넷, 벅스 세 곳의 시간대별 음원차트를 크롤링하여 확보한다.
- 해당 음원의 화제성 : 구글 트렌드 오픈 API를 이용한다.
- 아티스트의 유명세 정도 : 1)구글 트렌드 오픈 API를 이용한다.



5.3. 데이터베이스 모델(ERD)



6. 개발 환경

6.1. 데이터베이스

AWS의 RDS중 Mysql을 사용한다.

6.2 데이터 분석

IBM SPSS Statistics 와 python을 이용하여 데이터 분석

6.3 머신러닝

Scikit-Learn 기술을 이용하여 머신러닝 모델 생성

6.4 데이터 수집

자바와 파이썬을 이용한 데이터 크롤링(데이터 수집 자동화)

자바와 파이썬을 이용한 구글 트렌드 오픈 API 사용

6.5 협업, 버전관리

git, github 사용

7. 개발 과정에서의 유의점

7.1. 알고리즘 구현 난이도

데이터간의 상관관계와 데이터의 특정한 패턴을 파악하는 것이 핵심이며, 제대로 구해졌을 경우, 알고리즘 구현에 있어 큰 어려움이 없을 것으로 예상됨.

머신러닝 모델 구현의 경우 조작으로 의심되었던 음원의 수가 많지 않아 정교한 모델링이 필요할 것으로 보여짐.

7.2. 데이터 확보의 어려움

음원사이트들이 공개적인 API를 제공하고 있지 않다. 따라서 크롤링을 이용하여 원하는 데이터를 가지고 오는 작업이 필요하다. 크롤링 알고리즘을 각 음원순위 웹페이지에 맞게 구현하여 가져오는 작업에 어려움은 없을 것이라 보여진다.. 이 외의 데이터는 구글에서 제공하는 API혹은 오픈소스 API를 이용하여 데이터를 확보할 수 있다. 가져온 데이터들을 알맞게 수치화하는 등 정리하는 과정에서 어려움이 예상된다.

8. 참고문헌

곽아영,(2017) 데이터 분석을 통한 온라인 음원차트의 어뷰징 영향요인 탐색, *학위논문(석사) 서
울시립대학교 대학원 : 경영학과 2017. 2*