


캡스톤 디자인 I

종합설계 프로젝트

프로젝트 명	매칭 추천알고리즘 DB설계
팀 명	제로파운더스 팀
문서 제목	중간보고서

Version	1.0
Date	2018-04-12

팀원	이현지)
지도교수	윤 성혜 교수님

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12


CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 " **매칭 추천알고리즘 db설계**"를 수행하는 팀 "**제로파운더스**"의 팀원들의 자산이다. 국민대학교 컴퓨터공학부 및 팀 "**제로파운더스**"의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

문서 정보 / 수정 내역


Filename	중간보고서-매칭추천알고리즘db설계.doc
원안작성자	이현지
수정작업자	이현지

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2018-04-13	이현지	1.0	최초 작성	전체 내용작성

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12

목 차

1	프로젝트 목표.....	4
2	수행 내용 및 중간결과.....	5
2.1	계획서 상의 연구내용	5
2.2	수행 내용.....	8
2.2.1	하드웨어 데이터 설계	8
2.2.2	분석용 데이터 설계.....	9
3	수정된 연구내용 및 추진 방향	11
3.1	수정사항	11
3.1.1	데이터베이스 설계 수정.....	11
4	향후 추진계획.....	14
4.1	향후 계획의 세부 내용.....	14
4.1.1	데이터 설계에 따른 데이터 생성.....	14
5	고충 및 건의사항.....	15

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12

1 프로젝트 목표

더 아름다워지고 싶다는 욕망을 가진 사람들은 뷰티에 많은 관심이 있지만 흘러 넘치는 뷰티제품에서 나에게 맞는 뷰티 제품을 찾기 어렵다. 그렇다고 무조건 비싼 화장품, 남들이 좋다는 화장품을 따라 구매한다고 자신에게 성공적인 제품을 찾기가 어렵다. 자신의 피부에 맞지 않는 화장품을 사용하게 되어 오히려 바르지 않는 것보다 못하는 경우가 발생하기도 한다. 하지만 그러한 제품을 빗겨가면서 나에게 꼭 맞는 제품을 찾기란 너무나도 어려운 실정에 있다.

인턴기간 동안 많은 뷰티 관련 리서치를 통해서 실질적으로 뷰티 케어란 제품과 관리가 별개가 아닌 소비자 개개인에 맞는 화장품을 통해서 제품에 맞는 케어가 핵심이라는 결론을 내었고 소비자와 제품을 매칭하는 플랫폼을 목표로 하였다.

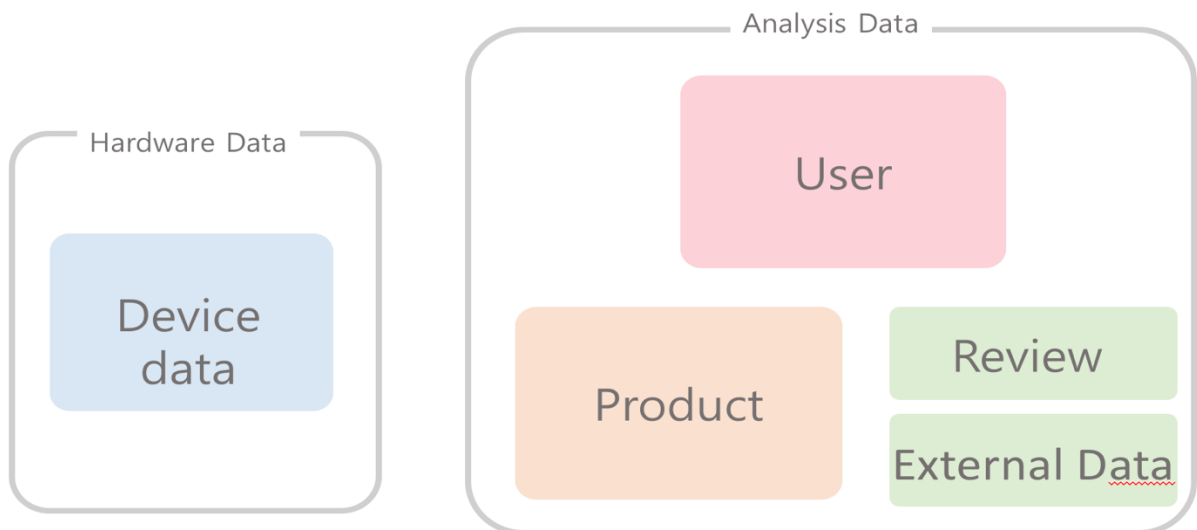
소비자와 제품 매칭 플랫폼의 핵심은 사용자의 여러 정보들과 맞는 화장품을 찾는 것이다. 인턴기간 동안 뷰티, 피부에 영향을 미치는 다양한 요소들과 어떤 영향을 미치는지에 대해서 리서치를 하였다. 그것을 바탕으로 화장품 선택 시 여러 요소들로 사용자와 제품의 매칭에 활용할 수 있는 방법을 구상하고 그것을 추천 알고리즘으로 기획하였다. 이러한 추천 알고리즘을 수동으로 시뮬레이션 했었을 때 실제 사용자가 매긴 제품 순위와 추천 알고리즘의 제품 추천 순위가 동일하였다.

기획하였던 이 추천 알고리즘이 유효하다고 판단하여서 이번 프로젝트를 통해서 뷰티에 영향을 주는 많은 요소들의 데이터 구조를 설계하고 이러한 요소들을 이용해서 데이터베이스에서 추천 알고리즘을 위한 Query를 구현하는 것이 목표이다.

2 수행 내용 및 중간결과

2.1 계획서 상의 연구내용

크게 두 가지 내용에 대한 데이터를 설계할 예정이다. 애플레이어를 통해 들어오는 피부 측정 하드웨어 Data와 추천 알고리즘 분석용 Data를 설계할 예정이다. 피부 측정 하드웨어에는 유 수분 데이터가 들어올 예정이고 추천 알고리즘 분석용 data는 유저, 후기, 제품 크게 세 개로 나누어 구조를 설계할 것이다.




(1) 애플레이어 하드웨어 data

:- 현재 하드웨어 data는 사용자의 피부 상태에 객관적 지표가 된다. 지금 하드웨어로 측정하는 사용자의 피부 상태는 local로만 저장되고 서버로는 아직 데이터구조가 없기 때문에 하드웨어에서 얻는 데이터를 구조화할 예정이다. 이후 이 데이터를 제품이나 관리에서 활용할 방법을 연구할 예정이다.

(2) 분석용 data

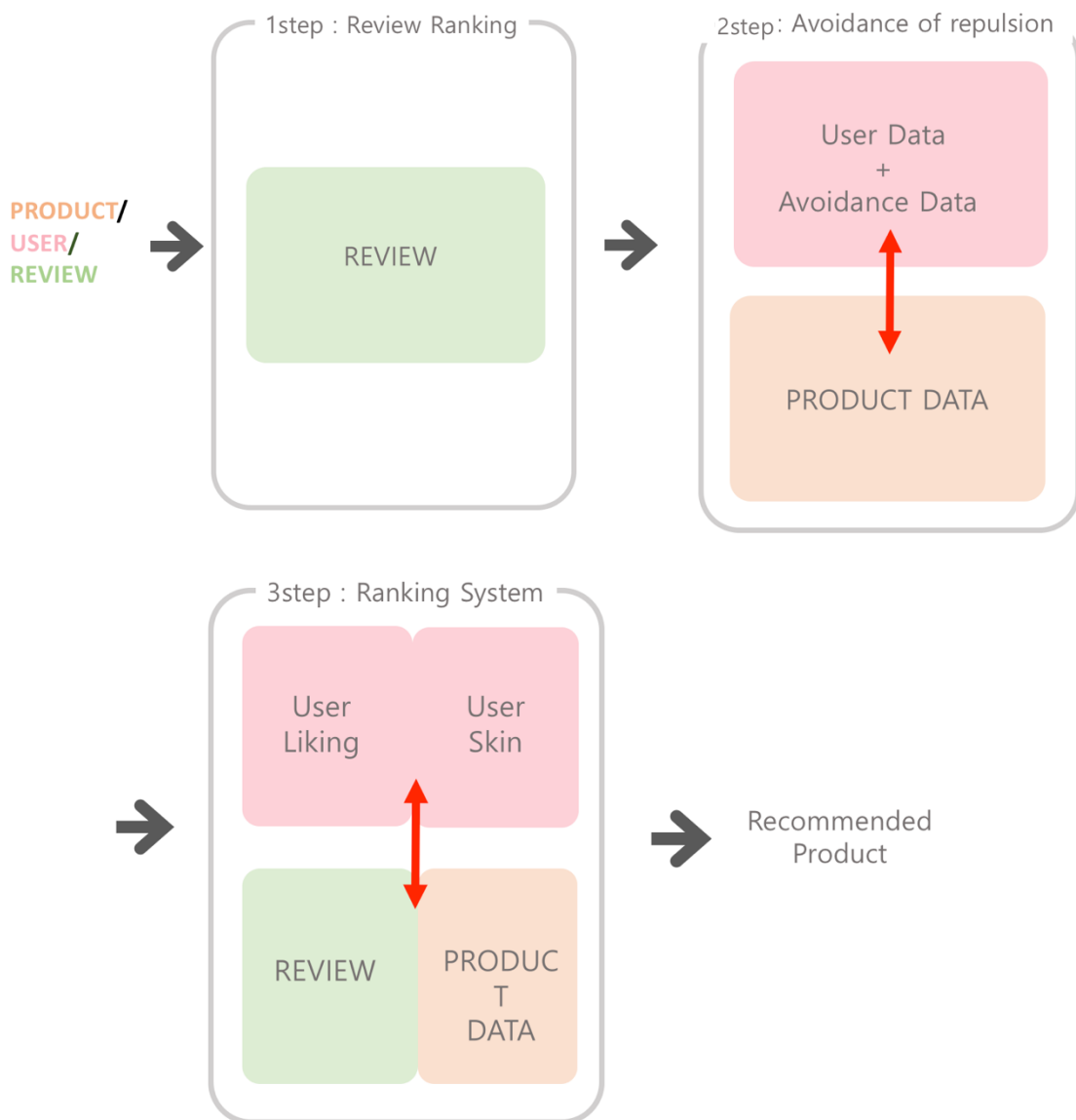
:- 분석용 data에는 크게 사용자, 제품, 후기, 외부 후기 데이터가 있습니다.


사용자는 회원가입은 간단하게 하고 이후 서비스를 이용하면서 취향을 분석하기 위한 기피 요소와 사용자 피부에 맞는 제품을 추천하기 위한 피부 정보가 들어간다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12

제품 데이터는 제품의 기본 정보와 케어에 관한 정보가 들어간다. 기본으로 크롤링을 이용하여 가져오나 실제 제품 정보와 실사용 정보가 완벽히 일치하는 경우는 적습니다.

후기 데이터는 제품의 성능 평가 정보와 사용자의 취향을 반영하는 정보로 제품 성능을 업데이트하고 사용자의 취향을 업데이트하는데 중요한 역할을 하는 데이터로 이루어진다.



 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12

추천 알고리즘은 크게 첫 번째 단계와 두 번째 단계 세 번째 단계로 이루어진다.

(1) **첫 번째 단계 - 단순 랭킹** : 사용자의 정보가 없기 때문에 제품과 후기 데이터에 의해서만 추천이 되는 단계이다. 타 사용자들의 후기 데이터 평가를 기반으로 높은 점수를 받은 데이터가 상위 추천되는 시스템이다.

(2) **두 번째 단계 - 기피 정보** : 추천 알고리즘에 사용할 많은 제품 중에서 소비자가 절대 고르지 않을 것 같은 제품은 미리 배제 해놓기 위한 단계이다. 뷰티 제품은 어느 특정 제품군이나 특징을 갖는 제품은 아예 사용을 안 하는 경우가 대부분이다. 이러한 특성을 살려서 1차 배제를 통해서 리소스 집단을 줄인다.

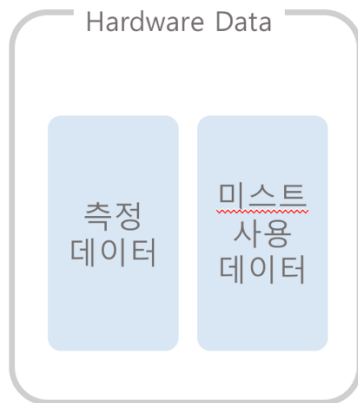
(3) **세 번째 단계 - 피부 정보** : 두 번째 단계에서 충분히 배제된 리소스를 가지고 기피 정보와 후기로 업데이트된 사용자 취향과 사용자의 피부 정보, 자신의 피부와 맞는 제품 추천, 유사한 사용자들의 제품 평가 등을 요소로 넣어서 제품마다 매칭 정도를 파악한다.

매칭한 것을 바탕으로 상위 5개의 제품을 뽑아서 추천할 것이다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12

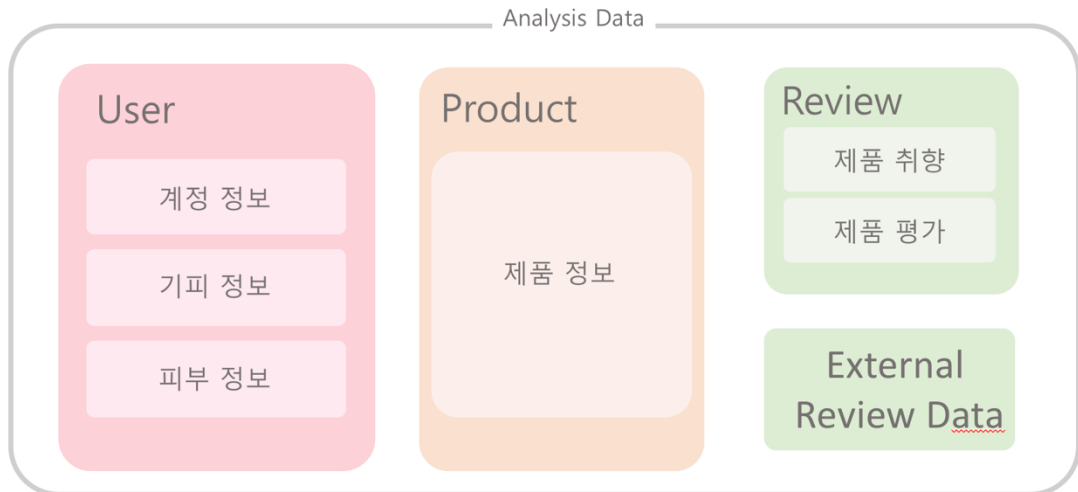
2.2 수행 내용

2.2.1 하드웨어 데이터 설계



기업에서 개발한 하드웨어 크게 두 가지 기능을 한다. 피부 유 수분을 측정하고 이후 즉각적으로 미스트를 사용하는 것이다. 현재 하드웨어에서 얻는 데이터는 local로만 저장되고 서버로는 아직 데이터구조가 없다. 기능을 바탕으로 하드웨어 데이터를 크게 측정 데이터 SET과 미스트 사용 데이터 SET으로 설계하였다.

2.2.2 분석용 데이터 설계




분석용 데이터는 크게 사용자, 제품, 후기, 외부데이터가 있다. 이 데이터를 조합해서 추천 알고리즘에 반영할 것이기 때문에 추천 알고리즘에 사용되는 요소들을 제대로 설계하는 게 중요하다.

(1) 사용자

사용자에는 크게 세 가지 데이터가 들어간다. 사용자 계정 정보로 앱을 통해서 유입되는 사용자를 관리하기 위한 데이터이다. 회원가입 시 기재하는 간단한 정보를 저장할 데이터 set이다. 사용자 기피 정보는 추천 알고리즘에서 두 번째 단계를 시행할 때 많이 사용될 데이터 set이다. 사용자의 기피 정보와 제품 정보를 매칭하여 기피 정보를 배제할 때 많이 쓰일 데이터 set이므로 각 요소와 많이 연결된 테이블이다. 사용자 피부 정보는 추천 알고리즘 세 번째 단계를 시행할 때 사용할 데이터 set이다. 뷰티 케어 서비스의 핵심일 뿐만 아니라 정확한 제품 매칭에 가장 핵심이 되는 부분이고, 피부 상태의 종류 또한 많기 때문에 여러 테이블로 설계하였다.

(2) 제품

제품에는 제품 정보에 관한 데이터 set이다. 제품에 대한 전반적인 데이터들을 담고 있고 그 정보들을 바탕으로 추천 알고리즘에서 사용자 기피 정보와 사용자 피부 정보와

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12


매칭 하는 부분이다.

(3) 리뷰

리뷰는 사용자가 제품에 대해서 남기는 데이터 set으로 크게 두 가지로 나뉜다. 화장품은 제품 기본 정보가 있지만, 실제 사용하는 부분에서 많은 차이가 있다. 그 부분을 보안을 유지하기 위한 실제 사용자들의 후기를 반영에 많은 비중이 있다. 제품 취향은 그러한 부분 중에서도 선호도로 많이 갈려지는 부분을 저장하는 데이터 set이다. 제품 평가는 제품을 평가하는 기준이 되는 요소들로 이루어지면서 성능적인 부분에서 좋다고 평가하는 데이터 set이다.

(4) 외부 데이터

외부 데이터는 실제 기업 측에서 받는 데이터 구조를 지원받았으면 더 좋았겠지만, 지원이 어려우므로 직접 서비스를 이용하고 데이터 구조를 설계해보았다. 해당 서비스는 사용자의 데이터는 피부 타입만 받고, 제품도 기본적인 정보만 기입하고 나머지는 리뷰 위주의 서비스 였다. 외부데이터 분석 결과 사용자와 제품데이터는 제로 파운더스에서 기획한 데이터 안에 포함되는 내용이었고, 후기 데이터를 집중적으로 분석하여 설계하였다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12

3 수정된 연구내용 및 추진 방향

3.1 수정사항

전반적으로 프로젝트의 범위가 많이 수정되었다.

3.1.1 데이터베이스 설계 수정


데이터베이스 분석용 데이터 설계에서 사용자, 제품, 후기, 앱 사용 데이터에 관한 데이터 설계를 연구 목표로 두었는데 앱 사용 데이터에서 외부 데이터로 변경되었다.

앱 사용 데이터에 대해서는 추천 알고리즘에 뚜렷하게 영향을 끼치는 요소가 없었다. 회사 측에서는 외부데이터를 받아서 사용할 예정이었기 때문에 이번 프로젝트의 데이터 설계를 위해서 외부데이터 구조와 샘플을 요청하였지만, 협력 기업 측은 보안상의 이유로 요청을 거부했다. 외부 리뷰 데이터는 매칭 추천알고리즘에 영향을 주는 요소라서 활용할 데이터 구조를 직접 설계해야 했기에 수정하였다.

3.1.2 추천 알고리즘 설계 수정

추천 알고리즘 설계를 기피 정보 배제하는 첫 번째 단계와 피부 정보를 바탕으로 여러 요소와 매칭하는 두 번째 단계로 설계하였는데 이 두 단계를 세 단계로 수정하였다.

사용자가 앱을 이용할 때 사용자의 기피 정보와 피부 정보 입력을 필수로 생각하였지만 정보 기재가 많아질 경우 이탈률이 높아지기 때문에 세단 계로 좀 더 세부화하였다. 첫 번째 단계로 정보 미 기재 시 축적된 리뷰데이터를 이용한 단순 랭킹 시스템을 활용하는 것을 추가하였다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12

3.1.3 개발 결과 산출물 수정

결과 범위와 산출물도 추천 알고리즘 구현하여 CSV로 프로토타입 개발에서 추천알고리즘을 위한 query 구현으로 수정하였다. 기획했던 데이터를 바탕으로 데이터 설계를 하는 거였기에 기획한 것과 실제 데이터 구조가 크게 다르지 않을 거라고 예상하여 추천 알고리즘을 구현하여 프로토타입으로 만들 수 있을 거라고 범위를 예상하였지만 실제로 프로젝트를 진행하면서 기획했던 데이터를 실제 데이터베이스로 구현하는데 많은 시간이 소요되면서 결과 범위와 산출물을 수정하였다.

결과 산출물로 알고리즘 유효성 확인을 추가하였습니다. 또 수동으로 시뮬레이션 돌린 추천 알고리즘이지만 알고리즘이 유효한지 정확하게 확인하지 못한 부분이기 때문에 추천 알고리즘 설계 후에 알고리즘 유효성을 확인할 것이다.


3.1.4 배경 기술의 수정

데이터베이스를 설계하기 위해서 MySQL을 사용하는 것은 여전하나 개발 결과 범위를 축소하면서 Python으로 추천 알고리즘 프로토타입 개발이 삭제되었다.

3.1.5 개발 일정의 수정

전반적인 프로젝트의 범위가 축소하면서 개발 일정도 그에 맞춰서 수정하였다. 기존의 기획한 데이터의 구조를 설계하는 일이 생각보다 커지면서 3월 말에 마무리되는 데이터 설계가 4월 초까지 수정이 되었다. 데이터 설계를 좀 더 세부화해서 4월까지의 데이터 설계 완성과 구조에 맞춰서 데이터를 넣고 수정하는 단계를 할 것이다.

그리고 추천 알고리즘 구현에 관해서도 4월부터 시작하려 하였으나 데이터 설계가 미완성이므로 4월 말 데이터 설계가 완성된 이후로 추천 알고리즘에 맞는


 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12

query 문을 구현할 예정이다. 알고리즘 구현 이후 알고리즘 유효성을 추가로 확인할 예정이다.

3.1.5 향후 수정에 대한 예정

사용자, 제품 매칭 추천 플랫폼 프로젝트는 향후 수정될 내용이 크게 많지 않다. 활용 데이터에 대해 인턴 기간 기획을 완료했고, 실제 추천 알고리즘에 대해서도 수동 시뮬레이션을 했기 때문에 내용적인 부분은 수정할 예정이 없다.

다만 해당 내용을 구현하는 방식에서 작은 변경이 있을 수 있다고 생각한다. 그러한 부분을 적절하게 변경하는 것이 이 프로젝트에서 맡은 역할의 핵심이라고 생각한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12

4 향후 추진계획

4.1 향후 계획의 세부 내용

4.1.1 데이터 설계에 따른 데이터 생성

데이터 설계한 것을 바탕으로 추천 알고리즘을 구현하기 위해서는 설계 위에 실질적으로 데이터를 생성해봐야 좀 더 명확하게 알 수 있다.

사용자는 인턴 때 시뮬레이션 참가한 사용자뿐만 아니라 추가로 사용자 데이터를 생성하도록 할 예정이고, 제품과 후기는 직접 작성한 것을 바탕으로 하기에는 데이터가 적고 정확성이 부족하므로 크롤링을 바탕으로 추천알고리즘 구현에 적절한 양의 데이터를 생성할 예정이다.


데이터 생성 이후에 잘못된 설계된 데이터 구조를 수정할 예정이다.

4.1.2 추천 알고리즘 query 생성

추천 알고리즘은 세 번째 단계로 이루어진다. 리뷰를 바탕으로 단순 랭킹 단계, 기피 정보 배제 단계, 피부 정보 활용한 단계이다. 모든 사용자가 추가 정보를 제대로 입력해주면 심화되고 정확한 추천이 가능하나 실제로 추가 정보 입력에서 이탈률이 가장 많기 때문에 단계마다 데이터를 입력할 때를 고려하여 각 단계에 맞는 데이터를 뽑을 수 있도록 할 것이다.

(1) **단순 랭킹** : 첫 번째 단계는 사용자의 정보가 없기 때문에 제품과 후기 데이터에 의해서만 추천이 되는 단계이다. 타 사용자들의 후기 데이터 평가를 기반으로 높은 점수를 받은 데이터가 상위 추천이 될 수 있도록 할 예정이다.

(2) **기피 정보 배제** : 추천 알고리즘에 사용할 많은 제품 중에서 소비자가 절대 고르지 않을 것 같은 제품은 미리 배제해놓기 위한 단계이다. 피부 정보를 매칭하기 위해서는 매칭할 리소스를 줄이는 것이 중요하다. 뷰티 제품을 선택할 때 어느 특정 제품군이나 특징을 갖는 제품은 아예 사용을 안 하는 경향이 있어서 배제하는 부분에서 원하는 제품을 찾는데 많은 영향을 줄 것이다. 이러한 특성을 살려서 1차 배제를 통해서 리소스 집단을 줄인다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	매칭 추천 알고리즘 db설계	
	팀 명	제로파운더스	
	Confidential Restricted	Version 1.0	2018-APR-12

(3) **피부 정보 추가** : 세 번째 단계는 두 번째 단계에서 충분히 배제된 리소스를 가지고 기피 정보와 후기로 업데이트된 사용자 취향과 사용자의 피부 정보, 자신의 피부와 맞는 제품 추천, 유사한 사용자들의 제품 평가 등을 요소로 넣어서 제품마다 매칭 정도를 파악한다.

매칭한 것을 바탕으로 상위 5개의 제품을 뽑아서 추천할 것이다.

4.1.2 추천 알고리즘 유효성 검사

추천알고리즘을 통해서 원하는 값을 얻었더라도 이 추천 알고리즘이 유효한 알고리즘인지를 확인하려면 유효성 체크를 해야 한다. 유효성 방법을 연구하고 유효성 검사 후 결과를 제출할 예정이다.

5 고충 및 건의사항

산학협력 캡스톤 1인 팀도 다른 팀과 유사하게 진행하는데 따로 캡스톤 과제 지원이 없습니다.