



Beer

Text Mining Project Final Presentation

Group Name : CHEERS 
Members : 곽민구, 김희영, 이창용

Contents

1. 프로젝트 개요
2. 협업 필터링 추천 시스템
3. 추천 시스템 제안

PART 01 >>

[프로젝트 개요]

>> 키워드 기반의 맥주 추천시스템

NAVER **특소는 맥주** 검색

통합검색 | Google **caramel tasting beer** 🔍

블로그

특소는 맥
어제도 모던
크림맥주를
닥터포유 출

특소는 맥
어제 하루종
잔벌컥벌컥
금갈치의 새

전체 이미지 동영상 지도 뉴스 더보기 ▼ 검색 도구

검색결과 약 21,300,000개 (0.36초)

관련검색: [tasting beer randy mosher pdf](#) [tasting beer randy mosher](#)
[tasting beer amazon](#) [beer tasting festival](#) [beer tasting serving paddle](#)

Most Caramel tasting beers? | Community | BeerAdvocate
[www.beeradvocate.com](#) > ... > General > Beer Talk > 이 페이지 번역하기
2015. 8. 9. - 댓글 39 - 작성자 39
I love when beers have strong caramel hints, weather it be malty IPAs or any other kind of beer. So far I think the one I consider strongest in this.

Caramel Cream Ale recipe?	게시물 27개	2016년 1월 21일
Adding Caramel Flavor	게시물 20개	2014년 9월 28일
Beer with caramel taste	게시물 40개	2013년 2월 7일

[www.beeradvocate.com](#) 검색결과 더보기

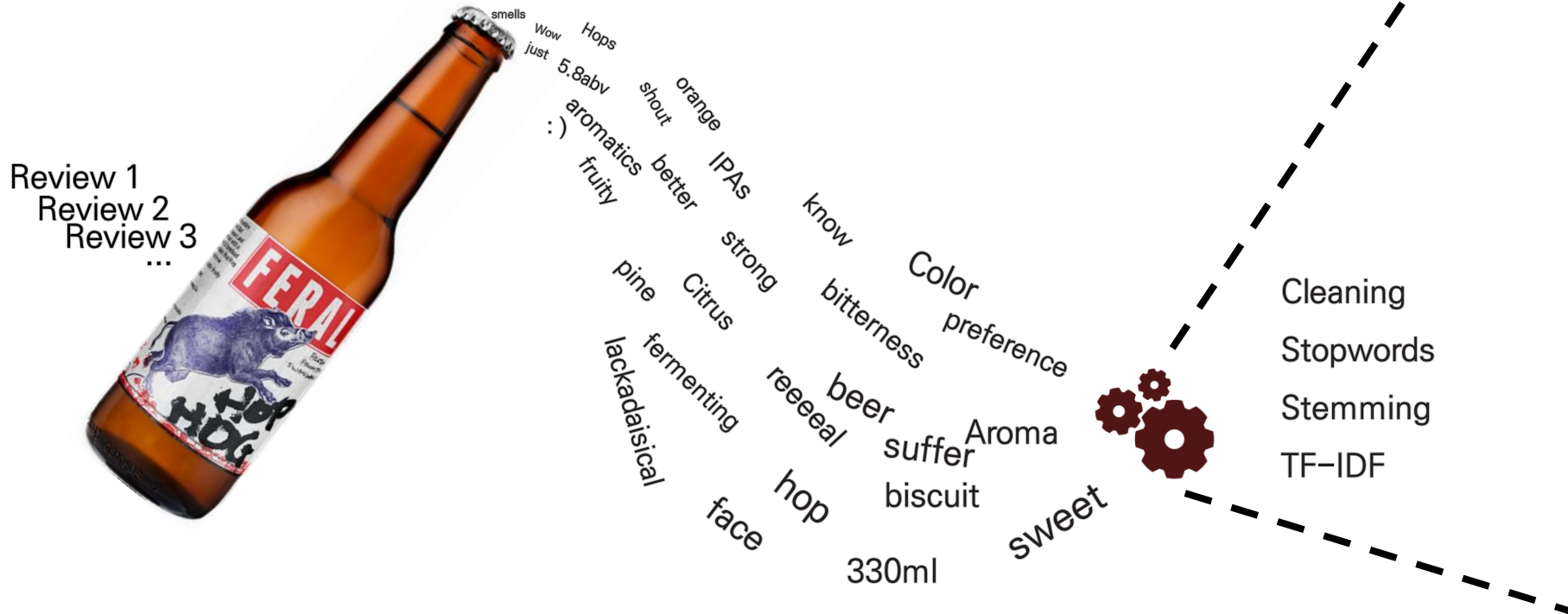


01

02

03

➤ 맥주별 키워드 추출



phrase	score
ipa	0.35
hop	0.31
bitter	0.2
citru	0.19
pine	0.18
passionfruit	0.14
orang	0.14
malt	0.12
great	0.11
flavour	0.11
bottl	0.11
aroma	0.1
carbon	0.1
well	0.1
bodi	0.09
littl	0.09
sweet	0.09
quit	0.09
american	0.09
tap	0.09

⋮
TF-IDF 상위 25개 단어

사전 작업: 국가별 맥주-키워드 매트릭스

키워드 250~500개

국가별 맥주
100개

	central	char	cherri	chewythi	chile	chili	chocol	cinnamon	citra	citru	cloudi	cocoa	coconut	coffe
Bourbon Barrel Age	0	0	0	0	0	0	0.269898	0	0	0	0	0	0	0.10353
Bourbon County Br	0	0	0	0	0	0	0.146706	0	0	0	0	0	0	0
Bourbon County Br	0	0	0	0	0	0	0.223887	0	0	0	0	0	0	0.711288
Bourbon County Br	0	0	0	0	0	0	0.332463	0	0	0	0	0	0	0.149385
Cable Car	0	0	0	0	0	0	0	0	0.090385	0	0	0	0	0
Citra Double IPA	0	0	0	0	0	0	0	0.220801	0.232615	0	0	0	0	0
Columbus Bodhi DI	0	0	0	0	0	0	0	0	0.272657	0	0	0	0	0
Congress Street IPA	0	0	0	0	0	0	0	0	0.240018	0	0	0	0	0
Dark Horse Bourbon	0	0	0	0	0	0	0.395878	0	0	0	0	0	0	0.129735
Dinner	0	0	0	0	0	0	0	0	0.233586	0	0	0	0	0
Doppelganger	0	0	0	0	0	0	0	0	0.13829	0	0	0	0	0
Double Citra	0	0	0	0	0	0	0	0.347932	0.254872	0	0	0	0	0
Double Dry Hopped	0	0	0	0	0	0	0	0	0.165225	0	0	0	0	0
Double Dry Hopped	0	0	0	0	0	0	0	0	0.220043	0	0	0	0	0
Double Galaxy	0	0	0	0	0	0	0	0	0.245872	0	0	0	0	0
Double Shot	0	0	0	0	0	0	0.272155	0	0	0	0.057882	0	0.811813	0
Double Shot - Sum	0	0	0	0	0	0	0.237189	0	0	0	0.101555	0	0.723473	0
Double Sunshine IP	0	0	0	0	0	0	0	0	0.255292	0	0	0	0	0
Duck Duck Goose	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ephraim	0	0	0	0	0	0	0	0	0.221625	0	0	0	0	0
Everett Porter	0	0	0	0	0	0	0.431584	0	0	0	0	0	0	0.329589
Evil Twin Double Ba	0	0	0	0	0	0	0.479937	0	0	0	0	0	0	0.154627
Focal Banger	0	0	0	0	0	0	0	0.095453	0.207177	0	0	0	0	0
Fort Point Pale Ale	0	0	0	0	0	0	0	0	0.239752	0	0	0	0	0
Founders Breakfast	0	0	0	0	0	0	0.453504	0	0	0	0	0	0	0.663528
Founders CBS Impe	0	0	0	0	0	0	0.312528	0	0	0	0	0	0	0.401653
Founders KBS (Kent	0	0	0	0	0	0	0.396543	0	0	0	0	0	0	0.464904
Framboise Du Fermi	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fundamental Obser	0	0	0	0	0	0	0.24816	0	0	0	0.069785	0	0	0
Fuzzy Baby Ducks IP	0	0	0	0	0	0	0	0.246182	0.234782	0	0	0	0	0

키워드 사전 구축

01

02

03

>> 도메인 지식과 결합



- 맥주 관련 서적
- Beer Advocate (데이터 출처)에
나와있는 예시 단어

>> 키워드 사전 구축

구분	Look	Smell	Taste	Feel
단어 수	140개	93개	214개	168개
키워드	brown	aroma	sweet	bodi
	black	smokey	spice	smooth
	pink	wood	lemon	rich
	dark	grass	vanilla	refresh
	amber	burn	mango	mild
	gold	fruit	acid	awesom
	red	lavend	orang	super

카테고리별 단어

총 615개 선정

PART 02>>

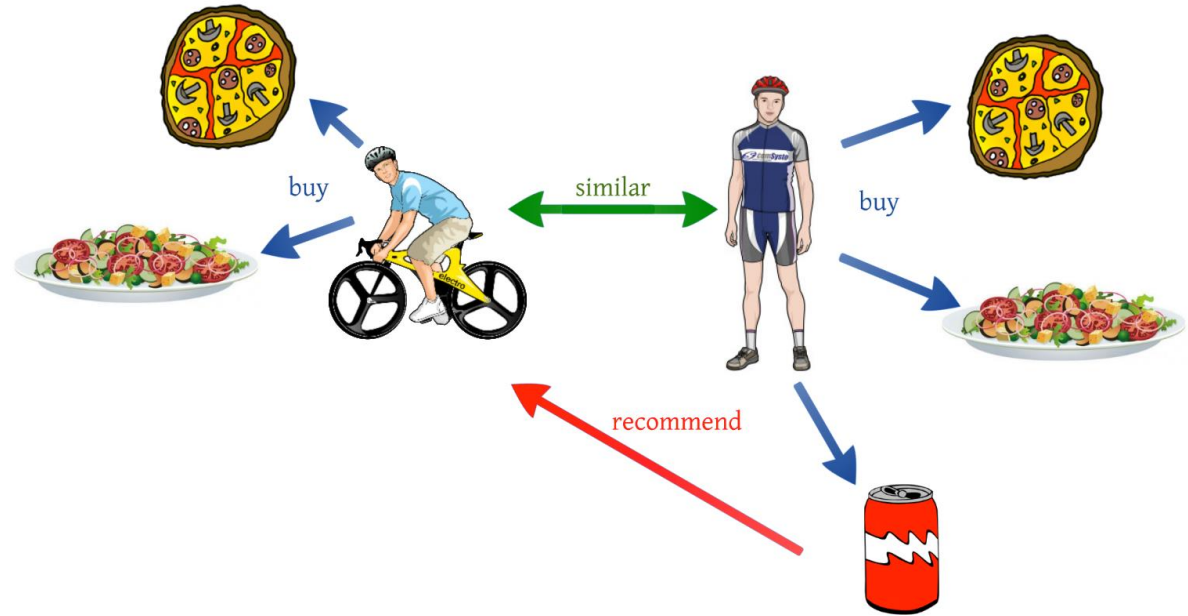
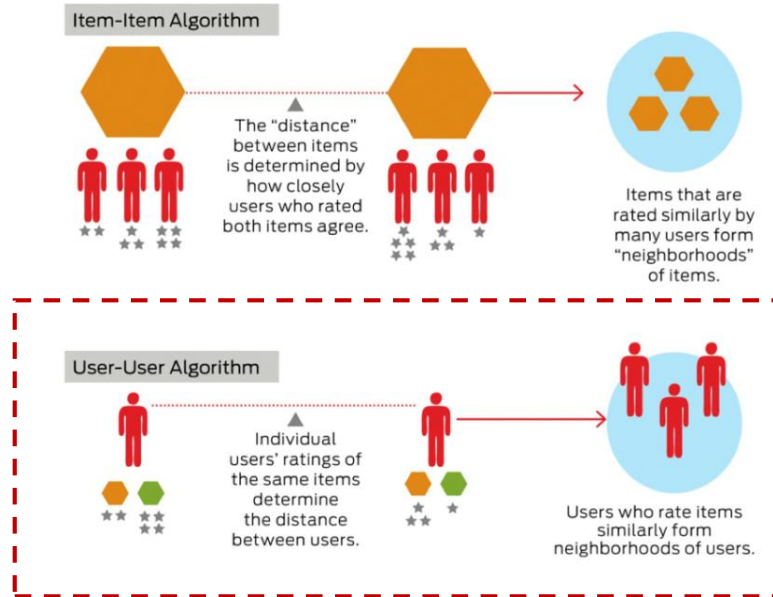
[협업 필터링 추천 시스템]

01

02

03

Memory-based



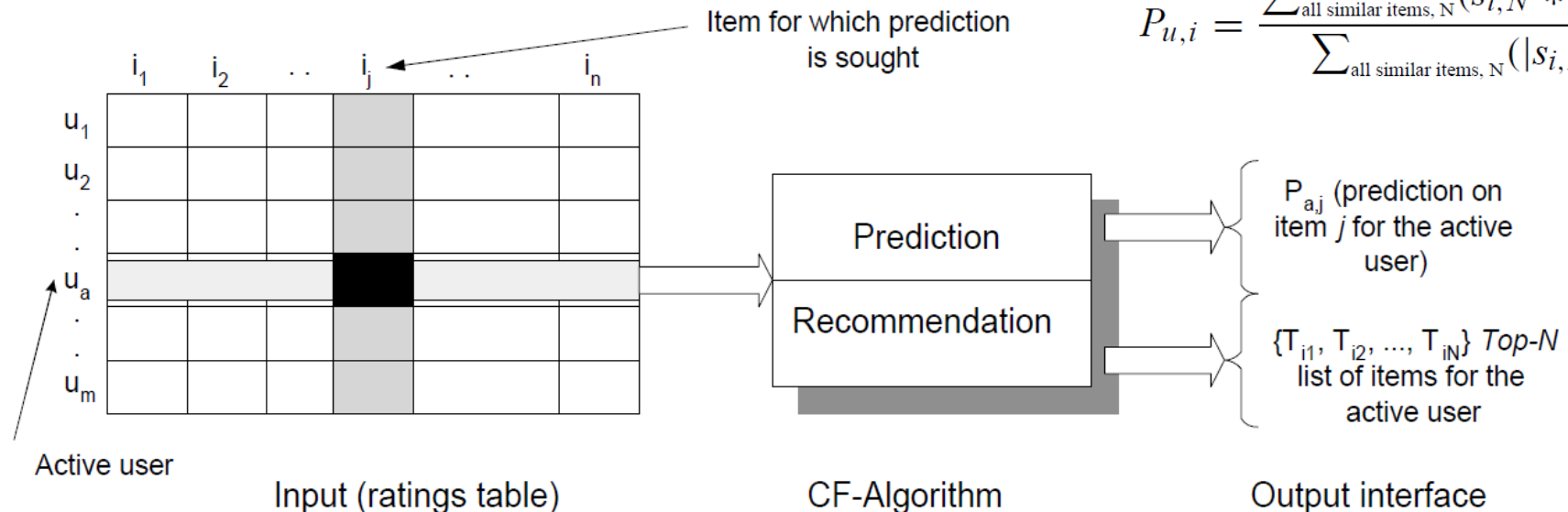
비슷한 구매 패턴을
보이는 사용자를 선정

다른 사용자의 평가에
기반하여 미사용 아이템에
대해 평가점수 측정

상위 N개 추천

»» 시스템 구현 내용

- 미국 리뷰 데이터: 리뷰 25개 이상 작성한 사용자 131명
- 사용자-아이템 매트릭스 구성
- 평가 점수의 코사인 유사도 (Cosine Similarity) 를 통해 평가 점수 예측
- 평가 하지 않은 맥주 중 평점이 높은 Top N개 (3개) 추천



$$P_{u,i} = \frac{\sum_{\text{all similar items, } N} (s_{i,N} * R_{u,N})}{\sum_{\text{all similar items, } N} (|s_{i,N}|)}$$

결과 및 한계

- Mean Absoluter Error: **1.23** (5점 척도)
- Sparsity : 리뷰 수가 적은 사용자가 Sparse Matrix 생성
- **Cold Start Problem**: 시스템에 충분한 정보가 모이지 않아, 사용자 또는 아이템에 대해 추정이 어려운 문제

Active User	1st	2nd	3rd
Alieniloquium	Stevedore	Huhzubendah	dar482
Stevedore	Alieniloquium	dar482	Huhzubendah
dar482	Stevedore	Alieniloquium	gcamparone
gcamparone	dar482	Stevedore	GreesyFizeek
GreesyFizeek	dar482	Stevedore	Damian

유사 사용자

Active User	1st	2nd	3rd
Alieniloquium	Evil Twin-	Double Dry-	Pliny The Elder
Stevedore	Assassin-	AleSmith-	Congress-
dar482	Supplication	Double Dry-	Artaic
gcamparone	Focal Banger	Abner	Barrel Aged-
GreesyFizeek	Atrial Rubicite	Double Galaxy	Black Tuesday

추천 맥주

PART 03>>

[추천 시스템 제안]

01

»» 사용자가 원하는 특징 (=키워드) 을 띄는 맥주를 추천

02

03

04



텍스트 전처리

기존 리뷰 데이터와
동일한 과정으로
전처리 진행

- Cleaning
- Stopwords
- Stemming

유의미한 키워드

키워드 매트릭스와 비교
»» 공통 키워드 추출

Meaningful
Words

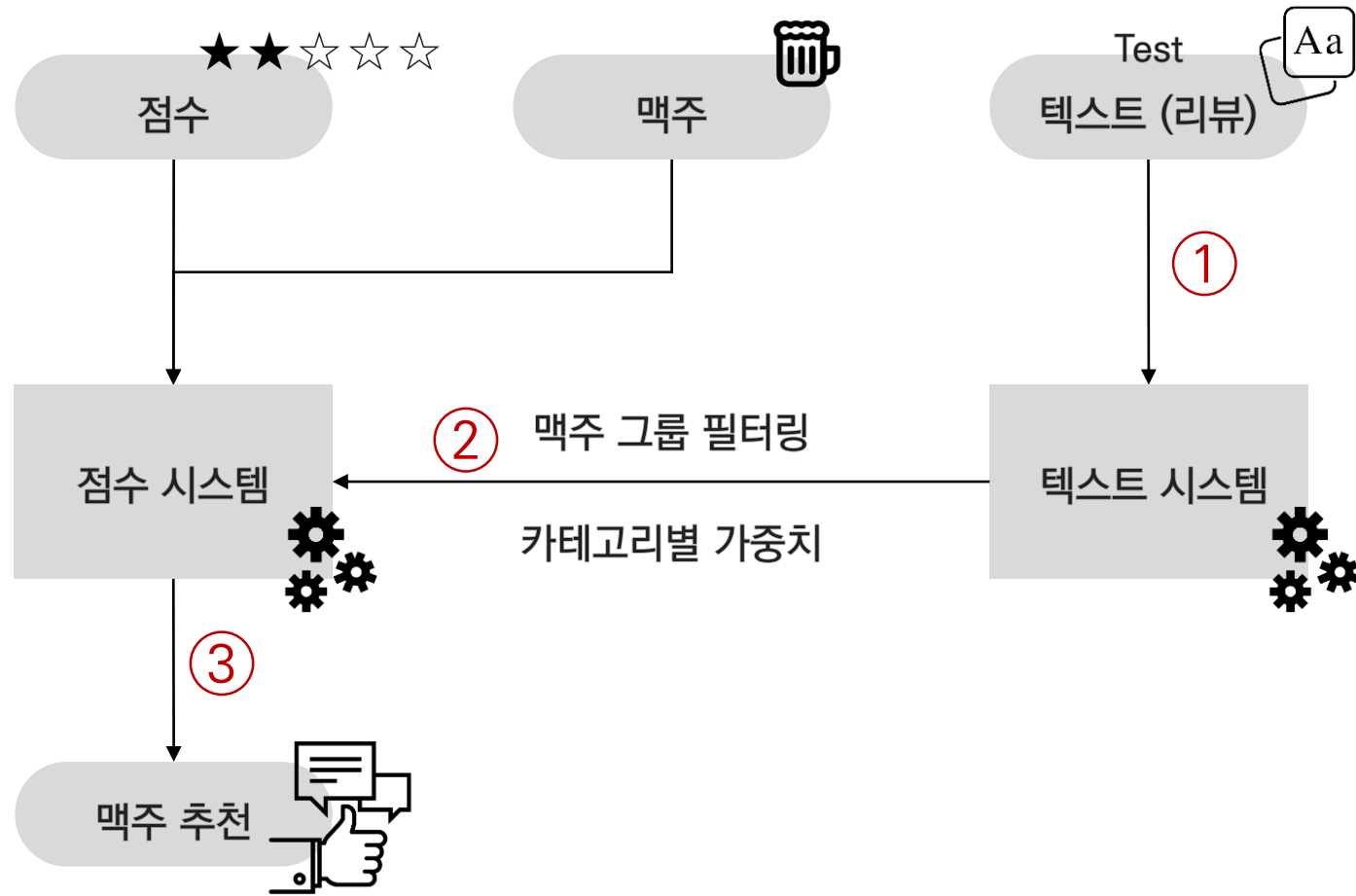
유사도 계산

맥주별 TF-IDF의
크기가 상이함
»» 코사인 유사도 사용

사용자가 원하는 맥주
특징에 대해 정량화

Top 3
맥주 추천

➤➤ 알고리즘 개요



텍스트 기반 추천 시스템

01

>> Step ① ②



Hazy golden; plenty of head.
cheese and rice...you can smell
the citrus and hops feet away
from the glass...more fruit than
you can shake your stick at.
Plenty of tropical, stone and citrus
fruit. Nice malt in the nose ~~~

Julius (Rank: 13) – American IPA



Overall 3.5
Feel 5.0
Look 4.0
Smell 2.0
Taste 3.0

카테고리별 가중치 획득

텍스트 전처리



	central	char	cherri	chewythi	chile	chilli	chocol	cinnamon	citra	citru	cloudi	cocoa	coconut	coffe
Bourbon Barrel Age	0	0	0	0	0	0	0.269898	0	0	0	0	0	0	0.10353
Bourbon County Br	0	0	0	0	0	0	0.146706	0	0	0	0	0	0	0.711288
Bourbon County Br	0	0	0	0	0	0	0.223887	0	0	0	0	0	0	0.149385
Bourbon County Br	0	0	0	0	0	0	0.332463	0	0	0	0	0	0	0.149385
Table Car	0	0	0	0	0	0	0	0	0.090385	0	0	0	0	0
Citra Double IPA	0	0	0	0	0	0	0	0	0.220801	0.232615	0	0	0	0
Columbus Bodhi D	0	0	0	0	0	0	0	0	0.272657	0	0	0	0	0
Congress Street IPA	0	0	0	0	0	0	0	0	0.240018	0	0	0	0	0
Dark Horse Bourbon	0	0	0	0	0	0	0.395878	0	0	0	0	0	0	0.129735
Dinner	0	0	0	0	0	0	0	0	0.233586	0	0	0	0	0
Doppelganger	0	0	0	0	0	0	0	0	0.13829	0	0	0	0	0
Double Citra	0	0	0	0	0	0	0	0	0.347932	0.254872	0	0	0	0
Double Dry Hoppe	0	0	0	0	0	0	0	0	0.165225	0	0	0	0	0
Double Dry Hoppe	0	0	0	0	0	0	0	0	0.220043	0	0	0	0	0
Double Galaxy	0	0	0	0	0	0	0	0	0.245872	0	0	0	0	0
Double Shot	0	0	0	0	0	0	0.272155	0	0	0	0	0.057882	0	0.811813
Double Shot - Sum	0	0	0	0	0	0	0.237189	0	0	0	0.101555	0	0	0.723473
Double Sunshine IP	0	0	0	0	0	0	0	0	0.255292	0	0	0	0	0
Duck Duck Goose	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ephraim	0	0	0	0	0	0	0	0	0.221625	0	0	0	0	0
Espresso Porter	0	0	0	0	0	0	0.431584	0	0	0	0	0	0	0.329589
Evil Twin Double B	0	0	0	0	0	0	0.479937	0	0	0	0	0	0	0.154627
Focal Banger	0	0	0	0	0	0	0	0	0.095453	0.207177	0	0	0	0
Fort Point Pale Ale	0	0	0	0	0	0	0	0	0.239752	0	0	0	0	0
Founders Breakfast	0	0	0	0	0	0	0.453504	0	0	0	0	0	0	0.663528
Founders CBS Impe	0	0	0	0	0	0	0.312528	0	0	0	0	0	0	0.401653
Founders KBS (Kent	0	0	0	0	0	0	0.396543	0	0	0	0	0	0	0.464904
Framboise Du Fermi	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fundamental Obser	0	0	0	0	0	0	0.24816	0	0	0	0	0.069785	0	0
Fuzzy Baby Ducks If	0	0	0	0	0	0	0	0.246182	0.234782	0	0	0	0	0

키워드 매트릭스와 비교

>> 공통 키워드를 포함하는

맥주 그룹 필터링



golden, malt, fruit,
tropic, citru, hazi

:

구축한 사전과 텍스트 비교



텍스트 기반 추천 시스템

>> Step ③



2가지 맥주를 동시에 평가한
유저 집합을 추출 (10명 이상)



Beer1	Beer2	Overall	Look	Smell	Taste	Feel
Julius	Pliny The Elder	-0.075	-0.131	-0.147	-0.066	-0.117
Julius	Pliny The Younger	0.385	0.384	0.482	0.152	0.237
Julius	PseudoSue	0.371	0.316	0.351	0.309	0.334
Julius	Rare Bourbon County Brand Stout (2015)	0.512	0.202	0.261	0.541	0.143
Julius	Resolute - Coconut	-10.000	-10.000	-10.000	-10.000	-10.000
Julius	Sip Of Sunshine	0.499	0.337	0.305	0.412	0.491
Julius	Snowed In - Bourbon Barrel Aged	-10.000	-10.000	-10.000	-10.000	-10.000
Julius	Society & Solitude #4	0.182	0.100	0.275	0.406	0.317
Julius	Society & Solitude #5	0.349	0.359	0.245	0.244	0.396
Julius	Sosus	0.406	0.693	0.567	0.064	0.333
Julius	Stickee Monkee	0.126	0.041	0.508	0.208	0.218
Julius	Susan	0.578	0.452	0.376	0.560	0.568
Julius	Swish	0.306	0.613	0.246	0.316	0.360
Julius	The Abyss	-0.050	0.369	-0.228	-0.085	0.169
Julius	Todd The Axe Man	0.205	0.143	0.236	0.115	0.112
Julius	Triple Sunshine IPA	0.500	0.086	0.603	0.348	-0.084
Julius	Upper Case	0.333	0.194	0.062	0.369	0.486
Julius	Very Green	0.502	0.808	0.064	0.218	0.225
Julius	Very Hazy	0.638	0.638	0.612	0.627	0.518

맥주 쌍에 대한 카테고리별
점수 유사도를 계산

>> Step ③

beer1	beer2	score
Julius	Very Hazy	10.48256
Julius	King Julius	10.25272
Julius	Doppelganger	10.00048
Julius	In Perpetuity	9.94483
Julius	Green	9.842391
Julius	Alter Ego	9.744494
Julius	Susan	9.101145

⋮

사전에 구해놓은 가중치를 사용해
가중합을 계산



Very Hazy (Rank: 14)
American Double / Imperial IPA



King Julius (Rank: 6)
American Double / Imperial IPA



Doppelganger (Rank: 39)
American Double / Imperial IPA

결과 예시

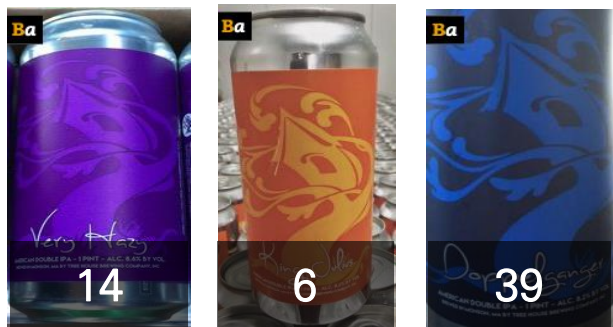
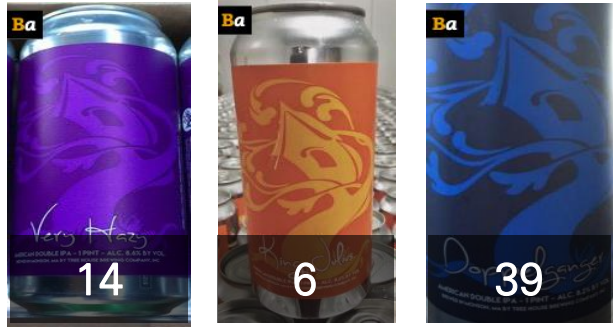
01

02

03

04

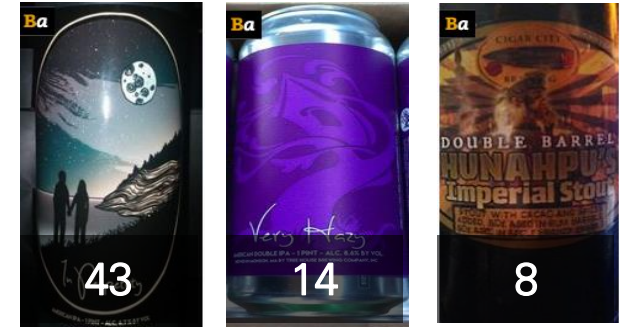
점수

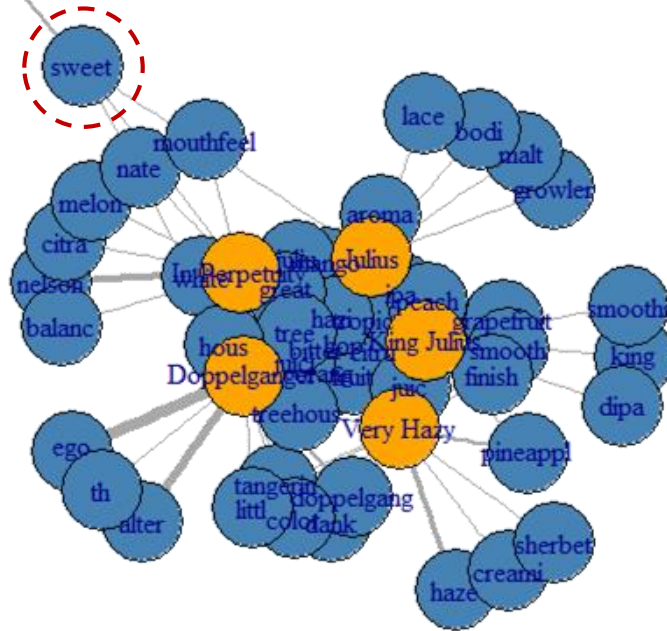
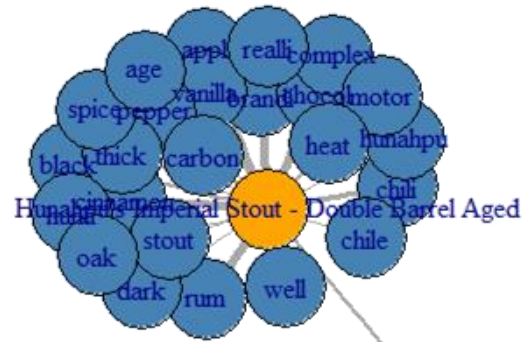


텍스트



제안 시스템





대표 키워드

Bitter	Orange
Citrus	Juicy
Peach	Mango

➤➤ 결론

- 획일적인 점수 기반 시스템 & 산발적인 텍스트 기반 시스템과 비교하여
비교적 **타당한 범위** 내에서 맥주를 추천

➤➤ Step

- 새로운 리뷰 데이터로 구현 모델을 **정량적으로 평가**하는데 어려움이 있음
- 사용자의 **개인 정보**를 활용하여 추천을 하는 요소가 필요함
- 사용자에게 요구하는 **정보의 양**이 많음