

세종대학원 인공지능언어공학과 빅파이크래프트 연구원 민아영

Index

- 1. 연구 소개
- 2. 연구 과정
- 3. 연구 결과
- 4. 연구 기여 방안

• 연구의 필요성

연구 소개

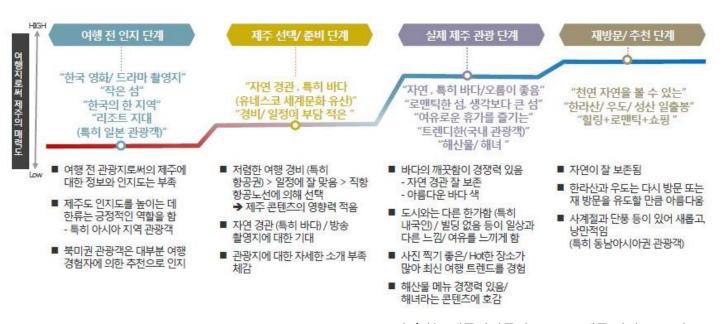
연구 과정

연구 결과

연구 기여 방안

제주 관광객의 여행 전~ 인터뷰 전까지의 여행 경험을 볼 때 여행 전 인지 단계 보다 Summary 1 여행이 이루어지면서의 매력도가 상승하고 있음

→ 여행 전과 준비 단계에서 제주에 대한 호감도 유발 요소 강화 필요



※그림 출처 : 제주관광공사 - 2018 제주 관광 보고 리포트

<u>동남아권 외국인 관광객의 제주도에 대한 관심 및 방문 증가</u>로 인한 새로운 관광객층을 위한 맞춤 마케팅 필요

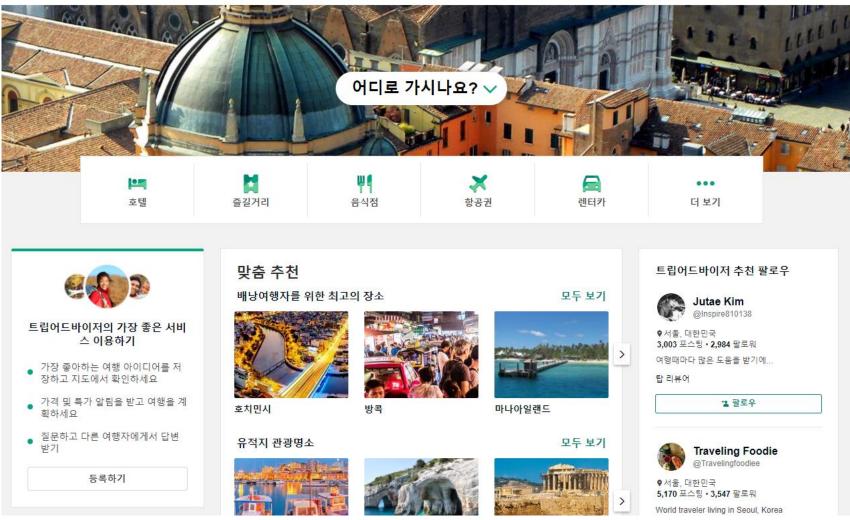
• 수집 데이터 - 트립어드바이저

연구 소개

연구과정

연구 결과





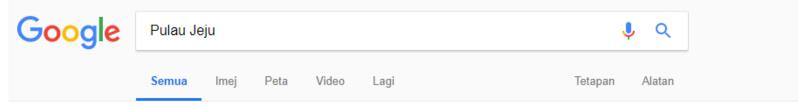
• 수집 데이터 - 구글 검색(Pulau Jeju)

연구 소개

연구 과정

연구 결과

연구 기여 방인



Halaman 2 daripada kira-kira 779,000 hasil carian (0.42 saat)

Pulau Jeju Korea | Harga Istimewa di Pulau Jeju | agoda.com

[lklan] www.agoda.com/Hotel/Promosi ▼

Harga Istimewa, -80% Diskon! Tempah Sekarang, Hanya di Agoda®

Kuala Lumpur Hotels - daripada USD14.00/hari - Up to 80% Off, Book Now · Lagi ▼

Lima Restoran Halal Yang Boleh Anda Singgah Di Pulau ...

www.mstar.com.my/jelajah/cerita/2018/02/01/makanan-halal-jeju/ 🔻

1 Feb 2018 - PULAU Jeju yang dinamakan semula kepada Wilayah Istimewa Khas Jeju memiliki pelbagai destinasi pelancongan bertaraf dunia yang sesuai dikunjungi sepanjang empat musim.

JEJU TOURISM ORGANIZATION | Cerita tentang Jeju > ...

ijto.or.kr/indonesian/?cid=27 ▼ Terjemahkan halaman ini

Setiap murid Korea mempelajari sejarah Jeju di sekolah karena Pulau Jeju merupakan sebuah tempat dimana sejarah masih hidup. Mari melihat sejarah Jeju melalui Periode Tiga Kerajaan yaitu dinasti Goryeo dan Joseon, serta periode penjajahan Jepang sampai dengan tahunan hiruk pikuk sejarah modern

7 Tempat ini jadi daya tarik wisata Pulau Jeju, gak salah ...

blog.airpaz.com/id/wisata-pulau-jeju/ ▼ Terjemahkan halaman ini

19 Jun 2017 - 7 Tempat ini jadi daya tarik wisata Pulau Jeju, gak salah jika Pulau Jeju masuk daftar UNESCO! Tahu gak, Airpazious kalau Jeju merupakan pulau terbesar yang berada di Korea Selatan. Jika Airpazious suka nonton drama korea, ternyata Jeju merupakan lokasi yang sering dijadikan tempat untuk syuting ...

• 수집 데이터 - 속성사전 구축(1차 사전)

연구 소개

연구 과정

연구 결과

접근성	분위기	문화	경험	관심	친절	가격	볼거리
address	atmosphere	accumulation	cognitive	enjoy	affectionate	affordable	garden
area	avant-garde	analects	creativity	exciting	aloof	cheap	breeze
automobile	breath	ancestor	experience	fashionable	amiable	cost	ability
banner	character	backbone	hearing	fun	anger	discount	aisle
barge	climate	ballad	illusion	funny	anxiety	expensive	animal
belt	condition	bible	illustration	happy	anxious	fee	aquarium
bicycle	conservation	ceremony	intelligence	hobby	arrogant	funds	art
bus	equipment	character	knowledge	humor	attitude	money	attractions
station	facility	civilization	learn	interest	bluntly	offer	bali
car	fashion	clothes	scene	interesting	blush	price	basalt

• 수집 데이터 - 속성사전 확장(WordNet)

연구 소개

연구 과정

연구 결과

연구 기여 방안

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: beautiful Search WordNet

Display Options: (Select option to change)

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Adjective

- <u>S</u>: (adj) **beautiful** (delighting the senses or exciting intellectual or emotional admiration) "a beautiful child"; "beautiful country"; "a beautiful painting"; "a beautiful party"
 - see also
 - similar to
 - S: (adj) beauteous ((poetic) beautiful, especially to the sight)
 - S: (adj) bonny, bonnie, comely, fair, sightly (very pleasing to the eye) "my bonny lass"; "there's a bonny bay beyond"; "a comely face"; "young fair maidens"
 - S: (adj) dishy ((informal British) sexually attractive) "a dishy blonde"
 - S: (adj) exquisite (of extreme beauty) "her exquisite face"
 - S: (adj) fine-looking, good-looking, better-looking, handsome, well-favored, well-favoured (pleasing in appearance especially by reason of conformity to ideals of form and proportion) "a fine-looking woman"; "a good-looking man"; "better-looking than her sister"; "very pretty but not so extraordinarily handsome"- Thackeray; "our southern women are well-favored"- Lillian Hellman
 - S: (adj) glorious, resplendent, splendid, splendiferous (having great beauty and splendor) "a glorious spring morning"; "a glorious sunset"; "splendid costumes"; "a kind of splendiferous native simplicity"
 - S. (adj) gorgeous (dazzlingly beautiful) "a gorgeous Victorian gown"
 - S: (adj) <u>lovely</u> (appealing to the emotions as well as the eye)
 - S: (adj) <u>picturesque</u> (suggesting or suitable for a picture; pretty as a picture) "a picturesque village"
 - S: (adj) pretty (pleasing by delicacy or grace; not imposing) "pretty girl"; "pretty song"; "pretty room"
 - <u>S.</u> (adj) <u>pretty-pretty</u> (ostentatiously or inappropriately pretty)
 - S: (adj) <u>pulchritudinous</u> (used of persons only; having great physical beauty) "<u>pulchritudinous movie stars</u>"
 - S: (adj) ravishing (stunningly beautiful) "a ravishing blonde"
 - S: (adj) scenic (used of locations; having beautiful natural scenery) "scenic drives"



동의어, 유의어 추출을 통한 속성사전 확장

• 수집 데이터 - 감성사전 구축 및 확장

연구 소개

연구 과정

연구 결과

idx	word	polarity_value	polarity_intense
1	colonial domination	negative	-0.82
2	macabre	negative	-0.88
3	amount water	positive	0.943
4	dreg	negative	-0.7
5	hard drink	positive	0.564
6	main bus terminal	positive	0.081
7	float upward	negative	-0.01
8	barrel-chested	positive	0.855
9	click link	positive	0.943
10	taobh	positive	0.875
11	ploughland	negative	-0.75
12	dextromethorphan	positive	0.871
13	deviance	positive	0.643
14	chirimoya	positive	0.888
15	hemipteran	negative	-0.72
16	northern coastline	positive	0.079

Work Flow

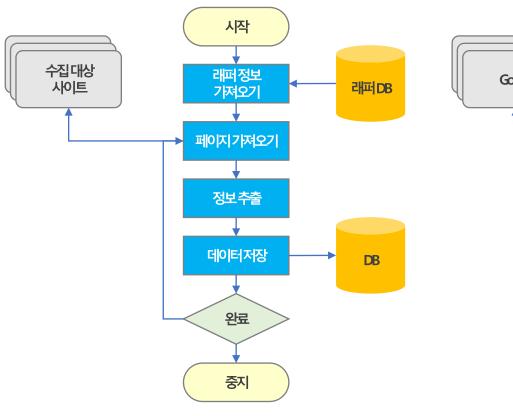
연구 소개

연구 과정

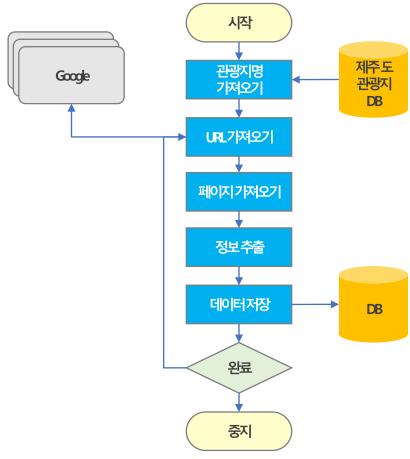
연구 결과

연구 기여 방안

● 래퍼기반웹데이터수집



● API기반웹데이터수집

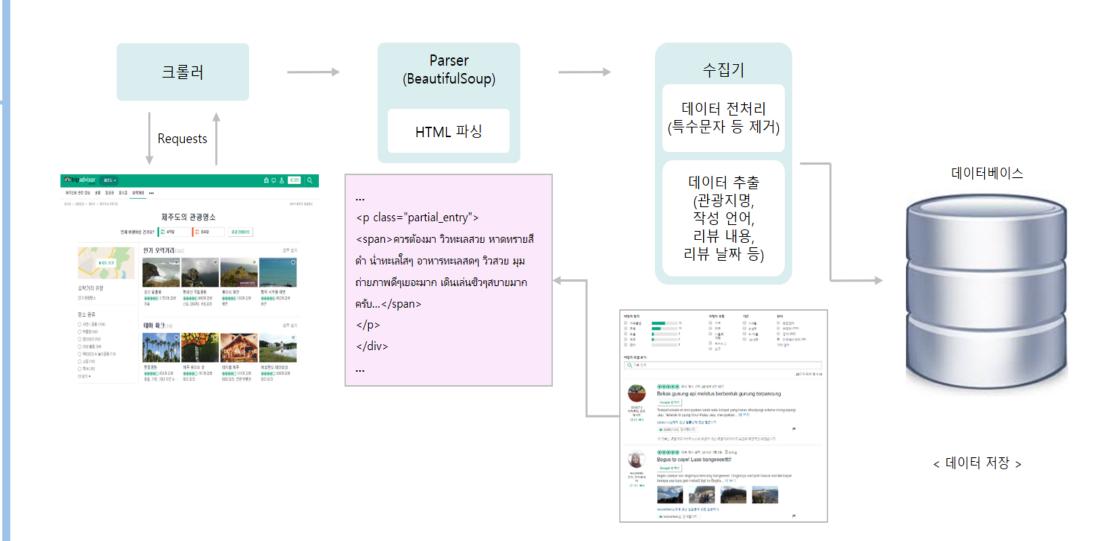


• 과정 1 - 크롤링

연구 소개

연구 과정

연구 결과

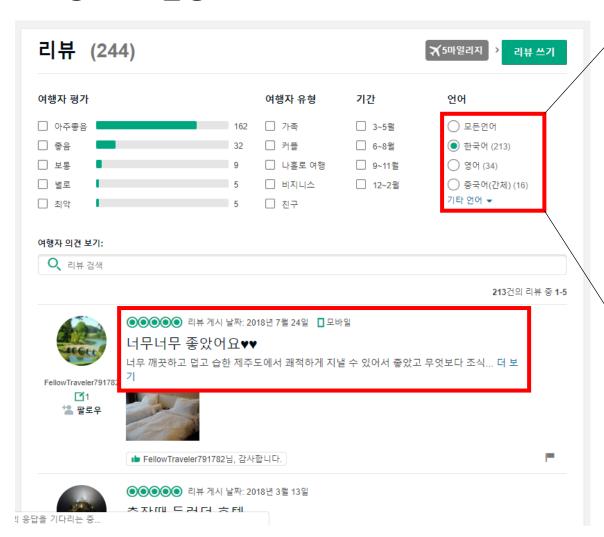


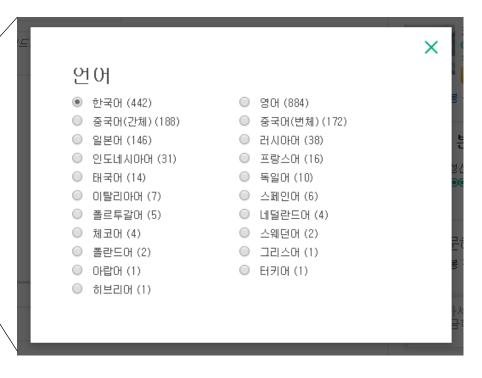
• 과정 1 - 크롤링

연구 소개

연구 과정

연구 결과





연구 소개

연구 과정

연구 결과

연구 기여 방안

• 과정 1 - 크롤링

* 카테고리 메인 접속

https://www.tripadvisor.{domain}/{activity}_Review-g{cityid}-d{activityid}



* 리뷰 메인 접속

https://www.tripadvisor.{domain}/ShowUserReviews-g{cityid}-d{activityid}-r{reviewid}

변수 명	변수 설명			
domain	언어별 도메인			
	오락거리 : Attractions ,			
activity	음식점 : Restaurants ,			
	호텔 : Hotels			
cityid	제주도 : 983296 , 297892 , 297885			
activityid	장소별 고유 ID			
reviewid	리뷰 고유 ID			

언어 종류	도메인	주소
태국어	com.tw	https://www.tripadvisor.com.tw
인도네시아어	co.id	https://www.tripadvisor.co.id
베트남어	com.vn	https://www.tripadvisor.com.vn
일본어	jp	https://www.tripadvisor.jp

• 과정 2 - 데이터베이스 저장

연구 소개

연구 과정

연구 결과

#	칼럼명	데이터 유형	기본 값	코멘트	Key
1	idx	int(11)	AUTO_INCRE MENT	인덱스	Primary
2	city_id	varchar(50)	null	도시 고유 코드	-
3	city_name	varchar(50)	null	도시 명	-
4	activity_type	varchar(50)	null	액티비티 종류	-
5	activity_id	varchar(50)	null	액티비티 고유코드	-
6	activity_name	varchar(50)	null	액티비티명	-
7	language	varchar(50)	null	작성된 리뷰 언어	-
8	review_id	varchar(50)	null	리뷰 고유 코드	-
9	review_date	varchar(50)	null	작성된 리뷰 날짜	-
10	review_title	varchar(1000)	null	작성된 리뷰 제목	-
11	review_conte nt	varchar(5000)	null	작성된 리뷰 내용	
12	review_rating	varchar(50)	null	작성된 리뷰 평점	
13	systime		CURRENT_TI	데이터 업데이트	

#	칼럼명	데이터 유형	기본 값	코멘트	Key
1	idx	INT	AUTO_IN CREMEN T	인덱스	Primary
2	word	VARCHAR (50)	null	감성 단어	-
3	pos_tag	VARCHAR (50)	null	품사	-
4	polarity_ value	VARCHAR (50)	null	감성 분류	-
5	polarity_ intense	DOUBLE	null	감성 수치	-

• 과정 2 - 데이터베이스 저장

연구 소개

연구 과정

연구 결과

연구 기여 방안

idx	city_id	city_name	activity_type	activity_id	activity_name	language	review_id	review_date	review_title	review_content
1	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	566608135	2018.03.15	Bekas gunung api meletus berbentuk	Tempat wisata ini mer
2	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	551742829	2018.01.05	Bagus tp cape! Luas bangeeeetttt!	begitu sampe sini ang
3	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	503556039	2017.07.19	Good and great view	Salah satu dari sekian
4	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	465999045	2017.03.10	BEST NATURE AND BEAUTIFUL VIEW	Subhanallah Besar Cip
5	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	396049109	2016.07.24	seven wonder, menguras keringat	lokasi ini merupakan s
6	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	393491904	2016.07.17	Tempat yg tepat untuk sedikit mengur	Lumayan bagus kiti
7	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	368342189	2016.04.28	Tempat ini amat menarik dan sesuai u	Kamu dapat menikma
8	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	366595127	2016.04.23	Hiking Pagi Hari untuk Mengejar Sunrise	Seongsan Ilchulbong
9	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	359688603	2016.03.29	Simbol pulau jeju	Jika kita melihat foto-
10	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	338372860	2016.01.08	Seongsan Ilchulbong	Tempat wisata ini mer
11	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	337994447	2016.01.07	Perkampungan nelayan/lady diver	Merupakan tempat wis
12	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	336343987	2016.01.01	Tertutup kabut	Datang kesana awal a
13	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	328889791	2015.11.25	Luar biasa	Tempat yang indah ur
14	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	327983782	2015.11.20	must see!	Kita akan menemukan
15	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	327344699	2015.11.17	Lokasi wisata yang menarik di jeju	untuk melihat sunrise,
16	983296	제주도	오락거리	1582693	성산 일출봉	인도네시아어	319985056	2015.10.19	Sunrise Peak	Lokasi yang cocok unt

1	접근성	address	accost
2	접근성	address	address
3	접근성	address	call
4	접근성	address	cover
5	접근성	address	deal
6	접근성	address	destination
7	접근성	address	direct
8	접근성	address	handle
9	접근성	address	plow
10	접근성	address	reference
11	접근성	address	savoir-faire
12	접근성	address	speak
13	접근성	address	speech
14	접근성	address	treat



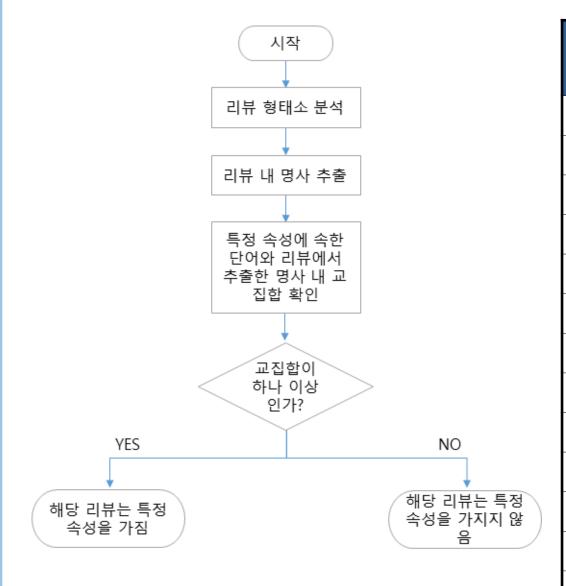
수집된 리뷰 데이터 및 속성사전을 DB에 업로드

• 과정 3 - 데이터 전처리 및 형태소 분석(POS) 과정

연구 소개

연구 과정

연구 결과



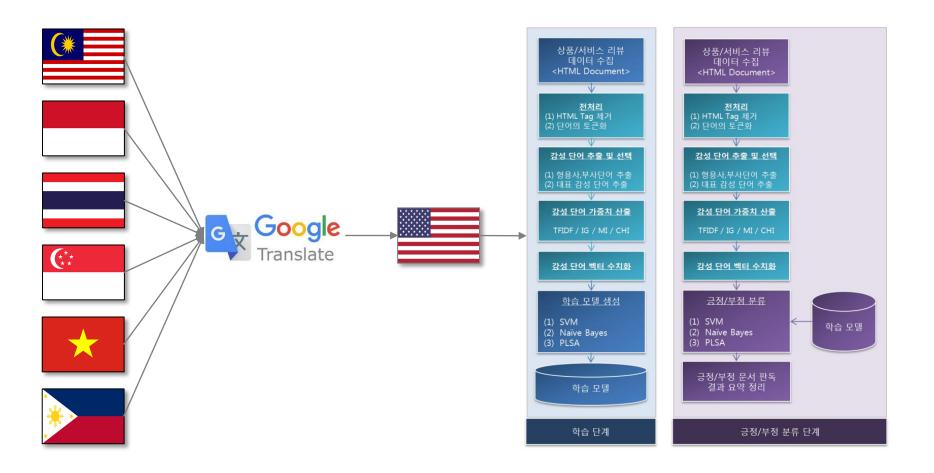
품사 태그	품사 설명
СС	coordinating conjunction
CD	cardinal digit
DT	determiner
EX	existential there
FW	foreign word
IN	preposition/subordinating conjunction
JJ	adjective
JJR	adjective, comparative
JJS	adjective, superlative
LS	list marker
MD	modal cloud
NN	noun
•••	※NLTK POS Tagger 참조 …

• 과정 4 - 감성분석

연구 소개

연구 과정

연구 결과



• 과정 4 - 감성분석 준비를 위한 속성 단어 여부 검토

연구 소개

연구 과정

연구 결과

review_id	accessibility	atmosphere	culture	experience	interest	kindness	price	attraction
566608135								0
551742829								0
503556039	0							0
465999045								
465999045				0				
465999045				0				
465999045								
465999045								
465999045								
465999045								
396049109	0							
396049109								
393491904								0
393491904								
393491904								
393491904								
368342189		0						0
368342189								0

연구 소개

연구 과정

연구 결과

연구 기여 방안

• 과정 4 - 감성분석 예시

- 문장 1 : Although it was allocated by tour two times, it is near from <u>airport</u>, convenience <u>store</u> is <u>convenient</u> near <u>hotel</u>, too.
- 속성: accessibility
- 감성:1(긍정)
- 문장 2: On the first day I went to a <u>shop</u>, it was cheap and very delicious.
- 속성: culture
- 감성:1(긍정)
- 문장 1 : The <u>hotel</u> was very kind and comfortable.
- 속성: kindness
- 감성:1(긍정)
- 각 리뷰가 가지고 있는 속성에 대하여 감성 분석을 한 결과를 아래와 같이 테이블에 저장함 각 칼럼 내 수치 값 1,0,-1은 각각 <u>긍정, 중립, 부정</u>을 의미함

• 과정 4 - 감성분석

연구 소개

연구 과정

연구 결과

accessibility	atmosphere	culture	experience	interest	kindness	price	attraction	etc	review_content
1	0	-1	0	0	0	1	1	1	it was a very cheap rate about 2,500 y
0	0	0	0	0	0	0	1	1	all the staff were able to stay kindly an
1	1	0	0	0	1	0	1	1	i was impressed by the luxury of the s
0	0	0	0	0	0	0	1	1	i descend a lot of stairs, arrive at the e
-1	1	0	0	0	1	-1	1	1	we used it on tour. although it is far fr
1	1	0	0	0	0	0	1	1	it is a big respectable hotel. because t
0	1	1	0	0	-1	0	1	1	it takes a while to coast. however, the
1	0	0	0	0	0	0	1	1	anyway it is a big, amusement image
0	1	0	0	0	1	0	1	1	it is introduced in guide books and sit
1	1	1	0	0	0	0	1	1	i stayed with my family for 5 days. in
1	0	0	0	0	0	1	1	1	if you are looking for a hotel in la da
1	0	1	0	0	0	0	1	1	i usually go to jeju island once, are no
1	0	-1	0	0	0	0	1	1	it is large and the number of people t
0	0	0	0	0	0	0	0	1	the lobby was large and the room was

연구 소개

연구 과정

연구 결과

연구 기여 방인

• 과정 4 - 리뷰 학습을 위한 LSTM 모델 설계

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 500, 32)	160000
conv1d_1 (Conv1D)	(None, 500, 32)	3104
max_pooling1d_1 (MaxPooling1	(None, 250, 32)	0
Istm_1 (LSTM)	(None, 100)	53200
dense_1 (Dense)	(None, 1)	101

Total params: 216,405 Trainable params: 216,405 Non-trainable params: 0

리뷰의 긍/부정을 예측할 LSTM 모델을 만들어 수집된 리뷰 및 사전들 등을 학습시켜준다.

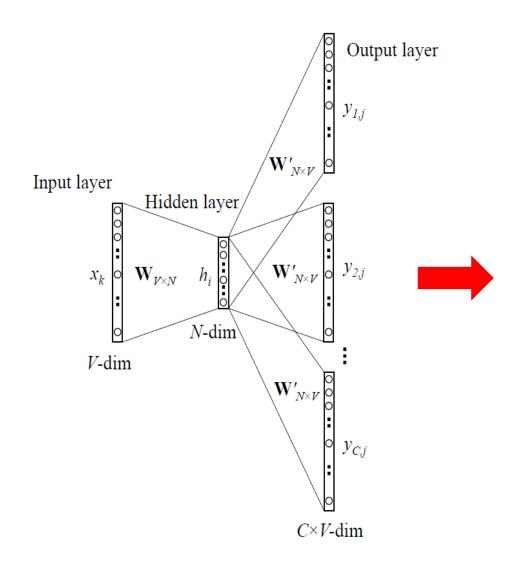
```
39968/40000 [==============].] - ETA: 0s - loss: 0.1515 - acc: 0.94
54Epoch 00004: val_loss improved from 0.34502 to 0.23646, saving model to ./model
7/04-0.2365.hdf5
c: 0.9454 - val_loss: 0.2365 - val_acc: 0.9068
Epoch 5/10
57Epoch 00005: val_loss did not improve
40000/40000 [======== ] - 493s 12ms/step - loss: 0.1248 - ac
c: 0.9557 - val_loss: 0.4895 - val_acc: 0.8361
Epoch 6/10
50Epoch 00006: val_loss did not improve
c: 0.9650 - val loss: 0.3232 - val acc: 0.8849
<keras.callbacks.History at 0x3e11b390>
```

• 과정 5 - 단어 간의 상관관계 시각화(skip-gram)

연구 소개

연구 과정

연구 결과



	source	target	value
0	store	convenience	85
1	stores	convenience	81
2	room	clean	60
3	shop	free	59
4	duty	free	54
5	problem	no	48
6	room	large	48
7	room	spacious	45
8	hotel	good	43
9	hotel	grand	36
10	location	good	36
11	time	long	35
12	sea	beautiful	34
13	view	good	32
14	scenery	beautiful	27
15	hotel	clean	24
16	hotel	convenient	23
17	black	pig	23
18	abalone	delicious	21
19	breakfast	delicious	21

• 결과 - User 맞춤 리뷰 분석 시스템 개발

연구 소개

연구 과정

연구 결과

연구 기여 방안

분석 프로세스 관리 시스템

분석 프로세스 상세

프로세스 목표	관광 리뷰 분석		
총 프로세스 개수	4 개	분석 중 프로세스 개수	0 개

분석 프로세스 목록

* 프로세스 상태 변경 후 새로고침을 통해 실시간 프로세스 단계를 확인할 수 있습니다

번호	분석 내용	프로세스 상태	* 프로세스 단계(진행률)	동작 관리
1	리뷰 데이터 전체 감성 분석	분석 멈춤	-	상태 변경 ▼
2	리뷰 속성별 감성 분석	분석 멈춤	-	상태 변경 ▼
3	리뷰 키워드 분석	분석 멈춤	-	상태 변경 ▼
4	리뷰 상관관계 분석	분석 멈춤	-	상태 변경 ▼

• 결과 - User 맞춤 리뷰 분석 시스템 개발

연구 소개

연구 과정

연구 결과

연구 기여 방안

※ 분석 프로세스 단계

분석 내용	단계	설명
리뷰 데이터 전체 감성 분석	1단계	분석 대상 데이터 준비
	2단계	감성 분석 및 DB 저장
리뷰 속성별 감성 분석	1단계	속성 사전 및 분석 대상 데이터 준비
	2단계	감성 분석
	3단계	분석 결과 처리 및 DB 저장
리뷰 속성별 키워드 분석	1단계	분석 대상 데이터 준비
	2단계	키워드 분석 및 결과 파일 저장
리뷰 속성별 상관관계 분석	1단계	분석 대상 데이터 준비
	2단계	상관관계 분석 및 결과 파일 저장

연구 소개

연구 과정

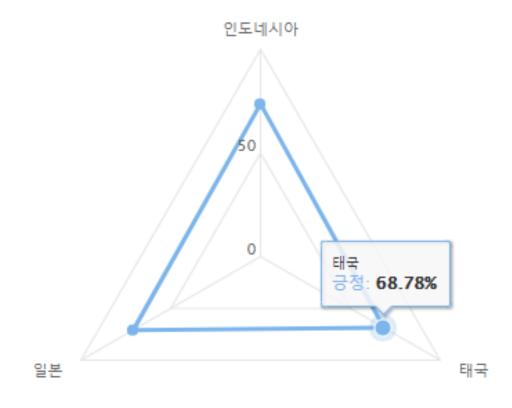
연구 결과

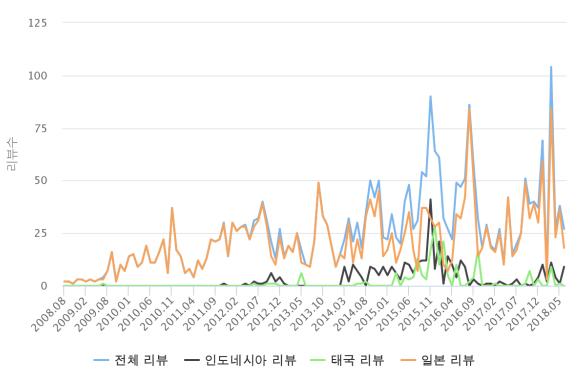
시각화 내용	시각화 설명	
국가별 긍정 리뷰 비율	관광객의 국적 별 긍정적으로 작성된 리뷰의 비율 시각화	
전체 리뷰 추이	관광객의 국적 별 리뷰 작성 추이 시각화	
속성별 대표 긍정/부정 리뷰	8가지 속성별 대표 긍정 및 부정 리뷰 시각화	
속성별 긍정 리뷰 비율	8가지 속성별 긍정 리뷰 비율 시각화	
속성별 워드 클라우드	8가지 속성별 중요 키워드 워드 클라우드 시각화	
속성별 키워드 네트워크	8가지 속성별 키워드 쌍 네트워크 시각화	

연구 소개

연구 과정

연구 결과





각 나라별 긍/부정도

각 나라별/시기별 리뷰 변동 추이

연구 소개

연구과정

연구 결과

연구 기여 방안

리뷰별 긍/부정 키워드 및 속성단어 색인

긍정 리뷰

The **location** is quite **attractive** and **beautiful**, there are all forms of teddy bear dolls.

Indonesian | 2016.07.17

부정 리뷰

Although it is located in the eastern end of the phoenix resort, there is a shuttle bus in the area, but **access** is **bad**, so you should take time with extra time.

Japanese | 2016.01.03

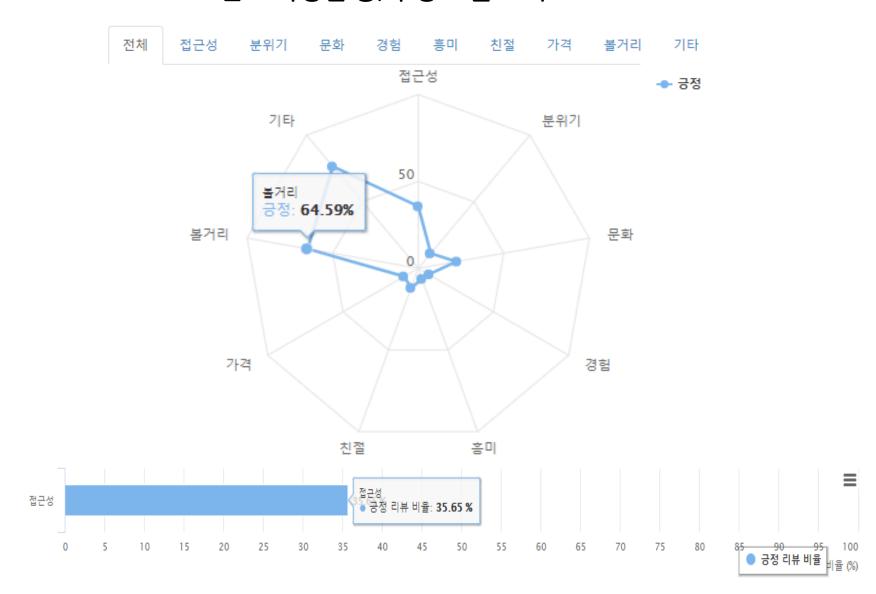
연구 소개

연구 과정

연구 결과

연구 기여 방안

전체 속성별 긍/부정 비율 시각화



연구 소개

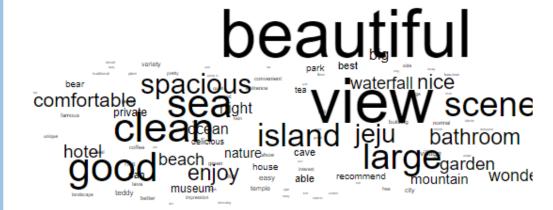
연구 과정

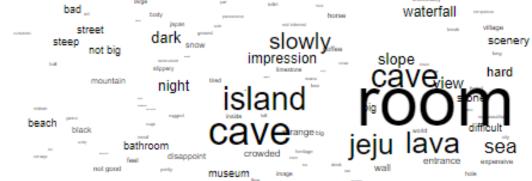
연구 결과

연구 기여 방안

속성별 긍/부정 워드 클라우드 시각화

전체 접근성 분위기 문화 경험 흥미 친절 가격 볼거리 기타





긍정 워드 클라우드

부정 워드 클라우드

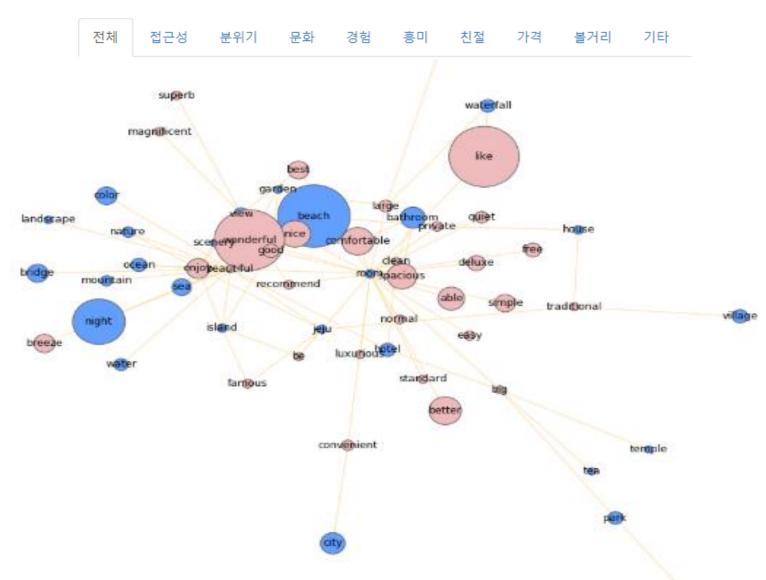
속성별 키워드 네트워크 시각화

연구 소개

연구과정

연구 결과

연구 기여 방안



amusement

• 결과 – 데이터 시각화 예시(제주도 지역별)

연구 소개

연구 과정

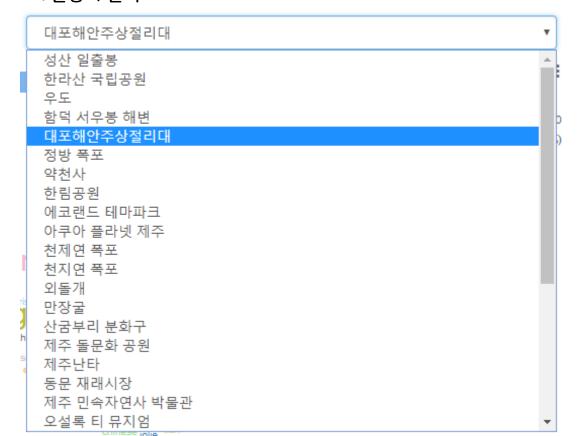
연구 결과

연구 기여 방안

1. 액티비티 선택



2. 관광지 선택

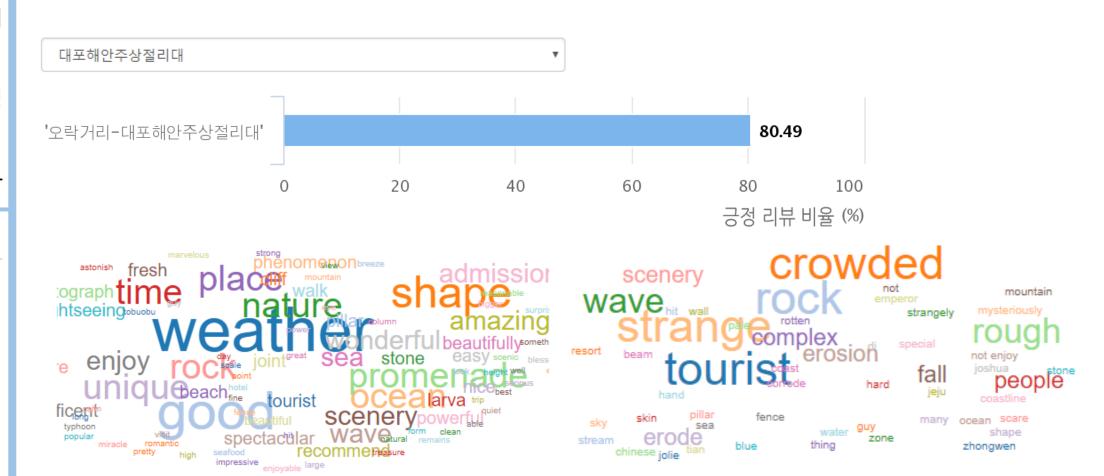


• 결과 - 데이터 시각화 예시(제주도 관광지별)

연구 소개

연구 과정

연구 결과



관광지별 긍정 리뷰 비율 및 긍/부정 워드 클라우드

연구 소개

연구 과정

연구 결과

연구 기여 방안

• 연구 기여 방안

- 1. User(제주 관광 마케터)의 필요에 맞는 맞춤 정보들을 제공함으로써 향후 고객들의 만족 및 불만족 사항에 대한 이해를 돕는다.
- 2. 실시간 수집 및 분석을 통하여 변화하는 트랜드에 빠르게 대응 및 보완이 가능하다.
- 3. 향후, 동남아 관광객 이외에도 다른 나라 관광객들을 위한 관광 마케팅 개발 및 연구가 가능하다.
- 4. 다양한 카테고리별/다각적 분석을 통해 전략적 마케팅을 가능하게 한다.

Q & A

감사합니다