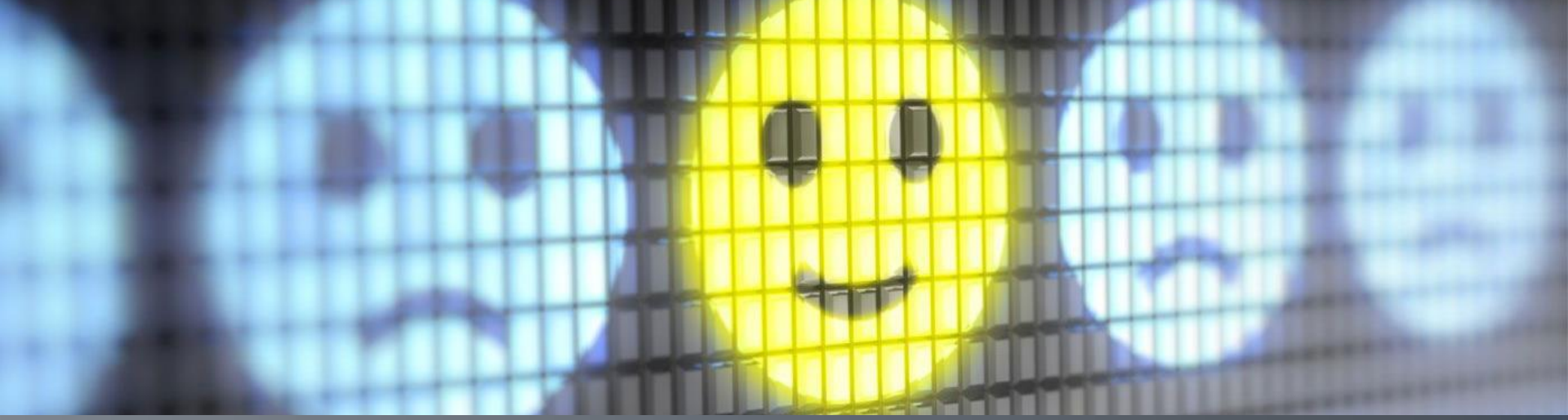


November 16, 2016

Method for Construction of Spatial Sentiment Lexicon Using Place Reviews

Case Study on Theme Parks

Youngmin Lee

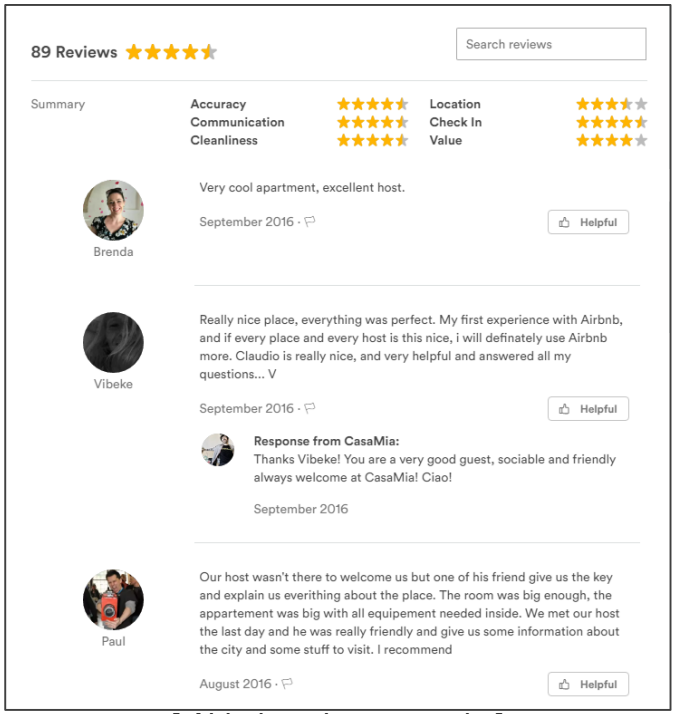
GIS/LBS LAB
Seoul National University

CONTENTS

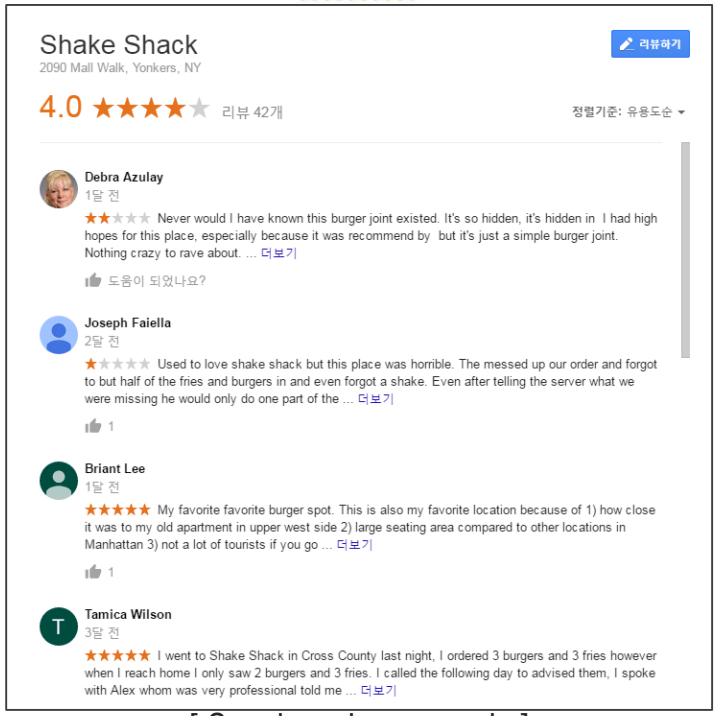
Background and purpose of the study

The trend of these days

- The use of location-based services(LBS) is constantly increasing. and, reviewing and grading places through the use of LBS has also become a common practice among users.



[Airbnb review example]



[Google review example]

- Previous reviews of a place can significantly affect their potential visitors.

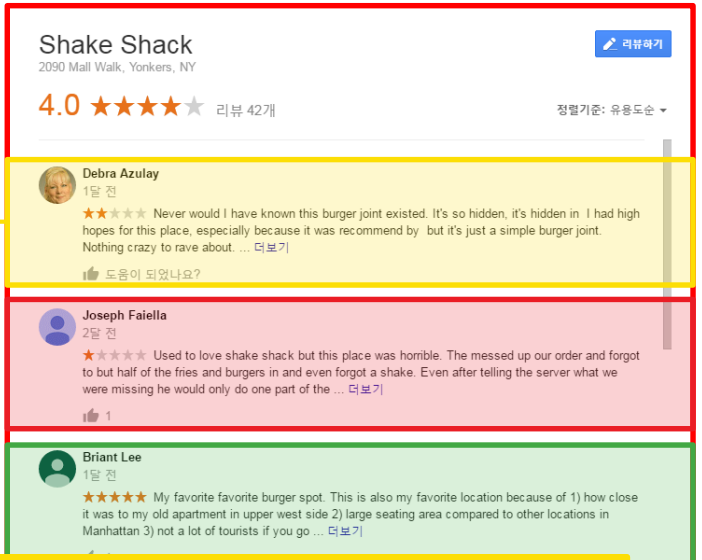
Background and purpose of the study

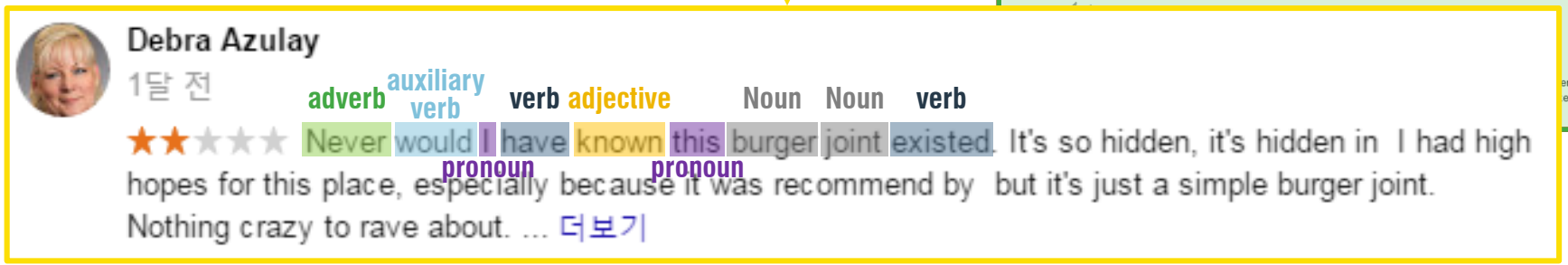
Visitor's sentiment expression

- A user-created review is the result of a visitor's actual positive or negative sentiment expression, and the sentiment could be expressed as a positive, negative, or neutral opinion.

- In order to perform sentiment analysis, each word should be separated by its POS (part-of-speech) through natural language processing.

- To do this, a database of spatial sentiment words should be constructed.



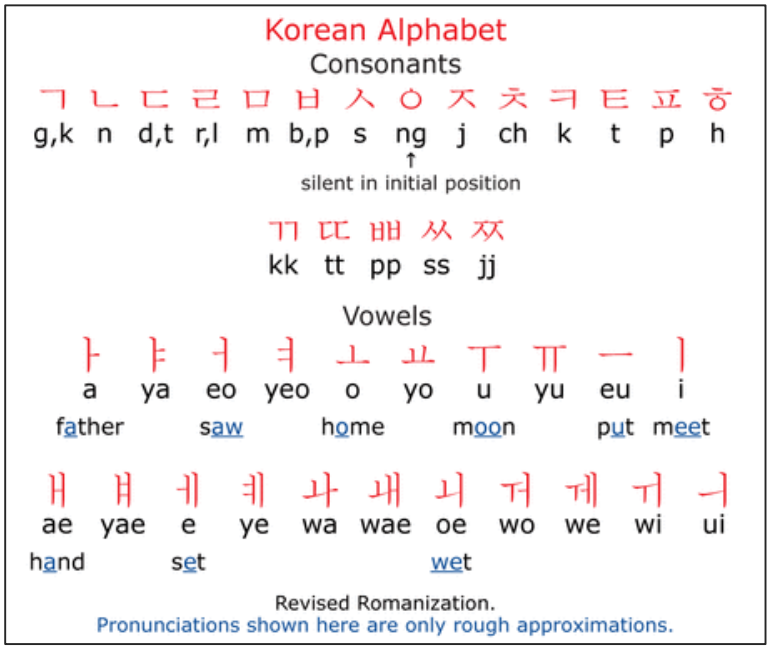


Background and purpose of the study

There is no spatial sentiment lexicon

- In the past, sentiment analysis has been used mainly for product reviews (Chang 2009, Hu & Liu 2004, Myung et al 2008, Scaffidi et al). No spatial sentiment lexicon for sentiment analysis of places has been constructed yet.

- Specifically, Korean lacks in terms of research on sentiment word analysis, compared to English. Korean, unlike English, has complicated characteristics and is composed of complex adjectives and suffixes (Jang et al. 2015).



[Korean(Hangeul) alphabet]

- Chang J (2009) A sentiment Analysis Algorithm for Automatic Product Reviews Classification in On-Line Shopping Mall. 14(4):19-33
- Hu M, & Liu B (2004) Mining and Summarizing Customer Reviews. In proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 168-177
- Myung J, Lee D, & Lee S (2008) A Korean Product Review Analysis System Using a Semi-Automatically Constructed Semantic Dictionary. Journal of KIISE:Software and Applications. 35(6): 392-403
- Scaffidi C, Bierhoff K, Chang E, Felker M, Ng H, & Jin C (2007) Red Opal: Product-Feature Scoring from Reviews. In proceedings of the 8th ACM conference on Electronic commerce. 182-191
- Jang K, Park S, & Kim W (2015) Automatic Construction of a Negative/positive Corpus and Emotional Classification using the Internet Emotional Sign. Journal of KIISE. 42(4):512-521

Background and purpose of the study

The purpose of the study

- Therefore, in this study, we propose a method to construct a spatial sentiment lexicon using place reviews written in Korean, and we focused primarily on a 'theme parks' out of the many possible place categories.

- Other types of spatial sentiment lexicon could be constructed using the same methodology.

How to construct spatial sentiment lexicon

Which are sentiment words?

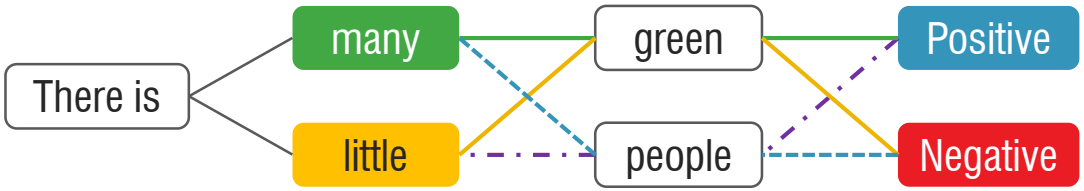
How to construct spatial sentiment lexicon

3 aspects to consider

1) The polarity of the sentiment word is analyzed.

- To do this, we calculated sentiment polarity and probability using the results of the survey.

2) It should be taken into account that some sentiment words are associated with properties of a place.

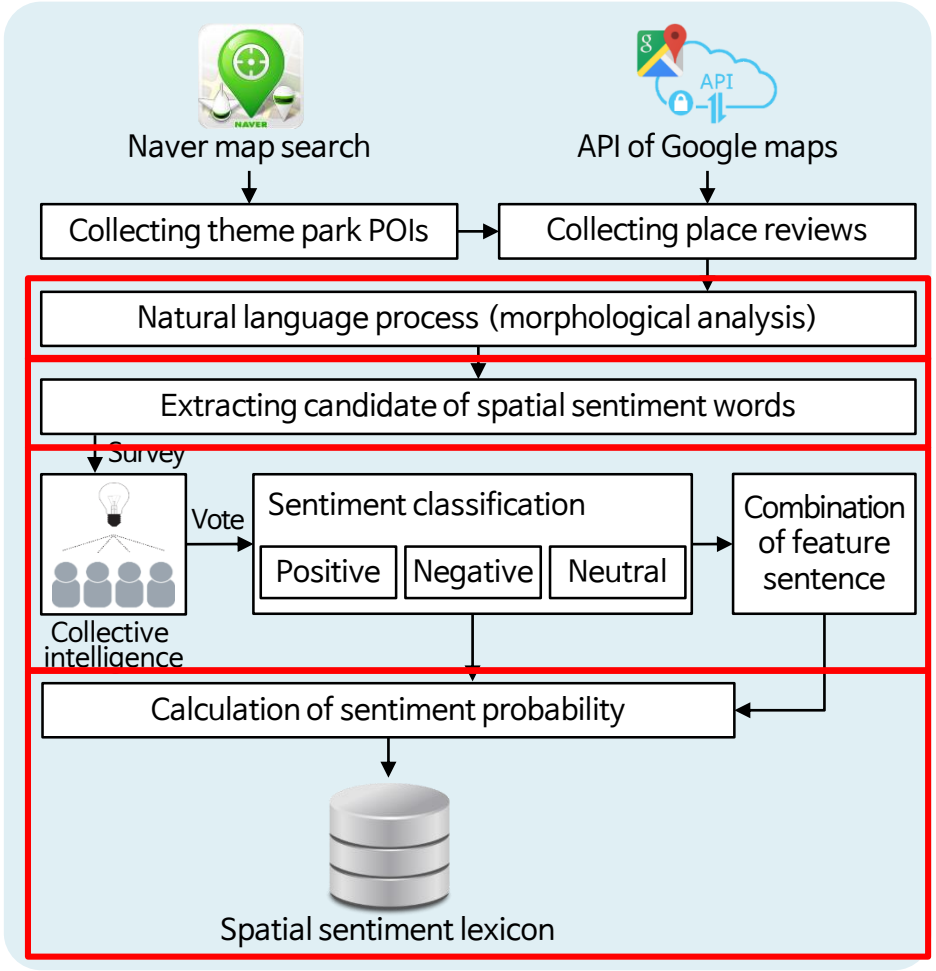


3) The spatial features and predicate characteristics of a place should be classified.

- We defined the combination of the spatial feature and the predicate as the 'spatial feature sentence'.

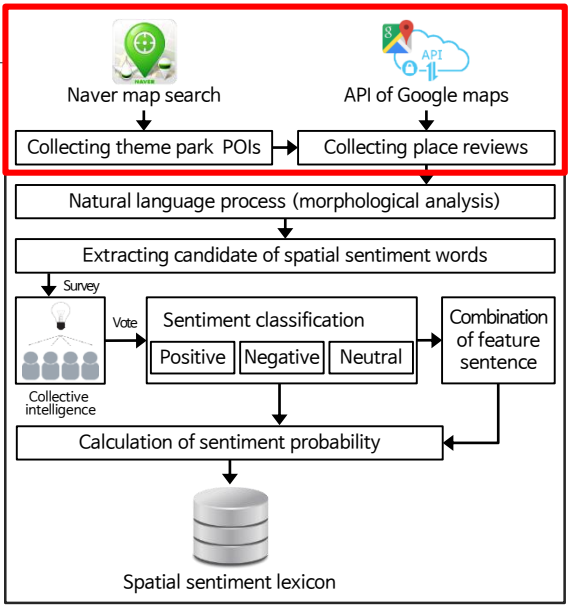
How to construct spatial sentiment lexicon

Research workflow



[Research workflow]

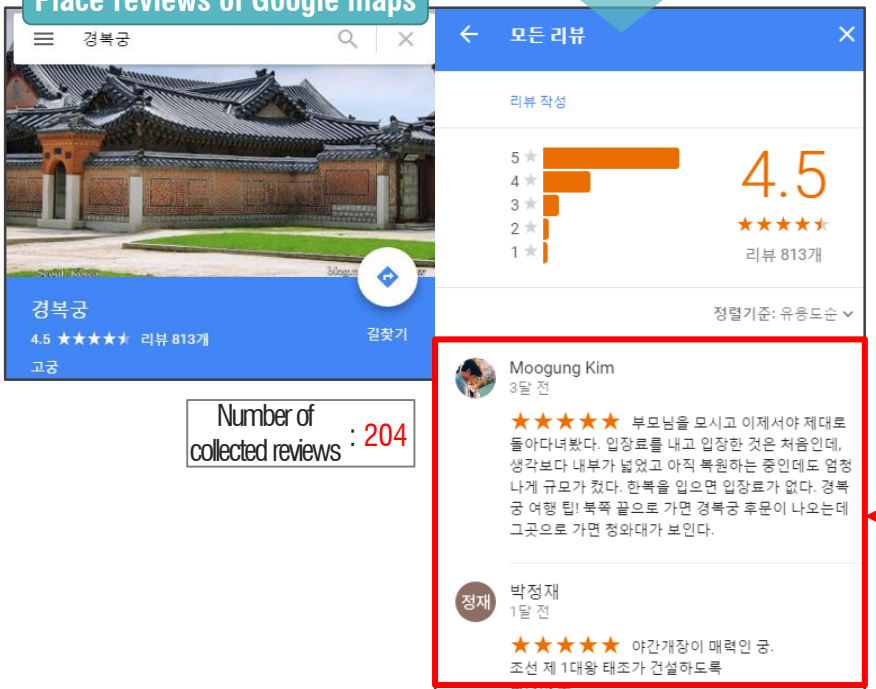
Experiment and result



Naver map searching

Theme park POIs

Place reviews of Google maps

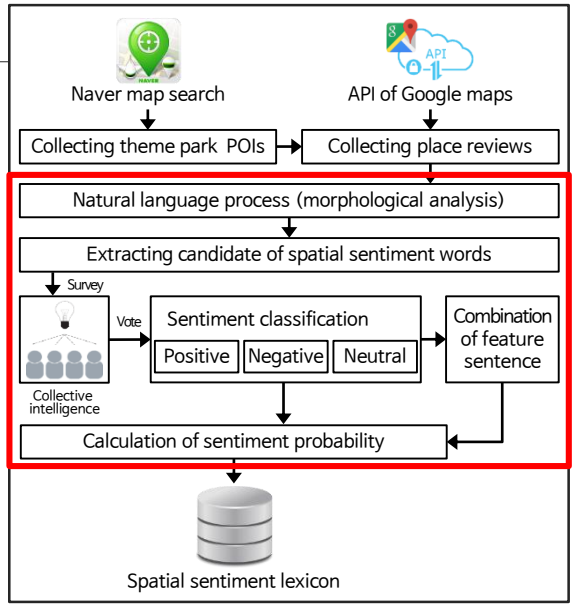


Number of collected reviews : 204

Collecting reviews using Google API using java script

```
<html>
  <head>
    <title>Place details</title>
    <meta name="viewport" content="initial-scale=1.0, user-scalable=no">
    <meta charset="utf-8">
    <style>
      html, body {
        height: 100%;
        margin: 0;
        padding: 0;
      }
      #map {
        height: 100%;
      }
    </style>
    <script src="http://code.jquery.com/jquery-latest.min.js"></script>    <script
src="https://maps.googleapis.com/maps/api/js?key=AIzaSyCwiMiugiEcSZF6jYtrvkeQC89Od06hgmo
&signed_in=true&libraries=places&callback=initMap" async defer></script>
    <script>
    placeIdArr = ['ChIJKZ9yGYyjfDURTts7CliwJoY',
'ChIJE0i-X0eZfDURGCuFxA4UFeo',
'ChIJ9TviAqadfDURzDMmyLGP0u8',
'ChIJt5HKlueifDURHONSIuFp_3A',
'ChIJa5tZTFyZfDURJ4smvd436Fk',
'ChIJhWv921iifDUReKwDFyWTVOg'];

    console.log(placeIdArr.length);
function initMap() {
  var map = new google.maps.Map(document.getElementById('map'), {
    center: {lat: -33.866, lng: 151.196},
    zoom: 15
```

Experiment and result

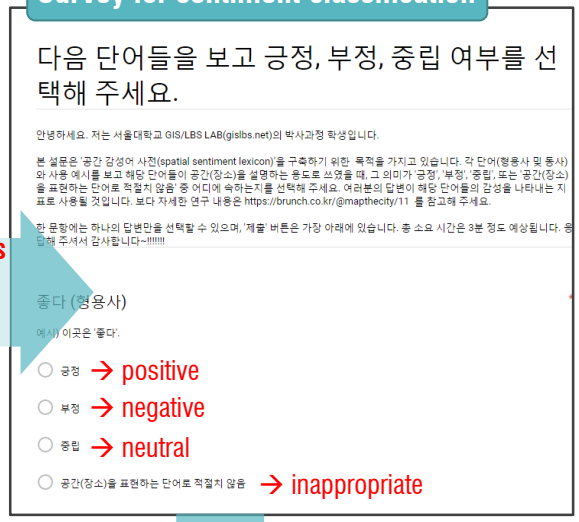


Extracting candidates of spatial sentiment words through NLP using R

```
library(KoNLP)
library(stringr)
library(tm)

setwd('C:/YM/02. Project/2016_여대학(원)생 공학연구팀제 사업/내연구')

useSejongDic()

d = scan(file = 'seoultour.txt', what='str', sep='\t', encoding='UTF-8')
d <- unlist(d)

d_01 <- NULL
for (i in seq(length(d))){
  if (nchar(d[i]) > 1){
  d_01 <- c(d_01, SimplePos22(d[i]))
  }
}

d_02 <- unlist(d_01)
class(d_02)
d_02 <- unlist(strsplit(d_02, "\\+")) # + 기준으로 문자열 나누기
d_02 <- unlist(strsplit(d_02, " ")) # + 기준으로 문자열 나누기

NC <- unlist(d_02[grep("NC$", d_02)]) # 보통명사(NC)만 추출
NQ <- unlist(d_02[grep("NQ$", d_02)]) # 고유명사(NQ)만 추출
PV <- unlist(d_02[grep("PV$", d_02)]) # 동사(PV)만 추출
PX <- unlist(d_02[grep("PX$", d_02)]) # 보조용언(PX)만 추출
PA <- unlist(d_02[grep("PA$", d_02)]) # 형용사(PA)만 추출
MA <- unlist(d_02[grep("MV$", d_02)]) # 부사(MA)만 추출

# lematization
NC_k <- gsub("/NC", "", NC)
NQ_k <- gsub("/NQ", "", NQ)
PV_k <- gsub("/PV", "", PV)
PX_k <- gsub("/PX", "", PX)
PA_k <- gsub("/PA", "", PA)
MA_k <- gsub("/MA", "", MA)

# 음절이 2개 이상인 것만 추출
NC_kk <- NULL
for (i in seq(length(NC_k))){
  if (nchar(NC_k[i]) > 1){
    NC_kk<- c(NC_kk, NC_k[i])
    # print (NC_kk)
  }
}

PV_kk <- gsub("/PV", "다", PV)
PX_kk <- gsub("/PX", "다", PX)
PA_kk <- gsub("/PA", "다", PA)

write.table(table(NC_kk), 'NC_kk.txt', sep="\t")
write.table(table(PV_kk), 'PV_kk.txt', sep="\t")
write.table(table(PX_kk), 'PX_kk.txt', sep="\t")
write.table(table(PA_kk), 'PA_kk.txt', sep="\t")
```

Candidates of spatial sentiment words

Survey for sentiment classification



Survey result

Sentiment polarity and probability calculation using R

```
# Step 1: inappropriate
if ((count(data[inappropriate]) > (count(data[positive]) &&
    count(data[inappropriate]) > (count(data[negative]) &&
    count(data[inappropriate]) > (count(data[neutral])){
            Score = 100 * count(data[inappropriate]) / count(data[positive]) + count(data[negative]
                    ) + count(data[neutral]) + count(data[inappropriate])
            Sentiment = 'inappropriate'}
# Step 2: neutral_1
else if ((count(data[neutral]) > (count(data[positive]) &&
    count(data[neutral]) > (count(data[negative])){
            Score = 100 * count(data[neutral]) / count(data[positive]) + count(data[negative]) + co
                    unt(data[neutral])
            Sentiment = 'neutral'}
# Step 3: positive & negative
else if ((count(data[positive]) > count(data[negative])){
            Score = 100 * count(data[positive]) / count(data[positive]) + count(data[negative])
            Sentiment = 'positive'}
else if ((count(data[negative]) > count(data[positive])){
            Score = 100 * count(data[negative]) / count(data[positive]) + count(data[negative])
            Sentiment = 'negative'}
# Step 4: when positive = negative = neutral
else
            Score = 100
            Sentiment = 'neutral'
```

Experiment and result

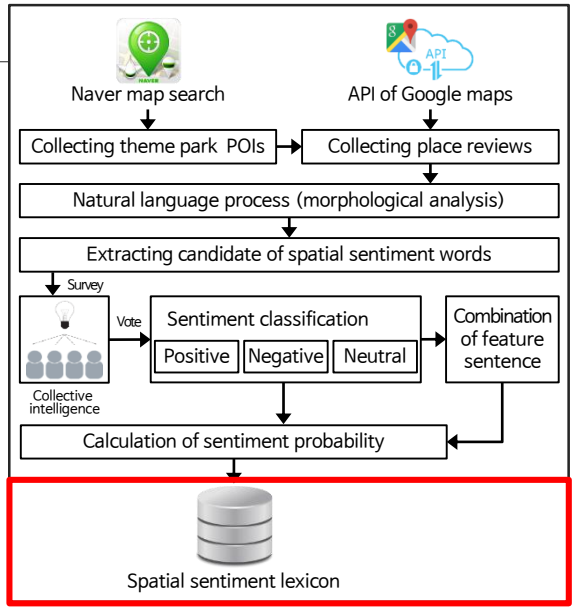
The result of spatial sentiment lexicon_01 (single adjective, verb, and noun)

- PA: Adjective, PV: Verb, NC: Noun
- Positive: 1, Negative: -1, Neutral: 0

No.	Sentiment words		Polarity	Probability
	Korean	English		
1	추천/NC	Recommendation	1	1.000
2	무료/NC	Free	1	1.000
3	굿/NC	Good	1	1.000
4	최고/NC	Best	1	1.000
5	만족/NC	Satisfaction	1	1.000
6	감탄/NC	Admiration	1	1.000
7	강추/NC	Strong recommendation	1	1.000
8	짱/NC	Super	1	1.000
9	최상/NC	The highest	1	1.000
10	최적/NC	Optimum	1	1.000
11	새롭/PA	New	1	1.000
12	소중/NC	Precious	1	1.000
13	따뜻/NC	Warm	1	1.000
14	수려/NC	Graceful	1	1.000
15	신기/NC	Amazing	1	1.000
16	아기자기/NC	Cute	1	1.000
17	아늑/NC	Cozy	1	1.000
18	안전/NC	Safety	1	1.000
19	친절/NC	Kind	1	1.000
20	쾌적/NC	Pleasant	1	1.000
21	특별/NC	Special	1	1.000
22	포근/NC	Snug	1	1.000
23	산책/NC	Walk	1	1.000
24	놀/PA	Play	1	1.000
25	나들이/NC	Trip	1	1.000
26	운동/NC	Exercise	1	1.000
27	트이/PV	Open	1	1.000
28	좋/PA	Nice	1	1.000
29	아름답/PA	Beautiful	1	1.000
30	가깝/PA	Near	1	1.000
31	편하/PA	Comfortable	1	1.000
32	맑/PA	Pure	1	1.000
33	예쁘/PA	Beautiful	1	1.000
34	멋지/PA	Wonderful	1	1.000

No.	Sentiment words		Polarity	Probability
	Korean	English		
35	새롭/PA	New	1	1.000
36	뛰어나/PA	Excellent	1	1.000
37	즐겁/PA	Pleasant	1	1.000
38	시원시원하/PA	Straightforward	1	1.000
39	알차/PA	Fruitful	1	1.000
40	[illegible]/PV	Funny	1	1.000
41	[illegible]	[illegible]	1	1.000
42	즐기/PV	Enjoy	1	1.000
43	쉬/NC	Rest	1	1.000
44	손꼽/PV	Look forward	1	1.000
45	신나/NC	Excited	1	1.000
46	멋있/PA	Nice	1	0.993
47	상쾌/NC	Fresh	1	0.993
48	활기차/PA	Active	1	0.992
49	빼어나/PA	Outstanding	1	0.992
50	엄청나/PA	Great	1	0.990
51	밝/PA	Bright	1	0.989
52	넓/PA	Large	1	0.988
53	배우/NC	Learn	1	0.987
54	괜찮/PA	Fine	1	0.984
55	색다르/PA	Different	1	0.981
56	대박/NC	Great	1	0.980
57	힐링하/NC	Heal	1	0.980
58	노닐/PV	Stroll	1	0.979
59	괜찮/PA	Good	1	0.979
60	천국/NC	Heaven	1	0.978
61	낫/PA	Better	1	0.976
62	귀하/NC	Precious	1	0.974
63	인정/NC	Admit	1	0.970
64	적당/NC	Proper	1	0.967
65	아담/NC	Neat	1	0.966
66	취하/PV	Drunk	1	0.964
67	싸/PA	Cheap	1	0.896
68	그럴듯하/PA	Plausible	1	0.840

100% positive words



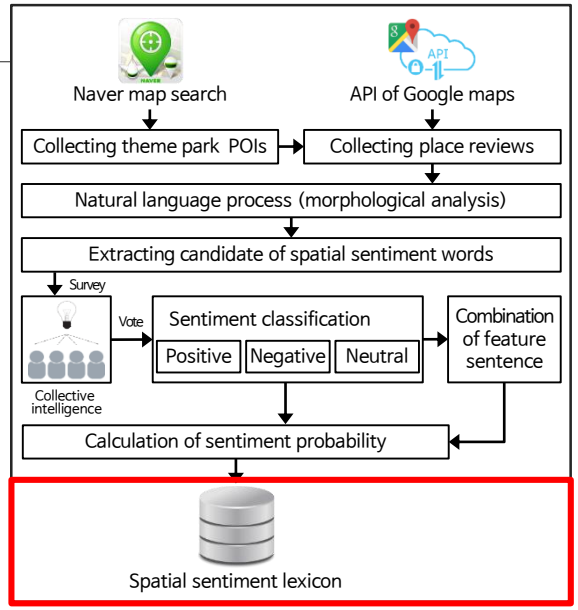
Experiment and result

The result of spatial sentiment lexicon_01 (single adjective, verb, and noun)

- PA: Adjective, PV: Verb, NC: Noun
- Positive: 1, Negative: -1, Neutral: 0

No.	Sentiment words		Polarity	Probability
	Korean	English		
69	실망/NC	Disappointment	-1	1.000
70	괴롭/PA	Painful	-1	1.000
71	딱딱하/PA	Hard	-1	1.000
72	불편/NC	Inconvenience	-1	1.000
73	삭막/NC	Desolate	-1	1.000
74	심심/NC	Bored	-1	1.000
75	싸늘/NC	Cold	-1	1.000
76	싫/PA	Hate	-1	1.000
77	힘들/PV	Hard	-1	1.000
78	비좁/PA	Small	-1	0.993
79	나쁘/PA	Bad	-1	0.992
80	무섭/PA	Scary	-1	0.992
81	아쉽/PA	Sad	-1	0.990
82	낡/PA	Old	-1	0.990
83	좁/PA	Narrow	-1	0.981
84	답답/NC	Stuffy	-1	0.980
85	번잡/NC	Complexity	-1	0.979
86	멀/PA	Distant	-1	0.978
87	괴상/NC	Strange	-1	0.977
88	시끄럽/PA	Noisy	-1	0.976
89	복잡/NC	Complicate	-1	0.974
90	습하/PA	Damp	-1	0.974
91	비싸/PA	Expensive	-1	0.968
92	최악/NC	The worst	-1	0.962
93	묘하/PA	Odd	0	0.831
94	높/PA	High	0	0.805
95	다르/PA	Different	0	0.802
96	느끼/PV	Feel	0	0.794
97	뛰/PV	Run	0	0.766
98	유료/NC	Paid	0	0.765
99	낯익/PA	Familiar	0	0.745
100	평범/NC	Ordinary	0	0.731
101	바라보/PV	Look	0	0.674
102	걷/PV	Walk	0	0.660

No.	Sentiment words		Polarity	Probability
	Korean	English		
103	촬영/NC	Shooting	0	0.653
104	돌아다니/PV	Go around	0	0.640
105	활동/NC	Activity	0	0.633
106	만나/NC	Meet	0	0.612
107	모이/NC	Gather	0	0.600
108	무난/NC	Easy	0	0.588
109	작/PA	Small	0	0.567
110	조용/NC	Silent	0	0.558
111	크/PA	Big	0	0.543
112	놀랍/PA	Surprising	0	0.519
113	북적대/PV	Crowded	0	0.480
114	붐비/PA	Crowded	0	0.477



Words determined by 'inappropriate'

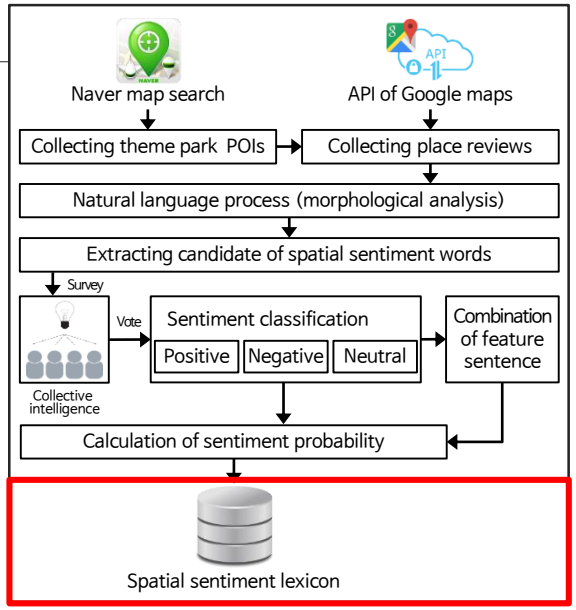
No.	Sentiment words		Polarity	Probability
	Korean	English		
1	쉽/PA	Easy	inappropriate	0.698
2	가볍/PA	Light	inappropriate	0.691
3	빠르/PA	Fast	inappropriate	0.664
4	고맙/PA	Thank	inappropriate	0.631
5	맛있/PA	Delicious	inappropriate	0.597
6	어렵/PA	Difficult	inappropriate	0.530
7	안타깝/PA	Regrettable	inappropriate	0.530
8	젊/PA	Young	inappropriate	0.497
9	슬프/PA	Sad	inappropriate	0.463
10	웬만하/PA	Fair	inappropriate	0.456
11	만지/PV	Touch	inappropriate	0.635

Experiment and result

The result of spatial sentiment lexicon_02 (combination with 'many' and 'little')

- PA: Adjective, PV: Verb, NC: Noun
- Positive: 1, Negative: -1, Neutral: 0

No.	Sentiment words_subject		Sentiment words_predicate		Polarity	Probability
	Korean	English	Korean	English		
1	구경거리/NC	Spectacle	많/PA	Many	1	1.000
2	주차장/NC	Parking lot	많/PA	Many	1	1.000
3	쉼터/NC	Rest area	많/PA	Many	1	1.000
4	음식점/NC	Restaurant	많/PA	Many	1	1.000
5	운동시설/NC	Sports facilities	많/PA	Many	1	1.000
6	문화재/NC	Cultural assets	많/PA	Many	1	1.000
7	산책로/NC	Walk	많/PA	Many	1	1.000
8	나무/NC	Tree	많/PA	Many	1	0.989
9	편의시설/NC	Amenities	많/PA	Many	1	0.979
10	꽃/NC	Flower	많/PA	Many	1	0.976
11	오르막/NC	Ascent	적/PA	Little	1	0.976
12	차/NB	Car	적/PA	Little	1	0.890
13	행사/NC	Event	많/PA	Many	1	0.871
14	구경거리/NC	Spectacle	적/PA	Little	-1	1.000
15	주차장/NC	Parking lot	적/PA	Little	-1	1.000
16	쉼터/NC	Rest area	적/PA	Little	-1	1.000
17	음식점/NC	Restaurant	적/PA	Little	-1	1.000
18	운동시설/NC	Sports facilities	적/PA	Little	-1	1.000
19	문화재/NC	Cultural assets	적/PA	Little	-1	1.000
20	산책로/NC	Walk	적/PA	Little	-1	1.000
21	나무/NC	Tree	적/PA	Little	-1	0.989
22	편의시설/NC	Amenities	적/PA	Little	-1	0.979
23	오르막/NC	Uphill road	많/PA	Many	-1	0.976
24	꽃/NC	Flower	적/PA	Little	-1	0.976
25	차/NB	Car	많/PA	Many	-1	0.890
26	행사/NC	Event	적/PA	Little	-1	0.871
27	건물/NC	Building	많/PA	Many	0	0.764
28	건물/NC	Building	적/PA	Little	0	0.764
29	학생/NC	Student	많/PA	Many	0	0.750
30	학생/NC	Student	적/PA	Little	0	0.750
31	상인/NC	Merchant	많/PA	Many	0	0.728
32	상인/NC	Merchant	적/PA	Little	0	0.728
33	사람/NC	People	많/PA	Many	0	0.685
34	사람/NC	People	적/PA	Little	0	0.685



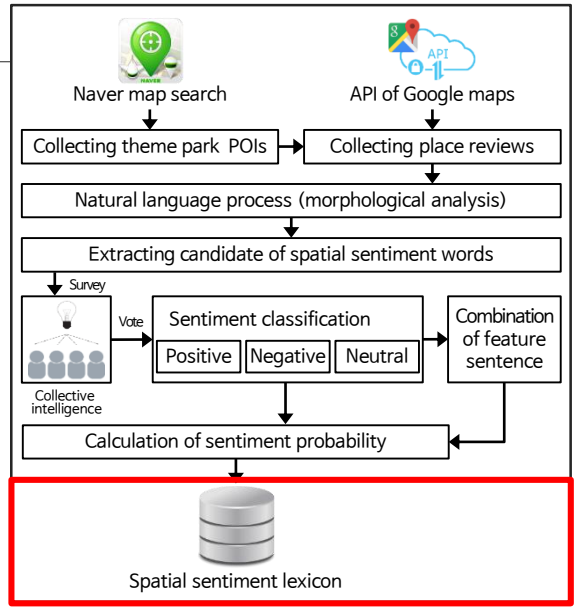
No.	Sentiment words_subject		Sentiment words_predicate		Polarity	Probability
	Korean	English	Korean	English		
35	어린이/NC	Child	많/PA	Many	0	0.653
36	어린이/NC	Child	적/PA	Little	0	0.653
37	가족/NC	Family	많/PA	Many	0	0.652
38	가족/NC	Family	적/PA	Little	0	0.652
39	아이들/NC	Children	많/PA	Many	0	0.651
40	아이들/NC	Children	적/PA	Little	0	0.651
41	상점/NC	Store	많/PA	Many	0	0.646
42	상점/NC	Store	적/PA	Little	0	0.646
43	연인/NC	Lover	많/PA	Many	0	0.611
44	연인/NC	Lover	적/PA	Little	0	0.611
45	언덕/NC	Slope	많/PA	Many	0	0.531
46	언덕/NC	Slope	적/PA	Little	0	0.531
47	관광객/NC	Tourist	많/PA	Many	0	0.431
48	관광객/NC	Tourist	적/PA	Little	0	0.431

Experiment and result

The result of spatial sentiment lexicon_03 (combination of feature sentence)

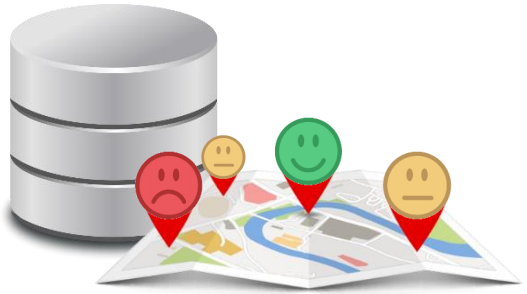
- PA: Adjective, PV: Verb, NC: Noun
- Positive: 1, Negative: -1, Neutral: 0

No.	Sentiment words_subject		Sentiment words_predicate		Polarity	Probability
	Korean	English	Korean	English		
1	역/NC	Station	가깝/PA	Close	1	1.000
2	탑/NC	Tower	멋지/PA	Wonderful	1	1.000
3	밤/NC	Night	아름답/PA	Beautiful	1	1.000
4	서울에서/NC	In Seoul	아름답/PA	Beautiful	1	1.000
5	야경/NC	Night view	아름답/PA	Beautiful	1	1.000
6	갈대/NC	Reed	예쁘/PA	Pretty	1	1.000
7	단풍/NC	Fall foliage	예쁘/PA	Pretty	1	1.000
8	야경/NC	Night view	예쁘/PA	Pretty	1	1.000
9	가족/NC	Family	좋/PA	Good	1	1.000
10	경치/NC	Scenery	좋/PA	Good	1	1.000
11	공원/NC	Park	좋/PA	Good	1	1.000
12	나들이/NC	Trip	좋/PA	Good	1	1.000
13	나들이하기/NC	Outing	좋/PA	Good	1	1.000
14	냄새/NC	Scent	좋/PA	Good	1	1.000
15	농구하기/NC	Playing basketball	좋/PA	Good	1	1.000
16	데이트하기/NC	Dating	좋/PA	Good	1	1.000
17	드라이브하기/NC	Driving	좋/PA	Good	1	1.000
18	벤치/NC	Bench	좋/PA	Good	1	1.000
19	분위기/NC	Atmosphere	좋/PA	Good	1	1.000
20	사진/NC	Photo	좋/PA	Good	1	1.000
21	산책/NC	Walk	좋/PA	Good	1	1.000
22	산책하기/NC	Walking	좋/PA	Good	1	1.000
23	시설/NC	Facility	좋/PA	Good	1	1.000
24	아이들/NC	Children	좋/PA	Good	1	1.000
25	야경/NC	Night view	좋/PA	Good	1	1.000
26	여유/NC	Spare	좋/PA	Good	1	1.000
27	운동하기/NC	Exercising	좋/PA	Good	1	1.000
28	이용/NC	Use	좋/PA	Good	1	1.000
29	일광욕하기/NC	Sunbathing	좋/PA	Good	1	1.000
30	일몰때/NC	Sunset	좋/PA	Good	1	1.000
31	저녁/NC	Evening	좋/PA	Good	1	1.000
32	전망/NC	View	좋/PA	Good	1	1.000
33	접근성/NC	Accessibility	좋/PA	Good	1	1.000
34	주차시설/NC	Parking lot	좋/PA	Good	1	1.000



No.	Sentiment words_subject		Sentiment words_predicate		Polarity	Probability
	Korean	English	Korean	English		
35	편의시설/NC	Amenities	좋/PA	Good	1	1.000
36	휴식/NC	Rest	좋/PA	Good	1	1.000
37	힐링하기/NC	Healing	좋/PA	Good	1	1.000
38	여유/NC	Spare	즐기/PV	Ejoy	1	1.000
39	시설/NC	Facility	편하/PA	Comfortable	1	1.000
40	장소/NC	Place	넓/PA	Large	1	0.988
41	주차장/NC	Parking lot	넓/PA	Large	1	0.988
42	분위기/NC	Atmosphere	색다르/PA	Different	1	0.981
43	공원/NC	Park	괜찮/PA	Fine	1	0.979
44	조용/NC	Silent	괜찮/PA	Fine	1	0.979
45	허브향/NC	Scent of herb	취하/PV	Drunk	1	0.964
46	휴식/NC	Rest	취하/PV	Drunk	1	0.964
47	주차비용/NC	Parking fee	싸/PA	Cheap	1	0.896
48	이용하기/NC	Using	어렵/PA	Difficult	-1	1.000
49	주차장/NC	Parking lot	좁/PA	Small	-1	0.981
50	땅값/NC	Land price	비싸/PA	Expensive	-1	0.968
51	물가/NC	Price	비싸/PA	Expensive	-1	0.968
52	향/NC	Scent	느끼/PV	Feel	0	0.794
53	여유/NC	Spare	느끼/PV	Feel	0	0.794
54	공원/NC	Park	작/PA	Small	0	0.567
55	공원/NC	Park	크/PA	Big	0	0.543
56	규모/NC	Scale	크/PA	Big	0	0.543

Conclusion



Contribution and future works

- Sentiment dictionaries have been constructed for analyzing product reviews, but no sentiment dictionary has been established for the places.
- Therefore, it is meaningful that this study constructed the sentiment lexicon for place and calculated the sentiment polarity and probability score of by word.
- The spatial sentiment lexicon could be utilized as a reference when performing sentiment analysis on the contents of various social media platforms, and could offer useful information to those who want to visit a place.

- In future work, we will study a method used to
  1) Extending the lexicon by adding synonyms for pre-constructed sentiment words.
  2) Developing the methodology to analyze syntax more precisely.

THANK YOU FOR LISTENING

Youngmin Lee
daldanka@snu.ac.kr