



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학박사 학위논문

리뷰 텍스트를 활용한  
토픽별 키워드 기반  
시맨틱 POI 검색

Topical Keywords based Semantic POI  
Retrieval using Review Texts

2021년 2월

서울대학교 대학원

건설환경공학부

이영민

리뷰 텍스트를 활용한  
토픽별 키워드 기반  
시맨틱 POI 검색

Topical Keywords based Semantic POI  
Retrieval using Review Texts

지도교수 유 기 윤

이 논문을 공학박사 학위논문으로 제출함

2021년 1월

서울대학교 대학원

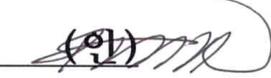
건설환경공학부

이 영 민

이영민의 박사 학위논문을 인준함

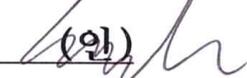
2021년 1월

위원장 김용익 (인) 

부위원장 유기훈 (인) 

위원 원준희 (인) 

위원 김현정 (인) 

위원 이원희 (인) 

## 국문초록

위치 기반 서비스 및 웹 지도 서비스에서 특정 장소나 위치에 대한 검색은 일상생활에서 매우 유용한 기능으로, 사용자들이 가장 많이 활용하는 서비스 중 하나이다. 장소 및 위치 검색 서비스는 빅데이터 및 인공지능 기술의 발달과 함께 맞춤형 서비스로 진화 중이며, 관심지점(point of interest, 이하 POI) 데이터는 이러한 서비스 제공을 위한 핵심 요소라고 할 수 있다. 그러나 기존 국내외 POI 검색 서비스들은 주로 POI 명칭, 카테고리, 태그 위주의 검색을 제공하고 있으며, 리뷰 검색 기능을 일부 지원하고 있으나 POI의 시맨틱(semantic) 정보를 활용하지 않아 POI 검색 시 사용자들의 다양한 상황과 선호를 반영하기 어려운 한계가 있다.

본 연구는 사용자들의 다양한 요구조건을 반영할 수 있는 토픽별 키워드 기반 시맨틱 POI 검색 기술을 개발하는 것을 목표로 한다. 이를 위해 사용자들이 POI에 대해 작성한 비정형의 리뷰 텍스트 데이터를 활용해 시맨틱 POI를 생성하는 프로세스를 제안하고, 시맨틱 POI를 검색에 활용할 수 있는 세 종류의 활용 구조를 설계하였다. 이를 위한 과정으로써 먼저, 데이터 수집 및 전처리 방법을 제시하고, 리뷰 텍스트로부터 토픽과 토픽별 키워드 추출을 위한 잠재 디리클레 할당(latent Dirichlet allocation, LDA) 기반 토픽 모델링(topic modeling) 방법을 제안하였다. 그리고 Word2vec 모델을 기반으로 토픽별 키워드를 임베딩(embedding)함으로써 최종적으로 POI를 임베딩하는 방법을 제안하였다. 이를 통해 토픽에 따라 다양하게 사용되는 키워드의 특성을 반영한 시맨틱 POI를 확정하였다. 또한, 시맨틱 POI 검색에 대한 활용 구조로써 자연어 기반 검색어에 대한 결과를 제공하는 자연어 기반 POI 검색, 토픽과 토픽별 키워드를 선택해 POI를 검색하는 선택 기반 POI 검색, 그리고 특정 POI와 유사한 POI를 탐색할 수 있는 유사 POI 검색을 제안하였다. 시맨틱 POI 검색에의 활용 결과에 대한 품질을 확인하기 위하여 서울시에 거주하는 남녀 400명을 대상으로 설문 조사를 시행하였다. 그 결과 만족도 평균이 5점 만점에 3.48점으로

산출되었으며, 이는 본 연구에서 제시한 평가 기준에 따라 만족의 범위에 속하는 것으로 판단할 수 있다. 또한, 사용자 만족도를 만족 정도에 따라 구분하여 분석한 결과, 모든 문항에서 ‘만족’을 선택한 응답 비율이 가장 높았으며, ‘매우 만족’ 및 ‘만족’을 선택한 비율이 ‘매우 불만족’ 및 ‘불만족’을 선택한 비율보다 평균 약 4.1배 더 많은 것으로 나타났다. 이에 따라 본 연구에서 개발한 시맨틱 POI 검색 기술의 품질을 확인할 수 있었다. 마지막으로, 본 연구에서 도출한 시맨틱 POI 검색 결과와 기존 POI 검색 서비스들과의 검색 결과 비교 분석을 통해 본 연구의 차별성을 확인하였다. 즉, 본 연구에서 개발한 시맨틱 POI 검색 기술을 활용하면 사용자가 검색어에 대한 가중치를 직접 지정할 수 있으므로 같은 검색어라도 가중치에 따라 다른 결과가 출력됨으로써 사용자 맞춤형 검색이 가능하다. 또한, 검색어와 정확히 일치하는 키워드가 없어도 동의어 및 유의어를 기반으로 한 결과도 도출되며, 같은 키워드라도 토픽에 따라 다르게 사용되는 키워드의 다양한 의미를 검색에 활용할 수 있다. 따라서 본 연구에서 개발한 시맨틱 POI 검색 기술을 통해 사용자들의 다양한 선호와 상황을 반영한 맞춤형 검색이 가능하며, 이는 기존의 POI 검색 시스템과 상호 보완적으로 활용할 수 있다.

**주요어** : 시맨틱 POI, POI 검색, POI 임베딩, 사용자 맞춤형 검색, Word2vec, LDA

**학 번** : 2014-30242

# 목 차

1. 서 론 .....	1
1.1 연구 배경 및 목적 .....	1
1.2 연구 동향 .....	11
1.3 연구 범위 및 방법 .....	17
2. 연구 방법 .....	20
2.1 데이터 수집 및 전처리 .....	21
2.1.1 데이터 수집 .....	21
2.1.2 데이터 전처리 .....	25
2.2 LDA 기반 토픽 모델링 .....	28
2.2.1 토픽별 키워드 추출을 위한 LDA .....	28
2.2.2 LDA 평가 방법 .....	34
2.3 토픽별 키워드 기반 POI 임베딩 .....	38
2.3.1 시맨틱 POI 생성을 위한 POI 임베딩 .....	38
2.3.2 POI 임베딩 평가 방법 .....	45
2.4 시맨틱 POI 검색에의 활용 구조 .....	49
2.4.1 자연어 기반 POI 검색 .....	49
2.4.2 선택 기반 POI 검색 .....	54
2.4.3 유사 POI 검색 .....	58
2.5 종합적인 사용자 평가 .....	60

3. 실험 적용 및 결과 .....	64
3.1 실험 환경 .....	64
3.2 데이터 수집 및 전처리 결과 .....	64
3.2.1 데이터 수집 결과 .....	64
3.2.2 데이터 전처리 결과 .....	66
3.3 LDA 기반 토픽 모델링 결과 .....	69
3.4 토픽별 키워드 기반 POI 임베딩 결과 .....	79
3.5 시맨틱 POI 검색에의 활용 결과 .....	91
3.5.1 결과 분석 개요 .....	91
3.5.2 자연어 기반 POI 검색에의 활용 결과 .....	91
3.5.3 선택 기반 POI 검색에의 활용 결과 .....	108
3.5.4 유사 POI 검색에의 활용 결과 .....	120
4. 결 론 .....	130
참 고 문 헌 .....	133
부 록 .....	147
Abstract .....	179

## 표 목 차

[표 1-1] POI 검색 관련 서비스 현황 .....	8
[표 1-2] POI 임베딩 관련 연구 요약 .....	15
[표 2-1] 코모란 형태소 분석기의 품사 종류 일부 .....	27
[표 2-2] 실험에 사용한 하이퍼 파라미터 조합 .....	47
[표 2-3] 키워드별 토픽 할당 알고리즘 의사 코드 .....	51
[표 2-4] 시맨틱 POI 검색 종류별 설문 구성(문항 1세트 기준) .....	61
[표 3-1] 망고플레이트 데이터 수집 현황 .....	65
[표 3-2] 망고플레이트 데이터 전처리 후 현황 .....	66
[표 3-3] 망고플레이트 리뷰 텍스트 형태소 분석 결과 일부 .....	67
[표 3-4] 토픽 모델링 분석 도구 비교(유예림(2017) p.73 <표 II-11> 재구성) ..	70
[표 3-5] 수집한 데이터의 음식 종류별 군집화 .....	71
[표 3-6] LDA 모델의 토픽 개수에 따른 복잡도 값 .....	72
[표 3-7] LDA 모델의 토픽 개수에 따른 CV 값 .....	73
[표 3-8] LDA 기반 토픽 모델링 결과 .....	74
[표 3-9] ‘중_VA(좋다)’의 토픽별 확률 분포 .....	76
[표 3-10] ‘덕화장’에 대한 리뷰 텍스트 형태소 분석 결과 일부 .....	77
[표 3-11] 리뷰([표 3-10] b 전체)에 대한 토픽 분포 .....	77
[표 3-12] 키워드별([표 3-10] b의 각 키워드) 토픽 분포 .....	78
[표 3-13] 하이퍼 파라미터 조합에 따른 POI의 음식 종류 카테고리 간 일치 비율 ..	81
[표 3-14] 토픽별 키워드 기반 POI 임베딩 관련 데이터 현황 .....	82
[표 3-15] ‘짜장면_NNG’ 벡터(크기: 250) .....	83
[표 3-16] 5번 토픽(중식) 벡터(크기: 250) .....	84
[표 3-17] 서울 서대문구 연희동에 위치한 POI ‘03InNOut’ 벡터(크기: 500) ..	87
[표 3-18] ‘분위기’ 토픽의 키워드 ‘좋다’와 관련성 높은 상위 10개 POI ..	89
[표 3-19] ‘가격’ 토픽의 키워드 ‘좋다’와 관련성 높은 상위 10개 POI ..	89
[표 3-20] ‘인생 피자’ 형태소 분석 결과 및 가중치 할당 .....	93
[표 3-21] ‘인생(w: 0.5)’, ‘피자(w: 0.5)’ 검색 결과(상위 10개) .....	93

[표 3-22] ‘인생_NNG’와 유사한 키워드 상위 20개 .....	94
[표 3-23] ‘피자_NNG’와 유사한 키워드 상위 20개 .....	94
[표 3-24] ‘인생(w: 0.9)’, ‘피자(w: 0.1)’ 검색 결과(상위 10개) .....	95
[표 3-25] 자연어 기반 검색 키워드 형태소 분석 결과 및 가중치 할당 (토픽별 키워드 3개) .....	96
[표 3-26] ‘식물(w: 0.5)’, ‘덕후(w: 0.3)’, ‘좋아하다(w: 0.2)’ 검색 결과(상위 10개) .....	97
[표 3-27] ‘스트레스(w: 0.25)’, ‘폴리논(w: 0.25)’, ‘매운맛(w: 0.5)’ 검색 결과 (상위 10개) .....	98
[표 3-28] ‘뷰(w: 0.4)’, ‘좋은(w: 0.3)’, ‘카페(w: 0.3)’ 검색 결과(상위 10개) ..	99
[표 3-29] 응답자 특성 .....	100
[표 3-30] 설문 의 신뢰도 분석 결과 .....	101
[표 3-31] 자연어 기반 POI 검색에 대한 사용자 만족도 평가 결과 ..	102
[표 3-32] 자연어 기반 POI 검색에 대한 사용자 만족도 결과 분포 ..	102
[표 3-33] 토픽: ‘이탈리안’, 키워드: ‘파스타’ 선택 결과 도출된 상위 100개 POI의 음식 종류 통계 .....	108
[표 3-34] 토픽: ‘이탈리안’, 키워드: ‘파스타’ 선택 결과(상위 10개) ..	109
[표 3-35] 토픽: ‘이탈리안’, 키워드: ‘파스타(w: 0.9)’, ‘맛있는(w: 0.1)’ 선택 결과(상위 10개) .....	110
[표 3-36] 토픽: ‘이탈리안’, 키워드: ‘파스타(w: 0.7)’, ‘맛있는(w: 0.3)’ 선택 결과(상위 10개) .....	111
[표 3-37] 토픽: ‘이탈리안’, 키워드: ‘파스타(w: 0.5)’, ‘맛있는(w: 0.5)’ 선택 결과(상위 10개) .....	112
[표 3-38] 토픽: ‘이탈리안’, 키워드: ‘파스타(w: 0.3)’, ‘맛있는(w: 0.7)’ 선택 결과(상위 10개) .....	113
[표 3-39] 토픽: ‘이탈리안’, 키워드: ‘파스타(w: 0.1)’, ‘맛있는(w: 0.9)’ 선택 결과(상위 10개) .....	114
[표 3-40] 토픽: ‘이탈리안’, 키워드: ‘맛있는’ 선택 결과(상위 10개) ..	115
[표 3-41] 선택 기반 POI 검색에 대한 사용자 만족도 평가 결과 .....	116

[표 3-42]	선택 기반 POI 검색에 대한 사용자 만족도 결과 분포	117
[표 3-43]	‘조용한(w: 0.8)’, ‘카페(w: 0.2)’ 검색 결과(상위 10개)	119
[표 3-44]	‘1201 커피카운터’와 유사한 POI 검색 결과(상위 10개)	120
[표 3-45]	‘고고초밥’과 유사한 POI 검색 결과(상위 10개)	121
[표 3-46]	‘129 라멘하우스’와 유사한 POI 검색 결과(상위 10개)	121
[표 3-47]	‘고대양꼬치집’과 유사한 POI 검색 결과(상위 10개)	122
[표 3-48]	‘1986베이커’와 유사한 POI 검색 결과(상위 10개)	123
[표 3-49]	‘계타는날 부암동치킨’과 유사한 POI 검색 결과(상위 10개)	123
[표 3-50]	‘익선잡방’과 유사한 POI 검색 결과(상위 10개)	124
[표 3-51]	‘저스틴스테이크’와 유사한 POI 검색 결과(상위 10개)	124
[표 3-52]	유사 POI 검색에 대한 사용자 만족도 평가 결과	125
[표 3-53]	유사 POI 검색에 대한 사용자 만족도 결과 분포	126
[표 3-54]	‘히메지’와 유사한 POI 검색 결과(상위 10개)	129
[표 A.2-1]	문항세트 a 검색어 1(자연어 기반 POI 검색, 토픽별 키워드 1개)	148
[표 A.2-2]	문항세트 a 검색어 2(자연어 기반 POI 검색, 토픽별 키워드 2개)	149
[표 A.2-3]	문항세트 a 검색어 3(자연어 기반 POI 검색, 토픽별 키워드 3개)	150
[표 A.2-4]	문항세트 a 검색어 4(자연어 기반 POI 검색, 토픽별 키워드 4개)	151
[표 A.2-5]	문항세트 a 검색어 5(자연어 기반 POI 검색, 토픽별 키워드 1개)	152
[표 A.2-6]	문항세트 a 검색어 6(자연어 기반 POI 검색, 토픽별 키워드 2개)	153
[표 A.2-7]	문항세트 a 검색어 7(자연어 기반 POI 검색, 토픽별 키워드 3개)	154
[표 A.2-8]	문항세트 a 검색어 8(자연어 기반 POI 검색, 토픽별 키워드 4개)	155
[표 A.2-9]	문항세트 a 검색어 9(유사 POI 검색)	156
[표 A.3-1]	문항세트 b 검색어 1(자연어 기반 POI 검색, 토픽별 키워드 1개)	157
[표 A.3-2]	문항세트 b 검색어 2(자연어 기반 POI 검색, 토픽별 키워드 2개)	157
[표 A.3-3]	문항세트 b 검색어 3(자연어 기반 POI 검색, 토픽별 키워드 3개)	158
[표 A.3-4]	문항세트 b 검색어 4(자연어 기반 POI 검색, 토픽별 키워드 4개)	158
[표 A.3-5]	문항세트 b 검색어 5(자연어 기반 POI 검색, 토픽별 키워드 1개)	159
[표 A.3-6]	문항세트 b 검색어 6(자연어 기반 POI 검색, 토픽별 키워드 2개)	159
[표 A.3-7]	문항세트 b 검색어 7(자연어 기반 POI 검색, 토픽별 키워드 3개)	159

[표 A.3-8] 문항세트 b 검색어 8(자연어 기반 POI 검색, 토픽별 키워드 4개) ..	160
[표 A.3-9] 문항세트 b 검색어 9(유사 POI 검색) .....	160
[표 A.4-1] 문항세트 c 검색어 1(자연어 기반 POI 검색, 토픽별 키워드 1개) ..	161
[표 A.4-2] 문항세트 c 검색어 2(자연어 기반 POI 검색, 토픽별 키워드 2개) ..	161
[표 A.4-3] 문항세트 c 검색어 3(자연어 기반 POI 검색, 토픽별 키워드 3개) ..	162
[표 A.4-4] 문항세트 c 검색어 4(자연어 기반 POI 검색, 토픽별 키워드 4개) ..	162
[표 A.4-5] 문항세트 c 검색어 5(자연어 기반 POI 검색, 토픽별 키워드 1개) ..	163
[표 A.4-6] 문항세트 c 검색어 6(자연어 기반 POI 검색, 토픽별 키워드 2개) ..	163
[표 A.4-7] 문항세트 c 검색어 7(자연어 기반 POI 검색, 토픽별 키워드 3개) ..	163
[표 A.4-8] 문항세트 c 검색어 8(자연어 기반 POI 검색, 토픽별 키워드 4개) ..	164
[표 A.4-9] 문항세트 c 검색어 9(유사 POI 검색) .....	164
[표 A.5-1] 문항세트 d 검색어 1(자연어 기반 POI 검색, 토픽별 키워드 1개) ..	165
[표 A.5-2] 문항세트 d 검색어 2(자연어 기반 POI 검색, 토픽별 키워드 2개) ..	165
[표 A.5-3] 문항세트 d 검색어 3(자연어 기반 POI 검색, 토픽별 키워드 3개) ..	166
[표 A.5-4] 문항세트 d 검색어 4(자연어 기반 POI 검색, 토픽별 키워드 4개) ..	166
[표 A.5-5] 문항세트 d 검색어 5(자연어 기반 POI 검색, 토픽별 키워드 1개) ..	167
[표 A.5-6] 문항세트 d 검색어 6(자연어 기반 POI 검색, 토픽별 키워드 2개) ..	167
[표 A.5-7] 문항세트 d 검색어 7(자연어 기반 POI 검색, 토픽별 키워드 3개) ..	167
[표 A.5-8] 문항세트 d 검색어 8(자연어 기반 POI 검색, 토픽별 키워드 4개) ..	168
[표 A.5-9] 문항세트 d 검색어 9(유사 POI 검색) .....	168
[표 B.1-1] ‘연인(w: 0.5)’, ‘기념일(w: 0.5)’ 검색 결과(상위 10개) .....	170
[표 B.1-2] ‘믿고(w: 0.5)’, ‘먹는(w: 0.1)’, ‘곱창(w: 0.4)’ 검색 결과(상위 10개) ..	172
[표 B.1-3] ‘바게트(w: 0.4)’, ‘맛있는(w: 0.3)’, ‘베이커리(w: 0.2)’, ‘맛집(w: 0.1)’ 검색 결과(상위 10개) .....	174
[표 B.1-4] ‘연인(w: 0.3)’, ‘화려한(w: 0.3)’, ‘와인바(w: 0.4)’ 검색 결과(상위 10개) ..	176
[표 B.1-5] ‘배경음악(w: 0.4)’, ‘선곡(w: 0.2)’, ‘센스(w: 0.2)’, ‘좋은(w: 0.2)’ 검색 결과 (상위 10개) .....	178

## 그림 목 차

[그림 1-1] 네이버 지도 웹 검색 결과 화면 .....	3
[그림 1-2] 네이버 웹 포털 검색 결과 화면(일부) .....	4
[그림 1-3] 다이닝코드 앱 검색 결과 화면(일부) .....	6
[그림 1-4] 구글 지도 웹 검색 결과 화면(일부) .....	7
[그림 1-5] 연구 흐름도 .....	19
[그림 2-1] 연구 방법의 흐름도 .....	20
[그림 2-2] 망고플레이트 웹 사이트 메인 화면 .....	21
[그림 2-3] 망고플레이트 검색 결과 화면 .....	23
[그림 2-4] POI 명에 대한 HTML .....	24
[그림 2-5] 리뷰 텍스트에 대한 HTML .....	24
[그림 2-6] LDA의 기본 가정(Blei(2012) p. 78 Figure 1 재구성) .....	29
[그림 2-7] LDA의 그래프 모델(Blei(2012) p.81 Figure 4 재구성) .....	31
[그림 2-8] 토픽 모델의 확률 생성 과정(Steyvers & Griffiths(2007) p.3 Figure 2 재구성) ..	32
[그림 2-9] Word2Vec의 두 가지 학습 방법(Tutubalina & Nikolenko, 2017) ..	39
[그림 2-10] TWE-1 및 skip-gram 모델 개념도(Fu et al.(2016) p.194 Figure 2; Liu et al.(2015b) p.2419 Figure 1 재구성) .....	41
[그림 2-11] 자연어 기반 POI 검색을 위한 활용 구조 .....	50
[그림 2-12] 자연어 기반 POI 검색 메인 화면 .....	52
[그림 2-13] 자연어 기반 POI 검색의 가중치 입력 화면 .....	53
[그림 2-14] 자연어 기반 POI 검색의 결과 출력 화면 .....	53
[그림 2-15] 선택 기반 POI 검색을 위한 활용 구조 .....	54
[그림 2-16] 선택 기반 POI 검색 메인 화면 .....	56
[그림 2-17] 선택 기반 POI 검색의 토픽별 키워드 선택 화면 .....	56
[그림 2-18] 선택 기반 POI 검색의 가중치 입력 화면 .....	57
[그림 2-19] 선택 기반 POI 검색의 결과 출력 화면 .....	57
[그림 2-20] 유사 POI 검색을 위한 활용 구조 .....	58
[그림 2-21] 유사 POI 검색의 결과 출력 화면 .....	59

[그림 2-22] 모바일 설문조사 화면 캡처 .....	63
[그림 3-1] LDA 모델의 토픽 개수에 따른 복잡도 값 변화 .....	72
[그림 3-2] LDA 모델의 토픽 개수에 따른 CV 값 변화 .....	73
[그림 3-3] 하이퍼 파라미터 조합에 따른 POI 임베딩 모델의 성능 평가 ..	80
[그림 3-4] ‘분위기 및 ‘가격 토픽에 속하는 상위 20개 키워드 벡터의 차원 감소 결과 ..	85
[그림 3-5] 토픽별 키워드 벡터 차원 축소 결과 .....	86
[그림 3-6] 토픽별 키워드 벡터와 POI 벡터 차원 축소 결과 .....	88
[그림 3-7] 분위기 및 가격 토픽별 키워드 및 관련 POI 벡터 차원 감소 결과 ..	90
[그림 3-8] 만족도 평가 기준 .....	101
[그림 3-9] ‘인생 피자’ 검색 결과(지도 중심 플랫폼) .....	104
[그림 3-10] ‘인생 피자’ 검색 결과(맛집 중심 플랫폼) .....	104
[그림 3-11] ‘식물 덕후가 좋아할 만한 곳’ 검색 결과(지도 중심 플랫폼) ..	105
[그림 3-12] ‘식물 덕후가 좋아할 만한 곳’ 검색 결과(맛집 중심 플랫폼) ..	105
[그림 3-13] ‘뷰 좋은 카페’ 검색 결과(지도 중심 플랫폼) .....	107
[그림 3-14] ‘뷰 좋은 카페’ 검색 결과(맛집 중심 플랫폼) .....	107
[그림 3-15] 선택 기반 POI 검색 결과(맛집 중심 플랫폼) .....	118
[그림 3-16] 세 가지 시맨틱 POI 검색에 대한 사용자의 만족 정도별 평균값 분포 차트 ..	127
[그림 3-17] 유사 POI 검색 결과(지도 중심 플랫폼) .....	128
[그림 3-18] 유사 POI 검색 결과(망고플레이트) .....	128
[그림 A.2-1] 문항세트 a의 검색어 1에 대한 설문 화면 캡처 .....	148
[그림 A.2-2] 문항세트 a의 검색어 2에 대한 설문 화면 캡처 .....	149
[그림 A.2-3] 문항세트 a의 검색어 3에 대한 설문 화면 캡처 .....	150
[그림 A.2-4] 문항세트 a의 검색어 4에 대한 설문 화면 캡처 .....	151
[그림 A.2-5] 문항세트 a의 검색어 5에 대한 설문 화면 캡처 .....	152
[그림 A.2-6] 문항세트 a의 검색어 6에 대한 설문 화면 캡처 .....	153
[그림 A.2-7] 문항세트 a의 검색어 7에 대한 설문 화면 캡처 .....	154
[그림 A.2-8] 문항세트 a의 검색어 8에 대한 설문 화면 캡처 .....	155
[그림 A.2-9] 문항세트 a의 검색어 9에 대한 설문 화면 캡처 .....	156
[그림 B.1-1] ‘연인과 기념일’ 검색 결과(지도 중심 플랫폼) .....	169

[그림 B.1-2] ‘연인과 기념일’ 검색 결과(맛집 중심 플랫폼) .....	170
[그림 B.1-3] ‘믿고 먹는 곱창집’ 검색 결과(지도 중심 플랫폼) .....	171
[그림 B.1-4] ‘믿고 먹는 곱창집’ 검색 결과(맛집 중심 플랫폼) .....	172
[그림 B.1-5] ‘바게트 맛있는 베이커리 맛집’ 검색 결과(지도 중심 플랫폼) ..	173
[그림 B.1-6] ‘바게트 맛있는 베이커리 맛집’ 검색 결과(맛집 중심 플랫폼) ..	174
[그림 B.1-7] ‘연인과 화려한 와인 바’ 검색 결과(지도 중심 플랫폼) .....	175
[그림 B.1-8] ‘연인과 화려한 와인 바’ 검색 결과(맛집 중심 플랫폼) .....	176
[그림 B.1-9] ‘배경음악 선곡 센스 좋은 곳’ 검색 결과(지도 중심 플랫폼) ....	177
[그림 B.1-10] ‘배경음악 선곡 센스 좋은 곳’ 검색 결과(맛집 중심 플랫폼) ..	177

## 용어 정의

### ■ 키워드(keyword)

: POI를 설명하는 비정형 텍스트 데이터에 대한 자연어 처리 시 형태소 분석 단계에서 일반명사(NNG), 고유명사(NNP), 동사(VV), 형용사(VA), 어근(XR)에 해당하는 품사만 추출해, 각 형태소와 해당 품사를 밑줄 기호(\_)로 연결한 것을 말한다(예: ‘짜장면\_NNG’, ‘분위기\_NNG’, ‘좋\_VA’, ‘아늑\_XR’).

### ■ 토픽(topic)

: ‘카테고리’, ‘분위기’, ‘가격’ 등 POI가 가지고 있는 여러 가지 속성을 의미한다. POI에 대한 토픽들은 POI를 설명하는 키워드들로부터 도출된다.

### ■ 토픽별 키워드(keywords by topic)

: 각 키워드에는 1개 이상의 토픽이 부여되는데, 이때 토픽이 부여된 키워드를 ‘토픽별 키워드’라고 지칭한다. 같은 키워드라도 다른 토픽이 부여되면 서로 다른 토픽별 키워드로 취급한다.

### ■ 단어 임베딩(word embedding)

: 임베딩은 자연어 처리 분야에서 주로 사용되는 용어로, 텍스트 데이터를 구성하는 단어의 의미와 맥락을 고려해 각 단어를 벡터 공간상의 한 점으로 대응시키는 기법을 말한다. 대표적인 단어 임베딩 기법인 Word2vec은 실숫값을 이용해 연구자가 설정한 크기로 단어 벡터를 밀집 표현(dense representation)한다. 텍스트 데이터에 임베딩 기법을 적용하면 입력 단어들을 모두 벡터화할 수 있으므로 이를 기반으로 단어 간의 관련성 및 유사성을 측정할 수 있다.

■ POI 임베딩(POI embedding)

: POI를 설명하는 리뷰 텍스트로부터 추출한 토픽별 키워드들의 의미와 맥락을 고려해 각 POI를 벡터 공간상의 한 점으로 대응시키는 기법을 말한다. 즉, 본 연구에서는 단어들이 모여 문서를 이루는 것처럼 POI를 설명하는 토픽별 키워드들이 모여 POI를 나타낼 수 있다고 간주하고 문서를 POI로 치환함으로써 POI를 임베딩한다. 이때 POI를 임베딩한 결과가 시맨틱 POI이다. 단어 임베딩과 마찬가지로, POI 임베딩 기법을 적용하면 POI를 모두 벡터화할 수 있으므로 이를 기반으로 POI 간 관련성 및 유사성을 측정할 수 있다.

■ 시맨틱 POI(semantic POI)

: POI를 설명하는 리뷰 텍스트로부터 추출한 토픽별 키워드들과 키워드 간 의미 관계에 기반한 가중치 정보를 포함하는 POI를 지칭한다. 시맨틱 POI가 확정되면 검색어와 POI 간 관련성에 기반한 검색(직접 검색) 및 POI 간 유사성을 기반으로 한 검색(추천검색)이 가능하다.

■ 관련성(relatedness) 및 유사성(similarity)

: 관련성과 유사성은 서로 연관된 개념으로, 관련성을 유사성의 상위 속성으로 볼 수 있다. 즉, 유사성이 있으면 관련성이 있으나 관련성이 있다고 해서 유사성이 있는 것은 아니다. 본 연구에서 개발한 시맨틱 POI 검색 기술 중 자연어 기반 POI 검색과 선택 기반 POI 검색이 관련성에 기반한 검색에 해당되며, 유사 POI 검색은 유사성에 기반한 검색에 해당된다.

■ 직접검색(direct retrieval)

: 사용자가 검색어를 직접 입력/선택함으로써 수행되는 검색으로, 본 연구에서 개발한 시맨틱 POI 검색 중 자연어 기반 POI 검색과 선택 기반 POI 검색이 직접검색 방식에 해당된다.

■ 추천검색(recommended retrieval)

- : 사용자가 입력/선택한 검색 결과로 도출된 POI를 기준으로 유사한 POI를 추천해 주거나 이전 POI 방문 이력 등 사용자 정보를 바탕으로 POI를 추천해 주는 방식으로, 본 연구에서 개발한 시맨틱 POI 검색 중 유사 POI 검색이 추천검색 방식에 해당된다.

# 1. 서론

## 1.1 연구 배경 및 목적

스마트폰 보급이 대중화되고, 이동통신 기술 및 범세계적 위치결정 체계(global positioning system, GPS), 와이파이(wireless fidelity, Wi-Fi) 등의 측위 기술이 발전하면서 위치 기반 서비스(location-based service, 이하 LBS) 시장도 함께 성장하였다. 이제 스마트 기기에서 LBS는 현대인의 필수 애플리케이션으로 자리 잡았으며, 과거 단순히 위치나 경로 안내를 제공하던 역할에서 빅데이터 분석과 인공지능에 기반한 맞춤형 서비스로 진화하고 있다.

LBS에서 특정 장소 및 위치에 대한 검색은 가장 기본적이고 핵심적인 기능으로, 이러한 검색에 대한 응답은 관심지점(point of interest, 이하 POI) 데이터를 기반으로 처리된다. POI는 위치, 명칭, 카테고리 등의 정보를 포함하는 점(point) 속성의 공간 객체로, 학교, 은행, 병원, 음식점 등 사람들의 삶과 밀접한 관련이 있는 지리적 요소를 지칭한다(Chuang *et al.*, 2016; Liu *et al.*, 2019; Zhang *et al.*, 2010; Zhu & Zhou, 2009). 실제로 POI 검색은 스마트폰에서 가장 많이 사용되는 서비스 중 하나로, 우리 일상생활과 밀접한 관련이 있다. 예를 들어 관광객은 스마트폰을 이용해 관광지 주변의 음식점을 검색하고, 새로운 지역으로 이사한 사람은 스마트폰으로 슈퍼마켓이나 약국 같은 지역 상점을 검색한다(Chuang *et al.*, 2016).

POI 검색을 제공하는 국내 서비스는 네이버 지도(Naver map)<sup>1)</sup>, 카카오맵(Kakao map)<sup>2)</sup> 등 포털 사이트 기반의 지도 중심 플랫폼과 망고플레이트(Mangoplate)<sup>3)</sup>, 다이닝코드(Diningcode)<sup>4)</sup> 등의 맛집 중심 플랫폼이 있으며, 해외 서비스로는 포털 기반의 지도 중심 플랫폼인 구글 지도(Google

---

1) 네이버 지도, <https://map.naver.com/> (접속일: 2020년 10월 24일)

2) 카카오맵, <https://map.kakao.com/> (접속일: 2020년 10월 24일)

3) 망고플레이트, <https://www.mangoplate.com/> (접속일: 2020년 10월 24일)

4) 다이닝코드, <https://www.diningcode.com/> (접속일: 2020년 10월 24일)

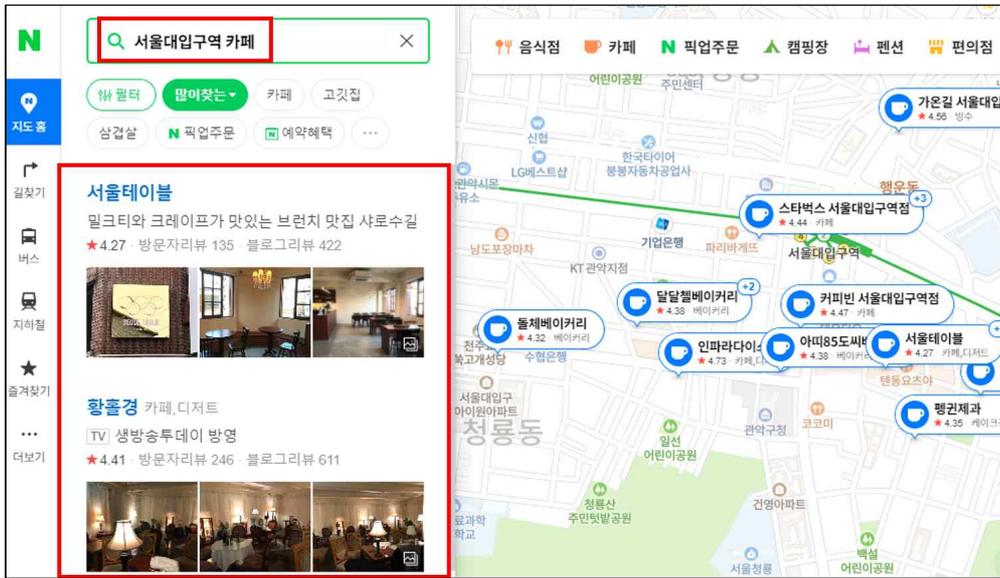
maps)<sup>5)</sup>가 있다. 이러한 기존 플랫폼들은 POI 명칭, 카테고리 또는 태그 위주의 검색 서비스를 제공하고 있고, 리뷰 검색 기능을 일부 지원하고 있으나 POI의 ‘시맨틱(semantic)’ 정보를 검색에 활용하지 않아, 사용자들의 다양한 선호나 상황을 반영한 검색이 어려운 한계가 있다.

시맨틱은 주로 정보 검색 분야에서 사용되는 용어로, 사전적으로는 ‘의미의’, ‘의미론적인’이라는 뜻이 있다. 일반적으로 단순 단어 매칭 검색과 달리 검색어의 의미를 파악해 결과를 반환하는 것을 ‘시맨틱 검색’이라 일컫는다. 이러한 개념을 바탕으로 본 연구에서는 POI와 관련된 시맨틱 개념을 다음과 같이 정의한다. ‘시맨틱 POI’란 POI를 설명하는 텍스트가 있을 때, 해당 텍스트에서 추출한 토픽별 키워드들과 이들의 의미 관계에 기반한 가중치 정보를 포함하는 POI를 지칭한다. 시맨틱 POI가 확정되면, 검색어와 POI 간의 관련성(relatedness)<sup>6)</sup>에 기반한 검색(직접검색)이나 POI 간 유사성(similarity)<sup>7)</sup>을 기반으로 한 검색(추천검색)이 가능하며, 본 연구에서는 이들을 시맨틱 POI 검색이라 통칭한다. 여기서 직접검색이란 사용자가 검색어를 직접 입력/선택함으로써 수행되는 검색을 의미하고, 추천검색은 사용자가 입력/선택한 결과로 도출된 POI를 기준으로 유사한 POI를 추천해 주거나, 이전 POI 방문 이력 등 사용자 정보를 바탕으로 추천해 주는 방식을 의미한다. 그러나 기존 서비스들은 시맨틱 POI를 검색에 활용하지 않고 있다. 가령, 국내 지도 서비스 중 가장 많은 이용자 수를 보유한<sup>8)</sup> 네이버 지도에서 ‘서울대입구역 카페(이하 ‘검색어 a’)’로 검색하면 관련 POI들이 여러 개 출력되지만, ‘서울대입구역 분위기 좋은 카페(이하 ‘검색어 b’)’로 검색했을 때는 아무런 결과가 출력되지 않는다([그림 1-1]).

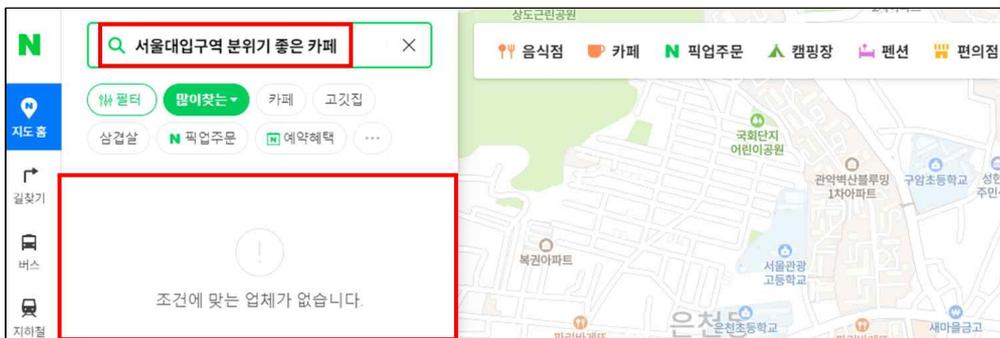
5) 구글 지도, <https://www.google.com/maps/> (접속일: 2020년 11월 24일)

6), 7) 관련성과 유사성은 서로 연관된 개념으로, 관련성을 유사성의 상위 속성으로 볼 수 있음. 즉, 유사성이 있으면 관련성이 있지만, 관련성이 있다고 유사성이 있는 것은 아님. 가령, ‘강감찬 공원’이라는 POI는 ‘강감찬’이라는 사람과 ‘관련’이 있지만 공원과 사람이 ‘유사’하지는 않음(Yan *et al.*, 2017)

8) 2020년 1월 기준, 네이버 지도의 순이용자 수는 약 1,380만 명, 카카오 맵의 순이용자 수는 약 840만 명으로 확인됨(닐슨코리아클릭, 2020)



(a) '서울대입구역 카페' 검색 결과(일부)

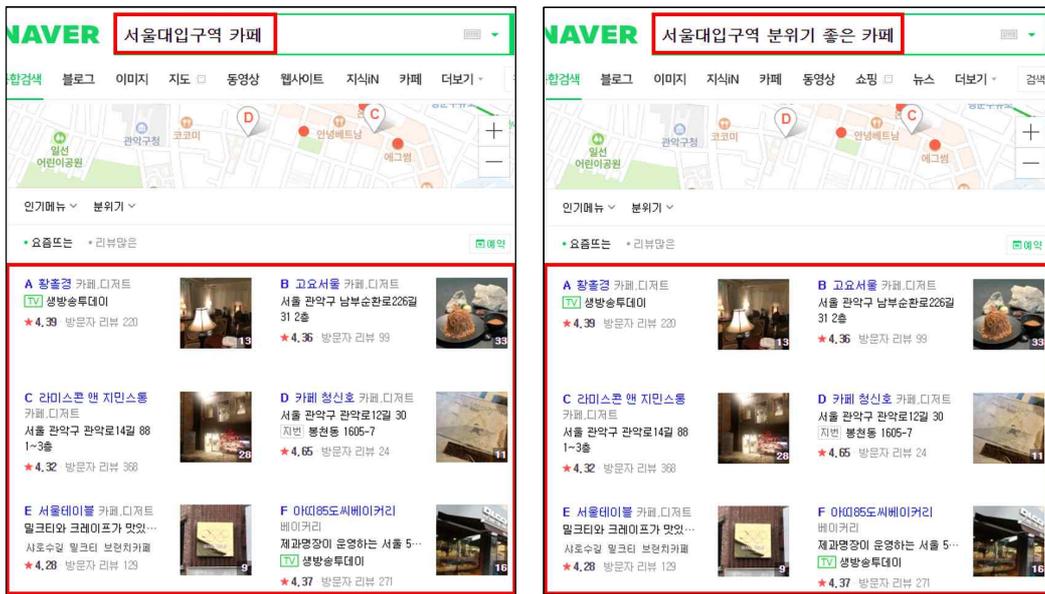


(b) '서울대입구역 분위기 좋은 카페' 검색 결과

[그림 1-1] 네이버 지도 웹 검색 결과 화면

네이버 지도에서 '서울대입구역 분위기 좋은 카페'가 아닌 '분위기 좋은 카페'로 검색 시 102개의 POI가 결과로 출력되는데, 이를 통해 네이버 지도가 리뷰 검색 기능을 일부 지원하고 있는 것으로 추측된다. 그러나 이러한 검색은 '맛집', '분위기' 등 사용자 수요가 많은 특정 검색어에 대해서만 가능하다는 한계가 있다. 가령, '인테리어 독특한 카페', '평일 저녁 데이트 장소 추천', '식물 덕후가 좋아할 만한 곳'과 같이 사용자들의 다양한 선호를 반영한 자연어 기반 검색어에 대해서는 결과가 출력되지 않았다. 이러한 자연어 기반 검색은 네이버 포털 사이트<sup>9)</sup>에서 수행할 수도 있는데, 네

이버 포털에서 ‘검색어 a’ 및 ‘검색어 b’로 각각 검색을 실행한 결과, 두 개의 검색어가 서로 다름에도 불구하고 결과로 반환되는 POI 목록이 같았다 ([그림 1-2]). 이는 사용자가 입력한 검색어 중 ‘분위기 좋은’이라는 자연어 기반 정보를 POI의 시맨틱 정보에 연결시키지 못하고, 단순히 ‘카페’라는 POI 카테고리 정보에만 중점을 둔 결과로 볼 수 있다.



(a) ‘서울대입구역 카페’  
검색 결과

(b) ‘서울대입구역 분위기 좋은 카페’  
검색 결과

[그림 1-2] 네이버 웹 포털 검색 결과 화면(일부)

네이버 지도에서는 검색어를 필터링(filtering)할 수 있는 ‘필터’라는 기능을 이용해 ‘인기 메뉴’, ‘방문 목적’, ‘분위기’ 등의 항목에 따라 원하는 검색 결과를 골라낼 수 있다. 그러나 항목별로 선택할 수 있는 단어의 개수가 한정적이고 하나의 항목당 하나의 단어만 선택해야 하는 제한이 있다. 또한, 본래 필터 기능은 검색어를 필터링하기 위한 목적으로 사용되므로 사용자가 특정 검색어를 입력했을 때만 이용할 수 있다는 한계가 있다. 그 밖에, 네이버 지도는 특정 POI에 대해 ‘함께 많이 찾아본 곳’으로 10개의

9) 네이버, <https://www.naver.com/> (접속일: 2020년 10월 25일)

POI 목록을 제공하는데, 이는 사용자가 해당 POI를 검색할 때 함께 검색한 다른 POI들을 추천해 주는 것으로, POI 간 시맨틱 정보에 기반한 유사성과는 차이가 있다.

카카오맵의 경우, 네이버 지도와 마찬가지로 ‘검색어 a’로 검색했을 때는 여러 개의 POI 목록을 결과로 출력해 주지만 ‘검색어 b’에 대해서는 아무 결과도 반환해 주지 못하였다. 또한, 앞서 언급한 ‘인테리어 독특한 카페’, ‘평일 저녁 데이트 장소 추천’, ‘식물 덕후가 좋아할 만한 곳’ 등과 같이 사용자들의 다양한 선호를 반영한 자연어 기반 검색도 불가능하다. 그 외에도 카카오맵은 시맨틱 POI에 기반한 검색 서비스를 제공하지 않고 있다.

맛집 중심 플랫폼인 망고플레이트의 경우, ‘검색어 a’에 대해서는 127개의 POI를 반환하는 반면, ‘검색어 b’에 대한 검색 결과로는 1개의 POI만 출력되었다. 또한, 네이버 지도나 카카오맵과 마찬가지로 다양하고 복잡한 자연어 기반 검색은 불가능하다. 망고플레이트에서도 ‘필터’ 기능을 통해 ‘음식 종류’, ‘가격’, ‘주차’ 등의 항목별로 주어진 단어를 선택해 POI를 검색할 수 있는 선택 기반 POI 검색 서비스를 일부 제공하고 있다. 그러나 항목별로 선택할 수 있는 단어의 개수가 적게는 2개에서 많게는 8개로 한정된다. 또한, 각 POI에 대해 ‘주변 인기 식당’ 4개 목록을 추천해 주고 있는데, 이는 단순히 기준이 되는 POI의 위치만을 고려한 것으로, 시맨틱 POI에 기반한 유사성과는 다르다.

빅데이터 기반 맛집 서비스를 지향하고 있는 다이닝코드의 경우, ‘검색어 a’에 대한 결과로 81개의 POI가 출력되었고, ‘검색어 b’에 대해서는 39개의 POI가 출력되었다. 다이닝코드에서는 앞서 언급한 POI 검색 서비스들에 비해 다양한 자연어 기반 검색이 가능하긴 하지만 검색할 수 있는 단어가 제한적이라는 한계가 있다. 가령, ‘평일 저녁 데이트하기 좋은 곳’으로 검색했을 때와 ‘주말 데이트 장소 추천’, ‘데이트’로 검색했을 때 결과로도출되는 상위 POI 목록이 모두 동일했는데, 이는 모든 검색어에서 ‘데이트’라는 단어만을 추출해 데이터베이스와 매칭한 결과로 보인다([그림 1-3]).



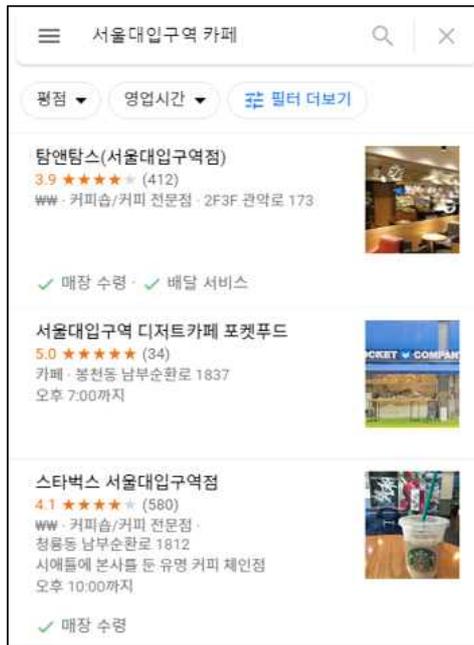
(a) '평일 저녁 데이트하기 좋은 곳' 검색 결과      (b) '주말 데이트 장소 추천' 검색 결과      (c) '데이트' 검색 결과

[그림 1-3] 다이닝코드 앱 검색 결과 화면(일부)

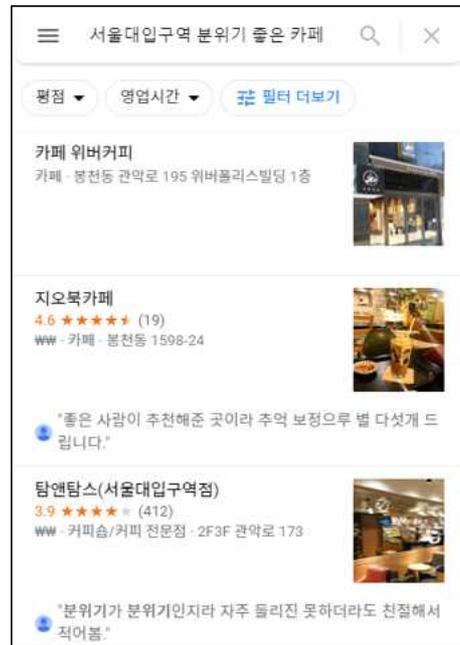
다이닝코드에서는 '테마' 기능을 이용해 '테마 맛집', '방문 목적', '분위기' 등의 항목별로 정해진 단어를 선택해 POI를 검색할 수 있는 선택 기반 POI 검색 서비스를 일부 제공하고 있다. 이때 항목별로 선택 가능한 단어의 개수는 9개 이상 20개 이하로, 망고플레이트에 비해서는 많은 편이지만, 항목 개수가 4개로 제한적이다. 또한, 다이닝코드는 시맨틱 POI 정보에 기반한 유사 POI 검색과 관련된 서비스는 제공하지 않고 있다.

구글 지도는 POI 검색 시 앞서 살펴본 국내 서비스들에 비해 텍스트 정보의 활용도가 높았다. 가령, '검색어 a' 및 '검색어 b'로 검색했을 때 도출되는 결과가 서로 달랐는데, 특히 '검색어 b'를 입력했을 때, 사용자 리뷰 텍스트와 해당 POI의 대표 웹 사이트에 기재된 텍스트 정보를 이용해 검색어와 일치하는 단어가 있는 POI를 우선적으로 반환해 주는 것을 확인할 수 있었다([그림 1-4]). 구글 지도는 국내 플랫폼들에 비해서는 리뷰 데이터를 적극적으로 활용하고 있으나 POI 명칭 검색과 리뷰 검색이 혼재되어 있기 때문에 검색에 대한 결과로, 검색어가 POI 명칭에 포함된 것과 리뷰에 포함된 것이 함께 출력되는 경향이 있다. 또한, 구글 지도에서는 검색어와 정확히 일치하는 단어가 포함된 리뷰를 위주로 검색하기 때문에 해당

검색어와 일치하는 단어가 없을 때는 검색이 이루어지지 않으며, 동의어 및 유의어를 활용한 검색에는 한계가 있는 것으로 파악되었다. 예를 들어, ‘평일 저녁 데이트 장소 추천’, ‘식물 덕후가 좋아할 만한 곳’과 같이 보다 다양한 자연어 기반 검색은 국내 서비스들과 마찬가지로 불가능했다.



(a) ‘서울대입구역 카페’  
검색 결과



(b) ‘서울대입구역 분위기 좋은 카페’ 검색 결과

[그림 1-4] 구글 지도 웹 검색 결과 화면(일부)

구글 지도에서는 ‘평점’ 및 ‘영업시간’ 정보를 기준으로 검색 결과를 필터링할 수 있는 기능을 제공하고 있으나, 이는 검색어를 입력한 결과가 있을 때만 사용 가능하다는 제한이 있다. 그 외에도 특정 POI에 대해 ‘함께 검색한 장소’로 4~5개의 POI 목록을 제공하는데, 이는 사용자가 해당 POI를 검색할 때 함께 검색한 POI들을 보여주는 것으로, POI 간 시맨틱 정보에 기반한 유사성과는 차이가 있다.

지금까지 POI 검색 기능을 제공하는 국내외 대표 서비스들의 현황을 살펴본 결과, 시맨틱 POI를 검색에 적극적으로 활용하고 있는 서비스가 부

족하다는 것을 확인할 수 있었다. 즉, 기존 검색 서비스들에서는 사용자들이 원하는 정보를 자유롭게 검색할 수 있는 자연어 기반 검색, 다양한 옵션을 선택해 POI를 검색할 수 있는 선택 기반 검색, 그리고 POI의 유사성에 기반한 유사 POI 검색이 일부 가능하거나 불가능한 것으로 확인되었다 ([표 1-1] 참조).

[표 1-1] POI 검색 관련 서비스 현황

구분		네이버 지도	카카오 맵	망고 플레이트	다이닝 코드	구글 지도
직접 검색	자연어 기반 POI 검색	일부 가능	불가능	일부 가능	일부 가능	일부 가능
	선택 기반 POI 검색	불가능 (필터 기능만 제공)	불가능	일부 가능 (카테고리, 음식 종류, 가격, 주차)	일부 가능 (테마 맛집, 방문 목적, 분위기, 편의시설)	불가능 (필터 기능만 제공)
추천 검색	유사 POI 검색	일부 가능	불가능	일부 가능	불가능	일부 가능

시맨틱 POI에 기반한 검색이 이루어지기 위해서는 POI를 나타내는 다양한 컨텍스트(context) 정보로부터 통합된 시맨틱 POI를 추출하는 것이 중요하다. 시맨틱 POI는 사용자 검색어와 POI 간의 관련성 및 POI 간 유사성을 올바르게 추론하기 위한 기반이 되며, 이것이 곧 POI 검색의 품질을 좌우하기 때문이다. POI에 대한 다양한 컨텍스트 정보를 통합해 시맨틱 POI를 추출하는 문제는 데이터로부터 POI 표현을 학습하는 방법에 대한 질문으로 이어진다. 텍스트 데이터에서 단어의 의미와 맥락을 고려해 각 단어를 벡터 공간상의 한 점으로 대응시키는 임베딩(embedding) 기법은 두 객체 간의 관련성 및 유사성 계측을 위한 좋은 도구로, 확장성 또한 뛰어나다(Yan *et al.*, 2017).

POI 관련 연구 분야에서는 추천검색 시스템에 활용하거나 도시의 토지 이용 현황을 분석하기 위한 목적으로 POI를 임베딩하는 연구가 수행되고 있다. POI 추천검색을 목표로 하는 연구 분야에서는 행렬 인수분해

(matrix factorization, 이하 MF) 기법을 기반으로 사용자가 POI에 체크인 (check-in)한 이력 데이터를 변환해 POI의 잠재 특성을 추출하고, 이를 임베딩에 활용하는 연구들(Cai *et al.*, 2018; Cheng *et al.*, 2012; Ding & Chen, 2018; Koren *et al.*, 2009; Lian *et al.*, 2014; Liu *et al.*, 2014)이 수행되었다. MF 기법은 희소(sparse) 데이터를 처리할 수 있다는 장점이 있으나 기법의 특성상 POI의 시맨틱 정보를 추출하는 것은 불가능하다. 또한, 신경망(neural network) 기반 임베딩 기법인 Word2vec(Mikolov *et al.*, 2013a; Mikolov *et al.*, 2013b) 모델을 기반으로 사용자가 POI에 체크인한 시퀀스(sequence) 데이터를 이용해 POI를 임베딩하는 연구들(Feng *et al.*, 2017; Liu *et al.*, 2016; Zhao *et al.*, 2017)도 다수 수행되었다. 최근에는 딥러닝(deep learning) 기술의 발달로, 여러 연구들(Lu & Huang, 2020; Lu *et al.*, 2019; Wang *et al.*, 2019; Yao *et al.*, 2017a)에서 POI 추천검색을 위한 딥러닝 모델의 훈련 성능을 높이기 위해 POI 임베딩을 활용하고 있다. 한편, 도시의 토지 이용 유형(type)을 파악하기 위한 목적의 연구들(Liu *et al.*, 2019; Yan *et al.*, 2017; Yao *et al.*, 2017b; Zhai *et al.*, 2019)에서는 주로 POI의 유형 정보를 활용해 POI를 임베딩하였다. 정리하자면, 추천검색 및 도시의 토지 이용 유형 분석을 목적으로 한 연구들에서는 POI 임베딩 시 체크인이나 유형 정보를 사용하기 때문에 POI에 대한 시맨틱 정보를 추출하지 못한다는 한계가 있다. 또한, 딥러닝 모델에 활용되는 POI 임베딩은 POI 추천검색을 위한 모델 훈련의 중간 단계로만 활용되며, 이 역시 POI의 시맨틱 정보를 표현하지 못한다.

POI는 체크인 및 유형 정보뿐 아니라 다양한 속성 정보를 포함하고 있는데, 그중에서도 사용자들이 작성한 리뷰 텍스트는 POI를 설명하는 가장 많은 양의 정보를 포함하고 있는 컨텍스트 데이터라고 할 수 있다. 사람들은 실세계에서 POI를 방문할 때 입소문이나 인기도의 영향을 많이 받는다 (Kefalas & Manolopoulos, 2017; Yin *et al.*, 2017; Zhang *et al.*, 2015). 모바일 관련 리서치<sup>10)</sup>에 의하면 소비자의 85%는 휴대폰에서 읽은 온라인

---

10) Quoracreative, <https://quoracreative.com/article/mobile-marketing-statistics> (접속 일: 2020년 10월 2일)

리뷰를 지인의 추천만큼 신뢰하며, 지역 검색을 하는 사용자 중 88%가 온라인에서 작성된 리뷰를 믿는 것으로 조사되었다<sup>11)</sup>. 이렇듯 음식점 등 방문을 원하는 POI에 대한 정보를 찾기 위해 스마트폰에서 관련 리뷰를 찾아보는 행위는 현대인의 일상이 되었다. 실제 사용자들은 POI에 대한 리뷰를 읽고 해당 POI를 방문할지 결정하기 때문에 리뷰 텍스트의 시맨틱 정보는 사용자의 POI 방문 여부에 영향을 미치는 중요한 요소라고 할 수 있다(Wang *et al.*, 2017; Yin *et al.*, 2017). 또한, POI 리뷰 데이터는 기존의 공간 데이터와 다르게 비정형의 시맨틱 정보를 포함하고 있는데, 데이터 자체가 사람들의 POI에 대한 인식, 감정, 정보 등의 다양한 측면을 직접적으로 나타낸 것이기 때문에 해당 POI를 대표하는 데이터로 볼 수 있다. 다양한 스마트폰 앱을 통해 작성되는 리뷰는 소셜 데이터와 유사하게 실시간적인 특징까지 갖추고 있으므로 POI에 대한 정보 제공 측면에서 활용 가치가 높은 데이터라고 할 수 있다.

종합하면, POI에 대해 작성된 리뷰 텍스트 데이터를 이용해 POI의 시맨틱 정보를 임베딩하거나, 임베딩된 POI를 시맨틱 검색에 활용하는 연구가 부족한 것을 확인할 수 있었다. 서비스 측면에서는 기존 서비스들이 POI 명칭, 카테고리, 또는 태그 위주의 검색을 우선 제공하고 있으며, 리뷰 검색 기능을 일부 지원하고는 있으나 POI의 시맨틱 정보는 활용하지 않기 때문에 사용자들의 다양한 요구사항을 반영할 수 있는 검색이 부족하다는 한계가 있었다. 따라서 본 연구에서는 기존 연구 및 서비스의 한계를 보완하기 위하여 리뷰 텍스트 데이터를 활용해 사용자들의 다양한 선호나 상황을 반영할 수 있는 시맨틱 POI 검색 기술을 개발하는 것을 목표로 하였다.

---

11) Brightlocal, [https://www.brightlocal.com/research/local-consumer-review-survey-2014/?SSAID=389818&SSCID=81k1\\_7ofbr](https://www.brightlocal.com/research/local-consumer-review-survey-2014/?SSAID=389818&SSCID=81k1_7ofbr) (접속일: 2020년 10월 2일)

## 1.2 연구 동향

최근 자연어 처리(natural language processing, NLP) 분야에서 Word2vec(Mikolov *et al.*, 2013a; Mikolov *et al.*, 2013b)으로 대표되는 신경망 기반 단어 임베딩 기법이 성공적으로 자리 잡으면서, 임베딩 결과를 활용한 다양한 연구가 수행되고 있다. 그중에서도 POI 관련 분야에서는 추천검색 시스템에 활용하기 위한 목적, 또는 도시의 토지 이용 현황을 파악하기 위한 목적으로 POI를 임베딩 하는 연구가 이뤄지고 있다.

POI를 추천검색 시스템에 활용하기 위한 연구로, 실제 사용자들이 POI에 체크인한 시퀀스 데이터를 이용해 POI를 임베딩한 연구들(Feng *et al.*, 2017; Liu *et al.*, 2016; Zhao *et al.*, 2017)이 수행되었다. 이러한 연구들은 사용자가 순차적으로 POI에 체크인한 일련의 기록을 문장으로, 각 POI는 해당 문장에 포함된 단어로 치환해 Word2vec 모델링을 적용한다. Liu *et al.*(2016)은 POI 추천검색을 위해 Word2vec의 skip-gram 모델을 이용해 POI를 벡터로 임베딩하였다. 해당 연구에서는 시퀀스 데이터를 다룰 때 기존에 주로 사용하던 마르코프 체인(Markov chain) 모델에 비해 신경망 기반 단어 임베딩 모델의 성능이 더욱 뛰어남을 밝혔으며, 이에 따라 Feng *et al.*(2017)은 Liu *et al.*(2016)의 연구에 지리적 영향을 통합한 임베딩 모델인 POI2vec을 제안하였다. 더 나아가 Zhao *et al.*(2017)은 Liu *et al.*(2016)의 연구에 지리적 영향 및 시간적 특성을 추가하여 POI를 임베딩하였다. 또한, He *et al.*(2019)은 여행 POI를 추천하기 위해 플리커(Flickr)<sup>12)</sup> 체크인 데이터를 이용해 Word2vec 모델을 기반으로 POI를 임베딩하였다. Word2vec 모델링 시 전체 여행 중에 방문한 POI 시퀀스를 문장으로, 각 POI를 단어로 치환했는데, 이는 여행 중 방문한 POI는 같은 여행(문장) 중에 함께 방문한 POI(단어)에 의해 결정된다고 가정한 것이다.

사용자 체크인 데이터를 이용해 POI와 사용자를 임베딩함으로써 이를 POI 추천검색에 활용하는 연구들(Qiao *et al.*, 2020; Yang & Eickhoff,

---

12) 플리커, <https://www.flickr.com/> (접속일: 2020년 10월 24일)

2018; Zhao *et al.*, 2018)도 진행되었다. Yang & Eickhoff(2018)는 대표적인 위치 기반 소셜 네트워크(location-based social networks, LBSN) 서비스인 포스퀘어(Foursquare)<sup>13)</sup>의 체크인 데이터를 이용해 POI를 임베딩한 후, 임베딩한 POI 벡터를 이용해 체크인 벡터를 생성하였으며, 마지막으로 체크인 벡터의 평균을 통해 POI와 사용자를 각각 임베딩하였다. Zhao *et al.*(2018)과 Qiao *et al.*(2020)은 체크인 시퀀스 데이터를 이용해 Word2vec 모델을 기반으로 POI와 사용자를 동시에 임베딩하는 방법을 제안하였다.

최근에는 딥러닝 기술의 발전으로 여러 연구(Lu & Huang, 2020; Lu *et al.*, 2019; Wang *et al.*, 2019; Yao *et al.*, 2017a)에서 POI 추천검색에 순환 신경망(recurrent neural network, 이하 RNN), 장단기 메모리(long short-term memory, 이하 LSTM) 등의 딥러닝 기술을 활용하고 있다. 심층 신경망(deep neural network, DNN)의 훈련 성능을 높이기 위해서는 입력 데이터의 특징(feature)에 대한 효과적인 벡터 표현을 확보하는 것이 필수적이기 때문에 임베딩 과정이 중요하게 다뤄진다(Covington *et al.*, 2016; Ding & Chen, 2018; Liang *et al.*, 2016). Yao *et al.*(2017a)은 사용자가 다음에 방문할 POI를 예측하기 위해 POI에 체크인한 시간, 위치 등의 데이터를 이용해 POI를 임베딩한 후, 이를 RNN의 입력 데이터로 활용하였다. Lu *et al.*(2019)은 Word2vec 모델을 기반으로 사용자의 연속적인 체크인 데이터를 이용해 POI를 임베딩하고, 사용자가 방문한 POI 시퀀스를 이용해 사용자를 임베딩한 후 이들을 POI 추천검색을 위한 RNN 모델의 학습 데이터로 사용하였다. Wang *et al.*(2019) 역시 Word2vec 모델을 이용해 체크인 시퀀스를 POI로 임베딩해 이를 RNN의 입력 데이터로 활용하였다. Lu & Huang(2020)은 POI의 유형 및 체크인 데이터를 이용해 POI를 임베딩하고, 이를 기반으로 LSTM을 사용해 POI를 추천검색에 활용하는 방법을 제안하였다.

지금까지 살펴본 바에 의하면 POI 추천검색을 목표로 하는 다수의 연구에서 POI를 임베딩하기 위해 주로 사용자의 체크인 데이터를 이용하고 있는 것을 알 수 있다. 그러나 시퀀스를 기반으로 한 체크인 데이터는 POI의

---

13) 포스퀘어, <https://foursquare.com/> (접속일: 2020년 10월 24일)

시맨틱 정보를 포함하지 않기 때문에 사용자 행위의 의미를 파악하지 못한다는 한계가 있다. 딥러닝 모델에 활용되는 POI 임베딩 역시 POI 추천 검색을 위한 중간 단계로만 활용되며, POI의 시맨틱 정보를 표현하지는 못한다. POI에 대한 시맨틱 정보를 추출할 수 있는 대상은 주로 텍스트 데이터로, POI에 포함된 텍스트 데이터를 POI 임베딩에 활용하는 연구들 (Chang *et al.*, 2018; Shoji *et al.*, 2018)이 수행된 바 있다. Chang *et al.*(2018)은 인스타그램(Instagram)<sup>14)</sup> 체크인 데이터에서 체크인 이력과 텍스트 데이터를 이용해 Word2vec 기반으로 POI를 임베딩하였다. 해당 연구에서는 사용자들이 POI에 체크인할 때 함께 작성한 텍스트 데이터를 활용해 POI를 임베딩했다는 점에서 POI의 시맨틱 정보를 표현하였다고 볼 수 있다. 그러나 각 단어를 단일 벡터로 표현하였기 때문에 동음이의어나 다의어를 구별하지 못하며, 임베딩된 POI는 사용자가 다음에 방문할 가능성이 높은 POI를 예측하기 위한 POI의 추천검색에만 사용되었을 뿐, 이를 직접검색에 활용하지는 못하였다는 한계가 있다. Shoji *et al.*(2018)은 위치 태그가 포함된 트위터(Twitter)<sup>15)</sup> 텍스트 데이터를 이용해 위치를 벡터로 임베딩하였는데, Word2vec 모델 적용 시 각 트위터 텍스트가 작성된 위치를 중심 단어로, 해당 위치 주변에 작성된 트위터 텍스트를 주변 단어로 치환하였다. 이는 Word2vec의 기본 개념에 따라 어떤 위치 주변에 작성된 텍스트가 다른 위치의 텍스트와 유사하면, 해당 위치들이 서로 유사한 것으로 간주한다는 점에서 위치에 대한 시맨틱 정보를 나타낸다고 볼 수 있다. 그러나 Shoji *et al.*(2018)은 Chang *et al.*(2018)의 연구와 마찬가지로 각 단어를 하나의 벡터로 나타냈기 때문에 임베딩 시 동음이의어 및 다의어를 구별하지 못하며, 특정 POI가 아닌 추상적인 위치를 임베딩함으로써 정확성과 활용성이 떨어진다는 한계점이 있다.

도시의 토지 이용 유형 및 기능 구역을 파악하기 위해 POI를 벡터 공간 상에 표현하는 연구들은 인간 활동이 도시 지역의 기능을 나타내는 POI에서 발생한다는 점에 착안해 POI를 분석함으로써 도시 구조를 이해하려는

---

14) 인스타그램, <https://www.instagram.com/> (접속일: 2020년 10월 24일)

15) 트위터, <https://twitter.com/> (접속일: 2020년 10월 24일)

목적은 가지고 있다. 해당 분야에서는 최근, NLP 분야의 단어 임베딩 기술을 도입해 POI 유형의 의미 관계를 측정하는 방향으로 연구(Liu *et al.*, 2019; Yan *et al.*, 2017; Yao *et al.*, 2017b; Zhai *et al.*, 2019)를 수행하고 있는 추세이다. Yao *et al.*(2017b)은 도시의 토지 이용 현황을 파악하기 위해 POI 유형의 빈도수만 취하던 기존 연구(Gao *et al.*, 2017; Liu *et al.*, 2015a; Yuan *et al.*, 2012; Zhong *et al.*, 2014)에서 벗어나, POI의 공간적 관계를 탐색하기 위해 Word2vec을 최초로 사용하였다. Word2vec 모델링 시 지역을 문서로, 그 안에 존재하는 POI 유형의 공간적 분포를 문서 내 단어 시퀀스로 간주하였다. 이때 토지 이용의 공간적 분포를 교통 분석 구역(traffic analysis zone, 이하 TAZ) 단위로 파악하였으며, 각 TAZ에 속하는 POI 벡터 평균의 합을 통해 TAZ 벡터를 추정하였다. Yan *et al.*(2017)은 POI 유형을 임베딩하기 위해 Word2vec을 공간적으로 확장한 모델인 Place2vec을 제안하였다. Word2vec이 학습을 위해 문장에서 윈도우(window) 크기를 기반으로 중심 단어와 주변 단어 쌍(pair)을 구성하는 것과 유사하게, Place2vec은 실세계에서 거리를 기반으로 중심 POI 유형과 가깝게 위치한 POI 유형을 주변 컨텍스트로 고려해 공간적 쌍을 구성한다. Place2vec은 Zhai *et al.*(2019)에 의해 도시의 기능 영역을 식별하는데 활용되기도 하였다. Zhai *et al.*(2019)의 연구에서는 Yao *et al.*(2017b)에서 사용한 TAZ보다 작은 단위인 이웃 영역(neighborhood area, 이하 NA) 규모의 기능 영역을 탐지하기 위해 NA를 임베딩하였는데, 이때 전체 도시를 거대한 말뚱치로 보고, 각 NA는 문서, NA 내 위치한 POI 유형은 단어로 간주하였다. Liu *et al.*(2019)은 주요 장소에 근접한 장소들의 집합이 지역 환경의 특징을 나타낸다고 보고 이것의 변동을 POI 유형을 통해 파악하고자 했으며, 이를 위해 Word2vec 기반으로 POI 유형을 임베딩하였다.

[표 1-2] POI 임베딩 관련 연구 요약

관련 연구	활용 데이터	임베딩 관련 활용 기술	연구 목적
Liu <i>et al.</i> (2016), Feng <i>et al.</i> (2017), Zhao <i>et al.</i> (2017), Yang & Eickhoff (2018), Zhao <i>et al.</i> (2018), Qiao <i>et al.</i> (2020), He <i>et al.</i> (2019), Yao <i>et al.</i> (2017a), Lu <i>et al.</i> (2019), Wang <i>et al.</i> (2019), Lu & Huang(2020)	POI 체크인	Word2vec	POI 추천검색
Chang <i>et al.</i> (2018)	POI 체크인 및 텍스트	Word2vec	POI 추천검색
Shoji <i>et al.</i> (2018)	위치 태그된 트위터 텍스트	Word2vec	POI 추천검색
Yan <i>et al.</i> (2017), Yao <i>et al.</i> (2017b), Zhai <i>et al.</i> (2019), Liu <i>et al.</i> (2019)	POI 유형	Word2vec	토지 이용 현황 분석
본 연구	POI 리뷰 텍스트	Word2vec	시맨틱 POI 검색 (직접검색+추천검색)

앞에서 살펴본 POI 임베딩 관련 선행 연구들을 요약 및 정리한 결과는 [표 1-2]와 같다. 정리하자면, POI 추천검색을 목표로 하는 다수의 연구들에서는 POI 체크인 데이터를 사용하고 있으며, 도시의 토지 이용 유형을 파악하기 위한 목적의 연구들에서는 POI 유형 데이터를 주로 사용하고 있다. 체크인 시퀀스를 기반으로 임베딩된 POI는 함께 체크인된 횟수가 많을수록 유사하기 때문에 POI 간 유사성에 기반한 추천검색이 일부 가능하다. 하지만 텍스트로부터 추출된 시맨틱 정보를 포함하지 않기 때문에 검색 가능한 유사성의 범위가 한정적이다. 또한, ‘분위기 좋은 카페’와 같은 자연어 기반 검색어가 입력되었을 때 검색어와 POI 간 관련성을 파악할 수 없으므로 관련성 기반의 직접검색이 불가능하다. POI 유형을 기반으로 임베딩된 POI는 유사한 유형의 POI일수록 벡터 공간상에 가깝게 위치할 확률이 높기에 POI 간 유사성에 기반한 추천검색이 일부 가능하다. 그러

나 텍스트 데이터에서 도출된 시맨틱 정보를 포함하지 않기 때문에 유사성 검색의 범위가 POI 유형에 한정된다는 한계가 있다. 또한, 검색어와 POI 간 관련성에 기반한 직접검색 역시 불가능하다.

POI는 체크인이나 유형 정보뿐 아니라 다양한 속성 정보를 포함하고 있으며, 그중에서도 사용자들이 작성한 리뷰 텍스트는 POI에 대한 가장 많은 양의 정보를 포함하고 있는 데이터라고 할 수 있다. 그럼에도 불구하고 지금까지 조사된 바에 의하면 첫째, POI에 대해 작성된 리뷰 텍스트 데이터를 이용해 POI의 시맨틱 정보를 임베딩하는 연구와 둘째, 임베딩된 POI를 시맨틱 검색, 특히 검색어와 POI 간 관련성에 기반한 직접검색에 활용하는 연구가 부족한 것을 확인할 수 있었다. 본 연구에서는 리뷰 텍스트 데이터를 활용해 시맨틱 POI를 생성하는 프로세스를 제안하고, 시맨틱 POI를 검색에 활용하기 위해 자연어 기반 POI 검색, 선택 기반 POI 검색, 그리고 유사 POI 검색에 대한 활용 구조를 설계한다. 본 연구에서 개발한 시맨틱 POI 검색 기술은 리뷰 데이터에서 추출한 시맨틱 정보를 기반으로 하므로 사용자들의 다양한 선호와 상황을 반영한 검색이 가능하며, 기존의 POI 검색 시스템과 상호 보완적으로 사용될 수 있다.

### 1.3 연구 범위 및 방법

본 연구는 사용자들의 다양한 선호와 상황을 반영할 수 있는 토픽별 키워드 기반 시맨틱 POI 검색 기술 개발을 목표로 한다. 이를 위해 리뷰 데이터를 활용해 시맨틱 POI를 생성하는 프로세스를 제안하고, 시맨틱 POI를 검색에 활용할 수 있는 세 종류의 활용 구조를 설계한다. 시맨틱 POI 검색을 위한 활용 구조는 자연어 기반 검색어에 대한 결과를 제공하는 자연어 기반 POI 검색, 사용자가 토픽과 토픽별 키워드를 선택해 검색하는 선택 기반 POI 검색, 그리고 특정 POI와 유사한 POI를 탐색할 수 있는 유사 POI 검색을 포함한다. 여기서 자연어 및 선택 기반 POI 검색은 사용자가 직접 검색어를 입력/선택하기 때문에 직접검색 방식에 해당하며, 유사 POI 검색은 사용자가 입력/선택한 결과로 도출된 POI를 기준으로 유사 POI를 추천해 주기 때문에 추천검색 방식에 해당한다. 특히, 자연어 및 선택 기반 POI 검색 환경에서는 사용자가 검색어에 대한 가중치를 직접 지정할 수 있으므로 같은 검색어라도 가중치에 따라 사용자별로 다른 결과가 도출된다. 따라서 사용자는 자신이 더 중요하게 생각하는 검색어에 높은 가중치를 부여함으로써 맞춤형 검색이 가능하다.

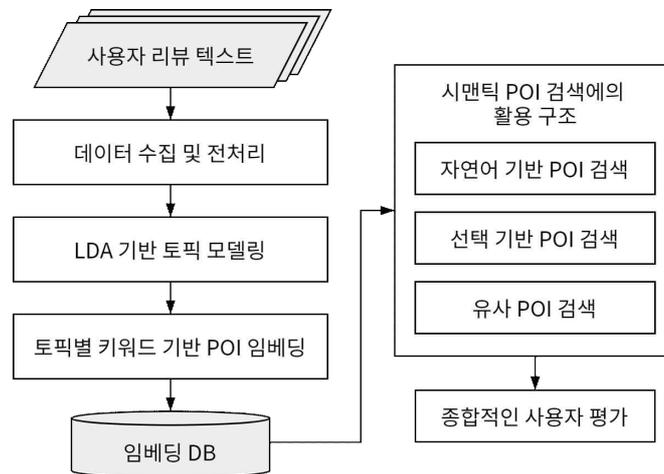
본 연구에서 사용한 데이터는 POI에 대해 한국어로 작성된 실제 사용자 리뷰 텍스트 데이터로, 플랫폼을 통해 직접 수집한 데이터를 활용한다. 데이터의 공간적 범위는 서울시로 한정하며, 데이터 수집을 위한 시간적 범위는 2019년 1월 14일부터 약 2주 간이다. 리뷰 텍스트 데이터는 리뷰를 하는 특정 대상에 대한 사람들의 직접적인 인식, 감정, 정보 등이 종합적·지속적으로 누적된 것이라는 점에서 단발성의 소개글, 객관적 사실을 다루는 뉴스기사 등의 다른 텍스트 데이터와 차별된다. 즉, 각 POI별 리뷰 데이터는 하나의 공통 대상인 특정 POI에 대한 여러 사람들의 의견이 누적된 것이기 때문에 POI와 관련된 특정 토픽들을 추출하기에 적합하다(Jin *et al.*, 2018). 예를 들어, 맛집 관련 POI에 대한 리뷰 데이터에는 음식, 분위기, 가격 등에 대한 내용이 반복적으로 등장할 확률이 높다.

POI 리뷰 데이터로부터 POI를 대표하는 토픽과 토픽별 키워드를 추출

하기 위한 방법으로, 텍스트 데이터에 내재된 토픽을 자동으로 찾아내는 토픽 모델링 기법을 적용하는 것을 제안한다. 본 연구에서는 여러 토픽 모델링 방법 중 대규모 비정형 데이터에서 토픽을 도출하는 데 적합한 것으로 알려진(Steyvers & Griffiths, 2007) 잠재 디리클레 할당(latent Dirichlet allocation, 이하 LDA)(Blei *et al.*, 2003) 기법을 활용한다. 또한, POI별로 누적된 리뷰 데이터 집합은 해당 POI를 둘러싼 컨텍스트 정보로 볼 수 있다(Chang *et al.*, 2018; Yan *et al.*, 2017). 컨텍스트 정보로부터 하나의 통합된 시맨틱 POI를 추출하기 위한 방법으로, 텍스트 데이터의 의미와 맥락을 고려해 각 키워드를 벡터 공간상의 한 점으로 대응시키는 임베딩 기법을 적용하는 것을 제안한다. 본 연구에서는 대표적인 임베딩 기법인 신경망 기반의 Word2vec 모델을 개선해, 토픽과 키워드를 모두 학습하도록 한 토픽 기반 단어 임베딩(topical word embedding, 이하 TWE)(Liu *et al.*, 2015b) 기법을 활용한다. TWE 방법을 이용한 임베딩을 통해 같은 키워드라도 토픽에 따라 다르게 사용되는 키워드의 다양한 의미를 반영할 수 있다.

특정 POI에 대해 작성된 리뷰 데이터에서 어떤 키워드가 빈번하게 사용됐다는 것은 그 키워드가 해당 POI의 중요한 특성을 나타낸다는 사실이 다수의 사람들에게 의해 검증된 것으로 간주할 수 있다. 그러므로 리뷰 데이터를 기반으로 시맨틱 POI를 생성해 검색에 활용한다면 다른 사람들의 평가를 기반으로 POI에 대한 의사결정을 하고자 하는 사용자들에게 유용한 정보를 제공할 수 있다. 본 연구에서는 시맨틱 POI를 검색에 활용한 결과에 대한 품질을 확인하기 위하여 설문 조사 기반의 사용자 만족도 평가 방법을 사용한다. 또한, 기존 유사 서비스들과의 검색 결과 비교를 통해 본 연구 결과의 차별성을 확인하고자 하였다. 본 연구에서 개발한 토픽별 키워드 기반 시맨틱 POI 검색 기술은 웹 기반 POI 검색 플랫폼에서 사용할 수 있는 응용적 범위를 갖는다.

본 연구의 흐름도는 [그림 1-5]와 같다. 먼저 사용자가 작성한 리뷰 텍스트 데이터를 수집하고, 향후 분석을 용이하게 하기 위해 수집한 데이터를 전처리한다. 그리고 LDA 기반 토픽 모델링을 통해 리뷰 텍스트로부터 POI에 대한 토픽과 토픽별 키워드를 추출하고, 추출된 토픽별 키워드를 기반으로 POI를 임베딩한다. 다음으로, 시맨틱 POI 검색에의 활용 구조로써 자연어 기반 POI 검색, 선택 기반 POI 검색, 그리고 유사 POI 검색을 제안하고, 마지막으로 종합적인 사용자 평가를 실시한다.

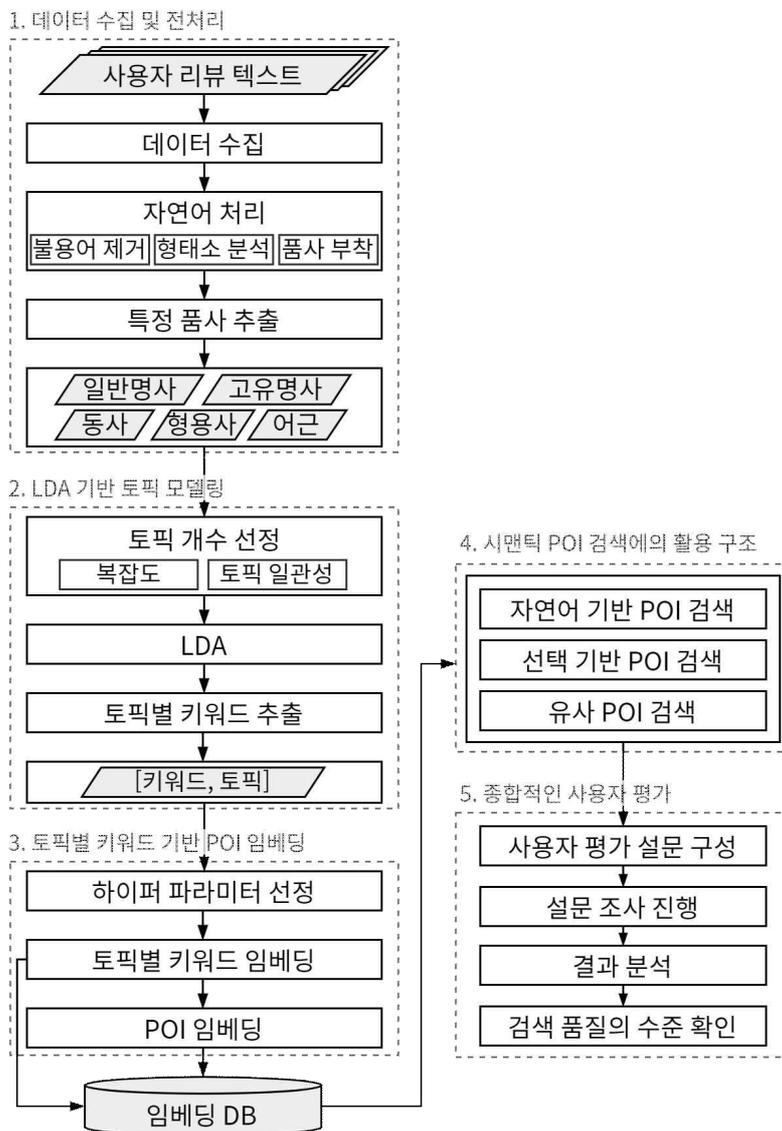


[그림 1-5] 연구 흐름도

본 연구의 구성은 다음과 같다. 1장에서는 연구 배경 및 목적과 연구 동향에 대해 고찰하고, 본 연구의 범위와 방법을 구체화하였다. 2장에서는 연구 방법인 데이터 수집 및 전처리 방법, LDA 기반 토픽 모델링 방법, 토픽별 키워드 기반 POI 임베딩 방법을 서술한다. 또한, 시맨틱 POI 검색에 대한 활용 구조로써 자연어 기반 POI 검색, 선택 기반 POI 검색, 그리고 유사 POI 검색에 관한 내용을 설명하고, 마지막으로 종합적인 사용자 평가 방법에 대해 기술한다. 3장에서는 실제 데이터에 적용한 결과를 각 프로세스에 따라 분석하고, 본 연구에서 개발한 토픽별 키워드 기반 시맨틱 POI 검색 결과에 대한 종합적인 사용자 평가를 실시한 결과를 기술한다. 마지막으로 4장에서는 본 연구의 결과를 요약하고, 그 의미와 향후 연구 계획을 종합적으로 서술한다.

## 2. 연구 방법

이 장에서는 토픽별 키워드 기반 시맨틱 POI 검색을 위한 연구 방법에 대해 자세히 설명한다. [그림 2-1]은 전체적인 연구의 이해를 돕기 위해 연구 방법의 흐름도를 나타낸 것이다.



[그림 2-1] 연구 방법의 흐름도

## 2.1 데이터 수집 및 전처리

### 2.1.1 데이터 수집

사용자들이 작성한 리뷰 데이터를 수집하기 위해 국내 맛집 검색 및 추천 서비스인 ‘망고플레이트’를 활용한다([그림 2-2]). 다양한 POI 관련 서비스 중 맛집에 특화된 플랫폼을 선택한 이유는 여러 POI 종류 중 사용자 검색 수요가 가장 많은 것이 맛집이기 때문이다. 망고플레이트는 국내 최대 규모의 사용자 참여형 맛집 추천 플랫폼으로, 맛집 약 20만 곳에 대한 사용자 리뷰 70만 건을 보유하고 있다. 리뷰 데이터는 서비스를 제공하는 업체가 특정 사용자에게 대가를 지불하는 등의 행위 등을 통해 허위로 작성될 여지가 있다. 그러나 리뷰 데이터의 진위 여부를 판단하는 것은 본 연구의 범위에 포함되지 않으며, 본 연구에서는 데이터를 수집한 플랫폼에서 작성된 리뷰가 사실이라는 가정하에 연구를 진행한다. 망고플레이트는 자체 개발한 알고리즘으로 광고성 콘텐츠나 허위 리뷰를 걸러내는 전략을 내세우며 사용자들로부터 좋은 평가를 받고 있기에<sup>16)</sup> 리뷰 데이터를 수집하기 적합한 플랫폼이라 판단하였다.

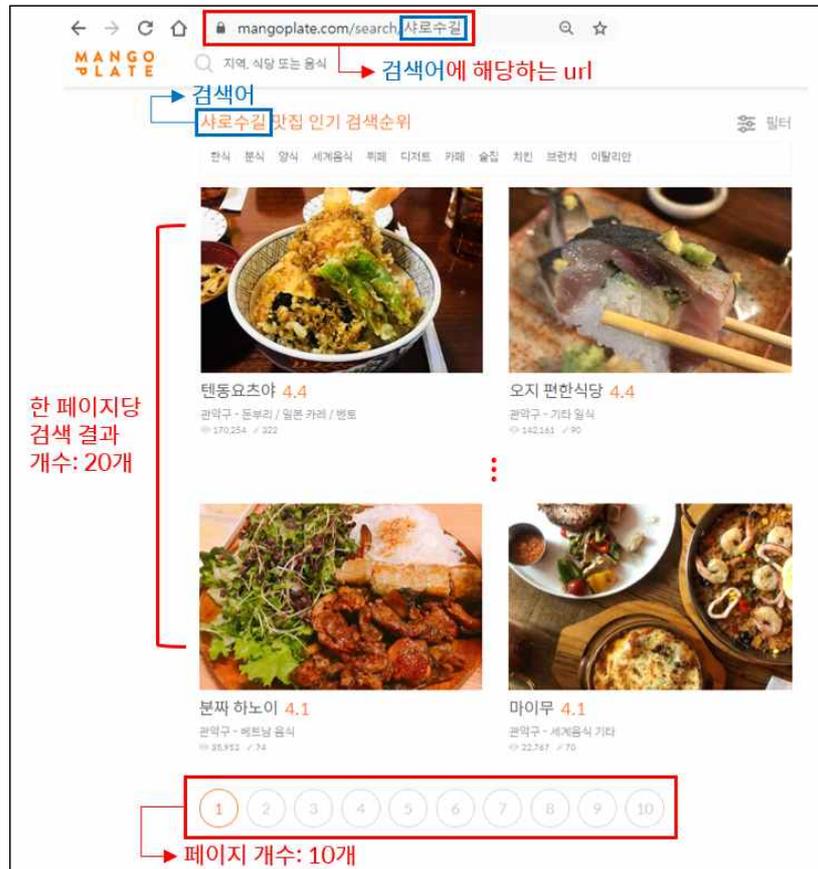


[그림 2-2] 망고플레이트 웹 사이트 메인 화면

16) 매일경제(2020-08-20), <https://www.mk.co.kr/news/it/view/2020/08/856654/> (접속일: 2020년 10월 5일)

망고플레이트에서 제공하는 POI 리뷰 관련 정보는 비정형 텍스트 형식이며, 이를 자동화된 방식으로 수집하기 위해 웹 스크래핑(web-scraping) 방식을 사용한다. 웹 사이트의 텍스트 데이터는 오픈 어플리케이션 프로그래밍 인터페이스(open application programming interface, 이하 오픈 API)나 웹 크롤링(web-crawling) 또는 웹 스크래핑 방식을 사용해 자동으로 수집할 수 있다. 그런데 현재 망고플레이트에서는 오픈 API를 제공하지 않기 때문에 자동화 방식으로 리뷰 데이터를 수집하기 위해서는 웹 크롤링이나 웹 스크래핑 방법을 사용해야 한다. 웹 크롤링은 자동화 봇(bot)이 정해진 규칙에 따라 여러 사이트를 돌아다니면서 정보를 추출하는 기술로, 주로 웹 페이지에 존재하는 모든 데이터를 수집한다. 반면 웹 스크래핑 방식은 웹 페이지 내에서 수집을 원하는 특정 위치의 정보만을 선택적으로 추출한다. 본 연구에서는 POI에 대한 모든 정보가 필요한 것이 아니라 POI 명칭, 주소, 음식 종류, 리뷰 텍스트에 대한 데이터만 수집하면 되므로 웹 스크래핑 방식을 활용한다.

망고플레이트에서 웹 스크래핑 기법을 이용해 POI 관련 정보를 수집하기 위해서는 먼저, 수집할 웹 페이지의 주소(url)를 알아야 하는데, 이는 망고플레이트 웹 사이트에서 지역, 식당 또는 음식에 해당하는 검색어를 입력하면 확인할 수 있다. 이때 검색 결과로 출력되는 POI 개수는 한 페이지당 20개, 페이지 개수는 최대 10개로 제한되므로, 1개의 검색어로 얻을 수 있는 POI 데이터의 개수는 최대 200개로 한정된다([그림 2-3] 참조).



[그림 2-3] 망고플레이트 검색 결과 화면

검색 결과로 출력된 POI들은 고유 주소를 가지고 있으므로 검색어 1개 당 최대 200개의 POI에 대한 하이퍼 텍스트 마크업 언어(hypertext markup language, 이하 HTML)를 수집할 수 있으며, 이를 이용해 원하는 정보를 추출할 수 있다. 즉, HTML 문서의 문법적 요소인 태그(tag)를 이용해 POI 명칭, 주소, 음식 종류, 리뷰 텍스트에 해당하는 정보를 추출하는 것이다. 예를 들어 [그림 2-4]와 [그림 2-5]는 ‘샤로수길’을 검색한 결과로 반환된 POI 중 하나인 ‘텐동요츠야’라는 음식점에 대한 HTML 정보를 열어본 것이다. [그림 2-4]를 보면, ‘h1’이라는 하이퍼링크(hyperlink) 태그에 ‘class’ 속성명이 ‘restaurant\_name’인 부분에 POI 명칭인 ‘텐동요츠야’가 들어가 있는 것을 확인할 수 있다. 또한, [그림 2-5]에서는 ‘p’라는 태그에 ‘class’ 속성명이 ‘RestaurantReviewItem\_ReviewText’인 부분에 사용자가 작성한 리뷰 텍스트에 해당하는 내용이 들어가 있는 것을 알 수 있다.

**텐동요츠야 4.4**

170,266 / 322 / 5,831

주소 서울특별시 관악구 남부순환로234길 18 1F  
지번 서울시 관악구 봉천동 1603-19 1F

전화번호 02-883-7974

음식 종류 돈부리 / 일본 카레 / 벤토

가격대 만원 미만

주차 주차공간없음

영업시간 12:00 - 21:00

쉬는시간 14:30 - 17:00

```

Elements Console Sources Network Performance Memory Application Set
<div class="column-contents">
  <div class="inner">
    <!-- 레스토랑 상세 -->
    <section class="restaurant-detail">
      <header>
        <div class="restaurant_title_wrap">
          <span class="title">
            <h1 class="restaurant_name">텐동요츠야</h1> == $0
          <strong class="rate-point">...</strong>
          <p class="branch"></p>
        </span>
        <div class="restaurant_action_button_wrap">...</div>
      </div>
    </section>
  </div>
</div>

```

[그림 2-4] POI 명에 대한 HTML

리뷰 (322) 전체 (322) | 맛있다 (237) | 괜찮다 (64) | 별로 (21)

2인 1차

제가 먹으러 갔을 때만 해도 웨이팅이 조금 있었는데 요새 가면 웨이팅 거의 없이 먹을 수 있다고 한  
당!! 정말 오랜만에 먹었는데 여전히 느끼하고 맛있었당 ㅎㅎㅎ 먹을 때마다 건강 안 좋아지는 느낌  
드는데 너무 맛있어서 그만 먹을 수가 없...

맛있다

1876 & 424

Elements Console Sources Network Performance Memory Application Security Lighthouse

```

<div class="RestaurantReviewList_ReviewList">
  <li class="RestaurantReviewItem RestaurantReviewList_ReviewItem">
    <a class="RestaurantReviewItem_Link" href="/reviews/Us8Njg2" target="_blank">
      <div class="RestaurantReviewItem_User">...</div>
      <div class="RestaurantReviewItem_ReviewContent">
        <div class="RestaurantReviewItem_ReviewText">
          <p class="RestaurantReviewItem_ReviewText"> == $0
            "
            제가 먹으러 갔을 때만 해도 웨이팅이 조금 있었는데 요새 가면 웨이팅 거의 없이 먹을 수 있다고 한당!! 정말 오랜
            만에 먹었는데 여전히 느끼하고 맛있었당 ㅎㅎㅎ 먹을 때마다 건강 안 좋아지는 느낌 드는데 너무 맛있어서 그만 먹을 수가
            없...
            "
          </p>
          <span class="RestaurantReviewItem_ReviewDate">2020-09-18</span>
        </div>
      </div>
    </li>
  </div>

```

[그림 2-5] 리뷰 텍스트에 대한 HTML

## 2.1.2 데이터 전처리

데이터 전처리는 앞서 수집한 텍스트 데이터에 자연어 처리 기법을 적용해 향후 분석이 가능한 형태로 변환하는 과정이다. 보통 한국어 기반 텍스트 데이터에 대한 전처리 작업은 크게 불용어(stopwords) 제거 단계와 형태소 분석 단계로 나뉘며, 순차적으로 진행된다.

불용어는 일반적으로 특수문자(~!@#\$%^&\*() 등)나 문장부호(마침표(.), 쉼표(,), 물음표(?) 등)와 같이 분석에서 의미를 부여할 필요가 없으므로 제거해야 하는 것들을 일컬으며, 이는 분석 목적에 따라 다르게 설정할 수 있다. 본 연구에서는 Python에서 제공하는 정규 표현식(regular expression) 패키지인 're'를 이용해 분석에 불필요하다고 판단되는 숫자, 영어, 특수문자, 문장부호를 제거한다.

데이터 전처리 과정에서 형태소 분석을 하는 이유는 한국어가 교착어<sup>17)</sup>라는 특징을 가지기 때문이다. 텍스트 데이터를 다루는 연구는 각 나라의 언어별로 활발히 진행되고 있는데 그중에서도 가장 많은 연구가 이루어진 언어는 영어다. 영어는 굴절어<sup>18)</sup>의 특성을 가지고 있기 때문에, 영어 기반 텍스트 데이터를 처리할 때는 띄어쓰기 기준으로 단어를 분리해도 무리가 없다. 그러나 조사와 어미가 발달한 한국어를 공백 기준으로 분리하면 동일한 의미의 단어를 다르게 구분할 가능성이 크다. 예를 들어, '카페에서'와 '카페는'은 모두 '카페'라는 형태소를 포함하고 있지만 서로 다른 조사가 붙어 있기 때문에 공백 기준으로 분리 시 다른 형태의 단어로 인식하게 된다(최상혁 등, 2016). 따라서 한국어 기반 텍스트 데이터를 다룰 때는 형태소 단위로 분리해 활용하는 것이 좋다(조현수 & 이상구, 2017; 최상혁 등, 2016).

---

17) 언어의 형태적 유형의 하나. 실질적 의미를 가진 단어 또는 어간에 문법적 기능을 가진 요소가 차례로 결합함으로써 문장 속에서의 문법적 역할이나 관계의 차이를 나타내는 언어로, 한국어, 터키어, 일본어, 핀란드어 등이 여기에 속함(네이버 표준국어대사전)

18) 형태론적 특징으로 본 언어의 한 유형. 어형과 어미의 변화로 단어가 문장 속에서 가지는 여러 가지 관계를 나타내는 언어로, 인도 및 유럽 어족에 속한 대부분의 언어가 이에 속함(네이버 표준국어대사전)

본 연구에서는 리뷰 텍스트의 형태소 분석을 위해 Python에서 제공하는 KoNLPy(Korean natural language processing in Python)<sup>19)</sup>의 코모란 형태소 분석기(Korean morphological analyzer)<sup>20)</sup>를 활용한다. 한국어 형태소 분석이 가능한 형태소 분석기는 코모란, 꼬꼬마<sup>21)</sup>, 한나눔<sup>22)</sup>, 오픈 한국어 텍스트(open-Korean-text)<sup>23)</sup> 등 다양하게 존재하는데, 여러 형태소 분석기 중 코모란을 사용한 이유는 본 연구에서 수집한 사용자 리뷰 텍스트 중 임의로 추출한 1,000개 문장에 대해 형태소 분석기별로 형태소 분류 작업을 진행한 결과, 코모란의 성능이 가장 우수하였기 때문이다. 특히 코모란은 타 형태소 분석기에 비해 어근과 고유명사를 잘 추출한다는 장점이 있다.

형태소 분석기는 주어진 문장을 형태소 단위로 나눠 준다. 즉, 어근과 조사, 어미 등을 분리하고, 어근도 명사, 동사, 형용사 등으로 분리시킨다. 코모란 형태소 분석기는 일반명사(NNG), 고유명사(NNP) 등 총 43개 품사 종류를 포함한다<sup>24)</sup>. 본 연구에서는 형태소 분석 과정을 마친 리뷰 텍스트 데이터에 LDA 기반 토픽 모델링 기법을 적용하기 위해 독립적으로 의미를 가지지 못하는 형태소를 제거하고, 체언과 용언처럼 실질적 의미를 가지는 형태소 단위만을 분석 대상으로 한다. 따라서 전체 43개 품사 중 체언에 해당하는 일반명사(NNG)와 고유명사(NNP), 용언에 해당하는 동사(VV)와 형용사(VA), 그리고 어근(XR) 품사에 해당하는 단어만을 추출해 분석에 활용한다([표 2-1] 참조).

---

19) KoNLPy, <http://konlpy.org/en/latest/> (접속일: 2020년 10월 5일)

20) 코모란, <https://www.shineware.co.kr/products/komorant/> (접속일: 2020년 10월 5일)

21) 꼬꼬마, <http://kkma.snu.ac.kr/> (접속일: 2020년 10월 5일)

22) 한나눔, <http://semanticweb.kaist.ac.kr/hannanum/index.html> (접속일: 2020년 10월 5일)

23) 오픈 한국어 텍스트, <https://github.com/open-korean-text/open-korean-text> (접속일: 2020년 10월 5일)

24) 코모란 문서, <https://docs.komorant.kr/firststep/postypes.html> (접속일: 2020년 10월 5일)

[표 2-1] 코모란 형태소 분석기의 품사 종류 일부

대분류	소분류	세분류	
		품사	태그
체언	명사	일반명사	NNG
		고유명사	NNP
		의존명사	NNB
	대명사	대명사	NR
		수사	NP
용언	동사	동사	VV
	형용사	형용사	VA
	보조용언	보조용언	VX
	지정사	긍정지정사	VCP
		부정지정사	VCN
의존형태	어근	어근	XR

형태소 분석을 통해 특정 품사(일반명사, 고유명사, 동사, 형용사, 어근)만 추출한 결과가 LDA 모델의 입력값으로 활용되는데, 추후 분석을 용이하게 하기 위해 형태소와 해당 품사를 밑줄 기호인 '\_'로 연결해 하나의 단어로 취급할 수 있도록 한다(예: '분위기\_NNG', '짜장면\_NNG', '좋\_VA'). 본 연구에서는 이렇게 형태소와 품사를 밑줄 기호로 연결한 형태를 일반적으로 사용되는 '단어'의 의미와 구분하기 위해 '키워드'라 지칭한다.

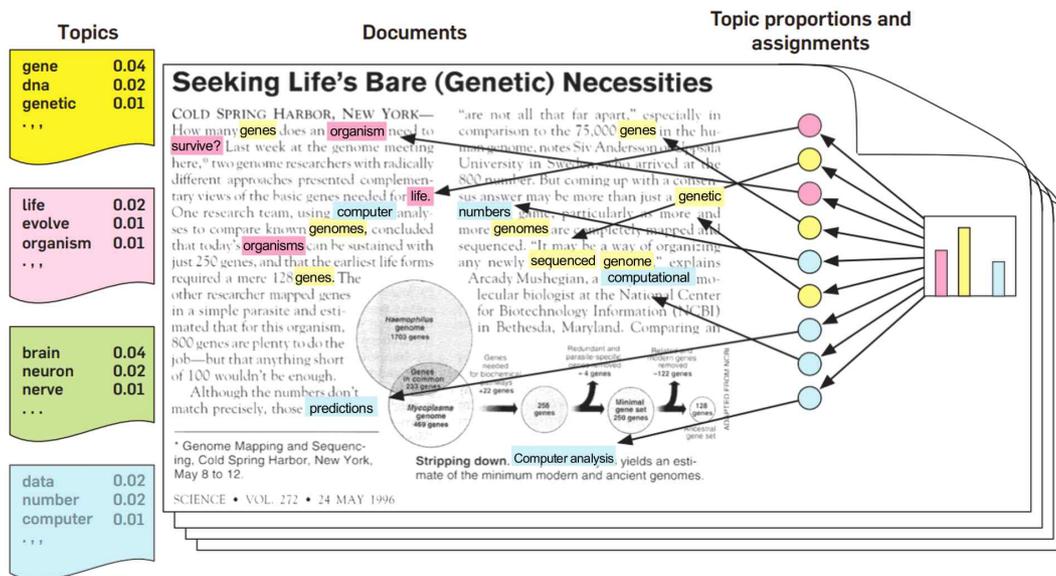
## 2.2 LDA 기반 토픽 모델링

### 2.2.1 토픽별 키워드 추출을 위한 LDA

전처리 작업 후 특정 품사(일반명사, 고유명사, 동사, 형용사, 어근)로 구성된 장소 리뷰 키워드들로부터 POI에 대한 시맨틱 정보를 추출하기 위한 첫 단계로, LDA 기반 토픽 모델링을 실시한다. 토픽 모델링은 텍스트 마이닝(text mining) 기법 중 하나로, 텍스트 데이터에 사용된 단어들의 동시 사용 패턴에 근거해 해당 단어들을 대표하는 특정 토픽을 자동으로 추출하는 분석 방법이다. 토픽 모델링 기법은 각 POI에 대해 작성된 리뷰 텍스트 집합이 분위기, 가격 등 다수의 토픽을 포함한다고 가정하므로 해당 텍스트 집합을 토픽들의 확률적 혼합체로 간주한다(유예림, 2017). 토픽 모델링 기법은 잠재 의미 분석(latent semantic analysis, 이하 LSA)(Deerwester *et al.*, 1990), 확률론적 LSA(probabilistic LSA, pLSA)(Hofmann, 1999), LDA(Blei *et al.*, 2003) 등의 모델이 있는데, 이 중 가장 일반적으로 사용되는 것이 LDA 기법이다. LDA는 다른 토픽 모델링 기법에 비해 결과에 대한 해석이 용이하고(Blei, 2012), 하이퍼 파라미터(hyper parameter) 값을 임의의 변수로 평활화(smoothing)해 과적합(overfitting)되지 않으므로 다수의 토픽을 도출하는 데 유리하다(Steyvers & Griffiths, 2007).

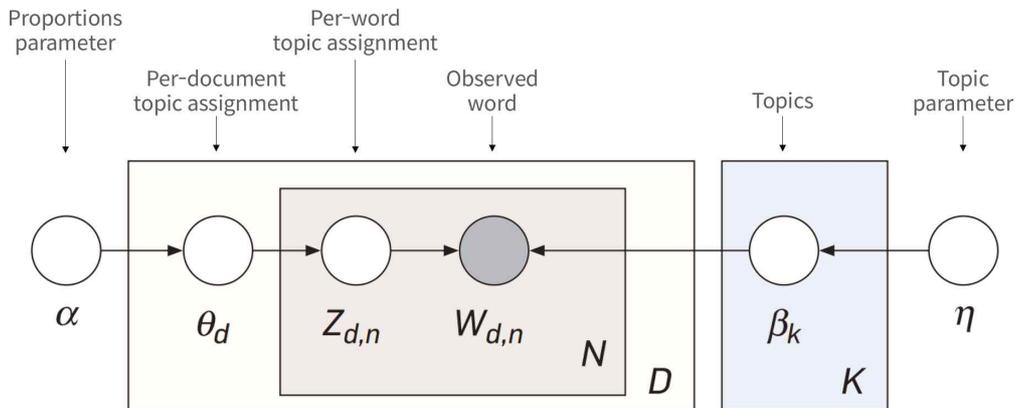
LDA 기법은 토픽에 따라 문서를 군집화한다는 측면에서 군집분석과 유사한데, LDA는 군집분석 기법과는 다르게 문서와 토픽 간 다대다 대응이 가능하다는 차이가 있다(Blei, 2012). 즉, 한 문서에 하나가 아닌 여러 개의 토픽이 포함될 수 있다는 점에서 실제 텍스트의 본질을 잘 반영한다는 장점이 있다(김용대 & 정구환, 2016; 김정수 & 이석준, 2016; 유예림, 2017).

LDA는 문서들이 잠재된(latent) 몇 개의 토픽들로 이루어져 있고, 각 토픽은 여러 단어로 구성된다고 가정한다. [그림 2-6]은 LDA의 기본 가정을 도식화한 것으로, 각 문서는 다음의 단계를 거쳐 생성된다고 가정한다. 먼저, [그림 2-6]의 가장 우측 히스토그램처럼, 토픽에 대한 분포를 무작위로 고른다. 그리고 해당 분포에 할당 가능한 토픽들([그림 2-6] 우측의 색칠된 원들)을 선택하고, 각 토픽에 대응되는 단어들 중 하나의 단어를 선택한다(Blei, 2012). 문서가 이러한 과정을 거쳐 생성되었다고 가정하고 [그림 2-6]을 좀 더 알기 쉽게 설명하자면, 좌측 상단의 박스는 해당 문서에 포함된 하나의 토픽이며, ‘gene’이라는 단어가 해당 토픽에 등장할 확률이 0.04, ‘dna’는 0.02, ‘generic’은 0.01이라는 의미다. 이러한 단어들을 통해 해당 토픽은 ‘유전자’와 관련되어 있다는 것을 추측할 수 있다. LDA에서는 각 문서를 구성하는 토픽의 비율은 문서마다 다르며, 이 비율을 결정하는 디리클레 분포(Dirichlet distribution)가 존재한다고 전제한다.



[그림 2-6] LDA의 기본 가정  
(Blei(2012) p. 78 Figure 1 재구성)

[그림 2-7]은 LDA가 가정하는 문서 생성 과정을 그래프 모델로 표현한 것이다. LDA는 주어진 텍스트 문서의 토픽을 파악하기 위해 베이지스(bayesian) 확률 모델로 사후 분포(posterior distribution)를 추정하는데, 이때 전체 텍스트 문서를 대표하는 토픽들과 각 토픽에 할당되는 주요 단어들의 집합을 찾아낸다(유예림, 2017). [그림 2-7]이 그 과정을 도식화한 것으로,  $D$ 는 전체 문서 수,  $K$ 는 전체 토픽 수,  $N$ 은  $d$ 번째 문서의 단어 수를 의미한다. 네모 상자는 해당 변수의 크기만큼 반복하라는 의미이며, 각 노드는 확률 변수를 뜻한다. 화살표가 시작되는 노드는 조건, 화살표가 향하는 노드는 결과에 해당하는 변수이다. 그중 음영 처리된 노드  $W_{d,n}$ 은 관찰된 데이터로,  $d$ 번째 문서에 등장한  $n$ 번째 단어를 뜻하는데, 해당 그림에서 화살표의 방향이 노드  $W_{d,n}$ 으로 향하고 있는 것을 알 수 있다. 즉, 관찰된 데이터인  $W_{d,n}$ 을 기준으로, 사용자가 지정하는 하이퍼 파라미터인  $\alpha$ 와  $\eta$ 를 제외한 모든 잠재 변수( $\beta_k, \theta_d, Z_{d,n}$ )를 추정해야 한다. 이는  $k$ 번째 토픽에 해당하는 벡터로, [그림 2-7]에서 화살표의 방향을 보면  $\beta_k$ 는  $\beta_k$ 의 사전 분포(prior distribution)를 결정하는 하이퍼 파라미터인  $\eta$ 의 영향을 받는 것을 알 수 있다.  $\theta_d$ 는  $d$ 번째 문서가 가진 토픽의 비율을 나타내는 벡터로,  $\theta_d$ 의 사전 분포를 결정하는 하이퍼 파라미터인  $\alpha$ 의 영향을 받는다.  $Z_{d,n}$ 은  $d$ 번째 문서의  $n$ 번째 단어를 각 토픽에 할당하는 변수로, 문서의 토픽 비중을 결정하는  $\theta_d$ 의 영향을 받는다. 모든 화살표의 종착지인  $W_{d,n}$ 은  $d$ 번째 문서의  $n$ 번째 단어가 어떤 단어인지 할당해 주는 변수로,  $Z_{d,n}$ 과  $\beta_k$ 의 영향을 동시에 받는 것을 알 수 있다. 즉, LDA는 토픽별 단어 분포( $\eta \rightarrow \beta_k$ )와 문서별 토픽 분포( $\alpha \rightarrow \theta_d \rightarrow Z_{d,n}$ )의 결합을 통해 문서 내 단어들이 생성된다고 가정하는 것이다.



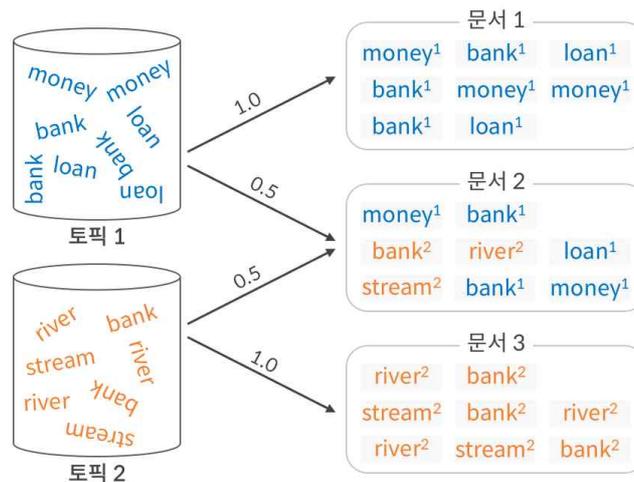
[그림 2-7] LDA의 그래프 모델  
(Blei(2012) p.81 Figure 4 재구성)

LDA의 하이퍼 파라미터인  $\alpha$ 와  $\eta$ 는 데이터를 기반으로 알아서 추정되는 값이 아니라 연구자의 사전 지식에 의해 결정되기 때문에 모델의 성능을 높이려면 하이퍼 파라미터의 값을 신중히 설정해야 한다. LDA의 하이퍼 파라미터를 결정하는 문제와 관련해 Wallach(Wallach, 2008; Wallach *et al.*, 2009)는 비대칭(asymmetric)의 하이퍼 파라미터를 LDA 모형에 최적화시키기 위한 방안으로 깃스 샘플링(gibbs sampling) 기법을 제안하였다. 해당 연구에서는 LDA의 하이퍼 파라미터에 대한 최적 조합을 실험적으로 입증한 바 있는데,  $\alpha$ 를 비대칭,  $\eta$ 는 대칭 파라미터로 설정했을 때 토픽 분류 성능이 최적화됨을 확인하였다. 비대칭은 토픽마다 해당 하이퍼 파라미터 값을 다르게 설정하는 것을 허용하는 것을 의미한다. 김은희 등(2012)도  $\alpha$ 를 비대칭,  $\eta$ 는 대칭으로 설정했을 때 LDA 모델이 최적화됨을 확인한 바 있다.

LDA 모델은 베イズ 확률 모델이기에 사후 분포를 추정해야 하는데, 이를 직접 다루는 것이 어렵기 때문에 깃스 샘플링 기법을 사용한다(김용대 & 정구환, 2016). 그러나 깃스 샘플링은 계산이 복잡하고 변수 간 의존성이 커 수렴이 늦기 때문에 대규모 텍스트 데이터에 적용하기 어려운 단점이 있다(김용대 & 정구환, 2016; 유예림, 2017). 이 같은 문제를 해결하기 위해 제안된 것이 붕괴된 깃스 샘플링(collapsed gibbs sampling)(Griffiths

& Steyvers, 2004) 방법이다. 붕괴된 깃스 샘플링 기법은 변수 간 의존성을 줄일 뿐 아니라 속도가 빠르다는 장점이 있다(김용대 & 정구환, 2016; 유예림, 2017).

LDA가 문서를 정의하는 방식에는 각 단어가 하나의 토픽에 속하도록 제한하는 상호 배타성의 개념이 없기 때문에 동음이의어(homonym)나 다의어(polysemy)를 구분할 수 있다(Steyvers & Griffiths, 2007). 예를 들어, [그림 2-8]의 문서 1과 문서 3은 각각 토픽 1과 토픽 2에서만 단어를 추출해 문서를 생성한 반면, 문서 2는 토픽 1과 토픽 2를 동일한 비율로 혼합해 생성한 것을 나타낸다. 각 단어의 위 첨자 번호는 단어를 추출하는데 사용된 토픽 번호를 의미한다. ‘bank’라는 단어는 두 가지 의미(은행/둑, 제방)를 가지고 있는데, 해당 단어가 독립적으로 존재하고 있으면 정확한 의미를 알기 어렵다. 그러나 [그림 2-8]의 문서 1과 같이 해당 단어가 ‘money’나 ‘loan’ 같은 단어와 함께 쓰였다면 이는 은행을 의미한다는 것을 알 수 있다. 반면 [그림 2-8]의 문서 3과 같이 해당 단어가 ‘river’나 ‘stream’ 같은 단어들과 함께 쓰인 경우에는 이것이 둑이나 제방을 의미한다는 것을 높은 확률로 추측할 수 있다.



[그림 2-8] 토픽 모델의 확률 생성 과정 (Steyvers & Griffiths(2007) p.3 Figure 2 재구성)

본 연구에서 장소 리뷰 키워드들에 대해 LDA를 적용하는 이점은 동음이의어, 다의어뿐 아니라 문맥에 따라 달라지는 키워드의 의미를 구분할 수 있다는 것이다. 리뷰에서 많이 사용되는 형용사인 ‘좋다’를 예로 들어 보면, ‘분위기가 좋다’에서의 ‘좋다’와 ‘가성비가 좋다’에서의 ‘좋다’는 모두 ‘대상의 성질이나 내용 따위가 보통 이상의 수준이어서 만족할 만하다<sup>25)</sup>’는 일반적 의미의 ‘좋다’로 해석할 수 있다. 그러나 일반적 의미의 ‘좋다’라는 키워드도 어떤 토픽에서 사용되었는지에 따라 그 세부 의미가 달라질 수 있다. 즉, ‘분위기’ 토픽의 ‘좋다’는 ‘아늑한’, ‘쾌적한’ 등의 단어와 유사한 의미를 가질 확률이 높고, ‘가격’ 토픽의 ‘좋다’는 ‘저렴한’, ‘싼’ 등의 단어와 유사한 의미를 가질 확률이 높다. 본 연구에서는 이렇게 사전적으로 같은 의미일지라도 토픽에 따라 세부 의미를 다르게 반영하기 위해 토픽별 키워드를 생성하고, 이를 POI 임베딩에 활용하기 위해 LDA 기법을 적용한다. 실제 ‘분위기’ 관련 토픽에서 ‘좋다’라는 단어를 임베딩한 것과 ‘가격’ 관련 토픽에서 같은 단어를 임베딩한 것의 차이는 3.4절 토픽별 키워드 기반 POI 임베딩 결과에서 더욱 상세히 기술한다.

형태소 분석을 통해 특정 품사(일반명사, 고유명사, 동사, 형용사, 어근)만 추출한 키워드에 LDA 모델링을 적용하면 그 결과로 토픽별 키워드 분포, POI에 작성된 각 리뷰별 토픽 분포, 그리고 키워드별 토픽 분포를 모두 추정할 수 있다. 그중에서도 본 연구에서는 LDA 모델링 다음 단계인 POI 임베딩을 위해 키워드별 토픽 분포를 활용한다. 즉, 각 POI에 대해 작성된 리뷰 키워드별로 가장 높은 확률의 토픽을 할당해 이를 [키워드, 토픽] 쌍으로 구성하며, 이것이 POI 임베딩을 위한 입력값으로 사용된다. 예를 들어, 어떤 POI에 대해 작성된 리뷰 텍스트에 LDA 모델링을 적용한 결과, ‘좋다’라는 키워드가 1번 토픽에 속할 확률이 가장 높다면 [좋다, 1] 형태의 값이 생성되며, 이러한 방식을 통해 누적된 [키워드, 토픽] 쌍이 해당 POI를 벡터로 표현하기 위한 입력 데이터로 활용된다. 이는 여러 단어가 모여 하나의 문서를 만드는 것과 같이 POI에 작성된 리뷰 텍스트로부터 추출한 토픽별 키워드들의 집합으로 POI를 나타낼 수 있다고 간주하는 것이다.

---

25) 네이버 표준국어대사전

## 2.2.2 LDA 평가 방법

LDA는 무감독 분류(unsupervised classification)에 속하는 분석 방법으로, 연구자가 사전에 토픽의 개수를 정해야 한다. 그러나 대규모 데이터 분석 시 연구자의 사전 지식만으로 토픽 개수를 정하기는 쉽지 않기 때문에 LDA 결과를 정량적으로 평가하는 과정이 필요하다. 토픽 개수에 따른 LDA 모델의 성능을 파악하기 위한 평가 방법으로는 복잡도(perplexity) 지수와 토픽 일관성(topic coherence) 지수가 있다. 일반적으로 LDA 모델링 시 두 값을 모두 산출한 후 이를 바탕으로 연구자가 적절한 토픽의 개수를 설정한다.

복잡도 지수(Grün & Hornik, 2011)는 크기가  $N$ 인 단어 전체 집합이  $W$ 라고 할 때 식 (2-1)과 같이 계산하며, 여기서  $H(W)$ 는 교차 엔트로피(cross entropy)의 근삿값을 의미한다. 해당 수식을 보면, 복잡도는 각 단어당 우도(likelihood)의 기하 평균(geometric mean)의 역수와 같은 것을 알 수 있다. 복잡도를 계산하는 가장 일반적인 방법은 데이터를 임의로 훈련 데이터(training data)와 시험 데이터(test data)로 나누어, 훈련 데이터에서 얻은 모델을 시험 데이터에 적용했을 때의 성능을 수치화하는 것이다(Battisti *et al.*, 2015; 김지은 & 백순근, 2016; 유예림, 2017).

$$\begin{aligned} \text{Perplexity}(W) &= 2^{H(W)} \\ &= 2^{-\frac{1}{N} \log_2 P(w_1, w_2, \dots, w_N)} \\ &= \sqrt[N]{\prod_{i=1}^N \frac{1}{P(w_i | w_1, w_2, \dots, w_{i-1})}} \end{aligned} \quad (2-1)$$

복잡도 값이 크다는 것은 훈련 데이터로 만들어진 토픽 모델이 시험 데이터에서 토픽별 단어를 잘못 추출한다는 것을 의미하기 때문에 그 값이 작을수록 모델의 성능이 우수한 것으로 평가할 수 있다(Battisti *et al.*,

2015; Blei *et al.*, 2003). 그러나 최근, 텍스트 데이터에서 토픽을 추출할 때 복잡도 산출을 통해 도출된 결과와 사람이 판단한 결과 간의 상관관계 (correlation)가 통계적으로 유의하지 않거나 적합하지 않은 경우가 있다는 것이 밝혀졌으며, 복잡도 값으로는 토픽의 해석이 불가능하다는 한계가 지적되고 있다(Chang *et al.*, 2009; Jacobi *et al.*, 2016; Newman *et al.*, 2011; O'callaghan *et al.*, 2015; Röder *et al.*, 2015). 이러한 한계를 극복하기 위해 개발된 것이 토픽 일관성 지수이다.

토픽 일관성 지수는 토픽에 할당된 단어들을 통해 해당 토픽을 해석할 수 있는지를 판단하기 위한 값이다. Röder *et al.*(2015)은 토픽 일관성을 측정하는 여러 방식 중 교차 검증(cross-validation) 기법을 기반으로 계산한  $C_V$ 의 성능이 가장 좋다고 평가하였다. Röder *et al.*(2015)에 의하면  $C_V$  기반 토픽 일관성 지수의 계산은 총 4단계를 거쳐 진행된다. 먼저, 1단계에서는 LDA로 추정된 토픽과 단어의 확률 분포로부터 가장 가능성이 큰 상위  $N$ 개의 단어 ( $W = \{w_1, w_2, \dots, w_N\}$ )를 정의한다. 이때 단어의 부분 쌍 집합  $S_i$ 는  $W' \in W$ 와  $W^* \in W$ 로 구성되므로  $S$ 는  $\{(W', W^*) | W' = \{w_i\}; w_i \in W; W^* = W\}$ 로 정의된 모든 쌍 집합이다.

2단계에서는 단어 집합  $W$ 에서 임의의 2개 단어가 동시에 출현하는 확률을 계산하기 위해 교차 검증 기법 중 하나인 불리언 슬라이딩 윈도우(boolean sliding window) 계산 방법을 사용한다. 예를 들어, 슬라이딩 윈도우의 크기를 50으로 정하면,  $W$  단어 집합에서 1부터 50번째 순서까지의 단어들을 선택해 부분 집합을 만들고, 임의의 단어 2개가 그 50개의 부분 집합에서 동시에 포함되어 있을 확률을 계산한다. 그리고 마치 창문이 미끄러지듯 단어 부분 집합들을 2부터 51번째 순서, 3에서 52번째 순서의 단어들로 이동해 가면서 훈련과 시험을 반복한다.

3단계에서는 각 단어 쌍 집합  $S_i = (W', W^*)$ 에 대해 확인 측도(confirmation measure)  $\phi$ 를 계산하는데, 이는  $W^*$ 가  $W'$ 를 얼마나 강력하게 지지하는지(support), 즉 두 단어 집합이 얼마나 자주 동시에 발생하는지를 계산하는 것이다. 유사성 계산을 위해  $W'$ 와  $W^*$ 는 각각 벡터  $\vec{v}(W')$

와  $\vec{v}(W^*)$ 로 표현된다(식 (2-2)). 또한, 단어  $w_i$ 와  $w_j$  사이의 유사성은 식 (2-3)과 같이 점 상호 정보(normalized pointwise mutual information, 이하 NPMI)를 통해 계산된다.  $\epsilon$ 은 분자가 0이 되는 것을 막기 위한 값으로, 보통  $10^{-12}$ 과 같이 아주 작은 값이 부여된다(Stevens *et al.*, 2012).  $\gamma$ 는 원래 더 높은 NPMI 값에 가중치를 부여하기 위한 것인데, 본 연구에서는 Röder *et al.*(2015)에 따라  $\gamma=1$ 로 설정한다. NPMI는 두 단어가 동시에 나타날 확률을 수치화한 점 상호 정보량(pointwise mutual information, PMI) 값의 범위를  $(-\infty, \infty)$ 에서  $[-1, 1]$ 이 되도록 정규화한 것이다. NPMI 값이 1이면 두 단어가 항상 같이 나타난다는 것을 의미하고, -1이면 두 단어가 절대 같이 나타나지 않는다는 것을 의미한다.

4단계에서 최종적인  $C_V$  값을 계산하는데, 이는 모든 확인 측도  $\phi$ 의 산술 평균값을 통해 구해진다.  $S_i$ 의 확인 측도  $\phi$ 는 식 (2-4)와 같이  $\vec{v}(W') \in \vec{u}$ 와  $\vec{v}(W^*) \in \vec{w}$ 를 기반으로,  $S_i$  내 모든 벡터  $\phi_{S_i}(\vec{u}, \vec{w})$ 의 코사인 유사도(cosine similarity)를 통해 계산할 수 있다.

$$\vec{v}(W') = \left\{ \sum_{w_i \in W'} NPMI(w_i, w_j)^\gamma \right\}_{j=1, \dots, |W|} \quad (2-2)$$

$$NPMI(w_i, w_j)^\gamma = \left( \frac{\log \frac{P(w_i, w_j) + \epsilon}{P(w_i) \cdot P(w_j)}}{-\log(P(w_i, w_j) + \epsilon)} \right)^\gamma \quad (2-3)$$

$$\phi_{S_i}(\vec{u}, \vec{w}) = \frac{\sum_{i=1}^{|W|} u_i \cdot w_i}{\|u\|_2 \cdot \|w\|_2} \quad (2-4)$$

$C_V$ 는 그 값이 클수록 토픽 내 단어들의 일관성이 높다는 것을 의미하므로, 이 값이 최대가 되는 토픽 개수를 채택하는 것이 좋다(Röder *et al.*, 2015).  $C_V$  값의 의미를 식 (2-2)~식 (2-4)를 이용해 간략히 설명하자면, 먼저 단어 쌍 집합  $S_i$ 에 포함된 단어 벡터  $\vec{v}(W') \in \vec{u}$ 는 해당 단어와 나머지 단어 간 계산된 NPMI 값의 집합을 통해 구성된다(식 (2-2), 식 (2-3)). 이렇게 구성된  $S_i$  내 모든 벡터  $\vec{u}$ 와  $\vec{w}$ 의 코사인 유사도 값(식 (2-3))의 합을 통해  $C_V$  값이 결정된다. 즉, 두 단어의 NPMI 값을 이용해 벡터를 구성했기 때문에 두 단어가 동시에 나타날 확률이 높을수록 코사인 유사도 값이 커지며, 이에 따라 코사인 유사도의 평균값인  $C_V$  값도 커지게 되는 것이다. 따라서  $C_V$  값이 크다는 것은 각 토픽에 속하는 단어들이 서로 유사하다는 것을 뜻하며, 이는 곧 토픽 내 단어들 간의 일관성이 높다는 것을 의미한다. 최종 토픽 개수는 복잡도 값과 토픽 일관성 값을 기준으로 연구자의 판단 하에 정해지는 것이 일반적이다.

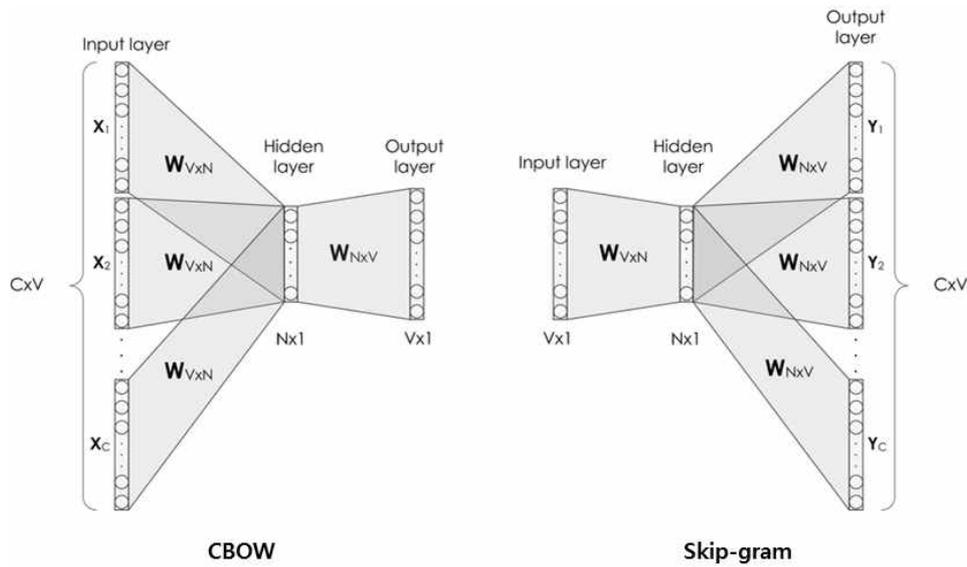
## 2.3 토픽별 키워드 기반 POI 임베딩

### 2.3.1 시맨틱 POI 생성을 위한 POI 임베딩

리뷰 데이터로부터 시맨틱 정보를 추출해 시맨틱 POI를 생성하기 위한 최종 단계로, POI를 임베딩하는 방법을 제안한다. POI를 임베딩하기 위해서는 우선적으로 LDA 결과로 얻어진 [키워드, 토픽] 쌍을 벡터로 임베딩하는 과정이 필요하다. 이는 여러 토픽별 단어들이 모여 하나의 문서를 만드는 것처럼 POI에 대한 토픽별 키워드들의 집합으로 POI를 나타낼 수 있다고 보는 것이다.

임베딩은 주로 NLP 분야에서 사용되는 용어로, 단어의 의미와 맥락을 고려해 각 단어를 벡터 공간상에 표현하는 방법을 통칭한다. 최근 딥러닝 등 인공지능 기반 기술이 발전하면서 NLP 관련 기술의 활용도 함께 증가하고 있는데, 그중 대표적인 것이 단어 임베딩 기법이다. 여러 임베딩 방법 중 가장 일반적으로 사용되는 것이 신경망 기반의 Word2vec 모델(Mikolov *et al.*, 2013a; Mikolov *et al.*, 2013b)이다. Word2vec은 연구자가 설정한 크기의 단어 벡터를 0이 아닌 실숫값을 이용해 밀집 표현(dense representation)하는 방식이다. 이는 단어 전체 개수의 크기를 가지는 벡터를 생성해 모두 0으로 채우고, 특정 단어만을 1로 표기하는 원-핫 인코딩(one-hot encoding) 기반의 희소 표현(sparse representation) 방식과 구분된다. Word2vec은 유사한 문맥의 단어는 유사한 의미를 가진다는 Harris(1954)의 분포가설에 기반해 언어에 내재된 의미(semantic) 관계를 파악하는 데 뛰어난 성능을 보인다(Liu *et al.*, 2017; Mikolov *et al.*, 2013a; Mikolov *et al.*, 2013b). Word2vec 모델은 단어 간 유사성 및 관련성을 측정하는 데 사용될 수 있어 NLP, 정보 검색 등 다양한 분야에서 활용되고 있다(Lai *et al.*, 2016; Liu *et al.*, 2015b). Word2vec은 continuous bag-of-words(이하 CBOW)와 skip-gram의 두 가지 방식이 있는데, CBOW는 주변(context) 단어를 이용해 중심(center) 단어를 예측하는 것이고, 반대로 skip-gram은 중심 단어를 기준으로 주변에 올 수 있는 단어

를 예측한다는 차이가 있다([그림 2-9]). 대규모 데이터에 적용 시 skip-gram 모델의 성능이 CBOW보다 더 뛰어난 것으로 증명되었기에 (Mikolov *et al.*, 2013a; Mikolov *et al.*, 2013b) 일반적으로 skip-gram이 더 많이 사용된다. [그림 2-9]에서  $V$ 는 단어 전체 개수,  $N$ 은 은닉층의 노드 개수,  $C$ 는 입력 벡터의 개수, 그리고  $W$ 는 해당 네트워크의 가중치 행렬을 의미한다.

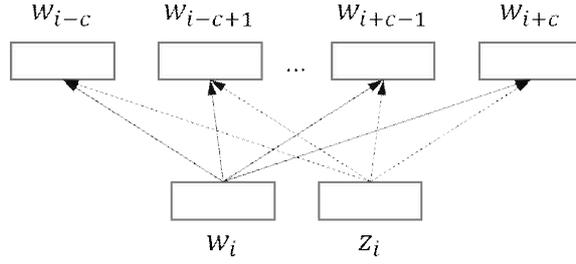


[그림 2-9] Word2Vec의 두 가지 학습 방법  
(Tutubalina & Nikolenko, 2017)

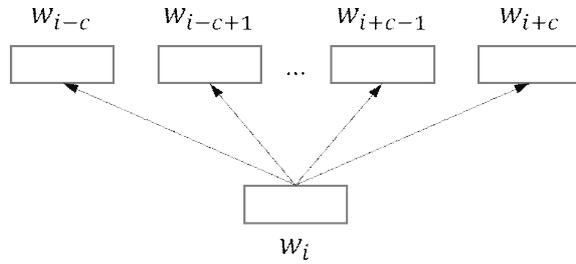
Word2vec 모델은 단일 벡터를 사용해 각 단어를 나타내기 때문에 동음이의어나 다의어를 구별하지 못한다는 단점이 있다. 이에 따라 Liu *et al.*(2015b)은 토픽과 토픽별 단어를 기반으로 학습하는 TWE 기법을 제안하였다. TWE는 skip-gram을 개선한 것으로, 기본 아이디어는 각 단어가 다른 토픽에서 다른 임베딩 값을 가지도록 하는 것이다. 가령 ‘apple’이라는 단어가 음식 토픽에 속할 때는 ‘과일’의 의미를 갖도록 하고, IT 토픽에 속하는 경우에는 ‘회사’의 의미를 갖도록 하는 것이다. Liu *et al.*(2015b)은 학습 방법에 따라 TWE-1, TWE-2, TWE-3의 세 가지 TWE를 제안하였

는데 확인 결과, TWE-1의 학습 방법이 단어의 문맥적 유사성 및 텍스트 분류에서 모두 가장 좋은 성능을 보였다. 따라서 본 연구에서도 TWE-1을 기반으로 POI 임베딩 작업을 수행한다.

POI를 임베딩하기 위해서는 LDA 결과로 추출된 토픽별 키워드에 대한 임베딩 작업이 우선되어야 한다. [그림 2-10 (a)]는 TWE-1 모델의 개념을 도식화한 것으로, [그림 2-10 (b)]와의 비교를 통해 TWE-1이 Mikolov *et al.*(2013b)의 skip-gram 모델을 확장한 것임을 알 수 있다. TWE-1에서 각 키워드  $w_i$ 는 LDA를 통해 할당된 토픽 정보를 포함하고 있으므로 [키워드( $w_i$ ), 토픽( $z_i$ )] 쌍을 구성할 수 있다. 또한, 윈도우 크기가  $c$ 일 때, 훈련을 위한 주변 [키워드, 토픽] 쌍은  $\{w_{i-c} : z_{i-c}, w_{i-c+1} : z_{i-c+1}, \dots, w_{i+c-1} : z_{i+c-1}, w_{i+c} : z_{i+c}\}$  로 주어진다. TWE-1은 각 토픽을 의사 단어(pseudo word)로 간주하고, 키워드 임베딩과 토픽 임베딩을 별도로 학습한다. 그런 다음, 키워드( $w_i$ )와 토픽( $z_i$ )의 임베딩에 따라 [키워드( $w_i$ ), 토픽( $z_i$ )]의 토픽별 키워드 임베딩을 수행한다. 주변 단어를 예측하기 위해 중심 단어( $w_i$ )만 사용하는 skip-gram과 비교해 보면, TWE-1은 주변 단어를 예측하기 위해 중심 단어(의사 단어)로 토픽( $z_i$ )도 함께 사용한다는 차이가 있다. TWE-1 모델에서 학습을 위한 목적 함수는 식 (2-5)와 같으며, 비교를 위해 skip-gram 모델의 목적 함수를 식 (2-6)에 함께 작성하였다.



(a) TWE-1 모델 개념도



(b) skip-gram 모델 개념도

[그림 2-10] TWE-1 및 skip-gram 모델 개념도

(Fu *et al.*(2016) p.194 Figure 2; Liu *et al.*(2015b) p.2419 Figure 1 재구성)

$$J_{TWE}(D) = \frac{1}{V} \sum_{i=1}^V \sum_{-c \leq j \leq c, j \neq 0} \log \Pr(w_{i+j}|w_i) + \log \Pr(w_{i+j}|z_i) \quad (2-5)$$

$$J_{SG}(D) = \frac{1}{V} \sum_{i=1}^V \sum_{-c \leq j \leq c, j \neq 0} \log \Pr(w_{i+j}|w_i) \quad (2-6)$$

식 (2-5) 및 식 (2-6)에서  $D$ 는  $D = \{w_1, \dots, w_V\}$ 로 주어지는 키워드 집합,  $V$ 는 키워드 집합의 크기,  $c$ 는 윈도우 크기를 의미한다.  $\Pr(w_{i+j}|w_i)$ 를 계산하기 위해서는 식 (2-7)과 같이 소프트맥스(softmax) 함수를 사용한다. 해당 식에서  $w_i$ 와  $w_k$ 는 각각 입력 벡터(중심 키워드)와 출력 벡터(주변 키워

드), 그리고  $W$ 는 키워드 집합을 의미한다. TWE-1에서 토픽  $z$ 에 속하는 키워드  $w$ 의 토픽 기반 임베딩은 두 벡터를 병합시킴으로써( $w^z = w \oplus z$ ) 수행된다. 따라서  $w^z$ 의 길이는  $w$  및  $z$  길이의 두 배가 된다.

$$\Pr(w_k | w_i) = \frac{\exp(w_k \cdot w_i)}{\sum_{w_j \in W} \exp(w_k \cdot w_j)} \quad (2-7)$$

Liu *et al.*(2015b)에 의하면 문서  $d$ 에 포함된 토픽별 단어 벡터  $w^z$ 의 합을 통해 문서를 임베딩할 수 있다. 본 연구에서는 단어들이 모여 문서를 이루는 것처럼 POI를 설명하는 키워드들이 모여 POI를 나타낼 수 있다고 간주하고, 문서를 POI로 치환한다. 즉, 각 POI  $l$ 에 속하는 각 키워드  $w$ 의 모든 토픽별 키워드 벡터  $w^z$ 를 합산해 POI의 시맨틱을 나타내는 것이다. 한 POI 내에서 동일한 벡터값을 가지는 토픽별 키워드가 2번 이상 등장할 수 있으므로 각 POI  $l$ 에서  $w$ 가 발생할 조건부 확률인  $\Pr(w|l)$ 을 토픽별 키워드 벡터  $w^z$ 에 곱해줌으로써 임베딩 값을 정규화한다. 본 연구에서 POI를 임베딩하기 위해 Liu *et al.*(2015b)의 문서 임베딩 수식을 재정의한 것은 식 (2-8)과 같다.

$$l = \sum_{w \in l} \Pr(w|l) w^z \quad (2-8)$$

식 (2-8)에서 조건부 확률로 정규화하는 대신 단어 빈도-역 문서 빈도 (term frequency-inverse document frequency, 이하 TF-IDF)를 이용해 가중치를 부여하는 방법도 있다. TF-IDF는 텍스트 마이닝 및 정보 검색 분야에서 사용하는 가중치로, POI별 리뷰 키워드들이 있을 때, 어떤 키워드가 특정 POI에서 얼마나 중요한 것인지를 통계적으로 나타낼 수 있다. 단어 빈도(term frequency, 이하 TF)는 특정 키워드가 한 POI 내에서 나

타나는 빈도수로, TF 값이 클수록 해당 키워드가 해당 POI에서 중요하게 사용된다고 추측할 수 있다. 예를 들어, POI-1에 ‘아늑한’이라는 표현이 10회 등장하고, POI-2에 동일 키워드가 1번 등장하였다면, 해당 키워드는 POI-1과의 관련성이 더 높다고 볼 수 있으므로 이를 가중치로 반영해 주는 것이다. 그러나 만약 해당 키워드가 거의 모든 POI에서 다 등장한다면 그만큼 흔하게 사용되는, 중요하지 않은 키워드일 것이라 짐작할 수 있는데, 이것을 반영하는 것이 문서 빈도(document frequency)의 역수인 역 문서 빈도(inverse document frequency, 이하 IDF)이다. 최종적으로 TF와 IDF의 곱으로 TF-IDF 값을 계산한다.

어떤 키워드에 대한 TF-IDF 값이 크다는 것은 다른 POI에 비해 해당 POI에서 그 키워드가 더 중요하게 사용되었다는 것을 의미한다. 따라서 사용자가 키워드를 기반으로 검색할 때 더욱 합리적인 결과를 반환해 줄 수 있다. 예컨대 사용자가 ‘아늑한’이라는 검색어를 입력한 경우, 이는 POI-2보다는 POI-1과의 관련성이 더 높으므로 검색 결과로써 POI-1을 우선으로 반환해 줘야 한다. 따라서 POI-2에 비해 POI-1이 ‘아늑한’이라는 키워드 벡터와 더 가깝게 위치하도록 TF-IDF 가중치를 부여하는 것이다. 토픽별 키워드 벡터  $w^z$ 에 TF-IDF 가중치를 부여해 POI를 임베딩하는 수식은 식 (2-9)와 같다. 해당 수식에서  $TF(w, l)$ 은 해당 키워드  $w$ 가 POI  $l$ 에 등장한 빈도수를 뜻한다.  $IDF(w)$ 는 식 (2-10)과 같이 계산하며, 여기서  $n$ 은 POI의 전체 개수를 의미한다.

$$l = \sum_{w \in l} (TF(w, l) \cdot IDF(w))w^z \quad (2-9)$$

$$IDF(w) = \log\left(\frac{1+n}{1+DF(w)}\right) + 1 \quad (2-10)$$

각 POI에 포함된 전체 키워드 개수가 다양하므로 TF-IDF 값을 가중치로 사용하였을 때 많은 키워드를 포함하고 있는 POI의 경우, 벡터값이 발산(divergence)될 수 있다. 이에 따라 POI마다 다양한 전체 키워드 개수의 영향을 제거하면서 서로 상대적인 키워드의 중요도를 고려하기 위하여  $L_2$ -norm 기반 정규화 작업을 수행한다.  $L_2$ -norm은 식 (2-11)과 같이 계산하며, 여기서  $v$ 는 각 POI에 포함된 토픽별 키워드들의 TF-IDF 값을 벡터로 치환한 것을 의미하고,  $k$ 는 각 POI에 포함된 토픽별 키워드들의 전체 개수를 뜻한다.

$$v_{norm} = \frac{v}{\|v\|_2} = \frac{v}{\sqrt{v_1^2 + v_2^2 + \dots + v_k^2}} \quad (2-11)$$

### 2.3.2 POI 임베딩 평가 방법

본 연구에서 POI 임베딩을 위해 사용한 TWE-1 모델의 성능은 하이퍼 파라미터의 설정에 따라 달라진다(Caselles-Dupré *et al.*, 2018). 신경망 기반의 TWE-1은 학습으로 진화하는 모델이기에 적절한 하이퍼 파라미터 값을 설정하는 것이 중요하다. 이에 따라 본 연구에서는 적절한 POI 임베딩을 위한 최적의 하이퍼 파라미터 조합을 찾기 위해 POI 임베딩 평가를 실시한다. 그러나 여러 개의 하이퍼 파라미터들이 서로 영향을 주고받기 때문에 최적의 하이퍼 파라미터 조합을 찾는 것은 쉽지 않은 일이다(강형석 & 양장훈, 2019). 이에 따라 대부분의 선행 연구(Feng *et al.*, 2017; Lai *et al.*, 2016; Yan *et al.*, 2017)에서는 예상되는 최적 조합을 기준으로 하이퍼 파라미터를 조절해 가면서 모델의 성능을 평가하는 식으로 하이퍼 파라미터를 최적화한다.

본 연구에서 사용한 TWE-1의 주요 하이퍼 파라미터는 임베딩할 벡터 차원(dimension)의 크기, 컨텍스트 윈도우(context window)의 크기, 그리고 키워드의 최소 출현 횟수다. 여기서 컨텍스트 윈도우 크기란 skip-gram이나 CBOW 기반의 Word2vec 모델에서 중심 단어를 기준으로 좌우에 있는 단어들을 최대 몇 개까지 주변 단어로 학습시킬 것인지를 결정하는 값이다. 영어의 경우, 윈도우 크기가 5일 때 좋은 성능을 내는 것으로 알려져 있다(Lai *et al.*, 2016; Levy *et al.*, 2015). 그러나 한국어 특성을 반영한 skip-gram 모델의 검증 방식은 거의 존재하지 않기 때문에 skip-gram을 포함한 Word2vec 모델의 하이퍼 파라미터 최적화 작업에는 어려움이 따른다(강형석 & 양장훈, 2019). 이에 따라 한국어 특성을 반영한 Word2vec 모델의 하이퍼 파라미터 최적값을 찾기 위한 시도가 있었다(강형석 & 양장훈, 2019; 조현수 & 이상구, 2017; 최상혁 등, 2016). 그 결과, 학습 방식은 skip-gram, 벡터 크기는 300, 윈도우 크기는 5~10 사이가 최적으로 도출된 바 있다. 다만 단어의 최소 출현 횟수는 연구에 따라 최적값으로 제시한 것이 다른데, 최상혁 등(2016)은 말뭉치 규모에 따라 적절히 큰 값으로 설정하는 것이 좋은 성능을 보인다고 주장한 반면, 강형석

& 양장훈(2019)은 말뭉치의 크기가 작은 경우에는 가능한 많은 어휘를 훈련에 포함시키는 것이 임베딩 성능 향상에 도움이 된다고 밝혔다. 이렇게 서로 다른 기준을 제시하는 이유는 훈련에 사용하는 데이터의 종류, 성격, 규모에 따라 임베딩 결과가 달라질 수 있기 때문이다. 따라서 연구에 활용되는 데이터에 여러 하이퍼 파라미터 조합을 적용함에 따라 임베딩 성능을 평가함으로써 최적값을 찾아야 한다.

skip-gram과 같은 임베딩 모델의 성능을 평가하기 위한 방법은 다양한데(Baroni *et al.*, 2014; Lai *et al.*, 2016; Schnabel *et al.*, 2015), 그중 가장 널리 쓰이는 방식은 유사도 검사(similarity test)와 유추 검사(analogy test)이다. 유사도 검사는 임베딩 결과인 단어 벡터 간의 유사도나 관련도를 평가하는데, 이때 평가의 기준이 되는 정답 데이터인 영어 기반 데이터셋<sup>26)</sup>이 존재한다. 유추 검사는 단어 벡터 간의 의미론적(semantic)·통사론적(syntactic) 유추 관계를 평가하며, 역시 기준이 되는 영어 데이터셋<sup>27)</sup>이 있다. 이러한 환경에서 한국어 기반 단어 임베딩 모델을 평가할 때 앞서 언급한 영어 데이터셋을 한국어로 번역해 활용한 경우가 많았다. 그러나 영어로 구축된 데이터를 단순 번역을 통해 한국어로 변환하는 것은 언어의 의미적·구조적 측면에서 한계가 따른다(강형석 & 양장훈, 2018).

본 연구는 단어나 문서가 아닌 POI를 임베딩하기 위해 임베딩 모델을 사용한다는 새로운 접근법을 제안하기에 기존의 임베딩 모델 평가 방법으로 POI 임베딩 모델의 성능을 평가하는 것은 적합하지 않다고 판단하였다. 따라서 본 연구에서는 POI 임베딩 결과를 비교하는 것으로 하이퍼 파라미터 설정에 따른 임베딩 모델의 성능 평가를 갈음하는 방법을 제안한다. 좀 더 구체적으로 설명하자면, 본 연구에서는 실험에 사용된 POI를 모두 벡터로 임베딩하므로 POI 간 코사인 유사도를 계산할 수 있다. 따라서 각 POI별로 코사인 유사도가 높게 산출된 상위 100개 POI를 추출해, POI 간 카테고리(음식 종류)가 일치하는 비율을 계산한다. POI의 여러 속성 중

---

26) ACLweb, [https://aclweb.org/aclwiki/WordSimilarity-353\\_Test\\_Collection\\_\(State\\_of\\_the\\_art\)](https://aclweb.org/aclwiki/WordSimilarity-353_Test_Collection_(State_of_the_art)) (접속일: 2020년 10월 24일)

27) ACLweb, [https://aclweb.org/aclwiki/Google\\_analogy\\_test\\_set\\_\(State\\_of\\_the\\_art\)](https://aclweb.org/aclwiki/Google_analogy_test_set_(State_of_the_art)) (접속일: 2020년 10월 24일)

카테고리는 POI에 대한 정보를 가장 직관적으로 판단할 수 있는 기준이기  
 에 POI 임베딩 결과를 평가하기 위한 근거로 활용하는 것이 적합하다고  
 판단하였다. 실험에 사용할 하이퍼 파라미터 값의 범위는 한국어 기반 임  
 베딩 모델의 하이퍼 파라미터 최적화 관련 선행 연구(강형석 & 양장훈,  
 2019; 조현수 & 이상구, 2017; 최상혁 등, 2016)을 참고로, 학습 알고리즘은  
 skip-gram으로 고정하고, 윈도우 크기가 3과 5일 때를 기준으로 단어 최  
 소 출현 횟수와 벡터 차원의 크기를 변화시키면서 결과를 비교한다. 윈도  
 우 크기를 3과 5로 설정한 이유는, 본 연구의 데이터 전처리 단계에서 형  
 태소 분석 후 특정 품사만을 추출하였기에, 기존에 최적값으로 도출된 최  
 소값인 5보다 작은 수의 윈도우 크기를 적용할 필요가 있다고 판단하였기  
 때문이다. 또한, 각 POI에 포함된 토픽별 키워드들의 벡터 합을 통해 POI  
 벡터를 확정할 때 토픽별 키워드에 대한 가중치인 TF-IDF 적용에 따른  
 성능도 함께 평가하고자 한다. 즉, TF-IDF 가중치 적용 여부에 따라 윈도  
 우 크기가 3과 5일 때, 단어 최소 출현 횟수를 1, 3, 5로 2씩 증가시키고,  
 벡터 크기를 50~400까지 50씩 증가시키면서 그 결과를 비교한다. 따라서  
 본 연구에서 임베딩 모델의 성능 평가를 위해 사용한 하이퍼 파라미터 조합  
 은 총 12개로, 전체 목록은 [표 2-2]와 같다.

[표 2-2] 실험에 사용한 하이퍼 파라미터 조합

연번	구분	윈도우 크기	키워드 최소 출현 횟수	벡터 크기
1	조건부 확률	3	1	50, 100, 150, 200, 250, 300, 350, 400
2			3	
3			5	
4		5	1	
5			3	
6			5	
7	TF-IDF 가중치	3	1	50, 100, 150, 200, 250, 300, 350, 400
8			3	
9			5	
10		5	1	
11			3	
12			5	

토픽별 키워드 및 POI 임베딩 결과를 시각적으로 평가하기 위해 고차원 벡터를 2차원으로 시각화하는 데 사용되는 t-stochastic nearest neighbor (이하 t-SNE)(Maaten & Hinton, 2008) 알고리즘을 활용한다. t-SNE는 원 공간인 고차원에서 유사하게 존재하는 두 벡터가 저차원인 2차원 공간에서도 유사할 수 있도록 원 공간의 벡터 간 유사도를 최대한 보존하는 방식으로 작동한다. 일반적으로 Word2vec을 기반으로 임베딩한 단어 벡터들은 그 크기가 200차원 이상이기 때문에 벡터값만 봐서는 해석하기 어렵다. 이에 따라 고차원의 원 벡터를 2차원의 공간으로 차원 감소시켜 각 벡터값을 시각적으로 확인하기 위한 목적으로 t-SNE를 사용한다.

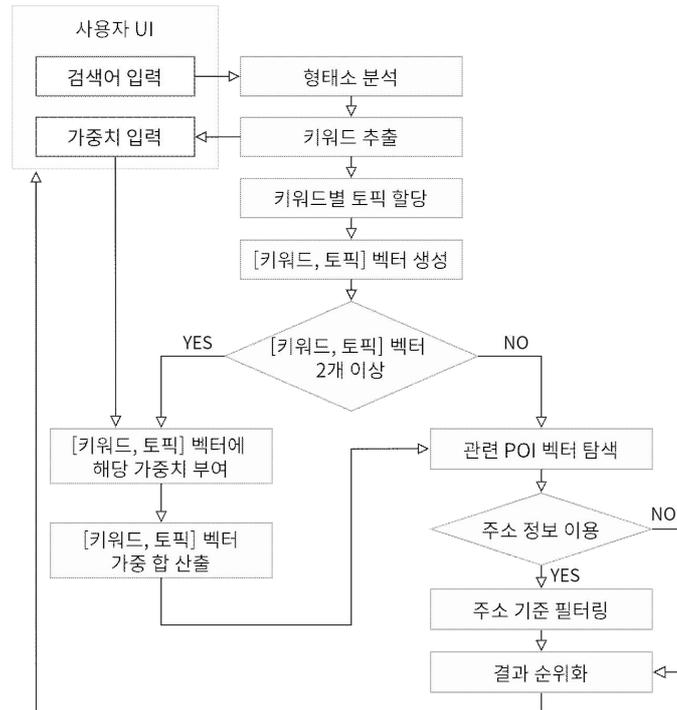
## 2.4 시맨틱 POI 검색에의 활용 구조

이번 절에서는 토픽별 키워드 기반 시맨틱 POI를 검색에 활용하기 위한 활용 구조에 대해 설명한다. 여기에는 자연어 기반 검색어에 대한 결과를 제공하는 자연어 기반 POI 검색, 사용자가 토픽과 토픽별 키워드를 선택하는 선택 기반 POI 검색, 그리고 특정 POI와 유사한 POI를 검색할 수 있는 유사 POI 검색에의 활용 구조가 포함된다. 자연어 및 선택 기반 POI 검색은 관련성에 기반한 직접검색 방식에 해당되며, 유사 POI 검색은 유사성에 기반한 추천검색 방식에 해당된다.

### 2.4.1 자연어 기반 POI 검색

본 연구에서 개발한 자연어 기반 POI 검색을 위한 활용 구조는 [그림 2-11]과 같다. 자연어 기반 검색과 선택 기반 검색의 가장 큰 차이점은 검색 가능한 키워드의 개수이다. 선택 기반 검색은 인터페이스의 특성상 사용자들에게 한 번에 보여줄 수 있는 키워드의 개수에 한계가 있다. 그러나 자연어 기반 검색에서는 사용자들의 다양한 검색어에 대응할 수 있도록 데이터베이스에 포함된 모든 토픽별 키워드를 활용한다. 자연어 기반 검색에서는 사용자가 직접 토픽을 선택하지 않아도 입력한 검색 키워드 조건에 맞는 토픽이 자동 부여되며, 이에 따른 결과가 출력된다. 사용자가 입력한 검색어 중 임베딩 데이터베이스에 포함된 키워드가 있는 경우, 이에 대한 가중치를 입력하도록 해당 키워드를 사용자에게 반환한다. 만약 사용자가 검색한 단어가 데이터베이스에 없는 경우에는 키워드가 반환되지 않으며, 따라서 검색을 수행할 수 없다. 만약 2개 이상의 키워드를 검색하는 경우에는 각 토픽별 키워드에 대한 가중치를 사용자가 직접 입력해야 하며, 이때 가중치의 합은 항상 1이 되어야 한다. 사용자가 부여한 가중치에 따라 같은 키워드일지라도 검색 결과로 도출되는 POI 목록은 달라질 수 있다. 즉, 사용자는 자신이 중요하게 생각하는 검색어에 더 높은 가중치를 부여함으로써 맞춤형 검색이 가능하다. 또한, 마지막 단계에서 관련성 높은

POI를 순위화해 사용자에게 보여줄 때 주소 정보를 이용해 원하는 지역의 POI만 필터링해 보여줄 수도 있다. 예를 들어, 검색 결과 중 강남구에 해당하는 POI만 필터링하고 싶은 경우, 주소 정보를 기반으로 강남구에 속하는 POI를 관련성 높은 순서대로 보여주는 것이다.



[그림 2-11] 자연어 기반 POI 검색을 위한 활용 구조

자연어 기반 검색과 선택 기반 검색의 두 번째 차이점은 토픽을 할당하는 방법에 있다. 선택 기반 검색의 경우, 토픽별 키워드가 사전에 정해져 있으므로 사용자가 직접 토픽과 그에 속하는 토픽별 키워드를 고를 수 있다. 그러나 자연어 기반 검색에서는 사용자가 자유롭게 검색어를 입력할 수 있도록 하면서 편의성을 높이기 위해 토픽 할당 알고리즘([표 2-3])을 통해 키워드에 대한 토픽이 자동으로 할당된다. [표 2-3]은 자연어 기반 POI 검색 프로세스에서 사용자가 입력한 검색 키워드에 토픽을 할당하는 알고리즘을 의사 코드(pseudo code)로 표현한 것이다. 복잡해 보이지만 원리는 간단하다. LDA 토픽 모델링 결과에서, 각 키워드는 1개 이상의 토픽

에 할당되는데, 문장에서 핵심적 역할을 하는 명사의 경우, LDA 결과로 얻은 키워드별 토픽 분포 중 가장 높은 확률값을 가지는 토픽으로 할당한다. 명사가 아닌 형용사, 동사, 어근의 경우, 명사와 함께 사용되었을 때는 해당 명사에 할당된 토픽과 같은 토픽에 할당할 수 있는지를 우선 검토한다. 이는 용언(형용사, 동사, 어근)이 함께 사용된 체언(명사)의 의미에 종속될 확률이 높기 때문이다. 만약 명사 외 키워드의 토픽 분포에, 함께 사용된 명사의 토픽이 포함되어 있지 않으면 해당 키워드의 토픽 분포 중 가장 높은 확률의 토픽으로 할당한다.

[표 2-3] 키워드별 토픽 할당 알고리즘 의사 코드

```

if (검색어 리스트에 명사만 있음):
    해당 명사의 토픽 분포 중 가장 높은 확률의 토픽 할당
elif (검색어 리스트에 1개의 명사와 1개 이상의 명사 아닌 키워드가 함께 있음):
    if (키워드가 명사):
        해당 명사의 토픽 분포 중 가장 높은 확률의 토픽(예: t1) 할당
    else (키워드가 명사가 아님):
        if (해당 키워드의 토픽 분포에 t1이 포함되어 있음):
            해당 키워드에 t1 토픽 할당
        else (해당 키워드의 토픽 분포 중 t1이 포함되어 있지 않음):
            해당 키워드의 토픽 분포 중 가장 높은 확률의 토픽 할당
elif (검색어 리스트에 1개 이상의 명사와 1개 이상의 명사 아닌 키워드가 함께 있음):
    if (키워드가 명사):
        해당 명사의 토픽 분포 중 가장 높은 확률의 토픽(예: t2(명사 1), t3
        (명사 2)) 할당
    else (키워드가 명사가 아님):
        if (해당 키워드의 토픽 분포에 t2 또는 t3가 포함되어 있음):
            t2와 t3 중에 확률값이 더 큰 토픽 할당
        else (해당 키워드의 토픽 분포에 t2와 t3가 모두 포함되어 있지 않음):
            해당 키워드의 토픽 분포 중 가장 높은 확률의 토픽 할당
else (검색어 리스트에 1개 이상의 명사 아닌 키워드만 있음):
    해당 키워드의 토픽 분포 중 가장 높은 확률의 토픽 할당

```

자연어 기반 POI 검색에 대한 서비스 화면을 설계한 결과는 [그림 2-12]~[그림 2-14]와 같다. 서비스 화면 설계 시 사용자가 메인 화면([그림 2-12])에서 검색어로 ‘식물 덕후가 좋아할 만한 곳’을 입력한 상황을 가정하였다. 이때 임베딩 데이터베이스에 포함되어 사용자에게 가중치를 입력하도록 반환되는 키워드는 ‘식물’, ‘덕후’, ‘좋아하다’로, 각각 0.5, 0.3, 0.2의 가중치를 입력한 경우를 예시로 사용하였다([그림 2-13]). 가중치의 합은 1이 되어야 하며, 검색 결과로는 입력된 토픽별 키워드에 부여된 가중치를 기준으로 관련도가 높은 상위 n개의 POI가 출력된다([그림 2-14]).



[그림 2-12] 자연어 기반 POI 검색 메인 화면



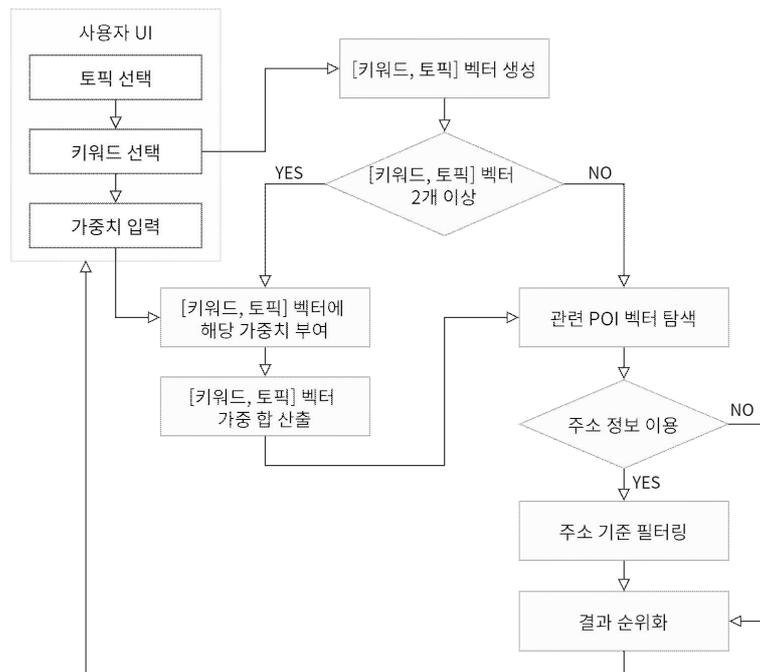
[그림 2-13] 자연어 기반 POI 검색의 가중치 입력 화면



[그림 2-14] 자연어 기반 POI 검색의 결과 출력 화면

## 2.4.2 선택 기반 POI 검색

선택 기반 POI 검색을 위한 활용 구조는 [그림 2-15]와 같다. 해당 그림을 보면, 선택 기반 POI 검색에서는 먼저 사용자가 검색을 원하는 토픽을 선택한 후, 해당 토픽에 포함되는 키워드를 선택해야 한다. 이때 1개 토픽에서 1개 이상의 키워드를 선택해도 되고, 2개 이상의 토픽에서 각각 키워드를 선택해도 된다. 다만 반드시 1개 이상의 토픽별 키워드를 선택해야 결과가 출력된다. 만약 2개 이상의 키워드를 선택하는 경우에는 각 토픽별 키워드에 대한 가중치를 사용자가 직접 입력해야 하며, 이때 가중치의 합은 항상 1이 되어야 한다. 자연어 기반 검색에서와 마찬가지로 사용자가 부여한 가중치에 따라, 동일 키워드라도 검색 결과로 도출되는 POI 목록은 달라질 수 있다. 또한, 마지막 단계에서 POI를 순위화하여 사용자에게 보여줄 때 주소 정보를 이용해 원하는 지역에 위치한 POI만 필터링해 보여줄 수도 있다.



[그림 2-15] 선택 기반 POI 검색을 위한 활용 구조

선택 기반 POI 검색에서 사용자에게 표출되는 토픽별 키워드는 LDA 모델링 과정에서 추출된 키워드 중 상위 20개이다. LDA와 같은 토픽 모델링 방법론의 목적은 무감독 방식으로 텍스트를 요약하기 위한 것이기 때문에 토픽별로 보여줘야 할 키워드 개수에 대한 기준이 따로 정해져 있지는 않다. 그러나 LDA를 활용하는 다수의 선행 연구(Chen *et al.*, 2015; Krasnashchok & Jouili, 2018; Maier *et al.*, 2018; Song *et al.*, 2009; Viegas *et al.*, 2019; Xu *et al.*, 2017)에서는 상위 5~20개의 단어를 사용하는 경향이 있는데, 20개 이상의 단어를 사용하게 되면 텍스트를 간략하게 요약하려는 목적이 무효화될 수 있기 때문이다. 또한, 웹 기반 POI 검색 시 화면상에 너무 많은 키워드를 노출시킬 경우 사용자에게 혼란을 야기할 수 있으므로 앞서 언급한 선행 연구들을 참고해 키워드의 개수를 최대 20개로 한정하였다. 더욱 다양한 토픽별 키워드에 대한 검색은 자연어 기반 POI 검색에서 가능하다.

선택 기반 POI 검색에 대한 서비스 화면을 설계한 결과는 [그림 2-16]~[그림 2-19]와 같다. 서비스 화면 설계 시 사용자가 ‘고기’ 토픽을 선택한 후 키워드로는 ‘삼겹살’, ‘짬’을 선택하고 가중치를 각각 0.8, 0.2로 할당한 상황을 가정하였다. 여기서 토픽과 토픽별 키워드를 표출할 때 격자형(grid) 메뉴([그림 2-16] 및 [그림 2-17] 참조)를 사용하였다. 격자형 구조는 모든 메뉴에 대한 항목을 한 화면에서 2차원 배열로 볼 수 있으므로 사용자로 하여금 빠른 선택을 가능하게 한다는 장점이 있다(Christie *et al.*, 2004; Shneiderman & Plaisant, 2010; 김명지 등, 2015; 정홍인, 2005).

사용자의 검색 시나리오를 순차적으로 설명하면 먼저, [그림 2-16]과 같이 메인 화면에서 토픽이 격자형 메뉴로 표출되는데, 여기서 검색을 원하는 토픽을 선택하면 [그림 2-17]과 같이 키워드에 대한 격자 메뉴가 확장된다. 사용자는 그것을 보고 검색을 원하는 토픽별 키워드를 선택해야 한다. 토픽별 키워드를 2개 이상 선택한 경우에는 [그림 2-18]과 같이 각 키워드에 대한 가중치를 직접 입력해야 하며, 가중치의 합은 1이 되어야 한다. 검색 결과로 [그림 2-19]와 같이, 입력된 토픽별 키워드와 관련도가 높은 상위 n개의 POI가 출력된다. 이때 검색 결과는 가중치를 높게 설정한 키워드와 더 많

은 관련성이 있다.



[그림 2-16] 선택 기반 POI 검색 메인 화면



[그림 2-17] 선택 기반 POI 검색의 토픽별 키워드 선택 화면



[그림 2-18] 선택 기반 POI 검색의 가중치 입력 화면

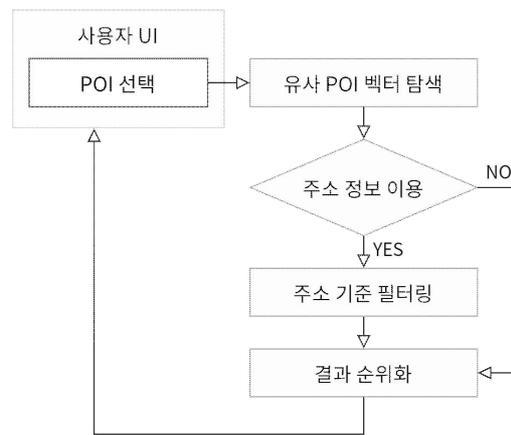


[그림 2-19] 선택 기반 POI 검색의 결과 출력 화면

### 2.4.3 유사 POI 검색

본 연구에서는 토픽별 키워드 기반 POI 임베딩 기술을 통해 주어진 모든 POI를 동일한 벡터 공간상에 임베딩한다. 이에 따라 코사인 유사도를 통해 POI 간 유사도를 계산할 수 있다. 유사한 장소를 검색하는 것은 지리 정보 검색 분야의 중요한 작업 중 하나로, 관광, 부동산, 이사 등 실생활과 밀접한 관련이 있다. 가령 사람들은 새로운 도시로 이주할 때, 새로운 장소에서 이전에 좋아하던 곳과 유사한 곳을 알고 싶어 할 수 있다(Gao *et al.*, 2017; Shoji *et al.*, 2018).

유사 POI 검색을 위한 활용 구조는 [그림 2-20]과 같다. 유사 POI 검색은 자연어 및 선택 기반 검색과 다르게 추천검색에 속한다. 추천검색은 사용자가 입력/선택한 결과로 도출된 POI를 기준으로 유사한 POI를 추천해주는 방식이다. 여기에서도 마지막 단계에서 유사도 높은 POI를 순위화해 출력할 때 주소 정보를 이용해 원하는 지역의 POI만 필터링해 보여줄 수도 있다.



[그림 2-20] 유사 POI 검색을 위한 활용 구조

유사 POI 검색에 대한 서비스 화면을 설계한 결과는 [그림 2-21]과 같다. 서비스 화면 설계 시 사용자가 이전 검색 결과로 도출된 POI 중 ‘고고초밥’이라는 POI를 선택한 상황을 가정하였다. 검색 결과로는 기준 POI와 유사도 가장 높은 상위 n개의 POI가 출력된다.



[그림 2-21] 유사 POI 검색의 결과 출력 화면

## 2.5 종합적인 사용자 평가

시맨틱 POI 검색에의 활용 결과에 대한 품질 검증을 위해 설문 조사를 통한 종합적인 사용자 만족도 평가를 실시한다. 본 연구에서는 서울에 위치한 음식점을 대상으로 실험을 진행하므로 모집단 크기를 서울시 경제활동인구인 5,322,000명<sup>28)</sup>으로 상정한다. 이때 신뢰수준 95%, 허용오차 5% 범위에서 요구되는 최소 표본 크기( $n$ )는 식 (2-12)에 따라 385로 산출되었기에, 본 연구에서는 총 400명에 대한 설문 조사를 실시한다. 식 (2-12)에서  $N$ 은 표본 크기,  $e$ 는 허용오차,  $Z_{\alpha/2}$ 는 신뢰수준에 대응하는 z-score를 의미하는데, 가장 많이 사용하는 95% 신뢰수준에서의 z-score 값은 1.96이다.  $p$ 는 관찰치로, 최대 허용오차를 구하는 경우, 0.5를 사용한다.

$$n = \frac{\frac{Z_{\alpha/2}^2 \cdot p(1-p)}{e^2}}{1 + \left(\frac{Z_{\alpha/2}^2 \cdot p(1-p)}{e^2 N}\right)} \quad (2-12)$$

설문은 총 4회로 나누어 실시하며, 설문 1회당 50개 문항에 대해 100명의 응답을 수집한다. 이때 각 문항세트는 [표 2-4]와 같이 POI 검색에의 활용 결과에 대한 종합적인 평가가 이루어질 수 있도록 검색 종류별로 동일한 비율의 문항으로 구성한다. 즉, 자연어 및 선택 기반 POI 검색에서는 키워드 개수(1~4개)별로 5개씩 각 20개 문항으로 구성하며, 유사 POI 검색에서는 10개 문항을 생성한다. 4개의 문항세트는 모두 [표 2-4]와 같은 비율로 구성하나 세부 문항 내용은 다르므로 결과적으로 총 200개의 서로 다른 문항에 대한 응답을 수집하게 된다.

28) 국가통계포털, [http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=INH\\_1DA7004\\_11](http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=INH_1DA7004_11)  
(접속일 : 2020년 10월 5일)

[표 2-4] 시맨틱 POI 검색 종류별 설문 구성(문항 1세트 기준)

구분	키워드 개수별 문항 개수				합계
	1개	2개	3개	4개	
자연어 기반 POI 검색	5	5	5	5	20
선택 기반 POI 검색	5	5	5	5	20
유사 POI 검색	문항 개수				
	10				10
문항 개수 합					50

설문 시 문항 개수를 50개씩 나눠 1명당 응답해야 할 질문의 수가 50개를 초과하지 않도록 하였다. 그 이유는 설문의 길이가 길어질수록 같은 응답을 반복할 가능성이 크고, ‘모르겠다’고 응답하는 비율과 무응답 비율이 증가하는 등 설문의 질이 떨어질 확률이 크기 때문이다(Deutskens *et al.*, 2004; Herzog & Bachman, 1981). 김권현 등(2015)은 설문지의 길이가 응답의 질에 미치는 영향에 대해 연구하였는데, 이에 따르면 설문 소요 시간이 약 50분 정도인 설문에서 최대 한 줄 응답 길이<sup>29)</sup>가 유의미하게 증가하는 등 반응의 질이 떨어지는 것으로 나타났다. 반면 설문 소요 시간이 약 20분 정도인 설문에서는 이러한 응답의 질 하락 현상이 발견되지 않았다. 또한, 온라인 설문 기관인 서베이몽키(Surveymonkey)<sup>30)</sup>에 의하면 87%의 사람들이 20분 이상 소요되는 설문에 응하는 것을 원하지 않는다. 본 연구에서 작성한 설문 문항 1개당 응답 소요 시간이 약 20~30초 정도로 예상됨에 따라, 20분 이내로 설문을 마치기 위해서는 전체 문항 수가 50개 이하여야 한다. 이에 따라 전체 문항 200개를 4개 세트로 나누어 각 50문항당 100명씩 응답하도록 함으로써 총 400명 대상의 설문 조사를 실시한다. 설문 진행은 모바일 전문 설문 조사 기관인 오픈서베이(opensurvey)<sup>31)</sup>

29) 연속된 리커트 척도 문항에서 같은 번호의 응답을 연속적으로 한 것 중 가장 긴 한 줄 응답의 길이(김권현 등, 2015)

30) 서베이몽키, <https://www.surveymonkey.com/curiosity/eliminate-survey-fatigue-fix-3-things-respondents-hate/> (접속일: 2020년 10월 5일)

31) 오픈서베이, <https://www.opensurvey.co.kr/> (접속일: 2020년 10월 5일)

플랫폼을 이용한다. 오픈서베이를 통해 실시하는 모든 설문은 스마트폰 환경에 최적화되어 있기 때문에 설문지를 구성할 때 1개 문항에 대한 내용이 한 화면에 모두 표출될 수 있도록 설계하는 것이 좋다. [그림 2-22]를 보면, 응답자는 각 문항에 주어진 검색 키워드에 대한 결과로 도출된 POI 목록을 보고 만족하는 정도에 따라 ‘매우 불만족(1점)’에서 ‘매우 만족(5점)’까지 리커트(Likert) 5점 척도로 평가해야 한다.

본 설문에서 응답자가 만족도를 평가할 때 근거가 되는 것은 결과로 도출된 각 POI에 대한 음식 종류 및 대표 리뷰 항목이다. 설문 구성 시 본 연구에서 제안한 시맨틱 POI 검색 결과를 실험군으로 놓고, 네이버 지도 등 다른 POI 검색 시스템의 검색 결과를 대조군으로 활용해 비교 평가하는 방법도 생각해 볼 수 있다. 그러나 본 연구에서는 그러한 비교 평가 방식을 사용하지 않았는데, 그 이유는 POI 명칭이 아닌 검색 키워드별 결과로 도출된 POI 목록에 대한 절대적 기준의 정답이 존재하지 않고, 본 연구에서 제안한 검색 기술은 기존의 POI 검색 기술과 상호 보완적으로 활용되도록 하는 목적이 있으므로 비교 평가의 의미가 없다고 판단하였기 때문이다. 설문에 대한 가장 정확한 평가를 할 수 있는 응답자는 결과로 도출된 모든 POI에 방문한 경험이 있거나 각 POI에 대해 잘 알고 있는 사람이다. 그러나 50개 문항에 포함된 최대 500개 POI를 모두 방문했거나 잘 알고 있는 사람들만을 대상으로 설문을 시행하는 것은 현실적으로 불가능하다. 따라서 그에 갈음하는 정보로 음식 종류 및 대표 리뷰 항목을 참고하도록 하였다. 더욱 자세한 설문 내용은 <부록 A.1>~<부록 A.5>에 첨부하였다.

맛집 검색 만족도 조사\_a 나가기 X

**'삼겹살(중요도 50%)', '맛있는(중요도 50%)'** 곳을 검색했을 때 관련 장소로 아래 같은 결과가 나온다면 만족하시나요?

\*검색 결과는 중요도가 높은 단어와 더 많은 관련이 있습니다.

순위	장소명	음식종류	대표리뷰
1	팔색삼겹살 (마포구서교동)	고기 요리	진짜 맛있다!
2	돼지 집 (강남구신사동)	고기 요리	생삼겹살 전문점입니다!! 우선 자갈 밑에 숯불을 넣고 절편들을 간 다음 자갈을 올립니다 그 뒤에 자갈이 달궈지면 소량씩 먹음만큼씩만 삼겹살을 올려서 먹는데!! 정말 맛있었어요!!!
3	코기코기 (강남구신사동)	고기 요리	허브 삼겹살이 맛있다고 해서 간 곳. 처음엔 냉동고기가 나와서 놀랐는데 먹어보니 대박 끝맛! 삼겹살 종류가 다양해서 골라먹는 재미도 쏠쏠하다.
4	행복추종령 활삼겹살 (강남구신사동)	고기 요리	고기도 맛있고 가격도 적당함! 압구정에서 통편하는 곳 중 하나
5	육전식당 (강남구역삼동)	고기 요리	태어나서 먹어본 돼지고기 중 가장 맛있었네요! 더 이상의 설명은 필요 없을 것 같습니다
6	고기요 (은평구수색동)	고기 요리	맛있어요! 반찬도 맛있고 (양념게장 나옵니다) 된장찌개도 훌륭합니다!
7	맛찬들 왕소금구이 (강서구마곡동)	고기 요리	숙성된 삼겹살이라 그런지 가격은 좀 비쌌지만 직접 구워주시고 고기와 반찬이 맛있던 집.
8	돌고기506 (강남구역삼동)	고기 요리	돼지고기 냄새가 이렇게 구수한 냄새와 맛있었구나!!! 재방문 의사 80%
9	고기스토리 (동작구상도동)	고기 요리	육식에서 삼겹살 제일 맛있고 싼곳 .. 그리고물냉면맛집임
10	양철집 (강남구신사동)	고기 요리	아... 생삼겹은 이래야 하는구나... 라는 기준을 정해 주는 집! 육사발과 김밥씩 안먹으면 섭섭함

개 선택

매우 불만족      보통      매우 만족

맛집 검색 만족도 조사\_a 나가기 X

**'건강한(중요도 70%)', '브런치(중요도 30%)'** 를 검색했을 때 관련 장소로 아래 같은 결과가 나온다면 만족하시나요?

\*검색 결과는 중요도가 높은 단어와 더 많은 관련이 있습니다.

순위	장소명	음식종류	대표리뷰
1	아까H (마포구연남동)	기타 양식	무엇보다 건강한 식사를 할 수 있어서 좋아요 :) 채식하는 분들을 위한 메뉴가 마련되어 있고, 채식을 하지 않더라도 맛있게 즐길 수 있는 맛!
2	페르에피스 (강남구도곡동)	브런치/버거/샌드위치	엔틱한 분위기에서 맛과 가격 모두 부담스럽지 않은 브런치를 먹을 수 있다
3	블랙랩 (강남구상성동)	카페/디저트	분위기가 있는 브런치 카페!
4	아티제 (서초구서초동)	카페/디저트	아티제는 베이커리가 맛있는 줄 알았는데 브런치도 맛있군요!
5	마고정 (은평구진관동)	국수/면 요리	한끼라도 건강해보자라고 계란샐러드가 아보카도 추가해 보았는데.. 깔끔한 외관에 비해 뭔가 굉장하... 엄청하고 특박한 느낌의 샐러드가 나왔지요.
6	그레인 (마포구연남동)	카페/디저트	5가지 브런치와 오후일 런치크를 먹었다. 브런치는 건강한 느낌이고 특별한 맛은 없었고 양이 적었다.
7	카페몽스 (종로구명륜4가)	브런치/버거/샌드위치	대학로에 있는 괜찮은 브런치 카페! 브런치 세트는 두 종류고 그 외에 다른 메뉴들도 전반적으로 맛있었다! 무엇보다 브런치 메뉴 시간 제한이 넉넉해서 자주 감.
8	스토리 (강서구화곡동)	카페/디저트	브런치 메뉴도 있지만 밤에는 술이랑 안주를 팔기도 하는 곳이다
9	씨니브레드 (영선구한남동)	베이커리	비건과 저탄수화물을 주제로 한 조그마한 베이커리!
10	찰떡김밥 (영천구목동)	베이커리	좋은재료로 건강한 찰떡밥을 만든곳.

개 선택

매우 불만족      보통      매우 만족

<      다음 문항 >

[그림 2-22] 모바일 설문조사 화면 캡처

## 3. 실험 적용 및 결과

### 3.1 실험 환경

실험은 Intel(R) Core(TM) i7-9700KF CPU @ 3.60GHz, 64.0GB 메모리, Windows 10 Home 64비트 환경에서 진행하였다. 해당 컴퓨터 사양은 본 연구에서 사용한 연구 방법에 대한 실험을 수행하기에 충분한 CPU 및 메모리를 제공한다. 데이터 수집, 전처리, 토픽 모델링, 그리고 POI 임베딩에 활용한 프로그래밍 언어는 Python 3.8.3이다. Python은 데이터 처리, 수치연산, 머신러닝, 딥러닝 등 다양한 라이브러리를 제공하고 있기에 관련 작업 수행 시 편의성이 좋은 장점이 있다.

### 3.2 데이터 수집 및 전처리 결과

#### 3.2.1 데이터 수집 결과

망고플레이트에서는 지역, 식당 또는 음식에 해당하는 검색어를 입력해야만 그 결과로 반환되는 POI 데이터를 수집할 수 있다. 본 연구에서는 실험 대상 지역을 서울시로 한정하고, 서울 내 맛집 관련 POI에 대해 작성된 리뷰 데이터를 수집하기 위해 서울시의 25개 구 명칭, 노선별 지하철역 604개 명칭<sup>32)</sup>, 음식 카테고리 등을 검색어로 활용하였다. 웹 스크래핑을 실시한 기간은 2019년 1월 14일부터 약 2주 간이다. 데이터 수집 결과 30,819개의 POI를 수집하였는데, 이 중 사용자가 작성한 리뷰가 1개 이상 존재하는 POI는 21,021개로, 이는 수집한 전체 POI의 약 68.2%에 해당하는 수치다. 1개 이상의 리뷰를 포함하는 21,021개의 POI

---

32) 서울 열린데이터 광장, <http://data.seoul.go.kr/dataList/OA-121/S/1/datasetView.do>  
(접속일: 2020년 10월 5일)

에 대해 사용자가 작성한 리뷰의 개수는 244,785개로, POI당 평균 리뷰 수는 약 11.7개인 것을 알 수 있다([표 3-1]).

[표 3-1] 망고플레이트 데이터 수집 현황

구분	개수	비고	
a	서울시를 대상으로 수집한 POI 수	30,819	-
b	a 중에서 리뷰 1개 이상을 포함하는 POI 수	21,021	a의 약 68.2%
c	b에 작성된 리뷰 수	244,785	POI당 평균 리뷰 수 약 11.7개

2017년의 서울시 통계에 의하면 서울 소재 음식점 수는 총 80,732개로, 한식, 치킨 전문점, 중식, 일식 순으로 많다<sup>33)</sup>. 또한, 서울 내 카페 수는 같은 해 4월 기준, 18,440개이다<sup>34)</sup>. 이러한 통계에 비추보았을 때 본 연구에서 수집한 맛집 POI(30,819개)는 서울시에 위치하는 음식점 및 카페 (99,172개)의 약 31.1%에 해당하는 수치일 것으로 추정된다.

33) 서울특별시(2020-02-11), [https://news.seoul.go.kr/gov/archives/511536?tr\\_code=sweb](https://news.seoul.go.kr/gov/archives/511536?tr_code=sweb) (접속일: 2020년 10월 5일)

34) 매일경제(2017-05-23), <https://www.mk.co.kr/news/realstate/view/2017/05/344342/> (접속일: 2020년 10월 5일)

### 3.2.2 데이터 전처리 결과

본 연구에서는 하나의 POI에 대해 한 명의 사용자가 작성한 리뷰 텍스트 전체를 한 개의 문서로 취급한다. 형태소 분석을 마친 각 리뷰는 일반 명사, 고유명사, 동사, 형용사, 그리고 어근 품사만 남게 되는데, 이때 품사의 종류와 관계없이 5개 이상의 단어를 포함하는 리뷰만을 분석 대상으로 삼았다. 그 결과 총 208,560개의 리뷰 데이터가 추출되었으며, 이에 해당하는 POI 개수는 19,902개이다. 앞서 작성한 [표 3-1]을 데이터 전처리를 마친 시점에서 다시 정리하면 [표 3-2]와 같다.

[표 3-2] 망고플레이트 데이터 전처리 후 현황

구분	개수	비고
a	30,819	-
b	21,021	a의 약 68.2%
c	244,785	POI당 평균 리뷰 수 약 11.7개
d	208,560	c의 약 85.2%
e	19,902	a의 약 64.6%, b의 약 94.7%

POI별 리뷰 텍스트 데이터에 형태소 분석을 시행한 후 특정 품사만 추출한 예시는 [표 3-3]과 같다. 이때 향후 분석을 용이하게 하기 위해 형태소와 해당 품사를 밑줄 기호(\_)로 연결시켜 하나의 키워드로 취급하였다.

[표 3-3] 망고플레이트 리뷰 텍스트 형태소 분석 결과 일부

POI 명칭 (주소)	음식 종류	a. 사용자 작성 리뷰 텍스트
		b. a에 대한 형태소 분석 후 특정 품사*만 추출한 결과
경춘자의 라면 뽕기는날 (종로구 화동 138-21)	기타 한식	<p>짬뽕라면 맛있게 매움 스트레스 풀고 싶을 때 매운게 뽕기면 어김없이 떠오르는 곳 ㅎㅎ 매운거 좋아하는 분이면 좋아할만한 라면집이에용 가격도 착합니다^^</p> <p>'짬뽕_NNG', '라면_NNG', '맛있_VA', '맵_VA', '스트레스_NNG', '풀_VV', '때_NNG', '맵_VA', '뽕기_VV', '떠오르_VV', '곳_NNG', '맵_VA', '좋아하_VV', '좋아하_VV', '라면_NNG', '집_NNG', '이_NNG', '가격_NNG', '착하_VA'</p>
덕화장 (중구 장충동2가 12-3)	정통 중식/ 일반 중식	<p>탕수육과 짜장면. 기본이 잘 갖춰져있는 맛집!</p> <p>'탕수육_NNG', '짜장면_NNG', '기본_NNG', '갖추_VV', '맛집_NNG'</p>
류 (용산구 이촌동 300-23)	정통 일식	<p>분위기 너무 좋고 친절하신 사장님 사모님 덕에 소소한 가족 모임이나 친한 친구들과 가게 되는 곳.. 스시는 말할 것도 없고 저녁에 도란도란 먹는 정식이 특히 정갈하고 맛있다 :) 추운 겨울날 두툼한 사시미에 따뜻한 도꾸리 한잔이 생각 나는 곳... 선도높은 사시미부터 반찬 하나하나까지 정갈하고 맛있다^^</p> <p>'분위기_NNG', '좋_VA', '친절_NNG', '사장_NNG', '사모님_NNG', '덕_NNG', '소소_XR', '가족_NNG', '모임_NNG', '친하_VA', '친구_NNG', '가_VV', '되_VV', '곳_NNG', '스시_NNP', '말_NNG', '없_VA', '저녁_NNG', '떡_VV', '정식_NNG', '정갈_XR', '맛있_VA', '츩_VA', '겨울날_NNG', '두툼_XR', '사시미_NNG', '따뜻_XR', '도_NNG', '꾸_VV', '한잔_NNG', '생각_NNG', '곳_NNG', '선도_NNG', '높_VA', '사시미_NNG', '반찬_NNG', '하나하나_NNG', '정갈_XR', '맛있_VA'</p>
명동피자 (송파구 문정동 634)	이탈 리안	<p>피자가 너무 먹고싶어서 갔던곳 피자파스타집은 다 거기서 거기라는 생각으로 갔는데 기대이상이었다ㅎ 적당한 가격에 적당한 양 적당한 맛이었다 느끼하지 않게 잡아주는 매콤함이 매력적이었던 요리!! 은근 생각나는 메뉴들이다 별점 4/5점</p> <p>'피자_NNG', '먹_VV', '가_VV', '곳_NNG', '피자_NNG', '파스타_NNP', '집_NNG', '생각_NNG', '가_VV', '적당_XR', '가격_NNG', '적당_XR', '적당_XR', '맛_NNG', '느끼_XR', '잡_VV', '매콤_NNP', '함_NNG', '매력_NNG', '요리_NNG', '은근_XR', '생각나_VV', '메뉴_NNG', '별점_NNG'</p>
미엘 (강남구 청담동 94-3)	카페/ 디저트	<p>카페에 볶음밥이????!?!?! 참신한데 맛있다 분위기도 좋음</p> <p>'카페_NNG', '볶음밥_NNP', '참신_XR', '맛있_VA', '분위기_NNG', '좋_VA'</p>

\*일반명사(NNG), 고유명사(NNP), 동사(VV), 형용사(VA), 어근(XR)

[표 3-3]에서 ‘경춘자의 라면 땡기는 날’의 결과를 보면, ‘매움’, ‘매운게’, ‘매운거’라는 서로 다른 단어가 형태소 분석을 통해 모두 같은 형태소인 ‘맵\_VA’로 분류된 것을 알 수 있다. 또한, ‘좋아하는’과 ‘좋아할만한’이라는 단어 역시 모두 같은 형태소인 ‘좋아하\_VV’로 분류되었다. 반면, ‘라면집이 예용’이라는 단어는 ‘라면\_NNG’, ‘집\_NNG’, ‘이\_NNG’로 분리되었는데, 여기서 원래 ‘이\_NNG’는 일반명사(NNG)가 아니라 긍정지정사(VCP)로 분리되어야 맞다. 이 같은 오류가 발생한 이유는 해당 단어가 표준 맞춤법에 맞지 않기 때문이다. 실제로 ‘라면집이에요’라는 맞춤법에 맞는 표현을 형태소 분석해 보면, ‘이’를 긍정지정사로 분리한다. 본 연구에서는 이렇게 표준 맞춤법을 지키지 않은 단어에 대해 품사가 오부착된 형태소들을 따로 제거하지 않았다. 이러한 품사 오부착 문제는 모든 형태소 분석기가 가지고 있는 한계로, 한국어에 대해 완벽한 형태소 분석을 수행하는 분석기는 아직 존재하지 않는다. 최근 자연어 처리 분야에서 신경망 기반으로 한국어 형태소 분석기의 성능을 개선하는 연구(이건일 등, 2017; 최병서 등, 2020; 최용석 & 이공주, 2020)가 진행됨에 따라 형태소 분석의 품질은 점차 향상될 수 있을 것으로 보인다.

### 3.3 LDA 기반 토픽 모델링 결과

LDA 기반 토픽 모델링 시 사후 확률의 근사치 추정 방법, 반복 수행(iteration) 횟수, 하이퍼 파라미터인  $\alpha$ 와  $\eta$  값, 그리고 토픽 개수를 사전에 설정해야 한다. 본 연구에서는 먼저, 사후 확률의 근사치 추정 시 정확성과 속도를 개선한 것으로 알려진 붕괴된 깃스 샘플링 방법(Griffiths & Steyvers, 2004; 김용대 & 정구환, 2016; 유예림, 2017)을 사용하였다. 반복 수행 횟수는 선행 연구(안주영 등, 2016; 유예림, 2017)에서 권고하는 값인 1,000으로 설정하였다. 또한, 하이퍼 파라미터 값은 선행 연구(Wallach, 2008; Wallach *et al.* 2009; 김은희 등, 2012)에서 최적 조합으로 입증된 비대칭  $\alpha$ 와 대칭  $\eta$  조합을 사용하였으며, 이때  $\alpha$  및  $\eta$  값은 가장 일반적으로 사용되는  $50/k$ (토픽 개수)과 0.01로 각각 설정하였다.

LDA를 구현하는 방법은 프로그래밍 언어별로 다양한데([표 3-4] 참조), 본 연구에서는 LDA 구현 시 속도가 빠르고 토픽 분리 성능이 비교적 뛰어난 것으로 알려진 Java 기반의 machine learning for language toolkit (이하 *mallet*)<sup>35)</sup>을 활용하였다. *mallet*은 통계적 자연어 처리, 문서 분류, 클러스터링, 토픽 모델링 등을 지원하는 패키지로, 원래 Java 기반으로 개발되었으나 Python의 자연어 처리 라이브러리인 *gensim*에서 *mallet* 기반으로 LDA를 구현하는 래퍼(wrapper)<sup>36)</sup>인 *wrappers.LdaMallet* API를 제공하고 있어, 이를 활용해 LDA 모델링을 실시하였다. 해당 API는 본 연구에서 설정한 사후 확률 근사치 추정 방법, 반복 수행 횟수, 그리고 하이퍼 파라미터인  $\alpha$ 와  $\eta$  값을 모두 기본값으로 설정하고 있다([표 3-4] 참조).

---

35) *mallet*, <http://mallet.cs.umass.edu/> (접속일: 2020년 10월 5일)

36) *gensim*, <https://radimrehurek.com/gensim/models/wrappers/ldamallet.html> (접속일: 2020년 10월 5일)

[표 3-4] 토픽 모델링 분석 도구 비교  
(유예림(2017) p.73 <표 II-11> 재구성)

구분	Java		Python	R	
	mallet		gensim	topicmodels	lda
사후 확률의 근사치 추정 방법	collapsed gibbs sampling		collapsed gibbs sampling, VEM <sup>37)</sup>	collapsed gibbs sampling, VEM	collapsed gibbs sampling
하이퍼 파라미터 (기본값)	$\alpha$	50/k, 비대칭적	1/k, 대칭적	50/k	-
	$\eta$	0.01	1/k, 대칭적	0.1	-

장소 리뷰 키워드로부터 토픽을 추출하기 위해서는 먼저 토픽의 개수를 정해야 한다. 본 연구에서는 LDA를 포함한 토픽 모델링 기법에서 적절한 토픽 개수를 추정하기 위한 평가 방법으로 주로 사용되는 복잡도와 토픽 일관성 측도 중 하나인  $C_V$  값을 모두 산출해 상호 보완적으로 활용하고자 하였다. 이때, 형태소 분석이 완료된 후 특정 품사(일반명사, 고유명사, 동사, 형용사, 어근)만 추출해 형태소와 품사를 밑줄 기호로 연결한 키워드 ([표 3-3]의 b)가 LDA의 입력값으로 활용된다. 토픽 개수에 따른 복잡도를 계산할 때 최대 토픽 수를 20개로 설정하였는데, 그 이유는 본 연구에서 수집한 POI 데이터를 음식 종류 기준으로 분류해 유사한 것들끼리 묶은 결과, 19개 정도로 압축되기 때문이다([표 3-5] 참조).

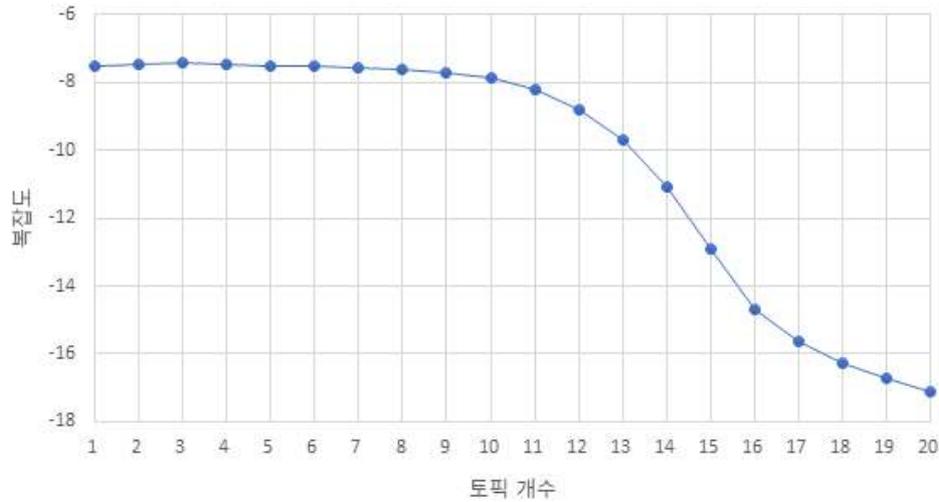
37) 변분 기댓값 최대화(variational expectation-maximization)

[표 3-5] 수집한 데이터의 음식 종류별 군집화

연번	음식 종류
1	한정식/백반/정통 한식, 탕/찌개/전골, 퓨전 한식, 기타 한식
2	정통 일식/일반 일식, 퓨전 일식, 기타 일식, 돈부리/일본 카레/벤토, 회/스시, 까스 요리
3	정통 중식/일반 중식, 퓨전 중식, 기타 중식, 딤섬/만두
4	이탈리안
5	프랑스 음식
6	고기 요리
7	닭/오리 요리
8	스테이크/바베큐, 퓨전 양식, 기타 양식
9	브런치/버거/샌드위치, 베이커리, 카페/디저트
10	국수/면 요리, 라멘/소바/우동
11	다국적 아시아 음식, 베트남 음식, 태국 음식, 인도 음식
12	다국적 퓨전, 세계음식 기타
13	남미 음식
14	해산물 요리, 시푸드 요리
15	철판 요리
16	패밀리 레스토랑
17	뷔페
18	칵테일/와인
19	전통 주점/포차, 일반 주점, 치킨/호프/팝, 이자카야/오텍/꼬치

본 연구에서는 복잡도 계산을 위해 Python의 gensim 라이브러리에서 제공하는 log\_perplexity<sup>38)</sup> 함수를 활용하였는데, 해당 함수는 식 (2-1)의 교차 엔트로피 근사값인  $H(W)$ 를 복잡도 값으로 반환한다. 이에 따라 복잡도를 계산한 결과는 [그림 3-1] 및 [표 3-6]과 같다. 해당 그림 및 표를 보면 토픽이 4개일 때부터 복잡도 값이 꾸준히 감소하고, 17개 이후부터는 비교적 완만한 감소세를 보이다가 20개에서 최솟값을 보이는 것을 알 수 있다. 복잡도는 그 값이 작을수록 모델의 성능이 우수한 것으로 평가하므로 해당 결과만 보면 토픽 수를 20개, 또는 그 이상으로 늘려야 할 것처럼 생각되지만 복잡도는 토픽 일관성 지수와 함께 상호 보완적으로 사용해야 효과적이다.

38) ldamodel of gensim, <https://radimrehurek.com/gensim/models/ldamodel.html> (접수일: 2020년 11월 24일)



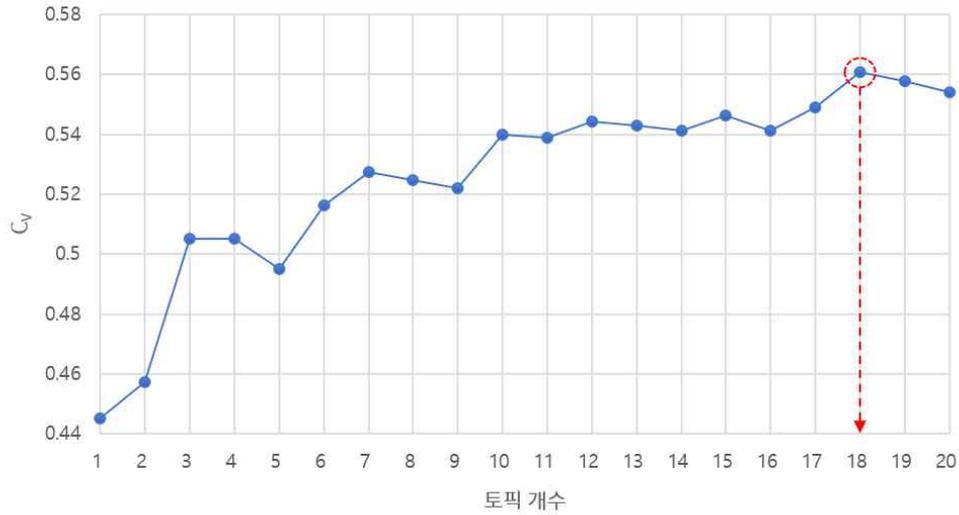
[그림 3-1] LDA 모델의 토픽 개수에 따른 복잡도 값 변화

[표 3-6] LDA 모델의 토픽 개수에 따른 복잡도 값

토픽 개수	복잡도	토픽 개수	복잡도
1	-7.51082234	11	-8.22898596
2	-7.45231351	12	-8.79538066
3	-7.44610872	13	-9.70340518
4	-7.46475918	14	-11.10584216
5	-7.50487654	15	-12.89634804
6	-7.54268687	16	-14.67093488
7	-7.57827489	17	-15.65183231
8	-7.63512062	18	-16.27496937
9	-7.71662180	19	-16.71264612
10	-7.89181484	20	-17.09508057

토픽 일관성 값을 계산한 결과는 [그림 3-2] 및 [표 3-7]과 같다. 토픽 개수가 증가함에 따라 토픽 일관성 값인  $C_v$ 가 증가와 감소를 반복하다가 토픽 수가 18개일 때 최댓값을 기록하고, 그 후로 다시 감소하는 추세를 보였다. 토픽 일관성 값이 크다는 것은 토픽 내 단어들의 연관성이 높다는 것을 의미하기에(Newman *et al.* 2010), 토픽 일관성 값이 가장 큰 토픽 개수를 최종 토픽의 수로 설정하는 것이 좋다. 앞서 살펴본 복잡도 값이 17

개 이후부터는 비교적 완만한 감소세를 보이므로 복잡도와 토픽 일관성 값을 종합적으로 판단해 토픽 개수를 18개로 설정하였다.



[그림 3-2] LDA 모델의 토픽 개수에 따른  $C_V$  값 변화

[표 3-7] LDA 모델의 토픽 개수에 따른  $C_V$  값

토픽 개수	$C_V$	토픽 개수	$C_V$
1	0.44515779	11	0.53876782
2	0.45712167	12	0.54430998
3	0.50527085	13	0.54286695
4	0.50527209	14	0.54107909
5	0.49512681	15	0.54644639
6	0.51638097	16	0.54125092
7	0.52753647	17	0.54898041
8	0.52474454	18	0.56064229
9	0.52192814	19	0.55769768
10	0.53984937	20	0.55393131

[표 3-8]은 LDA 모델링 결과로 추출된 18개의 토픽별 주요 키워드를 나타낸 것으로, 각 토픽의 명칭은 연구자가 직접 지정해야 한다. 예를 들어, 1번 토픽은 ‘가격\_NNG’, ‘괜찮\_VA(괜찮다)’, ‘비싸\_VA(비싸다)’ 등의 키워드를 포함하는데, 키워드들의 의미를 고려할 때 POI의 가격에 대한 속성과 관련되므로 토픽 명칭을 ‘가격’으로 부여하였다. 7번 토픽의 경우에는 ‘커피\_NNG’, ‘크림\_NNG’, ‘디저트\_NNG’ 등 카페 및 디저트와 관련된 키워드들을 다수 포함하므로 토픽 명칭을 ‘카페/디저트’로 설정하였다. 나머지 토픽들도 같은 방식으로 명칭을 부여하였다.

[표 3-8] LDA 기반 토픽 모델링 결과

구분	1	2	3	4	5	6
	가격	일시/혼잡도	맛집검색표현	긍정표현	중식	치킨/튀김류
1	가격_NNG	사람_NNG	맛집_NNG	맛있_VA	먹_VV	먹_VV
2	괜찮_VA	많_VA	찾_VV	먹_VV	맛있_VA	맛있_VA
3	비싸_VA	시간_NNG	근처_NNG	맛_NNG	양_NNG	맛_NNG
4	생각_NNG	자리_NNG	유명_XR	완전_NNG	시키_VV	소스_NNG
5	무난_XR	테이블_NNG	찾아가_VV	최고_NNG	많_VA	치킨_NNG
6	성비_NNG	기다리_VV	가게_NNG	굿_NNG	맵_VA	메뉴_NNG
7	저렴_XR	주문_NNG	동네_NNG	넘_NNG	볶음밥_NNP	세트_NNG
8	좋_VA	저녁_NNG	식당_NNG	맛나_VA	새우_NNG	우동_NNG
9	대비_NNG	얕_VV	요즘_NNG	제일_NNG	탕_NNG	돈까스_NNP
10	만족_NNG	가게_NNG	많_VA	좋아하_VV	짬뽕_NNG	카레_NNG
11	착하_VA	좁_VA	사람_NNG	존_NNP	탕수육_NNG	느끼_XR
12	적당_XR	평일_NNG	추천_NNG	인생_NNG	꼬치_NNG	좋아하_VV
13	수준_NNG	주말_NNG	가깝_VA	추천_NNG	소스_NNG	입맛_NNG
14	값_NNG	점심시간_NNG	골목_NNG	젤_NNG	매콤_NNP	괜찮_VA
15	싸_VA	혼자_NNG	최근_NNG	진짜_NNG	해물_NNG	튀기_VV
16	합리_NNG	인기_NNG	오래되_VV	맛나_NNP	요리_NNG	야채_NNG
17	장점_NNG	빠르_VA	인기_NNG	꿀맛_NNG	매운맛_NNP	매콤_NNP
18	할인_NNG	시끄럽_VA	단골_NNG	진리_NNG	푸짐_XR	갈릭_NNP
19	훌륭_XR	팀_NNG	발견_NNG	핵_NNG	볶음_NNG	감자튀김_NNP
20	만족도_NNG	회전_NNG	찾아오_VV	감동_NNG	짜장면_NNG	고로케_NNP
구분	7	8	9	10	11	12
	카페/디저트	브런치/베이커리	냉면/만두	멕시코칸	면류	이탈리안
1	커피_NNG	빵_NNG	먹_VV	타코_NNG	면_NNG	파스타_NNP
2	크림_NNG	치즈_NNG	맛있_VA	멕시코칸_NNP	국물_NNG	샐러드_NNG
3	디저트_NNG	피자_NNG	만두_NNG	퀘사디아_NNP	라멘_NNP	좋_VA
4	티_NNG	버터_NNG	좋아하_VV	부리또_NNP	맛_NNG	소스_NNG

5	달_VV	샌드위치_NNG	좋_VA	또띠아_NNP	칼끔_XR	스테이크_NNG
6	케이크_NNG	아보카도_NNP	냉면_NNG	멕시코_NNP	닭_NNG	음식_NNG
7	카페_NNG	건강_NNG	포장_NNG	친구_NNG	진하_VA	와인_NNG
8	라떼_NNP	촉촉_XR	생각나_VV	만족_NNG	쌀국수_NNP	요리_NNG
9	마시_VV	크림_NNG	입맛_NNG	비주얼_NNG	먹_VV	토마토_NNG
10	초코_NNP	느끼_XR	평양냉면_NNP	파히타_NNP	육수_NNP	버섯_NNG
11	딸기_NNG	부드럽_VA	추천_NNG	입맛_NNG	계란_NNG	리조또_NNP
12	아이스크림_NNG	스콘_NNP	유명_XR	넘_VV	칼국수_NNG	크림_NNG
13	음료_NNG	크림치즈_NNG	딤섬_NNP	특별_XR	짜_VV	만족_NNG
14	케익_NNG	식빵_NNG	군만두_NNG	극찬_NNG	시원_XR	레스토랑_NNG
15	밀크_NNG	팥_NNG	칼끔_XR	엔칠라다_NNP	추가_NNG	스프_NNG
16	티라미수_NNP	빵집_NNG	여름_NNG	여자친구_NNP	국수_NNG	메인_NNG
17	아메리카노_NNP	쫄깃_XR	새우_NNG	나초_NNP	쫄깃_XR	트러플_NNP
18	빙수_NNG	단호박_NNP	심심_XR	사위크림_NNP	담백_XR	감자_NNG
19	진하_VA	바게트_NNP	담백_XR	신기_XR	마늘_NNG	식전_NNG
20	우유_NNG	브런치_NNG	막국수_NNP	감탄_NNG	김치_NNG	느끼_XR
구분	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>
	<b>고기류</b>	<b>술집</b>	<b>서비스</b>	<b>분위기</b>	<b>일식</b>	<b>맛표현</b>
1	고기_NNG	좋_VA	친절_NNG	좋_VA	좋_VA	맛_NNG
2	밥_NNG	맥주_NNG	서비스_NNG	분위기_NNG	연어_NNG	느낌_NNG
3	나오_VV	추천_NNG	음식_NNG	인테리어_NNG	먹_VV	느끼_VV
4	먹_VV	분위기_NNG	직원_NNG	카페_NNG	초밥_NNG	향_NNG
5	굽_VV	다양_XR	사장_NNG	느낌_NNG	스시_NNP	강하_VA
6	양념_NNG	종류_NNG	기분_NNG	공간_NNG	새우_NNG	아쉽_VA
7	리필_NNG	마시_VV	손님_NNG	예쁘_VA	덮밥_NNG	부드럽_VA
8	반찬_NNG	메뉴_NNG	주문_NNG	넓_VA	신선_XR	재료_NNG
9	갈비_NNG	술_NNG	서빙_NNG	내부_NNG	재료_NNG	고소_XR
10	김치_NNG	안주_NNG	식당_NNG	조용_XR	만족_NNG	취향_NNG
11	곱창_NNP	맛있_VA	주방_NNG	위치_NNG	요리_NNG	특유_NNG
12	삼겹살_NNG	친구_NNG	상태_NNG	칼끔_XR	구성_NNG	맛보_VV
13	정식_NNG	팬찮_VA	위생_NNG	아늑_XR	생선_NNG	좋아하_VV
14	돼지_NNG	한잔_NNG	종업원_NNG	자리_NNG	참치_NNG	매력_NNG
15	소고기_NNG	즐기_VV	알바_NNP	매장_NNG	성비_NNG	단맛_NNG
16	기본_NNG	막걸리_NNG	메뉴판_NNP	편하_VA	홀룽_XR	스타일_NNG
17	짜_VV	카테일_NNG	알바생_NNP	음악_NNG	코스_NNG	풍미_NNG
18	돼지고기_NNG	가볍_VA	테이블_NNG	작_VA	롤_NNG	독특_XR
19	소금_NNG	기본_NNG	청결_NNG	사진_NNG	점심_NNG	깊_VA
20	칼끔_XR	소주_NNG	수준_NNG	따뜻_XR	구이_NNG	은은_XR

[표 3-8]을 보면, 같은 단어가 여러 토픽에 등장하는 것을 알 수 있다. 가령, ‘좋\_VA(좋다)’는 총 9개 토픽에 속하는 키워드이다([표 3-9] 참조). ‘좋\_VA’가 9개 토픽에 속한다는 것의 의미는 해당 키워드가 9개의 서로 다른 단어로 취급된다는 뜻이다. 이렇게 같은 키워드라도 서로 다른 토픽을 할당함으로써 얻는 이점은 함께 사용되는 키워드에 따라 달라지는 세부 의미를 파악할 수 있다는 것이다. 예를 들어 [표 3-9]에서 16번 토픽인 ‘분위기’에 속하는 ‘좋\_VA’와 1번 토픽인 ‘가격’에 속하는 ‘좋\_VA’는 사전적 의미는 유사하지만 세부 의미는 다를 수 있다. 즉, ‘분위기’ 토픽의 ‘좋다’는 ‘아늑한’, ‘쾌적한’ 등의 의미와 가까울 확률이 높고, ‘가격’ 토픽의 ‘좋다’는 ‘저렴한’, ‘싼’ 등의 의미와 근접할 확률이 높다. 이에 대한 사실은 3.4 절의 토픽별 키워드 기반 POI 임베딩 결과에서 확인할 수 있다.

[표 3-9] ‘좋\_VA(좋다)’의 토픽별 확률 분포

연번	토픽 번호	토픽	확률
1	14	술집	0.12631
2	16	분위기	0.10222
3	17	일식	0.03981
4	9	냉면/만두	0.02416
5	12	이탈리안	0.02244
6	1	가격	0.01872
7	13	고기류	0.00501
8	5	중식	0.00452
9	6	치킨/튀김류	0.00375

LDA 결과를 통해 리뷰별 토픽 분포([표 3-11]) 및 키워드별 토픽 분포([표 3-12])를 모두 추정할 수 있다. 예를 들어, [표 3-11]은 ‘덕화장’이라는 POI에 대해 작성된 리뷰에서 특정 품사 집합을 추출한 것([표 3-10] b)에 대한 토픽 분포로, 확률이 높은 순서로 정렬한 것이다. [표 3-11]을 보면, 해당 문서는 5번 토픽(중식)에 할당될 확률이 가장 높은 것을 알 수 있다. 이때 토픽별 확률값을 모두 더하면 1이 된다.

[표 3-10] '덕화장'에 대한 리뷰 텍스트 형태소 분석 결과 일부

POI 명칭 (주소)	음식 종류	a. 사용자 작성 리뷰 텍스트	b. a에 대한 형태소 분석 후 특정 품사만 추출한 결과
덕화장 (중구 장충동2가 12-3)	정통 중식/ 일반 중식	탕수육과 짜장면. 기본이 잘 갖춰져있는 맛집!	'탕수육_NNG', '짜장면_NNG', '기본_NNG', '갖추_VV', '맛집_NNG'

[표 3-11] 리뷰([표 3-10] b 전체)에 대한 토픽 분포

토픽 번호	토픽	확률	토픽 번호	토픽	확률
5	중식	0.088430	8	브런치/베이커리	0.051125
3	맛집검색표현	0.068689	17	일식	0.051081
14	술집	0.062377	1	가격	0.051044
16	분위기	0.057250	10	멕시칸	0.050508
13	고기류	0.056007	4	긍정표현	0.050507
18	맛표현	0.054546	7	카페/디저트	0.050507
11	면류	0.052910	9	냉면/만두	0.050507
6	치킨/튀김류	0.052111	2	일시/혼잡도	0.050507
15	서비스	0.051386	12	이탈리안	0.050507

[표 3-12]는 [표 3-10]의 b에 나열된 각 키워드에 대한 토픽 분포를 도출한 것이다. '탕수육\_NNG' 및 '짜장면\_NNG'이라는 키워드는 해당 문서와 마찬가지로 5번 토픽에 할당된 것을 알 수 있다. 또한, '맛집\_NNG'이라는 키워드는 3번 토픽에 할당되었다. 반면 '기본\_NNG' 및 '갖추\_VV'라는 단어는 '탕수육'이나 '짜장면'에 비해 다소 범용적 성격의 단어이기에 여러 토픽에 할당되었는데, 각각 13번 토픽(고기류) 및 14번 토픽(술집)에 할당될 확률이 가장 높았다. 따라서 각 키워드는 가장 높은 확률의 토픽으로 할당된다. 이렇듯 같은 키워드라도 리뷰의 문맥에 따라 다른 토픽에 할당될 수 있기에 문맥에 따라 달라지는 키워드의 의미를 반영할 수 있다. 각 리뷰별 키워드에 대해 가장 높은 확률값의 토픽을 할당해 이들을 [키워드, 토픽] 쌍으로 구성하며, 이것이 토픽별 키워드 기반 POI 임베딩을 위한 입력값으로 사용된다.

[표 3-12] 키워드별([표 3-10] b의 각 키워드) 토픽 분포

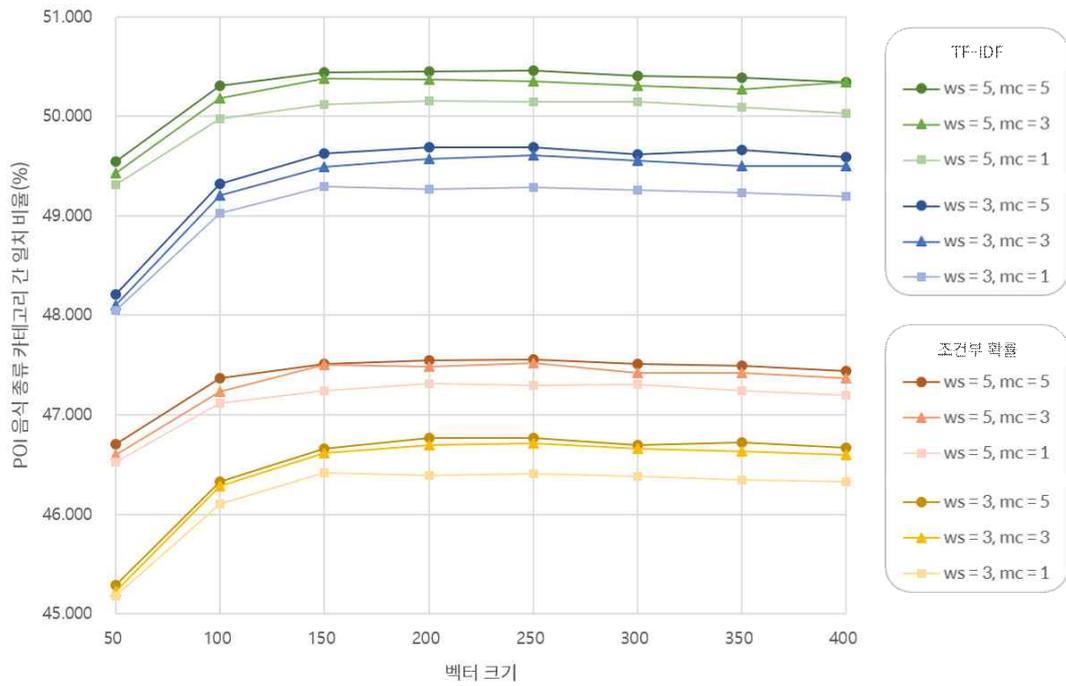
키워드	토픽 번호	토픽	확률
'탕수육_NNG'	5	중식	1.000
'짜장면_NNG'	5	중식	1.000
'기본_NNG'	13	고기류	0.302
	14	술집	0.277
	11	면류	0.132
	6	치킨/튀김류	0.088
	5	중식	0.086
	15	서비스	0.048
	8	브런치/베이커리	0.034
	17	일식	0.032
'갯추_VV'	14	술집	0.376
	16	분위기	0.371
	18	맛표현	0.222
	1	가격	0.030
'맛집_NNG'	3	맛집검색표현	1.000

### 3.4 토픽별 키워드 기반 POI 임베딩 결과

기존 단어 임베딩 기법인 skip-gram이 단어만을 입력값으로 사용했다면, 본 연구에서 활용하는 토픽별 키워드 기반 POI 임베딩에서는 [키워드, 토픽] 쌍이 입력값으로 활용된다. 실험을 위해 Liu *et al.*(2015b)이 Github<sup>39)</sup>을 통해 공개한 TWE-1 코드를 기본으로 활용하였다. 그러나 해당 코드는 Python 2를 기준으로 작성되었기에 일부 함수를 Python 3 버전으로 변환하는 과정을 거쳤으며, 하이퍼 파라미터도 본 연구에 맞게 수정하였다. 윈도우 크기(window size, ws), 키워드 최소 출현 횟수(minimum count, mc), 벡터 크기(vector size, vs), 그리고 TF-IDF 가중치 적용에 따른 POI 임베딩 모델의 성능 평가 결과는 [그림 3-3] 및 [표 3-13]과 같다. 전반적으로 TF-IDF 가중치를 사용한 조합의 성능이 더 우수한 것으로 나타났다. 특히 윈도우 크기 5, 단어 최소 출현 횟수 5, 벡터 크기 250일 때의 성능이 가장 좋았다. 참고로 이때의 벡터 크기는 키워드 벡터와 토픽 벡터가 병합되기 전의 크기이다. 키워드와 토픽을 기반으로 학습하는 임베딩 모델에서는 학습 결과로 얻어진 키워드 벡터와 토픽 벡터를 병합하는 과정을 통해 토픽별 키워드 벡터값을 산출하기 때문에 실제 POI 임베딩에 활용되는 벡터의 크기는 키워드 및 토픽 벡터 크기의 2배가 된다.

---

39) Github(TWE), [https://github.com/largelymfs/topical\\_word\\_embeddings](https://github.com/largelymfs/topical_word_embeddings) (접속일: 2020년 10월 5일)



[그림 3-3] 하이퍼 파라미터 조합에 따른 POI 임베딩 모델의 성능 평가

[표 3-13] 하이퍼 파라미터 조합에 따른  
POI의 음식 종류 카테고리 간 일치 비율

조건부 확률						
vs	ws = 3			ws = 5		
	mc = 1	mc = 3	mc = 5	mc = 1	mc = 3	mc = 5
	일치 비율(%)			일치 비율(%)		
50	45.182	45.224	45.291	46.526	46.599	46.703
100	46.103	46.284	46.331	47.114	47.235	47.365
150	46.417	46.615	46.662	47.240	47.500	47.510
200	46.393	46.692	46.768	47.314	47.486	47.545
250	46.412	46.714	46.765	47.298	47.519	47.553
300	46.378	46.659	46.697	47.303	47.417	47.512
350	46.350	46.633	46.725	47.239	47.424	47.490
400	46.332	46.597	46.667	47.193	47.364	47.441
TF-IDF						
vs	ws = 3			ws = 5		
	mc = 1	mc = 3	mc = 5	mc = 1	mc = 3	mc = 5
	일치 비율(%)			일치 비율(%)		
50	48.047	48.101	48.214	49.315	49.433	49.544
100	49.029	49.205	49.319	49.980	50.185	50.307
150	49.295	49.492	49.627	50.122	50.379	50.442
200	49.270	49.569	49.691	50.158	50.371	50.455
250	49.286	49.604	49.693	50.151	50.350	<b>50.463</b>
300	49.262	49.555	49.621	50.150	50.306	50.406
350	49.232	49.503	49.659	50.094	50.276	50.386
400	49.200	49.502	49.595	50.029	50.342	50.342

성능이 가장 우수하게 도출된 하이퍼 파라미터 조합인 TF-IDF 가중치, 윈도우 크기 5, 키워드 최소 출현 횟수 5, 벡터 크기 250을 채택해 토픽별 키워드 기반 POI 임베딩을 실시하였다. 임베딩에 활용한 데이터 현황은 [표 3-14]와 같다. 본 연구에서 최종 임베딩한 POI의 개수는 19,902개, 이때 활용한 리뷰 수는 208,560개이며, 여기에 포함된 중복되지 않은 키워드의 수는 17,694개이다. 그리고 POI 임베딩 모델에 포함된 토픽의 수는 18개이다. 따라서 가능한 [키워드, 토픽] 조합의 최대 개수는 318,492개이다. 실제로 추출된 [키워드, 토픽]은 42,307개인데, 하이퍼 파라미터 조합 시 키워드 최소 출현 횟수가 5회 이상인 것만 임베딩 모델에 포함됨에 따라 최종적으로는 24,682개의 [키워드, 토픽] 조합이 사용되었다. 이를 통해 키워드 1개당 평균 1.39개의 토픽이 부여되었다는 것을 알 수 있다.

[표 3-14] 토픽별 키워드 기반 POI 임베딩 관련 데이터 현황

구분	개수
임베딩된 POI 수	19,902
임베딩 모델에 활용된 리뷰 수	208,560
임베딩 모델에 포함된 키워드 수	17,694
임베딩 모델에 포함된 토픽 수	18
임베딩 모델에 포함된 [키워드, 토픽] 수	24,682

토픽별 키워드 임베딩 과정에서 얻어진 250차원의 키워드 벡터 중 ‘짜장면\_NNG’에 대한 벡터는 [표 3-15]와 같다. ‘짜장면\_NNG’이라는 키워드는 LDA 모델링 결과, 5번 토픽(중식)에 할당되었으며, 임베딩 결과 도출된 5번 토픽에 대한 250차원 벡터는 [표 3-16]과 같다.

[표 3-15] ‘짜장면\_NNG’ 벡터(크기: 250)

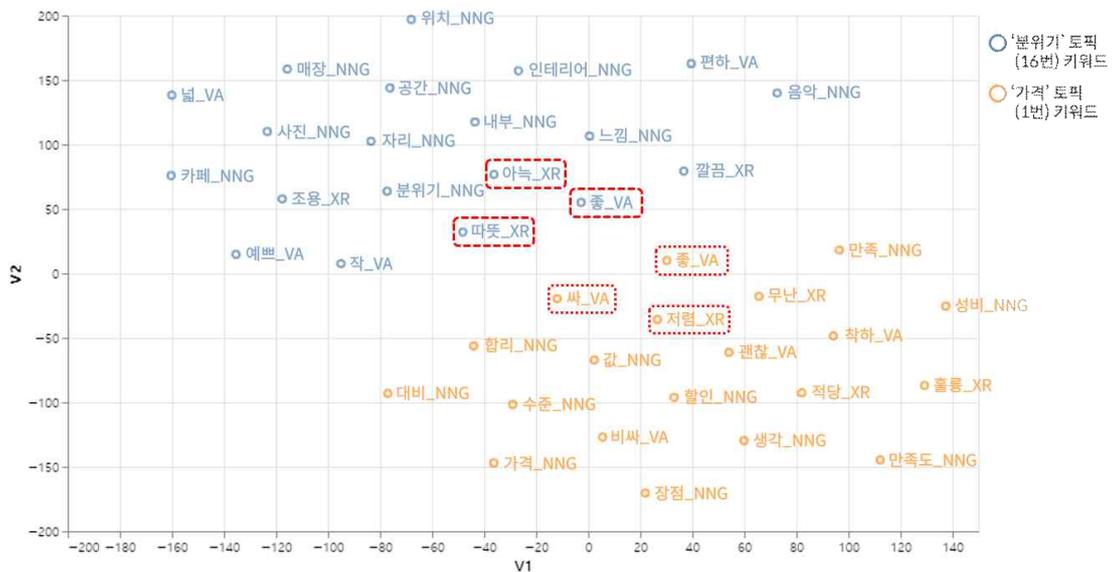
[	0.0170253,	-0.0254398,	0.1419052,	-0.0982857,	-0.698522 ,	-0.0078048,	0.0633295,
	-0.198412 ,	0.5955918,	-0.0184899,	-0.2152019,	-0.0936078,	0.1358424,	0.2119424,
	0.653968 ,	-0.3814785,	0.2262459,	0.1552938,	0.3249521,	-0.1699608,	0.2085667,
	0.2371351,	-0.2991742,	0.3765832,	0.2033286,	-0.1959593,	0.1890397,	-0.0280271,
	0.2153775,	-0.0083828,	0.1176957,	-0.1035556,	-0.4886742,	0.1933342,	-0.016565 ,
	-0.3403886,	0.197551 ,	-0.0909093,	-0.3763036,	-0.0242792,	-0.1564576,	-0.1747785,
	0.0184052,	0.3112377,	0.0650104,	0.3216728,	-0.1560524,	-0.2660529,	0.0750922,
	-0.1193108,	0.0896849,	-0.373093 ,	0.3258675,	-0.1359169,	-0.0372117,	-0.1877548,
	0.2059409,	0.1267562,	-0.0835416,	-0.0218622,	0.0793481,	0.3349297,	-0.0354749,
	0.1208575,	0.1130952,	-0.1275462,	0.3416471,	-0.1567855,	0.07336 ,	0.0309547,
	0.0796546,	-0.052143 ,	-0.2086324,	0.3505177,	0.0748341,	-0.2257348,	0.1124712,
	0.4969636,	-0.120893 ,	-0.1814059,	-0.4136057,	-0.189293 ,	0.1497228,	-0.0155968,
	0.2467246,	-0.2710654,	0.68285 ,	0.2096068,	-0.160049 ,	-0.1544522,	0.4865589,
	-0.2426554,	0.0476649,	0.3453662,	0.0480056,	-0.0718821,	-0.0640784,	0.6180654,
	-0.0594418,	-0.1248462,	-0.2444749,	0.2009857,	-0.1588776,	0.3239246,	0.2616117,
	-0.2342686,	-0.5414412,	0.0194561,	-0.2607719,	0.1180648,	0.1028125,	-0.2719174,
	0.1022465,	0.0811262,	0.0976619,	0.0518435,	0.0685956,	-0.1652216,	-0.1674308,
	0.2868887,	0.3077927,	-0.0898696,	0.2937886,	0.291154 ,	0.0263632,	0.263187 ,
	-0.175742 ,	0.543724 ,	0.4532242,	0.1592837,	-0.3718415,	0.1328225,	0.1277975,
	0.2371999,	-0.1713321,	0.2335608,	-0.2184573,	-0.5406097,	-0.0493544,	-0.3424273,
	-0.0960701,	0.1493559,	-0.3533568,	-0.030484 ,	-0.1543699,	0.1909248,	0.1263485,
	0.4703359,	0.0997994,	-0.1922624,	-0.0117427,	0.193514 ,	0.0465202,	-0.1410606,
	-0.0597234,	-0.4581314,	-0.3550029,	0.011046 ,	-0.1380565,	-0.2249445,	0.1569611,
	-0.0968728,	0.5103338,	0.3672104,	-0.1714304,	0.0196228,	-0.162085 ,	-0.0231919,
	0.0079426,	0.0526601,	0.3237966,	-0.2501471,	0.1714443,	0.5967727,	0.086485 ,
	0.3080105,	0.0596407,	-0.0079753,	-0.0358385,	-0.4146006,	-0.5911802,	-0.4343143,
	-0.2440174,	0.3872158,	-0.4294566,	-0.1892299,	-0.1521237,	0.0285368,	-0.0258503,
	0.045146 ,	0.1051972,	0.380597 ,	0.2535301,	-0.2696115,	-0.3788495,	-0.12929 ,
	0.1156777,	-0.0600108,	-0.1733346,	0.1228961,	0.3058206,	0.1160787,	-0.0774535,
	-0.365041 ,	0.0708411,	-0.2047675,	0.1955636,	0.1199581,	0.235995 ,	-0.2140147,
	0.0914373,	0.1772131,	-0.1426819,	0.1129255,	-0.1729424,	-0.2387033,	-0.0195465,
	-0.2079328,	0.4362938,	-0.3664069,	0.1308732,	-0.0707472,	0.0661139,	0.074537 ,
	0.39895 ,	0.0602111,	0.0590226,	-0.3660574,	0.2789994,	-0.1035192,	0.0899655,
	0.0828362,	0.2788256,	0.2219297,	0.0478672,	-0.286343 ,	-0.0467547,	0.2525316,
	-0.0652871,	-0.0699383,	-0.0117729,	-0.3806189,	-0.1229989,	-0.2066963,	0.2144102,
	-0.3684476,	-0.136662 ,	0.2168835,	0.3157521,	-0.0954386]		

[표 3-16] 5번 토픽(중식) 벡터(크기: 250)

0.0370205,	0.2438735,	0.1301384,	0.4198304,	-0.1912916,	0.0204597,	-0.2794322,
-0.0377619,	0.0811812,	0.1217049,	-0.0465027,	-0.0757644,	-0.1778168,	0.0710022,
-0.1097495,	0.2564043,	0.136179 ,	0.0686006,	0.221245 ,	-0.1806599,	-0.0297639,
0.2110134,	-0.1697943,	0.1943479,	0.195013 ,	-0.1750703,	-0.0189742,	0.0575283,
0.2401972,	-0.1195355,	0.1786139,	0.1279504,	0.1349487,	-0.3865013,	-0.1095522,
-0.1021257,	0.1336095,	0.0673992,	0.1088155,	0.320256 ,	-0.0905823,	-0.0679222,
0.06105 ,	0.1775795,	0.0238746,	0.095928 ,	-0.1361954,	-0.1992842,	-0.0468428,
-0.1825974,	-0.1091451,	-0.0730924,	0.1596209,	0.1479843,	0.3242753,	-0.0126911,
-0.0072974,	0.1356857,	-0.1955998,	-0.0106007,	-0.0413309,	0.2323852,	-0.1459282,
-0.2097691,	-0.1525728,	-0.2272083,	0.0428666,	-0.0506731,	0.0409306,	0.0625943,
0.172111 ,	0.0030043,	-0.3326395,	0.2145499,	0.0132587,	-0.2330091,	-0.0496127,
0.3603776,	-0.1174492,	0.1114232,	-0.09013 ,	-0.1377434,	0.0431847,	0.2144186,
0.1968016,	-0.1052691,	0.1405541,	0.2359425,	-0.0979577,	-0.0563649,	0.1087155,
-0.1010506,	-0.1259724,	-0.0468459,	-0.035645 ,	0.0052352,	-0.014397 ,	0.2652397,
0.0726122,	-0.3495345,	-0.0922279,	0.1205719,	-0.2452916,	0.1229193,	-0.0615669,
-0.1481546,	-0.1701072,	0.0214517,	-0.0599589,	0.0398337,	-0.1265439,	-0.2171273,
0.1606575,	-0.1733725,	0.212422 ,	0.1216831,	-0.0216461,	-0.0144905,	0.065963 ,
0.0293497,	0.0583368,	-0.0043675,	0.1705829,	-0.0138717,	0.0160594,	0.1213995,
-0.0201673,	0.1907009,	0.1809219,	0.1798242,	0.0633991,	0.1926427,	0.0695816,
-0.2268948,	-0.140171 ,	-0.0722916,	-0.1139451,	-0.1907747,	-0.1384683,	0.018975 ,
-0.2814059,	-0.0749107,	-0.2082985,	-0.0784341,	0.2728384,	0.080621 ,	-0.0120578,
-0.0091177,	-0.1069298,	-0.019602 ,	0.1161626,	-0.1560311,	-0.1083882,	0.1578324,
-0.2304701,	-0.0943774,	-0.0794726,	-0.1534444,	0.0867224,	-0.1591453,	0.0896087,
0.2321295,	0.2967121,	0.2931249,	-0.0504436,	0.0381436,	-0.1418208,	-0.1781672,
0.0389224,	-0.0201586,	0.153526 ,	-0.0062647,	0.1825331,	-0.0058554,	0.0604255,
-0.1966954,	-0.230741 ,	0.1462709,	-0.1730083,	0.0921085,	-0.041228 ,	-0.1250108,
-0.1650248,	-0.079747 ,	-0.0084463,	-0.034001 ,	-0.080888 ,	0.2642827,	0.1058261,
0.183737 ,	0.3579933,	0.3321008,	0.2178186,	-0.0671995,	-0.0336318,	-0.0716154,
0.1500349,	-0.1100409,	0.0493504,	0.2569656,	0.2032056,	-0.2566149,	-0.281578 ,
0.2354209,	-0.0787537,	-0.0418239,	-0.0108725,	-0.108184 ,	-0.0336721,	0.0302094,
-0.2976362,	0.1355501,	-0.0336775,	-0.009966 ,	-0.2049803,	-0.2752812,	-0.1476043,
0.0957735,	0.3750961,	-0.044768 ,	-0.2102735,	-0.0329046,	0.0545591,	-0.1454266,
-0.0552514,	0.0936616,	0.04818 ,	-0.2078523,	-0.2520476,	0.0011221,	0.0147769,
0.1856579,	0.1057245,	0.2816799,	-0.0066851,	0.00006 ,	0.0157384,	0.265763 ,
0.2677092,	-0.0435756,	-0.1616571,	-0.3713096,	0.0390967,	0.1904259,	-0.0119554,
-0.4860896,	0.0573044,	0.1295409,	0.0905282,	-0.0111437]		

결과적으로 5번 토픽에 할당된 키워드 ‘짜장면\_NNG’의 최종 임베딩 결과는 키워드 벡터인 [표 3-15]와 토픽 벡터인 [표 3-16]을 키워드, 토픽 순으로 병합함으로써 얻어지므로 500차원의 벡터가 된다. 이때 이렇게 나열된 숫자들만 보서는 각 벡터가 어떤 의미인지 알기 어렵기 때문에 t-SNE 알고리즘을 활용해 2차원 공간에 시각화하는 과정이 필요하다.

앞서 2.2절 및 3.3절에서 ‘좋\_VA(좋다)’라는 키워드가 9개 토픽에 속하며, 토픽에 따라 다른 의미로 사용된다고 언급한 바 있다. 가령, ‘분위기’ 토픽에 속하는 ‘좋\_VA’는 ‘아늑한’, ‘쾌적한’ 등의 의미와 가까울 확률이 높고, ‘가격’ 토픽에 속하는 ‘좋\_VA’는 ‘저렴한’, ‘싼’ 등의 의미와 근접할 확률이 높다. 이에 대한 사실 여부를 t-SNE 알고리즘 기반 시각화를 통해 확인해 보았다. 즉, ‘분위기’ 및 ‘가격’ 토픽에 속하는 키워드 벡터들의 차원을 감소시켜 2차원의 그래프상에 나타내었다. [그림 3-4]는 각 토픽의 대표 키워드 20개씩을 t-SNE를 이용해 기존 500차원에서 2차원으로 차원을 감소시킨 것이다. 이 그림에서 ‘분위기’ 토픽에 속하는 키워드 ‘좋\_VA’는 해당 토픽과 관련된 키워드인 ‘아늑\_XR’, ‘따뜻\_XR’ 등과 가까이 위치하고, ‘가격’ 토픽에 속하는 키워드 ‘좋\_VA’는 해당 토픽과 관련된 ‘저렴\_XR’, ‘싸\_VA(싸다)’ 등의 키워드와 가까이 위치한 것을 확인할 수 있다.

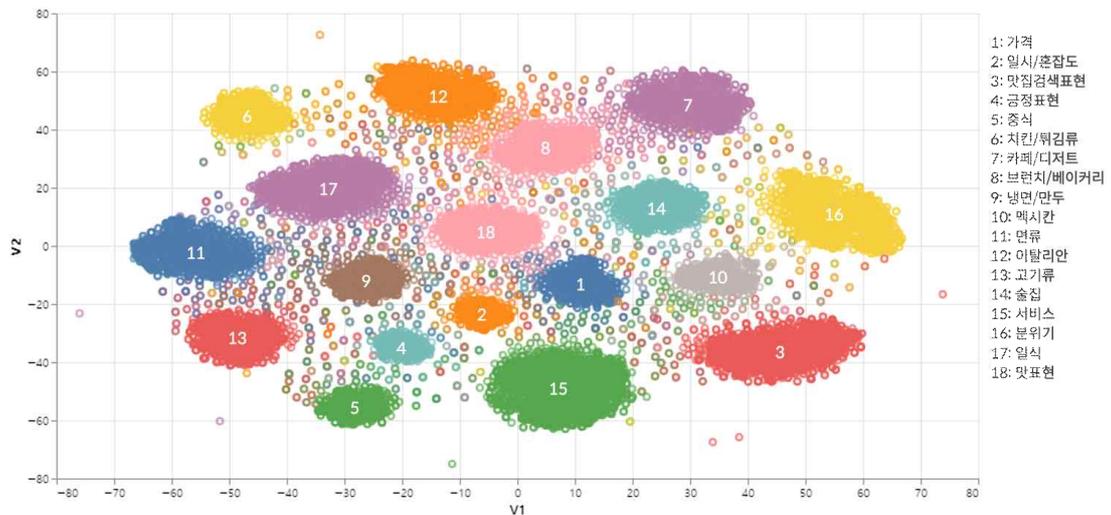


[그림 3-4] ‘분위기’ 및 ‘가격’ 토픽에 속하는 상위 20개 키워드 벡터의 차원 감소 결과

같은 키워드라도 토픽에 따라 서로 다른 키워드로 취급함으로써 토픽과 문맥에 따라 달라지는 키워드의 의미를 보다 세부적으로 반영할 수 있으

며, 이에 따라 POI의 시맨틱 정보를 기반으로 한 검색이 가능하다. 예를 들어, 사용자가 검색어로 ‘분위기 아늑한 카페’를 입력하고, ‘아늑한’에 높은 가중치를 부여하였을 때 해당 키워드와 관련된 POI뿐 아니라 이와 유사한 키워드인 ‘따뜻한’, ‘포근한’ 등과 관련된 POI들도 검색 결과로 반환해 줄 수 있다. 이러한 특성은 사용자가 입력한 검색어와 정확히 일치하는 키워드를 포함하는 POI가 없어도 유사 키워드를 통해 검색이 이루어지기 때문에 사용자에게 더욱 풍부한 검색 결과를 제공할 수 있다는 장점이 있다.

[그림 3-5]는 토픽별 키워드 벡터 24,682개를 2차원 공간에 시각화한 것이다. 해당 그림을 보면 18개 토픽별로 벡터들이 군집화된 것을 확인할 수 있다. 군집된 벡터들 사이에 분산되어 있는 벡터들은 2개 이상의 토픽에 할당된 키워드들로, 여러 토픽에 속하기 때문에 토픽별로 군집된 벡터들 사이에 분산되어 위치하는 것이다.



[그림 3-5] 토픽별 키워드 벡터 차원 축소 결과

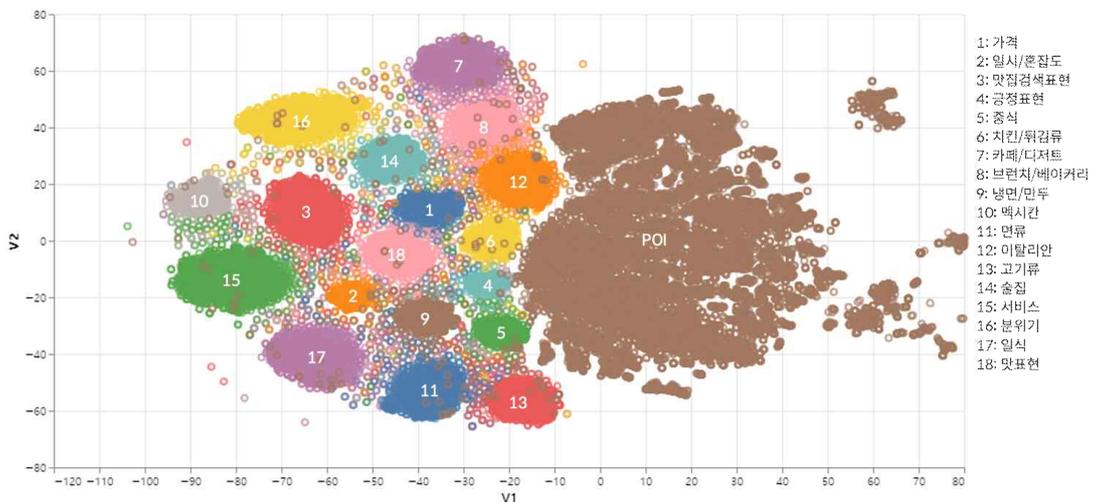
[키워드, 토픽] 벡터를 입력값으로 하여 토픽별 키워드 기반 POI 임베딩을 실시하였다. 즉, POI별로 [키워드, 토픽] 벡터에 TF-IDF 가중치를 고려한 합을 계산해(식 (2-9) 참조) 각 POI의 벡터를 확정하는 것이다. 결과적으로 총 19,902개의 POI를 벡터로 임베딩하였으며, 일례로 서울 서대문구 연희동에 위치한 장소인 '03InNOut'의 임베딩 결과는 [표 3-17]과 같다.

[표 3-17] 서울 서대문구 연희동에 위치한 POI '03InNOut' 벡터(크기: 500)

[-0.02274,	-0.0787,	-0.12199,	0.25158,	-0.24512,	-0.02912,	-0.14633,	-0.21023,	0.35329,
0.12935,	0.15367,	-0.10039,	0.10241,	-0.00342,	0.0364,	0.10317,	-0.09808,	-0.11068,
0.09268,	-0.05662,	0.12847,	0.09991,	-0.06451,	0.0251,	-0.04911,	-0.01005,	-0.18056,
-0.04336,	0.22672,	-0.03378,	0.26277,	0.2219,	0.0527,	-0.04317,	0.20179,	-0.068,
-0.11243,	-0.02011,	-0.06405,	0.15916,	0.03987,	-0.07262,	-0.02103,	0.04829,	0.09613,
0.12826,	-0.15749,	-0.14859,	-0.19326,	-0.17168,	-0.17826,	-0.09407,	0.08857,	0.13962,
0.22926,	-0.13022,	-0.13796,	0.09474,	-0.02467,	0.10542,	-0.15127,	0.06376,	-0.18798,
-0.13847,	-0.02177,	-0.14377,	-0.03685,	-0.17801,	-0.09649,	-0.08939,	0.06057,	0.0877,
0.17073,	0.24422,	0.11836,	-0.00148,	-0.20702,	0.25883,	0.04675,	-0.11031,	-0.20801,
-0.18425,	-0.06474,	-0.08066,	0.32543,	0.01832,	0.14533,	0.16443,	-0.16177,	-0.10283,
0.11692,	0.03825,	-0.16951,	0.09227,	-0.05067,	0.26398,	0.09838,	0.03672,	0.28562,
-0.29788,	0.00025,	0.0232,	-0.09965,	0.09299,	-0.04945,	-0.16429,	-0.06751,	-0.09905,
0.15076,	0.13532,	0.19981,	-0.28969,	-0.18685,	-0.03001,	0.0932,	0.11125,	0.04243,
0.01256,	-0.02728,	0.12371,	0.0473,	-0.17583,	0.13794,	0.01744,	0.01815,	0.11401,
-0.05124,	0.33532,	0.15254,	0.00714,	0.0656,	0.01323,	-0.07142,	-0.14014,	0.00556,
0.09966,	0.06275,	0.05154,	-0.07505,	0.05162,	0.10314,	-0.16266,	0.04809,	-0.16348,
-0.09035,	0.03674,	0.07923,	0.0804,	-0.07673,	0.04091,	-0.00573,	0.11354,	-0.2412,
0.09101,	-0.23761,	0.04441,	0.08014,	0.09407,	-0.12757,	0.2554,	-0.02662,	0.2058,
-0.06982,	0.22065,	-0.10145,	0.1605,	0.18152,	0.10763,	0.1879,	0.04849,	0.05115,
0.04916,	0.1728,	0.1228,	0.0532,	0.0447,	-0.03025,	0.22218,	-0.04644,	-0.00603,
-0.2826,	-0.11937,	-0.05994,	0.07805,	-0.15709,	-0.17281,	0.00861,	0.03434,	0.00152,
0.0656,	-0.08143,	0.15301,	0.08039,	-0.20243,	0.0683,	-0.11909,	0.26778,	-0.07632,
0.07324,	-0.08526,	0.12163,	-0.01184,	-0.08814,	0.07459,	0.08883,	-0.02668,	0.24643,
-0.10896,	-0.10054,	0.00313,	-0.18181,	0.09842,	-0.04508,	-0.1824,	-0.09655,	0.02747,
0.04532,	0.17169,	0.09687,	-0.21707,	-0.10667,	-0.02382,	0.10147,	-0.15632,	-0.08395,
-0.03212,	-0.04415,	-0.1132,	-0.05506,	-0.16707,	0.02809,	0.00112,	0.23923,	0.23814,
-0.1089,	0.20161,	0.08644,	-0.16192,	0.2594,	0.01741,	-0.03372,	-0.05556,	0.04846,
-0.09094,	0.1996,	-0.16399,	0.0818,	0.09645,	0.14259,	-0.00182,	-0.04633,	0.04908,
0.02219,	0.20347,	-0.2011,	-0.05273,	-0.23236,	-0.13427,	0.28464,	0.08369,	0.07691,
-0.0121,	0.03032,	0.12072,	-0.00219,	0.1395,	-0.07378,	0.06363,	0.17326,	-0.11503,
0.10942,	0.02335,	-0.05779,	0.08525,	0.06519,	-0.00007,	-0.11356,	0.12684,	0.0949,
-0.08392,	0.36151,	0.30216,	0.1663,	-0.04341,	0.00009,	-0.00145,	0.11023,	-0.11486,
-0.09889,	0.12415,	-0.06652,	-0.14954,	0.09644,	0.01708,	0.07198,	0.12147,	-0.11109,
-0.1211,	-0.10306,	-0.10167,	-0.02835,	-0.1211,	0.05957,	-0.05645,	0.21575,	-0.11751,
0.05501,	0.05437,	0.07403,	0.05011,	-0.10854,	0.13964,	-0.02738,	-0.21946,	-0.01863,
-0.17076,	-0.11172,	-0.00145,	-0.10875,	-0.02845,	0.05401,	-0.05536,	0.2232,	0.39593,
0.02688,	0.09106,	-0.0115,	0.32491,	0.00035,	0.03446,	-0.14122,	-0.18907,	-0.02761,
0.06514,	0.39504,	0.16981,	0.13471,	0.1776,	0.00495,	-0.16633,	0.17709,	0.01481,
-0.1773,	0.04704,	-0.00691,	0.16824,	-0.12986,	0.11734,	0.34911,	-0.38952,	0.01651,
0.01262,	-0.12984,	-0.06121,	-0.0285,	-0.15333,	-0.16359,	-0.07746,	0.08384,	0.08038,
0.22408,	-0.41844,	-0.09501,	-0.00802,	0.23858,	0.19442,	0.08286,	0.15255,	0.04022,
0.05272,	-0.06653,	-0.1533,	0.08738,	-0.08669,	-0.02368,	0.16642,	-0.12305,	0.23144,
0.08951,	-0.03427,	-0.01706,	-0.04083,	-0.03194,	-0.07007,	0.0538,	0.01902,	0.01841,
-0.07006,	-0.0726,	0.06475,	0.14606,	-0.02492,	0.0561,	-0.15803,	0.1048,	-0.15576,
0.08781,	0.02748,	0.00328,	0.03914,	0.16651,	0.03521,	-0.06602,	0.10113,	-0.157,
0.04914,	-0.01209,	-0.04292,	0.07509,	0.16055,	-0.04553,	0.04136,	-0.08642,	0.17807,

-0.10962,	0.01201,	0.00596,	0.01311,	0.09217,	0.0862,	0.0018,	0.06386,	0.00141,
0.22653,	0.01118,	-0.08638,	-0.1237,	0.10668,	0.0248,	0.09818,	-0.07369,	-0.19805,
-0.0776,	0.02752,	-0.10166,	-0.09116,	0.04462,	0.15984,	-0.09462,	0.22146,	-0.10244,
0.09437,	0.02145,	-0.10306,	-0.04536,	0.00273,	0.10765,	-0.1261,	0.15943,	0.07839,
0.16625,	-0.06883,	-0.22554,	0.15407,	0.11261,	-0.06671,	0.11986,	-0.03375,	0.01748,
0.01206,	-0.19326,	-0.0305,	0.01433,	-0.21984,	-0.26892,	0.0265,	0.01324,	0.22555,
0.29474,	-0.30701,	0.10704,	-0.04809,	-0.08948,	-0.0521,	-0.07453,	0.16223,	0.03627,
-0.0266,	-0.0755,	-0.08775,	0.01756,	-0.10638,	0.06874,	0.24377,	-0.08989,	-0.01026,
-0.00622,	0.06823,	0.28832,	-0.13931,	-0.1458,	-0.05075,	0.11364,	0.03245,	0.00254,
-0.25034,	0.00715,	0.19394,	0.05526,	0.00619]				

[그림 3-6]은 토픽별 키워드 벡터와 POI 벡터를 함께 시각화한 것으로, 좌측에는 [그림 3-5]와 유사하게 토픽별 키워드 벡터들이 토픽별로 군집화되어 있으며, 우측에는 POI 벡터들이 군집화된 것을 확인할 수 있다.



[그림 3-6] 토픽별 키워드 벡터와 POI 벡터 차원 축소 결과

앞서 사례로 설명하였던 ‘분위기’ 및 ‘가격’ 토픽에 모두 속하는 키워드인 ‘좋\_VA’와 가장 관련성 높은 상위 10개 POI를 토픽별로 각각 추출한 결과를 비교해 보았다. 먼저 [표 3-18]을 보면, ‘분위기’ 토픽에 속하는 ‘좋\_VA’ 벡터와 가장 관련성 높은 상위 10개 POI들의 음식 종류는 주로 ‘카페/디저트’가 많은 것을 알 수 있다. 반면 [표 3-19]를 보면, ‘가격’ 토픽의 ‘좋\_VA’와 관련성 높은 POI는 ‘닭/오리 요리’, ‘고기 요리’ 등 다양한 음식 종류에 속하는 POI들이 도출되었다. 이는 [표 3-18]의 상위 10개 POI들은 각 POI에 대해 작성된 리뷰에 ‘분위기’가 ‘좋다’와 유사한 단어들이 다른 POI들에 비해 중요하게 사용되었고, [표 3-19]의 상위 POI에서는 ‘가격’이 ‘좋다’와 유사한 단어들이 중요하게 사용되었다는 것을 의미한다.

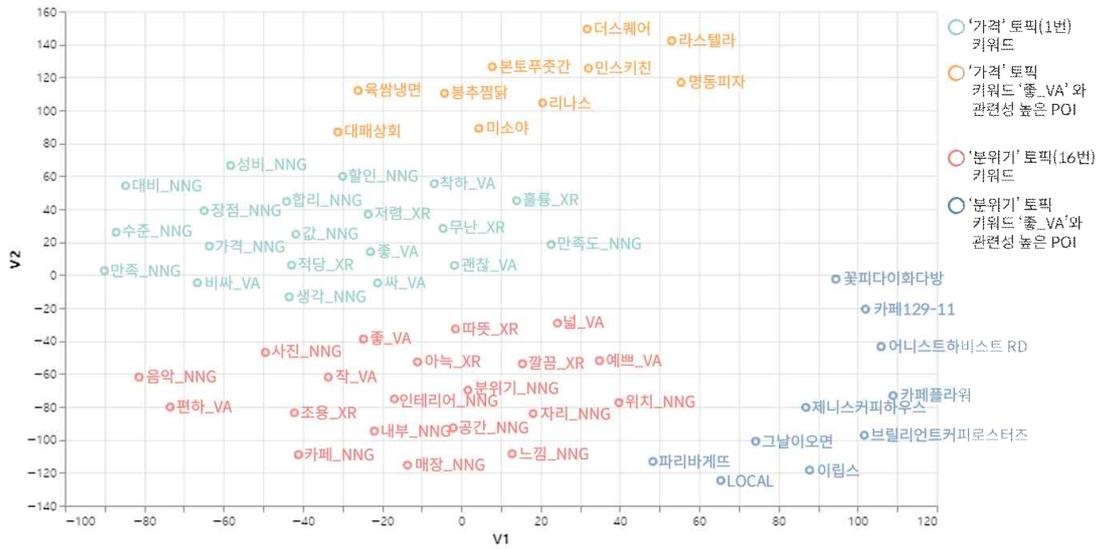
[표 3-18] ‘분위기’ 토픽의 키워드 ‘좋다’와 관련성 높은 상위 10개 POI

순위	관련도	POI 명칭	주소	음식 종류
1	0.95056	그날이오면	서대문구 창천동 53-24	칵테일/와인
2	0.90465	LOCAL	성북구 성북동 88	이탈리안
3	0.88661	카페플라워	중구 남산동2가 49-37	카페/디저트
4	0.88195	어니스트하비스트 RD	서초구 반포동 68-1	카페/디저트
5	0.87447	브릴리언트 커피로스터즈	서초구 잠원동 34-16	카페/디저트
6	0.87443	파리바게뜨	용산구 이촌동 300-19	베이커리
7	0.87032	카페129-11	서대문구 연희동 129-11	카페/디저트
8	0.86509	이립스	성북구 동선동2가 43	카페/디저트
9	0.86425	꽃피다이화다방	서대문구 대현동 37-1	카페/디저트
10	0.86334	제니스커피하우스	서대문구 연희동 128-5	카페/디저트

[표 3-19] ‘가격’ 토픽의 키워드 ‘좋다’와 관련성 높은 상위 10개 POI

순위	관련도	POI 명칭	주소	음식 종류
1	0.79090	봉추찜닭	강남구 신사동 541-14	닭/오리 요리
2	0.77174	본토푸줏간	강남구 신사동 661-18	고기 요리
3	0.75797	대패상회	은평구 갈현동 455-17	고기 요리
4	0.75556	미소야	종랑구 신내동 646	라멘/소바/우동
5	0.74793	명동피자	송파구 문정동 634	이탈리안
6	0.74213	육쌈냉면	도봉구 창동 74-7	국수/면 요리
7	0.74086	더스퀘어	강남구 역삼동 603	뷔페
8	0.73904	리나스	강남구 대치동 892	브런치/버거/샌드위치
9	0.73795	민스키친	강남구 신사동 651-4	한정식/백반/정통 한식
10	0.73715	라스텔라	마포구 도화동 169-1	뷔페

[표 3-18] 및 [표 3-19]의 POI들과 ‘분위기’ 및 ‘가격’ 토픽에 속하는 상위 20개 토픽별 키워드들을 한 벡터 공간에 시각화한 결과는 [그림 3-7]과 같다. 해당 그림을 보면, ‘분위기’ 토픽의 ‘좋\_VA’와 관련성 높은 POI들과 ‘분위기’ 토픽에 속하는 키워드들이 서로 근접해 위치하고, ‘가격’ 토픽의 ‘좋\_VA’와 관련성 높은 POI들과 ‘가격’ 토픽의 키워드들이 가까이 위치한 것을 알 수 있다.



[그림 3-7] 분위기 및 가격 토픽별 키워드 및 관련 POI 벡터 차원 감소 결과

## 3.5 시맨틱 POI 검색에의 활용 결과

### 3.5.1 결과 분석 개요

본 연구에서 개발한 시맨틱 POI 검색 기술 중 자연어 및 선택 기반 POI 검색에의 활용 결과 분석에서는 먼저, 검색 키워드와 결과가 합리적으로 매칭되는지와 가중치가 잘 작동하는지를 종합적으로 살펴보았다. 또한, 사용자 설문 조사를 이용해 검색 결과에 대한 품질을 확인하고, 최종적으로 기존 유사 서비스들과의 검색 결과 비교를 통해 본 연구 결과의 차별성을 확인하고자 하였다. 그리고 유사 POI 검색에의 활용 결과 분석에서는 먼저, 검색 기준이 되는 POI와 결과로 도출된 POI 간 유사성을 확인하고, 사용자 설문을 이용해 검색 결과에 대한 품질을 평가한 후, 기존 서비스들과의 결과 비교를 통해 본 연구 결과의 차별성을 확인하고자 하였다.

### 3.5.2 자연어 기반 POI 검색에의 활용 결과

#### 3.5.2.1 실험 결과 분석

자연어 기반 검색어는 크라우드 소싱(crowd sourcing)으로 지도를 큐레이션하는 플랫폼인 ‘진짜 서울<sup>40)</sup>’의 ‘테마지도’ 내용을 참고하였다. 이 플랫폼은 서울에 사는 사람들이 각자 자신의 취향을 반영한 장소를 공유하는 컨셉을 가지고 있으며, ‘식물 덕후가 좋아할 만한 곳’, ‘1인분 8천 원 이하 가성비 밥집’, ‘비가 내리는 날 올드팝이 듣고 싶어’ 등 다양하고 흥미로운 장소 관련 테마가 등록되어 있다. 그중 본 연구에서 검색어로 활용한 것은 ‘인생 피자’, ‘식물 덕후가 좋아할 만한 곳’, ‘스트레스 풀리는 매운맛’, ‘뷰 좋은 카페’이다.

자연어 기반 POI 검색에서는 사용자가 입력한 검색어에 대한 형태소 분

---

40) 진짜 서울, <https://jinjja-seoul.com/> (접속일: 2020년 10월 19일)

석이 우선 수행되며, 분석된 형태소가 임베딩 데이터베이스에 포함된 경우, 이를 다시 사용자에게 반환해 가중치를 입력하도록 한다. [표 3-20]은 ‘인생 피자’로 검색했을 때의 형태소 분석 결과와 이에 대한 가중치를 부여한 것이다. 해당 검색어의 형태소는 ‘인생\_NNG’와 ‘피자\_NNG’로 분리되며, 각각 ‘긍정표현’ 토픽과 ‘브런치/베이커리’ 토픽에 해당될 확률이 가장 높았다. 각 키워드에 0.5의 가중치를 부여해 관련도가 높은 상위 10개 POI를 출력한 결과는 [표 3-21]과 같다. 관련도는 토픽별 키워드들의 가중 합을 통해 도출된 벡터와 POI 벡터 간 코사인 유사도를 통해 계산된다. 상위 10개로 도출된 POI에 대해 작성된 리뷰를 살펴본 결과, ‘인생 피자’라는 직접적 표현이 언급된 리뷰가 다수 있었다. 해당 표현이 없음에도 상위로 도출된 POI의 경우, ‘인생’ 및 ‘피자’와 유사한 키워드들이 리뷰에 포함되어 있기 때문이다([표 3-22] 및 [표 3-23] 참조). [표 3-21]에서 2위로 도출된 ‘푸오꼬’라는 POI의 대표 리뷰를 보면, ‘인생\_NNG’라는 키워드는 없지만, ‘맛있어요’, ‘맛있는거’와 같이 형태소 분석을 통해 ‘맛있\_VA’로 추출되는 키워드가 포함되어 있다. 그리고 [표 3-22]를 보면, ‘인생\_NNG’와 유사한 단어 8위에 ‘맛있\_VA’가 도출된 것을 볼 수 있다. 즉, 시맨틱 POI를 기반으로 한 검색에서는 사용자가 입력한 검색 키워드와 정확히 일치하는 키워드가 없어도 다른 유사 키워드들을 통해 POI 검색이 이루어지므로 사용자에게 더욱 풍부한 검색 결과를 제공할 수 있다는 장점이 있다.

[표 3-20] ‘인생 피자’ 형태소 분석 결과 및 가중치 할당

검색어	키워드			
	1	가중치	2	가중치
인생 피자	인생_NNG	0.5	피자_NNG	0.5
	토픽: 긍정표현		토픽: 브런치/베이커리	

[표 3-21] ‘인생(w: 0.5)’, ‘피자(w: 0.5)’ 검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.85037	피자 보이시나	동작구 흑석동 190-5	이탈리안	1학년때중앙대 처음와서먹은 인생중피자중에 가장 맛있었다.
2	0.84524	푸오꼬	양천구 목동 917-9	이탈리안	완전 맛있어요! 피자맛이 환상적~^^ 푸오꼬 샐러드도 너무 좋구~ 맛있는거 먹고 싶을때마다 찾아가게 됩니다~
3	0.84482	디마떼오	종로구 동숭동 1-141	이탈리안	음식사진을 못찍어서 아쉬운데... 전반적으로 다 맛있었다ㅠ
4	0.84426	대장장이 화덕피자집	종로구 가회동 202-1	이탈리안	유일한 인생피자집입니다. 여기보다 더 맛있는 피자집은 아직 못 봤어요.
5	0.84420	카니발피자	서초구 방배동 421-1	이탈리안	위에 감자튀김 올라간 피자 맛있어용ㅎㅎ...
6	0.84400	스파카 나폴리	서대문구 창천동 57-13	이탈리안	피자는 도우가매우 쫄깃, 파스타는 그럭저럭 괜찮은 수준
7	0.84326	핏제리아오	종로구 동숭동 31-19	이탈리안	인생피자
8	0.84222	지노스 뉴욕피자	용산구 이태원동 457-3	기타 양식	인생피자집중하나.
9	0.84084	카니발피자	중구 남대문로5가 541	퓨전 양식	피자가 모두 하프엔하프가능해서 개이득!!
10	0.84063	모터시티	용산구 이태원동 56-30	이탈리안	인생 피자집. 바삭한 감자튀김에담백한 맛의 피자!

[표 3-22] ‘인생\_NNG’와 유사한 키워드 상위 20개

순위	유사도	단어	순위	유사도	단어
1	0.73621	존엄_NNG	11	0.68082	불변_NNG
2	0.72543	등극_NNG	12	0.68062	최고_NNG
3	0.71692	여태_NNG	13	0.67711	우왕_NNP
4	0.70607	역대_NNG	14	0.67681	최고봉_NNG
5	0.69932	종말_NNG	15	0.67606	기막히_VA
6	0.69693	박대박_NNP	16	0.67369	혈_VV
7	0.68988	젤로_NNP	17	0.67271	천상계_NNP
8	0.68817	맛있_VA	18	0.67204	진심_NNG
9	0.68725	다섯손가락_NNP	19	0.67191	기절초풍_NNG
10	0.68180	획기적_NNG	20	0.67150	개맛_NNP

[표 3-23] ‘피자\_NNG’와 유사한 키워드 상위 20개

순위	유사도	단어	순위	유사도	단어
1	0.81330	도우_NNP	11	0.76120	디아블로_NNP
2	0.79608	시카고피자_NNP	12	0.76057	뉴욕커_NNG
3	0.78918	페페로니_NNP	13	0.76052	콰트로_NNP
4	0.78592	알베르토_NNP	14	0.75621	슈프림_NNP
5	0.77786	피자헛_NNP	15	0.75492	디트로이트_NNP
6	0.77279	아티초크_NNP	16	0.75074	잭슨_NNP
7	0.76999	페퍼로니_NNP	17	0.75030	시카고_NNP
8	0.76524	도미노_NNG	18	0.74365	브룩클린_NNP
9	0.76223	피자스쿨_NNP	19	0.74312	콤비네이션_NNP
10	0.76165	화덕_NNG	20	0.74298	수프림_NNP

가중치에 따른 검색 결과를 살펴보기 위해 ‘인생\_NNG’에 가중치 0.9, ‘피자\_NNG’에 가중치 0.1을 적용해 보았다. 그 결과는 [표 3-24]와 같으며, ‘인생\_NNG’에 가중치를 높게 부여함으로써 ‘피자\_NNG’의 가중치가 낮아져 상위 10개로 도출된 POI의 음식 종류가 [표 3-21]에 비해 다양해진 것을 볼 수 있다.

[표 3-24] ‘인생(w: 0.9), ‘피자(w: 0.1)’ 검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.82002	빠네에 쌀레	송파구 방이동 205-4	기타 양식	친구가 런치를 만원에 먹을수있다고 해서 같이 갔었는데 가서 인생 프렌치토스트와 인생 로제파스타를 먹고왔다....
2	0.69569	원조 우리분식 멸치국수	양천구 신월동 520-1	국수/면 요리	신월동 토박이라면 한번쯤은 먹어봐야지 진짜 대존맛탱 양도 겁나많고 겁나 싸요 진짜♡♡ 내사랑
3	0.67499	25번지 김밥	은평구 진관동 25	기타 한식	맞은최고 김치볶음밥은 100번 넘게 먹어도 안질리고 비빔밥도 완전 맛있다 질리지않는 맛 최고최고 !!
4	0.67121	서강 껌테기	성북구 동선동2가 52	고기 요리	최고로 맛있습디다. 브아솔의 영준도 단골이라고 합니다 :)
5	0.66858	갈매기의 꿈	성동구 도선동 185-27	고기 요리	처음먹고 너무 맛있어서 일주일에 두세번도 갔었는데 갈매기살은 여기가 제일 맛있음! 강추맨 한번가서 먹어보세여
6	0.66546	오꾸미	양천구 신월동 546-25	해산물 요리	우리동네 체고 맛집. 치즈 꾸꾸미 대존맛이고 생맥 맛있음 ㄷㄷ
7	0.66394	PNB 풍년제과	양천구 목동 916	베이커리	맛있다 맛있다
8	0.65994	빠빠용 곱창	중구 황학동 628	기타 한식	친구가 인생 곱창이라고 데리고 갔는데 정말 인생 곱창집이 생겼어요.
9	0.65868	꾸꾸미 사랑	강동구 성내동 438-11	해산물 요리	양대박 가격착하고 맛은 더대박. 완전맛있음! 짱짱
10	0.65742	생소금 구이	성동구 용답동 60-9	고기 요리	숨어있는 맛집 가격도 저렴 맛도 최고. 삼겹,항정이 맛나고 서비스도 최고. 내인생맛집

[표 3-25]는 자연어 검색 시 토픽별 키워드가 3개인 경우(‘식물 덕후가 좋아할 만한 곳’, ‘스트레스 풀리는 매운맛’, ‘뷰 좋은 카페’)에 대한 형태소 분석 결과와 이에 대한 가중치를 임의로 부여한 것이다.

[표 3-25] 자연어 기반 검색 키워드 형태소 분석 결과 및 가중치 할당 (토픽별 키워드 3개)

연번	검색어	키워드					
		1	가중치	2	가중치	3	가중치
1	식물 덕후가 좋아할 만한 곳	식물_NNG	0.5	덕후_NNP	0.3	좋아하_VV	0.2
		토픽: 분위기		토픽: 긍정표현			
2	스트레스 풀리는 매운맛	스트레스_NNG	0.25	풀리_VV	0.25	매운맛_NNP	0.5
		토픽: 중식					
3	뷰 좋은 카페	뷰_NNP	0.4	좋_VA	0.3	카페_NNG	0.3
		토픽: 분위기			토픽: 카페/디저트		

‘식물 덕후가 좋아할 만한 곳’을 검색했을 때 ‘식물\_NNG’, ‘덕후\_NNP’, ‘좋아하\_VV(좋아하다)’에 가중치를 각각 0.5, 0.3, 0.2로 부여한 것과 관련도 높은 상위 10개 POI를 도출한 결과는 [표 3-26]과 같다. 상위 10개로 도출된 POI에 대해 작성된 대표 리뷰들을 살펴본 결과, ‘식물\_NNG’에 가장 높은 가중치를 부여했기에 해당 키워드와 관련된 표현인 ‘꽃’, ‘플라워’, ‘초록’ 등의 단어가 포함되어 있는 것을 확인할 수 있다.

[표 3-26] '식물(w: 0.5)', '덕후(w: 0.3)', '좋아하다(w: 0.2)'

검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.80980	코발트 무드	마포구 연남동 382-19	카페/디저트	실내는 마치 빈티지 취향의 의류브랜드 쇼룸같다. 오버사이즈 꽃자수니트가디건과 도트무늬 원피스가 벽면에 걸려있었고, 컵은 아기자기한 플라워 일러스트의 유리잔이라 귀엽게 데코된 케익들과 잘 어우러지고
2	0.80390	베르테	강서구 화곡동 1072-21	카페/디저트	커피향과맛이 좋아요 식물과 꽃이 많아 힐링되는느낌이에요
3	0.80110	플로릭 카페	동대문구 이문동 264-375	카페/디저트	크... 꽃이랑 식물 가득가득.. 넘나 힐링플레이스!
4	0.79951	VERS HOUSE	마포구 연남동 246-22	카페/디저트	플라워카페처럼 꽃 진짜 많고 앉을곳도많다.
5	0.79658	아리아페	마포구 동교동 164-30	카페/디저트	화려한 외관만큼 내부도 드라이플라워와 수많은 꽃들 소품들로 예쁘게 장식되어 있었습니다.
6	0.79528	까페이도	종로구 가회동 10-6	카페/디저트	윗층은도자기 갤러리 아래층은 공방 겸 카페를 운영. 그래서인지 분위기도 좋고 안락.
7	0.79398	Silhouett e Coffee	마포구 합정동 358-21	카페/디저트	사장님이 실루엣 덕후인것은 확실해요. 인테리어에 흰색 커튼이 전부니까요
8	0.79316	케플러 커피	서대문구 창천동 5-32	카페/디저트	친구말로는 엄청 하얗고 초록초록하다길래무슨 말이지 했는데 말그대로 하얀~~~배경에 초록 식물로 포인트를 준 인테리어더라고요
9	0.79273	카페가또	마포구 연남동 566-67	카페/디저트	일본의 작은 시골에 있는 가게 같기도 하고 뷰티인사이드에 나올 거 같은 카페였어요!
10	0.79142	카페325	마포구 동교동 190-1	카페/디저트	분위기도 식물이 많아서 너무 좋았음

‘스트레스 풀리는 매운맛’을 검색했을 때 ‘스트레스\_NNG’, ‘풀리\_VV(풀리는)’, ‘매운맛\_NNP’에 가중치를 각각 0.25, 0.25, 0.5로 부여한 것과 관련도 높은 상위 10개 POI를 도출한 결과는 [표 3-27]과 같다. 상위 10개로 도출된 POI에 작성된 대표 리뷰들의 내용을 보면, ‘매운맛\_NNP’에 가장 높은 가중치를 부여했기에 해당 키워드와 관련된 표현인 ‘매운’, ‘매운거’, ‘칼칼한맛’, ‘청양고추’ 등의 단어가 포함되어 있는 것을 확인할 수 있다.

[표 3-27] ‘스트레스(w: 0.25)’, ‘풀리는(w: 0.25)’, ‘매운맛(w: 0.5)’  
검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.88424	신맷당	마포구 서교동 346-41	고기 요리	매운맛 조절이 가능한 매운갈비찜
2	0.88201	신전떡볶이	영등포구 영등포동1가 144	기타 한식	매운거잘 못먹어도먹히는 매운맛인데, 고추기름같은 매운맛은 아니고 캡사이신넣은 매운맛..?
3	0.84991	열봉찜닭	종로구 종로2가 6	닭/오리 요리	매운거좋아한다면 추천!
4	0.84386	라화쿵부	관악구 신림동 1637-2	기타 중식	매운맛은 점바점인지얼얼한 맛으로 했는데도 별로 안매웠다
5	0.84234	부뚜막 매운갈비찜	송파구 석촌동 3-1	고기 요리	계란후라이랑 매운맛 돼지갈비찜이 제일 맛있음!술땡기는칼칼한맛
6	0.83961	두만강 양꼬치	관악구 신림동 1429-17	기타 중식	고기도 맛있고 직원분들도 친절하셔서 기분 좋았던 곳. 양념꼬치는매운맛 덜매운맛두 종류인데 반반 섞어주시기도한다.
7	0.83917	신림매운 등갈비찜	관악구 신림동 1432-53	고기 요리	지금까지 갔던 매운갈비찜 집 중에 맛도 가장 일관성있고 알바생도 매우 친절함. 매운맛 단계는 3단계는 좀 맵고 2단계가 무난한 듯.
8	0.83786	성하루	강남구 신사동 525-24	정통 중식/ 일반 중식	깐풍기는 기분 좋게 매운 맛이다.
9	0.83626	맷당	성동구 행당동 4-33	고기 요리	총4단계까지있는데4단계가 매운맛 세계1위 부트졸로키아인가하는 고추종류였음, 청양고추는 2단계밖에안됨.
10	0.83558	여의정	영등포구 여의도동 23-3	한정식/ 백반/ 정통 한식	아 정말 내가 이런데맷글 웬만해서 안다는데 매운갈비찜이랑 뚝배기닭갈비먹고나서 맷을안달수가 없어서 로그인했어요. 진짜 개맛있음ㅎ. 특히 뚝닭이 완전 내취향.

[표 3-28]은 ‘뷰 좋은 카페’ 검색 시 ‘뷰\_NNP’, ‘좋은\_VA’, ‘카페\_NNG’에 가중치를 각각 0.4, 0.3, 0.3으로 부여한 것과 관련도 높은 상위 10개 POI를 도출한 결과이다. 상위 10개로 도출된 POI에 대해 작성된 대표 리뷰들의 내용을 보면, ‘뷰\_NNP’에 가장 높은 가중치를 부여했기에 해당 키워드와 관련된 표현인 ‘전망’, ‘야경’ 등의 단어가 포함되어 있는 것을 확인할 수 있다.

[표 3-28] ‘뷰(w: 0.4)’, ‘좋은(w: 0.3)’, ‘카페(w: 0.3)’  
검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.92123	카페쵸쵸드	종로구 삼청동 63-31	카페/디저트	무엇보다 한적하고 좋은 뷰도 보여주면서 고즈넉한 느낌을 가진 카페 분위기가 너무 좋았다..!
2	0.91300	티티티 (TTT)	송파구 신천동 29	카페/디저트	석촌호수를볼 수 있는 전망 좋은 카페
3	0.90344	서울루덴스	용산구 이태원동 5-12	카페/디저트	많은 루프탑카페나 술집 중에서도 야경이 안 좋은 곳도 많은데 여기의 야경은 가히 최고 라고할 만 하다! 시원한 음료와 함께 남산 뷰를 보고 싶은 사람은 강추!
4	0.89427	연	종로구 삼청동 63-20	카페/디저트	팬찮아요!! 다락방에있었어요 분위기가너무좋아요
5	0.89315	개뿔	종로구 이화동 9-443	카페/디저트	낙산공원의약간 아래 쪽에 가장 좋은 자리를 떡하니차지하고 있는 개뿔 이라는 카페. 커피 맛도, 서비스도 (특히 이름이 구름) 그냥저냥인데, 뷰 하나 만큼은 끝!
6	0.89243	플로렛카페	강남구 도곡동 454-15	카페/디저트	밀크티와쿠키가 맛있는 카페. 양재천을 바라보는 뷰도 좋다.
7	0.88973	84라운지	용산구 이태원동 211-2	카페/디저트	뷰가 좋았는데 좀 좁다
8	0.88775	불끼	종로구 명륜2가 135-3	카페/디저트	개인카페음료는 거기서 거기라고 생각하고 실제로 거기서 거기. 그러나 2층에서 성균관대를 바라보는 뷰가 꽤나 좋다
9	0.88552	파리바게뜨	용산구 이촌동 300-19	베이커리	브런치하기 딱 좋은 분위기^^
10	0.88407	카페 한옥	종로구 원서동 20-1	카페/디저트	맛은 무난하지만 컨셉이 좋은 - 한옥카페

### 3.5.2.2 사용자 평가 결과 분석

본 연구에서는 시맨틱 POI 검색에의 활용 결과에 대한 품질을 확인하기 위하여 설문 조사를 통해 종합적인 사용자 평가를 수행하였다. 설문 조사 기간은 2020년 10월 14일부터 16일까지이고, 조사 대상은 서울시에 거주하는 20대 이상 50대 이하의 남녀이다. 설문 조사는 오픈서베이 플랫폼을 이용하였으며 모바일 애플리케이션을 통해 응답을 수집하였다. 문항세트 1개 당 50개 문항에 대한 100명의 응답을 수집하였으며, 총 4개의 문항세트(a~d)에 대한 조사를 동시다발적으로 실시하였다. 자료 분석을 위한 통계 처리는 SPSS 26 프로그램을 활용하였다.

전체 응답자 400명 중 여성이 201명(50.25%), 남성은 199명(49.75%)이며, 응답자의 직업군은 사무/기술직 34.5%, 자유/전문직 15.75%, 전업주부 11.5%, 경영/관리직 9.75% 순으로 많았다([표 3-29]). 결과적으로 총 20,000개(50문항x100명x4세트)의 응답을 수집한 것인데 대체로 남성이 여성에 비해 부정적 응답을 더 많이 하는 경향을 보였다. 통계로 보면, 남성이 여성에 비해 ‘매우 불만족’ 항목을 1.67배 더 많이 선택한 반면, 여성이 남성에 비해 ‘매우 만족’을 1.23배 더 많이 선택하였다.

[표 3-29] 응답자 특성

응답자 직업군	응답자 수(명)	응답자 비율(%)	응답자 직업군	응답자 수(명)	응답자 비율(%)
사무/기술직	138	34.50	자영업	20	5.00
자유/전문직	63	15.75	무직	12	3.00
전업주부	46	11.50	기능/작업직	11	2.75
경영/관리직	39	9.75	대학원생	6	1.50
판매/영업 서비스직	33	8.25	농/임/어/축산업	1	0.25
대학생	31	7.75	합계	400	100.00

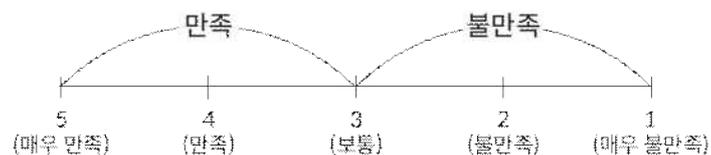
설문 결과 분석에 앞서, 설문 자료의 신뢰도 추정에 가장 많이 활용되는 크론바흐 알파(Cronbach's alpha) 계수를 이용해 문항세트 a~d에 대한 신뢰도 검정을 시행하였다. 크론바흐 알파는 한 설문지를 구성하고 있는 문

항 간의 일관성을 일컫는 내적 일관성(internal consistency)을 추정할 때 사용되는 것이다. 일반적으로 크론바흐 알파값이 0.7 이상이면 조사 도구의 신뢰성이 적정하다고 판단한다(Nunally, 1994). 문항세트 a~d의 POI 검색 종류별 크론바흐 알파값은 [표 3-30]과 같다. 해당 표를 보면, 모든 문항세트의 검색 종류별 신뢰도 값이 0.9 이상으로 나타났으며, 이는 신뢰성의 판단 기준이 되는 0.7 이상이므로, 본 연구에서 구성한 설문지의 신뢰도는 매우 높다고 판단할 수 있다.

[표 3-30] 설문지의 신뢰도 분석 결과

구분	문항세트별 크론바흐 알파값			
	a	b	c	d
자연어 기반 POI 검색 (문항 수: 20개)	0.926	0.900	0.916	0.922
선택 기반 POI 검색 (문항 수: 20개)	0.934	0.917	0.928	0.931
유사 POI 검색 (문항 수: 10개)	0.946	0.951	0.947	0.929

본 연구에서는 사용자 만족도 평가 결과 해석 시 [그림 3-8]과 같이 리커트 5점 척도의 중간값인 3점(보통)을 기준으로 사용자 평가 점수가 3점을 초과하는 경우에는 만족의 범위에 속하는 것으로 판단하고, 3점 미만인 경우에는 불만족의 범위에 속하는 것으로 판단하였다. [표 3-31]은 자연어 기반 POI 검색에 대한 사용자 만족도 평가 결과를 나타낸다. a~d의 모든 문항 세트에 대한 만족도 평균이 3.48점으로 산출됨에 따라 자연어 기반 POI 검색 결과에 대한 사용자 평가는 만족의 범위에 속하는 것으로 판단할 수 있다.



[그림 3-8] 만족도 평가 기준

[표 3-31] 자연어 기반 POI 검색에 대한 사용자 만족도 평가 결과

구분		a	b	c	d	평균
자연어 기반 POI 검색 (문항 수: 20개)	평균	3.52	3.49	3.50	3.42	<b>3.48</b>
	최솟값	3.04	2.96	3.00	2.92	2.98
	최댓값	3.73	3.73	3.76	3.61	3.71
	분산	0.03	0.04	0.03	0.03	0.03

사용자 만족도 평가 결과에서 만족 정도에 따른 값의 분포는 [표 3-32]와 같다. a~d의 모든 문항세트에서 ‘만족’을 선택한 응답자 비율이 가장 높았으며, ‘매우 만족’ 및 ‘만족’을 선택한 응답자가 ‘매우 불만족’ 및 ‘불만족’을 선택한 응답자보다 약 3.3~5.2배 더 많았다. 이를 통해 자연어 기반 POI 검색 결과에 대한 서비스 이용자들의 만족 정도가 불만족 정도에 비해 더 높다는 것을 알 수 있다.

[표 3-32] 자연어 기반 POI 검색에 대한 사용자 만족도 결과 분포

단위: %(명)

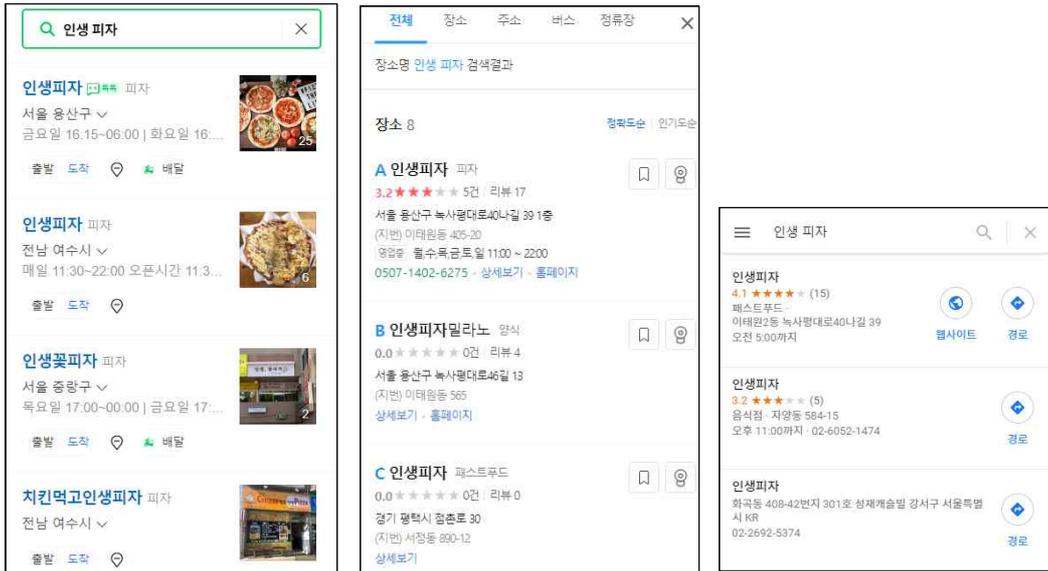
구분		매우 만족	만족	보통	불만족	매우 불만족
자연어 기반 POI 검색 (문항 수: 20개)	a	11.35(227)	43.45(869)	32.45(649)	11.30(226)	1.45(29)
	b	8.15(163)	43.8(876)	38.10(762)	8.35(167)	1.60(32)
	c	9.80(196)	44.65(893)	33.85(677)	9.00(180)	2.70(54)
	d	11.30(226)	37.05(741)	36.85(737)	11.85(237)	2.95(59)
		10.15	42.24	35.31	10.13	2.18
평균(%)		<p>A pie chart illustrating the distribution of user satisfaction levels. The chart is divided into five segments: '매우 만족' (Very Satisfied) at 2.18%, '만족' (Satisfied) at 42.24%, '보통' (Average) at 35.31%, '불만족' (Dissatisfied) at 10.13%, and '매우 불만족' (Very Dissatisfied) at 10.15%. A legend on the right side of the chart identifies each category with a corresponding color: dark blue for '매우 만족', light blue for '만족', green for '보통', yellow for '불만족', and orange for '매우 불만족'.</p>				

### 3.5.2.3 기존 서비스와의 비교 분석

본 연구에서 개발한 시맨틱 POI 검색 기술을 활용하면 사용자들의 다양한 선호와 상황을 반영한 검색이 가능하다. 기존 국내 POI 검색 서비스들은 POI 명칭, 카테고리, 또는 태그 위주의 검색을 제공하고 있으며, 네이버 지도는 리뷰 검색을 일부 지원하고 있으나 ‘맛집’, ‘분위기’, ‘뷰’ 등 사용자 수요가 많은 특정 검색어에 한해 검색 가능하므로 검색할 수 있는 키워드의 범위가 한정적이다. 구글 지도는 국내 서비스들에 비해서는 사용자 리뷰 데이터를 검색에 적극적으로 활용하고 있으나 POI 명칭 검색과 리뷰 검색이 혼재되어 있다. 이에 따라 검색어에 대한 결과로, 해당 검색어가 POI 명칭에 포함된 결과와 리뷰에 포함된 결과가 함께 출력되는 경향이 있다. 그리고 구글 지도에서는 검색어와 정확히 일치하는 단어가 포함된 리뷰를 위주로 검색하기 때문에 해당 검색어와 일치하는 단어가 없으면 검색이 이루어지지 않으며, 동의어 및 유의어를 활용한 검색에는 한계가 있다. 또한, 국내외 POI 검색 서비스는 사용자가 검색 키워드에 가중치를 직접 적용해 검색을 수행할 수 있는 기능을 제공하지 않아 사용자들의 다양한 요구사항을 반영한 맞춤형 검색이 어려운 한계가 있다. 본 연구에서 제안하는 시맨틱 POI를 검색에 활용하면 기존 서비스들이 가지고 있는 한계점들을 보완할 수 있다. 이를 입증하기 위하여 본 연구에서 도출한 시맨틱 POI 검색 결과와 기존 유사 서비스들에서의 검색 결과를 비교 분석하였다.

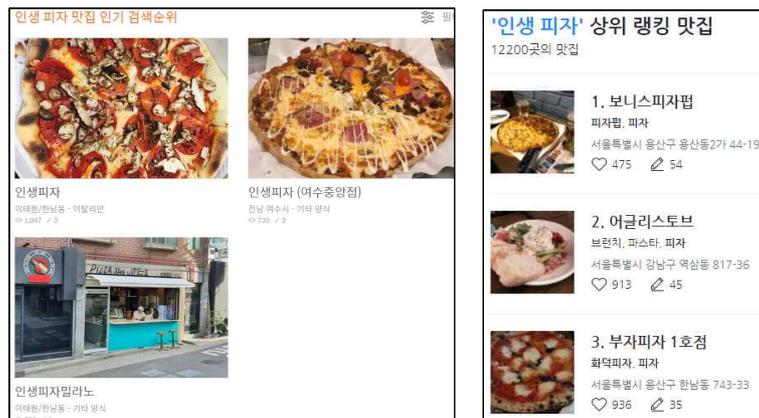
본 절에서는 자연어 기반 POI 검색 결과를 지도 중심 플랫폼(네이버 지도, 카카오맵, 구글 지도) 및 맛집 중심 플랫폼(망고플레이트, 다이닝코드)에서의 검색 결과와 비교해 보았다. 첫 번째로 비교 분석할 자연어 기반 키워드는 ‘인생 피자’이다. 해당 키워드로 지도 중심 플랫폼과 맛집 중심 플랫폼에서 각각 검색을 수행해 보면, [그림 3-9] 및 [그림 3-10]과 같이 모든 플랫폼에서 POI 명칭에 ‘인생’과 ‘피자’라는 키워드가 포함된 결과를 반환하는 것을 알 수 있다. 그러나 본 연구에서 개발한 자연어 기반 POI 검색 결과에서는 앞에서 살펴본 [표 3-21] 및 [표 3-24]와 같이 리뷰에 ‘인

생' 및 '피자'와 동일하거나 유사한 키워드가 다른 POI에 비해 중요하게 사용된 POI들이 결과로 출력되며, 해당 결과는 사용자가 입력한 가중치에 따라 달라진다.



(a) 네이버 지도(일부)      (b) 카카오맵(일부)      (c) 구글 지도

[그림 3-9] '인생 피자' 검색 결과(지도 중심 플랫폼)



(a) 망고플레이트      (b) 다이닝코드(일부)

[그림 3-10] '인생 피자' 검색 결과(맛집 중심 플랫폼)

두 번째로 비교 분석할 자연어 기반 키워드는 '식물 덕후가 좋아할 만한

곳'이다. 해당 키워드로 지도 중심 플랫폼에서 검색을 수행한 결과는 [그림 3-11]과 같은데, 모든 플랫폼에서 검색 결과가 출력되지 않았다. 동일 키워드로 맛집 중심 플랫폼에서 검색을 수행한 결과는 [그림 3-12]와 같으며, 유일하게 결과가 출력된 다이닝코드의 경우 POI 명칭에 검색 키워드가 포함된 결과를 반환해 주는 것을 확인할 수 있었다. 그러나 본 연구에서 개발한 자연어 기반 POI 검색 결과에서는 앞에서 살펴본 [표 3-26]과 같이 리뷰에 '식물', '덕후', '좋아하다'와 동일하거나 유사한 키워드가 다른 POI에 비해 중요하게 사용된 POI들이 결과로 출력되며, 해당 결과는 사용자가 입력한 가중치에 따라 달라진다.



[그림 3-11] '식물 덕후가 좋아할 만한 곳' 검색 결과(지도 중심 플랫폼)

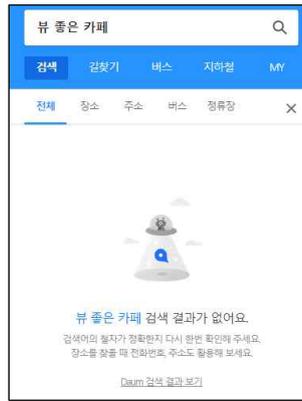


[그림 3-12] '식물 덕후가 좋아할 만한 곳' 검색 결과(맛집 중심 플랫폼)

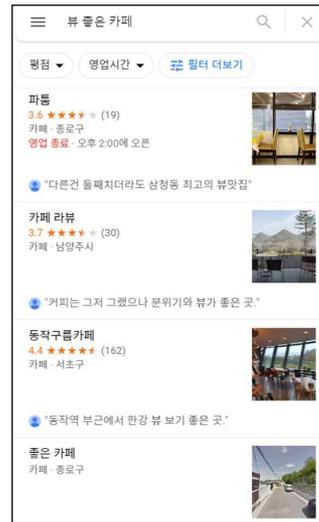
세 번째로 비교 분석할 자연어 기반 키워드는 ‘뷰 좋은 카페’이다. 해당 키워드로 지도 중심 플랫폼에서 검색을 수행한 결과는 [그림 3-13]과 같으며, 카카오맵을 제외한 나머지 2개 플랫폼에서 검색 키워드에 대한 결과로 POI 목록이 출력되었다. 네이버 지도와 구글 지도에서 결과로 출력된 POI 들에 대한 정보를 살펴본 결과, 검색 키워드인 ‘뷰’, ‘좋은’, ‘카페’와 일치하는 단어가 리뷰에 포함된 POI 목록을 결과로 반환해 준 것으로 보인다. 그러나 네이버 지도는 검색어에 대한 결과로 단 4개의 POI만 출력되었으며, 구글 지도는 POI 명칭이 ‘좋은 카페’인 결과가 출력되는 한계가 있었다. 또한, 구글 지도에서는 검색 키워드와 정확히 일치하는 단어가 포함된 리뷰를 위주로 검색하므로 동의어 및 유의어에 대한 고려는 부족한 것으로 보인다. ‘뷰 좋은 카페’로 맛집 중심 플랫폼에서 검색을 수행한 결과는 [그림 3-14]와 같다. 다이닝코드의 경우, 검색어와 일치하는 태그가 포함된 POI 목록을 반환해 주는 것으로 파악되는데, 검색 키워드 중 ‘카페’ 태그와 관련된 POI들이 주로 출력된 것을 확인할 수 있었다. 반면 망고플레이트의 경우에는 결과로 반환된 POI들의 정보를 살펴봤을 때 검색 키워드와 관계 없이 사용자 평점이 높은 ‘카페’ 관련 POI 목록을 출력해 준 것으로 보인다. 그러나 본 연구에서 개발한 자연어 기반 POI 검색 결과에서는 앞에서 살펴본 [표 3-28]과 같이 리뷰에 ‘뷰’, ‘좋은’, ‘카페’와 정확히 일치하는 키워드뿐 아니라 가중치가 높게 부여된 키워드인 ‘뷰’와 유사한 단어인 ‘전망’, ‘야경’ 등이 포함된 POI들도 결과로 출력되었다. 해당 결과는 사용자가 입력한 가중치에 따라 달라지기 때문에 기존 서비스들과 다르게 사용자 맞춤형 검색이 가능하다. 더욱 다양한 자연어 기반 검색 키워드를 이용해, 본 연구에서 도출된 결과와 기존 국내외 POI 검색 서비스에서 도출된 결과를 비교한 내용은 <부록 B.1>에 첨부하였다.



(a) 네이버 지도



(b) 카카오맵



(c) 구글 지도(일부)

[그림 3-13] '뷰 좋은 카페' 검색 결과(지도 중심 플랫폼)



(a) 망고플레이트(일부)



(b) 다이닝코드(일부)

[그림 3-14] '뷰 좋은 카페' 검색 결과(맛집 중심 플랫폼)

### 3.5.3 선택 기반 POI 검색에의 활용 결과

#### 3.5.3.1 실험 결과 분석

선택 기반 POI 검색에서는 먼저 사용자가 토픽을 선택하고, 해당 토픽에 속하는 키워드를 선택한 다음, 각 토픽별 키워드에 가중치를 부여해야 한다. 먼저, 검색어와 그 결과가 합리적으로 매칭되는지 확인하기 위해 토픽별로 선택한 키워드에 대한 결과로 도출된 POI들의 음식 종류 항목을 살펴보았다. [표 3-33]은 토픽 중에서 ‘이탈리안’을 선택하고, 키워드로는 ‘파스타\_NNP’를 선택하였을 때 가장 관련도 높은 상위 100개 POI의 음식 종류에 대한 통계다. 총 9개의 음식 종류가 도출되었는데, ‘이탈리안’이 58개로 가장 많았고, ‘기타 양식(27개)’, ‘퓨전 양식(6개)’, ‘스테이크/바베큐(3개)’, ‘세계음식 기타(2개)’ 순으로 많았다. 그 외에도 ‘시푸드 요리’, ‘브런치/버거/샌드위치’, ‘전통 주점/포차’, ‘기타 한식’ 카테고리에 해당하는 POI가 각각 1개씩 추출되었다. 비록 4개 POI가 ‘이탈리안’에 속하는 음식점은 아니지만 파스타 메뉴를 판매하고 있었고, 사용자들이 리뷰에 ‘파스타’ 관련 내용을 언급하였기에 관련 장소로 추출된 것을 확인할 수 있었다. 만약 기존 검색 시스템에서 파스타 파는 곳을 찾기 위해 ‘이탈리안’ 카테고리에 속하는 POI만 필터링하였다면 보다 좁은 범위의 검색 결과만 얻을 수 있었을 것이다.

[표 3-33] 토픽: ‘이탈리안’, 키워드: ‘파스타’ 선택 결과 도출된 상위 100개 POI의 음식 종류 통계

연번	음식 종류	관련 POI	개수
1	이탈리안	몽고네, 살롱드쥬	58
2	기타 양식	루이루이, 도시락파스타	27
3	퓨전 양식	이층양옥, 소온테이블	6
4	스테이크/바베큐	미니자이언트, 호텔로프트레스토랑	3
5	세계음식 기타	지중해, 단밤	2
6	시푸드 요리	양키랍스터	1
7	브런치/버거/샌드위치	알렉스키친	1
8	전통 주점/포차	곰포차	1
9	기타 한식	분스타	1
합계			100

[표 3-34]는 토픽 중 ‘이탈리안’을 선택하고, 키워드로 ‘파스타\_NNP’를 선택하였을 때 가장 관련도가 높은 상위 10개 POI를 출력한 것이다. 상위 10개 도출된 POI들의 음식 종류는 ‘이탈리안’ 및 ‘기타 양식’인데, 사용자들이 작성한 리뷰 내용을 통해 ‘기타 양식’으로 분류된 POI에서도 파스타를 판매하고 있는 것을 확인할 수 있었다.

[표 3-34] 토픽: ‘이탈리안’, 키워드: ‘파스타’ 선택 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.94309	루이루이	중구 신당동 370-120	기타 양식	피자는 매우 굿굿. 파스타는 알단테 삶기가 아니여서 아쉽. 이견 취향차이일듯
2	0.93035	도시락 파스타	종로구 계동 8-2	기타 양식	컨셉도 정겹고 파스타 맛도 좋았어요~ 뽀모도로랑 날치알 크림 파스타 먹었는데 느끼하지않고 적당해요ㅎㅎ
3	0.92484	몽고네	서대문구 연희동 192-29	이탈리안	이렇게 작은 가게가 맛하나로 명성이 대단하네요. 바삭 촉촉한 식전빵부터 시작, 알단테 이상으로 환상의 식감이었던 파스타들 !!
4	0.92480	제이비 파스타	동대문구 회기동 348-3	기타 양식	오랜만에 갔는데 역시나 맛있다 :) 식전 빵도 촉촉하고 특히 크림 파스타 핵존맛 :) 감베로니 완전 추천한다!
5	0.91597	라고 파스타	양천구 목동 917-9	기타 양식	많은 파스타를 먹어봤지만 까르보나라는 여기가 최고예요. 중학교때부터 벌써 15년째 단골이네요.
6	0.91503	살롱드쥬	송파구 방이동 165-2	이탈리안	여기 봉골레 파스타는 파스타 안먹는 사람도 먹게하는 재주가있다. 저번에 비해서 양도 많아진거같은 느낌적인 느낌.
7	0.91157	다옴383	마포구 연남동 383-44	이탈리안	부담없이 괜찮은 파스타를 먹을 수 있는 곳.
8	0.91028	방카	도봉구 창동 664-25	이탈리안	떡볶이파스타는 맛있었는데 크림파스타는 별로였음ㅎㅎ 친절하긴 했지만 내스타일은 아니었던 곳
9	0.90988	봉지미 테헤란 16번가	강남구 역삼동 735-24	이탈리안	역삼러라면 꽤 알고있을 파스타집. 길은 허름해보이고 내부도 작아보이는데 들어가면 생각보다 넓고 분위기도 그럭저럭 괜찮다. 그리고 파스타류는 다 맛이 괜찮고
10	0.90809	씨푸드 살롱 마레스타	마포구 동교동 203-19	이탈리안	가성비 최고의 파스타집. 양 메뉴 모두 9500원. 진짜 맛있고 매콤하고 감칠맛난다. 연남동 잘 안오는데 올 때마다 갈 것 같다.

가중치(w)가 제대로 작동하는지 확인하기 위해 ‘이탈리안’ 토픽에서 선택한 기존 키워드 ‘파스타\_NNP’에, 동일 토픽 내 키워드인 ‘맛있\_VA’를 추가 선택해 가중치를 달리 적용하면서 결과를 비교해 보았다. 가장 먼저, ‘파스타\_NNP’에 가중치 0.9, ‘맛있\_VA’에 가중치 0.1을 적용한 결과는 [표 3-35]와 같다. ‘파스타\_NNP’에 1에 가까운 가중치를 적용했기에 ‘파스타\_NNP’만 선택해 검색했을 때의 결과([표 3-34])와 상위 7개로 도출된 POI들의 순위가 동일한 것을 확인할 수 있었다.

[표 3-35] 토픽: ‘이탈리안’, 키워드: ‘파스타(w: 0.9)’, ‘맛있는(w: 0.1)’  
선택 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.95518	루이루이	중구 신당동 370-120	기타 양식	[표 3-34] 1위 내용 참고
2	0.94085	도시락 파스타	종로구 계동 8-2	기타 양식	[표 3-34] 2위 내용 참고
3	0.93828	몽고네	서대문구 연희동 192-29	이탈리안	[표 3-34] 3위 내용 참고
4	0.93617	제이비 파스타	동대문구 회기동 348-3	기타 양식	[표 3-34] 4위 내용 참고
5	0.92966	라고파스타	양천구 목동 917-9	기타 양식	[표 3-34] 5위 내용 참고
6	0.92835	살롱드쥬	송파구 방이동 165-2	이탈리안	[표 3-34] 6위 내용 참고
7	0.92581	다움383	마포구 연남동 383-44	이탈리안	[표 3-34] 7위 내용 참고
8	0.92253	봉지미 테헤란 16번가	강남구 역삼동 735-24	이탈리안	[표 3-34] 9위 내용 참고
9	0.92145	씨푸드살롱 마레스타	마포구 동교동 203-19	이탈리안	[표 3-34] 10위 내용 참고
10	0.92064	지중해	종로구 익선동 146	세계음식 기타	오랜만에 정말 맛있게 파스타를 먹을 수 있었던 곳! 바질페스토파스타는 진짜 너무 맛있었다TT 내 인생 바질페스토폼는 곳이 두군데있었는데 이제 여기가 아마 1등...?

‘파스타\_NNP’, ‘맛있\_VA’에 대한 가중치를 각각 0.7과 0.3으로 적용했을 때의 결과인 [표 3-36]을 보면, 상위 10개로 도출된 POI 목록은 동일한데, 그 안에서 순위가 변동된 것을 알 수 있다.

[표 3-36] 토픽: ‘이탈리안’, 키워드: ‘파스타(w: 0.7)’, ‘맛있는(w: 0.3)’  
선택 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.96649	루이루이	중구 신당동 370-120	기타 양식	[표 3-35] 1위 내용 참고
2	0.95301	몽고네	서대문구 연희동 192-29	이탈리안	[표 3-35] 3위 내용 참고
3	0.94874	도시락 파스타	종로구 계동 8-2	기타 양식	[표 3-35] 2위 내용 참고
4	0.94623	지중해	종로구 익선동 146	세계음식 기타	[표 3-35] 10위 내용 참고
5	0.94614	제이비 파스타	동대문구 회기동 348-3	기타 양식	[표 3-35] 4위 내용 참고
6	0.94508	라고파스타	양천구 목동 917-9	기타 양식	[표 3-35] 5위 내용 참고
7	0.94295	살롱드쥬	송파구 방이동 165-2	이탈리안	[표 3-35] 6위 내용 참고
8	0.94257	다움383	마포구 연남동 383-44	이탈리안	[표 3-35] 7위 내용 참고
9	0.93628	씨푸드살롱 마레스타	마포구 동교동 203-19	이탈리안	[표 3-35] 9위 내용 참고
10	0.93570	봉지미 테헤란 16번가	강남구 역삼동 735-24	이탈리안	[표 3-35] 8위 내용 참고

[표 3-37]은 두 키워드에 동일한 가중치 0.5를 적용한 결과로, 이전 가중치 적용 결과와 거의 유사하나 순위에 변동이 있고, 상위 10개 목록에 새로운 POI인 '523'과 '시장에간남자'가 등장한 것을 확인할 수 있다.

[표 3-37] 토픽: '이탈리안', 키워드: '파스타(w: 0.5)', '맛있는(w: 0.5)'  
선택 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.95000	루이루이	중구 신당동 370-120	기타 양식	[표 3-36] 1위 내용 참고
2	0.94710	지중해	종로구 익선동 146	세계음식 기타	[표 3-36] 4위 내용 참고
3	0.94093	몽고네	서대문구 연희동 192-29	이탈리안	[표 3-36] 2위 내용 참고
4	0.93676	523	마포구 연남동 383-82	이탈리안	새우 게살 크림 파스타는 가장최근에갔을때먹었는데 맛있다.
5	0.93407	라고파스타	양천구 목동 917-9	기타 양식	[표 3-36]의 6위 내용 참고
6	0.93319	다움383	마포구 연남동 383-44	이탈리안	[표 3-36]의 8위 내용 참고
7	0.93103	살롱드쥬	송파구 방이동 165-2	이탈리안	[표 3-36]의 7위 내용 참고
8	0.92878	도시락 파스타	종로구 계동 8-2	기타 양식	[표 3-36]의 3위 내용 참고
9	0.92870	제이비 파스타	동대문구 회기동 348-3	기타 양식	[표 3-36]의 5위 내용 참고
10	0.92678	시장에간 남자	서대문구 창천동 288	이탈리안	맛 최고 분위기 최고 기념일맞이해서또갔다 너무 맛있어서 여기서 같이 밥 먹은 사람이랑은 평생 친하게 지낼 수 있게 만드는 맛

[표 3-38]는 ‘파스타\_NNP’, ‘맛있\_VA’에 대한 가중치를 각각 0.3과 0.7로 적용했을 때의 결과로, [표 3-34]~[표 3-37]에서 계속 1위를 차지했던 ‘루이루이’가 8위로 밀려나고 ‘지중해’라는 POI가 1위로 올라왔다. 그 외 상위 10위권에 새로운 POI 6개(‘클래식당’, ‘22번가키친’, ‘티톨로’, ‘이층양옥’, ‘더그리드’, ‘고메트리’)가 등장한 것을 볼 수 있다.

[표 3-38] 토픽: ‘이탈리안’, 키워드: ‘파스타(w: 0.3)’, ‘맛있는(w: 0.7)’  
선택 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.91126	지중해	종로구 익선동 146	세계음식 기타	[표 3-37] 2위 내용 참고
2	0.90410	523	마포구 연남동 383-82	이탈리안	[표 3-37] 4위 내용 참고
3	0.90386	클래식당	마포구 연남동 383-95	이탈리안	늦게까지하는파스타맛집. 새벽2시까지영업. 조용히식사하기좋은집
4	0.90205	22번가 키친	성북구 동선동2가 135-26	기타 양식	파스타가 괜찮다고 보기도 해서 겸사겸사 가게되었어요. 새우 알리오올리오는 페퍼론치노가 들어가서 그런지 약간 매콤하고 짭짤해서 느끼하지 않고 좋았습니다.
5	0.89955	티톨로	강남구 청담동 79-8	이탈리안	강추 천만개
6	0.89589	시장에간 남자	서대문구 창천동 288	이탈리안	[표 3-37] 10위 내용 참고
7	0.89497	이층양옥	종로구 익선동 110-4	퓨전 양식	스파이시 로제파스타는 제가 좋아하는 스타일의 (다이닝랩이랑비슷해여!) 소스였어요 맛이떠여
8	0.89482	루이루이	중구 신당동 370-120	기타 양식	[표 3-37] 1위 내용 참고
9	0.89476	더그리드	마포구 서교동 404-8	이탈리안	직원분들도 너무 핵친절하시고 음식도 뭐이건너무 맛있잖아요ㅠㅠ
10	0.89323	고메트리	성동구 금호동4가 235	프랑스 음식	파스타는 비싼집에서 먹어야하나보다 이런 식감이나오다니 너무 조화롭고 넘 넘엄 좋다. 최고

‘파스타\_NNP’에 0.1, ‘맛있\_VA’에 0.9의 가중치를 적용한 결과인 [표 3-39]를 보면, 이전 가중치 적용 결과에서 5위였던 ‘티톨로’가 1위로 올라왔고, POI 5개(‘Give Me Fever’, ‘오늘그릴’, ‘레스트로’, ‘페닌술라’, ‘마렘마트라토리아’)가 새롭게 등장한 것을 볼 수 있다. 키워드별로 가중치를 다르게 적용함에 따라 상위권에 새로운 POI들이 등장하는 것의 의미는 상위권에서 사라진 POI에 비해 새롭게 등장한 POI에서 해당 키워드가 더 중요하게 사용되었다는 것을 뜻한다.

[표 3-39] 토픽: ‘이탈리안’, 키워드: ‘파스타(w: 0.1)’, ‘맛있는(w: 0.9)’  
선택 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.84962	티톨로	강남구 청담동 79-8	이탈리안	[표 3-38] 5위 내용 참고
2	0.84882	Give Me Fever	마포구 연남동 566-60	이탈리안	엔초비 파스타는 면도 잘 삶아지고 깔끔한게 좋았고 무엇보다 노끼가 정말 쫄쫄맛있었다!
3	0.84802	오늘그릴	중랑구 신내동 410-69	스테이크/바베큐	여기 좀 짱인 듯. 메뉴는 파스타랑 플래터같은 것들 팔더라구요!!!! 닭고기 빨간소스 묻은거 꼭 드셔보세요 소스 맛이 대박입니다 ♥
4	0.84795	클래식당	마포구 연남동 383-95	이탈리안	[표 3-38] 3위 내용 참고
5	0.84691	레스트로	서초구 반포동 90-3	이탈리안	서래마을에 위치한 레스트로 작고 아담한 사이즈 음식의 밸런스와 스타일이 아주 좋았어요.
6	0.84546	고메트리	성동구 금호동4가 235	프랑스 음식	[표 3-38] 9위 내용 참고
7	0.84411	더그리드	마포구 서교동 404-8	이탈리안	[표 3-38] 10위 내용 참고
8	0.84266	페닌술라	중구 소공동 1	이탈리안	맛있었어요. 한우 카르파초는 구운치즈 과자에 올려 먹으니 더 맛있는듯 했습니다. 어니언스프도 괜찮았어요.
9	0.84133	22번가 키친	성북구 동선동2가 135-26	기타 양식	[표 3-38] 4위 내용 참고
10	0.83877	마렘마트라토리아	용산구 한남동 739-17	이탈리안	맛있어요! 파스타랑 리조또 그리고 아란치니 먹었는데 다 정말 맛있었어요 파스타 진짜 대박대박 맛있어요 파스타 추천해요!

마지막으로, ‘이탈리안’ 토픽에서 키워드 ‘맛있\_VA’를 선택하였을 때의 결과는 [표 3-40]과 같다. 이전 가중치 적용 결과에서 3위였던 ‘오늘그릴’이 1위로 올라왔고, POI 3개(‘르포트13’, ‘꿈모아’, ‘엘레브’)가 새로 등장한 것을 볼 수 있다. [표 3-34]~[표 3-40] 결과를 종합해 보면, 동일한 키워드 일지라도 가중치를 다르게 적용함에 따라 다른 결과를 보이는 것을 알 수 있다. 즉, POI 검색 결과는 가중치가 더 높은 키워드와 더 많은 관련성을 갖는다.

[표 3-40] 토픽: ‘이탈리안’, 키워드: ‘맛있는’ 선택 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.82703	오늘그릴	중랑구 신내동 410-69	스테이크/ 바베큐	[표 3-39] 3위 내용 참고
2	0.81808	Give Me Fever	마포구 연남동 566-60	이탈리안	[표 3-39] 2위 내용 참고
3	0.81592	페닌술라	중구 소공동 1	이탈리안	[표 3-39] 8위 내용 참고
4	0.81404	르포트13	강남구 신사동 525-13	브런치/버거 /샌드위치	너무 맛있어요
5	0.81340	레스트로	서초구 반포동 90-3	이탈리안	[표 3-39] 5위 내용 참고
6	0.81218	티톨로	강남구 청담동 79-8	이탈리안	[표 3-39] 1위 내용 참고
7	0.81201	꿈모아	용산구 용산동2가 26-10	프랑스 음식	친구들끼리 맛있는 프렌치 편하게 먹으러 가기 좋은 식당!
8	0.80914	고메트리	성동구 금호동4가 235	프랑스 음식	[표 3-39] 6위 내용 참고
9	0.80779	엘레브	마포구 연남동 240-29	프랑스 음식	시그니처인 덕 브레스트가 정말 맛있었어요
10	0.80763	클래식당	마포구 연남동 383-95	이탈리안	[표 3-39] 4위 내용 참고

본 연구에서 개발한 시맨틱 POI 검색 기술 중 선택 기반 POI 검색은 해시태그(hashtag)를 이용한 검색과 비교해 볼 수 있는데, 두 검색은 사용자의 선택을 기반으로 검색이 시작된다는 공통점이 있기는 하지만 결과를 도출하는 과정에 차이가 있다. 해시태그는 인스타그램, 트위터 등의 소셜 미디어에서 주로 사용되는 것으로, 해시태그(#) 뒤에 특정 단어를 함께 작성함으로써 해당 단어에 대한 글을 모아 볼 수 있는 기능을 제공하며, 사

용자들이 관심 있는 주제에 대한 내용을 쉽게 찾을 수 있도록 도와준다. 특정 해시태그를 클릭하면 해당 해시태그가 포함된 내용이 모두 출력되는데, 이러한 검색 형태는 토픽에 따라 다양하게 사용되는 단어의 의미, 동의어 및 유의어, 그리고 단어별 가중치를 고려한 검색을 수행하는 것이 불가능하기 때문에 본 연구에서 제안한 시맨틱 POI를 기반으로 하는 검색과는 다르다.

### 3.5.3.2 사용자 평가 결과 분석

선택 기반 POI 검색에 대한 사용자 만족도 평가 결과는 [표 3-41]과 같다. a~d의 모든 문항세트에 대한 만족도 평균이 5점 만점에 3.49점으로 산출되었다. 이는 [그림 3-8]의 만족도 평가 기준에 따라, 만족의 범위에 속하는 것으로 판단할 수 있다.

[표 3-41] 선택 기반 POI 검색에 대한 사용자 만족도 평가 결과

구분		a	b	c	d	평균
선택 기반 POI 검색 (문항 수: 20개)	평균	3.53	3.47	3.48	3.46	<b>3.49</b>
	최솟값	3.04	3.09	2.97	3.11	3.05
	최댓값	3.76	3.70	3.64	3.67	3.69
	분산	0.05	0.02	0.03	0.03	0.03

선택 기반 POI 검색에 대한 사용자 만족도 평가 결과에서 만족 정도에 따른 값의 분포는 [표 3-42]와 같다. 모든 문항세트에서 ‘만족’을 선택한 응답자 비율이 가장 높았으며, ‘매우 만족’ 및 ‘만족’을 선택한 응답자가 ‘매우 불만족’ 및 ‘불만족’을 선택한 응답자보다 약 3.7~4.5배 더 많았다. 이와 같은 결과를 통해 선택 기반 POI 검색 결과에 대한 서비스 이용자들의 만족 정도가 불만족 정도에 비해 더 높다는 것을 확인할 수 있었다.

[표 3-42] 선택 기반 POI 검색에 대한 사용자 만족도 결과 분포

단위: %(명)

구분		매우 만족	만족	보통	불만족	매우 불만족
선택 기반 POI 검색 (문항 수: 20개)	a	11.30(226)	44.85(897)	31.45(629)	10.70(214)	1.70(34)
	b	7.90(158)	43.85(877)	37.05(741)	9.50(190)	1.70(34)
	c	7.75(155)	47.30(946)	32.70(654)	9.95(199)	2.30(46)
	d	11.80(236)	38.15(763)	36.35(727)	11.10(222)	2.60(52)
평균(%)		9.69	43.54	34.39	10.31	2.08

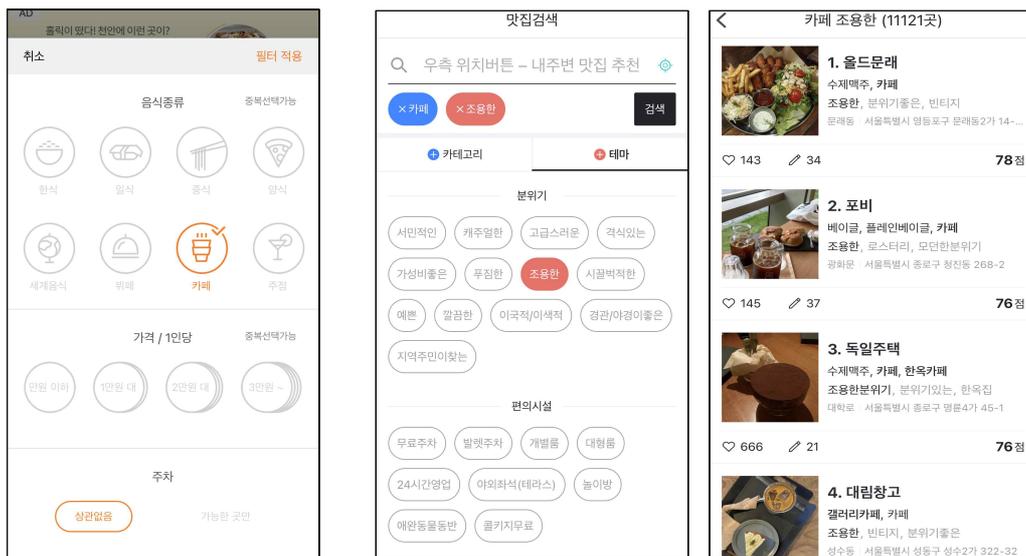
  

Detailed description of the pie chart: The pie chart illustrates the average user satisfaction levels for the selected-based POI search. The largest segment is '만족' (Satisfied) at 43.54%, followed by '보통' (Average) at 34.39%, '불만족' (Dissatisfied) at 10.31%, '매우 불만족' (Very Dissatisfied) at 9.69%, and the smallest segment is '매우 만족' (Very Satisfied) at 2.08%.

### 3.5.3.3 기존 서비스와의 비교 분석

본 연구에서 도출한 선택 기반 POI 검색 결과와 기존 유사 서비스들에서의 검색 결과를 비교 분석하였다. 기존의 국내외 POI 검색 서비스 중 선택 기반 검색 기능을 제공하는 플랫폼은 맛집 중심 플랫폼인 망고플레이트와 다이닝코드가 있다([표 1-1] 참조). 망고플레이트에서는 ‘필터’라는 기능을 통해 ‘음식 종류’, ‘가격’, ‘주차’ 항목에 대해 주어진 옵션을 선택해 POI를 검색할 수 있는 서비스를 제공하고 있는데, 항목별로 선택할 수 있는 옵션의 개수가 2~8개로 한정된다. 따라서 망고플레이트에서는 본 연구에서 선택 기반 POI 검색 키워드로 사용한 ‘파스타 맛있는’, ‘삼겹살 가성비 맛집’, ‘주말 오후 데이트’와 같은 검색이 불가능하다([그림 3-15 (a)]). 다이닝코드에서는 ‘테마’라는 기능을 이용해 ‘테마 맛집’, ‘방문 목적’, ‘분위기’, 그리고 ‘편의시설’의 항목별로 정해진 단어를 선택해 POI를 검색할 수 있는 서비스를 제공하고 있는데, 항목별로 선택 가능한 단어의 개수는

9~20개로, 망고플레이트에 비해서는 많은 편이지만 항목 개수가 4개로 제한적이다([그림 3-15 (b)]). 반면, 본 연구에서는 18개 토픽별로 각 20개의 키워드에 대한 선택 기반 POI 검색이 가능하다. 선택 기반 POI 검색에서 다이닝코드와의 비교를 위해 사용한 검색 키워드는 ‘조용한 카페’이다. 다이닝코드에서의 검색 결과인 [그림 3-15 (b)]를 보면, 검색 키워드와 관련된 태그가 포함된 POI들을, 다이닝코드 내부 알고리즘으로 계산한 점수가 높은 순서대로 반환해 준 것으로 추측된다. 따라서 검색 키워드와는 직접적인 관련이 없는 ‘수제맥주’를 메인으로 판매하는 ‘카페’ 관련 POI들이 상위권에 포함된 것을 볼 수 있다. 반면에 본 연구에서 ‘조용한’에 0.8, ‘카페’에 0.2의 가중치를 부여한 결과인 [표 3-43]을 보면, ‘조용한’에 높은 가중치를 부여했기에 리뷰에 검색 키워드와 동일하거나 유사한 키워드들이 포함된 POI들이 출력된 것을 알 수 있다. 즉, 본 연구에서 개발한 시맨틱 POI 검색 기술을 활용하면 다양한 토픽별 키워드에 대한 선택 기반 검색이 가능하며, 사용자가 직접 키워드별 가중치를 부여함으로써 개인의 선호와 상황을 반영한 검색을 할 수 있다.



(a) 망고플레이트

(b) 다이닝코드(일부)

[그림 3-15] 선택 기반 POI 검색 결과(맛집 중심 플랫폼)

[표 3-43] '조용한(w: 0.8)', '카페(w: 0.2)' 검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.90741	카페샤인	강남구 신사동 634-11	카페/ 디저트	압구정로테오와 압구정역사이에 있는 조용한 카페.
2	0.90062	카페잔디의 숲	용산구 이태원동 321-2	카페/ 디저트	공간이 제법 넓고 자리간격이넓고 조용해요~
3	0.86298	카페통인	종로구 통인동 132	카페/ 디저트	여유롭게, 조용하게 머물 수 있는 곳 :)
4	0.85839	산수화	용산구 한남동 68-27	카페/ 디저트	차 맛이 훌륭했고 조용하니 분위기가 좋아서 만족스러웠습니다
5	0.85710	보나르페 플라워카페	강남구 신사동 552-1	카페/ 디저트	카페 꽃향기도가득하고 조용한 매력이 있어요~
6	0.85519	핑크테이블	마포구 서교동 362-12	카페/ 디저트	홍대 내 위치한 스터디 카페! 조용하고 가격도 저렴하다.
7	0.85007	어반테이블	종로구 명륜4가 188-14	카페/ 디저트	공부하기 좋은 스터디카페 엄청 엄숙함
8	0.84845	낭만고양이	광진구 자양동 553-531	카페/ 디저트	조용하고 괜찮아요
9	0.84664	카페로쏘	종로구 팔판동 40-1	카페/ 디저트	삼청동 메인길뒷골목(삼청교회쪽)에 위치한 조용하고 아늑한 카페. 정신없지않아서 너무 좋음.
10	0.84405	연	종로구 삼청동 63-20	카페/ 디저트	좌식으로 딱딱 채워넣지않은 접이 좋았고 편안하게 쉬고 차한잔할만한 고즈넉한 곳~

### 3.5.4 유사 POI 검색에의 활용 결과

#### 3.5.4.1 실험 결과 분석

본 연구에서는 실험에 사용한 19,902개의 POI를 모두 벡터화하므로 코사인 유사도를 통해 POI 간 유사도를 계산할 수 있다. [표 3-44]~[표 3-51]은 19,902개의 POI 중 일부에 대해 해당 POI와 코사인 유사도가 가장 높은 상위 10개 POI를 추출한 결과이다. 먼저, [표 3-44]~[표 3-47]은 각각 ‘1201 커피카운터’, ‘고고초밥’, ‘129 라멘하우스’, ‘고대양꼬치집’과 유사한 상위 10개 POI를 도출한 결과이다. 각 표를 살펴보면, 기준이 되는 POI와 동일한 음식 종류에 해당하는 POI들이 출력된 것을 확인할 수 있다.

[표 3-44] ‘1201 커피카운터’와 유사한 POI 검색 결과(상위 10개)

순위	유사도	POI 명칭	주소	음식 종류	가격대
기준	-	1202 커피카운터	강남구 신사동 518-15	카페/디저트	1만원 미만
1	0.95048	greyt coffee	성동구 성수동1가 685-324	카페/디저트	1만원 미만
2	0.94707	오클리카피스탠드	성동구 성수동1가 225-2	카페/디저트	1만원 미만
3	0.94526	티라노	관악구 봉천동 1613-7	카페/디저트	1만원 미만
4	0.94521	프리퍼커피로스터스	서초구 서초동 1475-5	카페/디저트	1만원 미만
5	0.94474	쿠오레 에스프레소	서초구 서초동 1357-66	카페/디저트	1만원 미만
6	0.94359	카페톤	마포구 망원동 57-12	카페/디저트	1만원 미만
7	0.94233	수수커피	강남구 신사동 512-8	카페/디저트	1만원 미만
8	0.94225	커피폴리	마포구 서교동 378-19	카페/디저트	1만원 미만
9	0.94185	커피가게동경	마포구 망원동 410-1	카페/디저트	1만원 미만
10	0.94118	수수커피	종로구 청진동 246	카페/디저트	1만원 미만

[표 3-45] '고고초밥'과 유사한 POI 검색 결과(상위 10개)

순위	유사도	POI 명칭	주소	음식 종류	가격대
기준	-	고고초밥	강남구 역삼동 827-69	회/스시	1-2만원
1	0.98605	가네미	동작구 상도1동 265-1	회/스시	1-2만원
2	0.98564	대작	서대문구 창천동 46-7	회/스시	1-2만원
3	0.98538	미세기	서대문구 창천동 91-6	회/스시	1-2만원
4	0.98421	효자동초밥	종로구 통의동 59	회/스시	1-2만원
5	0.98371	건대스시	광진구 화양동 11-2	회/스시	1-2만원
6	0.98359	히든스시	종로구 명륜4가 53	회/스시	1-2만원
7	0.98348	유정상초밥	강동구 성내동 449-13	회/스시	1-2만원
8	0.98276	스시산	광진구 중곡동 93-50	회/스시	2-3만원
9	0.97936	린스시	서초구 서초동 1585-12	회/스시	1-2만원
10	0.97855	현초밥	종로구 명륜4가 166-1	회/스시	1-2만원

[표 3-46] '129 라멘하우스'와 유사한 POI 검색 결과(상위 10개)

순위	유사도	POI 명칭	주소	음식 종류	가격대
기준	-	129 라멘하우스	중구 신당동 374-27	라멘/소바/우동	1만원 미만
1	0.98632	켄비멘리키	영등포구 영등포동4가 434-5	라멘/소바/우동	1-2만원
2	0.98486	사루카메 라멘	마포구 연남동 566-67	라멘/소바/우동	1만원 미만
3	0.98448	켄비멘리키	마포구 상수동 313-3	라멘/소바/우동	1만원 미만
4	0.98411	잇텐바리	마포구 서교동 395-78	라멘/소바/우동	1만원 미만
5	0.98209	멘야산다이메	양천구 목동 917-9	라멘/소바/우동	1만원 미만
6	0.98105	라멘트릭	마포구 상수동 328-7	라멘/소바/우동	1만원 미만
7	0.98078	가야가야	동대문구 이문동 305-130	라멘/소바/우동	1만원 미만
8	0.98018	라멘모토	강남구 신사동 541-14	라멘/소바/우동	1만원 미만
9	0.97998	멘야히가시	마포구 연남동 387-6	라멘/소바/우동	1만원 미만
10	0.97985	지로우라멘	마포구 서교동 343-13	라멘/소바/우동	1만원 미만

[표 3-47] ‘고대양꼬치집’과 유사한 POI 검색 결과(상위 10개)

순위	유사도	POI 명칭	주소	음식 종류	가격대
기준	-	고대양꼬치집	성북구 안암동5가 146-4	기타 중식	-
1	0.97447	양치는아저씨1929	종로구 관철동 6-5	기타 중식	1-2만원
2	0.96794	제메이양꼬치	강남구 신사동 539-2	기타 중식	1-2만원
3	0.96296	한연양꼬치	중구 을지로6가 18-63	기타 중식	2-3만원
4	0.96164	논현양꼬치	강남구 논현동 138-22	기타 중식	1-2만원
5	0.95755	동북화과왕	종로구 창신동 463-1	기타 중식	1-2만원
6	0.95751	북경양꼬치	성북구 안암동5가 102-46	기타 중식	1-2만원
7	0.95380	신사양꼬치	강남구 신사동 515-10	기타 중식	2-3만원
8	0.95116	동방미식성	영등포구 영등포동6가 10	기타 중식	1-2만원
9	0.94968	매화반점	광진구 자양동 5-11	기타 중식	1-2만원
10	0.94925	만경양육관	관악구 남현동 602-1	기타 중식	-

[표 3-48]~[표 3-51]은 각각 ‘1986베이커’, ‘계타는날 부암동치킨’, ‘익선잡방’, ‘저스틴스테이크’라는 POI와 유사한 POI를 검색한 결과이다. 각 표의 상위 10개 목록을 보면, 기준 POI와 동일하거나 유사한 음식 종류에 해당하는 POI들이 도출된 것을 확인할 수 있다. 이러한 결과를 통해, 유사 POI 검색 결과는 POI의 음식 종류 속성과 관련이 있다는 것을 추측할 수 있다.

[표 3-48] '1986베이커'와 유사한 POI 검색 결과(상위 10개)

순위	유사도	POI 명칭	주소	음식 종류	가격대
기준	-	1986베이커	노원구 상계동 430-21	베이커리	1만원 미만
1	0.98039	바게트K	강남구 역삼동 723-30	베이커리	1만원 미만
2	0.97544	장티크	서초구 서초동 1641-13	카페/디저트	1만원 미만
3	0.97120	블랑제리코팡	마포구 망원동 455-17	베이커리	1만원 미만
4	0.96665	쿄베이커리	마포구 상수동 317-4	베이커리	1만원 미만
5	0.96404	폴앤폴리나	영등포구 여의도동 37	베이커리	1만원 미만
6	0.96365	베이커리구스토	광진구 광장동 582	베이커리	1만원 미만
7	0.96113	비에뜨반미	광진구 화양동 94-4	브런치/버거/ 샌드위치	1만원 미만
8	0.96072	블랑제르	성동구 성수동1가 14-46	베이커리	1만원 미만
9	0.96070	폴앤폴리나	종로구 내수동 74	베이커리	1만원 미만
10	0.96032	몽소	용산구 청과동2가 90-18	베이커리	1만원 미만

[표 3-49] '계타는날 부암동치킨'과 유사한 POI 검색 결과(상위 10개)

순위	유사도	POI 명칭	주소	음식 종류	가격대
기준	-	계타는날 부암동치킨	종로구 통인동 109-2	치킨/호프/뽕	1-2만원
1	0.94462	꼬꼬순이	마포구 동교동 147-6	닭/오리 요리	-
2	0.94355	쓰리로보스	성북구 동선동1가 85-84	일반 주점	1-2만원
3	0.94289	아웃닭	서대문구 대현동 56-141	닭/오리 요리	1만원 미만
4	0.94104	바른치킨	마포구 서교동 369-44	닭/오리 요리	1-2만원
5	0.93949	BHC치킨	종로구 관철동 7-4	닭/오리 요리	1-2만원
6	0.93558	맛닭꼬	노원구 공릉동 633-17	닭/오리 요리	1-2만원
7	0.93497	미친닭	종로구 관철동 44-1	치킨/호프/뽕	1-2만원
8	0.93464	한국통닭	종로구 낙원동 177	닭/오리 요리	1만원 미만
9	0.93421	치킨처럼	마포구 아현동 275-1	닭/오리 요리	1-2만원
10	0.93381	아웃닭	광진구 화양동 46-55	닭/오리 요리	1-2만원

[표 3-50] '익선잡방'과 유사한 POI 검색 결과(상위 10개)

순위	유사도	POI 명칭	주소	음식 종류	가격대
기준	-	익선잡방	종로구 익선동 166-80	이탈리안	1-2만원
1	0.95987	사루비아	강남구 신사동 542-3	이탈리안	1-2만원
2	0.95788	비스트로폴포	서초구 방배동 13-8	이탈리안	2-3만원
3	0.95536	소년서커스	종로구 청진동 246	이탈리안	2-3만원
4	0.95368	다이닝텐트	강남구 삼성동 114-44	이탈리안	2-3만원
5	0.95183	꿈앤편	용산구 이태원동 131-8	스테이크/바베큐	1-2만원
6	0.95131	더466키친	광진구 구의동 213-1	브런치/머거/ 샌드위치	1-2만원
7	0.95114	도미닉	용산구 한남동 683-133	스테이크/바베큐	4만원 이상
8	0.95081	아피아	용산구 이태원동 184-5	이탈리안	3-4만원
9	0.95073	더고져스키친	강남구 신사동 545-1	이탈리안	2-3만원
10	0.94993	아르모니움	용산구 한남동 657-37	이탈리안	3-4만원

[표 3-51] '저스틴스테이크'와 유사한 POI 검색 결과(상위 10개)

순위	유사도	POI 명칭	주소	음식 종류	가격대
기준	-	저스틴스테이크	용산구 이태원동 129-5	스테이크/바베큐	1-2만원
1	0.99181	에이컷스테이크	강동구 성내동 468-3	패밀리 레스토랑	2-3만원
2	0.99072	라구	용산구 용산동2가 45-4	스테이크/바베큐	2-3만원
3	0.98944	2046팬스테이크	양천구 목동 917-9	스테이크/바베큐	1-2만원
4	0.98582	도쿄스테이크	마포구 서교동 358-9	스테이크/바베큐	1-2만원
5	0.98535	파스치노팬스테이크	마포구 서교동 358-8	스테이크/바베큐	1-2만원
6	0.98382	더스테이크하우스	서초구 양재동 89-2	스테이크/바베큐	4만원 이상
7	0.97942	스테이크레이브	종로구 청진동 70	스테이크/바베큐	1-2만원
8	0.97864	부첼리하우스	용산구 한남동 1-90	스테이크/바베큐	4만원 이상
9	0.97635	붓처스컷	중구 태평로1가 84	스테이크/바베큐	4만원 이상
10	0.96751	장스테이크하우스	강남구 신사동 644-18	스테이크/바베큐	3-4만원

### 3.5.4.2 사용자 평가 결과 분석

유사 POI 검색에 대한 사용자 만족도 평가 결과는 [표 3-52]와 같다. a~d의 모든 문항세트에 대한 만족도 평균이 3.46점으로 산출되었으며, 이는 [그림 3-8]의 평가 기준에 따라 만족의 범위에 속하는 것으로 판단할 수 있다.

[표 3-52] 유사 POI 검색에 대한 사용자 만족도 평가 결과

구분		a	b	c	d	평균
유사 POI 검색 (문항 수: 10개)	평균	3.46	3.47	3.39	3.52	<b>3.46</b>
	최솟값	3.36	3.39	3.17	3.44	3.34
	최댓값	3.60	3.64	3.51	3.67	3.61
	분산	0.01	0.01	0.01	0.01	0.01

유사 POI 검색에 대한 사용자 만족도 평가 결과에서 만족 정도에 따른 값의 분포는 [표 3-53]과 같다. 모든 문항세트에서 ‘만족’을 선택한 응답자 비율이 가장 높았으며, ‘매우 만족’ 및 ‘만족’을 선택한 응답자가 ‘매우 불만족’ 및 ‘불만족’을 선택한 응답자보다 약 3~5배 더 많았다. 이러한 결과를 통해 유사 POI 검색 결과에 대한 서비스 이용자들의 만족 정도가 불만족 정도에 비해 더 높다는 것을 알 수 있다.

[표 3-53] 유사 POI 검색에 대한 사용자 만족도 결과 분포

단위: %(명)

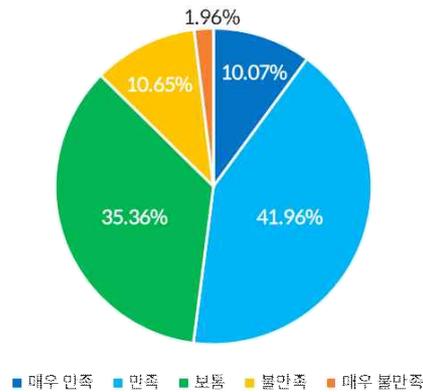
구분		매우 만족	만족	보통	불만족	매우 불만족
유사 POI 검색 (문항 수: 10개)	a	10.00(100)	42.00(420)	32.70(327)	15.00(150)	0.30(3)
	b	8.90(89)	40.80(408)	40.30(403)	8.50(85)	1.50(15)
	c	9.50(95)	39.00(390)	35.40(354)	12.70(127)	3.40(34)
	d	13.10(131)	38.60(386)	37.10(371)	9.90(99)	1.30(13)
평균(%)		10.38	40.10	36.38	11.53	1.63

Detailed description of the pie chart: The pie chart illustrates the percentage distribution of user satisfaction levels. The largest segment is '만족' (Satisfied) at 40.10%, followed by '보통' (Average) at 36.38%, '불만족' (Dissatisfied) at 11.53%, '매우 만족' (Very Satisfied) at 10.38%, and '매우 불만족' (Very Dissatisfied) at 1.63%. A legend on the right side of the chart identifies each category with a corresponding color: dark blue for '매우 만족', light blue for '만족', green for '보통', yellow for '불만족', and orange for '매우 불만족'.

종합하자면, 자연어 및 선택 기반 POI 검색과 유사 POI 검색 결과에 대한 만족도 평균은 각각 3.48점, 3.49점, 3.46점으로 도출되었으며, 이는 모두 만족 범위에 속하는 값으로 해석할 수 있다. 또한, 세 가지 시맨틱 POI 검색 결과에 대한 종합적인 사용자 만족도 평균은 5점 만점에 3.48점으로 산출되므로 본 연구에서 개발한 자연어 및 선택 기반 POI 검색과 유사 POI 검색 결과는 종합적으로 만족 범위에 속하는 것으로 판단할 수 있다. 사용자 평가 결과를 만족 정도별로 구분하여 분석한 결과에서는 세 가지 시맨틱 POI 검색의 모든 문항세트에서 ‘만족’을 선택한 응답자 비율이 가장 높았으며, 전체적으로 ‘매우 만족’ 및 ‘만족’을 선택한 응답자가 ‘매우 불만족’ 및 ‘불만족’을 선택한 응답자보다 약 3~5.2배 더 많았다. [그림 3-16]은 세 가지 시맨틱 POI 검색에 대한 사용자의 만족 정도별 평균값을 종합적으로 나타낸 것이다. 결과적으로, ‘매우 만족’ 및 ‘만족’을 선택한 응답자 비율이 52.03%로, ‘매우 불만족’ 및 ‘불만족’을 선택한 응답자 비율인 12.61%보다 약 4.1배 더 많은 것을 알 수 있다. 모든 문항세트에 대한 사용

자 평가 결과에서 만족 정도가 불만족 정도에 비해 높게 도출됨에 따라 본 연구에서 개발한 시맨틱 POI 검색 기술의 품질을 확인할 수 있었다.



[그림 3-16] 세 가지 시맨틱 POI 검색에 대한 사용자의 만족 정도별 평균값 분포 차트

### 3.5.4.3 기존 서비스와의 비교 분석

국내외 POI 검색 서비스 중 유사 POI 검색 기능을 제공하는 플랫폼은 지도 중심 플랫폼인 네이버 지도 및 구글 지도와 맛집 중심 플랫폼인 망고플레이트가 있는데([표 1-1] 참조), 네이버 지도의 ‘함께 많이 찾아본 곳’, 구글 지도의 ‘함께 검색한 장소’, 그리고 망고플레이트의 ‘주변 인기 식당’은 본 연구에서 제안한 POI 간 시맨틱 정보에 기반한 유사성과는 차이가 있다. 먼저, 본 연구에서 도출된 유사 POI 검색 결과([표 3-54])와 지도 중심 플랫폼에서 도출된 결과([그림 3-17])를 비교해 보았다. 결과 비교를 위해 사용한 POI는 마포구 연남동에 위치한 ‘히메지’이다. 네이버 지도의 ‘함께 많이 찾아본 곳’의 결과([그림 3-17 (a)])를 보면, ‘퓨전음식’, ‘맥주, 호프’, ‘샌드위치’ 등 다양한 카테고리의 POI들이 출력된 것을 볼 수 있다. 네이버와 유사한 기능인 구글 지도의 ‘함께 검색한 장소’ 결과([그림 3-17 (b)])에서도 ‘중식’ 및 ‘한식’에 해당하는 POI들이 결과로 도출되었다. 맛집 중심 플랫폼인 망고플레이트의 경우에는 ‘주변 인기 식당’으로 ‘카페/디저

트’, ‘베이커리’ 등에 해당하는 POI들이 출력되었다([그림 3-18]). 반면 본 연구에서 ‘히메지’와 유사한 상위 10개 POI를 도출한 결과([표 3-54])를 보면, ‘히메지’와 동일한 음식 종류인 ‘돈부리/일본 카레/벤토’, 또는 ‘정통 일식/일반 일식’ 및 ‘라멘/소바/우동’에 해당하는 POI들이 출력된 것을 알 수 있다. 즉, 본 연구에서 제안한 유사 POI 검색 기술은 기존 POI 검색 플랫폼에서는 제공하지 않는 기능을 제공하며, 이는 기존 서비스들과 상호 보완적으로 사용 가능하다.



(a) 네이버 지도(일부)

(b) 구글 지도

[그림 3-17] 유사 POI 검색 결과(지도 중심 플랫폼)



[그림 3-18] 유사 POI 검색 결과(망고플레이트)

[표 3-54] '히메지'와 유사한 POI 검색 결과(상위 10개)

순위	유사도	POI 명칭	주소	음식 종류	가격대
기준	-	히메지	마포구 연남동 227-15	돈부리/일본 카레/벤토	1만원 미만
1	0.99286	노다메	광진구 화양동 2-10	정통 일식/일반 일식	1만원 미만
2	0.99249	히메지	서대문구 연희동 129-1	돈부리/일본 카레/벤토	1만원 미만
3	0.99223	라운드어바웃	용산구 한남동 739-4	돈부리/일본 카레/벤토	1-2만원
4	0.99197	카타코토	마포구 서교동 463-22	돈부리/일본 카레/벤토	1만원 미만
5	0.99044	소소카레	서대문구 남가좌동 7-5	돈부리/일본 카레/벤토	1만원 미만
6	0.98976	키이로메시아	송파구 송파동 8-9	돈부리/일본 카레/벤토	1만원 미만
7	0.98881	코나야	구로구 신도림동 692	돈부리/일본 카레/벤토	1만원 미만
8	0.98877	거북이의주방	마포구 노고산동 31-50	돈부리/일본 카레/벤토	1만원 미만
9	0.98799	코나야	중구 을지로6가 17-2	라멘/소바/우동	1만원 미만
10	0.98795	고씨네	동작구 신대방동 395-69	돈부리/일본 카레/벤토	1-2만원

## 4. 결 론

스마트 기기의 대중화와 측위 기술의 발달에 따라 POI 검색은 점차 맞춤형 서비스로 진화하고 있다. 이에 본 연구에서는 사용자들의 다양한 선호와 상황을 반영할 수 있는 시맨틱 POI 검색 기술을 개발하였다. 본 연구에서는 사용자가 작성한 리뷰 텍스트 데이터를 활용하여 시맨틱 POI를 생성하는 프로세스를 제안하고, 시맨틱 POI를 검색에 활용할 수 있도록 하는 자연어 기반 POI 검색, 선택 기반 POI 검색, 그리고 유사 POI 검색을 위한 활용 구조를 설계하였다.

본 연구에서 사용한 방법론은 데이터 수집 및 전처리, LDA 기반 토픽 모델링, 토픽별 키워드 기반 POI 임베딩, 그리고 시맨틱 POI 검색을 위한 활용 구조로 구성된다. 그리고 시맨틱 POI 검색에의 활용 결과에 대한 품질 검증을 위해 설문 조사 방법을 사용하였다. 구체적으로, 사용자 리뷰 데이터는 망고플레이트에서 한국어로 작성된 텍스트 데이터를 웹 스크래핑 방식으로 수집하였으며, 전처리 단계에서 한국어의 특성을 반영해 형태소 단위까지 분석해 사용하였다. 전처리를 마친 텍스트 데이터로부터 POI의 시맨틱 정보를 추출하기 위해 LDA 기반 토픽 모델링을 실시하였으며, 이때 복잡도 및 토픽 일관성 지수를 근거로 POI와 관련된 18개의 토픽을 선정하고, 토픽별 키워드를 추출하였다. 다음으로, 토픽 기반 키워드 임베딩 기법인 TWE-1을 활용해, 본 연구에서 최적 하이퍼 파라미터 조합으로 선정된 TF-IDF 가중치, 윈도우 크기 5, 단어 최소 출현 횟수 5, 벡터 크기 250을 기준으로 총 19,902개의 시맨틱 POI를 생성하였다. 시맨틱 POI를 검색에 활용하기 위하여 자연어 검색어에 대한 결과를 제공하는 자연어 기반 POI 검색, 사용자가 토픽과 토픽별 키워드를 선택하는 선택 기반 POI 검색, 그리고 특정 POI와 유사한 POI를 검색할 수 있는 유사 POI 검색을 위한 활용 구조를 설계하였다. 그리고 세 가지 시맨틱 POI 검색에의 활용 결과에 대한 품질을 검증하기 위하여 서울에 거주하는 20대 이상 50대 이하 남녀 400명을 대상으로 설문 조사를 시행하였다. 설문 조사 결과

종합적인 사용자 만족도 평균이 5점 만점에 3.48점으로 산출되었으며, 이는 본 연구에서 제시한 만족도 평가 기준에 의해 만족의 범위에 속하는 것으로 판단할 수 있다. 또한, 사용자 만족도 평가를 만족 정도별로 구분하여 분석한 결과에서는 세 가지 시맨틱 POI 검색에 대한 모든 설문 문항에서 ‘만족’을 선택한 응답 비율이 가장 높았다. 평균적으로 ‘매우 만족’ 및 ‘만족’을 선택한 응답자가 ‘매우 불만족’ 및 ‘불만족’을 선택한 응답자보다 약 4.1배 더 많은 것으로 도출됨에 따라 시맨틱 POI 검색 기술의 품질을 확인할 수 있었다. 마지막으로, 본 연구에서 도출한 시맨틱 POI 검색 결과와 기존 POI 검색 서비스들과의 검색 결과 비교 분석을 통해 본 연구의 차별성을 확인하였다.

본 연구의 의의는 첫째, 리뷰 텍스트 데이터를 활용해 시맨틱 POI를 생성하는 프로세스를 제안한 것이다. 시맨틱 POI를 생성하기 위해 정보 검색 및 자연어 처리 분야에서 사용되는 토픽 모델링 기법과 토픽별 키워드 임베딩 기법을 POI 임베딩에 적용하는 방법을 제안하였다. 둘째, 시맨틱 POI를 검색에 활용할 수 있도록 하는 활용 구조를 설계한 것이다. 시맨틱 POI 검색은 검색어와 POI 간 관련성에 기반한 자연어 및 선택 기반 POI 검색(직접검색)과 POI 간 유사성에 기반한 유사 POI 검색(추천검색)을 포함한다.

본 연구의 결과인 시맨틱 POI를 검색에 활용함으로써 얻을 수 있는 이점은 첫째, 사용자가 생각하는 검색어의 중요도에 따라 다양한 가중치를 부여할 수 있으므로 같은 검색어라도 가중치에 따라 다른 결과가 출력됨으로써 사용자 맞춤형 POI 검색이 가능하다. 둘째, 검색어와 정확히 일치하는 키워드가 없어도 해당 검색어의 동의어 및 유의어를 기반으로 한 POI 검색이 가능하다. 셋째, 같은 키워드라도 토픽에 따라 다르게 사용되는 키워드의 다양한 의미를 POI 검색에 활용할 수 있다. 따라서 본 연구에서 개발한 시맨틱 POI 검색 기술을 활용하면 사용자들의 다양한 선호와 상황을 반영한 검색이 가능하며, 이는 기존의 검색 시스템과 상호 보완적으로 사용될 수 있다.

본 연구가 가지는 한계점은 POI에 대해 작성된 사용자 리뷰의 개수가

적을수록 시맨틱 POI를 구성하는 키워드의 영향력이 강해지기 때문에 검색 시 결과가 편향될 수 있다는 것이다. 그러나 이 같은 문제는 리뷰 데이터가 시간이 흐름에 따라 점차 축적되는 경향이 있기에 주기적인 모델 업데이트를 통해 해결 가능할 것으로 생각된다. 학습 기반 모델은 데이터가 많이 축적될수록 견고해지는 특징이 있으므로 데이터를 기반으로 모델을 업데이트함으로써 더욱 정교한 검색 결과를 제공할 수 있을 것으로 기대된다. 이렇게 새로운 리뷰 데이터나 신조어 등이 추가될 때 이를 반영하기 위한 LDA 및 POI 임베딩 모델의 재학습이 필요한데, 새로 추가된 데이터에 대해서만 학습이 이루어지기 때문에 서비스 측면에서 실용성을 확보할 수 있다. LDA는 잘 정의된 생성 모형이기에 새로운 데이터에 대해 쉽게 일반화될 수 있으며, POI 임베딩 시에는 본 연구에서 제안한 방법에 따라 학습한 토픽별 키워드 벡터를 초깃값으로 이용해 모델을 훈련하는 전이학습(transfer learning) 방법을 사용할 수 있다. 즉, 새로운 데이터가 추가될 때 완전히 새로운 모델링을 하는 것이 아니라 기존 모델을 기반으로 새 데이터만 학습하는 방법을 통해 효율적인 모델 업데이트가 가능하다.

본 연구에서 생성한 시맨틱 POI는 기존 POI 추천검색 연구에서 입력값으로 활용될 수 있다. 특히, RNN, LSTM 등의 딥러닝 기반 모델에서는 입력 데이터의 특징에 대한 효과적인 벡터 표현을 확보하는 것이 모델의 성능을 높이는 데 중요하므로 본 연구에서 생성한 POI 벡터를 입력값으로 활용하면 효율적인 모델 학습이 가능할 것으로 보인다. 이와 관련하여 향후에는 본 연구 결과로 생성한 시맨틱 POI를 기반으로 위치, 시간, 사용자 등의 정보를 함께 활용함으로써 POI 직접검색 및 추천검색을 위한 딥러닝 모델의 성능을 향상시키는 방법에 관한 연구가 수행되어야 할 것으로 생각된다. 또한, NLP 분야의 성장을 발판 삼아 조사와 어미가 발달한 한국어의 특성을 잘 반영하고, POI 리뷰에 특화된 NLP 기법을 활용해 시맨틱 POI의 품질을 향상시킬 수 있는 연구도 함께 수행되어야 할 것이다.

## 참 고 문 헌

- Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*.
- Battisti, F. D., Ferrara, A., & Salini, S. (2015). A decade of research in statistics: A topic model approach. *Scientometrics*, *103*(2), 413-433.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993-1022.
- Cai, L., Xu, J., Liu, J., & Pei, T. (2018). Integrating spatial and temporal contexts into a factorization model for POI recommendation. *International Journal of Geographical Information Science*, *32*(3), 524-546.
- Caselles-Dupré, H., Lesaint, F., & Royo-Letelier, J. (2018). Word2vec applied to recommendation: Hyperparameters matter. In *Proceedings of the 12th ACM Conference on Recommender Systems*.

- Chang, B., Park, Y., Park, D., Kim, S., & Kang, J. (2018). Content-aware hierarchical point-of-Interest embedding model for successive POI recommendation. In *Proceedings of the International Joint Conferences on Artificial Intelligence*.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Proceedings of the Advances in Neural Information Processing Systems*.
- Chen, W., Wang, J., Zhang, Y., Yan, H., & Li, X. (2015). User based aggregation for biterm topic model. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*.
- Cheng, C., Yang, H., King, I., & Lyu, M. R. (2012). Fused matrix factorization with geographical and social influence in location-based social networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Christie, J., Klein, R. M., & Watters, C. (2004). A comparison of simple hierarchy and grid metaphors for option layouts on small-size screens. *International Journal of Human-Computer Studies*, 60(5-6), 564-584.
- Chuang, H., Chang, C., Kao, T., Cheng, C., Huang, Y., & Cheong, K. (2016). Enabling maps/location searches on mobile devices: Constructing a POI database via focused crawling and information extraction. *International Journal of Geographical Information Science*, 30(7), 1405-1425.

- Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391-407.
- Deutskens, E., De Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters*, 15(1), 21-36.
- Ding, R., & Chen, Z. (2018). RecNet: a deep neural network for personalized POI recommendation in location-based social networks. *International Journal of Geographical Information Science*, 32(8), 1631-1648.
- Feng, S., Cong, G., An, B., & Chee, Y. M. (2017). POI2Vec: Geographical latent representation for predicting future visitors. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Fu, X., Wang, T., Li, J., Yu, C., & Liu, W. (2016). Improving distributed word representation and topic model by word-topic mixture model. In *Proceedings of the Asian Conference on Machine Learning*.
- Gao, S., Janowicz, K., & Couclelis, H. (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, 21(3), 446-467.

- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. In *Proceedings of the National Academy of Sciences*.
- Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1–30.
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2–3), 146–162.
- He, J., Qi, J., & Ramamohanarao, K. (2019). A joint context-aware embedding for trip recommendations. In *Proceedings of the 2019 IEEE 35th International Conference on Data Engineering*.
- Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, 45(4), 549–559.
- Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM Conference on Research and Development in Information Retrieval*.
- Jacobi, C., Van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1), 89–106.
- Jin, M., Luo, X., Zhu, H., & Zhuo, H. H. (2018). Combining deep learning and topic modeling for review understanding in context-aware recommendation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

- Kefalas, P., & Manolopoulos, Y. (2017). A time-aware spatio-textual recommender system. *Expert Systems with Applications*, 78, 396–406.
- Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37.
- Krasnashchok, K., & Jouili, S. (2018). Improving topic quality by promoting named entities in topic modeling. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*.
- Lai, S., Liu, K., He, S., & Zhao, J. (2016). How to generate a good word embedding. *IEEE Intelligent Systems*, 31(6), 5–14.
- Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3, 211–225.
- Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., & Rui, Y. (2014). GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Liang, D., Altosaar, J., Charlin, L., & Blei, D. M. (2016). Factorization meets the item embedding: Regularizing matrix factorization with item co-occurrence. In *Proceedings of the 10th ACM Conference on Recommender Systems*.

- Liu, B., Xiong, H., Papadimitriou, S., Fu, Y., & Yao, Z. (2014). A general geographical probabilistic factor model for point of interest recommendation. *IEEE Transactions on Knowledge Data Engineering*, 27(5), 1167–1179.
- Liu, Q., Huang, H., Gao, Y., Wei, X., & Geng, R. (2017). Leveraging pattern associations for word embedding models. In *Proceedings of the International Conference on Database Systems for Advanced Applications*.
- Liu, X., Andris, C., & Rahimi, S. (2019). Place niche and its regional variability: Measuring spatial context patterns for points of interest with representation learning. *Computers, Environment Urban Systems*, 75, 146–160.
- Liu, X., Liu, Y., & Li, X. (2016). Exploring the context of locations for personalized location recommendations. In *Proceedings of the International Joint Conferences on Artificial Intelligence*.
- Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., & Shi, L. (2015a). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105(3), 512–530.
- Liu, Y., Liu, Z., Chua, T.-S., & Sun, M. (2015). Topical word embeddings. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- Lu, Y.-S., & Huang, J.-L. (2020). GLR: A graph-based latent representation model for successive POI recommendation. *Future Generation Computer Systems*, 102, 230–244.

- Lu, Y.-S., Shih, W.-Y., Gau, H.-Y., Chung, K.-C., & Huang, J.-L. (2019). On successive point-of-interest recommendation. *World Wide Web*, 22(3), 1151-1173.
- Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov), 2579-2605.
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., & Häussler, T. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods Measures*, 12(2-3), 93-118.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the Advances in Neural Information Processing Systems*.
- Newman, D., Bonilla, E. V., & Buntine, W. (2011). Improving topic coherence with regularized topic models. In *Proceedings of the Advances in Neural Information Processing Systems*.
- Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. In *Proceedings of the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

- Nunnally, J. C. (1994). *Psychometric theory 3E*: Tata McGraw-hill education.
- O'callaghan, D., Greene, D., Carthy, J., & Cunningham, P. (2015). An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, *42*(13), 5645–5657.
- Qiao, Y., Luo, X., Li, C., Tian, H., & Ma, J. (2020). Heterogeneous graph-based joint representation learning for users and POIs in location-based social network. *Information Processing Management*, *57*(2), 102151.
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*.
- Schnabel, T., Labutov, I., Mimno, D., & Joachims, T. (2015). Evaluation methods for unsupervised word embeddings. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*.
- Shneiderman, B., & Plaisant, C. (2010). *Designing the user interface: strategies for effective human-computer interaction*: Pearson Education India.
- Shoji, Y., Takahashi, K., Dürst, M. J., Yamamoto, Y., & Ohshima, H. (2018). Location2Vec: Generating distributed representation of location by using geo-tagged microblog posts. In *Proceedings of the International Conference on Social Informatics*.

- Song, Y., Pan, S., Liu, S., Zhou, M. X., & Qian, W. (2009). Topic and keyword re-ranking for LDA-based topic modeling. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*.
- Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). Exploring topic coherence over many models and many topics. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.
- Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of Latent Semantic Analysis*, 427(7), 424-440.
- Tutubalina, E., & Nikolenko, S. (2017). Demographic prediction based on user reviews about medications. *Computación y Sistemas*, 21(2), 227-241.
- Viegas, F., Canuto, S., Gomes, C., Luiz, W., Rosa, T., Ribas, S., Rocha, L., & Gonçalves, M. A. (2019). CluWords: exploiting semantic word clustering representation for enhanced topic modeling. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*.
- Wallach, H. M. (2008). *Structured topic models for language*. University of Cambridge Cambridge, UK,
- Wallach, H. M., Mimno, D. M., & McCallum, A. (2009). Rethinking LDA: Why priors matter. In *Proceedings of the Advances in Neural Information Processing Systems*.

- Wang, F., Qu, Y., Zheng, L., Lu, C.-T., & Philip, S. Y. (2017). Deep and broad learning on content-aware POI recommendation. In *Proceedings of the 2017 IEEE 3rd International Conference on Collaboration and Internet Computing*.
- Wang, M.-F., Lu, Y.-S., & Huang, J.-L. (2019). SPENT: A successive POI recommendation method using similarity-based POI embedding and recurrent neural network with temporal influence. In *Proceedings of the 2019 IEEE International Conference on Big Data and Smart Computing*.
- Xu, Y., Yin, Y., & Yin, J. (2017). Tackling topic general words in topic modeling. *Engineering Applications of Artificial Intelligence*, *62*, 124-133.
- Yan, B., Janowicz, K., Mai, G., & Gao, S. (2017). From ITDL to Place2Vec - Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.
- Yang, J., & Eickhoff, C. (2018). Unsupervised learning of parsimonious general-purpose embeddings for user and location modeling. *ACM Transactions on Information Systems*, *36*(3), 1-33.
- Yao, D., Zhang, C., Huang, J., & Bi, J. (2017). Serm: A recurrent model for next location prediction in semantic trajectories. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*.

- Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K. (2017). Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science*, 31(4), 825-848.
- Yin, H., Wang, W., Wang, H., Chen, L., & Zhou, X. (2017). Spatial-aware hierarchical collaborative deep learning for POI recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 29(11), 2537-2551.
- Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Zhai, W., Bai, X., Shi, Y., Han, Y., Peng, Z.-R., & Gu, C. (2019). Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Computers, Environment Urban Systems*, 74, 1-12.
- Zhang, H.-P., Mo, Q., & Huang, H. (2010). Structured POI data extraction from internet news. In *Proceedings of the 2010 4th International Universal Communication Symposium*.
- Zhang, X., Hu, X., & Li, Z. (2015). Learning geographical hierarchy features for social image location prediction. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*.

- Zhao, S., Zhao, T., King, I., & Lyu, M. R. (2017). Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In *Proceedings of the 26th International Conference on World Wide Web Companion*.
- Zhao, W. X., Fan, F., Wen, J.-R., & Chang, E. Y. (2018). Joint representation learning for location-based social networks with multi-grained sequential contexts. *ACM Transactions on Knowledge Discovery from Data*, 12(2), 1-21.
- Zhong, C., Huang, X., Arisona, S. M., Schmitt, G., & Batty, M. (2014). Inferring building functions from a probabilistic model using public transportation data. *Computers, Environment Urban Systems*, 48, 124-137.
- Zhu, X., & Zhou, C. (2009). POI inquiries and data update based on LBS. In *Proceedings of the 2009 International Symposium on Information Engineering and Electronic Commerce*.
- 강형석, 양장훈. (2018). 한국어 단어 임베딩 모델의 평가에 적합한 유추 검사 세트. *한국디지털콘텐츠학회 논문지*, 19(10), 1999-2008.
- 강형석, 양장훈. (2019). 한국어 단어 임베딩을 위한 Word2vec 모델의 최적화. *한국디지털콘텐츠학회 논문지*, 20(4), 825-833.
- 김권현, 유동주, 김형준, 김청택. (2015). 설문지의 길이가 응답의 질에 미치는 영향. *조사연구*, 16(1), 1-48.
- 김명지, 윤수경, 최준호. (2015). 스마트워치의 효율적인 탐색을 위한 메뉴 구조 연구: 메뉴 유형과 색상단서를 중심으로. *디지털디자인학 연구*, 15(3), 395-406.

- 김용대, 정구환. (2016). 토픽모형을 이용한 빅데이터 기반마이크로 세그멘테이션 방법론 연구. *통계연구, 특별호*, 17-34.
- 김은희, 표신지, 김문철. (2012). LDA 기반 은닉 토픽 추론을 이용한 TV 프로그램 자동 추천. *방송공학회논문지, 17(2)*, 270-283.
- 김정수, 이석준. (2016). 취업준비생 토픽 분석을 통한 취업난 원인의 재탐색. *35(1)*, 85-116.
- 김지은, 백순근. (2016). 텍스트 빅데이터 분석 기법을 활용한 대학구조개혁 평가의 쟁점 분석. *아시아교육연구, 17(3)*, 409-436.
- 안주영, 안규빈, 송민. (2016). 텍스트 마이닝을 이용한 매체별 에볼라 주제 분석: 바이오 분야 연구논문과 뉴스 텍스트 데이터를 이용하여. *한국문헌정보학회지, 50(2)*, 289-307.
- 유예림. (2017). *빅데이터 분석 기법을 활용한 2015 개정 교육과정 정책에 대한 언론보도 분석* (박사학위논문, 서울대학교 대학원).
- 이건일, 이의현, 이종혁. (2017). Sequence-to-sequence 기반 한국어 형태소 분석 및 품사 태깅. *정보과학회논문지, 44(1)*, 57-62.
- 정홍인. (2005). 비교 연구를 통한 그리드 메뉴의 효율성 평가. *디자인학연구, 18(3)*, 191-198.
- 조현수, 이상구. (2017). FastText 를 적용한 한국어 단어 임베딩. *한국소프트웨어 종합학술대회 학술발표논문집*.
- 최병서, 이익훈, 이상구. (2020). 신조어 및 띄어쓰기 오류에 강인한 시퀀스-투-시퀀스 기반 한국어 형태소 분석기. *정보과학회논문지, 47(1)*, 70-77.

- 최상혁, 설진석, 이상구. (2016). 한국어에 적합한 단어 임베딩 모델 및 파라미터 튜닝에 관한 연구. *제 28회 한글 및 한국어 정보처리 학술대회논문집*.
- 최용석, 이공주. (2020). 트랜스포머와 BERT 로 구현한 한국어 형태소 분석기의 성능 분석. *정보과학회논문지*, 47(8), 730-741.
- 한혁, 금현섭. (2017). 만족도 측정 방법의 비교 가능성 연구: 행정서비스 만족도의 4, 5, 11점 리커트형 문항을 중심으로. *조사연구*, 18(1), 61-96.
- 닐슨코리아클릭. (2020-02-20). 네이버 VS 카카오: 지도앱 서비스 비교, [http://www.koreanclick.com/insights/newsletter\\_view.html?code=topic&id=563&page=1&utm\\_source=board&utm\\_medium=board&utm\\_campaign=topic&utm\\_content=20200220](http://www.koreanclick.com/insights/newsletter_view.html?code=topic&id=563&page=1&utm_source=board&utm_medium=board&utm_campaign=topic&utm_content=20200220) (접속일: 2020년 10월 5일).

## 부 록

### A.1 설문 조사 개요

사용자 만족도에 대한 설문 조사는 모바일 전문 설문 기관인 오픈서베이 플랫폼을 이용해 총 4회로 나누어 실시하였으며, 설문 1회당 50개 문항에 대한 100명의 응답을 수집하였다. 각 문항세트는 시맨틱 POI 검색에의 활용 결과에 대한 종합적인 평가가 이루어질 수 있도록 검색 종류별로 동일한 비율의 문항으로 구성하였다([표 2-4] 참조). 즉, 자연어 및 선택 기반 POI 검색에서는 키워드 개수(1~4개)별로 5개씩 각 20개 문항으로 구성하였으며, 유사 POI 검색에서는 10개 문항을 생성하였다. 따라서 1명의 응답자는 총 50개 문항을 읽고 답해야 한다. 4개의 문항세트는 검색 종류별로 모두 같은 비율로 구성되었으나, 세부 문항 내용은 각각 다르므로 결과적으로 총 200개의 서로 다른 문항에 대한 응답을 수집한 것이다. 문항세트 4개(a~d)에 대한 각 설문 내용은 <부록 A.2>~<부록A.5>에 작성하였다.

### A.2 문항세트 a 설문 조사 내용

문항세트 a에서 자연어 기반 POI 검색 결과에 대한 문항은 키워드 개수(1~4개)별로 5개씩이므로 총 20개 문항을 포함한다. 키워드 개수에 따른 검색어는 [표 A.2-1]~[표 A.2-4]와 같다. 실제 오픈서베이를 통해 실시한 설문 화면 중 일부는 [그림 A.2-1]~[그림 A.2-4]와 같다.

[표 A.2-1] 문항세트 a 검색어 1  
(자연어 기반 POI 검색, 토픽별 키워드 1개)

연번	검색어	키워드
1	떡볶이	떡볶이_NNG
2	평양냉면	평양냉면_NNP
3	시카고피자	시카고피자_NNP
4	칼국수	칼국수_NNG
5	소개팅	소개팅_NNG



[그림 A.2-1] 문항세트 a의 검색어 1에 대한 설문 화면 캡처



[표 A.2-3] 문항세트 a 검색어 3  
(자연어 기반 POI 검색, 토픽별 키워드 3개)

연번	검색어	키워드					
		1	가중치	2	가중치	3	가중치
1	식물 덕후가 좋아할 만한 곳	식물_NNG	0.5	덕후_NNP	0.3	좋아하_VV	0.2
2	스트레스 풀리는 디저트 가게	스트레스_NNG	0.3	풀리_VV	0.3	디저트_NNG	0.4
3	직장인 점심 맛집	직장인_NNG	0.3	점심_NNG	0.3	맛집_NNG	0.4
4	국물 맛 끝내주는 곳	국물_NNG	0.4	맛_NNG	0.2	끝내_VV	0.4
5	여자친구랑 데이트 하기 좋은 곳	여자친구_NNP	0.45	데이트_NNG	0.45	좋_VA	0.1

**'식물(중요도 50%)', '덕후(중요도 30%)', '좋아할 만한(중요도 20%)' 곳을 검색했을 때 관련 장소로 아래 같은 결과가 나온다면 만족하시나요?**

\*검색 결과는 **중요도가 높은 단어와 더 많은 관련**이 있습니다.

순위	장소명	음식종류	대표리뷰
1	코발트무드 (마포구연남동)	카페/디저트	뜨끈한 플랫 화이트랑 연유?들어간 따뜻한 라떼를 시켰어요. 플랫 화이트는 제가 좋아하는 고수운 맛은 별로 없었어요.
2	베르데 (강서구화곡동)	카페/디저트	커피향과 맛이 좋아요 식물과 꽃이 많아 힐링되는 느낌이에요
3	플로릭카페 (동대문구이문동)	카페/디저트	크... 꽃이랑 식물 가득가득.. 넘나 힐링 플레이스!
4	VERS HOUSE (마포구연남동)	카페/디저트	플라워 카페처럼 꽃 진짜 많고 앉을곳도 많다.
5	아리아데 (마포구등교동)	카페/디저트	허려한 외관만큼 내부도 드라이플라워와 수많은 꽃들 소품들로 예쁘게 장식되어 있었습니다.
6	까페이도 (종로구기회동)	카페/디저트	윗층은 도자기 갤러리 아래층은 공방 겸 카페를 운영, 그래서인지 분위기도 좋고 안락.
7	Silhouette Coffee (마포구합정동)	카페/디저트	사장님이 실루엣 덕후인 것은 확실해요. 인테리어에 흰색 커튼이 전부니까요
8	케플러커피 (서대문구정선동)	카페/디저트	친구말로는 엄청 하얗고 초록초록 하다길래 무슨 말이지 했는데 말 그대로 하얀~~~~배경에 초록 식물로 포인트를 준 인테리어더라고요
9	카페가도 (마포구연남동)	카페/디저트	일본의 작은 시골에 있는 가게 같기도 하고 뷰티인 사이드에 나올 거 같은 카페였어요!
10	카페325 (마포구등교동)	카페/디저트	분위기도 식물이 많아서 너무 좋았음

1개 선택

매우 불만족                      보통                      매우 만족

**'스트레스(중요도 30%)', '풀리는(중요도 30%)', '디저트(중요도 40%)' 가게를 검색했을 때 관련 장소로 아래 같은 결과가 나온다면 만족하시나요?**

\*검색 결과는 **중요도가 높은 단어와 더 많은 관련**이 있습니다.

순위	장소명	음식종류	대표리뷰
1	에니롱 (마포구합정동)	카페/디저트	최애 마카롱 가게예요! 선물용 디저트 필요할 때면 항상 에니롱을 이용합니다.
2	커피빈 (강남구삼성동)	카페/디저트	커피빈 브라우니 엄청 달아요..! 스트레스 풀고싶을 때 먹으면 머리가 멍 할 정도로 달달합니다...
3	골든 크림블 (용산구후암동)	카페/디저트	라멘집처럼 디저트가게에서 기계로 주문을 하는 방식이라 새로웠어요.
4	달콤한위로 (성북구성북동1가)	카페/디저트	스트레스 받는 날엔 달콤한 위로에서 마카롱 사들고 집어가는 길이 그렇게 행복할수가 없네염..!♡
5	곰이커피 (서대문구대원동)	카페/디저트	신촌이대 주변에서 디저트 가성비 원탑 아니니까?
6	요거프레소 (구로구신도림동)	카페/디저트	저렴이 카페중에 맛이나 메뉴다양성으로 보면 가장 훌륭한 카페인듯!
7	고릴랄라 (성동구옥수동)	카페/디저트	자바칩쿠키 프라페랑 바나나요거트 시켜서 먹었는데 쿠키 처음 먹을때부터 다 먹을때까지 입안에서 계속 씹히고 바나나요거트는 달고 신 맛이 적절히 조화를 이루네요.
8	소나 (강남구신사동)	카페/디저트	여자들이 좋아할만한 디저트 전문 카페 소나! 스트레스가 날라가는맛.
9	오코티디앙 (송파구송파동)	카페/디저트	담엔 꼭 낮에 와서 커피메뉴랑 디저트랑 먹어보고싶어요
10	고디바 (연남구연남동)	카페/디저트	자주 먹기에는 가격이 조금 부담스럽긴 하지만 단 거 맹길 때나 스트레스 왕창 받을 때 가끔 먹기 좋을 것 같다

1개 선택

매우 불만족                      보통                      매우 만족

[그림 A.2-3] 문항세트 a의 검색어 3에 대한 설문 화면 캡처



문항세트 a에서 선택 기반 POI 검색 결과에 대한 문항은 키워드 개수 (1~4개)별로 5개씩이므로 총 20개 문항을 포함한다. 키워드 개수에 따른 검색어는 [표 A.2-5]~[표 A.2-8]과 같다. 실제 오픈서베이를 통해 실시한 설문 화면 중 일부는 [그림 A.2-5]~[그림 A.2-8]과 같다.

[표 A.2-5] 문항세트 a 검색어 5  
(선택 기반 POI 검색, 토픽별 키워드 1개)

연번	토픽 번호	토픽	키워드
1	13	고기류	고기_NNG
2	17	일식	초밥_NNG
3	5	중식	마라_NNG
4	12	이탈리안	리조또_NNP
5	12	이탈리안	샐러드_NNG



[그림 A.2-5] 문항세트 a의 검색어 5에 대한 설문 화면 캡처

[표 A.2-6] 문항세트 a 검색어 6  
(선택 기반 POI 검색, 토픽별 키워드 2개)

연번	토픽 번호	토픽	키워드			
			1	가중치	2	가중치
1	12	이탈리안	파스타_NNP	0.3	맛있_VA	0.7
2	13	고기류	삼겹살_NNG	0.5	맛있_VA	0.5
3	17	일식	연어_NNG	0.4	신선_XR	0.6
4	8	브런치/베이커리	브런치_NNG	0.3	건강_NNG	0.7
5	7	카페/디저트	아메리카노_NNP	0.5	케이크_NNG	0.5

**'파스타(중요도 30%)', '맛있는(중요도 70%)'** 곳을 검색했을 때 관련 장소로 아래 같은 결과가 나온다면 만족하시나요?

\*검색 결과는 중요도가 높은 단어와 더 많은 관련이 있습니다.

순위	장소명	음식종류	대표리뷰
1	지중해 (종로구익선동)	세계음식 기타	오랜만에 정말 맛있게 파스타를 먹을 수 있었던 곳! 바질페스토 파스타는 진짜 너무 맛있었다ㅠㅠ 내 인생 바질페스토 굽는 곳이 두군데 있었는데 이제 여기가 아마 1등...?
2	523 (마포구연남동)	이탈리안	새우 게살 크림 파스타는 가장최근에 갔을때 먹었는데 맛있다.
3	클래식당 (마포구연남동)	이탈리안	늦게까지하는 파스타맛집. 새벽2시까지 영업. 조용히식사하기좋은집
4	22번가 키친 (성북구동선동)	기타 양식	파스타가 괜찮다고 보기도 해서 겸사겸사 가게되었어요. 새우 알리오올리오는 페퍼oni치노가 들어가서 그런지 약간 매콤하고 짭짤해서 느끼하지 않고 좋았습니다.
5	티틀로 (강남구청동)	이탈리안	강추 천만개
6	시장에간 남자 (서대문구창천동)	이탈리안	맛 최고 분위기 최고 기념일맞이해서 또갔다 너무 맛있어서 여기서 같이 밥 먹은 사람이랑은 평생 친하게 지낼 수 있게 만드는 맛
7	이층양육 (종로구익선동)	퓨전 양식	스파이시 로제 파스타는 제가 좋아하는 스타일의 (다이너 렉이랑 비슷해여!) 소스였어요 맛있어여
8	루이루이 (중구신당동)	기타 양식	피자는 매우 굿. 파스타는 알던데 살기가 아니여서 아쉽. 이젠 취향차이일듯
9	더그리드 (마포구서교동)	이탈리안	직원분들도 너무 핵친절하시고 음식도 뭐이건 너무 맛있잖아요 ㅠㅠ
10	고메트리 (성동구금호동)	프랑스 음식	파스타는 비싼집에서 먹어야하나보다 이런 식감이 나오다니 너무 조화롭고 넘 넘엄 좋다. 최고

1개 선택

매우 불만족                      보통                      매우 만족

**'삼겹살(중요도 50%)', '맛있는(중요도 50%)'** 곳을 검색했을 때 관련 장소로 아래 같은 결과가 나온다면 만족하시나요?

\*검색 결과는 중요도가 높은 단어와 더 많은 관련이 있습니다.

순위	장소명	음식종류	대표리뷰
1	팔색삼겹살 (마포구서교동)	고기 요리	진짜 맛있다!
2	돼지집 (광진구지양동)	고기 요리	생삼겹살 전문점입니다!! 우선 자갈 밑에 숯불을 넣고 찰판들을 칸 다음 자갈을 옮깁니다 그 뒤에 자갈이 달궈지면 소량씩 먹을만큼씩만 삼겹살을 올려서 먹는데여!! 정말 맛있었어요!!!
3	코기코기 (강남구신사동)	고기 요리	허브 삼겹살이 맛있다고 해서 간 곳. 처음엔 냉동고기가 나와서 놀랐는데 먹어보니 대박 끝맛! 삼겹살 종류가 다양해서 골라먹는 재미도 쏠쏠하다. 강주메뉴:파프리카 삼겹살
4	행복추평형 칼삼겹살 (강남구신사동)	고기 요리	고기도 맛있고 가격도 적당한 업구정에서 통편하는 곳 중 하나
5	육전식당 (강남구역삼동)	고기 요리	태어나서 먹어본 돼지고기 중 가장 맛있었네요 더이상의 설명은 필요 없을 것 같습니다
6	고기요 (은평구수색동)	고기 요리	맛있어요! 반찬도 맛있고 (양념게장 나옵니다) 된장찌개도 훌륭합니다!
7	맛찬들 왕소금구이 (강서구마곡동)	고기 요리	숙성된 삼겹살이라 그런지 가격은 좀 비쌌지만 직접 구워주시고 고기와 반찬이 맛있던 집.
8	돌고기506 (강남구역삼동)	고기 요리	돼지고기 냄새가 이렇게 구수한 냄새와 맛있었구나!!! 재방문 의사 80%
9	고기스토리 (동작구상도동)	고기 요리	축적에서 삼겹살 제일 맛있고 싼곳 .. 그리고물냉맛집임
10	양철집 (강남구신사동)	고기 요리	아... 생삼겹은 이래야 하는구나... 라는 기준을 정해 주는 집! 목사발과 김범벅 안먹으면 섭섭함

1개 선택

매우 불만족                      보통                      매우 만족

[그림 A.2-6] 문항세트 a의 검색어 6에 대한 설문 화면 캡처

[표 A.2-7] 문항세트 a 검색어 7  
(선택 기반 POI 검색, 토픽별 키워드 3개)

연번	토픽 번호			토픽			키워드					
	1	2	3	1	2	3	1	가중치	2	가중치	3	가중치
1	12	1	3	이탈리안	가격	맛집 검색표현	스테이크_ NNG	0.3	성비_NNG	0.5	맛집_NNG	0.2
2	15	17	3	서비스	일식	맛집 검색표현	친절_NNG	0.3	스시_NNP	0.3	맛집_NNG	0.4
3	11	3	면류			맛집 검색표현	국수_NNG	0.3	매콤_NNP	0.5	맛집_NNG	0.2
4	11	3	면류			맛집 검색표현	쌀국수_ NNP	0.3	칼끔_XR	0.4	맛집_NNG	0.3
5	7	16	카페/ 디저트	분위기			카페_NNG	0.3	내부_NNG	0.3	넓_VA	0.4

**'스테이크(중요도 30%)', '가성비(중요도 50%)', '맛집(중요도 20%)'** 을 검색했을 때 관련 장소로 아래 같은 결과가 나온다면 만족하시나요?

\*검색 결과는 **중요도가 높은 단어와 더 많은 관련**이 있습니다.

순위	장소명	음식종류	대표리뷰
1	스파게티 스토리 (서초구 서초동)	이탈리안	가성비 최고♡스파게티 맛, 풀리지 않아요♡
2	버니 스테이크 하우스 (관악구 신림동)	스테이크/바비큐	가성비있는 신림대학동 고서촌근처 스테이크 집. 주머니가 가벼운 분들 스테이크 외식시 추천
3	티바두마리치킨 (강북구 번동)	치킨/호프/펍	초기엔 가성비로 자주 먹었지만 이젠 근처에 가성비로 더 나은 치킨집이 많아...
4	산동관 (관진구 자양동)	정통 중식/일반 중식	가성비 짱!
5	세븐 스테이크 (관진구 화양동)	스테이크/바베큐	학교에 있는 곳이다 보니 스테이크치고는 저렴한 편이다. 하지만 맛있다.
6	육회한연어 (종로구 명동2가)	고기 요리	가성비 좋다.
7	달달보드레 (구로구 구로동)	이탈리안	가격대비 괜찮은곳..!
8	헤비 스테이크 (종로구 계동)	스테이크/바비큐	이 동네에 가성비 좋은집이 거의없는데 최근에 역근처에 오픈하면서 가성비로 유명세를 타고 있는 스테이크집
9	칼 (서대문구 창천동)	퓨전 양식	부채살스테이크는 가성비 좋은 소고기 스테이크 느낌이다~
10	스킬렛 (강남구 신사동)	스테이크/바베큐	역시 이 집은 가성비 좋게 스테이크를 먹으러 오기 좋은 집이었다.

1개 선택

매우 불만족                      보통                      매우 만족

**'친절함(중요도 30%)', '스시(중요도 30%)', '맛집(중요도 40%)'** 을 검색했을 때 관련 장소로 아래 같은 결과가 나온다면 만족하시나요?

\*검색 결과는 **중요도가 높은 단어와 더 많은 관련**이 있습니다.

순위	장소명	음식종류	대표리뷰
1	스시마에 (강동구 상계동)	회/스시	완전 맛있어서 한달에 두번찾는 맛집 엄마 월급날 한번 + 내 월급날 한번
2	기왓스시 (강서구 화곡동)	회/스시	화곡역에서 스시를 기대하기는 좀 그렇다. 하지만 꼭 가야한다면 이곳을 간다. 나가사키 짬뽕은 불맛을 입혀서 꽤 괜찮다. 서비스도 좋다
3	스시고 (종로구 삼청동)	회/스시	진짜 맛나게 먹고왔어요. 직원분 모두 친절하시고.
4	유쾌한스시 (서대문구 연희동)	회/스시	사시미 코스 감추합니다.
5	스시 이타마에 (마포구 홍강동)	회/스시	식전 죽-연어덮밥-우동 이렇게 나오는데 전부 맛있어요. 사장님도 엄청 친절하시고 ㅠ
6	스시현 (강남구 개포동)	회/스시	정말 친절하신 사장님 이 집에 가시면 연어초밥은 무조건, 반드시! 드셔야합니다!! 정말 정말 맛있습니니다.
7	엔스시 (용산구 한남동)	회/스시	초밥 자체는 맛있고 서비스도 좋은데 비슷한 다른 초밥집에 비해서 가격대가 쎈 느낌이에요
8	김스시 (종로구 정진동)	회/스시	이 가격에 이 정도로 서울에서 먹을 수 있구나ㅠㅠ 스시와 우동 세트가 8500원이었다. 둘다 매우 맛있었다.
9	유정상초밥 (마포구 상암동)	회/스시	퐁기는 분위기가 포스가 범상치 않은 듯 해서 방문한 곳이지만 재료의 한계인지 맛은 그냥 괜찮은 정도였음. 하지만 가성비는 최고 수준!
10	굿모닝스시 (관진구 자양동)	회/스시	스시하나하나 정성이 들어가서 맛있다. 특히 구운생선인 아부리류의 초밥은 훌륭한 느낌 사장님도 매우 친절하시고 갈때마다 서비스를 주신다. 추천!

1개 선택

매우 불만족                      보통                      매우 만족

[그림 A.2-7] 문항세트 a의 검색어 7에 대한 설문 화면 캡처

[표 A.2-8] 문항세트 a 검색어 8  
(선택 기반 POI 검색, 토픽별 키워드 4개)

연번	토픽 번호				토픽				키워드							
	1	2	3	4	1	2	3	4	1	가중치	2	가중치	3	가중치	4	가중치
1	11	3	4		면류	맛집 검색 표현	긍정 표현		닭_NNG	0.3	육수_NNP	0.3	맛집_NNG	0.2	추천_NNG	0.2
2	4	17			긍정 표현		일식		즌_NNP	0.25	맛_NNG	0.25	언어_NNG	0.3	초밥_NNG	0.2
3	16	7			분위기	카페/디저트		인테리어_NNG	0.2	예쁘_VA	0.4	케이크_NNG	0.3	카페_NNG	0.1	
4	4	8			긍정 표현	브런치/ 베이커리		감동_NNG	0.2	꿀맛_NNG	0.2	브런치_NNG	0.2	샌드위치_NNG	0.4	
5	1	17			가격		일식	성비_NNG	0.4	좋_VA	0.1	사시미_NNG	0.3	코스_NNG	0.2	

**'닭(중요도 30%)', '육수(중요도 30%)', '맛집(중요도 20%)', '추천(중요도 20%)'** 을 검색했을 때 관련 장소로 아래 같은 결과가 나온다면 만족하시나요?

\*검색 결과는 **중요도가 높은 단어와 더 많은 관련**이 있습니다.

순위	장소명	음식종류	대표리뷰
1	서평 기사식당 (강남구 일원동)	닭/오리 요리	이미 동네에서는 알사람 다 아는 맛집이지만 변화가 없었으면 방송타고 출서지역을 것 같은 맛집! 저렴한 가격에 항상 같은 맛과 질을 유지하는 닭육수에 24시간 영업이라서 출출할때 가기에 부담 없다.
2	다락투 (마포구 서교동)	닭/오리 요리	깔끔한 닭육수에 적당히 토렴된 밥알과 한입 크기로 찢긴 닭고기가 좋습니다.
3	소문 난원할매 닭한마리 (종로구 종로5가)	닭/오리 요리	닭 한마리가 통째로 들어가고 살짝 더 끓인 후 잘라주신다. 육수는 맑지 않고 살짝 진하게 우려낸 맛이라 밥도 많이 먹고 싶을 정도 ㅎㅎ
4	백부장집 닭한마리 (종로구 공평동)	닭/오리 요리	맛있다. 시원한 육수에 파, 떡 그리고 닭 이렇게만 들어가 있어서 깔끔하고 맛있다. 마성의 육수!
5	다시올치킨 (성동구 사근동)	닭/오리 요리	왕십리 올 일요일엔 자주 들리는 맛집이다. 닭갈육수와 치킨, 여름메뉴로 초계냉면 등등을 내놓는 집이다.
6	시조명동 닭한마리 (종로구 종로5가)	닭/오리 요리	매번 찾아가는 닭한마리입니다. 추천드려요!!
7	진원조 닭한마리 (서대문구 창천동)	닭/오리 요리	닭육수가 좋아. 진한국물, 소주, 죽
8	공릉 닭한마리 (서대문구 대원동)	닭/오리 요리	비실비실하면 먹는거 추천 닭육수가 갈국수될쯤 완성되서 그때부터 죽까지 먹으면 기분도 좋고 양도 많고 최고다
9	동대 닭한마리 (중구 필동3가)	닭/오리 요리	친구랑 닭한마리 맛집 새로 알아냈다가 서로 추천했던 맛집~
10	돈화문 닭한마리 (종로구 돈의동)	탕/찌개/전골	건물 좋은 이쪽으로 갔는데 잘먹고음!

1개 선택

매우 불만족                      보통                      매우 만족

**'존맛(중요도 50%)', '언어(중요도 30%)', '초밥(중요도 20%)'** 을 검색했을 때 관련 장소로 아래 같은 결과가 나온다면 만족하시나요?

\*검색 결과는 **중요도가 높은 단어와 더 많은 관련**이 있습니다.

순위	장소명	음식종류	대표리뷰
1	램프 샌드위치 (마포구 서교동)	브런치/버거/샌드위치	언어샐러드에 발사믹소스 완전 짱짱맛
2	스시이찌방 (노원구 상계동)	회/스시	언어초밥 맛있어서 자주포장해먹어요
3	알래스카에서 온언어가 맛있는집 (동대문구 제기동)	회/스시	언어존맛 밤에 사람 마는
4	육회먹은언어 (은평구 길원동)	고기 요리	다이어트를 결심해서 그냥 깔끔하게 언어를 먹었쥬~ 맛있어서 다시칼랜..ㅠ
5	동차밥 (마포구 연남동)	퓨전 일식	사케동 먹었는데 언어좋아해서 완전 맛있게 먹었어요!!
6	오징어나라 (서초구 서초동)	해산물 요리	언어가 맛있는 오징어나라
7	빽스 (성북구 하월곡동)	패밀리 레스토랑	언어메뉴 중 가장 맛있던 건 레몬크림언어베이크!
8	동신참치 (중구 서소문동)	회/스시	맛있었습니다.
9	홍대돈부리 (송파구 선전동)	돈부리/일본 카레/벤토	언어낫또동 먹었는데 넘나 맛있었어요...
10	살모니스트 (용산구 이태원동)	시푸드 요리	분위기도 좋고 언어가 너무 싱싱하고 맛있었어요♡

1개 선택

매우 불만족                      보통                      매우 만족

[그림 A.2-8] 문항세트 a의 검색어 8에 대한 설문 화면 캡처

문항세트 a에서 유사 POI 검색 결과에 대한 문항은 총 10개로 구성되어, 문항별 기준 POI는 [표 A.2-9]와 같다. 실제 오픈서베이를 통해 실시한 설문 화면 중 일부는 [그림 A.2-9]와 같다.

[표 A.2-9] 문항세트 a 검색어 9(유사 POI 검색)

연번	POI 명칭	주소	음식 종류	가격대
1	1202 커피카운터	강남구 신사동 518-15	카페/디저트	1만원 미만
2	고고초밥	강남구 역삼동 827-69	회/스시	1-2만원
3	정라멘	송파구 삼전동 2-4	라멘/소바/우동	1만원 미만
4	고대양꼬치집	성북구 안암동5가 146-4	기타 중식	-
5	익선잡방	종로구 익선동 166-80	이탈리안	1-2만원
6	전통쌀국수 Pho Bo	은평구 녹번동 43-39	베트남 음식	1만원 미만
7	히메지	서대문구 연희동 129-1	돈부리/일본카레/벤토	1만원 미만
8	전봇대	강남구 역삼동 669-2	국수/면 요리	1만원 미만
9	정광수의돈까스가게	마포구 합정동 457-3	기타 한식	1만원 미만
10	전대감덕	종로구 체부동 184	전통 주점/포차	1-2만원



[그림 A.2-9] 문항세트 a의 검색어 9에 대한 설문 화면 캡처

### A.3 문항세트 b 설문 조사 내용

문항세트 b에서 자연어 기반 POI 검색 결과에 대한 문항은 키워드 개수 (1~4개)별로 5개씩이므로 총 20개 문항을 포함한다. 키워드 개수에 따른 검색어는 [표 A.3-1]~[표 A.3-4]와 같다. 실제 오픈서베이를 통해 실시한 설문 화면은 지면 관계상 생략한다.

[표 A.3-1] 문항세트 b 검색어 1  
(자연어 기반 POI 검색, 토픽별 키워드 1개)

연번	검색어	키워드
1	짜장면	짜장면_NNG
2	맥주	맥주_NNG
3	초코케익	초코케익_NNP
4	연어	연어_NNG
5	기념일	기념일_NNG

[표 A.3-2] 문항세트 b 검색어 2  
(자연어 기반 POI 검색, 토픽별 키워드 2개)

연번	검색어	키워드			
		1	가중치	2	가중치
1	인생 타코	인생_NNG	0.4	타코_NNG	0.6
2	삼선짜장 맛집	삼선짜장_NNP	0.8	맛집_NNG	0.2
3	한적한 카페	한적_XR	0.8	카페_NNG	0.2
4	상견례 장소 추천	상견례_NNG	0.9	추천_NNG	0.1
5	유명한 고기집	유명_XR	0.4	고기_NNG	0.6

[표 A.3-3] 문항세트 b 검색어 3  
(자연어 기반 POI 검색, 토픽별 키워드 3개)

연번	검색어	키워드					
		1	가중치	2	가중치	3	가중치
1	간단하게 혼밥하기 좋은 곳	간단_XR	0.4	혼_NNG	0.3	밥_NNG	0.3
2	수제 버거 맛있는 곳	수제_NNG	0.4	버거_NNP	0.4	맛있_VA	0.2
3	뜨끈한 국물 한그릇	뜨끈_XR	0.45	국물_NNG	0.45	그릇_NNG	0.1
4	인스타그램 감성 카페	인스타그램_NNP	0.4	감성_NNG	0.3	카페_NNG	0.3
5	조용한 분위기 카페	조용_XR	0.5	분위기_NNG	0.25	카페_NNG	0.25

[표 A.3-4] 문항세트 b 검색어 4  
(자연어 기반 POI 검색, 토픽별 키워드 4개)

연번	검색어	키워드							
		1	가중치	2	가중치	3	가중치	4	가중치
1	혼자 노트북 들고 작업하러 가기 좋은 곳	혼자_NNG	0.25	노트북_NNG	0.25	작업_NNG	0.25	좋_VA	0.25
2	가성비 좋은 사시미 맛집	성비_NNG	0.2	좋_VA	0.2	사시미_NNG	0.4	맛집_NNG	0.2
3	주말 저녁에 데이트하기 좋은 곳	주말_NNG	0.2	저녁_NNG	0.3	데이트_NNG	0.4	좋_VA	0.1
4	인테리어 멋있고 한적한 카페	인테리어_NNG	0.2	멋있_VA	0.3	한적_XR	0.4	카페_NNG	0.1
5	맛있는 가성비 브런치 식당	맛있_VA	0.3	성비_NNG	0.3	브런치_NNG	0.3	식당_NNG	0.1

문항세트 b에서 선택 기반 POI 검색 결과에 대한 문항은 키워드 개수 (1~4개)별로 5개씩이므로 총 20개 문항을 포함한다. 키워드 개수에 따른 검색어는 [표 A.3-5]~[표 A.3-8]과 같다. 실제 오픈서베이를 통해 실시한 설문 화면은 지면 관계상 생략한다.

[표 A.3-5] 문항세트 b 검색어 5  
(선택 기반 POI 검색, 토픽별 키워드 1개)

연번	토픽 번호	토픽	키워드
1	12	이탈리안	와인_NNG
2	10	멕시코	퀘사디아_NNP
3	6	치킨/튀김류	돈까스_NNP
4	11	면류	쌀국수_NNP
5	9	냉면/만두	냉면_NNG

[표 A.3-6] 문항세트 b 검색어 6  
(선택 기반 POI 검색, 토픽별 키워드 2개)

연번	토픽 번호	토픽	키워드			
			1	가중치	2	가중치
1	12	이탈리안	스테이크_NNG	0.4	맛있_VA	0.6
2	5	중식	짬뽕_NNG	0.2	맛있_VA	0.8
3	6	치킨/튀김류	치킨_NNG	0.3	맛있_VA	0.7
4	11	면류	국물_NNG	0.4	깔끔_XR	0.6
5	8	브런치/베이커리	브런치_NNG	0.5	건강_NNG	0.5

[표 A.3-7] 문항세트 b 검색어 7  
(선택 기반 POI 검색, 토픽별 키워드 3개)

연번	토픽 번호			토픽			키워드					
	1	2	3	1	2	3	1	가중치	2	가중치	3	가중치
1	13	1	3	고기류	가격	맛집 검색표현	삼겹살_NNG	0.4	성비_NNG	0.4	맛집_NNG	0.2
2	1	4	3	가격	공정 표현	맛집 검색표현	저렴_XR	0.4	맛있_VA	0.4	식당_NNG	0.2
3	6	4	3	치킨/ 튀김류	공정 표현	맛집 검색표현	돈까스_NNP	0.5	맛있_VA	0.3	식당_NNG	0.2
4	13	15		고기류		서비스	불고기_NNG	0.4	맛있_VA	0.5	친절_NNG	0.1
5	2	16		일시/혼잡도		분위기	주말_NNG	0.2	오후_NNG	0.2	데이트_NNG	0.6

[표 A.3-8] 문항세트 b 검색어 8  
(선택 기반 POI 검색, 토픽별 키워드 4개)

연번	토픽 번호				토픽				키워드							
	1	2	3	4	1	2	3	4	1	가중치	2	가중치	3	가중치	4	가중치
1	16	13	3	4	분위기	고기류	맛집 검색 표현	긍정 표현	칼끔_XR	0.3	돼지고기 _NNG	0.3	식당_ NNG	0.1	추천_ NNG	0.3
2	2	8	4	3	일식/ 혼잡도	브런치/ 베이커리	긍정 표현	맛집 검색 표현	주말_ NNG	0.3	브런치_ NNG	0.4	추천_ NNG	0.2	맛집_ NNG	0.1
3	16	17	3	4	분위 기	일식	맛집 검색 표현	긍정 표현	조용_XR	0.4	스시_ NNP	0.4	식당_ NNG	0.1	추천_ NNG	0.1
4	2	16	4	3	일식/ 혼잡도	분위 기	긍정 표현	맛집 검색 표현	평일_ NNG	0.3	데이트_ NNG	0.4	추천_ NNG	0.2	맛집_ NNG	0.1
5	13	4	3		고기류	긍정 표현	맛집 검색 표현	김치_ NNG	0.3	삼겹살_ NNG	0.4	추천_ NNG	0.2	식당_ NNG	0.1	

문항세트 b에서 유사 POI 검색 결과에 대한 문항은 총 10개로 구성되며, 문항별 기준 POI는 [표 A.3-9]와 같다. 실제 오픈서베이를 통해 실시한 설문 화면은 지면 관계상 생략한다.

[표 A.3-9] 문항세트 b 검색어 9(유사 POI 검색)

연번	POI 명칭	주소	음식 종류	가격대
1	1986베이커	노원구 상계동 430-21	베이커리	1만원 미만
2	전영준프로초밥	종로구 내자동 4	회/스시	1-2만원
3	전티마이베트남쌀국수	관악구 신림동 601-15	베트남 음식	1만원 미만
4	저스틴스테이크	용산구 이태원동 129-5	스테이크/바베큐	1-2만원
5	전골떡볶이덕미가	서대문구 대현동 54-9	기타 한식	1-2만원
6	전곱	종로구 낙원동 110	고기 요리	1-2만원
7	전주영양돌솥밥	성동구 옥수동 218-1	한정식/백반/ 정통 한식	1-2만원
8	정 한뿌리죽	송파구 문정동 634	기타 한식	1-2만원
9	전봇대곱창	영등포구 신길동 99-7	고기 요리	1-2만원
10	히피히피쉐이크	마포구 동교동 154-1	칵테일/와인	1-2만원

## A.4 문항세트 c 설문 조사 내용

문항세트 c에서 자연어 기반 POI 검색 결과에 대한 문항은 키워드 개수 (1~4개)별로 5개씩이므로 총 20개 문항을 포함한다. 키워드 개수에 따른 검색어는 [표 A.4-1]~[표 A.4-4]와 같다. 실제 오픈서베이를 통해 실시한 설문 화면은 지면 관계상 생략한다.

[표 A.4-1] 문항세트 c 검색어 1  
(자연어 기반 POI 검색, 토픽별 키워드 1개)

연번	검색어	키워드
1	빵집	빵집_NNG
2	아이스크림	아이스크림_NNG
3	스테이크	스테이크_NNG
4	한우	한우_NNG
5	가성비	성비_NNG

[표 A.4-2] 문항세트 c 검색어 2  
(자연어 기반 POI 검색, 토픽별 키워드 2개)

연번	검색어	키워드			
		1	가중치	2	가중치
1	인생 마카롱	인생_NNG	0.5	마카롱_NNP	0.5
2	아내와 기념일	아내_NNG	0.2	기념일_NNG	0.8
3	더치커피 맛집	더치커피_NNP	0.8	맛집_NNG	0.2
4	플라워 카페	플라워_NNG	0.9	카페_NNG	0.1
5	회식하기 좋은 곳	회식_NNG	0.7	좋_VA	0.3

[표 A.4-3] 문항세트 c 검색어 3  
(자연어 기반 POI 검색, 토픽별 키워드 3개)

연번	검색어	키워드					
		1	가중치	2	가중치	3	가중치
1	믿고 먹는 곱창집	믿_VV	0.5	먹_VV	0.1	곱창_NNP	0.4
2	뷰 좋은 카페	뷰_NNP	0.4	좋_VA	0.3	카페_NNG	0.3
3	깔끔하고 캐주얼한 한식집	깔끔_XR	0.3	캐주얼_NNG	0.3	한식_NNG	0.4
4	공부하기 좋은 카페	공부_NNG	0.5	좋_VA	0.2	카페_NNG	0.3
5	맵고 맛있는 닭발집	맵_VA	0.4	맛있_VA	0.3	닭발_NNG	0.3

[표 A.4-4] 문항세트 c 검색어 4  
(자연어 기반 POI 검색, 토픽별 키워드 4개)

연번	검색어	키워드							
		1	가중치	2	가중치	3	가중치	4	가중치
1	조용하게 혼술하기 좋은 곳	조용_XR	0.3	혼_NNG	0.3	술_NNG	0.3	좋_VA	0.1
2	부모님이랑 저녁 먹기 좋은 곳	부모님_NNP	0.4	저녁_NNG	0.3	먹_VV	0.2	좋_VA	0.1
3	바게트 맛있는 베이커리 맛집	바게트_NNP	0.4	맛있_VA	0.3	베이커리_NNG	0.2	맛집_NNG	0.1
4	타르트 맛있는 디저트 카페	타르트_NNP	0.4	맛있_VA	0.3	디저트_NNG	0.2	카페_NNG	0.1
5	겁나 매운 갈비찜 맛집	겁나_VV	0.2	맵_VA	0.3	갈비찜_NNP	0.3	맛집_NNG	0.2

문항세트 c에서 선택 기반 POI 검색 결과에 대한 문항은 키워드 개수 (1~4개)별로 5개씩이므로 총 20개 문항을 포함한다. 키워드 개수에 따른 검색어는 [표 A.4-5]~[표 A.4-8]과 같다. 실제 오픈서베이를 통해 실시한 설문 화면은 지면 관계상 생략한다.

[표 A.4-5] 문항세트 c 검색어 5  
(선택 기반 POI 검색, 토픽별 키워드 1개)

연번	토픽 번호	토픽	키워드
1	13	고기류	불고기_NNG
2	11	면류	닭고기_NNG
3	8	브런치/베이커리	샌드위치_NNG
4	7	카페/디저트	카페라떼_NNP
5	14	술집	막걸리_NNG

[표 A.4-6] 문항세트 c 검색어 6  
(선택 기반 POI 검색, 토픽별 키워드 2개)

연번	토픽 번호		토픽		키워드			
	1	2	1	2	1	가중치	2	가중치
1	10	3	멕시코칸	맛집검색표현	타코_NNG	0.5	생각_NNG	0.5
2	14		술집		맥주_NNG	0.4	추천_NNG	0.6
3	9		냉면/만두		냉면_NNG	0.5	맛있_VA	0.5
4	1		가격		성비_NNG	0.5	좋_VA	0.5
5	16		분위기		분위기_NNG	0.2	아늑_XR	0.8

[표 A.4-7] 문항세트 c 검색어 7  
(선택 기반 POI 검색, 토픽별 키워드 3개)

연번	토픽 번호			토픽			키워드					
	1	2	3	1	2	3	1	가중치	2	가중치	3	가중치
1	14			술집			가볍_VA	0.3	한잔_NNG	0.4	괜찮_VA	0.3
2	9			냉면/만두			평양냉면_NNP	0.4	만두_NNG	0.4	맛있_VA	0.2
3	8			브런치/베이커리			치즈_NNG	0.4	촉촉_XR	0.4	빵_NNG	0.2
4	11			면류			감칠맛_NNG	0.3	쫄깃_XR	0.4	국수_NNG	0.3
5	12			이탈리안			트리플_NNP	0.4	리조또_NNP	0.4	맛있_VA	0.2

[표 A.4-8] 문항세트 c 검색어 8  
(선택 기반 POI 검색, 토픽별 키워드 4개)

연번	토픽 번호				토픽				키워드							
	1	2	3	4	1	2	3	4	1	가중치	2	가중치	3	가중치	4	가중치
1	7	3	4	카페/디저트	맛집 검색 표현	긍정 표현	우유_ NNG	0.4	빙수_ NNG	0.3	맛집_ NNG	0.2	추천_ NNG	0.1		
2	17	3	4	일식	맛집 검색 표현	긍정 표현	연어_ NNG	0.4	스시_ NNP	0.3	맛집_ NNG	0.2	추천_ NNG	0.1		
3	17	4	일식		긍정 표현	재료_ NNG	0.2	신선_XR	0.3	초밥_ NNG	0.2	맛있_ VA	0.3			
4	7	4	카페/디저트		긍정 표현	얼그레이_ NNP	0.4	케이크_ NNG	0.1	카페_ NNG	0.1	맛있_ VA	0.4			
5	11	3	면류		맛집 검색 표현	라멘_ NNP	0.3	국물_ NNG	0.3	칼끔_XR	0.3	맛집_ NNG	0.1			

문항세트 c에서 유사 POI 검색 결과에 대한 문항은 총 10개로 구성되며, 문항별 기준 POI는 [표 A.4-9]와 같다. 실제 오픈서베이를 통해 실시한 설문 화면은 지면 관계상 생략한다.

[표 A.4-9] 문항세트 c 검색어 9(유사 POI 검색)

연번	POI 명칭	주소	음식 종류	가격대
1	절묘한커피그리고 눈꽃빙수연구소	마포구 도화동 290-4	카페/디저트	1만원 미만
2	129 라멘하우스	중구 신당동 374-27	라멘/소바/우동	1만원 미만
3	계타는날 부암동치킨	종로구 통인동 109-2	치킨/호프/뽕	1-2만원
4	달궁	중구 신당동 141-2	태국 음식	1만원 미만
5	정다래함박스택	동대문구 제기동 137-426	스테이크/바베큐	1만원 미만
6	김밥천국	광진구 구의동 220-218	기타 한식	1만원 미만
7	정대포	마포구 도화동 183-16	고기 요리	1-2만원
8	1인식당봄	관악구 신림동 241-111	한정식/백반/ 정통 한식	1만원 미만
9	전라도맛집	종로구 관수동 154-2	한정식/백반/ 정통 한식	1만원 미만
10	전주순대국	강남구 삼성동 143-19	탕/찌개/전골	1만원 미만

## A.5 문항세트 d 설문 조사 내용

문항세트 d에서 자연어 기반 POI 검색 결과에 대한 문항은 키워드 개수 (1~4개)별로 5개씩이므로 총 20개 문항을 포함한다. 키워드 개수에 따른 검색어는 [표 A.5-1]~[표 A.5-4]와 같다. 실제 오픈서베이를 통해 실시한 설문 화면은 지면 관계상 생략한다.

[표 A.5-1] 문항세트 d 검색어 1  
(자연어 기반 POI 검색, 토픽별 키워드 1개)

연번	검색어	키워드
1	타코	타코_NNG
2	아포가토	아포가토_NNP
3	파스타	파스타_NNP
4	데이트	데이트_NNG
5	브런치	브런치_NNG

[표 A.5-2] 문항세트 d 검색어 2  
(자연어 기반 POI 검색, 토픽별 키워드 2개)

연번	검색어	키워드			
		1	가중치	2	가중치
1	가족 모임 장소	가족_NNG	0.5	모임_NNG	0.5
2	외국인 친구랑	외국인_NNG	0.7	친구_NNG	0.3
3	뷔페 추천	뷔페_NNG	0.6	추천_NNG	0.4
4	양념치킨과 맥주	양념치킨_NNP	0.6	맥주_NNG	0.4
5	심심하고 건강함	심심_XR	0.5	건강_NNG	0.5

[표 A.5-3] 문항세트 d 검색어 3  
(자연어 기반 POI 검색, 토픽별 키워드 3개)

연번	검색어	키워드					
		1	가중치	2	가중치	3	가중치
1	스트레스 풀리는 매운맛	스트레스_ NNG	0.25	풀리_VV	0.25	매운맛_ NNP	0.5
2	분위기 아늑한 카페	분위기_ NNG	0.1	아늑_XR	0.6	카페_ NNG	0.3
3	인테리어 독특한 카페	인테리어_ NNG	0.3	독특_XR	0.4	카페_ NNG	0.3
4	남친이랑 데이트 하기 좋은 곳	남친_ NNP	0.45	데이트_ NNG	0.45	좋_ VA	0.1
5	연인과 화려한 와인 바	연인_ NNG	0.3	화려_XR	0.3	와인바_ NNG	0.4

[표 A.5-4] 문항세트 d 검색어 4  
(자연어 기반 POI 검색, 토픽별 키워드 4개)

연번	검색어	키워드							
		1	가중치	2	가중치	3	가중치	4	가중치
1	소개팅하는 날 저녁 식사하기 좋은 곳	소개팅_ NNG	0.4	저녁_ NNG	0.3	식사_ NNG	0.2	좋_ VA	0.1
2	진한 삼계탕 국물 땡길 때	진한_ NNP	0.2	삼계탕_ NNP	0.3	국물_ NNG	0.3	땡기_ VV	0.2
3	완전 매운 떡볶이 맛집	완전_ NNG	0.2	맵_ VA	0.2	떡볶이_ NNG	0.5	맛집_ NNG	0.1
4	배경음악 선곡 센스 좋은 곳	배경음악_ NNP	0.4	선곡_ NNG	0.2	센스_ NNG	0.2	좋_ VA	0.2
5	자리 넓은 한적한 카페	자리_ NNG	0.2	넓_ VA	0.3	한적_ XR	0.3	카페_ NNG	0.2

문항세트 d에서 선택 기반 POI 검색 결과에 대한 문항은 키워드 개수 (1~4개)별로 5개씩이므로 총 20개 문항을 포함한다. 키워드 개수에 따른 검색어는 [표 A.5-5]~[표 A.5-8]과 같다. 실제 오픈서베이를 통해 실시한 설문 화면은 지면 관계상 생략한다.

[표 A.5-5] 문항세트 d 검색어 5  
(선택 기반 POI 검색, 토픽별 키워드 1개)

연번	토픽 번호	토픽	키워드
1	6	치킨/튀김류	치킨_NNG
2	8	브런치/베이커리	브런치_NNG
3	7	카페/디저트	얼그레이_NNP
4	9	냉면/만두	만두_NNG
5	13	고기류	닭갈비_NNP

[표 A.5-6] 문항세트 d 검색어 6  
(선택 기반 POI 검색, 토픽별 키워드 2개)

연번	토픽 번호		토픽		키워드			
	1	2	1	2	1	가중치	2	가중치
1	11	3	면류	맛집검색표현	라멘_NNP	0.4	맛집_NNG	0.6
2	6	3	치킨/튀김류	맛집검색표현	감자튀김_NNP	0.7	맛집_NNG	0.3
3	7	4	카페/디저트	긍정표현	커피_NNG	0.5	맛있_VA	0.5
4	4	13	긍정 표현	고기류	인생_NNG	0.6	갈비_NNG	0.4
5	16		분위기		내부_NNG	0.2	조용_XR	0.8

[표 A.5-7] 문항세트 d 검색어 7  
(선택 기반 POI 검색, 토픽별 키워드 3개)

연번	토픽 번호			토픽			키워드					
	1	2	3	1	2	3	1	가중치	2	가중치	3	가중치
1	13			고기류			고기_NNG	0.5	무한_NNG	0.25	리필_NNG	0.25
2	6			치킨/튀김류			돈까스_NNP	0.6	떡_VV	0.2	좋_VA	0.2
3	5			중식			짜장면_NNG	0.3	맛있_VA	0.6	중국집_NNG	0.1
4	16			분위기			인테리어_NNG	0.2	예쁘_VA	0.4	깔끔_XR	0.4
5	8			브런치/베이커리			건강_NNG	0.5	샌드위치_NNG	0.3	브런치_NNG	0.2

[표 A.5-8] 문항세트 d 검색어 8  
(선택 기반 POI 검색, 토픽별 키워드 4개)

연번	토픽 번호				토픽				키워드							
	1	2	3	4	1	2	3	4	1	가중치	2	가중치	3	가중치	4	가중치
1	16		7		분위기			카페/디저트	음악_NNG	0.2	좋_VA	0.3	작_VA	0.3	카페_NNG	0.2
2	5				중식				푸짐_XR	0.2	맛있_VA	0.3	중국_NNP	0.3	요리_NNG	0.2
3	14				술집				안주_NNG	0.5	다양_XR	0.3	술집_NNG	0.1	추천_NNG	0.1
4	11				면류				라멘_NNP	0.4	국물_NNG	0.1	깔끔_XR	0.2	시원_XR	0.3
5	6				치킨/튀김류				치킨_NNG	0.2	돈까스_NNP	0.3	카레_NNG	0.4	맛있_VA	.01

문항세트 d에서 유사 POI 검색 결과에 대한 문항은 총 10개로 구성되며, 문항별 기준 POI는 [표 A.5-9]와 같다. 실제 오픈서베이를 통해 실시한 설문 화면은 지면 관계상 생략한다.

[표 A.5-9] 문항세트 d 검색어 9(유사 POI 검색)

연번	POI 명칭	주소	음식 종류	가격대
1	전통다원	종로구 관훈동 30-1	카페/디저트	1만원 미만
2	정돈	강남구 역삼동 811-4	까스요리	1-2만원
3	적사부	중구 중림동 463	정통 중식/일반 중식	1-2만원
4	219	서대문구 미군동 99-1	이탈리안	1만원 미만
5	고가빈커리하우스	종로구 내수동 110-48	인도 음식	1-2만원
6	전호메밀우동	관악구 봉천동 1664-14	국수/면 요리	1만원 미만
7	정가네 고려왕족발	동작구 흑석동 50-37	고기 요리	1-2만원
8	전주식당	중구 남창동 34-38	탕/찌개/전골	1만원 미만
9	전주옥한식당	중구 필동2가 17-1	탕/찌개/전골	2-3만원
10	전주콩나물국밥	용산구 한강로1가 150	탕/찌개/전골	-

## B.1 자연어 기반 POI 검색 결과 비교

본 연구에서 도출한 자연어 기반 POI 검색 결과와 기존 유사 서비스인 지도 중심 플랫폼(네이버 지도, 카카오맵, 구글 지도) 및 맛집 중심 플랫폼(망고플레이트, 다이닝코드)에서의 검색 결과를 비교 분석해 보았다. 비교에 사용한 검색 키워드는 ‘연인과 기념일’, ‘믿고 먹는 곱창집’, ‘바게트 맛 있는 베이커리 맛집’, ‘연인과 갈 만한 와인 바’, 그리고 ‘배경음악 선곡 센스 좋은 곳’이다. 각 검색 키워드에 대한 본 연구에서의 검색 결과는 [표 B.1-1]~[표 B.1-5]와 같으며, 동일 키워드에 대한 국내외 POI 검색 플랫폼에서의 결과는 [그림 B.1-1]~[그림 B.1-10]과 같다. 먼저, 지도 중심 플랫폼에서 ‘연인과 기념일’이라는 키워드로 검색을 실행한 결과인 [그림 B.1-1]을 보면 검색 결과가 출력되지 않은 것을 확인할 수 있다. 동일한 키워드로 맛집 중심 플랫폼에서 검색한 결과인 [그림 B.1-2]을 보면, 다이닝코드에서만 검색 결과가 도출된 것을 알 수 있다. 해당 결과 목록을 보면, 검색 키워드 중 ‘기념일’과 일치하는 태그가 포함된 POI들을, 다이닝코드 내부 알고리즘으로 계산한 POI별 점수가 높은 순서대로 반환해 준 것으로 파악된다. 즉, 검색어와 POI 간의 관련정보는 자체적으로 계산된 POI별 점수가 검색 결과에 더 많은 영향을 주는 것으로 보인다. 따라서 다이닝코드에서는 돈까스, 꼬막무침, 물회 등을 판매하는 POI들이 상위로 도출된 반면, 본 연구에서 ‘연인’에 0.5, ‘기념일’에 0.5의 가중치를 부여한 결과로는 스테이크/바비큐, 퓨전양식, 칵테일/와인을 판매하는 POI들이 상위로 도출된 것을 확인할 수 있다([표 B.1-1]). 본 연구에서의 검색 결과는 사용자가 키워드별로 입력한 가중치 값에 따라 달라질 수 있다.



(a) 네이버 지도

(b) 카카오맵

(c) 구글 지도

[그림 B.1-1] ‘연인과 기념일’ 검색 결과(지도 중심 플랫폼)

'연인과 기념일'에 대한 검색 결과가 없습니다.

검색한 식당이 망고플레이트에 보이지 않을 땐??

1. 망고플레이트 모바일 앱을 설치한다
2. '+'버튼을 눌러 식당 등록하기를 터치!
3. 등록할 식당의 정보를 입력한 후 등록 완료!



(a) 망고플레이트

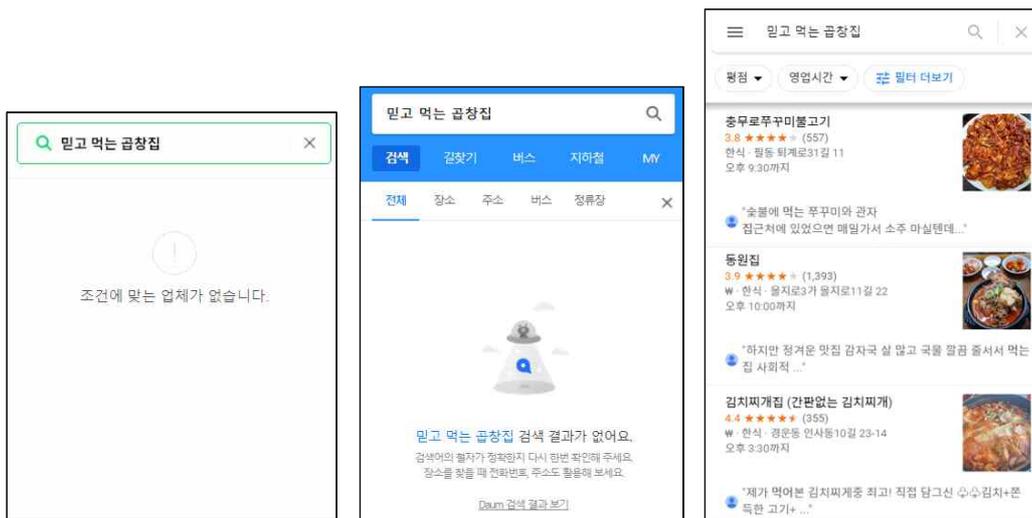
(b) 다이닝코드(일부)

[그림 B.1-2] '연인과 기념일' 검색 결과(맛집 중심 플랫폼)

[표 B.1-1] '연인(w: 0.5)', '기념일(w: 0.5)' 검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.77777	골든불	용산구 한남동 683-37	스테이크/바비큐	소중한날친구, 연인이 함께 스테이크와 와인을 즐길수있는스테이크 맛집~!! 기념일에 다시 찾아갈거예요^^
2	0.77307	랩24	강남구 청담동 100-14	퓨진 양식	코스로 진행되고 맛도 좋아요 기념일에 가기엔 딱입니다
3	0.74298	쿠킨 스테이크	종로구 관철동 12-12	스테이크/바베큐	저렴한 가격으로 분위기 내기에 좋은 곳 주머니 가벼운 우리 커플에겐 기념일 단골 식당ㅎ
4	0.73497	브이오	강남구 청담동 22-1	칵테일/와인	방으로 된 와인바임, 소믈리에분들이 넉넉히 자리하심. 모임용이지캐주얼스팟은일층만 있는듯
5	0.71452	헬로우	강남구 신사동 541-2	칵테일/와인	분위기 좋아요
6	0.71265	슈퍼마켓	용산구 한남동 620-163	기타 양식	분위기 씹오짐
7	0.70924	희망정	강남구 신사동 663-16	회/스시	진짜 깔끔하고 맛있네요 친구 연인사이에 와도 기분좋은곳
8	0.70630	로맨틱포차	광진구 화양동 9-76	칵테일/와인	소개팅 장소로도 너무 좋구! 친구끼리 분위기 좋은 레스토랑 갈때도너무 좋을꺼같아요!
9	0.69967	뚜르뒤뱅	서초구 반포동 96-8	칵테일/와인	아늑하고 조용해서 크리스마스에 오면 좋을 분위기의 와인바. 키워드: 와인/2차, 직장인/친구/연인, 로맨틱/아늑/어둑
10	0.69873	살롱드달구	강남구 신사동 517-25	칵테일/와인	조용하고 잔잔하고 분위기가 있었던곳

두 번째 자연어 기반 검색 키워드는 ‘믿고 먹는 곱창집’으로, 해당 키워드로 지도 중심 플랫폼에서 검색을 실행한 결과, 구글 지도에서만 검색 결과가 출력되었다([그림 B.1-3]). 구글 지도는 리뷰에 검색 키워드가 포함된 POI들을 반환하는데, 키워드별로 가중치를 부여할 수 없으므로 전체 검색 키워드 중에서 비교적 중요도가 낮은 ‘먹는’이라는 키워드가 포함된 POI들이 주로 도출된 것을 볼 수 있다. 이에 따라 ‘곱창’과 관련 없는 한식집이나 김치찌개를 판매하는 POI들도 결과에 포함되어 있었다. 동일 키워드로 맛집 중심 플랫폼에서 검색을 수행할 결과, 다이닝코드에서만 검색 결과가 반환되었다([그림 B.1-4]). 다이닝코드는 검색 키워드 중 ‘곱창’ 관련 태그가 포함된 POI들을, 다이닝코드 내부에서 부여한 POI별 점수가 높은 순서대로 반환해 준 것으로 보인다. 반면에, 본 연구에서 ‘믿고’에 0.5, ‘먹는’에 0.1, 그리고 ‘곱창’에 0.5의 가중치를 부여한 결과인 [표 B.1-2]를 보면, ‘믿고’와 ‘곱창’에 0.5의 가중치를 부여했기에 검색 결과로 모두 곱창을 판매하는 POI들이 출력된 것을 확인할 수 있다.



(a) 네이버 지도                      (b) 카카오맵                      (c) 구글 지도(일부)

[그림 B.1-3] ‘믿고 먹는 곱창집’ 검색 결과(지도 중심 플랫폼)

'민고 먹는 곱창집'에 대한 검색 결과가 없습니다.

검색한 식당이 망고플레이트에 보이지 않을 땐??

1. 망고플레이트 모바일 앱을 설치한다
2. '+'버튼을 눌러 식당 등록하기를 터치!
3. 등록할 식당의 정보를 입력한 후 등록 완료!



(a) 망고플레이트

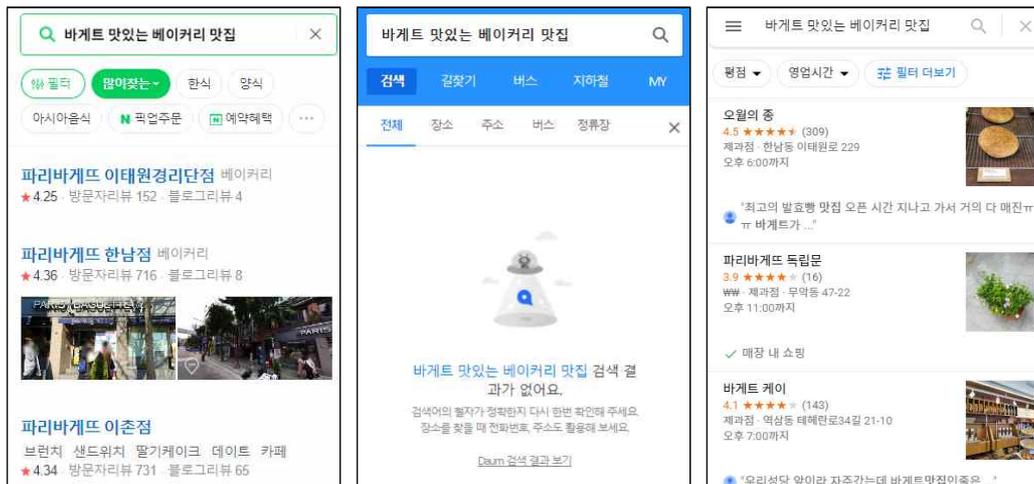
(b) 다이닝코드(일부)

[그림 B.1-4] '민고 먹는 곱창집' 검색 결과(맛집 중심 플랫폼)

[표 B.1-2] '민고(w: 0.5)', '먹는(w: 0.1)', '곱창(w: 0.4)' 검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.87781	옛날한우 곱창전문	강서구 등촌동 647-9	고기 요리	여기 양, 곱창은 삼각지 평양집과비교해도 뒤지지않음.
2	0.86443	원조신사 양곱창	강남구 신사동 510-11	고기 요리	맛있음!!!
3	0.85661	60년전통신촌황소곱창	강남구 역삼동 619-5	고기 요리	짱맛! 곱창만 먹으면 여기가 왜 유명...? 하는데 부추랑같이 드시면 짱맛!! 볶음밥도 맛있어용ㅎㅎ
4	0.85568	고바우 소곱창	노원구 상계동 735-2	고기 요리	먹어본 곱창중에가장 맛있었다.
5	0.85559	노루목 황소곱창	마포구 서교동 357-2	고기 요리	2년째 곱창먹으러가면노루목에감
6	0.85532	일번지 곱창대창	중구 을지로2가 199-64	고기 요리	남자친구랑 곱창먹으러다니는걸 무척 좋아해서 맛있다는 곳을 여기저기 찾아다니는데, 진짜 여기 정말 맛있어요.
7	0.85495	부추곱창	종로구 명륜4가 16-2	고기 요리	가본 곱창집중대학로에서는 여기가 제일 좋은 것 같다. 기본 안주(?) 서비스로 나오는 육회가 생각보다 너무 맛있어서 친구랑 따로 시켜먹었다ㅎ
8	0.85361	주부곱창	중구 신당동 242-4	고기 요리	화사때문에 먹고 싶었던 곱창!
9	0.85318	신촌 황소곱창	강동구 천호동 454-13	고기 요리	뭐 신촌황소곱창이야 내가 유일하게 민고 가는 곱창 체인점
10	0.85120	호남곱창	종로구 종로5가 225-27	고기 요리	곱창 좋아해서 자주 먹는데 정말 곱창이 신선하고 맛있었다 잡내 하나 나지 않았고

세 번째 자연어 기반 키워드는 ‘바게트 맛있는 베이커리 맛집’으로, 해당 키워드로 지도 중심 플랫폼에서 검색을 실행한 결과, 네이버 지도와 구글 지도에서 검색 결과가 출력되었다([그림 B.1-5]). 네이버 지도는 검색 키워드 중 ‘바게트’만을 인식했기 때문에 POI 명칭에 해당 키워드가 포함된 ‘파리바게트’만이 도출된 것으로 보인다. 구글 지도에서는 네이버 지도와 유사하게, POI 명칭에 검색 키워드가 포함된 ‘파리바게트’와 리뷰에 검색 키워드와 유사한 키워드인 ‘빵’과 ‘빵집’이 포함된 POI들을 함께 반환해 주었다. 동일 키워드로 맛집 중심 플랫폼에서 검색을 수행한 결과인 [그림 B.1-6]을 보면, 망고플레이트에서는 검색 결과로 2개의 베이커리 관련 POI만 출력된 것을 알 수 있다. 다이닝코드는 검색 키워드 중 ‘베이커리’를 ‘빵집’과 연결시켜 ‘바게트’ 및 ‘빵집’ 태그가 포함된 POI들을, 다이닝코드 내부에서 부여한 POI별 점수가 높은 순서대로 반환해 준 것으로 보인다. 반면에 본 연구에서 ‘바게트’에 0.4, ‘맛있는’에 0.3, ‘베이커리’에 0.2, 그리고 ‘맛집’에 0.1의 가중치를 부여한 결과인 [표 B.1-3]을 보면, ‘바게트’에 가장 높은 가중치 값을 부여했기에 리뷰에 해당 키워드 및 해당 키워드와 유사한 키워드들이 포함된 POI들이 출력된 것을 알 수 있다.



(a) 네이버 지도(일부)

(b) 카카오맵

(c) 구글 지도(일부)

[그림 B.1-5] ‘바게트 맛있는 베이커리 맛집’ 검색 결과(지도 중심 플랫폼)



(a) 망고플레이트



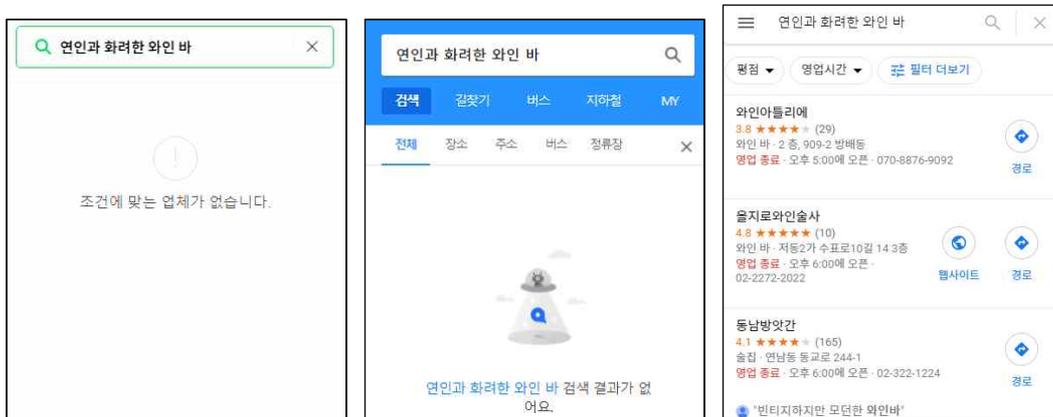
(b) 다이닝코드(일부)

[그림 B.1-6] '바게트 맛있는 베이커리 맛집' 검색 결과(맛집 중심 플랫폼)

[표 B.1-3] '바게트(w: 0.4)', '맛있는(w: 0.3)', '베이커리(w: 0.2)', '맛집(w: 0.1)' 검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.95044	1986베이커	강서구 등촌동 647-9	베이커리	겉바속촉바게트의 대명사...!
2	0.94993	바게트K	강남구 신사동 510-11	베이커리	바게트케이이미 소문난집이라서 더 덧붙일 것도 없지만 빵 담백하고 맛있어요~
3	0.93094	장티크	강남구 역삼동 619-5	카페/디저트	여기 바게트는 소리부터 다릅니다...
4	0.91118	브라운 브레드	노원구 상계동 735-2	베이커리	제일 사랑하는 빵집♡♡♡♡♡♡♡♡ 이대앞에있을때부터갔지만 여기만큼 바삭바삭한 바게트를 못봤다.
5	0.90923	김영모 과자점	마포구 서교동 357-2	베이커리	와 역시 빵 천국이네요... 다 너무 맛있어 보이고 다 사가고 싶었어요 ㅠㅠ
6	0.90865	블랑제르	중구 을지로2가 199-64	베이커리	언제 먹어도 맛있는 빵 감사합니다!
7	0.90851	코 베이커리	종로구 명륜4가 16-2	베이커리	블로그 맛집 같은거안민는데세상에 내가 먹어본 연유 바게트 중에 가장 맛있었다.
8	0.90743	블랑 제리코팡	중구 신당동 242-4	베이커리	이렇게 맛있는 바게트 처음 먹어봤다.
9	0.90572	디어브레드	강동구 천호동 454-13	베이커리	안암에서 빵으로 제일 맛있는 곳. 바게트, 치아바타가주력이며 굉장히 맛있다
10	0.90553	폴앤폴리나	종로구 종로5가 225-27	베이커리	갓나온 바게트가 제일 맛있는듯.

네 번째 자연어 기반 키워드는 ‘연인과 화려한 와인 바’로, 해당 키워드로 지도 중심 플랫폼에서 검색을 실행한 결과, 구글 지도에서만 검색 결과가 출력되었다([그림 B.1-7]). 구글 지도는 검색 키워드 중 하나인 ‘와인’이 POI 명칭에 포함된 POI들과 리뷰에 ‘와인바’가 포함된 POI들을 함께 반환해 주었다. 동일한 키워드로 맛집 중심 플랫폼에서 검색을 수행한 결과인 [그림 B.1-8]을 보면 다이닝코드에서만 검색 결과가 출력되었는데, 검색 키워드 중 ‘와인바’와 일치하는 태그가 포함된 POI들을, 다이닝코드 내부 알고리즘으로 계산한 POI별 점수가 높은 순서대로 반환해 준 것으로 파악된다. 반면에 본 연구에서 ‘연인’에 0.3, ‘화려한’에 0.3, 그리고 ‘와인’에 0.4의 가중치를 부여한 결과인 [표 B.1-4]를 보면, 각 키워드별로 거의 동일한 가중치를 부여했기에 ‘연인’, ‘분위기’, ‘와인’ 등 검색 키워드와 관련된 다양한 키워드가 포함된 POI들이 출력된 것을 알 수 있다.



(a) 네이버 지도

(b) 카카오맵

(c) 구글 지도(일부)

[그림 B.1-7] ‘연인과 화려한 와인 바’ 검색 결과(지도 중심 플랫폼)

'연인과 화려한 와인 바'에 대한 검색 결과가 없습니다.

검색한 식당이 망고플레이트에 보이지 않을 땐??

1. 망고플레이트 모바일 앱을 설치한다
2. '+'버튼을 눌러 식당 등록하기를 터치!
3. 등록할 식당의 정보를 입력한 후 등록 완료!



(a) 망고플레이트

(b) 다이닝코드(일부)

[그림 B.1-8] '연인과 화려한 와인 바' 검색 결과(맛집 중심 플랫폼)

[표 B.1-4] '연인(w: 0.3)', '화려한(w: 0.3)', '와인바(w: 0.4)' 검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.88909	뚜르뒤뱅	서초구 반포동 96-8	칵테일/ 와인	키워드: 와인/2차, 직장인/친구/연인, 로맨틱/아늑/어둑
2	0.87538	엘세뇨르둘세	강남구 청담동 128-6	칵테일/ 와인	북적거리는 청담동 쪽이 아닌 곳에 위치해서 상대적으로 조용하면서 어둑하니세련된 느낌의 분위기 있는 와인바:)
3	0.87494	뱅가	강남구 신사동 634-1	이탈리안	풍부한 와인리스트, 분위기에 더하는 라이브뮤직좋아요
4	0.86743	헬로우	강남구 신사동 541-2	칵테일/ 와인	분위기가 진짜 진짜좋은 와인바. 데이트로 딱이에요.
5	0.86609	비노바인	서초구 서초동 1540-5	이탈리안	와인바겸 비스트로. 와인샵을같이 운영해서 저렴하고 맛있는 와인과 괜찮은 음식을 즐길 수 있음.
6	0.85855	몽상가들	종로구 이화동 28-18	칵테일/ 와인	이화벽화마을 계단을 내려오면 보이는 분위기 좋은 와인카페다
7	0.85834	더젤	용산구 이태원동 241-2	칵테일/ 와인	우선 뷰가 너무 좋고, 봄비지 않아 조용하게 지는 해를 바라보며 대화하며 와인을 마실 수 있는 곳.
8	0.85676	에이오씨	강남구 청담동 95-3	이탈리안	분위기 좋고 와인, 파스타도 괜찮아요!
9	0.85502	하프파스트텐	용산구 이촌동 302-46	칵테일/ 와인	사장님이 소믈리에대회에서 우승하신 분이라 와인 추천을 아주 잘해주심.
10	0.85290	When	마포구 연남동 390-28	세계음식 기타	연남동 반지하와인바. 참 아늑한 분위기다.

마지막 키워드는 ‘배경음악 선곡 센스 좋은 곳’으로, 해당 키워드로 지도 중심 플랫폼에서 검색을 실행한 결과, 모든 플랫폼에서 결과가 출력되지 않았다([그림 B.1-9]). 동일한 키워드로 맛집 중심 플랫폼에서 검색을 수행한 결과인 [그림 B.1-10]을 보면, 다이닝코드에서만 검색 결과가 반환되었는데 POI 명칭에 ‘센스’가 포함된 POI들을 도출해 주었다. 반면 본 연구에서 ‘배경음악’에 0.4, ‘선곡’에 0.2, ‘센스’에 0.2, 그리고 ‘좋은’에 0.2의 가중치를 부여한 결과인 [표 B.1-5]를 보면, ‘배경음악’에 가장 높은 가중치를 부여했기에 리뷰에 해당 키워드 및 해당 키워드와 유사한 키워드들이 포함된 POI들이 출력된 것을 알 수 있다.



[그림 B.1-9] ‘배경음악 선곡 센스 좋은 곳’ 검색 결과(지도 중심 플랫폼)



[그림 B.1-10] ‘배경음악 선곡 센스 좋은 곳’ 검색 결과(맛집 중심 플랫폼)

[표 B.1-5] '배경음악(w: 0.4)', '선곡(w: 0.2)', '센스(w: 0.2)',  
'좋은(w: 0.2)' 검색 결과(상위 10개)

순위	관련도	POI 명칭	주소	음식 종류	대표 리뷰
1	0.90892	제이앤 제이슨	강남구 신사동 656-25	칵테일/와인	일단 여기 음악이 힙합인데 아주 음악 선곡이 끝내줍니다ㅋㅋ분위기도 좋고요
2	0.90818	아트포 라이프	종로구 부암동 29-4	카페/디저트	음악에 조예가 깊은 사장님이 계신 곳. 분위기도 좋고와인가격도착해요.
3	0.90716	인더무드 포러브	강남구 청담동 119-5	칵테일/와인	멋있는 분위기에 좋은술 분위기가 너무 특이하고 노래도 좋았다
4	0.89758	샹들리에	마포구 연남동 228-28	칵테일/와인	분위기가 좋고친절하신 편입니다.
5	0.89086	Ando Vice Versa	용산구 한남동 736-8	칵테일/와인	안도카페가운영하는 바. 분위기가 좋아데이트하기 좋다.
6	0.88681	스테어웨이	용산구 이태원동 112-2	칵테일/와인	분위기 좋고 나오는 음악도 다 너무 좋았던 곳
7	0.88671	그날이오면	서대문구 창천동 53-24	칵테일/와인	분위기 너무 좋고사장님 친절도 너무 좋고 음악도 너무 좋아요
8	0.88653	서울바이닐	용산구 용산동2가 45-13	치킨/호프/뽕	2차로 방문한 곳인데 분위기와 음악이 특히 좋았다.
9	0.88570	카페몽돌	동작구 대방동 394-1	카페/디저트	턴테이블위에 올려진 레코드판에서 나오는 음악이 레트로한분위기를 풍기는 이 카페에 들러 차 한잔과 간식거리 주문하고 비치된 인형을 보는 것만으로도 심심찮게 시간을 보낼 수 있을 것 같다.
10	0.88532	ILZAJIP coffee	광진구 자양동 552-3	카페/디저트	음악도 잔잔해서 편안하게 시간 잘 때웠습니다ㅎㅎ커피맛이나가격도 물론 좋았구요.

Abstract

# Topical Keywords based Semantic POI Retrieval using Review Texts

Youngmin Lee

Department of Civil and Environment Engineering

The Graduate School

Seoul National University

Retrieving a specific place or location in location based and web map services is beneficial in everyday life and is one of the most accessed services by users. These services evolve into customized services with the recent development of big data and artificial intelligence technologies and point of interest(POI) data are crucial for providing these retrieval services. The existing domestic and overseas POI retrieval services mainly provide a retrieval function focusing on POI name, category, tag and some support a review retrieval function. However since the semantic information of the POI is not used, it is not easy to reflect users' various situations and preferences when retrieving for a POI.

This study aims to develop a semantic POI retrieval technology

based on keywords by topics that reflect users' various requirements. To this end we proposed a process of generating semantic POIs using unstructured review text data written by users about POIs and designed three types of structures that can utilize semantic POIs for retrieval. First, we present a data collection and pre-processing method and then propose a topic modeling method based on latent Dirichlet allocation(LDA) for extracting topics and keywords for each topic from the review text. We finally propose a method of embedding POI by embedding keywords for each topic based on the Word2vec model. This process makes it possible to determine a semantic POI that reflects the characteristics of various keywords used based on the topic. In addition we propose a utilization structure for natural language based POI retrieval of natural language based terms for selection based POI retrieval by selecting topics and keywords for each topic to retrieve POI, and for similar POI retrieval that can retrieve POIs similar to a specific POI.

We surveyed 400 men and women residing in Seoul to verify the semantic POI retrieval results. The average satisfaction was calculated as 3.48 out of 5, which was within the range of satisfaction based on the study's evaluation criteria. In terms of user satisfaction, 'satisfied' was the most frequent response. On average, 'very satisfied' and 'satisfied' were selected about 4.1 times more than 'very unsatisfied' and 'unsatisfied', confirming the practicality of the semantic POI retrieval technology developed in this study. Lastly, a comparison of the semantic POI retrieval results of this study and those of existing POI retrieval services confirmed this study's differences. Specifically, with the semantic POI retrieval technology developed in this study, a user can directly designate a weight for a retrieval word. Any retrieval word can be customized by displaying different results according to the

weight. In addition, results based on synonyms can be derived, even when no keyword exactly matches the retrieval word. Even various meanings of keywords that differ depending on the topic can be used for retrieval. Therefore, through the semantic POI retrieval technology developed in this study, users' various preferences and situations can complement the existing retrieval system.

**keywords : semantic POI, POI retrieval, POI embedding,  
customized retrieval, Word2vec, LDA**

*Student Number : 2014-30242*