



데이터 마이닝 프로젝트

- 로봇 서비스를 제공하는 호텔들의 리뷰를 중심으로 토픽 모델링 -

3 조

17010254 조병주

18010161 박 훈

19010248 하온비

목 차

1. 비즈니스에 대한 이해

- 1) 주제 선정 배경
- 2) 데이터 마이닝 목적
- 3) 기대효과

2. 데이터 이해

- 1) 수집해야 하는 데이터
- 2) 데이터 수집 및 분석 방안
- 3) The readiness level of data

3. 모델링 및 분석 결과 보고서

- 1) 리뷰데이터 수집 및 전처리
- 2) EDA 탐색적 데이터 분석
- 3) LDA 토픽 모델링 및 결과

4. 결론

0. 주제

토픽 모델링을 통해 로봇을 활용하는 호텔들의 리뷰 분석

1. 비즈니스에 대한 이해

1) 주제 선정 배경

■ 4 차 산업 혁명과 호텔 산업

4 차 산업혁명이란 인공 지능(AI), 인터넷(IoT), 클라우드 컴퓨팅, 빅데이터, 모바일 등의 기술이 기존 산업과 서비스에 융합되거나 3D 프린팅, 로봇공학, 생명공학, 나노기술 등 여러 분야의 신기술과 결합되어 모든 제품, 서비스를 네트워크로 연결하고 사물을 지능화한 것을 말한다. 현재 여러 산업 분야에서 이러한 4 차 산업혁명 기술들을 적용하려는 시도가 계속되고 있으며, 이는 호텔 산업 또한 예외는 아니다. 호텔 산업은 호텔 내에 4 차 산업 혁명 기술을 접목하여 서비스를 제공하려는 시도를 해왔다. 예를 들어 클라우드 컴퓨팅 기술의 경우 호텔의 예약 관리 및 고객 관리에 쓰인다. 또, 객실 내에서 커튼을 열고 닫거나, 알람을 맞추거나 하는 서비스가 IoT 서비스를 통해 제공된다. 룸서비스 배달 및 식음료 서빙 등의 서비스도 로봇을 통해 제공한다.

이러한 기술을 제공하는 국내 기업으로는 ‘KT enterprise’가 있다. KT enterprise 의 ‘AI 호텔’과 ‘AI 호텔 로봇’ 서비스는 호텔 산업에 KT의 4 차 산업 기술들을 접목한 서비스이다. 첫 번째로 ‘AI 호텔’의 서비스로는 크게 2 가지가 있다. 먼저 전용 AI 단말기를 통해 고객은 룸서비스를 요청할 수 있다. 또 고객이 음성을 통해 AI 단말기에 명령을 한다면 단말기가 음성을 인식하여 IoT 화 된 객실 내 시설을 제어한다. 이 서비스를 사용하는 호텔 브랜드로는 노보텔, 레스케이프, 엠버서더, 베이몬트, 해밀턴, 메이필드 등이 있다.

다음으로 ‘AI 호텔 로봇’ 서비스는 AI 호텔 로봇을 통해 호텔 어메니티를 배달해주는 서비스이다. 만약 객실에서 앞서 설명한 단말기를 통해 어메니티를 요청한다면, 단말기는 이 정보를 호텔 관리 서버로 전송한다. 이를 관제실에서 확인한 다음 하우스 키퍼가 로봇에게 요청사항을 전달하고 로봇이 객실로 룸서비스를 제공하는 방식이다. KT에서는 이 호텔 로봇을 이용한다면 1 인 인건비 미만의 비용으로 1~1.5 명의 인력을 서포트할 수 있을 거라고 기대한다. 현재 AI 호텔 로봇을 상용화한 호텔로는 서울 동대문구에 위치한 ‘노보텔 엠버서더 서울 동대문 호텔 & 레지던스’가 있다.

4 차 산업 기술을 적용한 호텔 중에는 세계 첫 로봇 호텔로 유명한 일본의 ‘헨나 호텔’이 있다. 헨나 호텔은 2021 년 한국 서울에 명동점을 개관했다. 헨나 호텔 또한 KT 의 ‘AI 호텔’ 서비스를 사용하고 있다. 헨나 호텔의 가장 큰 특징은 체크인부터 체크아웃까지 모든 과정에서 사람 대신 로봇이 그 일을 대신한다는 점이다. 고객은 프런트 데스크 로봇을 통해 체크인과 체크아웃 수속을 진행하고, 키오스크에서 카드키를 발급받을 수 있다. 객실에서는 기가 지니를 활용해 룸서비스를 주문하면 이를 룸 딜리버리 로봇이 객실까지 배달한다.

■ 호텔 산업에서 리뷰가 미치는 영향

온라인 여행사(OTA)는 예비 투숙객이 호텔에 대한 정보를 찾기 위해 주로 이용하는 곳이다. 스마트폰의 등장으로 대부분의 사람들이 언제 어디서든 온라인에 접속할 수 있는 현재에는 OTA 에서의 평가가 예비 투숙객의 의사 결정에 미치는 영향이 매우 크다. 예비 투숙객이 정보를 찾을 온라인은 오프라인에서 정보를 찾을 때 보다 시간과 비용을 줄이면서 다양한 정보에 대해 보다 비교를 편리하게 하고, 더 다양하고 많은 정보를 찾을 수 있게 해준다. 시간이 갈수록 소비자들은 온라인 채널을 사용하여 제품 정보를 수집할 가능성이 높아졌다. 이를 통해 기존의 소비자 경험을 참고할 수 있으므로, 구매 전 탐색과정에서 온라인 정보가 호텔 산업의 고객들에게 미치는 영향력은 더욱 커지고 있다.

■ 조사 대상 호텔

호텔명	노보텔 앰버서더 서울 동대문 호텔&레지던스	레스케이프 호텔	헨나 호텔
적용된 서비스	KT AI 호텔, KT AI 호텔 로봇	KT AI 호텔	KT AI 호텔, KT AI 호텔 로봇
등급	5성급	4성급	미정(3등급 이하 추정)
가격대	260,000 원 이상	160,000 원 이상	90,000 원 이상
객실 수	523 객실	204 객실	122 객실
개관년도	2018	2018	2021

2) 데이터 마이닝 목적

위에서 볼 수 있듯이 호텔 산업 내에서 4 차 산업 기술을 적용하려는 여러 시도가 있었다. 위에서 설명한 “이러한 시도를 해당 호텔을 묵은 고객들의 평가는 어떠할까?” 라는 궁금증을 해결하기 위해 데이터 마이닝을 해보자 한다.

■ 구체적으로 확인할 내용 혹은 가설

로봇 호텔을 사용하는 호텔과 해당 서비스를 제공하는 기업 입장에서 쓰여진 기사를 보게 되면, 고객들과 호텔 내의 직원들이 만족한다고 쓰여 있다. 실제로 메이필드 호텔 주임의 인터뷰에 따르면, 로봇이 가져다주는 것 자체가 콘텐츠로 기능하기에 어린이와 어른 모두 긍정적인 반응을 보인다고 한다. 또 인건비 문제로 최소 근무 인원만 남겨둔 상황에서 딜리버리를 로봇이 대신하기 때문에 직원들의 반응도 긍정적이라고 한다. 그렇다면 이러한 반응이 정말 사실일까? 5성급 호텔을 묵은 고객들이 사람이 아닌 로봇이 제공하는 서비스에 100% 만족할 수 있을까? 고객의 입장에서 사람이 해주는 서비스가 아니니 섬세하지 못하다고 생각해 불만족하지 않을까? 호텔의 고객들이 로봇, 인공지능에게 기대하는 서비스는 무엇일까? 서비스에 대한 고객의 기대를 호텔 로봇이 만족시킬 수 있을까? 하는 여러 의문들에 대해 우리가 선정한 세 개의 호텔에 대한 다양한 OTA 의 리뷰를 통해 구체적으로 확인해보고자 한다.

3) 기대효과

4 차 산업 혁명의 기술 중 로봇 공학 기술과 AI, IoT 기술이 적용된 호텔들의 리뷰들을 통해 호텔 산업에서의 디지털 트랜스포메이션에 대한 고객들의 반응을 확인해본다. 고객들의 리뷰를 분석하여 고객들이 실제로 만족하는 로봇 및 인공지능 서비스가 무엇인지 확인하고, 이를 토대로 앞으로 호텔 산업의 바람직한 디지털 트랜스포메이션에 대해 방향성을 제시할 수 있을 것이다.

2. 데이터 이해 : 데이터 수집 및 분석 방안

1) 수집해야 하는 데이터

우리는 로봇 서비스를 활용하는 호텔 중 ‘KT AI 호텔’ 서비스를 활용하는 호텔에 대한 OTA 리뷰를 조사하려고 한다. 그 중에서도 ‘노보텔 앰배서더 서울 동대문 호텔 & 레지던스’ ‘레스케이프 호텔’, ‘헨나 호텔’을 조사할 것이다.

먼저 ‘노보텔 앰배서더 서울 동대문 호텔 & 레지던스’를 조사하는 이유는 ‘KT AI 호텔 로봇’ 서비스를 최초로 상용화한 호텔이기 때문이다. 2019 년 12 월 최초로 상용화에 성공했다. 다음으로 ‘레스케이프 호텔’을 조사한 이유는 ‘KT AI 호텔’ 서비스를 이용하는 호텔 중 하나이고, 5성급인 노보텔 앰배서더 동대문과는 다르게 4성급 호텔이기 때문이다. 가격대 또한 10만원정도 차이가 나기에 이러한 차이가 고객들의 리뷰에도 유의미한 영향을 미쳤을 것이라 생각한다. 마지막으로 ‘헨나 호텔’을 조사하게 된 이유는 세계 최초의 로봇 호텔임과 동시에 ‘KT AI 호텔’ 서비스를 이용하고, 중국 업체의 호텔 로봇을 사용하기 때문이다. 또 아직 호텔 등급이 정해지지 않았고, 가격대 또한 9만원대로 우리가 조사하는 호텔 중에 가장 낮기 때문에 위 두 개의 호텔과 차이를 보여주는 리뷰를 얻을 수 있을 거라고 생각한다. 우리가 분석해야 할 데이터는 리뷰 데이터로 이 데이터의 유형은 비정형 데이터 중에 텍스트 데이터이다.

2) 데이터 수집 및 분석 방안

데이터를 분석하기 위해서는 먼저 데이터 수집을 해야 한다. 데이터 수집은 python 의 selenium 라이브러리를 활용해 해당 호텔의 OTA 리뷰 데이터를 수집한다. 텍스트 데이터 전처리를 위해 형태소 분석기인 KoNLPY 의 mecab 을 활용하여 형태소 분석을 진행한다. 위 과정을 통해 수집 및 전처리된 리뷰를 토픽 모델링을 통해 분석할 것이다. 토픽 모델링은 문서의 집합에서 토픽을 찾아내는 프로세스를 말한다. 단어가 특정 토픽에 존재할 확률과 문서에 특정 토픽이 존재할 확률을 통해 토픽을 추출한다. 이는 검색 엔진, 고객 민원 시스템 등과 같이 문서의 주제를 알아내는 일이 중요한 곳에서 사용된다. 하여 리뷰에서 해당 호텔을 경험한 고객의 만족도 및 로봇 경험에 대한 분석 결과를 도출해내려는 우리의 주제에 토픽 모델링이 적합하다.

위의 방법을 통해 수집한 데이터를 분석할 때, 크게 두 가지의 분석결과가 도출될 수 있다. 첫번째는 리뷰 데이터 내에 ‘로봇 경험’에 대한 언급이 없는 경우인데, 이러한 경우에는 해당 호텔에서 실제로 로봇을 통한 서비스가 이루어지고 있는지에 대한 확인이 필요하다. 언급이 없는 경우 생각해볼 수 있는 다른 상황은 앞서 살펴본 기사와는 다르게 실제로는 로봇

서비스가 고객에게 인상적인 요소로 작용하지 않았다는 것일 수도 있다. 때문에 로봇 서비스가 실효성이 있는지에 대해 조사가 필요할 것이다. 두번째는 리뷰 데이터에서 ‘로봇 경험’에 대한 내용을 확인할 수 있는 경우이다. 해당 경우에는 함께 도출되는 다른 단어들과 함께 고객의 반응을 정성적으로 분석할 수 있을 것이다. 예를 들어 로봇 서비스와 긍정적인 반응이 함께 나오는 경우, 해당 호텔에 성공적으로 적용되고 있다는 의미이기 때문에 다른 호텔에 이 로봇 서비스가 어떻게 적용될 수 있는지 추가적으로 연구해 볼 수 있다. 로봇 서비스와 부정적인 반응을 나타내는 단어가 함께 나오는 경우, 로봇 서비스의 개선이 필요함을 의미하기 때문에 이와 관련된 구체적인 개선 사항에 대해 고민할 필요가 있다.

3) The readiness level of data

■ Data content accuracy

본 프로젝트에 필요한 데이터는 위에 언급한 세 개 호텔에 대한 리뷰이다. 리뷰의 호텔에 대한 고객들의 평가에는 AI 로봇과 관련된 서비스를 제공받은 경험이 포함될 것으로 기대한다. AI 로봇과 관련된 서비스에는 로봇이 룸서비스를 배달하는 경우가 많을 것이므로 룸서비스에 대한 리뷰를 참고할 필요가 있다. 만약 로봇과 관련된 리뷰를 찾을 수 없다면 호텔에서 로봇 서비스가 제대로 이루어지고 있지 않는 상황으로 해석할 수도 있을 것이다. 또 리뷰 데이터가 긍정 혹은 부정 한 쪽 극단으로 치우쳐지는 현상에 대해서도 유의하며 데이터 분석을 진행해야 한다. 지나치게 긍정적인 리뷰만 많거나, 부정적인 리뷰만 많다면 그 리뷰에 대한 신뢰성이 떨어질 것이다. 이러한 부분을 유의하여 분석을 진행해야 한다.

■ Data accessibility

본 프로젝트에서 필요한 데이터의 경우 OTA 에 호텔을 검색하거나, Search Engine 을 통해 얻을 수 있을 것으로 기대한다. 이렇게 오픈 되어 있는 OTA 의 리뷰를 수집할 경우 리뷰의 소유권이 기업에게 있는 것은 아니지만, 개인을 특정할 수 있는 신상정보가 드러나지 않게 수집해야 한다는 점을 유의해야 한다.

■ Data richness

리뷰 데이터를 얻을 때 다양한 OTA 에서 리뷰를 크롤링 해야 한다. 한가지 OTA 만을 조사할 경우 해당 OTA 를 선호하는 고객들의 리뷰만을 반영하게 되어 토픽이 특정 토픽으로 편향될 수 있다. 따라서 여러 OTA 의 리뷰를 참고해야 한다. 예를 들어 ‘Tripadvisor’를 이용할 경우 ‘노보텔 엠베서더 서울 동대문 호텔’, ‘레스케이프 호텔’의 리뷰는 쉽게 접근할 수 있었으나, ‘헨나 호텔’에 대한 리뷰는 찾아볼 수 없었다. 하지만 ‘Expedia’를 이용할 경우 세

호텔에 대한 리뷰에 쉽게 접근할 수 있었다. 이처럼 각 OTA마다 호텔 리뷰의 수나 종류가 다르기 때문에 다양한 OTA에 접근하여 데이터를 수집해야 한다.

■ Data consistency

세 호텔의 리뷰에 대해 크롤링할 때 다양한 OTA를 이용하게 된다. OTA 별로 리뷰의 형식이 다를 수 있고, 리뷰의 양 또한 차이가 날 수 있다. 만약 평점의 있는 경우, 실제 OTA의 평점 분포 비율이 반영될 수 있게 리뷰를 수집해야 하고, 이를 유지해야 한다. 또 행과 열 값이 OTA마다 다를 수 있는데, 모두 병합할 때 결측치나 열의 개수 등을 고려해서 병합하여야 한다.

■ Data currency/ data timeliness

데이터를 크롤링할 때 모든 리뷰를 크롤링할 것이므로, 리뷰 데이터의 작성이 시작된 시점은 호텔 개관일 시점일 것이다. ‘노보텔 엠베서더 서울 동대문 호텔 & 레지던스’는 2019년 12월에 AI 호텔 로봇서비스를 상용화하였다. 하여 이 시점 이전과 이후 데이터를 비교하며 분석한다면, 유의미한 결과를 도출해낼 수 있을 것이다. ‘레스케이프 호텔’은 2018년 7월 19일에 호텔 AI 서비스와 함께 개관하였다. 또 ‘헨나 호텔’은 2021년에 8월 1일에 AI 호텔 서비스와 호텔 로봇 서비스와 함께 개관하였다. 위 두 호텔은 AI 서비스가 도입되기 전후 차이를 비교하기에는 어려울 것이다.

■ Data granularity

리뷰 데이터 크롤링을 통해 구체적으로 수집할 데이터는 다음과 같다; “OTA 종류 / 리뷰 작성 날짜(년/월/일) / 호텔 이름 / 리뷰 내용 / 작성자 아이디 / 고객이 남긴 평점”. 먼저 OTA 특성상 해당 플랫폼을 이용하는 주 고객층이 달라질 수 있으며, 이에 따라 리뷰 내용도 OTA 별로 다르게 나타날 가능성이 존재한다. 또한 시간이 지남에 따라 호텔의 서비스가 개선 또는 악화될 수 있으며, 동일한 작성자가 특정 의도를 가지고 동일한 내용의 리뷰를 반복해서 업로드할 수 있다. 따라서 LDA를 위해 필요한 데이터 외에도, 위의 내용들을 고려하기 위해 크롤링한 데이터를 세분화하여 다른 데이터로 저장해야 한다.

■ Data relevancy

본 프로젝트를 통해 호텔에서 4차 산업 혁명의 기술 중 AI와 로봇이 제공하는 서비스가 고객 만족도에 미치는 영향에 대해 파악해보고자 한다. 호텔 등급별로 고객이 기대하는 서비스의 질이 다를 것이라고 예상하며, 구체적인 비교를 위해 호텔 등급 및 숙박 평균 가격을 고려해줄 것이다. 로봇 서비스와 고객 만족도 간의 관련성을 파악하기 위해 리뷰 데이터를 활용할 것이며 이를 LDA를 통해 분석해보고자 한다.

3. 모델링 및 분석 결과 보고서

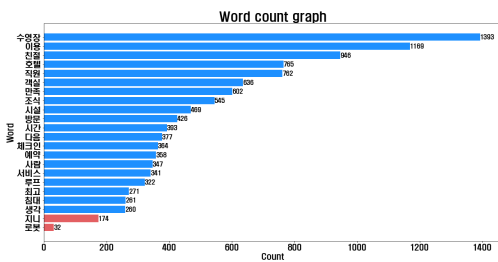
1) 리뷰 데이터 수집 및 전처리

본 연구 대상인 ‘노보텔 앰버서더 서울 동대문 호텔 & 레지던스’(이하 노보텔 호텔), ‘레스케이프 호텔’, ‘헨나 호텔’의 이용객 리뷰를 다양한 OTA 플랫폼에서 스크래핑했다. 수집 당일인 2022년 11월 25일을 기준으로 booking.com, expedia, 여기어때, trip.com, 야놀자, hotelscombined 에서 호텔의 모든 리뷰 데이터를 수집했다. 이를 통해 노보텔 호텔에 대해 총 3914 개의 리뷰를 수집하였고, 레스케이프 호텔에 대해 1914 개, 헨나 호텔에 대해 508 개의 리뷰를 수집했다. 2021 년 오픈한 헨나호텔의 리뷰가 다른 호텔에 비해 상대적으로 적다. 하지만 본 연구의 목적은 지정된 호텔의 리뷰들의 토픽을 파악하려는 것이기 때문에 크게 문제가 되지 않을 거라고 판단하고 자료수집을 마무리하였다.

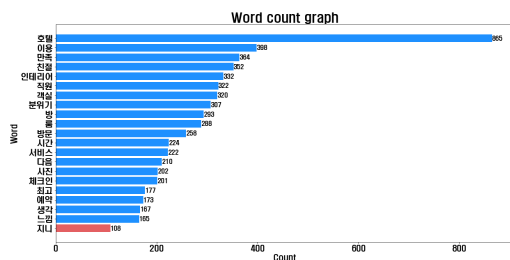
리뷰 데이터는 비정형 데이터 중 텍스트 데이터의 일종이다. 컴퓨터는 일반적으로 우리가 쓰는 형태의 문장을 이해하지 못하기 때문에 전처리 과정이 필요하다. 텍스트 전처리는 주어진 텍스트에서 노이즈와 같이 불필요한 부분을 제거하고, 문장을 단어로 분리하는 것까지를 의미한다. 먼저 정제과정을 통해 분석에 불필요한 노이즈를 제거해주었다. 정규표현식을 활용해 escape sequences, 이모티콘, nan 등을 노이즈로 판단하고 제거했다. 다음으로 네이버 맞춤법 검사기를 이용한 파이썬용 한글 맞춤법 검사 라이브러리인 hanspell 을 사용하여 노이즈가 제거된 리뷰 데이터에 맞춤법 검사 및 수정을 진행했다. 다음으로 konlpy 라이브러리가 제공하는 형태소 분석기 중 Mecab 을 이용하여 문장 형태의 리뷰 데이터를 단어 형태로 토큰화 해주었다. 토픽 모델링에 적합하다고 생각되는 명사만을 반환하게 했다. 위 과정을 거쳐 노보텔 호텔의 리뷰 데이터는 35638 개의 토큰으로, 레스케이프 호텔은 21926 개의 토큰으로, 헨나 호텔은 4363 개의 토큰으로 전처리 되었다.

2) EDA(Exploratory Data Analysis, 탐색적 데이터 분석)

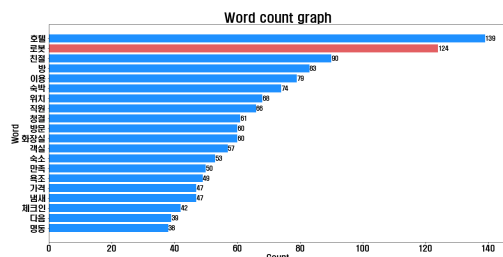
전처리된 데이터를 바탕으로 EDA 를 진행했다. EDA 는 탐색적 데이터 분석으로 데이터를 탐색하고 이해하는 과정이다. EDA 를 위해 전처리된 데이터를 가지고 단어 별 빈도수 (상위 20 개 단어 및 로봇, 인공지능 등의 단어)를 나타내는 그래프를 그려보고, 한 눈에 파악하기 위해 워드 클라우드를 그려보았다.



위 그래프는 노보텔 호텔의 전체 데이터에 대한 단어 빈도수 그래프 및 워드 클라우드이다. 로봇에 대한 언급이 많을 거라고 기대했던 것과는 다르게 로봇에 대한 언급 빈도는 35638 개 중 32 회로 저조했다. 하지만 지니에 대한 언급이 174 회로 로봇에 비해 상대적으로 많은 것을 알 수 있다. 단순히 리뷰 내 언급된 빈도만을 기준으로 생각했을 때, 노보텔 호텔을 찾는 고객들은 노보텔 호텔을 이용하고 나서 AI 로봇 서비스가 리뷰에 남길만큼 인상깊진 않았거나, 해당 서비스가 활성화되지 않아 호텔을 투숙하며 못 경험해봤을 확률이 높다는 것을 추측할 수 있다.



다음으로 레스케이프 호텔의 단어 빈도수 그래프와 워드 클라우드이다. 레스케이프 호텔은 AI 호텔 서비스만을 상용 중인 호텔이다. 하여 로봇에 대한 언급은 없다. 대신 KT AI 호텔 서비스를 상용 중인데, KT 의 인공지능 스피커인 ‘지니’에 대한 언급은 21926 개 중 108 번으로 비율 상 노보텔의 AI 서비스에 비해 높지만 기대한 것만큼 큰 수치는 아니다.



마지막으로 헨나 호텔의 단어 빈도수 그래프와 워드 클라우드이다. 헨나 호텔의 그래프를 보면 호텔 다음으로 로봇에 대한 언급 빈도가 위의 두 호텔보다 많은 것을 볼 수 있다. 헨나 호텔의 로봇 서비스가 고객들에게 인상 깊은 서비스였다는 것으로 추측해 볼 수 있다.

다른 시점에서 리뷰의 단어 빈도수에 대해 분석해보자면 노보텔 호텔과 레스케이프 호텔의 핵심 서비스가 AI 호텔, AI 로봇이 아니라 수영장, 인테리어 등이라 해당 호텔을 방문하는 고객들의 방문 동인이 수영장과 인테리어라는 것으로 해석할 수 있고, 헨나 호텔의 경우에는 핵심 서비스가 AI 스피커, 호텔 로봇이라 헨나 호텔을 방문하는 고객들은 AI 스피커, 호텔 로봇을 경험하러 방문하는 것이라는 해석을 할 수 있다.

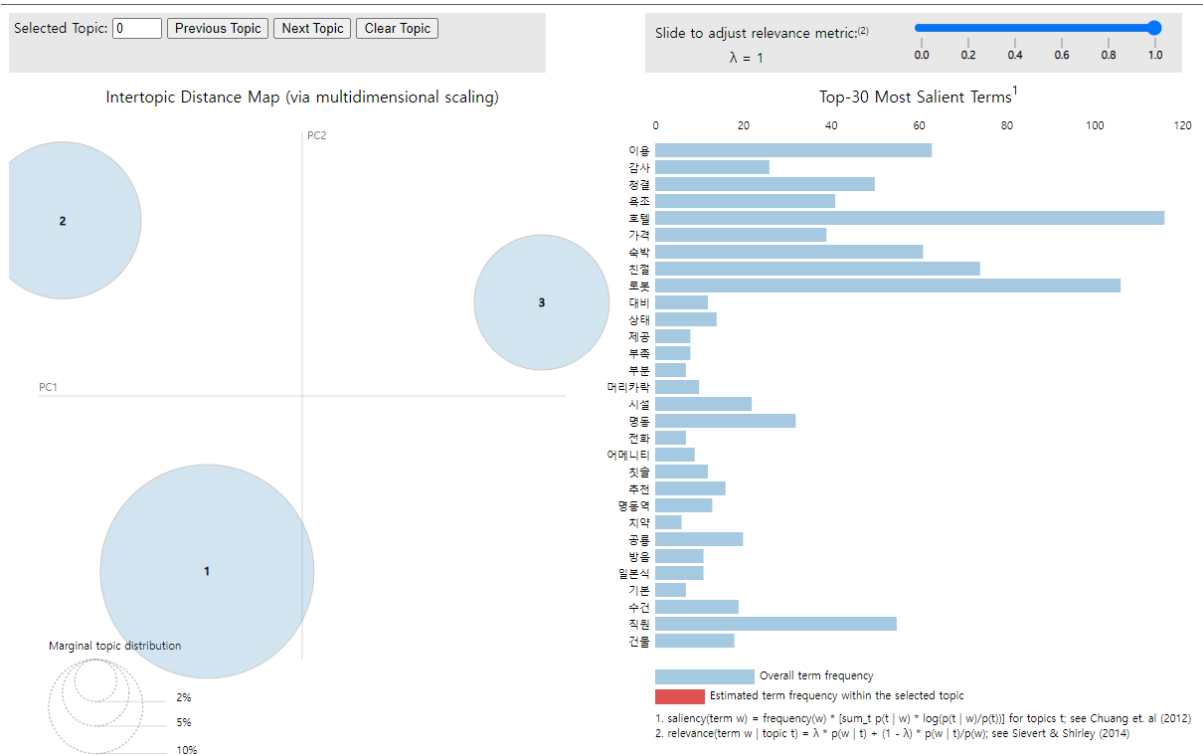
3) LDA(Latent Dirichlet Allocation) 토픽 모델링

LDA 는 토픽 모델링의 대표적인 알고리즘이다. LDA 는 단어와 단어(토큰)사이의 순서는 고려하지 않고, 빈도수를 기반으로 토픽을 추정한다. LDA 는 토픽에 대해 ‘문서를 작성할 때 어떤 토픽을 정해서 문서를 작성할 것이고, 해당 토픽에는 특정한 단어가 포함될 것이다’ 라는 가정을 하여 각 문서가 어떤 토픽에 속할 확률과 특정 토픽에 어떤 단어의 확률을 구하는 것을 여러 번 반복하여 최종적으로 토픽을 구한다. LDA 에서 하이퍼 파라미터로 토픽의 개수가 있는데 이를 성능이 높아지는 방향으로 조정해줘야 한다. LDA 의 성능을 평가하는 지표에는 Perplexity 와 Coherence 가 있다. Perplexity 는 혼란도, 혼잡도로 실제 토픽과 얼마나 유사하게 예측하는지 평가한다. Coherence 는 응집도로 각 토픽에서 상위 비중을 차지하는 단어들이 의미적으로 얼마나 유사한지를 나타낸다. 이번 연구에서는 Coherence 를 우선적으로 반영하여 헨나 호텔, 레스케이프 호텔, 노보텔 호텔 모두 3 개의 토픽을 갖도록 설정했다. LDA 를 시각화 하는 방법으로 LDAvis 라는 라이브러리를 사용하여 토픽들 간의 관계를 보았고, 워드클라우드를 활용해 토픽에서 단어가 갖는 비중을 직관적으로 관찰하였다. LDAvis 는 차원 축소 방법인 PCA 와 키워드 추출 방법을 이용하여 LDA 를 통해 알아낸 토픽 간의 관계와 토픽 키워드를 손쉽게 이해할 수 있도록 도와준다. 좌측의 좌표평면은 토픽들 간의 관계를 보여주고, 우측의 수평막대그래프는 해당 토픽에서 단어가 가지는 비율을 보여준다.

EDA 과정에서 확인했듯이 리뷰에 대한 단어 빈도 기반 그래프에서 유의미한 ‘로봇’ 빈도를 보였던 헨나 호텔 부터 LDA 결과를 확인하고 정성적으로 분석해보았다.

3-1) 헨나 호텔 LDA 결과

헨나호텔의 리뷰에 대해 LDA 를 진행한 결과로 도출된 3 개의 토픽과 그 토픽에 해당하는 키워드의 비중은 LDAvis 와 워드 클라우드로 표현해주었다.



■ Henna Topic 1 : 헨나 호텔의 위치



토픽 1 은 상위 10 개 단어로 '0.036*"로봇" + 0.023*"호텔" + 0.017*"위치" + 0.016*"만족" + 0.015*"화장실" + 0.014*"숙소" + 0.014*"냄새" + 0.014*"객실" + 0.013*"방문" + 0.013*"친절"'을 나타낸다. 이러한 단어들이 한 토픽으로 묶인 것을 보아 이 토픽은 헨나 호텔의 위치로 분석할 수 있다. 더불어 호텔 위치에 만족하나 신축 건물의 특성상 숙소에서 냄새가 난다는 리뷰가 있다고 볼 수 있다. 이 토픽에 해당하는 단어들은 전체 토픽의 51.6%에 해당한다.

■ Henna Topic 2 : 로봇의 룸서비스



토픽 2 는 상위 10 개 단어로 ' $0.043 \times \text{"이용"} + 0.019 \times \text{"로봇"} + 0.019 \times \text{"감사"} + 0.013 \times \text{"호텔"} + 0.012 \times \text{"친절"} + 0.012 \times \text{"직원"} + 0.010 \times \text{"숙소"} + 0.009 \times \text{"시설"} + 0.008 \times \text{"남산"} + 0.008 \times \text{"방문"}$ '을 갖는다. 이러한 단어들이 한 토픽으로 묶인 것으로 보아 이 토픽은 로봇의 룸 서비스로 볼 수 있다. 또 이 토픽에는 로봇을 수건을 가져다주는 데에 이용했다는 내용이나 호텔의 어메니티와 관련된 내용이 있다고 분석할 수 있다. 이 토픽에 해당하는 단어들은 전체 토픽의 27.8%에 해당한다.

■ Henna Topic 3 : 호텔에 만족



토픽 3 의 상위 10 개 단어는 ' $0.043 \times \text{"호텔"} + 0.030 \times \text{"친절"} + 0.027 \times \text{"숙박"} + 0.027 \times \text{"청결"} + 0.022 \times \text{"욕조"} + 0.019 \times \text{"가격"} + 0.015 \times \text{"직원"} + 0.013 \times \text{"위치"} + 0.012 \times \text{"방문"} + 0.012 \times \text{"체크인"}$ '이다. 이 토픽들이 한 토픽으로 묶인 것을 보아 이는 호텔에 대한 고객들의 만족을 의미한다고 볼 수 있다. 더불어 이 토픽안에는 호텔이 청결하고 위치가 좋아 만족스럽다는 내용이 있다고 볼 수 있다. 이 토픽에 해당하는 단어들은 전체 토픽의 20.6%에 해당한다.

특히, 결과값에서 주목해야 할 부분은 헨나 호텔의 리뷰에는 '로봇'에 대한 언급이 많다는 것이다. 이는 앞서 본 연구를 통해 확인하고자 한 내용인 '로봇 경험이 고객에게 인상적인 요소로 작용하였는가?'를 설명할 수 있는 내용이다. 우선 헨나 호텔의 리뷰에서 '로봇'에 대한

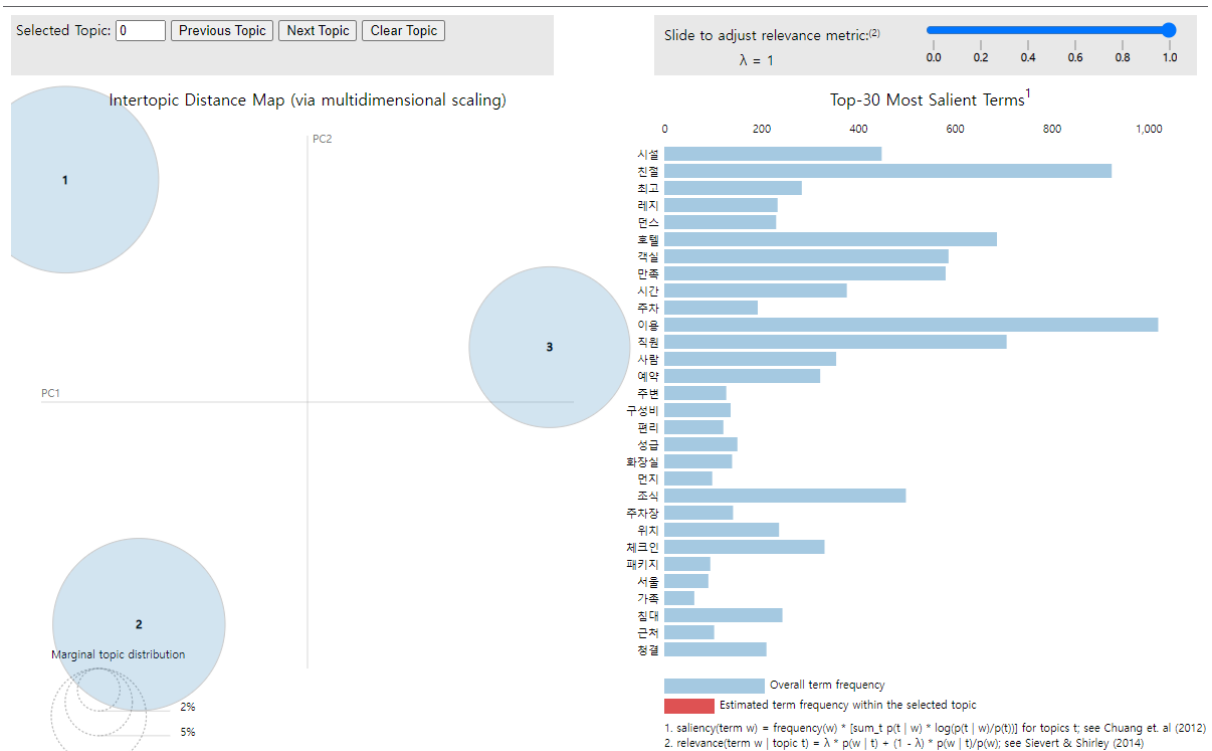
언급은 전체 토큰 3982 개 중 1375 개에 해당하는 것으로 도출되었으며 이는 전체의 34%의 비중을 차지한다. 또한 토픽별 키워드가 도출될 확률에 일부 차이가 있을 뿐 모든 토픽에서 키워드로 도출되었다. 이는 방문객들에게 로봇이 인상깊은 요소로 작용했음을 의미한다고 분석할 수 있다.

이러한 결과는 아마 헨나 호텔의 마케팅 방향과 연관이 있는 것으로 보인다. 헨나 호텔은 세계 최초의 로봇호텔로 일본에서 유명세가 있던 호텔이다. 이 호텔의 마스코트는 체크인과 체크아웃을 도와주는 공룡로봇이다. 하여 이곳에 방문하는 사람들은 호텔의 로봇이 체크인과 체크아웃을 도와주고 룸서비스를 가져다주는 등의 서비스를 기대하고 해당 호텔을 방문하였을 것이다. 그렇다면 자연스럽게 로봇에 대한 호텔 리뷰의 토픽이 많아진 것으로 추측된다.

헨나 호텔 외의 두 호텔(노보텔 호텔, 레스케이프 호텔)에 대해서는 LDA 를 진행했을 때 로봇 서비스와 관련한 단어가 토픽들에서 유의미한 분포를 갖지 못했다. EDA 에서 볼 수 있듯이 노보텔 호텔의 리뷰 중에서 ‘로봇’ 또는 ‘지니’에 대한 언급은 전체 토큰 중 0.6%에 해당하는 비중이 언급되었고, 레스케이프 호텔의 리뷰 중에서 ‘지니’에 대한 언급은 전체 토큰의 0.5%에 해당하는 양이 언급되었다. 이를 통해 노보텔 호텔과 레스케이프 호텔에 방문한 고객에게는 로봇이 제공하는 서비스가 리뷰에 남길 만큼의 인상적인 요소로 작용하지 않았다고 분석할 수 있다. 혹은 두 호텔을 방문할 때 로봇이나 인공지능 스피커가 고객들이 기대하는 서비스가 아니었을 것으로 판단된다. 하여 본 연구에서는 이러한 분석에 그치지 않고 더 나아가 어떤 요소가 해당 호텔의 방문객들에게 인상적인 요소로 작용하였는지 파악하고자 하였다.

3-2) 노보텔 호텔 LDA 결과

노보텔 호텔의 리뷰에 대해 LDA 를 진행한 결과 도출된 3 개의 토픽과 그 토픽에 해당하는 키워드는 다음과 같이 워드 클라우드와 LDAvis 로 표현해주었다.



■ Novotel topic 1 : 호텔 부대시설과 서비스



토픽 1 의 상위 10 개 단어는 '0.052*"**이용**" + 0.046*"**수영장**" + 0.025*"**조식**" + 0.024*"**시간**" + 0.021*"**사람**" + 0.019*"**예약**" + 0.018*"**체크인**" + 0.012*"**다음**" + 0.010*"**친절**" + 0.010*"**직원**"이다. 이를 보았을 때 이 토픽은 호텔 부대시설과 서비스로 정할 수 있다. 더 자세히 살펴보면 노보텔 호텔의 주요 부대시설인 수영장에 대한 내용과 함께 조식, 루프 탑, 해피 아워 등의 서비스 내용이 있다고 파악할 수 있다. 이 토픽에 해당하는 단어들은 전체 토픽의 38.5%에 해당한다.

■ Novotel topic 2 : 호텔 객실 만족



토픽 2 는 상위 10 개 단어로 ' $0.036 \times \text{"객실"} + 0.035 \times \text{"만족"} + 0.026 \times \text{"수영장"} + 0.021 \times \text{"친절"} + 0.019 \times \text{"레지"} + 0.018 \times \text{"던스"} + 0.017 \times \text{"방문"} + 0.017 \times \text{"직원"} + 0.015 \times \text{"서비스"} + 0.014 \times \text{"호텔"}$ '를 갖는다. 이는 객실에 만족하고 레지던스 시설에 대한 내용으로 분석할 수 있다. 또 이 토픽 안에서는 지니라는 단어가 보인다. 이를 보았을 때 지니가 객실 만족에 영향을 준다고 추측이 가능하다. 이 토픽에 해당하는 단어들은 전체 토픽의 32.9%에 해당한다.

■ Novotel Topic 3 : 호텔 직원의 친절

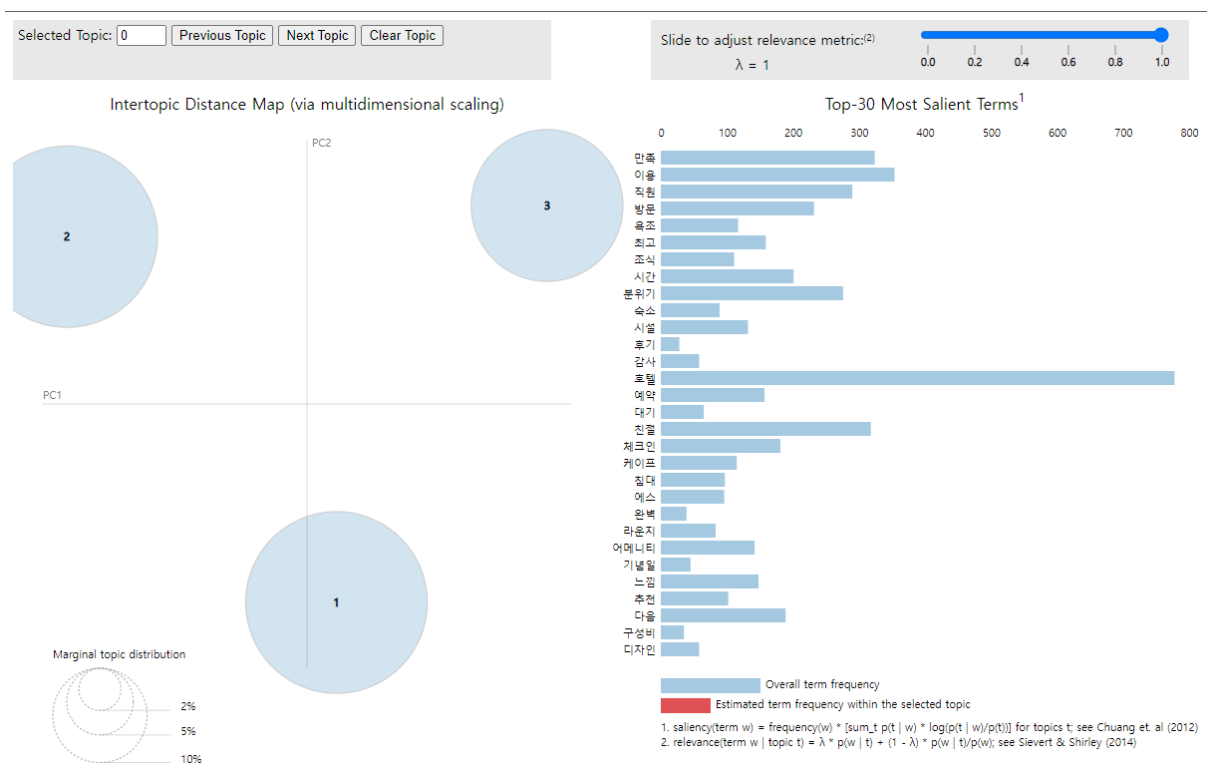


토픽 3 은 ' $0.053 \times \text{"친절"} + 0.040 \times \text{"호텔"} + 0.039 \times \text{"시설"} + 0.038 \times \text{"수영장"} + 0.037 \times \text{"직원"} + 0.025 \times \text{"최고"} + 0.017 \times \text{"주차"} + 0.017 \times \text{"이용"} + 0.014 \times \text{"루프"} + 0.014 \times \text{"위치"}$ '를 상위 10 개 단어로 갖는다. 이를 종합해서 판단해보면 이 토픽은 호텔 직원의 친절로 파악할 수 있다. 자세하게는 호텔 직원들이 친절하다는 내용 및 호텔 시설에 대한 내용으로 분석할 수 있다. 이 토픽에 해당하는 단어들은 전체 토픽의 28.6%에 해당한다.

노보텔 호텔의 LDA 분석 결과 노보텔 호텔에 대한 리뷰를 남기게 하는 주된 서비스는 루프탑 수영장과 직원들의 친절한 서비스였다. 노보텔 호텔은 KT AI 호텔 로봇 서비스를 사용하고 있기는 하지만 해당 서비스가 헨나 호텔과 같이 주된 서비스로 전환되지는

못하였다. 과연 5성급 노보텔 호텔을 방문한 고객은 로봇의 서비스에 만족하였을까? 이번 LDA 결과로 보면 로봇에 대한 기대를 하고 해당 호텔을 방문하지도(헨나를 방문하게 하는 요인이 로봇이었다면 반대로), 로봇의 서비스에 깊은 인상을 갖지도 못하였다고 분석된다. 하여 노보텔 호텔과 같이 기존에 로봇으로 마케팅이 되지 않았고, 더 나아가 해당 호텔에 강력한 방문 동인이 있다면 AI 로봇 도입과 같이 추가적인 비용이 발생하는 서비스를 수행하는 것에 깊은 고민이 필요할 것이다.

3-3) 레스케이프 호텔 LDA 결과



레스케이프 호텔의 리뷰에 대해 LDA 를 진행한 결과, 도출된 3 개의 토픽과 그 토픽에 해당하는 키워드는 다음과 같이 LDAvis 와 워드 클라우드로 표현해주었다.

■ Lescape Topic 1 : 호텔 이용 만족



토픽 1 은 '0.033*"이용" + 0.030*"만족" + 0.028*"호텔" + 0.014*"직원" + 0.014*"분위기" + 0.013*"시간" + 0.012*"객실" + 0.011*"서비스" + 0.011*"친절" + 0.011*"시설"'를 상위 10 개 단어로 갖는다. 이를 종합해서 판단해보면 이 토픽은 호텔 이용 만족에 대한 내용으로 파악된다. 자세히 파악해보면 호텔을 이용하는 시간이 만족스러웠다는 내용과 호텔 서비스가 친절했다는 내용이 있는 걸로 분석된다. 이 토픽에 해당하는 단어들은 전체 토픽의 37.1%에 해당한다.

■ Lescape Topic 2 : 호텔 체크인, 객실 최고



토픽 2 의 상위 단어 10 개는 '0.049*호텔' + 0.020*방문" + 0.014*체크인" + 0.014*최고" + 0.013*객실" + 0.013*다음" + 0.013*인테리어" + 0.013*예약" + 0.012*시간" + 0.010*이용"이다. 이를 종합적으로 보면 이 토픽은 호텔 체크인 및 객실 최고의 내용을 담고 있는 걸로 보인다. 또, 호텔을 방문해서 체크인부터 객실까지 만족했다는 내용도 있는 걸로 분석된다. 이 토픽에 해당하는 단어들은 전체 토픽의 37%에 해당한다.

■ Lescape Topic 3 : 호텔 분위기 인테리어

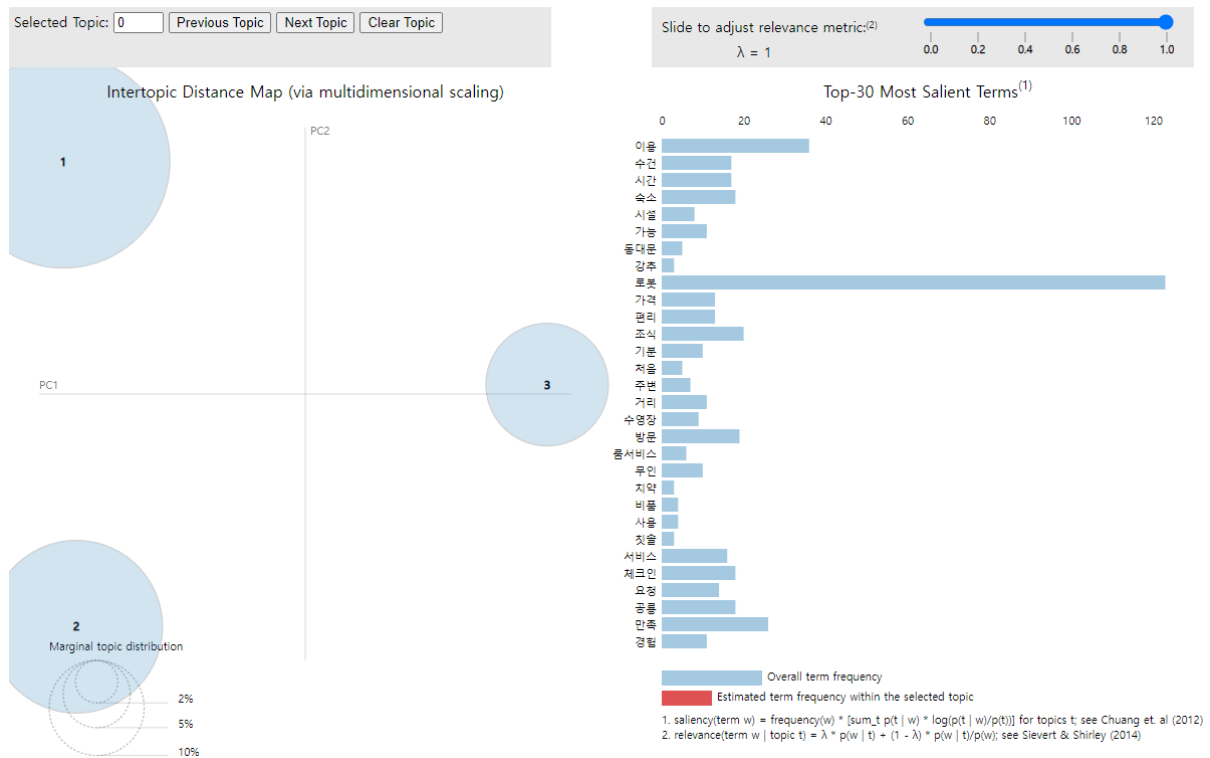


토픽 3의 상위 단어 10개는 '0.027*호텔' + 0.022*친절' + 0.021*직원' + 0.019*인테리어' + 0.018*분위기' + 0.014*만족' + 0.014*객실' + 0.011*욕조' + 0.010*이용' + 0.009*친구"이다. 이를 종합적으로 보면 이 토픽은 호텔 분위기 인테리어에 대한 주제로 파악된다. 호텔의 인테리어와 분위기에 만족했다는 내용도 있는 걸로 분석된다. 이 토픽에 해당하는 단어들은 전체 토픽의 25.9%에 해당한다.

LDA 결과 레스케이프 호텔을 찾는 고객들은 호텔의 분위기와 인테리어, 친절한 직원들의 서비스를 경험하기 위해 방문하는 것으로 보인다. AI 인공지능 스피커인 ‘지니’에 대한 언급이 보이긴 하지만 전체 토픽에서 큰 비중을 차지하지는 못한다. 이 호텔에서 ‘지니’는 룸서비스를 요청하거나 알람을 울려주고, 커튼을 열고 닫아주는 등의 자동화 역할 또한 수행한다. 하지만 이렇게 편리한 서비스임에도 불구하고 리뷰에 언급될 정도로 인상깊지 못했다는 사실은 AI 인공지능 서비스의 상용화에 대한 고민을 한 번 더 하게 한다.

3-4) 로봇에 대한 LDA 결과

예상과 달리 로봇과 지니에 대한 토픽이 적어 로봇과 지니를 포함하는 리뷰들에 대해서만 토픽 모델링을 해서 로봇과 지니에 대한 리뷰의 토픽을 파악해보고자 했다.



로봇 리뷰에 대해 LDA 를 진행한 결과, 도출된 3 개의 토픽과 그 토픽에 해당하는 키워드는 다음과 같이 LDAvis 와 워드 클라우드로 표현해주었다.

■ Robot Topic 1 : 로봇의 룸서비스



토픽 1 의 상위 단어 10 개는 '0.051*"로봇" + 0.018*"수건" + 0.017*"호텔" + 0.016*"직원" + 0.016*"만족" + 0.014*"화장실" + 0.013*"방문" + 0.012*"친절" +

$0.011 * \text{"객실"} + 0.010 * \text{"시간"}$ 이다. 이를 종합적으로 보면 이 토픽은 로봇의 룸서비스에 대한 주제로 파악된다. 이 토픽에 해당하는 단어들은 전체 토픽의 50.4%에 해당한다.

■ Robot Topic 2 : 객실에서 로봇



토픽 2 의 상위 단어 10 개는 ' $0.056 * \text{"로봇"} + 0.029 * \text{"호텔"} + 0.019 * \text{"지니"} + 0.018 * \text{"객실"} + 0.015 * \text{"친절"} + 0.014 * \text{"직원"} + 0.013 * \text{"공룡"} + 0.012 * \text{"이용"} + 0.010 * \text{"가격"} + 0.010 * \text{"만족"}$ '이다. 이를 종합적으로 보면 이 토픽은 로봇을 쓰는 호텔 객실에서 지니가 사용된다는 내용으로 파악된다. 이 토픽에 해당하는 단어들은 전체 토픽의 33%에 해당한다.

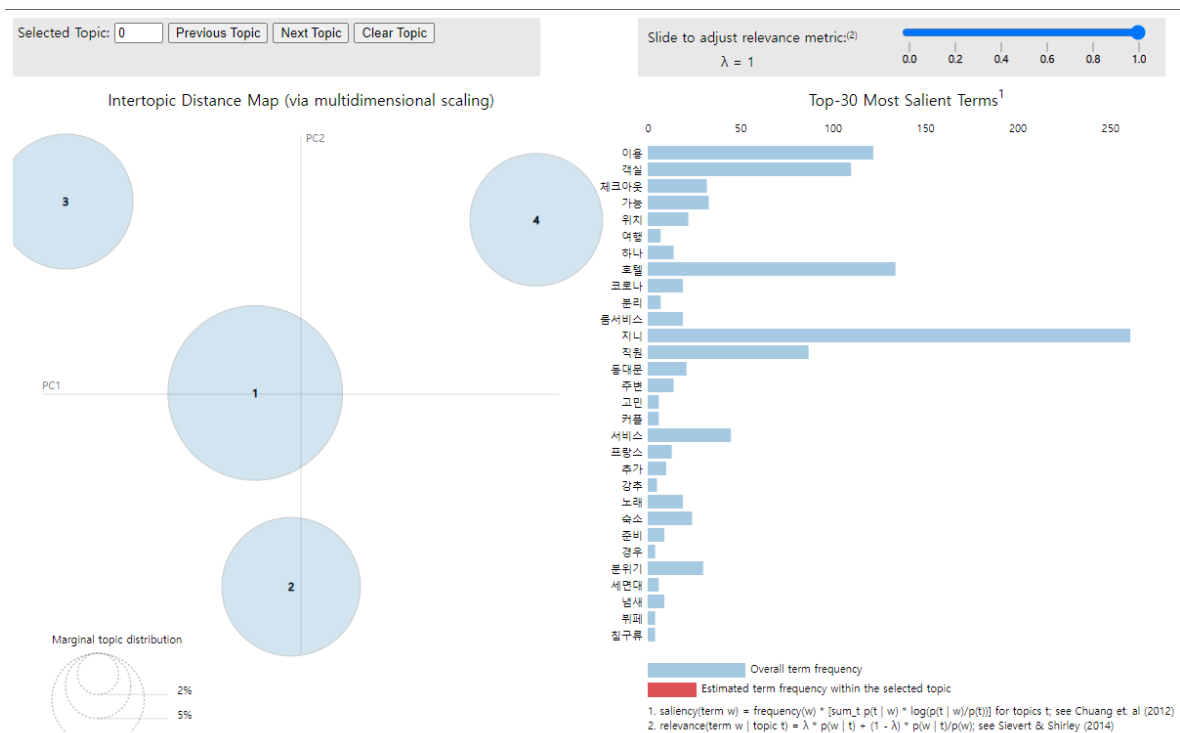
■ Robot Topic 3 : 호텔 이용에 만족



토픽 3 의 상위 단어 10 개는 ' $0.036 * \text{"로봇"} + 0.034 * \text{"이용"} + 0.014 * \text{"숙소"} + 0.013 * \text{"시간"} + 0.012 * \text{"직원"} + 0.011 * \text{"조식"} + 0.011 * \text{"가능"} + 0.010 * \text{"시설"} + 0.010 * \text{"지니"} + 0.009 * \text{"친절"}$ '이다. 이를 종합적으로 보면 이 토픽은 로봇을 쓰는 호텔 숙소를 이용하는 시간에 만족했다는 내용으로 파악된다. 이 토픽에 해당하는 단어들은 전체 토픽의 16.6%에 해당한다.

로봇의 룸서비스를 나타내는 토픽인 1 번 토픽이 가장 많은 비중을 차지하였다. 여기서 알 수 있는 것은 로봇이 룸서비스를 해주는 경험을 리뷰로 많이 남겼다는 것이다. 또 LDA 결과 로봇의 룸서비스에 고객들은 만족도가 꽤 높은 것으로 보인다.

3-5) 지니에 대한 LDA 결과



■ Jini Topic 1 : 지니와 호텔 서비스



토픽 1 의 상위 단어 10 개는 '0.043*"지니" + 0.029*"호텔" + 0.020*"이용" + 0.020*"직원" + 0.017*"친절" + 0.015*"수영장" + 0.014*"방문" + 0.011*"다음" + 0.011*"만족" + 0.011*"서비스"이다. 이를 종합적으로 보면 이 토픽은 지니와 호텔

서비스에 대한 주제로 파악된다. 이 토픽에 해당하는 단어들은 전체 토픽의 35.7%에 해당한다.

■ Jini Topic 2 : 지니를 이용한 객실 컨트롤



토픽 2 의 상위 단어 10 개는 ' $0.038 \times \text{"지니"} + 0.035 \times \text{"이용"} + 0.023 \times \text{"객실"} + 0.020 \times \text{"호텔"} + 0.014 \times \text{"수영장"} + 0.012 \times \text{"가능"} + 0.012 \times \text{"방문"} + 0.009 \times \text{"만족"} + 0.009 \times \text{"서비스"} + 0.009 \times \text{"조명"}$ '이다. 이를 종합적으로 보면 이 토픽은 지니를 이용해 커튼을 치거나 조명을 조절하는 등의 객실 컨트롤을 했다는 내용으로 파악된다. 이 토픽에 해당하는 단어들은 전체 토픽의 22.5%에 해당한다.

■ Jini Topic 3 : 지니와 직원 서비스



토픽 3 의 상위 단어 10 개는 ' $0.046 \times \text{"지니"} + 0.025 \times \text{"객실"} + 0.017 \times \text{"직원"} + 0.015 \times \text{"친절"} + 0.014 \times \text{"체크아웃"} + 0.012 \times \text{"수영장"} + 0.012 \times \text{"이용"} + 0.011 \times \text{"방문"} + 0.009 \times \text{"만족"} + 0.009 \times \text{"체크인"}$ '이다. 이를 종합적으로 보면 이 토픽은 지니와 직원 서비스의 주제로 파악된다. 이 토픽에 해당하는 단어들은 전체 토픽의 21.3%에 해당한다.

■ Jini Topic 4 : 지니와 호텔 분위기



토픽 4 의 상위 단어 10 개는 $0.037 \times \text{"지니"} + 0.024 \times \text{"호텔"} + 0.016 \times \text{"객실"} + 0.013 \times \text{"친절"} + 0.011 \times \text{"이용"} + 0.009 \times \text{"만족"} + 0.009 \times \text{"분위기"} + 0.009 \times \text{"조식"} + 0.009 \times \text{"직원"} + 0.008 \times \text{"기분"}$ 이다. 이를 종합적으로 보면 이 토픽은 지니와 호텔 분위기에 대한 주제로 파악된다. 아마 레스케이프 호텔에 대한 리뷰로 추측된다. 이 토픽에 해당하는 단어들은 전체 토픽의 20.6%에 해당한다.

지니 또한 지니에 대한 리뷰만 찾아보았을 때 지니 사용에 만족 중인 것으로 추측된다.

4. 결론

본 연구는 노보텔 호텔, 레스케이프 호텔, 헨나 호텔의 리뷰에 대한 LDA 를 진행하였다. 이를 통해 해당 호텔들의 리뷰가 어떤 주제로 쓰였는지 파악할 수 있었다. 본래 기대했던 것은 4 차 산업 기술들이 도입된 호텔들이기 때문에 해당 내용의 토픽들이 다수 포함되어 있을 것이라고 생각했지만 헨나 호텔을 제외하고는 4 차 산업 기술에 대한 토픽을 찾기 힘들었다. 이는 아직 4 차 산업 기술들이 기존 호텔에서 크게 활약하고 있지는 못한 것으로 분석된다. 하지만 로봇과 지니에 대해 따로 LDA 를 해보았을 때는 로봇과 지니에 대한 고객들의 반응이 좋은 것을 보았을 때 점진적으로 호텔 산업에 4 차 산업 기술을 도입한다면 긍정적인 결과를 얻을 수 있을 거라고 생각한다.

추가적으로 이번에 진행한 LDA 는 2 가지를 고려하지 못했다. 먼저 시간에 대한 개념이다. 노보텔 호텔의 경우 2019 년 12 월 로봇 서비스가 상용화 되어 이 기간 전후로 리뷰 토픽의 변화에 대해 분석해 본다면 유의미한 결과를 도출해낼 수 있었을 것이라 생각한다. 또 시간 개념을 고려할 수 있다면 2020 년 전후로 발병한 COVID-19 로 인한 토픽의 변화 또한 발견할 수 있을 거라고 기대한다. 다음으로 단어가 쓰인 문맥을 고려하지 못했다. LDA 는 단순히 문서에 단어들이 쓰이는 빈도를 가지고 토픽을 추정하는 알고리즘이기에 해당 리뷰의 토픽에서 쓰이는 단어가 긍정적으로 쓰인 건지 부정적인 쓰인 건지 파악하기 힘들었다. 해서 발전 방향으로 시계열을 고려한 LDA 와 contextual 분석을 진행해 발전시킨다면 더 의미있는 결과를 얻을 수 있을 거라고 생각한다.