

제주 특화 맞춤형 여행 추천 서비스 개발을 위한 숙박 데이터 수집 및 통합

임유정*, 김수균**, 윤여찬*

*제주대학교 인공지능전공, **제주대학교 컴퓨터공학전공, *제주대학교 인공지능전공

*routing0214@stu.jejunu.ac.kr, **kimsk@jejunu.ac.kr, *ycyoon@jejunu.ac.kr

Collection and Integration of Accommodation Data for the Development of a Jeju-Specific Personalized Travel Recommendation Service

Yu Jeong Lim*, Soo Kyun Kim**, Yeo Chan Yoon*

*Dept. of Artificial Intelligence, Jeju National Univ.

**Dept. of Computer Engineering, Jeju National Univ.

*Dept. of Artificial Intelligence, Jeju National Univ.

요 약

본 논문은 제주 특화 맞춤형 여행 추천 서비스를 개발하기 위해 다양한 출처에서 제주 지역의 숙박 데이터를 수집, 정제, 통합하여 양질의 데이터셋을 구축하는 것을 목표로 하였다. 제주 관광의 디지털 대전환을 위해 인공지능을 활용한 개인화된 관광 서비스를 제공하는 것이 필수적임을 인식하고, 제주 관광 협회, 공공데이터 포털, AI Hub 등에서 데이터를 확보하였다. 데이터 정제 및 통합 과정을 통해 최종 통합된 데이터셋은 호텔의 기본 정보와 리뷰 및 평점 데이터를 포함하며, 추후 AI 모델 학습에 활용될 수 있다. 이를 통해 개인화된 관광 경험을 제공하여 관광객의 만족도를 높임으로써 제주 관광 산업의 경쟁력을 강화하는데 기여할 수 있을 것으로 기대된다.

I. 서 론

제주 관광의 디지털 대전환은 MZ세대를 중심으로 한 디지털 세대가 주요 관광 소비층으로 부상함에 따라 필수적으로 추진되어야 할 과제이다. MZ세대는 디지털 환경에 익숙하고 개인화된 경험을 중시하기 때문에, 이들의 요구에 부응하기 위해 인공지능 기술을 활용한 서비스 개발이 필요하다. 인공지능 기술을 통해 제공되는 개인화된 맞춤형 서비스는 제주 관광 산업의 부가가치를 크게 향상시킬 수 있다[1].

AI 모델의 성능을 극대화하기 위해서는 양질의 데이터가 필수적이다. 그러나 현재 제주 관광 데이터는 여러 저장소에 흩어져 있어 분석 및 활용에 결핍되어 있다. 제주 특화 맞춤형 여행 추천 서비스 개발을 위해서는 다양한 저장소에 분산되어 있는 제주 관광 데이터를 통합하고, 이를 체계적으로 정리할 필요가 있다. 따라서 본 연구는 여러 제주 숙박 관광 데이터를 통합하고 정제하여 이를 기반으로 사용자에게 제주에서 머무는 숙박시설에 대한 상세하고 신뢰성 있는 정보를 제공하는 제주 특화 맞춤형 여행 추천 서비스[2]를 개발하기 위한 기초 데이터를 마련하는 것을 목표로 한다.

II. 본론

1. 데이터 수집 및 원본 데이터셋의 특징

제주지역 특화 관광 AI 에이전트를 개발하기 위해 제주 관광 협회, 서귀포시 시청, 공공데이터 포털, AI Hub, 문화 빅데이터 플랫폼, kaggle과 같은 다양한 출처에서 데이터를 수집하였다. 각 데이터셋은 다음 표와 같이 숙박업소에 대한 정보를 포함하고 있었으나 제공하는 구체적인 정보의 특성이 달라 일관되지 않다는 문제점이 있었고, 일부 데이터셋에는 누락된 정보나 비어있는 필드가 많아 완전한 데이터를 얻는데 어려움이 있었다.

원본 데이터	특징
도내 숙박업 현황(2024년 7월말 기준)	제주 도내 7,484개 숙박시설 현황(업종, 상호명, 소재지, 객실 수, 연락처, 비고) 농어촌민박과 같은 지역 특화 숙박시설도 모두 포함 출처: 제주특별자치도 관광 협회 http://www.visitjeju.or.kr/web/bbs/bbsDtl.do?pageIndex=2&sBbsId=NOTICE&bbsId=NOTICE&noticeNum=3287&authNum=&sKeyOpt=1&sKey=
서귀포시 숙박시설 현황	서귀포시 내 1,868개의 숙박업소에 대한 구체적인 정보(예: 상호명, 주소, 객실 수 등) 농어촌민박과 같은 지역 특화 숙박업소도 포함 출처: 서귀포시 시청 https://www.seogwipo.go.kr/communication/notify/accommodation/status.htm?act=view&seq=145423993
VISIT JEJU 숙박콘텐츠	1,373개의 숙박시설에 대한 숙박 옵션의 다양성(예: 애완동물 동반 가능 여부, 조식 제공 여부 등)과 편의시설에 대한 정보 출처: 공공데이터 포털 https://www.data.go.kr/data/15041985/fileData.do
관광 KVQA 데이터	텍스트, 이미지, 그리고 질의응답 형식으로 구성된 다중 데이터셋으로, 제주지역의 1,614개 숙박 시설에 대한 정보(예: 홈페이지, 입·퇴실시간 등)를 포함 출처: AI Hub https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realmd&dataSetSn=71605
제주 숙박시설 주변 식당 선호도 정보	253개 숙박시설의 주변 식당에 대한 정보 제공 출처: 문화 빅데이터 플랫폼 https://www.bigdata-culture.kr/bigdata/user/data_market/detail.do?id=5c294ba0-7f9c-11ee-88b4-13846e2c3c9
트립어드바이저 호텔 리뷰	고객 경험을 반영한 텍스트 리뷰와 평점 데이터를 포함하여 사용자 경험 기반의 평가를 제공 출처: kaggle https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews

표 1 원본 데이터의 특징

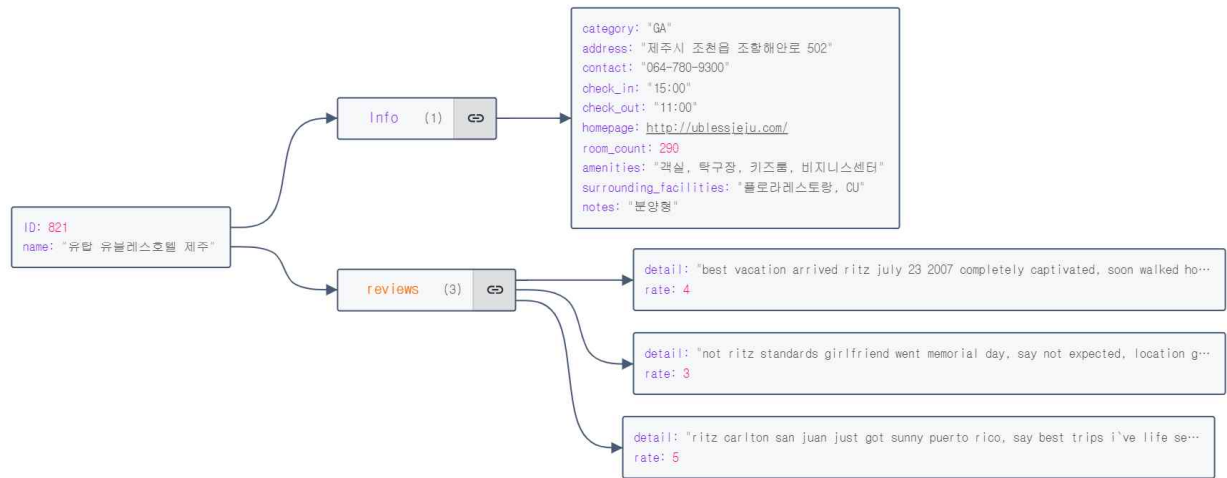


그림 1 JSON 최종 데이터셋 예시

참 고 문 헌

2. 데이터 정제 및 통합 과정

각 데이터셋이 지니고 있는 문제점을 보완하기 위해, 데이터 정제 및 통합의 과정을 거쳤다. 동일한 호텔이 여러 데이터셋에 중복되어 나타나는 경우가 많았기 때문에 총 7,484개의 고유한 호텔 ID를 생성하여 중복을 처리하였다. 없는 항목이나 불명확한 정보는 null로 설정하여 추가 입력을 고려하였다. 추가로 데이터셋 안에서 각기 다른 명칭으로 흩어져있는 숙박 시설에 대한 정보를 편의시설, 인접시설로 분류하였다. 추후 보다 효과적인 RAG 수행을 위해 필요한 리뷰 및 평점 데이터는 실제 데이터를 이용할 수 없어 트립어드바이저에서 제공된 데이터를 가상으로 연결하였다[3].

3. 최종 데이터셋의 구성 및 활용

통합된 최종 데이터셋 JSON 형태로 다음과 같은 구성 요소를 포함한다.

- 호텔의 기본 정보: 호텔 ID, 호텔명, 업종분류, 주소, 연락처, 입실시간, 퇴실시간, 홈페이지, 객실 수, 편의시설, 인접시설, 비고 등의 기본적인 정보.

- 리뷰 및 평점 정보: 각 호텔에 대한 가상 리뷰 내용과 평점.

최종 데이터셋은 제주 특화 맞춤형 여행 추천 서비스를 위한 AI 모델 학습에 활용될 수 있으며, 이는 제주를 방문하는 관광객에게 개인화된 호텔 추천 및 여행 경로를 제시하는 데 중요한 역할을 한다.

III. 결론 및 향후 연구

본 논문에서는 제주 특화 맞춤형 여행 추천 서비스를 개발하기 위한 기초 데이터를 마련하기 위해, 다양한 출처에서 제주 지역의 숙박 관련 데이터를 수집하고 이를 정제하여 통합된 제주 숙박 데이터셋을 구축하였다. 통합된 최종 데이터셋은 제주를 방문하는 관광객들에게 개인화된 호텔 추천과 여행 경로를 제시하는 생성형 AI의 학습에 활용될 수 있다. 본 연구의 결과는 제주 지역의 관광 데이터 활용에 새로운 접근을 제시하며, 향후 더욱 고도화된 제주 특화 맞춤형 여행 추천 서비스 개발에 중요한 기초 자료로 활용될 것으로 기대된다.

ACKNOWLEDGMENT

본 과제(결과물)는 2024년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다.(2023RIS-009)P

- [1] 정광민·한희정·박준희·박병호(2023), 미래유망기술의 관광산업 분야 적용 전망과 대응 방향, 한국문화관광연구원, pp.211-220
- [2] 박정현, & 이경민. (2024). 리뷰 데이터 기반의 사용자 친화적인 호텔 추천 챗봇 시스템. 한국통신학회 학술대회는문집, 1419-1422.
- [3] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in Neural Information Processing Systems, 33, 9459-9474.