Development of the Rule-based Smart Tourism Chatbot using Neo4J graph database

Dong-Hyun Kim, Hyeon-Su Im, Jong-Heon Hyeon, Jeong-Woo Jwa

Professor, Department of Telecommunication Engineering, Jeju National University, Korea
{ehsgusdh , dlagustn55, rkskwhdgh12, lcr02}@jejunu.ac.kr

Abstract

We have been developed the smart tourism app and the Instagram and YouTube contents to provide personalized tourism information and travel product information to individual tourists. In this paper, we develop a rule-based smart tourism chatbot with the khaiii (Kakao Hangul Analyzer III) morphological analyzer and Neo4J graph database. In the proposed chatbot system, we use a morpheme analyzer, a proper noun dictionary including tourist destination names, and a general noun dictionary including containing frequently used words in tourist information search to understand the intention of the user's question. The tourism knowledge base built using the Neo4J graph database provides adequate answers to tourists' questions. In this paper, the nodes of Neo4J are Area based on tourist destination address, Contents with property of tourist information, and Service including service attribute data frequently used for search. A Neo4J query is created based on the result of analyzing the intention of a tourist's question with the property of nodes and relationships in Neo4J database. An answer to the question is made by searching in the tourism knowledge base. In this paper, we create the tourism knowledge base using more than 1300 Jeju tourism information used in the smart tourism app. We plan to develop a multilingual smart tour chatbot using the named entity recognition (NER), intention classification using conditional random field(CRF), and transfer learning using the pretrained language models.

1. Introduction

We need to develop a chatbot that can activate the smart tourism industry by using the main technologies of the 4th industrial revolution such as Internet of Things (IoT), communication infrastructure including 5G mobile communication, big data, and artificial intelligence (AI) [1]-[3]. Smart tourism services must be able to provide personalized tourism services to tourists before, during, and after travel. Smart tourism services should provide tourists in the travel preparation stage with recommended tourist destinations and recommended travel products that can easily and conveniently create personalized travel itineraries. During travel, tourists must be able to provide a smart tour guide service according to the travel itinerary, and after travel, provide their travel itinerary as a recommended travel product to other tourists, including acquaintances. To efficiently provide competitive smart tourism services to domestic and foreign tourists, it

is necessary to develop a smart tourism system that can provide services through various channels such as chatbots, mobile apps, Instagram, YouTube, Facebook, and Kakao Talk. Google [4], Amazon [5], IBM [6], Microsoft [7], and Naver [8] provide chatbot APIs and chatbot agents so that developers can easily develop chatbots. The tourism chatbots had been proposed to provide the storytelling and photo services to tourists [9]-[10].

We have built tourism contents for more than 1,300 tourist destinations in Jeju Island and developed smart tourism app, Instagram and YouTube services using them. Tourism contents include tourist information, storytelling contents, photos, VR and drone video contents, etc. We have developed pedestrian network data in the tourist attractions [11] and displaying the walking route in a map view using Google API. The smart tourism app provides weather information of tourist destinations using the weather API of the Korea Meteorological Administration [12]. We also have developed the location-based audio tour guide system using speech synthesis provided by the server-based multilingual text-to-speech (TTS) engine [13]. The tour guide system provides a personalized audio guidance service in the tourist attractions based on a personal travel experience to resolve the disadvantages of relatively inexpensive group tours. Before the trip, the user can select the geotagged voice guidance sentence and use the tour guide service with the app during the trip. Voice guidance sentences have been constructed by classifying them as POI (point of interest), PORP (point of route plane), POS (point of story), POR (point of risk), POP (point of photo), and POE (point of event). In the tour guide system, we can analyze tourist attraction data visited by tourists and travel trajectory data of tourists, and provide them as chatbot services.

The graph database can be used in recommendation systems in various fields. In this paper, we develop a rule-based smart tourism chatbot service using the Neo4J tourism graph database to efficiently provide tourism information to tourists. In section 2, we explain how to grasp the intention in a user's question sentence using the khaiii (Kakao Hangul Analyzer III) morphological analyzer [14] and dictionaries. In section 3, we describe a tourism knowledge base built using the Neo4J graph database [15] and explains Neo4J queries for user's questions. Finally, in section 4, conclusions and future research methods are described.

2. Understanding the user's intent using the morpheme analyzer and dictionaries

Smart tourism chatbot systems typically consist of main six components: automatic speech recognition (ASR), text-to-speech synthesis (TTS), natural language understanding (NLU), natural language generation (NLG), dialog management (DM), and a tourism knowledge base as shown in Figure 1. NLU consists of named entity recognition and intent classification, and DM consists of dialog state tracking and dialog policy.
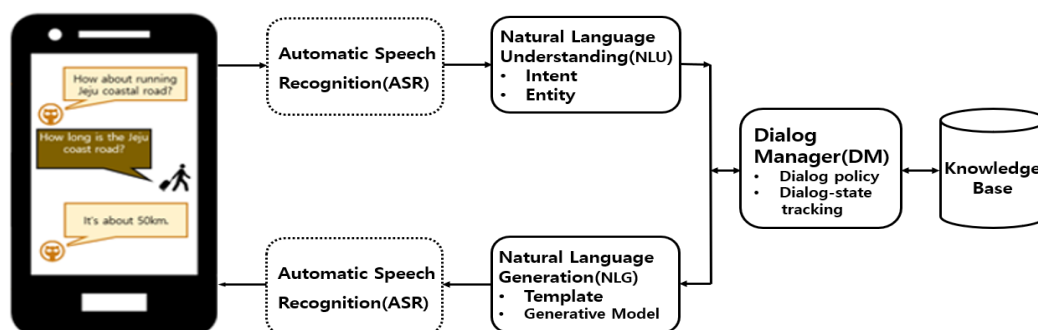

Figure 1. Smart tourism chatbot system.

In the preprocessor stage, we refine the user query sentence by checking for typos, merging or splitting words, processing special characters, and checking for spaces. We construct a dictionary with typo and space to refine the user query sentence as shown in Fig. 2. The dictionary was created to correct the problem of spacing in the results of analyzing the names of more than 1,300 tourist attractions.
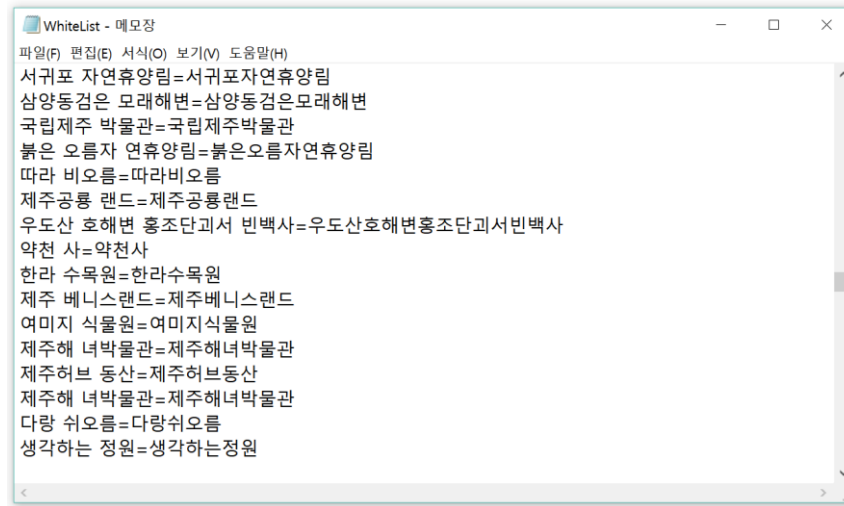

Figure 2. The typo and space dictionary for correcting user's question sentences.

We use a morpheme analyzer to create a Neo4J query statement by grasping the intention of a user's question from a refined question sentence during the user preprocessing process. We are writing a query statement using proper nouns (NNP) and nouns (NNG) among the analysis results of the morpheme analyzer for the user's question sentence. In this paper, we use Kakao's khaiii (Kakao Hangul Analyzer III) to morphologically analyze user's question sentences. We modified khaiii's pre-analyzed dictionary and error patch dictionary for accurate morphological analysis of user's questions. The error patch dictionary as shown in Fig. 3 is created based on the result of the morpheme analyzer to prevent erroneous analysis in the morpheme analyzer. In particular, the names of tourist destinations are often composed of compound nouns, so if the error patch dictionary is not used, the intention of the user's question cannot be grasped.
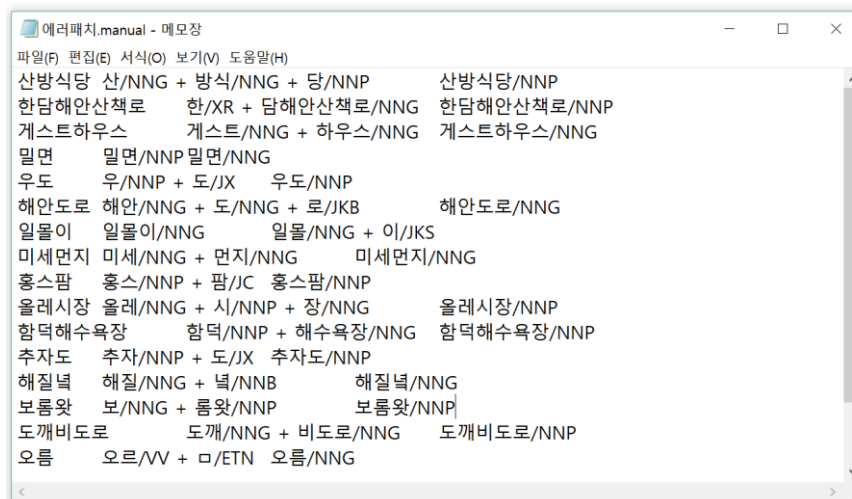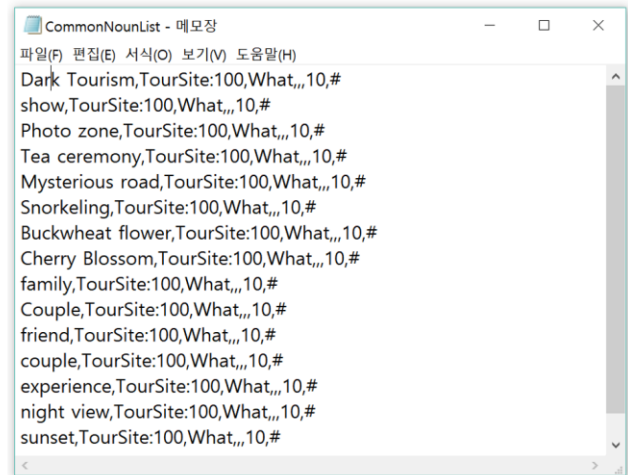

Figure 3. The error patch dictionary used in Khaiii morpheme analyzer.

After morphological analysis of the user's question sentence, category analysis is performed on the user's question based on the analysis result. For proper nouns (NNPs) including tourist destination names, we efficiently set the category by using a proper noun dictionary that specifies the category as shown in Fig. 4(a). For general nouns (NNG) in morpheme analysis, we create and use a general noun dictionary with a category as shown in Fig. 4(b). From the result of morphological analysis of user query sentences, we set up a category for proper nouns and common nouns using the proper nouns and common nouns dictionary and create a Neo4J query statement by analyzing the properties of tourist attractions from common nouns.
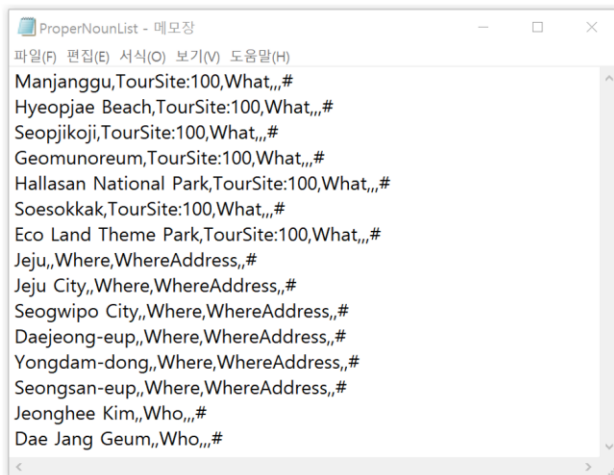
(a) Proper nouns dictionary   (b) Common nouns dictionary
Figure 4. Proper nouns and common nouns dictionaries used to understand the intent of the user's question.

3. The smart tourism knowledge base with Neo4J graph database

We created the Neo4J tourism graph database for the smart tourism chatbot service by using information on about 1,300 tourist destinations built in the Jeju area to provide smart tourism app services as shown in the Fig. 5.
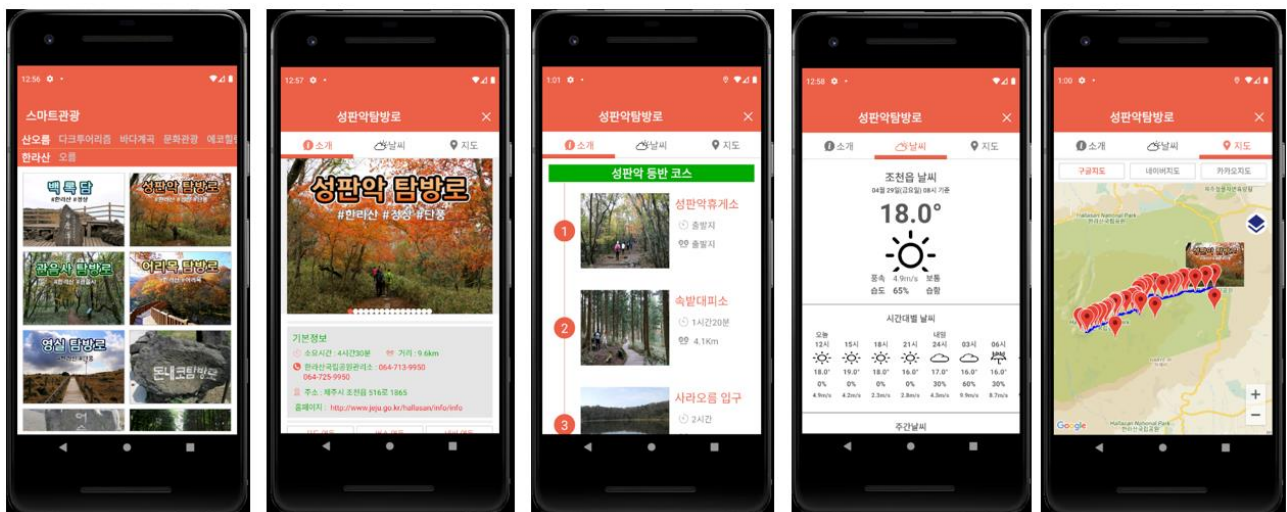

Figure 5. Smart tourism app that provides tourist information to individual tourists.

The smart tourism app is building tourism information to have a phase relationship with Hallasan national park, Hallasan trails, and tourist points within the trails. In the smart tourism app, weather information of tourist destinations uses the weather API of the Korea Meteorological Agency, and maps and navigation information use Google, Daum, and Naver map APIs. In the smart tourism chatbot service, weather information of tourist destinations is also provided from the smart tourism app. Tourist attraction information for the Neo4J tourist graph database is classified into tourist destination content and service attribute data.

The tourist destination content is composed of tourist destination information such as tourist destination name, address, phone number, homepage, description, coordinates, etc. provided as an answer to use as a chatbot for user query words. The tourist destination content can be provided by sorting the results of Neo4J queries in the order of Views, including data on the number of Views provided by the VISITJEJU website [16]. Tourist attraction service attribute data can be assigned to 520 important tourist destinations among 1300 tourist destinations, and customized tourist information can be provided to users through Neo4J query search. Tourist attraction attribute data is a service tag used by tourists to select a tourist destination and determines the quality of search service for Neo4J queries. We added the information of 520 tourist destinations to provide the recommended tourist destinations, such as tourist destinations according to tourist companions, famous photo zone tourist destinations, tourist destinations famous for flowers, and experiential tourist destinations such as marine sports, as an answer as a tourist destination service attribute data.

Table 1. CSV file of smart tourism knowledge base.

name	addr1	addr2	tel	homepage	latitude	longitude	desc	category	service	servic	service9
Seongsan Ilchulbong Pe	Jeju Seogwipo C	284-12, Ilchul-ro, S	064-783-0959	https://www.v	33.45826	126.942479	eongsan Ilchulbong Peak is	4000000	family		mountain
Seongsan Pohang Passe	Jeju Seogwipo C	347-9, Seongsan-ri	1544-8884		33.472873	126.934573	Seongsanpo General Passer	4000000	family	parent	experience
Seongeup Land	Jeju Seogwipo C	Seongeup-ri 2045-	064-787-5324	jejusland.com	33.39853	126.787629	Seongeup Land is a theme p	4000000	family		experience
Seongsidol Farm	Jeju Jeju City	142 Geumak-ri, Ha	064-796-0396		33.348703	126.327648	싱 Ishidol Ranch is a ranch l	4000000	family		experience
Seongpanak	Jeju Jeju City	137-2, Gyorae, J	064-713-9950	http://www.je	33.375561	126.585234	It is located in Hallasan Nati	4000000	family	parent	mountain
Seongpanak Service Are	Jeju Jeju City	137-24, Gyorae-ri,	064-713-9950	http://www.je	33.38510619	126.6198204	It is a rest area at the entra	4000000	family	parent	mountain
Shep Island	Jeju Seogwipo C	Bomok-dong San 1	064-710-6656		33.230605	126.598781	It is an island designated as	4000000	family		Jeju Island
World Liquor Museum	Jeju Seogwipo C	1814-1, Hacheon-r	064-787-9500		33.354244	126.817753	The World Liquor Museum,	4000000	family	parent	Cultural touri
World Automobile Jeju	Jeju Seogwipo C	1610, Jungsangans	064-792-3000	koreaautomuse	33.283374	126.349665	In the Automobile Museum	4000000	family	parent	Cultural touri
World Natural Heritage	Jeju Jeju City	569-36, Seongyo-r	1800-2002	jeju.go.kr	33.456948	126.714457	The Jeju World Natural Her	4000000	family	parent	Cultural touri
World Seashell Museum	Jeju Seogwipo C	Seohong-dong 557	064-762-5551	wsmuseum.co.	33.247812	126.551687	Artists directed the beautifu	4000000	family		Cultural touri
Sehwa Beach	Jeju Jeju City	27, Haenyeo Muse	064-728-7752		33.52533	126.861119	Sehwa Beach is a small bea	4000000	family		Sea
Soam Memorial Hall	Jeju Seogwipo C	15, Soam-ro, Seog	064-760-3511	culture.seogwi	33.24511	126.567566	The Soam Memorial Hall, lo	4000000	family	parent	Cultural touri
Soingook Theme Park	Jeju Seogwipo C	725, Seogwang-ri,	064-794-5400	http://www.so	33.290392	126.322406	Soingook Theme Park is a m	4000000	family		Theme tourist
소정방폭포(소라의성)	Jeju Seogwipo C	Donghong-dong 27	064-733-1530		33.245263	126.577529	Jeongbang Falls is located o	4000000	family	parent	Sea
Sokbat Shelter	Jeju Jeju City	Gyorae-ri, Jocheon	064-713-9950	http://www.je	33.38026802	126.581033	It is a resting place after goi	4000000	family	parent	mountain
Songdang Horseriding C	Jeju Jeju City	126, Songdang-ri,	064-782-1199		33.230384	126.498641	Songdang Horseriding Cours	4000000	family		experience

After grasping the intention of the question from the user query sentence, the answer is searched through the Neo4J query in the tourism knowledge base. Figure 6 shows nodes and relationships in the Neo4J tourism graph database. The nodes being created are Area, Contents, and Service. The relationships being created are IN_AREA, IN_CATEGORY, and IN_SVC. The property of the Contents node includes tourist information, and the Service node includes service attribute data for searching for a tourist destination as shown in Table 1. The following shows Neo4J queries for the user's question such as "Tell me about tourist attractions in Jeju". Neo4J queries for "Tell me about tourist attractions in Seogwipo City", and "Tell me about tourist attractions in Seongsan-eup" can be done by replacing a.name with Seogwipo City and Seongsan-eup instead of Jeju.

```
MATCH (a:Area)<-[:IN_AREA]-(c:Contents)-[:IN_CATEGORY]->(ca:Category{name:'4000000'}),
(s:Service)<-[:IN_SVC]-(c) WHERE (a.name =~ ' Jeju.*' ) AND (c.name =~ '.* tourist attraction.*'   OR
s.name =~ '.* tourist attraction.*')
RETURN DISTINCT c.name AS name, c.addr AS addr, c.tel AS tel, c.homepage AS homepage, c.lat AS lat, c.lot
AS lot, c.desc AS desc, c.vcount as vcount ORDER BY toInteger(c.vcount) DESC LIMIT 10
```
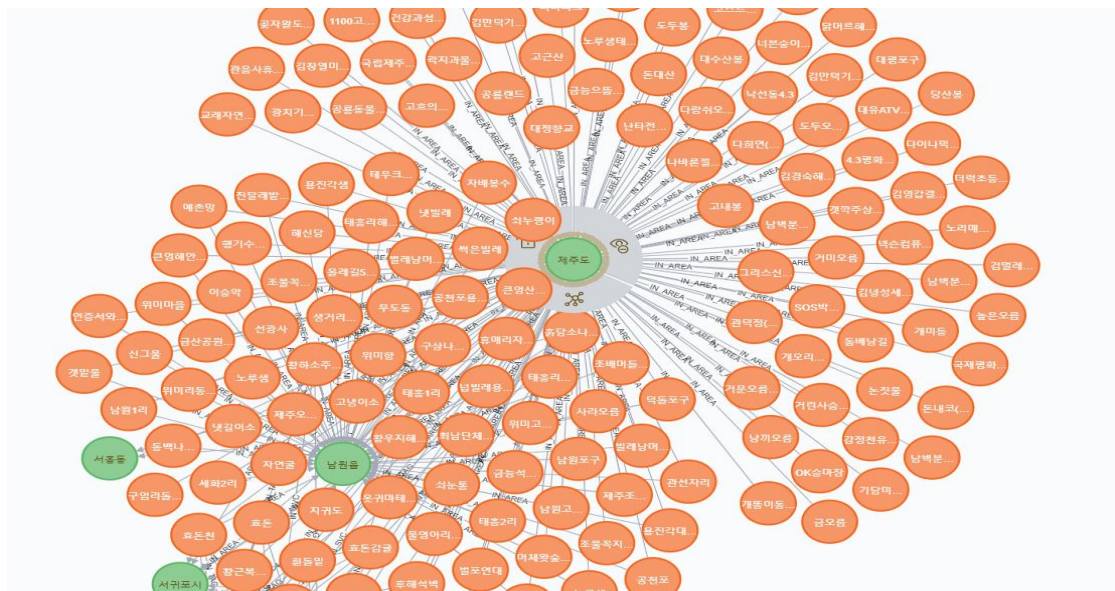

Figure 6. The smart tourism knowledge base built using Neo4J graph database.

The tourist destination content is constructed in three stages: province, city, eup, myeon-dong, depending on the address of the tourist destination, and provides the recommended tourist destination in the tourism knowledge base for the search for the name of the Neo4J query. The following is a Neo4J query for a user's question, "Please tell me a tourist attractions to go with a friend."

```
MATCH (a:Area)<-[:IN_AREA]-(c:Contents)-[:IN_CATEGORY]->(ca:Category{name:'4000000'}),
(s:Service)<-[:IN_SVC]-(c),(s1:Service)<-[:IN_SVC]-(c)
WHERE (a.name =~ 'Jeju.*' ) AND (s.name =~ '.* friend.*') and (s1.name =~ '.* tourist attraction.*')
RETURN DISTINCT c.name AS name, c.addr AS addr, c.tel AS tel, c.homepage AS homepage, c.lat AS lat, c.lot
AS lot, c.desc AS desc, c.vcount as vcount ORDER BY toInteger(c.vcount) DESC LIMIT 10
```

It shows the Neo4J query for the user's question "Please tell me the best coastal road with my family". To answer this, the tourist destination is searched through the AND operation of service attribute data such as "family" and "coast road".

```
MATCH (a:Area)<-[:IN_AREA]-(c:Contents)-[:IN_CATEGORY]->(ca:Category{name:'4000000'}),
(s:Service)<-[:IN_SVC]-(c), (s1:Service)<-[:IN_SVC]-(c)
WHERE (a.name =~ ' Jeju.*') AND (s.name =~ '.* family.*') and (s1.name =~ '.* coast road.*')
RETURN    DISTINCT c.name AS name, c.addr AS addr, c.tel AS tel, c.homepage AS homepage, c.lat AS lat, c.lot
AS lot, c.desc AS desc, c.vcount as vcount ORDER BY toInteger(c.vcount) DESC LIMIT 10
```

As described above, the tourist attraction attribute data includes information on the attributes of companions such as family, parents, couples, friends, and alone, and thus provides an answer to the Neo4J search query above. The figure 7 shows the search results for the above Neo4J query and the screen provided by the smart chatbot app.
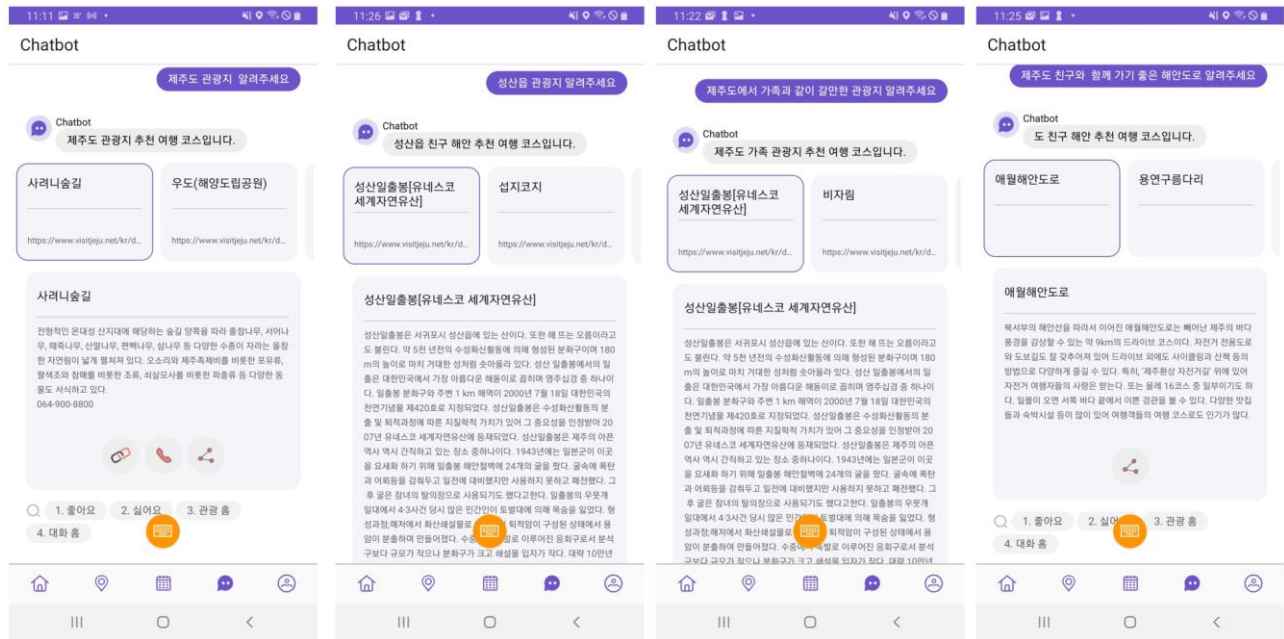

Figure 7. Smart tourism chatbot app screen showing search results in Neo4J tourism graph DB.

4. Conclusions and Discussions

We have been developed smart tourism services such as mobile apps and the Instagram and YouTube services to provide personalized travel information services to tourists. In this paper, we develop the rule-based smart tourism chatbot service to provide tourism information to individual tourists with a mobile app. To develop a tourism chatbot, named entity recognition (NER) is required to understand the intention of a tourist's question. We perform dictionary-based NER using Khaiii morpheme analyzer, the proper noun dictionary, and the common noun dictionary to understand the intention of tourists' questions. The proper noun dictionary contains the name of the tourist destination and the general noun dictionary includes words frequently used in sentences for searching tourist information. We also develop dictionaries to correct typos and spacing in user's question sentences, and the error patch dictionary to prevent misanalysis of Khaiii morpheme analyzer.

This paper uses a graph database to provide tourism information reply to the intention of the tourist's question. We create the tourism knowledge base to provide appropriate answers to user's questions based on more than 1300 Jeju tourism information used in the smart tourism app built using Neo4J graph database. Neo4J tourism graph database is created by adding words frequently used in search of tourist destinations to existing tourism information as service attribute data. It provides answers to tourists' questions by using the property of nodes and relationships in Neo4J tourism graph database. We plan to provide a convenient chatbot service to tourists using the location-based audio tour guide system [5]. We also plan to upgrade the chatbot system to apply AI-based NER and intention classification using CRF and transfer learning using a

pretrained language model [17]-[21].

Acknowledgement

References

[1]   Gretzel, U., Sigala, M., Xiang, Z. and Koo, C, "Smart Tourism: Foundations and Developments", *Electronic Market*, pp.179-188, 2015.
      DOI:10.1007/s12525-015-0196-8

[2]   Sameera A Abdul-Kader and JC Woods. "Survey on chatbot design techniques in speech conversation systems", *International Journal of Advanced Computer Science and Applications*, 6(7), pp.72-80, 2015.
      DOI : 10.14569/IJACSA.2015.060712

[3]   F.Clarizia, F. Colace, M. Lombardi, F. Pascale, "A Context Aware Recommender System for Digital Storytelling", *IEEE 32nd International Conference on Advanced Information Networking and Applications*, pp.542-549, 2018.
      DOI: 10.1109/AINA.2018.00085

[4]   Dialogflow, https://dialogflow.com/

[5]   Amazon Lex, https://aws.amazon.com/ko/lex/

[6]   Watson Assistant, https://www.ibm.com/watson/kr-ko/how-to-build-a-chatbot/

[7]   Azure Bot, https://azure.microsoft.com/ko-kr/solutions/architecture/information-chatbot/

[8]   Naver Clova Chatbot, https://www.ncloud.com/product/aiService/chatbot

[9]   Fabio Clarizia, Francesco Colace, Marco Lombardi, F. Pascale, "A Context Aware Recommender System for Digital Storytelling",2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), pp. 542-549, May 2018
      DOI: 10.1109/AINA.2018.00085

[10] Daishi Suzuki, Kazuya Nunotani, Keisuke Fukusato, Mikiko Sode Tanaka, "A Study of Tourism Proposal System Using AI", IEEE 9th Global Conference on Consumer Electronics (GCCE), pp.634-635, Oct. 2020.
      DOI: 10.1109/GCCE50665.2020.9292070

[11] Jeong-Woo Jwa, "Pedestrian Network Models for Mobile Smart Tour Guide Services," International Journal of Internet, Broadcasting and Communication, vol.8, no.1, pp.73-78. 2016. https://doi.org/10.7236/IJIBC.2016.8.1.27

[12] KiBeom Kang, JeongWoo Jwa, SangDon Earl Park, "Smart Audio Tour Guide System using TTS", International Journal of Applied Engineering Research, pp.9846-9852, 2017.
      https://www.ripublication.com/ijaer17/ijaerv12n20_81.pdf

[13] JeongWoo Jwa, "Development of Personalized Travel Products for Smart Tour Guidance Services", International Journal of Engineering & Technology, 7 (3.33) 58-61, 2018. DOI: 10.14419/ijet.v7i3.33.18524

[14] Kakao khaiii(Kakao Hangul Analyzer III), https://tech.kakao.com/2018/12/13/khaiii/

[15] Neo4j graph database, https://neo4j.com/

[16] Visit Jeju Website, https://www.visitjeju.net/kr

[17] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer "Deep contextualized word representations," arXiv preprintarXiv:1802.05365, 2018.
      DOI:10.18653/v1/N18-1202

[18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pretraining of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.
      DOI: 10.18653/v1/N19-1423

[19] ETRI KorBERT, http://aiopen.etri.re.kr/service_dataset.php

[20] AWS-SKT KoGPT-2, https://github.com/SKT-AI/KoGPT2

[21] SKT KoBERT, https://github.com/SKTBrain/KoBERT