



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](#)

碩士學位論文

다차원 순위 모델 기반의  
관광후기 분석

濟州大學校 大學院

관광융합소프트웨어학과

유 상 욱

2017 年 8 月

# 다차원 순위 모델 기반의 관광후기 분석

지도교수 김 근 형

유상욱

이 論文을 工學 碩士學位 論文으로 提出함

2017 年 6 月

유상욱의 工學 碩士學位 論文을 認准함

審査委員長 \_\_\_\_\_ 印

委 員 \_\_\_\_\_ 印

委 員 \_\_\_\_\_ 印

濟州大學校 大學院

2017 年 6 月

## 차례

표 차례 .....	iii
그림 차례.....	iv
ABSTRACT.....	v
<b>제 1 장 서론.....</b>	<b>1</b>
제 1 절 연구배경 및 목적 .....	1
1. 연구의 배경 .....	1
2. 연구의 목적 .....	2
제 2 절 연구의 방법 및 구성.....	3
1. 연구의 방법 .....	3
2. 논문의 구성 .....	3
<b>제 2 장 이론적 배경 .....</b>	<b>5</b>
제 1 절 데이터 마이닝 .....	5
1. 데이터 마이닝의 개념 .....	5
2. 데이터 마이닝 추진 단계.....	5
3. 데이터 마이닝의 기법 .....	7
1) 연관분석 .....	7
2) 다차원분석(OLAP).....	8
(1) OLAP의 특징 .....	9
4. 데이터 마이닝 관련 기존 연구 고찰.....	9
제 2 절 제주 관광지에 대한 선행연구 .....	11

<b>제 3 장 연구설계.....</b>	<b>12</b>
제 1 절 연구문제.....	12
제 2 절 분석 절차.....	13
제 3 절 분석 모델(다차원 순위 모델).....	14
<b>제 4 장 데이터 수집 및 분석.....</b>	<b>18</b>
제 1 절 데이터 원천.....	18
제 2 절 데이터 수집.....	19
제 3 절 텍스트 전처리(Text Preprocessing).....	20
제 4 절 분석 결과.....	22
1. 데이터 수집 결과 및 표본의 특성.....	22
2. 분석 결과.....	23
1) 제주에서 관심도가 높은 관광지.....	23
2) 제주에서 호감도가 높은 관광지.....	25
3) 제주 관광에서의 핵심 키워드.....	27
4) 관광지별 핵심 키워드.....	29
5) 키워드 연관성 분석.....	31
<b>제 5 장 결론.....</b>	<b>34</b>
제 1 절 연구 요약 및 시사점.....	34
제 2 절 연구의 한계.....	36
참고문헌.....	38

## 표 차례

<표 4-1> 데이터 수집 결과.....	22
<표 4-2> 차원별 관광지 개수 .....	22
<표 4-3> 계절별 제주에서 관심도가 높은 관광지 .....	23
<표 4-4> 연도별 제주에서 관심도가 높은 관광지 .....	24
<표 4-5> 계절별 제주에서 호감도가 높은 관광지 .....	26
<표 4-6> 연도별 제주에서 호감도가 높은 관광지 .....	26
<표 4-7> 계절별 제주 관광의 핵심 키워드.....	27
<표 4-8> 연도별 제주 관광의 핵심 키워드.....	28
<표 4-9> 계절별 관광지 핵심 키워드 .....	29
<표 4-10> 연도별 관광지 핵심 키워드.....	30
<표 4-11> 계절별 핵심 키워드 연관성.....	31
<표 4-12> 연도별 핵심 키워드 연관성.....	32

## 그림 차례

<그림 1-1> 연구의 흐름도.....	4
<그림 2-1> 데이터 마이닝의 추진 절차.....	6
<그림 3-1> 계절, 연도 차원에서의 연구문제.....	12
<그림 3-2> 연구의 분석 절차.....	13
<그림 3-3> 비 정형 빅 데이터의 정형화.....	14
<그림 3-4> 다차원 순위 모델의 개념도.....	15
<그림 3-5> 객체형 속성의 정형화.....	16
<그림 3-6> 재정형화된 테이블의 속성 유형.....	17
<그림 4-1> Tripadvisor 의 후기 작성 예시.....	18
<그림 4-2> 웹 크롤링의 흐름도.....	19
<그림 4-3> 데이터 수집 절차.....	20
<그림 4-4> 형태소 분석기 간의 처리 속도 비교.....	21
<그림 4-5> twitter-korean-text 를 이용한 형태소 분석 예시.....	21

## ABSTRACT

# Analysis of Tourism review based on Multi-dimensional Ranking Model

With the development of ICT, the continuous spread of smart phones, and the increase of SNS users, a large amount of data is being generated in real time, and the kind and quantity thereof is rapidly. In particular, Review Data through smartphones and SNS are important big data that can be used to derive the actual consumption, purchase pattern, and preference. Although the research is proceeding with the use of Review Big Data in the tourism society, the research of Jeju tourism is mainly composed of the questionnaires based on the questionnaire and there is a limited amount of data to analyze the perception of tourists visiting Jeju, because it is a lack of data and structured format. In addition, various data mining techniques for analyzing Review Data have been studied, but it focused on efficiency-oriented technology improvement, and research on analytical models to derive more effective and valuable information by fusing existing technologies is limited.

In this study, we propose a multidimensional ranking model that is applicable not only to Review data but also to analyze other types of data. The multidimensional ranking model enables ranking of the evaluation subject and the domain in various perspectives so that meaningful information can be derived. The multidimensional ranking model is of value in that it is a generalized model that can be applied to ranking data as well as Review data as well as other types of big data.



In particular, this study analyzed a 4,037 Jeju tourist Review data from 185 tourist sites that posted on Trip Advisor from 2015 to 2016 based on the multidimensional ranking model. We conducted a multidimensional ranking analysis of the rankings of tourist attractions, key keyword rankings, seasons, and years with high level of interest and favorability. As a result of the analysis, it received attention by tourists in the order of Seongsan ilchulbong, O'sulloc tea museum and Udo Island and be liked in the order of Hallasan National Park, Hamdeok Beach and Udo Island. The key keyword analysis of Jeju tourism has been used to refer to many keywords related to tourists and natural sightseeing places accompanied with children, and it has been analyzed that key words indicating the identity of each sightseeing spot are mainly mentioned through analysis of key keyword frequency by sightseeing spot. Through the keyword association analysis, the result of the analysis that leads to the analysis of the key keyword frequency of the previous tourist spots, and also, it was analyzed that tourists use coupons more frequently in order to receive discounts. Based on the results of the analysis, it was possible to derive the interest, liking, and perception of the tourist in the tourist attraction.

# 제 1 장 서론

## 제 1 절 연구배경 및 목적

### 1. 연구의 배경

정보통신기술(Information and Communications Technologies : ICT) 발달과 함께 스마트폰의 보급, SNS 사용자의 증가 등으로 다양한 데이터가 실시간으로 빠르게 생성되고 있다(윤지환, 이영진, 2014). 2016 년에 생성된 데이터는 16.1ZB 으로 추산되며, 2025 년도까지 10 배인 163ZB 까지 생성될 것이라 전망된다(David Reinsel, John Gantz, Rydning, 2017). 이처럼 대규모의 데이터와 다양한 데이터 형태, 생성 주기가 짧은 데이터 모두를 포괄하는 개념을 빅 데이터라 하며, 기존의 소규모 데이터에서 창출할 수 없는 가치를 빅 데이터에서 창출할 수 있기 때문에 전 세계적으로 빅 데이터에 주목하고 있다(한국데이터베이스진흥원, 2014). 빅 데이터 중에서도 후기(review) 데이터는 해당 분야 소비자들의 실질적인 소비 및 구매 패턴, 선호도 등을 도출할 수 있는 중요한 빅 데이터이기 때문에 중요성은 높아지고 있다. 하지만 이러한 후기 데이터는 대부분이 비구조화된 데이터이기 때문에 분석하기엔 쉽지 않지만(김근형, 오성열, 2009), 최근에는 하드웨어의 발달과 기술의 발달 등으로 분석이 용이해지고 있다.

문화체육관광부에서 실시한 2015 국민여행실태조사에 따르면, 개인 여행 선택 시 주로 블로그, 트위터, 관광 리뷰 사이트 등 소셜 미디어를 참고한다는 응답이 57.1%로 선택한다는 조사 결과가 나왔다. 이러한 결과는 관광정보를 획득하고, 공유하는 데 있어서 소셜미디어가 중요한 역할을 하고 있으며 관광 관련 데이터도 증가하고 있다는 것을 의미한다. 대표적인 예로 세계적인 관광 웹사이트인 Tripadvisor 는 관광과 관련된 후기 및 평가가 4 억 3,500 만 건을 보유하고 있으며(Tripadvisor, 2017) 이는 기존의 제한적이고 일관적인 정보에서 벗어나, 관광객이 직접 정보를 공유하고, 함께 정보를 만들어 가고자 하는 욕구가 반영된 결과이다(오성수, 2013). 실시간으로 데이터가 증가하며 전에는 파악하기 힘

들었던 관광객의 인식 및 욕구 분석이 가능해짐에 따라 최근 관광학회에서도 소셜 빅 데이터에 나타난 후기 데이터를 활용한 연구가 진행되고 있다. 한편, 제주 관광 관련 연구들은 설문 조사 기반의 연구들이 주를 이루고 있으며 이러한 연구 방법은 제주를 방문하는 관광객의 인식을 분석하기엔 데이터의 수가 부족하고, 구조화된 형식이기 때문에 한계가 있다(심영석, 2016). 따라서 후기 빅 데이터를 이용하여 제주 관광객의 인식을 분석하는 연구의 필요성이 제기되고 있다. 후기 데이터를 분석하기 위한 다양한 데이터 마이닝 기술들이 연구되고 있지만, 효율성 중심의 기술 향상에 치우쳐 있으며, 기존 기술들을 융합함으로써 보다 효과적이고 가치 있는 정보를 도출하기 위한 연구는 미진한 실정이다. 본 연구에서는 후기 데이터뿐만 아니라 다른 유형의 데이터를 분석할 때도 적용 가능한 분석 모델을 제안하고, Tripadvisor 에 게시된 제주 관련 후기를 분석하여 관광객의 인식을 파악하고자 한다.

## 2. 연구의 목적

본 연구는 관광지별 관광객의 후기를 제공해주는 관광 정보 웹사이트인 Tripadvisor 에 나타난 제주 관광지 후기를 분석하여 관광객의 인식을 알아보는 것이다. 이를 위해 웹 어플리케이션에서 제공하는 웹페이지의 자원을 자동으로 수집하는 웹 크롤러(송호길, 2011)를 이용하여 2015 ~ 2016 년간의 제주 관광지 후기 데이터를 수집하고, 형태소 분석기를 통해 명사 추출 후 데이터 마이닝 기법을 이용하여 다차원적으로 분석한다. 특히, 본 연구에서는 후기데이터를 효과적이고 가치 있는 정보를 도출하기 위한 분석 모델로 다차원 순위 모델을 제안한다. 다차원 순위 모델은 특정 속성값에 대한 순위를 여러 가지 관점에서 분석하여 유의미한 정보를 도출할 수 있도록 한다. 다차원 순위 모델은 순위 평가 영역과 대상이 관광 후기 데이터뿐만 아니라 타 유형의 빅 데이터에도 적용될 수 있는 일반화 모델이라는 점에서 그 가치와 의의가 있다. 또한, 제주 관광지 후기 데이터를 다차원 순위 분석을 하여 나온 결과를 통하여 관광 전략 수립을 위한 시사점을 제공하는 데에 그 목적이 있다.

## 제 2 절 연구의 방법 및 구성

### 1. 연구의 방법

본 연구에서는 연구의 목적을 달성하기 위해 문헌적 연구 방법과 실증적 연구 방법을 병행하여 진행하였으며, 구체적인 연구 방법은 다음과 같다.

첫째, 전문서적과 선행연구의 이해, 각종 연구 보고서 등을 통해 데이터 마이닝의 개념, 기법에 관한 문헌 연구를 진행하였다. 또한, 분석 대상이 될 후기 데이터를 이용한 선행연구의 흐름을 이해하고, 제주 관광 선행연구를 고찰한다.

둘째, 세계적인 관광 정보 웹사이트인 Tripadvisor 에 게시된 제주 관광지의 후기 데이터를 웹 크롤링하여 데이터를 수집한다. 수집된 데이터를 분석 가능한 형태로 변환하기 위해 전처리 과정을 거쳐 데이터베이스에 저장한다.

셋째, 선행연구를 바탕으로 효과적이고 가치 있는 정보를 추출하기 위한 분석 모델을 구축하고 분석을 실시한다.

### 2. 논문의 구성

본 논문은 총 5 개의 장으로 구성되어 있으며 세부내용은 다음과 같다. 제 1 장 서론에서는 연구의 배경 및 연구 목적, 연구의 방법, 본 논문의 구성에 대해 설명한다.

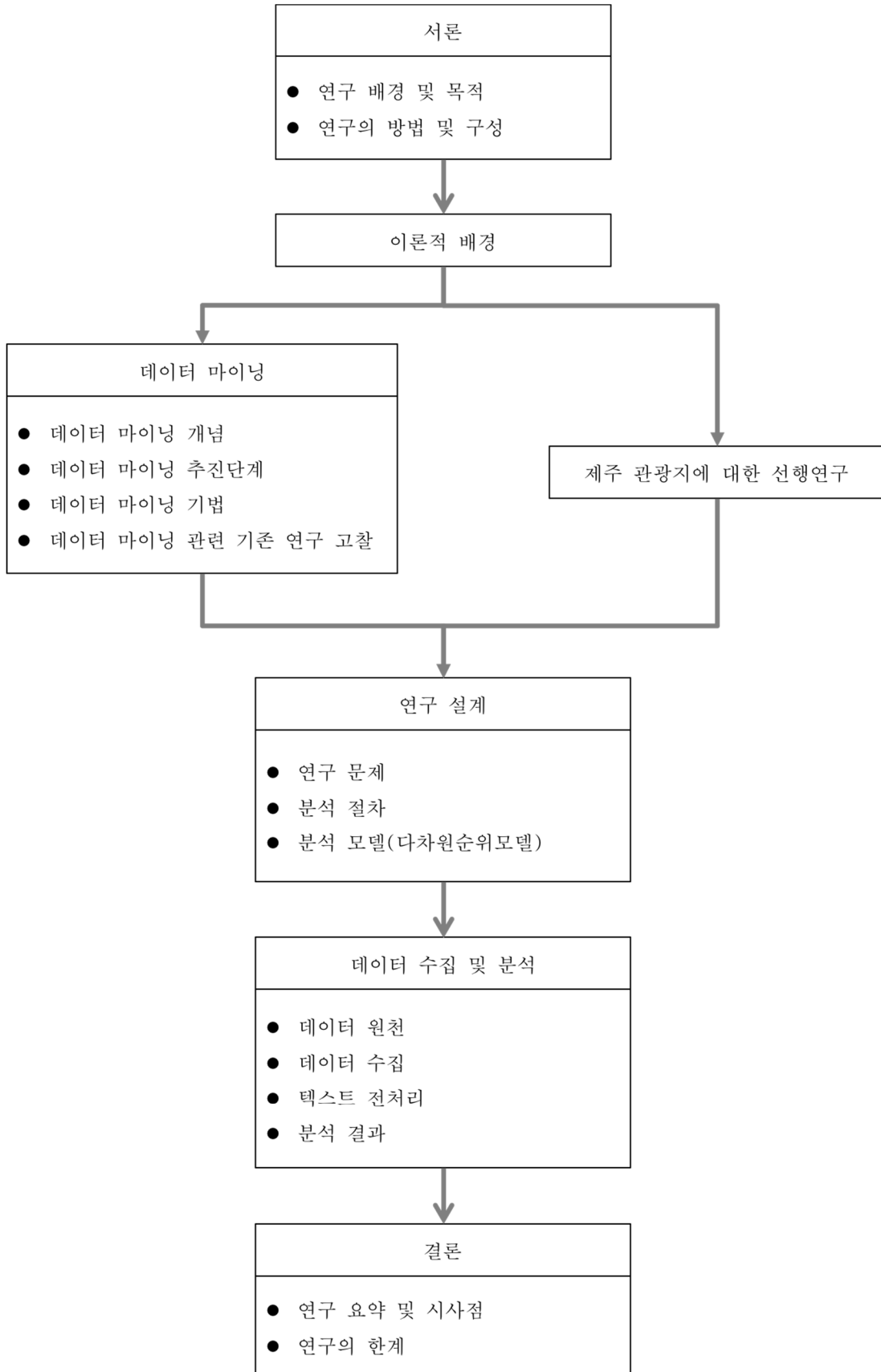
제 2 장 이론적 배경에서는 데이터 마이닝의 개념과 절차, 기법에 대해 이해하고 데이터 마이닝과 제주 관광 선행연구들을 고찰한다.

제 3 장 연구 설계에서는 연도별, 계절별 차원에서 연구 문제와 분석 절차를 설정하고, 분석 모델인 다차원 순위 모델을 제시한다.

제 4 장 데이터 수집 및 분석에서는 수집 데이터를 설정하고 수집 절차와 텍스트 전처리 과정을 거쳐 분석결과를 제시한다.

제 5 장 결론에서는 연구의 요약 및 시사점을 제시하고, 본 연구 과정에서 나타난 한계점을 설명한다.

<그림 1-1> 연구의 흐름도



## 제 2 장 이론적 배경

### 제 1 절 데이터 마이닝

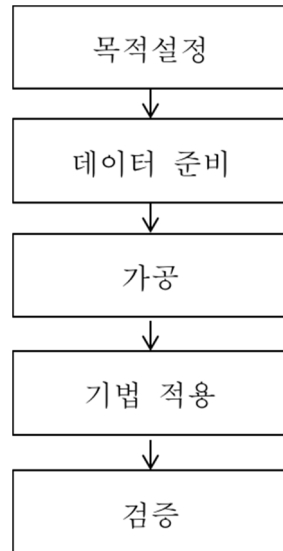
#### 1. 데이터 마이닝의 개념

빅 데이터 시대를 맞이하여 데이터 양은 실시간으로 기하급수적으로 증가하고 있다. 과거에는 기술의 한계, 비용 등의 이유로 분석할 수 없었던 빅 데이터를 분석 가능해짐에 따라 기존에 창출하지 못하였던 가치를 빅 데이터 속에서 창출하고 있다. 그 대표적인 기술의 예로 데이터 마이닝을 들 수 있다. 데이터 마이닝이란 광산에서 광물을 찾아내기 위해 많은 땅을 채굴하듯이, 대용량 데이터에서 탐색 및 분석하여 의미 있는 데이터 패턴을 찾아내고 의사결정에 도움을 주는 방법론이다. SNS 를 제공하는 웹사이트에 기술되어 있는 관광 관련 리뷰와 같은 원천 데이터에서 탐색 및 분석하여 가치 있는 정보를 찾아내 의사결정에 활용할 수 있도록 데이터 마이닝을 이용할 수 있다(송수미, 2001). 데이터 마이닝이 다루는 데이터 형태에 따라 구조화된 데이터(Structured data)와 비구조적 데이터(Unstructured data)로 나뉜다. 비구조적 데이터는 구조화가 어려운 영상, 음성 정보, 후기 데이터와 같은 일반 텍스트와 같은 데이터로 전 세계 데이터의 약 80% 이상을 차지한다(이정민, 2017). 기존의 데이터 마이닝 기법들은 구조화된 데이터를 전제로 하여 분석하기 때문에, 후기 데이터와 같은 비구조화된 데이터를 바로 분석하기엔 적합하지 않다(김근형, 오성열, 2009). 따라서 비구조화된 데이터를 구조화하는 전처리 과정이 필요하며, 전처리된 데이터에서 유의미한 데이터 패턴을 찾기 위해서는 데이터 특성에 맞는 데이터 마이닝 기법을 적용하는 것이 중요하다(허성완, 2016).

#### 2. 데이터 마이닝 추진 단계

데이터 마이닝의 추진 단계는 <그림 2-1>와 같이 진행 되며(한국데이터베이스진흥원, 2014), 각 단계의 세부 내용은 다음과 같다.

<그림 2-1> 데이터 마이닝의 추진 절차



자료 : 한국데이터베이스진흥원(2014). 연구자 재정리

#### [1 단계] 목적설정

데이터 마이닝의 목적을 분명하게 정의해야 한다. 목적이 없다면 데이터를 분석한 결과 값이 의사결정에 도움을 주지 못하는 경우가 발생 할 수 있기 때문이다. 또한 목적은 관련 관계자들 모두의 동의가 있어야 하며, 이해할 수 있어야 한다.

#### [2 단계] 데이터 준비

목적설정 단계에서 정한 목적에 부합하는 데이터를 수집하는 단계이다. 수집된 데이터 중에서 오류가 있는 데이터 또는 중복데이터가 있는 경우가 있을 것이다. 이러한 데이터를 제거하는 작업을 통해 데이터 품질을 확보하고 데이터 마이닝 기법을 적용하는데 문제가 없도록 한다.

#### [3 단계] 가공

데이터 마이닝 기법에 적용할 수 있도록 적합한 형식으로 데이터를 가공한다. 데이터 준비 단계에서 수집된 데이터는 대부분이 구조화 되어 있지 않은 비정형 데이터이기 때문에 데이터 마트 형식으로 만드는 시간이 많이 소요된다.

#### [4 단계] 기법 적용

준비된 데이터에 기법을 적용하여 목적에 맞는 정보를 추출하는 단계이다. 적용할 데이터 마이닝 기법 목적설정단계에서 정해졌어야 바람직하다.

#### [5 단계] 검증

데이터 마이닝으로 추출된 정보가 실제 업무에 적용할 수 있는가 검증하는 단계이다. 검증이 되었으면 업무에 적용할 수 있도록 해야한다.

### 3. 데이터 마이닝의 기법

데이터 마이닝은 목적에 따라 서로 다른 작업유형이 적용되게 된다. 예측이 목적일 땐 분류규칙(Classification), 설명이 목적일 때는 연관규칙(Association), 연속규칙(Sequence), 데이터군집화(Clustering) 유형으로 나뉜다. 네 가지 작업유형에 사용되는 기법으로는 회귀분석, Neural Networks, Decision Tree, 동시발생 매트릭스, K-means Clustering, 연관분석, OLAP 등이 있다(최지애, 2002). 본 연구에서 사용되는 연관분석, OLAP 를 중심으로 세부내용에 대해 살펴보도록 하겠다.

#### 1) 연관분석

연관분석은 데이터 마이닝 기법 중 가장 활발하게 연구가 되는 분야이다(김근형, 황병웅, 김민철, 2004). 대용량 데이터베이스에서 항목들간의 관련성 및 규칙을 찾아내기 위해 연관분석 기법을 사용한다. 항목 A 와 B 사이에서 'A → B' 와 같은 형식으로 나타나며 항목 A 가 나타나면 항목 B 가 나타날 확률이 높은 것을 의미한다. 이때 항목 A 는 조건부가 되며 항목 B 는 결과부가 된다. 예를 들어 맥주를 산 사람은 기저귀를 사는 비율이 높을 때 맥주는 조건부가 되며 기저귀는 결과부가 된다. 이러한 규칙은 모든 항목에 적용되지 않는다. 즉, 사용자가 지정하는 일정수준을 넘는 빈번하게 일어나는 규칙이어야 한다. 연관규칙이 생성이 되었을 때 유용한 규칙인지 뒷받침해 주는 기본적인 척도로 지지도(Support), 신뢰도(Confidence), 향상도(Lift)가 사용된다(김온실, 2001).



지지도(Support) : 항목 A 와 항목 B 사이에 연관규칙이 생성이 되었을 때 전체 트랜잭션에서 항목 A 와 항목 B 를 포함하는 트랜잭션의 비율을 나타낸다. 즉, 전체 트랜잭션에서 항목 A 와 항목 B 를 동시에 포함하고 있는 트랜잭션이 얼마나 되는지를 나타내주는 척도이다.

$$\text{지지도} = \frac{\text{항목 A 와 항목 B 를 포함하는 트랜잭션 수}}{\text{전체 트랜잭션 수}}$$

신뢰도(Confidence) : 항목 A 를 포함한 트랜잭션 중에서 항목 A 와 항목 B 가 동시에 포함될 수 있는 확률을 의미한다. 즉, 생성된 규칙에 대한 강도를 나타내는 척도이다.

$$\text{신뢰도} = \frac{\text{항목 A 와 항목 B 를 동시에 포함하는 트랜잭션 수}}{\text{항목 A 를 포함하는 트랜잭션 수}}$$

향상도(Lift) : 항목 B 가 단독으로 나타날 확률에 비해 항목 A 가 주어졌을 때 항목 B 가 나타날 확률의 증가 비율이다. 항목 A 와 B 간의 향상도의 값이 1 이면 독립적인 관계이며, 1 보다 크면 이 규칙은 양의 상관관계를 가지는 것이며, 결과를 예측하는데 있어서 우연적이 아닌 우수하다는 것을 의미한다. 마지막으로 1 보다 작으면 음의 상관관계를 가지고 있으며, 우연적 기회보다 도움이 되지 않는 결과값으로 분석이 가능하다.

$$\text{향상도} = \frac{\text{항목 A 와 항목 B 를 동시에 포함하는 트랜잭션 수}}{\text{항목 A 를 포함하는 트랜잭션 수} * \text{항목 B 를 포함하는 트랜잭션 수}}$$

## 2) 다차원분석(OLAP)

OLAP(Online Analytical Processing)는 1993년 Edgar Frank Codd에 의해 제안되었으며 "동적모델로부터 정보를 생성, 조작, 활성화를 종합하는데 필요한 역동적 기업분석"으로 정의하였다. 즉, 최종 사용자가 여러 차원으로 이뤄진 정보에 직접 접근하여 정보를 대화 형태로 분석하고 의사결정에 활용하는 과정이다 (위키백과, 2017).

#### (1) OLAP 의 특징

첫째, 분석에 활용되는 정보는 다차원적 형태이다. 이러한 특징은 다른 시스템과 OLAP 를 구분하는 가장 큰 개념이고, 때문에 OLAP 를 '다차원 분석'이라고도 불린다(한영임, 2005).

둘째, 최종 사용자는 정보처리 담당자와 같은 중간 매개자 없이 온라인 상에서 직접 데이터에 접근이 가능하다(김지민, 2003).

셋째, 최종 사용자는 정보를 분석할 때 대화식으로 분석 한다. 이런 특징때문에 시스템은 최종 사용자의 사고 흐름에 방해되지 않도록 질의 결과를 신속하게 제시하여야 한다(유한주, 2008).

넷째, 정의에 나온 것과 같이 최종사용자가 질의를 하여 제시된 결과를 통해 의사결정을 지원할 수 있어야 한다.

#### 4. 데이터 마이닝 관련 기존 연구 고찰

기존의 데이터 마이닝 연구에서는 효율성 중심의 기술 향상에 중점을 둔 연구들이 대부분이었으며, 또한 구조화된 데이터를 활용하고 있다. 김산성·김명원(2002)의 연구에서는 기존의 연관성 기법으로 분석하지 못하는 속성값을 함으로 가지고 있는 데이터베이스를 입력 형식에 대한 자료구조를 제안함으로써 효율적인 규칙을 생성할 수 있도록 하였다.

이해각(2004)의 연구에서는 연관성 규칙 생성 도중 통계적 추론에 따라 주어진 조건을 만족하는 항목이 있을 때 의사 결정을 내리는 알고리즘을 제안함으로써 기존의 연관성 규칙 알고리즘보다 탐색 시간을 단축하였다.

최근 연구들을 살펴보면 비구조화된 데이터를 활용한 연구들이 많이 이루어지고 있으며, 기술 향상의 연구에서 벗어나 기존의 기술, 연구와 융합한 연구들이 점차 늘어가고 있는 추세이다. 김근형·오성열(2009)의 연구에서는 온라인 고객 리뷰를 텍스트 마이닝 기술로 분석하기 위한 방법론을 제안하였으며, 보다 효과적이고 효율적으로 분석하기 위해 시장세분화 개념을 도입하였다.

장재영(2009)의 연구에서는 온라인 쇼핑몰에 나타난 소비자들의 후기를 저자가 제안하는 감성 분석 알고리즘을 이용하여 긍정 및 부정으로 자동 분류하는 시스템을 소개하고 알고리즘의 효율성을 검증하였다.

김유영·송민(2016)의 연구에서는 영화 리뷰와 평점을 이용하여 감성 분류기를 구축하였고, 이를 통해 감성점수와 영화 흥행 간의 상관관계를 밝혀냈다.

관광 학회에서도 비구조화된 데이터를 활용한 연구들이 점차 늘어가고 있는데 선행연구들을 살펴보면, 조완섭·조아·권가은·류관희(2015)는 뉴스, 블로그와 같은 비정형데이터와 충북 관광 관련 공공데이터를 이슈분석, 연관분석, 감성분석을 수행하여 충북 관광의 마케팅 전략 및 관광 진흥 정책 수립 방안을 제시하였다.

오익근·이태숙·전채남·윤영일(2015)은 주요 포털사이트(네이버, 다음, 구글)에서 '관광/여행' 키워드로 검색했을 때 나오는 웹페이지, 블로그, 뉴스 등에서 자료 수집을 하여 네트워크 분석을 하였다. 분석결과 1인 관광이 부각되고 있고, 제주도의 웹 가시성이 독보적임을 밝혀냈다.

류시영·유선욱(2017)은 블로그, 포털사이트 게시판, 트위터에 게시된 강원도 관광 관련 게시글을 수집하여 강원도 관광 및 지역에 대한 인식 및 관심사를 파악하고자 했다. 이를 통해 강원도 관광의 정책 수립 및 마케팅 전략을 위한 기초 자료를 제공하였다.

기존의 데이터 마이닝의 연구들을 살펴보면 분석 시간의 단축, 알고리즘 개선과 같이 효율성 중심의 기술향상을 하는 것을 알 수 있다. 최근 ICT의 발달과 스마트폰 보급의 확대로 SNS가 부각되었고, 이로 인해 실시간으로 다양한 데이터양은 급증하였다. 빅 데이터를 이용하여 새로운 가치를 찾고자 하는 연구가 활발히 진행되고 있으며, 최근 연구들을 살펴보면 기존과 달리 비구조화된 데이터를 활용한 연구들이 증가하고 있고 다른 분야와 데이터 마이닝을 융합한 연구뿐만 아니라 후기데이터를 이용한 감성 분석이 주를 이루고 있음을 알 수 있었다. 최근 관광 분야에서도 이런 추세에 빅 데이터를 이용한 연구들이 진행되고 있다. 선행연구를 살펴보면 주요포털사이트에서 제주도 관광이 관광객들 사이에서 주요 이슈로 떠오르고 있음을 알 수 있었고 데이터 마이닝 기법으로 소셜 빅 데이터를 분석하여 관광 정책을 위한 기초자료를 제공 및 시사점을 제공하는 연구들이 연구되고 있음을 알 수 있다. 제주도가 웹 상에서 주요 이슈로 떠오르는 만큼 이와 관련된 충분한 데이터가 있음에도 이를 활용한 연구가 미진하고 관련 연구들이 이루어져야 할 필요성이 제기된다.

## 제 2 절 제주 관광지에 대한 선행연구

관광지란 일상적인 생활공간에서 벗어난 관광객의 관광욕구 및 동기를 충족시켜주는 숙박시설, 편의시설, 다양한 관광활동이 가능한 기반시설 등의 제반시설이 갖춰진 공간을 의미한다(김정준, 2010). 관광지의 정의에서 나온 것과 같이 관광자의 관광욕구 및 동기에 관한 연구 또는 관광지 장소에 대한 연구에 대해 제주 지역을 중심으로 살펴보도록 하겠다.

송성신(2009)의 연구에서는 제주 관광지가 양적인 측면에서는 많은 성장을 이루었으나 질적인 측면에서는 관광지별로 차별화가 없어 경쟁력이 적다고 주장하고 있다. 이를 위해 관광객의 다양한 관광욕구를 충족하고, 경쟁력을 갖추기 위한 여섯 가지의 방안들을 제시하고 있다.

현홍준(2007)의 연구에서는 테마관광지를 자연형, 인공형, 문화형 세 가지로 분류하여 유형별 관광객 욕구, 만족 속성의 차이점과 관광객 욕구와 만족 속성 간의 상관관계가 있는지를 파악하는 연구를 진행하였다. 관광객의 욕구에서는 대동소이 하지만 문화형 테마관광지를 찾는 관광객은 학습 욕구가 가장 높게 평가가 되었고, 문화형 테마관광지에서 시설 편의성 요인과, 매력성요인에서 가장 높은 만족을 얻은 것을 밝혀냈다. 마지막으로 이용성, 매력성, 기념품 독특성, 서비스와 기분전환 욕구 요인과 학습욕구 만족요인에서 영향을 미치고 있다는 연구결과를 도출해 내었다.

정규렬(2013)는 관광지의 장소 정체성에 따라 방문 동기와 만족이 달라진다고 주장하고 있다. 이를 제주 올레 4 개 코스 중심으로 연구를 실시하였다. 연구결과 제주 올레 방문 동기는 일상탈출, 자연감상, 건강증진 등의 6 개 요인으로 도출되었으며, 제주올레의 정체성은 자연미, 문화학습, 심리적 안정 3 개 요인으로 도출되었다. 제주 올레 코스별로 다른 정체성을 가지고 있다고 연구가 되었는데 이에 따라 방문 동기와 만족이 달라짐을 확인되었다.

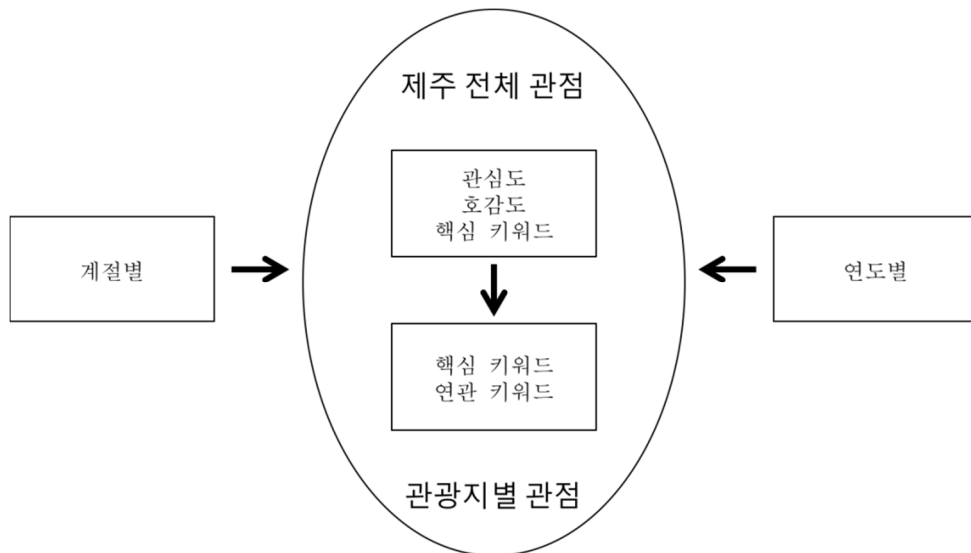
앞선 선행연구를 살펴보면 전반적으로 특정 장소 혹은 유형별로 나누어 연구가 이루어지고 있거나 관광객의 어떠한 욕구를 가졌는지에 대한 연구가 없이 관광지에 발전 방향을 제시하는 연구들이 대부분이었기 때문에 전반적인 제주 관광지에 대한 관광객의 인식 분석을 통한 제주 관광 전략 수립의 필요성이 제기되고 있다.

## 제 3 장 연구설계

### 제 1 절 연구문제

본 연구에서는 제주 관광지에 대한 관광객의 인식을 알아보고자 <그림 3-1>과 같은 차원에서 연구문제를 설정하고자 하였다.

<그림 3-1> 계절, 연도 차원에서의 연구문제



계절과 연도를 차원으로 제주 관광지 전체의 관심도, 호감도, 핵심키워드와 관광지별 관점에서 핵심 키워드, 연관 키워드를 통해 관광객의 인식을 알아보고자 하였으며, 도출된 연구문제는 다음과 같다.

[연구문제 1] 제주에서 관심도가 높은 관광지는 어디인가?

[연구문제 2] 제주에서 호감도가 높은 관광지는 어디인가?

[연구문제 3] 제주관광에서 핵심 키워드는 무엇인가?

[연구문제 4] 관광지별 핵심 키워드는 무엇인가?

[연구문제 5] 관광지별 핵심 키워드들 사이의 연관성은 어떻게 달라지는가?

## 제 2 절 분석 절차

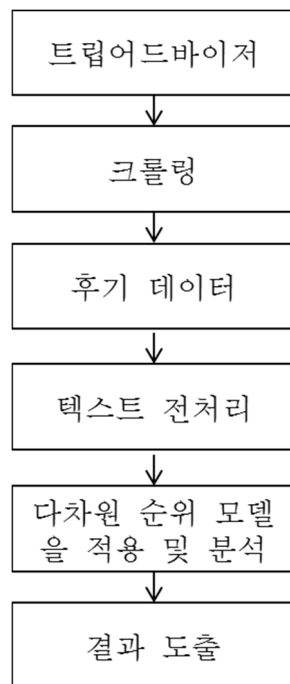
다차원 순위 모델을 기반으로 관광 후기 분석을 하기 위한 분석 절차로 <그림 3-2>와 같이 설계하였다. 구체적인 절차는 다음과 같다.

첫째, Tripadvisor 의 후기 데이터를 웹 크롤링 기법을 이용하여 수집한다.

둘째, 수집된 후기 데이터를 다차원 순위 모델에 적용할 수 있는 형태로 변환하기 위해 구문 분석 및 명사 추출과 같은 텍스트 전처리를 한다.

셋째, 수집되고 변환된 데이터를 다차원 순위 모델을 이용하여 분석한다. 마지막으로 분석 결과를 통한 시사점을 제시한다.

<그림 3-2> 연구의 분석 절차

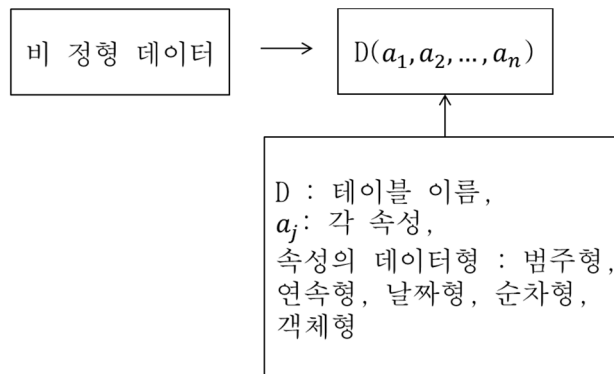


### 제 3 절 분석 모델(다차원 순위 모델)

빅 데이터를 다양한 방식으로 분석할 수 있지만, 특정 속성을 대상으로 순위를 비교하는 것은 기본적으로면서 효과성이 높은 정보가 된다. 본 논문에서 제안하는 다차원 순위 모델(Multi-dimensional Ranking Model)은 빅 데이터 속의 특정 속성값에 대한 순위를 다차원 관점에서 분석할 수 있는 모형이다.

빅 데이터를 다차원 순위 모델로 분석하기 위해서는 빅 데이터를 정형화된 테이블 구조 형태인  $D(a_1, a_2, \dots, a_n)$ 로 변환하여야 한다(D: 임의의 테이블 이름,  $a_i$ : 임의의 각 속성(attribute)). 임의의 속성  $a_i$ 의 데이터형은 크게 범주형, 연속형, 날짜형, 순차형, 객체형으로 구분하여 정의하여야 한다. 범주형은 속성값이 한정된 값 중에서만 올 수 있는 데이터형이며 주로 문자형의 형태를 갖는다. 연속형은 속성값으로 연속된 숫자 값을 가질 수 있으며 산술연산이 가능한 데이터형이다. 날짜형은 년도/월/일/시간으로 표현되는 데이터형이며 순차형은 일련번호 등과 같이 연속해서 증가하는 값을 갖는 데이터형이다. 객체형은 텍스트, 이미지, 동영상 등과 같은 복잡한 데이터형을 의미한다.

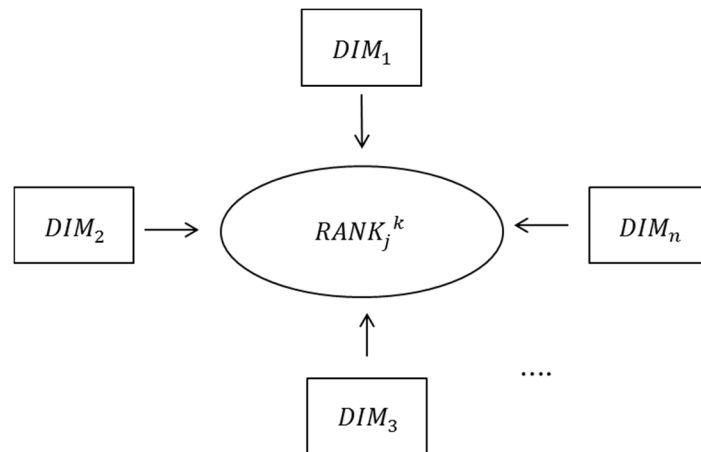
<그림 3-3> 비 정형 빅 데이터의 정형화



관광 후기 빅 데이터를 T(일련번호, 관광지, 후기 텍스트, 평가점수, 작성일자)와 같은 스키마(schema)로 정형화했을 경우, 일련번호 속성은 순차형, 관광지 속성은 범주형, 후기 텍스트 속성은 객체형, 평가점수 속성은 연속형, 작성 일자는 날짜형으로 변환된다.

<그림 3-4>은 다차원 순위 모델의 개념도를 나타내고 있다.  $RANK_j^k$ 는 순위평가 대상 속성  $a_j(j=1, \dots, n)$ 의 속성값(그룹)들에 대한  $m_k(k=1, \dots, m)$ 형 순위평가결과를 의미한다.  $DIM_j$ 는 순위평가결과를 바라보는 임의의 차원을 의미한다. 예를 들면, 정형화된 관광 후기 빅 데이터 T에서 관광지 호감도 순위 ( $RANK_{\text{관광지}}^{\text{호감도}}$ )를 계절 차원( $DIM_{\text{계절}}$ )에서 보거나 관광지 관심도 순위 ( $RANK_{\text{관광지}}^{\text{관심도}}$ )를 연도 차원( $DIM_{\text{연도}}$ )에서 분석하는 경우가 이에 해당된다. 상품평 빅 데이터에서는 상품순위( $RANK_{\text{상품}}^{\text{호감도}}$ )를 지역 차원( $DIM_{\text{지역}}$ ) 또는 성별 차원( $DIM_{\text{성별}}$ )로 바라보기 위하여 다차원 순위 모델을 응용할 수 있다.  $RANK_j^k$ 는 관광지 속성값(그룹)의 정렬된 형태로 표현된다. 즉,  $RANK_{\text{관광지}}^{\text{호감도}}$ 는 {관광지 1, 관광지 2, 관광지 3, ...}과 같은 형태로 나타나는데, 관광지 1의 호감도가 제일 높다는 의미가 된다.

<그림 3-4> 다차원 순위 모델의 개념도



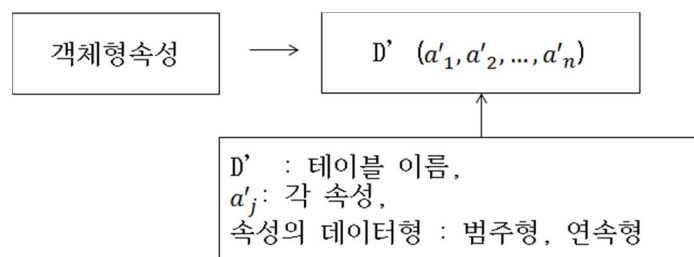
$RANK_j^k$ 에서  $j$ 는 범주형 속성 중에서 선택된다. 정형화된 관광 후기 빅 데이터 T(일련번호, 관광지, 텍스트리뷰, 평가점수, 작성일자)에서  $RANK_{\text{관광지}}^{\text{호감도}}$ 의 '관광지'는 T의 관광지 속성으로부터 선택된 것이다.  $RANK_j^k$ 의  $k$ 는 속성  $j$ 의 속성값 빈도수를 계산하거나 연관된 연속 속성으로부터 유도된다. T에서  $RANK_{\text{관광지}}^{\text{관심도}}$ 의 '관심도'는 속성값 그룹에 대한 텍스트 리뷰의 빈도수로부터 도



출 될 수 있으며  $RANK_{\text{관광지}}^{\text{호감도}}$ 의 '호감도'는 관광지 속성값 그룹에 대한 평가점수의 평균으로부터 계산될 수 있다.  $DIM_j$ 는 범주형 속성이나 낱짜형 속성으로부터 유도될 수 있다. T에서  $DIM_{\text{계절}}$ 은 작성일자 속성으로부터 유도할 수 있다.

정형화된 테이블의 객체형속성은 <그림 3-5>처럼 재정형화 후 순환적으로 다차원 순위 모델을 적용할 수 있다. 객체형속성을 정형화한 테이블의 속성은 범주형과 연속형으로 구성된다. 객체형 속성이 텍스트인 경우 명사, 동사, 형용사 등을 추출하여 각각에 대응하는 범주형 속성에 담을 수 있다. 필요한 경우 연속형 속성을 정의하여 각 범주형 속성값들의 출현 빈도수를 대응시킬 수 있다. 객체형 속성이 이미지나 동영상인 경우 우선적으로 이미지 및 동영상의 내용을 설명하는 텍스트를 생성한 후 텍스트 전처리 과정을 순환적으로 적용하여야 한다.

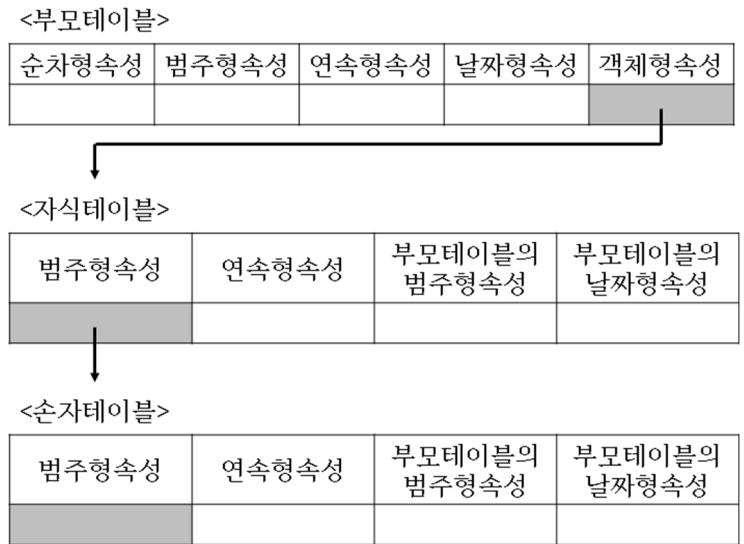
<그림 3-5> 객체형 속성의 정형화



객체형 속성을 재정형화한 테이블을 다차원적으로 분석하기 위해서는 <그림 3-6>처럼, 기존 부모 테이블의 범주형 속성과 낱짜형 속성이 재정형화 테이블에 포함되거나 참조될 수 있도록 하여야 한다. 정형화된 관광 후기 빅 데이터 T(일련번호, 관광지, 텍스트리뷰, 평가점수, 작성일자)에서 텍스트리뷰는 객체형 속성이며 각 리뷰 값들은 형태소분석기에 의하여 명사, 형용사, 동사 등으로 추출·분할되어 <그림 3-6>처럼 <자식테이블>의 범주형 속성에 들어간다. 명사, 형용사, 동사 등의 각 출현빈도 수를 담기 위한 연속형 속성을 정의할 수 있다. <자식테이블>의 범주형 속성의 속성값들은 범주형 데이터의 마이닝(mining)에 적합한 연관규칙탐사기법을 적용하여 분석할 수 있으며 분석결과인 연관규칙들은 <손자테이블>의 범주형 속성값이 된다. 연관규칙의 지지도(support)와 신뢰도(confidence) 등과 같은 규칙 평가 척도 값은 <손자테이블>의 연속형 속성의 값이 되며 순위에 기반한 연관규칙 검색에 활용된다. <손자테이블>에도 범주형 속

성인 연관규칙에 대한 다차원분석이 가능할 수 있도록 부모테이블의 범주형 속성과 날짜형 속성이 재정형화 테이블에 포함되거나 참조될 수 있도록 하여야 한다.

<그림 3-6> 재정형화된 테이블의 속성 유형



## 제 4 장 데이터 수집 및 분석

### 제 1 절 데이터 원천

본 연구에서는 세계적인 관광 웹사이트인 Tripadvisor 에 게시된 256 개 제주 명소 중 후기가 있는 195 개 관광지를 대상으로 삼았다. Tripadvisor 는 2000 년에 설립되어 호텔뿐만 아니라 음식점, 관광명소 등 실제 사용자의 후기와 정보를 제공해 준다. Tripadvisor 사이트는 매월 약 3 억 9 천만 명의 사용자가 이용하며, 4 억 3,500 만 건에 이르는 후기 및 평가를 보유하고 있다(tripadvisor, 2017). 2012 년부터 한국어 서비스를 제공해 주고 있지만, 한국어 리뷰 수는 많지 않은 실정이다. <그림 4-1>와 같이 사용자는 관광지별로 후기제목, 평점, 후기작성일, 내용, 방문날짜로 작성할 수 있으며, 본 연구에서는 평점, 내용, 방문날짜를 수집 대상으로 삼았다.

<그림 4-1> Tripadvisor 의 후기 작성 예시

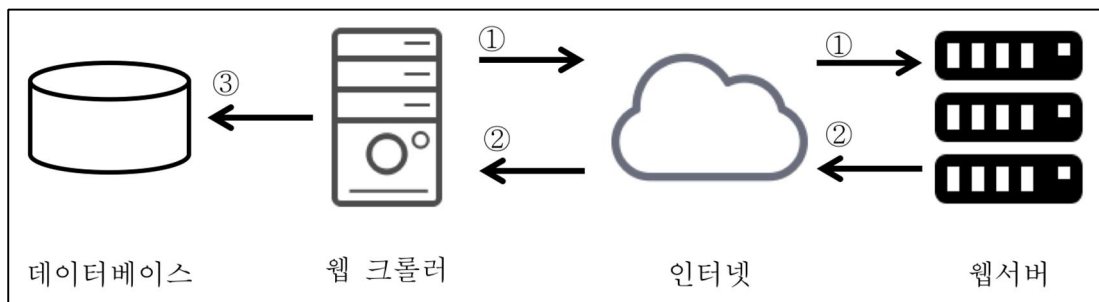


자료: Tripadvisor(2017)

## 제 2 절 데이터 수집

본 연구에서는 데이터를 수집하기 위해 웹 크롤러 기법을 이용하였다. 웹 크롤러는 웹 어플리케이션에서 제공하는 웹 페이지의 자원을 자동으로 수집하는 기술로(송호길, 2011) ants, automatic indexer, bots, worms 등으로 불린다. 웹 크롤링의 흐름도를 간략하게 나타내면 <그림 4-2>와 같다. 웹 크롤러가 웹 서버에 웹 페이지를 요청하면, 웹 서버는 요청된 웹 페이지 정보를 웹 크롤러에 응답하고, 수집된 웹 페이지를 분석하여 데이터베이스에 저장하게 된다.

<그림 4-2> 웹 크롤링의 흐름도



하지만 기존의 웹 크롤링 기법으로는 자바스크립트의 Ajax 등으로 구현한 동적 페이지에서 놓치는 자원들이 많으므로 자바스크립트 연산을 수행할 수 있는 웹 크롤러가 필요하고, 본 연구에서는 이를 지원해주는 JAVA API 인 selenium 을 사용하였다. 데이터 수집을 위한 절차로는 <그림 4-3>와 같으며, 상세 내용은 다음과 같다.

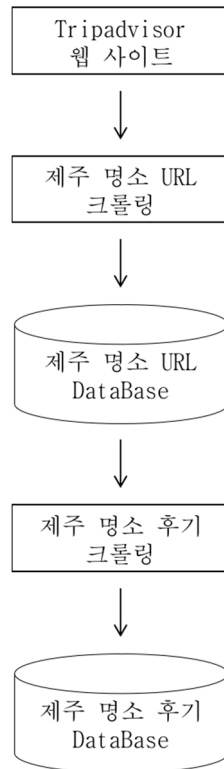
첫째, 웹 크롤러를 이용하여 제주 명소 리스트가 나와 있는 웹 페이지를 Tripadvisor 의 웹 서버에 요청한다.

둘째, 응답받은 웹 페이지에서 리스트의 앵커 태그에 설정된 URL 을 크롤링하여 제주 명소 URL Database 에 저장한다.

셋째, 제주 명소 URL Database 에 저장된 URL 을 하나씩 불러와서 지정된 웹 페이지를 Tripadvisor 의 웹 서버에 요청한다.

넷째, 응답받은 웹 페이지에서 후기데이터만을 크롤링하여 제주 명소 후기 Database 에 저장한다.

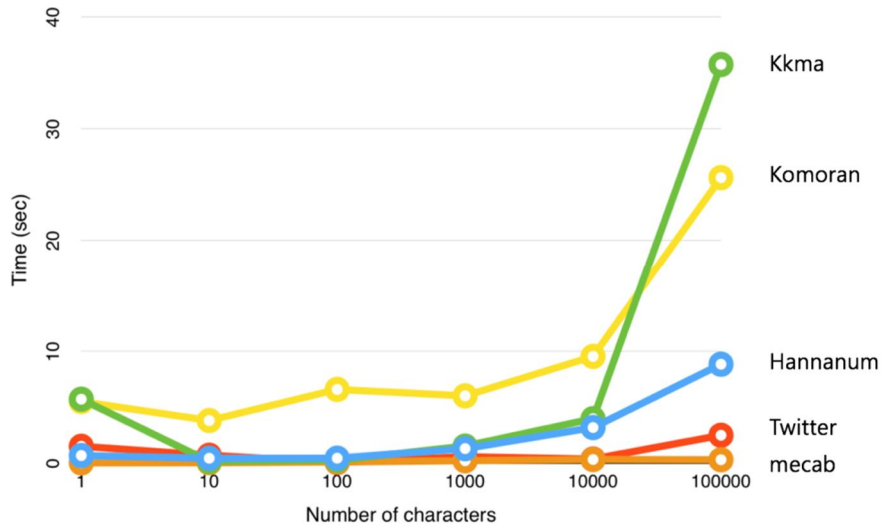
<그림 4-3> 데이터 수집 절차



### 제 3 절 텍스트 전처리(Text Preprocessing)

텍스트 전처리는 텍스트 데이터를 컴퓨터가 처리할 수 있도록 변환하는 작업이다(김근형, 오성열, 2009). 텍스트 데이터를 전처리하는 과정으로 문장을 나눠주는 토큰화와 나눠진 문장의 품사를 결정하는 품사 태깅 순으로 진행된다. 이를 위한 형태소 분석기로 Kkma, Komoran, Hannanum, Twitter, Mecab 등이 있다. 본 연구에서는 Twitter 에서 제공하는 JAVA API 인 twitter-korean-text 를 이용하였다. twitter-korean-text 는 <그림 4-4>에서 보듯이 타 형태소 분석기와 비교했을 때 빠른 분석 속도를 가지고 있으며, 형태소 분석 첫 단계인 정규화를 통해 "입니닥ㅋㅋ"를 "입니다ㅋㅋ"와 같이 후기데이터에서 흔히 나오는 오타자를 교정해주기 때문에 후기 데이터를 전처리 하기에 적합하다.

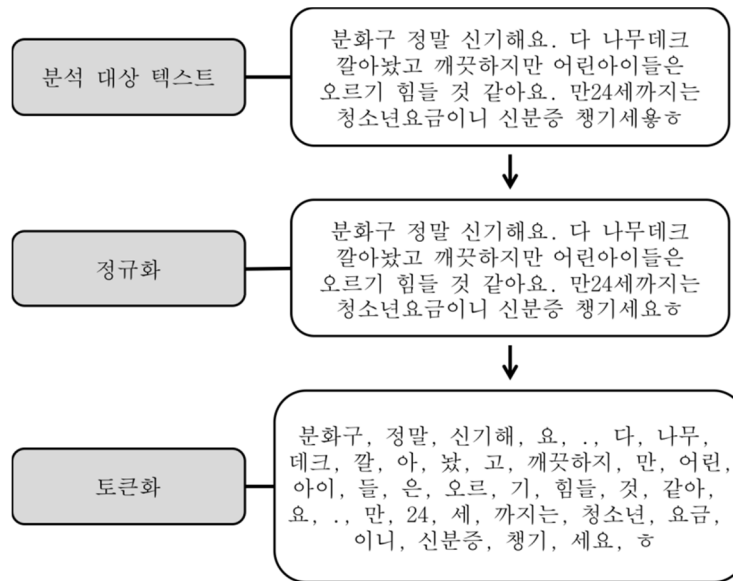
<그림 4-4> 형태소 분석기 간의 처리 속도 비교



자료: KoNLPy(2017)

twitter-korean-text 는 크게 오타자를 교정해주는 정규화, 띄어쓰기를 기준으로 문장을 나눠주고, 품사 태깅을 해주는 토큰화, 형용사와 동사를 사전에 나와 있는 형태로 바꿔주는 어근화, 형용사, 명사, 접속사를 추출하는 어구추출 단계로 이뤄진다. 본 연구에서는 명사 추출을 목적으로 하였기 때문에 토큰화 단계까지 진행하였으며, twitter-korean-text 의 실제 예시는 <그림 4-5>와 같다.

<그림 4-5> twitter-korean-text 를 이용한 형태소 분석 예시



## 제 4 절 분석 결과

### 1. 데이터 수집 결과 및 표본의 특성

본 연구에서는 웹 크롤러를 이용하여 Tripadvisor 에 게시된 256 개 제주 명소 중 후기가 있는 195 개 관광지를 대상으로 후기데이터를 수집하였다. 수집된 4,807 개의 후기 데이터 중 계절별, 연도별 차원으로 분석하기 위해 비교적 데이터 수가 많은 2015 년과 2016 년의 후기 데이터를 분석하였고, 형태소 분석을 통해 지시대명사와 같이 관광과 관련 없는 명사를 제외하여 43,640 개의 명사를 추출하였다. 수집 결과는 다음 <표 4-1>과 같다.

<표 4-1> 데이터 수집 결과

	전체 데이터	분석 데이터
관광지	195	185
후기 데이터	4,807	4,037
명사	96,436	43,640
방문 년도	2008 년 ~ 2017 년	2015 년, 2016 년

분석 대상인 185 개의 관광지를 다차원 순위 모델에 적용하기 위하여 차원을 계절별(봄, 여름, 가을, 겨울)과 연도별(2015 년, 2016 년)로 분류하였다. <표 4-2>는 분류한 결과이며, 계절별 후기 수를 살펴보면 여름이 1,208 개로 가장 많았으며, 겨울이 727 개로 가장 낮게 나왔다. 연도별 후기 수는 2015 년이 2,279 개로 2016 년 1,758 개보다 많이 나왔음을 알 수 있다.

<표 4-2> 차원별 관광지 개수

	관광지 수	비율
봄	1,049	25.98%
여름	1,208	29.92%
가을	1,053	26.08%
겨울	727	18.01%
전체	4,037	100%

2015 년	2,279	56.45%
2016 년	1,758	43.55%
전체	4,037	100%

## 2. 분석 결과

앞서 제시한 연구문제를 해결하기 위해 분석 도구로 R 를 이용하였으며, R 의 주요 패키지로는 그룹별 빈도분석을 위한 dplyr 와 연관성 분석을 위한 arules 를 사용하였다.

### 1) 제주에서 관심도가 높은 관광지

제주에서 관심도가 높은 관광지를 분석하기 위해 계절과 연도를 차원으로 하는 관광지별 후기 데이터 개수 중 상위 10 개를 파악해 보았으며 결과는 <표 4-3>, <표 4-4>와 같다.

<표 4-3> 계절별 제주에서 관심도가 높은 관광지

순위	봄			여름		
	관광지명	리뷰수	비율	관광지명	리뷰수	비율
1	성산일출봉	66	6.29%	성산일출봉	67	5.55%
2	우도	52	4.96%	우도	67	5.55%
3	오설록 티 뮤지엄	47	4.48%	오설록 티 뮤지엄	61	5.05%
4	쇠소깍	37	3.53%	협재 해수욕장	48	3.97%
5	주상절리대	34	3.24%	쇠소깍	41	3.39%
6	에코랜드	33	3.15%	만장굴	37	3.06%
7	협재해수욕장	32	3.05%	함덕 서우봉 해변	37	3.06%
8	아쿠아플라넷제주	30	2.86%	주상절리대	33	2.73%
9	천지연 폭포	28	2.67%	천지연폭포	33	2.73%
10	용두암	27	2.57%	정방폭포	41	2.57%
	...	...	...	...	...	...
	전체	1,049	100%	전체	1,208	100%



순위	가을			겨울		
	관광지명	리뷰수	비율	관광지명	리뷰수	비율
1	성산일출봉	63	5.98%	성산일출봉	49	6.74%
2	우도	59	5.60%	오설록 티 뮤지엄	41	5.64%
3	오설록 티 뮤지엄	51	4.84%	동문 재래시장	23	3.16%
4	에코랜드	32	3.04%	쇠소깍	23	3.16%
5	함덕 서우봉 해변	31	2.94%	우도	22	3.03%
6	쇠소깍	27	2.56%	테디베어 박물관	22	3.03%
7	주상절리대	27	2.56%	한라산 국립공원	22	3.03%
8	협재 해수욕장	26	2.47%	아쿠아플라넷제주	21	2.89%
9	산굼부리	25	2.37%	주상절리대	20	2.75%
10	아쿠아플라넷제주	25	2.37%	비자림	19	2.61%
	...	...	...	...	...	...
	전체	1,053	100%	전체	727	100%

<표 4-4> 연도별 제주에서 관심도가 높은 관광지

순위	2015년			2016년		
	관광지명	리뷰수	비율	관광지명	리뷰수	비율
1	성산일출봉	153	6.71%	성산일출봉	92	5.23%
2	오설록 티 뮤지엄	114	5.00%	우도	92	5.23%
3	우도	108	4.74%	오설록 티 뮤지엄	86	2.89%
4	쇠소깍	78	3.42%	함덕 서우봉 해변	63	3.58%
5	주상절리대	71	3.12%	동문 재래시장	54	3.07%
6	협재해수욕장	64	2.81%	협재 해수욕장	51	2.90%
7	아쿠아플라넷제주	55	2.41%	쇠소깍	50	2.84%
8	에코랜드	51	2.24%	에코랜드	50	2.84%
9	테디베어 박물관	50	2.19%	용두암	49	2.78%
10	천지연 폭포	49	2.15%	천지연 폭포	48	2.73%
	...	...	...	...	...	...
	전체	2,279	100%	전체	1,758	100%

분석결과 모든 계절과 연도에서 성산일출봉이 가장 많은 관심을 받는 것으로 나왔으며, 이어서 봄, 여름, 가을에 우도와 오설록 티 뮤지엄, 겨울에 오설록 티 뮤지엄, 동문 재래시장, 2015 년도에 우도와 오설록 티 뮤지엄, 2016 년도에 오설록 티 뮤지엄, 우도가 관심을 받는 것으로 분석이 되었다. 만장굴을 보면 여름철에 비해 비교적 낮은 순위(봄 : 27 위, 가을 : 31 위, 겨울 : 36 위)를 받던 만장굴이 여름철에는 상위 10 위 안에 포함된 것으로 분석이 되었는데 이는 만장굴이 사계절 내내 섭씨 10 ~ 21 도를 유지하고 있기 때문에 더운 여름날 더위를 피하기 위해 많이 찾는 것으로 파악된다. 실제로 여름철 만장굴 후기 데이터를 보면 '시원하다', '춥다'와 같은 단어를 포함한 후기데이터는 27 개로 더위를 피하기 위해 만장굴을 많이 찾았다. 봄, 여름, 가을에 상위 2 번째를 기록한 우도인 경우 겨울철에는 상위 5 번째로 하락한 것으로 분석이 되었는데, 이는 동절기에는 배운항시간이 하절기보다 2 시간 가량 짧아지기 때문이라 파악되고, 겨울철에는 동문 재래시장, 테디베어 박물관, 아쿠아플라넷제주와 같이 추위를 피할 수 있는 인공형 관광지가 다른 계절에 비해 많이 순위에 올라와 있는 것으로 분석되었다. 연도 차원에서 상위 10 위권 안에 있는 관광지 중 가장 순위 변동이 높았던 관광지는 동문재래시장으로 2015 년도에 17 위에서 2016 년도에 5 위를 기록하였다. 동문재래시장의 후기데이터 분석해 보면 공항 근처에서 기념품을 사기 위해 관광객들이 많이 찾고 있으나, 주차공간 문제, 본래의 모습을 잃은 듯한 관광객 중심의 시장 등의 아쉬운 점을 갖고 있는 것으로 분석되었으며, 이는 앞으로 동문재래시장이 개선해야 할 것으로 판단된다.

## 2) 제주에서 호감도가 높은 관광지

제주에서 호감도가 높은 관광지를 분석하기 위해 계절과 연도를 차원으로 하는 관광지별 평가 점수의 평균 중 상위 10 개를 파악해 보았다. 평가 점수는 5 점 만점으로 사용자가 해당 관광지의 평가점수를 내린 것이며, 좀 더 정확한 분석을 위해 후기데이터가 15 개 이상인 관광지를 대상으로 분석하였다. 결과는 <표 4-5>, <표 4-6>과 같다.

<표 4-5> 계절별 제주에서 호감도가 높은 관광지

	봄	여름	가을	겨울
순위	관광지명 (평점평균)	관광지명 (평점평균)	관광지명 (평점평균)	관광지명 (평점평균)
1	비자림 (4.73)	한라산 국립공원 (5.00)	한라산 국립공원 (4.77)	한라산 국립공원 (4.73)
2	함덕서우봉해변 (4.73)	함덕서우봉해변 (4.70)	함덕서우봉해변 (4.65)	함덕서우봉해변 (4.67)
3	협재 해수욕장 (4.72)	우도 (4.64)	비자림 (4.60)	우도 (4.64)
4	우도 (4.71)	에코랜드 (4.58)	우도 (4.58)	성산일출봉 (4.59)
5	한림공원 (4.67)	주상 절리대 (4.55)	성산일출봉 (4.41)	비자림 (4.53)
6	성산일출봉 (4.58)	용머리 해안 (4.53)	주상 절리대 (4.37)	주상 절리대 (4.25)
7	에코랜드 (4.55)	비자림 (4.52)	협재 해수욕장 (4.31)	쇠소깍 (4.09)
8	쇠소깍 (4.46)	협재 해수욕장 (4.44)	산굼부리 (4.24)	아쿠아플라넷제주 (4.00)
9	한라산 국립공원 (4.44)	천지연 폭포 (4.42)	정방 폭포 (4.23)	테디베어 박물관 (3.95)
10	천지연 폭포 (4.39)	성산 일출봉 (4.39)	천지연 폭포 (4.21)	동문 재래시장 (3.83)

<표 4-6> 연도별 제주에서 호감도가 높은 관광지

	2015년	2016년
순위	관광지명 (평점평균)	관광지명 (평점평균)
1	한라산 국립공원 (4.71)	한라산 국립공원 (4.76)
2	송악산 (4.65)	함덕 서우봉 해변 (4.73)
3	우도 (4.65)	비자림 (4.70)
4	플레이케이팝 (4.65)	김녕 성세기 해변 (4.68)
5	함덕 서우봉 해변 (4.62)	우도 (4.63)
6	제주 절물 자연 휴양림 (4.57)	용머리 해안 (4.54)

7	협재 해수욕장 (4.52)	협재 해수욕장 (4.53)
8	우도등대공원 (4.50)	한림공원 (4.52)
9	성산 일출봉 (4.49)	천지연 폭포 (4.50)
10	비자림 (4.49)	성산일출봉 (4.48)

분석결과 봄에는 비자림(4.73), 함덕 서우봉해변(4.73), 협재 해수욕장(4.72), 여름에는 한라산 국립공원(5.00), 함덕 서우봉 해변(4.70), 우도(4.64), 가을에는 한라산 국립공원(4.77), 함덕 서우봉 해변(4.65), 비자림(4.60), 겨울에는 한라산국립공원(4.73), 함덕 서우봉 해변(4.67), 우도(4.64), 2015년에는 한라산 국립공원(4.71), 송악산(4.65), 우도(4.65), 2016년에는 한라산 국립공원(4.76), 함덕 서우봉 해변(4.73), 비자림(4.70) 순으로 분석 되었다. 전체적으로 한라산 국립공원이 평점 평균 4.73 으로 가장 높은 호감도를 받는 관광지로서 후기데이터를 분석한 결과 등산 과정은 힘들었지만, 한라산이 주는 빼어난 자연 경관으로 관광객의 마음을 사로 잡은것으로 분석되었다. 앞선 관심도 분석에서 가장 많은 후기데이터를 보유한 성산일출봉인 경우 계절, 연도 차원에서 관심에 비해 낮은 호감도를 받는 것으로 분석되었으며 관광객이 가장 많이 표현한 불만 사항으로는 뛰어난 경관을 가지고 있지만 많은 중국인으로 인해 감상이 불편했다는 점을 말하고 있다.

### 3) 제주 관광에서의 핵심 키워드

제주 관광에서 핵심 키워드를 분석하기 위해 계절과 연도를 차원으로 하는 키워드 빈도 분석을 하였으며 관광지 명을 제외한 상위 10 개를 추출하였다. 결과는 <표 4-7>, <표 4-8>과 같다.

<표 4-7> 계절별 제주 관광의 핵심 키워드

순위	봄		여름	
	키워드	언급 횟수	키워드	언급 횟수
1	아이	148	바다	175
2	바다	145	아이	164

3	폭포	94	해변	150
4	산책	91	폭포	120
5	해변	87	산책	96
6	가족	69	녹차	74
7	자연	68	자연	69
8	녹차	63	박물관	65
9	박물관	45	가족	62
10	카약	37	카페	59

순위	가을		겨울	
	키워드	언급 횟수	키워드	언급 횟수
1	아이	139	아이	109
2	바다	118	바다	92
3	해변	114	산책	67
4	폭포	88	한라산	52
5	산책	84	녹차	50
6	녹차	66	폭포	48
7	가족	63	해변	44
8	자연	60	가족	43
9	오름	45	박물관	39
10	박물관	41	시장	33

<표 4-8> 연도별 제주 관광의 핵심 키워드

순위	2015년		2016년	
	키워드	언급 횟수	키워드	언급 횟수
1	아이	324	바다	253
2	바다	277	아이	236
3	산책	199	해변	217
4	폭포	179	폭포	171
5	해변	178	산책	139
6	녹차	138	녹차	115
7	자연	138	가족	113

8	박물관	134	카페	89
9	가족	124	자연	84
10	한라산	80	시장	67

분석 결과 모든 차원에서 아이 또는 바다가 첫 번째, 두 번째 순위를 받고 있으며, 이어서 봄에는 폭포(94), 산책(91), 여름에는 해변(150), 폭포(120), 가을에는 해변(114), 폭포(88), 겨울에는 산책(67), 한라산(52), 2015년에는 산책(199), 폭포(179), 2016년에는 해변(217), 폭포(171) 순으로 나타났다. 모든 차원에서 아이가 언급되는 빈도가 높았는데 제주도를 찾은 관광객의 유형 중 가족형 관광객의 비중이 높은 것으로 분석된다. 2016 제주특별자치도 방문관광객 실태조사에 따르면 동반자 유형 중 가족/친척이 2015 년도에는 22.4%로 2 번째로 높았으며 2016 년도에는 31.9%로 가장 높게 나왔는데 본 연구결과와 제주특별자치도의 분석결과가 일치함을 알 수 있다. 또한 바다, 폭포, 해변과 같은 자연과 관련된 키워드가 박물관, 시장에 비해 많이 언급되었는데 제주를 찾는 관광객의 주된 관심이 자연형 관광지에 집중된 것으로 분석된다.

#### 4) 관광지별 핵심 키워드

앞선 선행연구를 통해 관광객 인식에 따른 관광지의 정체성이 방문 동기에 영향을 미친다는 것을 확인했다. 만약 관광객이 해당 관광지에 대한 긍정적 또는 부정적 인식을 한다고 하면 인식의 강화 또는 인식 전환 등을 위한 포지셔닝 전략이 필요할 것이다. 이에 대한 기초자료를 제공하기 위해 관광지별 핵심 키워드를 계절과 연도를 차원으로 분석하였으며, 관광지별 키워드 빈도 분석 관광지명을 제외한 상위 10 개를 추출하였다. 결과는 <표 4-9>, <표 4-10>과 같다.

<표 4-9> 계절별 관광지 핵심 키워드

순위	봄		여름	
	관광지	키워드 (언급 횟수)	관광지	키워드 (언급 횟수)
1	오설록 티 뮤지엄	녹차(56)	오설록 티 뮤지엄	녹차(71)
2	쇠소깍	카약(37)	쇠소깍	카약(35)
3	에코랜드	기차(32)	오설록 티 뮤지엄	아이스크림(23)

4	성산 일출봉	일출(24)	우도	땅콩(23)
5	오설록 티 뮤지엄	아이스크림(16)	에코랜드	기차(18)
6	천지연 폭포	밤(16)	우도	아이스크림(17)
7	아쿠아플라넷제주	아이(15)	우도	자전거(17)
8	성산 일출봉	계단(13)	마라도	짜장면(15)
9	성산 일출봉	중국인(13)	오설록 티 뮤지엄	이니스프리(15)
10	아쿠아플라넷제주	돌고래(13)	성산 일출봉	일출(14)

순위	가을		겨울	
	관광지	키워드 (언급 횟수)	관광지	키워드 (언급 횟수)
1	오설록 티 뮤지엄	녹차(63)	오설록 티 뮤지엄	녹차(50)
2	에코랜드	기차(46)	성산 일출봉	일출(22)
3	쇠소깍	카약(24)	아쿠아플라넷제주	아이(16)
4	산굼부리	역새(22)	카멜리아힐	동백꽃(14)
5	오설록 티 뮤지엄	아이스크림(16)	넥슨컴퓨터박물관	게임(13)
6	우도	자전거(15)	비자림	산책(12)
7	에코랜드	아이(13)	쇠소깍	카약(12)
8	아쿠아플라넷제주	공연(12)	오설록 티 뮤지엄	아이스크림(11)
9	우도	아이스크림(12)	에코랜드	기차(9)
10	아쿠아플라넷제주	아이(11)	용두암	공항(9)

<표 4-10> 연도별 관광지 핵심 키워드

순위	2015년		2016년	
	관광지	키워드 (언급 횟수)	관광지	키워드 (언급 횟수)
1	오설록 티 뮤지엄	녹차(132)	오설록 티 뮤지엄	녹차(108)
2	쇠소깍	카약(73)	에코랜드	기차(40)
3	에코랜드	기차(55)	성산일출봉	일출(36)
4	오설록 티 뮤지엄	아이스크림(37)	쇠소깍	카약(35)
5	성산일출봉	일출(32)	오설록 티 뮤지엄	아이스크림(29)
6	우도	땅콩(31)	월정리해변	카페(26)

7	아쿠아플라넷제주	아이(29)	비자림	산책(24)
8	마라도	짜장면(27)	우도	자전거(24)
9	우도	아이스크림(26)	넥슨컴퓨터박물관	게임(23)
10	에코랜드	아이(25)	아쿠아플라넷제주	아이(23)

모든 차원에서 오설록 티 뮤지엄의 녹차가 가장 많이 언급되었으며, 이어서 봄에는 쇠소깍의 카약(37), 에코랜드의 기차(37), 여름에는 쇠소깍의 카약(35), 오설록 티 뮤지엄의 아이스크림(23), 여름에는 에코랜드의 기차(46), 쇠소깍의 카약(24), 겨울에는 성산 일출봉의 일출(22), 아쿠아플라넷제주의 아이(16), 2015년에는 쇠소깍의 카약(73), 에코랜드의 기차(55), 2016년에는 에코랜드의 기차(40), 성산일출봉의 일출(36) 순으로 분석되었다. 전반적으로 오설록 티 뮤지엄의 녹차와 같이 해당 관광지를 생각하면 떠오르는 키워드가 많이 언급되었으며, 앞선 분석 결과와 같이 중국인 관광객으로 인한 관람의 불편함을 받았던 성산일출봉에서 봄철에 중국인이 언급된 것을 확인할 수 있다. 천지연 폭포의 밤인 경우 봄에 6 번째로 많이 언급되었는데 봄철에만 22 시까지 개장하는 야간 개장이 관광객에게 좋은 반응을 얻은 것으로 분석되었다. 천지연 폭포의 야간개장과 같이 관광지의 새로운 모습을 보여주는 전략을 세운다면 관광객의 긍정적인 반응을 이끌 수 있을 것이다.

##### 5) 키워드 연관성 분석

키워드 사이의 연관성을 알아보기로 최소 지지도 0.005, 최소 신뢰도 0.6 으로 설정 하여 계절과 연도 차원으로 연관성 분석을 실시하였으며 향상도 기준으로 정렬한 결과는 <표 4-11>, <표 4-12>와 같다.

<표 4-11> 계절별 핵심 키워드 연관성

봄					
순위	조건부	결과부	support	confidence	lift
1	쿠폰	할인	0.0067	1.0000	74.9286
2	쇠소깍	카약	0.0086	0.6923	33.0105
3	아이스크림, 오설록	녹차	0.0057	1.000	32.7813
4	에코랜드	기차	0.0076	0.8889	32.1533



5	대기	카약	0.0057	0.6667	31.7879
---	----	----	--------	--------	---------

여름					
순위	조건부	결과부	support	confidence	lift
1	짜장면	마라도	0.0066	0.8889	67.1111
2	아이스크림, 우도	땅콩	0.0058	0.8750	50.3333
3	쇠소각	카약	0.0091	0.6875	29.6607
4	오설록 티 뮤지엄	녹차	0.0057	0.7000	20.1333
5	에메랄드	해변	0.0075	0.6000	7.1762

가을					
순위	조건부	결과부	support	confidence	lift
1	아이스크림, 우도	땅콩	0.0066	1.0000	95.7273
2	산굼부리	억새	0.0066	1.0000	47.8636
3	쇠소각	카약	0.0066	0.70000	40.9500
4	등반	한라산	0.0066	0.70000	28.3500
5	산굼부리	가을	0.0057	0.7500	25.4758

겨울					
순위	조건부	결과부	support	confidence	lift
1	아이스크림, 우도	땅콩	0.0069	1.0000	103.8571
2	휴게소	성판악	0.0055	0.8000	52.8727
3	카트	야외	0.0055	1.0000	51.9286
4	오메기떡	시장	0.0068	0.8333	28.8492
5	설경	한라산	0.0055	0.6667	13.8476

<표 4-12> 연도별 핵심 키워드 연관성

2015년					
순위	조건부	결과부	support	confidence	lift
1	아이스크림, 우도	땅콩	0.0057	1.0000	75.9667
2	짜장면	마라도	0.0057	7.2222	58.7837

3	쇠소깍	카약	0.0079	0.6429	29.3014
4	등반	한라산	0.0061	0.6667	25.7514
5	오설록 티 뮤지엄	녹차	0.0070	0.6667	19.7316

2016 년					
순위	조건부	결과부	support	confidence	lift
1	쿠폰	할인	0.0051	0.9000	75.3429
2	오설록 티 뮤지엄	녹차	0.01194	0.6774	19.8484
3	바다, 카페	해변	0.0051	0.7500	8.7900

전반적으로 앞선 관광지별 키워드 분석결과와 같이 관광지와 해당 관광지에서 가장 많이 언급된 키워드가 연관이 높은 것으로 나왔다. 이외에 봄철과 2016 년도에 '쿠폰'이 나왔을 때 '할인'이라는 키워드가 연관되었는데 관광지 입장권 구입 할 때 쿠폰으로 할인을 받는 것으로 추론되며, 봄철에 대기가 언급되었을 때 카약이 연관성이 있는 것으로 나왔는데 쇠소깍이 유명세를 얻으면서 카약을 이용하려는 관광객이 많아진 것으로 판단된다. 가을에는 산굼부리와 관련된 연관분석 결과가 상위 5 개 중 2 건이 나왔는데 역사를 구경하기 위해 산굼부리를 찾으며, 가을철을 대표하는 관광지로 자리 잡은 것으로 분석된다.

## 제 5 장 결론

### 제 1 절 연구 요약 및 시사점

다양한 데이터가 실시간으로 급증하며 빅 데이터 시대가 열렸고 기존의 소규모 데이터에서는 창출하지 못했던 가치를 빅 데이터에서 창출이 가능해 짐에 따라 전 세계적으로 빅 데이터에 주목하고 있다. 이에 관광 관련 연구에서도 빅 데이터를 활용한 연구들이 진행되고 있으나 아직 초기 단계에 머물고 있으며, 특히, 제주 관광지 연구는 관광객의 인식을 파악하기 위한 연구 방법으로 설문 조사와 같이 구조화되어 있고 데이터 수가 부족하여 파악이 어렵다는 한계에도 불구하고 후기데이터를 이용한 연구가 미진하다. 또한, 빅 데이터를 분석하기 위한 다양한 데이터 마이닝 기술들이 연구되고 있지만, 효율성 중심의 기술 향상에 치우쳐 있으며, 기존 기술과 융합하여 더욱 효과적이고 가치 있는 정보를 이끌어내기 위한 연구도 미진한 실정이다.

이러한 배경 하에 본 연구에서는 후기 데이터뿐만 아니라 다른 유형의 데이터를 분석할 때도 적용 가능한 다차원 순위 모델을 제시하였다. 특히, 다차원 순위 모델을 활용하여 세계적인 관광 웹사이트인 Tripadvisor 에 게시된 제주 관광지 후기에 나타난 관광객의 인식을 파악하고자 하였다. 연구의 목적을 달성하기 위해 이론적 배경에서는 빅 데이터 분석 기술 중 하나인 데이터 마이닝에 대한 이해와 선행연구를 고찰하였으며 제주 관광지 선행연구를 통해 현재 제주 관광지에 관한 연구 현황에 대해 알아보았다.

본 연구에서는 Tripadvisor 제주 관광지 후기 중 관광객이 작성한 평점, 후기, 방문날짜를 수집대상으로 삼아 JAVA API 인 selenium 으로 만든 웹 크롤러를 이용하여 195 개 관광지의 4,807 개의 후기 데이터를 수집하였으며, 계절별, 연도별 차원으로 분석하기 위해 충분한 데이터가 있는 2015 년과 2016 년도 데이터를 분석대상으로 삼았다. 다차원 순위 모델에 적용하기 위해 JAVA API 인 twitter-korean-text 를 이용하여 전처리 과정을 거쳐 43,640 개의 명사를 추출하였다.

정제된 데이터를 분석 도구인 R 을 이용하여 계절과 연도 차원으로 빈도분석, 평점 분석을 하였으며, 키워드 사이의 연관 분석을 하였다.

본 연구 결과를 통해 나온 주요 내용은 다음과 같다.

첫째, 관광객의 관광지 관심도는 아쿠아플라넷제주, 동문재래시장과 같은 인공형 관광지 순위에 있었지만, 성산일출봉, 우도, 쇠소깍 등과 같이 자연형 관광지에 집중되어 있었으며 계절별로 관광지의 관심도가 달라지는 것으로 분석되었다. 여름철 같은 경우 만장굴과 같이 더위를 피할 수 있는 관광지를 선호하고 있었으며, 겨울철에는 오설록과 동문재래시장과 같이 추위에 상관없이 즐길 수 있는 관광지의 관심도가 올라갔다. 이에 따라 자연형 관광지에는 오설록과 같이 자연 + 실내형 관광지의 수립 필요성이 제기 되고 인공형 관광지는 관광객의 관심을 이끌 마케팅의 필요성이 제기된다.

둘째, 관광객의 관광지 호감도는 관심도와 큰 상관관계가 없는 것으로 나타났다. 대표적인 예로 모든 차원에서 가장 많은 관심을 받았던 성산일출봉인 경우 관심에 비해 호감도가 낮게 나왔는데, 이는 입소문은 많이 퍼져 관광객의 관심이 높으나 관광객의 만족을 이끌어내는 것에는 실패했다는 것을 의미한다. 송성진(2009)의 연구결과와 같이 관광지의 질적 성장이 필요해 보인다.

셋째, 제주 관광에서의 핵심 키워드로는 아이, 바다, 해변, 폭포 등이 가장 많이 언급되었으며 앞선 결과와 같이 자연형 관광지에 대한 관심이 높다는 것을 알 수 있다. 특이점으로는 아이가 언급되는 빈도가 높다는 것인데, 이는 아이를 동반한 가족형 관광객의 비중이 높다는 것을 의미한다. 실제로 2016 제주특별자치도 방문관광객 실태조사에 따르면 동반자 유형 중 가족/친척이 2015 년도에는 22.4%로 2 번째로 높았으며 2016 년도에는 31.9%로 가장 높게 나왔는데 본 논문의 분석 결과와 제주관광공사의 조사 결과와 일치함을 의미한다. 이를 통해 가족 관광객들 대상으로 서비스 및 마케팅이 필요할 것으로 사료된다.

넷째, 관광지별 핵심키워드로 오설록은 녹차, 쇠소깍은 카약, 에코랜드는 기차 등으로 도출되었으며, 이를 통해 관광객이 해당 관광지를 어떻게 인식하고 있는지를 파악할 수 있었다. 해당 분석 결과를 타 관광지와의 차별성을 위한 포지셔닝 전략으로 활용했을 때 효과적일 것이라 사료된다.

마지막으로 키워드 간의 연관분석결과, 쇠소깍이 언급되었을 때 카약이 언급될 확률이 높아지는 것과 같이 앞선 관광지별 핵심키워드와 비슷한 결과가 나왔

다. 그 외에 봄철과 2016 년도에 할인을 목적으로 쿠폰을 사용하는 관광객의 비중이 높았고, 최소각의 카약을 이용하기 위해선 대기시간이 긴 것으로 분석되었다.

본 연구 결과의 시사점은 다음과 같다.

첫째, 효율성 중심의 기술 향상이 주 목적이었던 기존의 데이터 마이닝 연구에서 기존 기술들을 융합함으로써 보다 효과적이고 가치 있는 정보를 도출하기 위한 다차원 순위 모델을 제시함으로써 후기 데이터뿐만 아니라 다른 유형의 데이터를 분석할 수 있는 일반화 모델이라는 점에서 다양한 방법으로 활용될 수 있을 것이다.

둘째, 기존의 설문조사 위주의 구조화된 연구 방법에서 알기 힘들었던 관광객의 인식을 관광 후기 빅 데이터를 이용한 양적 연구방법으로 관광객이 실제로 인식하고 있는 제주 관광지에 대해 다각도로 파악함으로써 앞으로의 제주 관광 정책 수립에 유용하게 이용될 것이다.

셋째, 아직 초기 단계인 빅 데이터를 활용한 관광학 내에서 후기 빅 데이터 분석을 시도했다는 점에서 가치를 찾을 수 있다. 이는 향후 진행될 관광학 연구에서 토대가 되어줄 것으로 본다.

## 제 2 절 연구의 한계

본 연구는 제주 관광지를 방문한 관광객의 인식을 파악하기 위해 Tripadvisor 의 후기 데이터를 수집하였다. 그러나 온라인상에서 관광 후기를 남기는 주요 사용자 층이 20 대에서 30 대에 집중되어 있으므로 본 연구에서 인구 통계적 특성을 반영하지 못한 점에서 한계가 있다고 볼 수 있다. 향후 연구에서는 이러한 특성을 반영하여 후기 데이터를 분석한다면 좀 더 정확한 연구결과를 도출할 수 있을 것이다. 또한, 내국인을 대상으로 한글 후기만을 분석 대상으로 설정하였기 때문에 외국인 관광객이 느끼는 인식은 이와 다를 것으로 생각된다. 이에 따라 외국인 관광객 대상까지 확대했을 때 내국인 관광객의 인식과 외국인 관광객의 인식이 어떻게 다른지에 대한 연구도 추가로 이뤄져야 할 것으로 보인다. 마지막으로 본 연구에서는 후기데이터에서

명사만을 추출하여 키워드 분석을 하였기 때문에 후기에 나타난 관광객의 감성을 파악하지 못했다는 점이다. 형태소 분석기를 통해 형용사를 추출하고 평점을 이용하여 감성 사전을 구축하여 후기데이터에 나타난 감성을 분석한다고 하면 더욱 다채로운 연구결과를 얻을 수 있을 것이며, 이와 같은 연구도 추가로 필요할 것으로 보인다.

## 참고문헌

- 김근형, 오성열, "온라인 고객리뷰 분석을 통한 시장세분화에 텍스트마이닝 기술을 적용하기 위한 방법론", *한국콘텐츠학회논문지*, 제 9 권, 제 8 호, 2009, pp.272-284.
- 김근형, 황병용, 김민철, "데이베이스 : 중요지지도를 고려한 연관규칙 탐사 알고리즘", *정보처리학회논문지D*, 제 11 권, 제 3 호, 2004, pp.545-552.
- 김산성, 김명원, "데이터 마이닝을 위한 연관규칙의 다중 값 속성 처리방법", *한국정보과학회 학술발표논문집*, 제 29 권, 제 2 호(Ⅱ), 2002, pp.340-342.
- 김은실, "일반화된 연관규칙 발견을 위한 레벨 기반 데이터마이닝 시스템", 이화여자대학교 석사학위논문, 2001.
- 김유영, 송민, "영화 리뷰 감성분석을 위한 텍스트 마이닝 기반 감성 분류기 구축". *지능정보연구*, 제 22 권, 제 3 호, 2016, pp.71-89.
- 김정준, "관광지 유형에 따른 관광동기, 관여도와 관광매력성 연구", 경기대학교 박사학위논문, 2010.
- 김지민, "효율적인 olap 시스템 구현을 위한 rolap 과 molap 의 비교분석", 연세대학교 석사학위논문, 2003.
- 류시영, 유선옥, "소셜미디어에 나타난 강원도 관광에 대한 인식 연구", *관광연구저널*, 제 31 권, 제 2 호, 2017, pp.63-81.
- 문화체육관광부, "2015 국민여행실태조사", 2016.
- 송성진, "제주지역 관광지의 발전 방안", *제주관광학회 학술대회*, 제 2009 권, 제 1 호, 2009, pp.25-49.
- 송수미, "데이터 마이닝을 이용한 증권사 위탁고객 이탈모형 구축", 이화여자대학교 석사학위논문, 2001.
- 송호길, "크롤링을 이용한 웹 애플리케이션의 접근 제어 취약점 자동 탐지", 한양대학교 석사학위논문, 2012.
- 심영석, "텍스트 마이닝(Text Mining)을 이용한 관광지 이미지 결정요인에 관한 연구", 세종대학교 석사학위논문, 2016.
- 오성수, "SNS 관광정보가 관광지선택속성과 관광만족에 미치는 영향에 관한 연구", 전남대학교 박사학위논문, 2013.

- 오익근, 이태숙, 전체남, 운영일, "빅 데이터 분석을 통한 한국관광 인식에 관한 연구", *한국관광학회 국제학술발표대회집*, 제 78 권, 제 3 호, 2015, pp.595-602.
- 위키백과, "olap", [https://ko.wikipedia.org/wiki/온라인\\_분석\\_처리](https://ko.wikipedia.org/wiki/온라인_분석_처리), 2017.
- 유한주, "트랜잭션기반 분석용 olap 큐브구조 설계", *숭실대학교 박사학위논문*, 2008.
- 윤지환, 이영진, "관광분야에서 SNS 빅 데이터의 활용 방법 모색", *관광연구저널*, 제 28 권, 제 3 호, 2014, pp.5-14.
- 이정민, "무용학의 지적 구조 분석 연구", *성균관대학교 박사학위논문*, 2017.
- 이해각, "연관성규칙 발견을 위한 효율적인 데이터 마이닝 알고리즘 설계", *한국멀티미디어학회 학술발표논문집*, 제 2004 권, 제 1 호, 2004, pp.587-591.
- 장재영, "온라인 쇼핑몰의 상품평 자동분류를 위한 감성분석 알고리즘", *한국전자거래학회지*, 제 14 권, 제 4 호, 2009, pp.19-33.
- 정규렬, "관광지 방문동기와 장소 정체성, 만족에 관한 연구", *제주대학교 석사학위논문*, 2013.
- 조완섭, 조아, 권가은, 류관희, "SNS 데이터 분석을 통한 스마트 충북관광 구축", *한국데이터정보과학회지*, 제 26 권, 제 2 호, 2015, pp.409-418.
- 최지애, "데이터 마이닝을 이용한 구매예측 모형", *이화여자대학교 석사학위논문*, 2002.
- 한국데이터베이스진흥원, "데이터 분석 전문가 가이드", *한국데이터베이스진흥원*, 2014.
- 한영임, "다차원분석을 지원하는 지능형 olap 모형의 개발", *서강대학교 석사학위논문*, 2005.
- 허성완, "한국어 웹 문서 범주화를 위한 텍스트 마이닝 기법 연구", *아주대학교 석사학위논문*, 2016.
- 현홍준, "제주지역 테마관광지 유형별 관광객의 욕구 및 만족에 관한 연구", *제주대학교 석사학위논문*, 2007.
- Codd, E. F., Codd, S. B., Salley, C. T., "Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate". Codd and Associates, Technical Report, 1993.



David Reinsel, John Gantz, Rydning, J, "Data Age 2025", 2017.

KoNLPy, "KoNLPy: 파이썬 한국어 NLP", "<http://konlpy-ko.readthedocs.io/ko/v0.4.3/>", 2017.

Tripadvisor, <https://www.tripadvisor.co.kr/>, 2017.

twitter, "twitter/twitter-korean-text", <https://github.com/twitter/twitter-korean-text>, 2017