

산경논집 제 37 집 (2016年 8月)
The Journal of Industry and Economy, Vol.
37, August 2016.

논문 접수일 : 2016. 07. 11.

논문 심사일 : 2016. 08. 12.

게재 확정일 : 2016. 08. 23.

온라인 관광후기 데이터 분석*
- 트립어드바이저의 우도 데이터를 중심으로 -

Analysis of online tourist reviews about Udo in Tripadvisor web site

김 근 형** (Keun-Hyung Kim)

< 목 차 >

I. 서론	참고문헌
II. 이론적 배경 및 선행연구	Abstract
III. 데이터 분석	
IV. 결론 및 시사점	

주제어: 온라인 관광후기, 우도, 트립어드바이저, 귀납적 방법론

* 이 연구과제는 JCC그룹 발전기금으로 수행되었음

** 제주대학교 경상대학 교수(제주대학교 관광과경영경제연구소 전임연구원), khkim@jejunu.ac.kr

I. 서론

제주도 방문 관광객은 1983년에 100만 명을 기록한 후 2005년 500만 명을 돌파하였고, 2013년에는 1천만 명을 돌파하여 2005년 이후 8년 만에 약 2배가 증가한 것으로 나타났다. 대표적인 섬 관광지인 우도 역시 관광객이 크게 증가하여 2015년에 206만 명이 다녀간 것으로 나타났다(호남지방통계청 제주사무소·보도자료, 2016). 유력한 미디어에 따르면, 2016년도에는 우도 관광객의 100만명 돌파 시점이 2015년보다 25일 빨라졌다고 밝혔다(연합뉴스, 2016년 6월 28일).

우도 방문객이 해마다 늘어나는 것은 청정한 자연환경뿐만 아니라 독특한 관광인프라 등 우도만의 매력이 널리 알려졌기 때문으로 관계자들은 풀이하고 있다. 현재까지는 제주를 방문한 관광객이 부차적으로 우도 관광을 하거나 제주도민이 주로 방문하는 상황이지만, 역으로 우도방문을 주목적으로 하는 관광객을 늘릴 수 있다면 제주 방문 관광객을 더욱 증가시킬 수 있을 것이다. 우도만의 특별한 매력을 극대화시킨다면 우도 방문을 주목적으로 하는 외래 관광객이 늘어날 것이고 이는 결국 제주 방문 관광객의 증가를 유발하는 효과로 귀결될 것이기 때문이다. 즉, 제주의 많은 관광명소들 각각에 대한 특징, 가치, 매력 등을 분석하고 이들 각각을 개별적으로 극대화시키는 것은 창의적인 제주관광 활성화 전략이 될 수 있다.

제주관광의 활성화와 발전 방향에 대한 많은 연구들이 있지만, 우도 등과 같은 제주도 내의 특정 명소를 대상으로 한 관광객들의 행동연구는 많지 않은 실정이다. 제주도 내 관광명소들의 가치와 매력 등을 탐색적으로 연구하여 이들에 대한 구체적인 이론 연구의 주춧돌이 될 수 있도록 하는 것이 필요한 상황이다.

관광객들의 관광활동이나 관광 패턴 등을 발견하기 위한 기존의 연구방법들은 관광객들을 대상으로 한 설문지 분석을 주로 이용하지만, 이러한 방법은 탐색적 연구에서는 자연스럽지 못한 측면이 있다. 왜냐하면, 설문지

는 기존 연구들을 기반으로 작성되어야 그 타당성이 입증될 것이기 때문이다. 따라서 선행연구가 별로 없거나 전무한 주제를 대상으로 한 탐색적 연구는 이미 존재하는 데이터 등을 분석하여 그 의미를 파악하는 방식으로 진행되는 것이 효율적이고 효과적인 접근방법일 수 있다. 관광객들의 여행 경험이나 의견들이 자유롭게 기록된 관광후기 데이터는 기획된 설문지와 별개로 생산된 것이며 관광객들의 솔직한 생각들이 내재되어 있다.

본 논문에서는 트립어드바이저(www.tripadvisor.co.kr)라는 세계적인 여행 사이트의 관광후기 데이터를 이용하여 제주 관광명소 중 하나인 우도의 가치와 매력을 도출하는 탐색적 연구를 수행하고자 한다. 트립어드바이저 사이트에 게시된 텍스트(text) 형식의 우도 관련 관광후기 데이터를 빅데이터 분석기법에 의하여 분석하고 그 결과에 의한 시사점을 제시하고자 한다.

II. 이론적 배경 및 선행연구

2.1. 사회과학 연구방법

2.1.1 양적 방법과 질적 방법

양적연구방법¹⁾은 경험적 자료를 수집하고 계량화하여 사회·문화 현상을 통계적으로 분석하는 연구방법이다. 자연현상과 사회·문화현상은 본질적으로 다르지 않을 것이라는 전제를 바탕으로 하며 자연 과학적 연구방법을 사회·문화현상에 적용할 수 있다는 방법론적 일원론이라 할 수 있다. 자연 과학에서처럼 일반적인 법칙을 발견하는 것을 목적으로 하나, 그 대상이 사회·문화에 국한되는 차이가 있다. 사회·문화 현상의 법칙을 발견하기 위하여 추상적인 개념을 객관적인 관찰이 가능하도록 조작적으로 정의하는

1) ZUM 학습 백과 - "사회·문화 현상의 연구 방법" 을 토대로 정리

과정이 필요하다. 개념들 사이의 관계를 표현한 가설을 검증함으로써 법칙을 발견하거나 수립할 수 있다. 양적 연구방법은 계량화가 어려운 분야에서의 연구방법으로는 부적절하며 인간의 의식과 행위에 대한 깊이 있는 접근이 어렵다는 한계가 있다. 양적 연구방법의 일반적인 절차는 문제인식이 전제된 상태에서 이론적 논리를 바탕으로 가설설정을 하며 가설검증을 위한 연구설계, 자료수집, 자료분석의 순서로 진행된다.

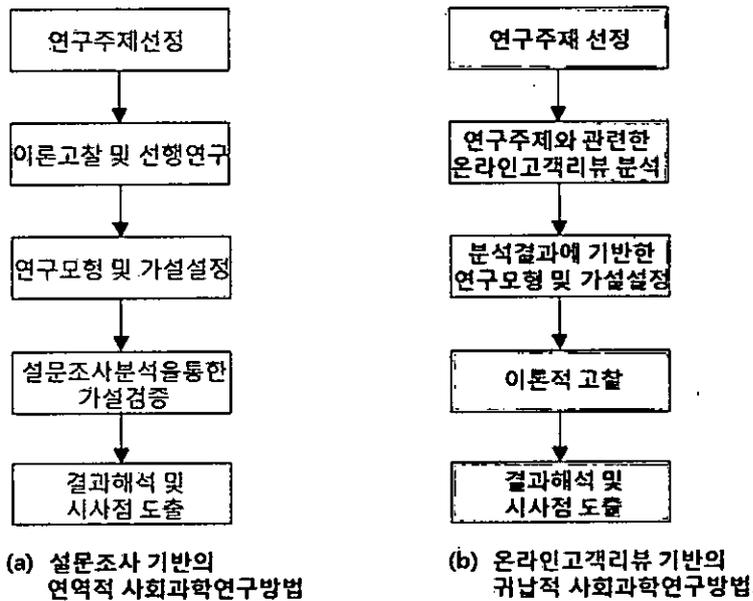
질적연구방법²⁾은 연구자의 직관적 통찰로 사회·문화 현상의 의미를 해석하고 이해하려는 연구방법이다. 자연현상과 사회·문화 현상은 본질적으로 다르므로 자연 과학적 연구 방법을 가치 함축적인 사회·문화 현상에 적용할 수 없다는 방법론적 이원론에 기반하는 연구방법으로서, 사회·문화 현상에 대한 심층적 이해를 목적으로 한다. 대화록이나 관찰일지, 비공식적 문서 등의 자료를 활용한다. 질적 연구방법은 주관적 의식에 대한 심층적 이해와 인간행위의 동기와 의미를 중시한다. 질적 연구방법의 일반적인 절차는 문제인식을 통하여 연구설계를 하며 질적 연구에 적합한 자료를 수집하고 분석 및 해석 등의 순서로 진행된다.

2.1.2 연역적 방법과 귀납적 방법

김근형(2015)의 연구에서는 사회과학연구방법론을 2가지로 구분하고자 주장하고 있다. 전통적 방식의 사회과학 연구방법은 일반적으로 기존 이론모형을 기반으로 새로운 연구모형과 가설을 도출하고 설문조사를 통하여 데이터를 수집 및 분석하는 방식으로 가설검증을 하는 연역적(deductive) 방법이다. 선행연구와 가설설정 과정에서 가설의 논리적 정당성을 제시해야 한다는 측면에서 연역적인 이론 정립 방식인 것이다. 아래 <그림1>의 (a)는 연역적 연구방법의 개념을 나타내고 있다. 이에 반하여, 귀납적 방법은 이미 확보된 데이터를 분석하여 데이터의 주요 패턴을 발견하고 이 패턴을 근거로 가설과 시사점을 도출하는 방식이다. 아래 <그림 1>의 (b)는 귀납적 연구방

2) ZUM 학습 백과 - "사회·문화 현상의 연구 방법" 을 토대로 정리

법의 개념을 나타내고 있다. 정보기술이 발전하지 않고 고객들의 후기 데이터 즉, 온라인고객리뷰(online customer review)가 많지 않았던 시절에는 연역적 방법이 주로 사용되었으나 트립어드바이저(www.tripadvisor.com)에서처럼 온라인 고객리뷰가 넘쳐나는 빅데이터 시대에는 온라인고객리뷰기반의 귀납적 연구방법이 필요하며 그 중요성이 커졌다.



<그림 1> 사회과학연구방법의 비교

2.2. 선행연구

김민철·장희정(2008)의 연구에서는 보길도와 선유도, 우도 등의 섬 지역을 중심으로 관광매력요인을 파악하고 그 유형에 따른 중요도의 차이가 있음을 주장하고 있다. 섬지역의 관광콘텐츠 유형(정태형, 동태형, 혼합형)에 따라 여행 전 섬 관광 매력요인의 중요도는 차이가 있음을 밝히고 있다. 또한, 매력요인의 중요도와 방문 후의 성취도 사이에는 유의한 차이가 있

으며 성취도는 관광만족도에 유의한 영향을 미치는 것으로 나타났다. 그러나 이 연구에서는 다른 선행연구의 상대적인 차별성을 바탕으로 관광유형을 분류한 후 이에 따른 중요도를 고려하고 있기 때문에 우도에 대한 구체적인 매력요인을 파악할 수 없다는 한계가 있다. 양영순·박재민(2013)의 연구에서는 온라인 여행후기의 정보특성을 파악하고, 이러한 정보특성이 수용의도 및 구전효과에 어떠한 영향을 미치는지 분석하고 있다. 온라인 여행후기 정보특성 중 발신자의 진실성은 정보의 유용성을 매개로 수용의도에 유의한 영향을 미치는 것으로 나타났다. 정보특성 중 정보 공감성은 수용의도에 직접적인 영향을 미치는 것으로 나타났고 발신자 경험성은 수용의도 및 구전효과 모두에 유의한 영향을 미치는 것으로 나타났다. 이 연구는 온라인 여행후기의 정보특성을 일반화하고 일반화된 정보특성과 여행후기의 수용의도 및 구전효과와의 관계를 파악하고 있으며 설문지 분석에 의한 가설검증을 수행하고 있다. 이정현(2013)의 연구에서는 베니키아호텔을 대상으로 호텔이용후기에 대한 질적내용분석 결과를 제시하고 있다. 전국 체인망 형태로 흩어져 있는 베니키아호텔의 웹사이트에 등록된 호텔이용후기를 수집하여 고객가치 차원의 내용분석을 실시하였다. 고객가치 유형을 기능적 가치, 정서적 가치, 가격대비 가치, 인식적 가치, 조건적 가치 등으로 구분하고 있으며 각 여행후기들이 어떤 가치에 중점을 두고 있는지 분석하였다. 이 연구에서는 고객가치 관점에서 여행후기 분석을 위한 분석틀을 미리 정하고 있기 때문에 분석결과의 다양성 도출에 대한 한계가 존재한다. 강동언(2008)의 연구에서는 우도의 전통돌담에 대한 문화적 이미지와 경제적 가치를 평가하고 관광자원화하기 위한 방안을 제시하고 있다. 조완섭(2015)의 연구에서는 충북관광에 관한 온라인 상의 글들을 수집하고 텍스트 마이닝 기법을 활용하여 분석함으로써 충북관광을 위한 홍보전략과 관광진흥 정책을 수립하는 방안을 제시하고 있다. 분석에 이용된 데이터는 주요 블로그, 뉴스, 관광관련 공공 데이터 등이며, 충북전역의 관광지를 대상으로 하고 있다. 특히 다차원 분석 기법을 연계함으로써 관광관련 텍스트 정보를 다각도로 분석하여 전략을 수립하는데 활용할 수 있도록 하였다.

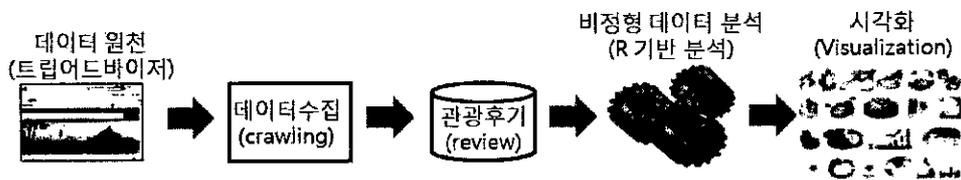
이상에서 살펴본 바와 같이, 제주도 우도와 관련한 연구들은 많지 않으며 몇몇 연구들이 있다 할지라도 연역적 연구방법으로 접근하거나 질적 연구방법을 사용하고 있다. 조완섭(2015)의 연구는 귀납적 연구방법을 사용하고 있으나 충북 관련 데이터를 분석하고 있다.

본 연구에서는 우도 관광에 대한 관광객들의 행동양식을 탐색적으로 밝히기 위하여 트립어드바이저의 관광후기 데이터를 바탕으로 귀납적 방식의 양적 접근을 한다는 측면에서 기존 연구들과 차별점이 있다.

Ⅲ. 데이터 분석

3.1. 데이터 수집 및 분석 절차

제주도 우도에 대한 탐색적 연구를 수행하기 위하여 <그림 2>와 같이 연구설계를 하였다. 트립어드바이저 웹사이트로부터 우도 관련 관광후기를 웹크롤러를 이용하여 수집하였으며, R기반의 비정형데이터 분석 기법을 통하여 분석 한 후 그 결과를 시각화하였다.

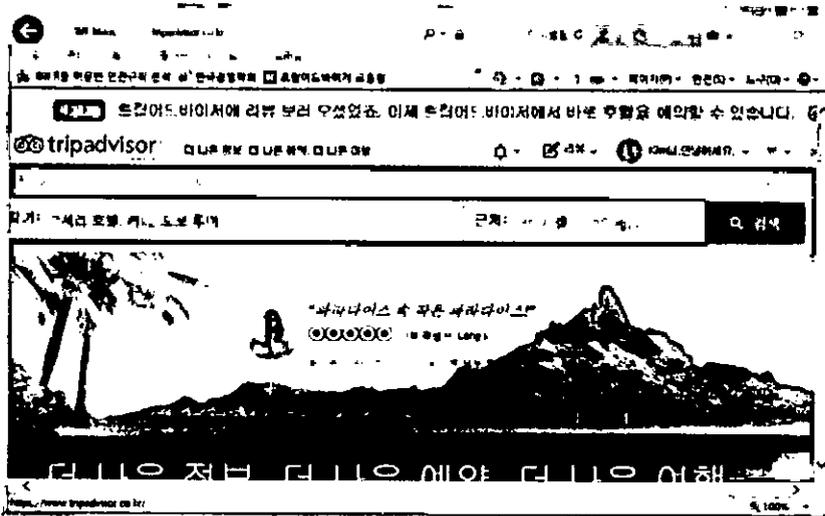


<그림 2> 데이터 수집 및 분석 절차

3.1.1 데이터 원천

본 연구에서는 우도 관련 관광후기를 수집하기 위하여 관광시설 및 명소

에 대한 여행후기(review)를 제공하는 트립어드바이저(www.tripadvisor.com, 한국여행 관련은 www.tripadvisor.com)를 활용하였다.



<그림 3> 트립어드바이저 사이트 메인 화면

트립어드바이저에는 6천여만명의 회원이 가입되어 있으며 1억7천여 건의 여행관련 후기를 보유하고 있다. 이 중에서 우도 관련 후기는 450여건이 존재했으며 한국어 후기는 230여건이 있었다. 영문버전의 트립어드바이저는 2000년에 오픈되었지만, 한국어 버전은 2012년에 오픈되었기 때문에 한국어 기반 관광후기는 아직까지는 많지 않다.

3.1.2 데이터 수집

트립어드바이저 웹사이트로부터 우도 관련 여행후기를 수집하기 위하여 파이썬 기반의 웹크롤러를 제작하여(서진수, 2016) 데이터를 수집하였다. <그림 4>는 수집한 우도 관련 여행후기 데이터를 나타내고 있다.



<그림 4> 우도 관련 관광후기

3.1.3 데이터 분석

수집한 우도 관광후기(한국어 기반)는 233건이며 텍스트(text) 기반의 비정형 데이터이다. 따라서, 데이터 분석 소프트웨어인 R을 사용하여 텍스트 데이터의 전체 분포를 한눈에 파악할 수 있는 워드클라우드(wordcloud), 출현단어 빈도수 등을 우선 분석하였다. 텍스트를 보다 심층적으로 분석하기 위하여 각 후기 별 텍스트 문장들로부터 명사들만을 추출하여 정형 데이터로 변환하였으며, 정형데이터의 연관규칙탐사 기법을 이용하여 각 후기 별 명사들 사이의 연관성을 분석하였다.

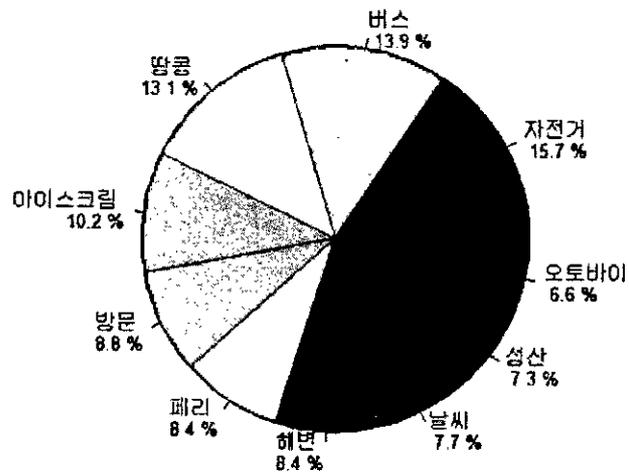
3.2. 분석 결과

<그림 4>는 우도 관광후기의 워드클라우드를 나타내고 있다. 워드클라우드는 텍스트에서 많이 출현한 단어를 더 크게 나타내는 방식으로 전체 텍스트의 윤곽을 나타내는 기법이다. 233건의 우도 관광후기에서는 ‘버스’,

‘자전거’, ‘땅콩’, ‘해변’ 등이 대한 단어들 이 많음을 알 수 있다.

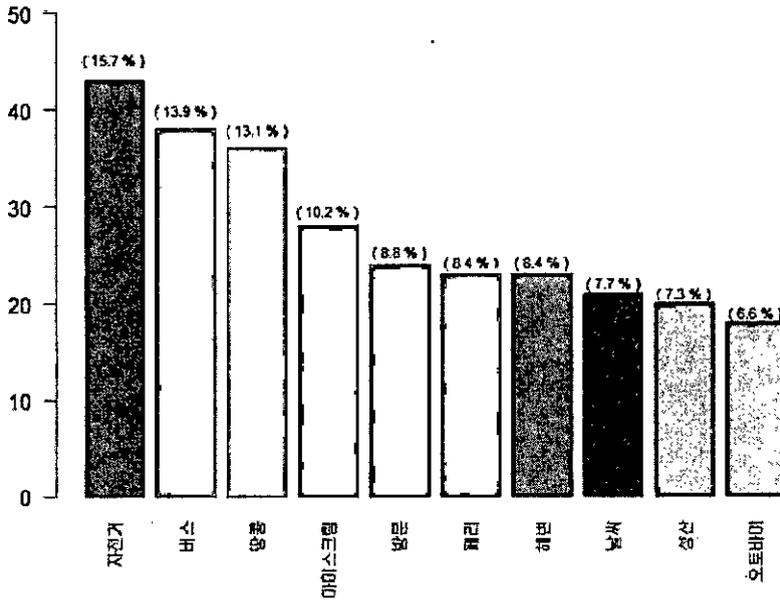


<그림 4> 우도 관광후기의 워드클라우드



<그림 5> 출현 단어들의 출현 빈도 비율

<그림 4>는 우도 관광후기의 워드클라우드를 나타내고 있다. 워드클라우드는 텍스트에서 많이 출현한 단어의 크기를 더 크게 나타내는 방식으로 전체 텍스트의 윤곽을 나타내는 기법이다. 우도 관광후기에서는 ‘버스’, ‘자전거’, ‘땅콩’, ‘해변’ 등이 대한 단어들이 많음을 알 수 있다. <그림 5>와 <그림 6>에서는 관광후기 텍스트에 출현한 단어들의 상대적인 비율과 순위를 나타내고 있다. 가장 많이 언급된 단어는 ‘자전거’, ‘버스’, ‘땅콩’ 등순이었다. 관광객들을 우도 관광에서 자전거와 버스를 많이 이용하고 있으며 우도땅콩도 많이 사먹고 있음을 짐작할 수 있다.



<그림 6> 출현 단어들의 출현 순위

<그림 7>은 233건의 우도 관광후기에서 각 후기별 출현 단어들 사이의 연관성을 나타내는 25개의 규칙을 도출(최소지지도:0.012, 최소신뢰도:0.6)한 결과를 보여주고 있다. 규칙1에서 confidence가 0.75라는 의미는 233건의

각 관광후기에서 '완벽' 이라는 단어가 나오면 '자전거' 라는 단어가 나올 확률이 75% 라는 의미이다. 규칙1에서 support가 0.01293103이라는 의미는 '완벽' 과 '자전거' 라는 단어를 동시에 포함한 관광후기가 전체 233건 중 3건이라는 의미이다. <그림 7>에 있는 25개의 연관규칙은 최소지지도가 너무 낮고 규칙의 왼쪽부분(lhs)와 오른쪽 부분(rhs)에 나타나는 단어쌍이 상식 수준에 머물고 있기 때문에 유의미한 규칙은 없는 것으로 해석할 수 있다.

	lhs	rhs	support	confidence	lift
1	{완벽}	⇒ {자전거}	0.01293103	0.7500000	4.833333
2	{서빈백사}	⇒ {모래}	0.01293103	0.6000000	27.840000
3	{모래}	⇒ {서빈백사}	0.01293103	0.6000000	27.840000
4	{관광객}	⇒ {자전거}	0.01293103	0.7500000	4.833333
5	{진짜}	⇒ {딸콩}	0.01293103	0.7500000	5.800000
6	{해안}	⇒ {도로}	0.01293103	0.6000000	23.200000
7	{순환}	⇒ {버스}	0.01293103	0.7500000	5.612903
8	{특별}	⇒ {바닷가}	0.01293103	0.7500000	34.800000
9	{바닷가}	⇒ {특별}	0.01293103	0.6000000	34.800000
10	{명소}	⇒ {관광}	0.01293103	1.0000000	33.142857
11	{15}	⇒ {생산}	0.01724138	0.6666667	8.592593
12	{일출}	⇒ {생산}	0.02155172	0.8333333	10.740741
13	{티켓}	⇒ {페리}	0.02155172	1.0000000	11.600000
14	{티켓}	⇒ {버스}	0.01293103	0.6000000	4.490323
15	{관광버스}	⇒ {추천}	0.01724138	0.6666667	8.592593
16	{국제}	⇒ {페리}	0.01293103	0.7500000	8.700000
17	{국제}	⇒ {자전거}	0.01293103	0.7500000	4.833333
18	{투어}	⇒ {버스}	0.03448276	0.8000000	5.987097
19	{티켓, 페리}	⇒ {버스}	0.01293103	0.6000000	4.490323
20	{버스, 티켓}	⇒ {페리}	0.01293103	1.0000000	11.600000
21	{음식, 투어}	⇒ {버스}	0.01293103	1.0000000	7.483871
22	{버스, 음식}	⇒ {투어}	0.01293103	1.0000000	23.200000
23	{방문, 사람}	⇒ {페리}	0.01293103	0.7500000	8.700000
24	{사람, 페리}	⇒ {방문}	0.01293103	0.6000000	6.960000
25	{방문, 페리}	⇒ {사람}	0.01293103	0.7500000	13.384615

<그림 7> 출현단어들 사이의 연관성 분석 결과

IV. 결론 및 시사점

온라인상에 남겨진 관광 소비자들의 생생한 여행 경험담인 관광후기는 그 후기를 접하는 관광객들에게 중요한 의사결정 정보가 된다. 관광후기에는 관광객이 지불한 돈과 시간적, 정신적, 육체적 노고의 대가인 상품과 서비스에 대한 소비자의 가치판단이 허심탄회하게 담겨있기 때문이다.

본 연구에서는 트립어드바이저에 게시된 우도 관련 관광후기를 분석하였다. 우도 여행에 대하여 관광객들이 어떻게 생각하는지에 관한 탐색적 연구를 수행하였으며 다음과 같은 연구결과들을 얻을 수 있었다. 즉, 우도 관광객들은 자전거와 버스를 이용하는 경우가 많으며 우도 땅콩 등에 대해서도 관심이 많음을 알 수 있었다. 일반적으로 우도는 자연경관이 좋아서 많은 관광객들이 방문하는 것으로 알고 있었지만 우도 방문객들은 버스와 자전거와 같은 관광인프라와 우도 특산물 중의 하나인 우도땅콩이나 아이스크림 등 먹거리에도 많은 관심이 있음을 알 수 있다. 본 연구의 결과를 통하여 우도의 관광인프라를 확충하는 것과 스토리텔링에 기반한 특화된 먹거리를 개발하는 것이 중요할 것임을 시사한다.

특히, 본 연구는 텍스트 형태의 비정형데이터를 기반으로 양적 접근 방식의 귀납적 연구방법이라는 새로운 연구방법론을 적용하였다는 측면에서 타 연구들과의 차별성이 있다. 그럼에도 불구하고, 본 연구에서 사용한 관광후기는 한글 기반의 텍스트 데이터이기 때문에 한글문장을 파싱(parsing, 구문분석)하고 명사 등을 추려내는 과정에서의 정확성이 떨어지는 측면이 있다. 한글문장은 많은 조사들이 포함되어 있고 띄어쓰기 과정도 복잡하므로 영어문장에 비하여 자연어 처리기법에 의한 구문분석 정확도가 떨어지기 때문이다. R에서의 한글처리 패키지인 KoNLP도 이러한 한계에 직면한 상태이다. 관광후기 문장에서 명사들을 추출할 때 띄어쓰기가 제대로 되어 있지 않으면 명사로 선택되지 않거나 엉뚱한 조사들이 명사로 추출되는 경우들이 있었다는 것은 본 연구의 한계점 중의 하나이다.

본 연구의 또 다른 한계점으로는 데이터의 양적 충실도가 떨어졌다는 점이다. 일반적으로 워드클라우드나 데이터마이닝의 결과들은 처리 데이터가 대용량일수록 그 신뢰성이 높게 되지만, 본 연구에서 사용한 텍스트 데이터는 충분한 신뢰성을 확보하기에는 미흡한 측면이 있다.

참고문헌

- 김근형, 『귀납적 사회과학연구를 위한 확장형 오피니언마이닝시스템 설계』, 산경논집 제 37집, 2015.
- 조완섭, 조아, 권가은, 류관희, 『SNS 데이터 분석을 통한 스마트 충북관광 구축』, 한국데이터정보과학회, 26권,2호, 2015.
- 김민철, 장희정, 『도서의 관광콘텐츠 유형에 따른 관광매력 요인-보길도, 선유도, 우도를 중심으로』, 관광레저연구, 20권,4호, 2008,pp.123-137.
- 양영순, 박재민, 『온라인 여행후기 정보특성과 수용의도 및 구전효과의 영향관계 연구』, 관광레저연구, 25권,7호, 2013,pp.195-214.
- 이정현, 『온라인 호텔이용후기의 질적 내용분석에 의한 고객가치 연구』, 한국콘텐츠학회논문지, 13권,10호, 2013, pp.533-546.
- 강동언, 『전통제주 돌담의 문화관광자원화 방안 연구-우도지역을 중심으로』, 제주발전연구원 연구보고서, 2008.
- 서민구, 『R을 이용한 데이터분석 처리&실무』, 길벗, 2014.
- 서진수, 최경우, 『파이썬 배워 크롤러 DIY하다』, 알음 2016.

Analysis of online tourism reviews about Udo in Tripadvisor

Keun-Hyung Kim

<Abstract>

In order to establish creative tourism strategies of Jeju, it would be necessary to analyze and maximize features, values and attractions of sight spots in Jeju. In this paper, we implemented exploratory study to derive the values and the attractions of Udo by analyzing tourism reviews posted to Tripadvisor which would be world-famous tourism web site. The analysis results are following. First, many people in Udo tourism used bicycles and bus. Second, the tourists were interested in peanuts and ice cream produced in Udo. While we have known generally that many tourists would visit the Udo because of beautiful nature, we discovered that the tourists would be not only interested in convenient tourism infra like a means of transportation, but also would be ready to enjoy special foods. we recognized that it would be important to develop more tourism infra and special food based on story telling in order to activate tourism of Udo.

Key words: Udo tourism, online tourism reviews, Tripadvisor, tourism infra, special food