

SNS 환경에서 동시출현 해시태그 분석 기법

A Co-Occurring HashTag Analysis Technique In SNS EnvironMents

김 세 진, 이 상 돈
목포대학교

Kim se-jin, Lee sang-don
Mokpo Univ.

요약

최근 빅데이터 시대에 다가와서 소셜 네트워크 서비스(Social Network Service)가 중요한 정보 공유의 수단으로 발전함에 따라 그에 따른 예측분석, 동향분석, 이슈탐지 등이 증가하고 있으며, 콘텐츠 분야에서 빅데이터 기법 사례가 증가하는 추세이다. 모바일기기 보급이 빠르게 확산되면서 SNS 활성화와 함께 많은 양의 데이터가 증가하고 있으며, 인스타그램과 같은 해시태그 사용 가능 SNS 서비스에서 해시태그의 동시출현은 해시태그만의 연관성이 있음을 의미한다. 본 논문에서는 대상 SNS의 동시출현 해시태그를 분석하기 위해 발생하는 데이터를 가지고 현재 트렌드에 맞게 분석하여 정보를 제공하는 방법을 제시한다.

I. 서론

오늘날 소셜 네트워크 서비스(Social Network Service)가 활성화 되면서 전 세계적으로 수많은 데이터가 급증하고 있다. 웹상에서의 데이터뿐만 아니라 모바일 기기를 통한 SNS 접근 방법에서 발생하는 데이터 양은 우리 사회에 전반적으로 많은 영향을 미치고 있다.

빠르게 변화하는 시대에 대처해야 하는 문화산업에서 빅데이터는 비정형 데이터까지 분석을 하여 미래의 트렌드, 이미지 분석 등 주로 기업의 마케팅 활동이나 공공부문에서 수집된 정보 가공 등에 집중되어 왔다[1]. 그러나 현재는 다양하고 많은 데이터를 모으는 것보다 데이터를 의미 있는 방향으로 해석하고, 시사점을 도출하는 작업이 중시되고 있다[2]. 최근 콘텐츠 분야에서의 빅데이터 기법 활용 사례가 늘어나는 추세이고, 콘텐츠 분야에서 관광산업은 정보기술(Information Technology) 발전으로 새로운 전환기를 맞이하고 있다. 관광의 패턴이 빠르게 변화하면서 최근에는 목적관광이나 개별관광이 급격히 변하고 있는 것을 알 수 있다[3].

웹 검색과 정보 공유가 가까워지면서 사람들은 빠른 시간 내에 짧고 간결한 정보를 얻고자 한다. 사람들이 관광정보를 얻고자 할 때 웹 사이트나 블로그 등에서 정보를 얻었는데, 최근에는 SNS 상에서 정보를 얻는 추세이다.

본 논문에서는 SNS 대상을 인스타그램으로 하였다. 인스타그램에서 하루 평균 올라오는 이미지는 약 6천만 장이며, лай크(좋아요) 수는 16억개, 지금까지 누적된 이미지 개수는 약 200억장이라고 한다. 업로드된 사진에서 발생하는 해시태그 데이터를 가지고 분석을 하여, 지역 관광, 문화, 예측분석 등 현재 트렌드에 맞게 활용 및 제

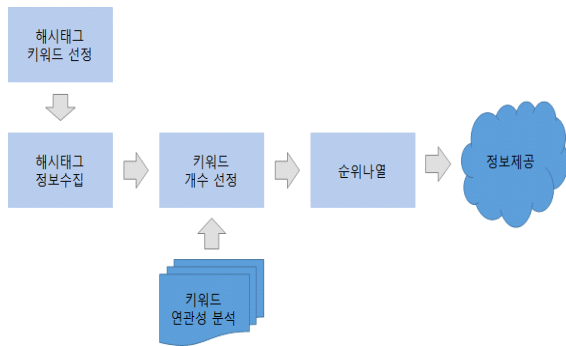
공하는 방법을 제안한다.

II. 본론

모바일 기기의 보급이 빠르게 확산이 되면서 기존 웹 사이트, 블로그 등 제공 방식의 관광 정보 보다는 SNS 서비스를 활용한 제공 방식이 활발하여 사용자들이 많이 접하는 추세이다. 스몰토크(small talk) 기반으로 편리하고, 일대다(一對多) 커뮤니케이션으로 웹 접속 없이도 사용이 가능한 특성을 잘 보여주고 있다[4].

SNS에서 얻는 관광정보를 보면 이용자가 의사결정을 내리는데 필수적으로 수반되는 관광욕구와 관광동기를 불러 일으키는 중요한 매체 역할이다. 다양한 경로와 자료들을 가지고 관광지를 결정을 하는데, 이 과정에서 위험요인을 최소화 하고 만족도는 높이려는 경향이 수반된다. 인스타그램 사용자들의 사용현황을 보면 댓글과 공유하기 기능보다는 '좋아요' 단추를 클릭하는 것으로 반응을 한다. 이렇게 실시간으로 생성되는 데이터는 매우 방대하다.

본 논문에서는 대상 SNS를 인스타그램로 선정하였고, 성격 유형이 '일상 공유'로 주 사용자가 되는 젊은 연령층(10대~30대초반)을 대상으로 동시 발생하는 해시태그 분석을 통한 관광지 정보 제공 데이터 구성도를 제안하였다.



▶▶ 그림 1. 해시태그 분석을 통한 데이터 제공 구성도

기본적으로 해시태그로 검색을 하면 총 개수는 명시가 되는데, 일정기간에 몇 개의 사진이 업로드 되었는지는 알 수 없다. 인스타그램 API를 이용하여 일정 기간을 설정 후 관심 키워드를 지역(A)로 선정한 후 해시태그 정보를 수집한다. 두 번째로 함께 연결되어 동시발생하는 해시태그 정보를 수집하여, 서로 직/간접적 연관성을 가진다고 가정 후 연관성을 분석한다. 세 번째로 동시발생 키워드에서 불필요한 정보를 제거하고, 관심 키워드에 대해서만 부분 추출 후 빈도수가 높은 키워드부터 순위 나열을 한다. 마지막으로 설정된 기간 동안의 정보에 대하여 볼거리, 먹을거리 등 상위 카테고리를 지정하고 정보를 제공한다.

예제로 [표1.]을 보면 관광정보를 수집하기 위해서 지역(A)의 해시태그의 게시물 개수를 검색한 결과 2014년 10월 현재까지 30,668의 게시물이 업로드 되었다. 지역(A)에 대해 해시태그 검색을 하면 연결성이 있는 여러 가지 키워드가 있는데, 공통적으로 해당되는 키워드가 생성되므로 데이터 분석 후 카테고리를 설정하고, 순위 나열 된 키워드 결과 값에 따라서 정보를 얻을 수 있다.

표 1. 구현예시의 태그추출 결과

키워드 명	결과 데이터	카테고리 분류
지역(A) (30,668)	a(4,987)	관광지
	b(2,680)	
	c(943)	
	d(779)	
	e(479)	
	...	
	f(713)	맛집
	g(331)	
	h(250)	
	i(223)	
	j(143)	
	...	체험
	...	

III. 결론

본 논문에서는 빠른 트렌드의 SNS 인스타그램을 이용하여, 해시태그 데이터를 가지고 동시출현의 연관성을 분석을 하고, 현재 트렌드에 맞게 제공하는 방법에 대하여 데이터 제공 구성도, 관심키워드 우선적용 방법을 설명하였다. 제시되는 방법으로 구현을 할 때 이벤트성 게시물, 연령대, 성별 등에 대한 문제점들을 해결 해야한다. 또한 합리적인 정보제공과 만족에 대한 극대화에 대해서도 고민을 해봐야 한다. 이러한 연구는 시대적 흐름에 맞추어 진행이 되고 있는 추세이므로 선행연구 조사 및 세부적인 데이터 추출방법을 연구가 필요하다. 향후 지역 관광 정보제공을 실제 구현하기 위하여 연구할 예정이다.

■ 참고 문헌 ■

- [1] 윤홍근 “문화산업에서 빅데이터 활용방안에 관한 연구”, 글로벌문화콘텐츠, 제10호, pp.157-180, 2013.
- [2] 한국콘텐츠진흥원, “콘텐츠 분야에서의 빅데이터 기법 활용 사례”, CT 이슈분석 II, 2월호, 통권34호, pp. 50-62, 2014.
- [3] 오성수, 김성후, “관광지선택에 SNS관광정보 특성이 미치는 영향”, 산업경제연구 제25권 제5호, pp. 3255-3271, 2012.
- [4] 한국콘텐츠진흥원, “SNS의 진화, ‘마이크로 블로그’의 성장”, 2009.