

해시태그[#]를 이용한 연관 규칙 분석 연구

이종화^a, Hoanh-Su Le^b, 이현규^c

^a 부경대학교 경영대학 경영학부 박사
Tel: +82-51-629-5177, E-mail: newjwcom@daum.net

^b 호치민대학교 정보시스템학과 조교수
Tel: +82-51-629-5177, E-mail: lehoanh-su@gmail.com

^c 부경대학교 경영대학 경영학부 교수
Tel: +82-51-629-5177, E-mail: hyunqlee@pknu.ac.kr

Abstract

고객이 니즈(Needs)를 찾기 위한 빅데이터 분석은 이미 많은 연구가 이루어지고 있다. 비정형적 데이터를 정형화 하는 노력들은 웹 마이닝, 텍스트 마이닝, 오피언 마이닝 등으로 연구가 이루어진다. 일상 생활에서 이슈가 되는 것들을 네트워크에서 함께 공유하며 의견을 나누어가는 소셜 네트워크 서비스(SNS)에 대한 관심도 더욱 높아지고 있다. 고객의 SNS 반응도가 높을수록 수많은 비정형 데이터가 분석되어야 한다. 사용자 또한 원하는 주제의 검색을 편리하게 돕는 기능으로 해시태그(Hash Tag)인 [#]를 사용하여 해당 주제에 글들을 검색과 의견을 공유하고 있다. 이런 주제어는 의미가 함축적이고 많은 사람이 공유할 수 있는 단어로 사용자들간의 정제된 키워드를 사용함으로 보다 많은 의미를 가지고 있다. 이런 해시태그를 활용하여 SNS의 시간 자료 수집과 태그간의 연관 분석이 가능한 시스템 설계와 웹 페이지 구현을 통하여 실시간으로 고객들의 니즈를 파악하고자 한다.

Keywords:

Hash Tag; Association Analysis; Real-time Processing; Web Mining; Big Data

I. 서 론

소셜 네트워크 서비스(Social Network Service)는 자신의 일상이나 생각을 주고 받을 수 있는 인터넷 서비스이다. 글과 사진 동영상 다양한 관심이나 활동을 공유하는 사람들 사이의 네트워크를 구축하는 온라인 서비스이다(나누리 외, 2017; 윤여경, 2016; 남민지 외, 2015).

“오늘은 어떤 글들이 올라왔나 한 번 볼까?”라는 일상 생활에서 소셜 네트워크 서비스가 시작된다.

아이의 웃긴 영상과 정치이야기, 사회 이슈가 되는 영상들 등 우리가 “좋아요”가 많은 글들에 더욱 눈길을 보내고 흥미가 간다. 또한, 내가 “좋아요”라고 누르면 내 친구도 그 글을 볼 수 있게 되어 내 친구도 그 글이 재미있다면 “좋아요”를 선택하고 되며 그렇게 “좋아요”는 점점 늘어나게 된다.

SNS의 홍보는 친구를 통해서 이루어진다.

200명의 친구와의 SNS를 하고 있다면 나의 “좋아요”는 200명들에게 해당 콘텐츠(Contents)를 연결할 것이다. 또한 그 친구들이 100명의 친구들로 연결되어 있다면 무려 20,000명의 친구가 같은 콘텐츠와 연결된다고 할 수 있다.

만약 같은 방법으로 5,000명의 친구와의 네트워킹이 되어 있고 각 100명의 친구가 연결 되어 있다면 500,000명에게 해당 콘텐츠를 보여줄 수 있는 효과가 있을 것이다. 이런 식의 과급효과가 SNS들이 갖고 있는 장점이기도 하다. “좋아요”의 가치라 볼 수 있다.

이러한 네트워크의 평창으로 SNS는 새로운 마케팅 채널로 활용되고 있다. 어떤 콘텐츠에 어떤 사람이 있는지 마케팅 채널은 어떻게 해야 하는지 등 소셜 네트워크 서비스를 이용한 마케팅이다. 인스타그램(<https://www.instagram.com/>)은 사진 1장, 60초 동영상, [#]사용 소통하며 메시지 하나를 담아서 전달하는 SNS이다.

“나를 검색해 봐”라는 뜻으로 해시태그를 게시물 사용하여 고객을 찾아 팔로우 게시물에 “좋아요”를 활용하여 반응을 이끌어 내기도 한다.

표현하고자 하는 말을 함축적으로 단어화 시켜 공감대 형성의 단어로 만들어 SNS상에서 그들만의 언어로 사용하고 있다. 본 연구는 이러한 해시태그 분석을 통하여 보다 다양한 마케팅 전략을 살펴보고자 한다. 인스타그램의 모든 콘텐츠를 대상으로 해시태그 검색과 관련 콘텐츠의 해시태그를 실시간 크롤링(Crawling)하여 한글의 명사 추출과정 없이 해당 해시태그를 빈도 분석 및 워드클라우드(Word Cloud) 분석 결과를 웹 페이지에

구현하고자 한다. 또한 각 콘텐츠의 해시태그 분석을 통하여 해시태그간 서로의 관련성을 분석하고자 한다.

II. 이론적 배경

1. 데이터마이닝

데이터 마이닝은 대용량 데이터 내에서 의미 있는 패턴을 찾아 집단을 분류하고 예측하며, 유사집단으로 묶거나 동시 또는 순차적으로 발생하는 의미 있는 연관관계를 찾는 것을 목표로 하고 있다. 마케팅 전략으로 보면 기업이 보유하고 있는 방대한 데이터에 존재하는 유용한 정보를 발굴하여 경영자의 의사 결정에 도움이 되는 지식을 제공할 수 있도록 하는 과정으로, 데이터가 아닌 지식 수준의 고객 통찰력을 기반으로 수행되어야 하는 고객관계마케팅 전략에 있어 핵심적인 기반 기술이다.

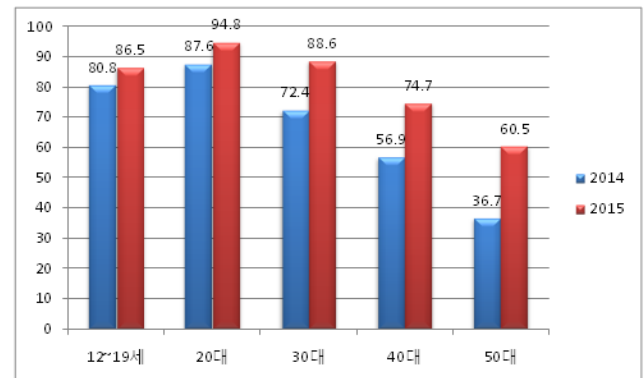
정형 데이터 마이닝은 의사 결정 나무(Decision Trees)와 같이 입력데이터를 통해 학습함으로써 새로운 데이터에 대해 실제로 어떤 그룹에 속해있는지를 구분하는 분류분석(Classification Analysis)이 있으며, 연속형 값을 예측하는 것으로 시계열 변수를 이용해 예측하는 예측분석(Prediction Analysis), 특성에 따라 고객을 여러 개의 집단으로 나누어 가까운 데이터 끼리 하나의 그룹으로 분류하는 군집분석(Clustering Analysis), 장바구니 분석이라 하며 구매 품목간의 관계를 알아보는 연관 분석(Association Analysis) 등이 있다(lee et al., 2016).

비정형 데이터 마이닝은 문서의 요약, 분류, 군집, 추출기능을 하는 텍스트 마이닝(Text Mining)과 서로의 관계가 확산되어 형성된 사람들 사이의 네트워크를 분석하는 사회 연결망 분석(Social Network Analysis)로 구분 된다(Lee and Lee, 2015).

2. 해시태그(HashTag)

SNS의 변화는 모바일 디바이스의 사용량 증가와 함께 팽창하고 있다. <Fig. 1>와 같이 모바일 SNS 이용율을 보여주고 있다.

모바일 SNS 이용률은 67.8%(2014년)에서 80.9%(2015년)으로 전년대비 대폭(13.1%p) 상승한 가운데, 특히 50대의 모바일 SNS 이용률은 36.7%(2014년)에서 60.5%(2015년)로 증가폭(23.8%p)이 가장 높았다(한국인터넷진흥원, 2015).



<Figure 1> 연령별 모바일 SNS 이용률(%)

웹 1.0인 월드 와이드 웹(WWW)은 사용자가 신문이나 방송처럼 일방적으로 정보를 받는 것이었고, 웹 2.0은 참여, 공유, 개방의 플랫폼 기반으로 정보를 함께 제작하고 공유하는 것이었다. 그리고 웹 3.0은 개인화, 지능화된 웹으로 진화하여 개인이 중심에서 모든 것을 판단하고 추론하는 방향으로 개발되고 활용될 것이다. 이런 웹 환경에 맞게 SNS 또한 변화가 이루어진다. 1세대는 미니홈피를 통한 기존 오프라인 인맥을 확장하는 형태이다. 2세대는 모바일 디바이스의 등장으로 단순한 내용을 실시간으로 전달하는 페이스북(Facebook)이나 트위터(Twitter)를 활용하는 형태이다. 3세대는 작은 단위의 소셜 플랫폼을 연결하고 특정 주제를 중심으로 관심사 공유하는 큐레이션(Curation)이다. 사진이나 영상 등 시각물을 중심으로 한 이용자 맞춤형 서비스로 링크트인, 인스타그램, 텀블러 등이 대표적이다(남민지 외, 2015).

[#]은 C언어에서 헤더 파일을 호출하기 위한 기호로 C언어가 우선적으로 처리해야 하는 명령어 앞에 사용되었던 기호이다. SNS에서 해시태그는 게시물에 일종의 꼬리표이며 검색어에 링크를 설정하는 기호로 사용되고 있다. SNS에 게시물을 올리고 해시태그를 달면, 다른 사용자가 그 게시물과 같은 해시태그를 단 게시물을 함께 찾아 볼 수 있다는 것이다. 사진과 글을 입력 그리고 ‘#’를 하는 순간 사람들 속으로 들어간다. 나이, 직업도 모르는 누군가의 세상으로 그들과 나를 연결하는 것이 해시태그이다(윤여경, 2016). 트위터, 인스타그램 등 소셜 미디어에서 특정 단어들을 편리하게 검색할 수 있도록 하는 메타데이터의 한 형태 “#오늘은필떡지”, “#공유합니다”, #가 불을 경우 다른 사용자에게 공유되는 비율 두 배로 증가한다고 한다. 2014년 8월 미국에서 열여덟 흑인 소년 백인 경관이 쏜 총에 사망하자 누군가의 SNS에서 “#BlackLivesMatter”, “#흑인의생명도소중하다” 해시태그가 등장과 공유가 시작되었고 공감하는 시민들이 급속히 생겨 나게 되었다. 미국 법 집행 기관의 인종차별적인 형태에 분노한 수많은 사람들이 이런 마음을 널리 알리고 공유하기를 원했다는 것을 의미한다고 볼 수 있다.

이런 해시태그의 영향으로 자신이 좋아하는 분야에 심취하여 그와 관련된 것들을 모으거나 찾아보는 “덕질”, “덕후”등이 있고 어떤 단어나 문장에 [#]를 하는 순간 확장되는 가능성과 몇 분 지나지 않아 [#]된 사진들은 다양한 지구촌의 반응으로 이끌어낸다. 본 연구는 이런 해시태그에 대한 연구를 진행하였다.

3. 연관 분석

연관 분석은 흔히 장바구니 분석(Market Basket Analysis)이라고 하며 재화나 서비스인 항목 A를 구매한 경우 B도 구매하게 될 것이라는 상품간의 연관성을 분석하는 방법이다.

기업의 데이터베이스에서 상품, 서비스의 구매 등 일련의 거래간의 규칙을 발견하기 위해 적용되는 방법이다. 마케팅에서는 손님의 장바구니에 들어 있는 물품 간 관계를 알아본다는 의미에서 장바구니 분석이라 한다.

연관 분석의 측정은 얼마나 자주 구매했는가 하는 빈도를 사용하여 지지도(Support), 신뢰도(Confidence), 향상도(Lift)의 결과로 유용한 지식을 가늠할 수 있다.

지지도(Support)는 전체 거래 중 항목 A와 항목 B를 동시에 포함하는 거래의 비율로 정의된다. 식(1)은 전체 거래 중 항목 A와 항목 B를 동시에 포함하는 거래가 어느 정도인지를 나타낸다.

$$\text{Support} = \frac{\text{A와 B를 포함하는 거래수}}{\text{전체 거래수}(N)} \quad (1)$$

식(2)의 신뢰도(Confidence)는 항목 A를 포함한 거래 중에서 항목 A와 항목 B가 같이 포함될 확률이 어느 정도인가를 나타내며 연관성의 정도를 확인할 수 있다.

$$\text{Confidence} = \frac{\text{A와 B를 포함하는 거래수}}{\text{A를 포함하는 거래수}} \quad (2)$$

식(3)의 향상도(Lift)는 A가 주어지지 않았을 때의 품목 B의 확률에 비해 A가 주어졌을 때의 품목 B의 확률의 증가 비율이다. 다시 정리하면 B만 구매할 비율과 A와 B를 함께 구매할 비율을 나타낸다. 향상도가 ‘1’ 이면 두 품목이 서로 독립적인 관계이며 ‘1’ 보다 크면 두 품목이 서로 양의 상관관계, ‘1’ 보다 작으면 서로 음의 상관관계가 있다고 볼 수 있다.

$$\text{Lift} = \frac{\text{A와 B를 포함하는 거래수}}{\text{A를 포함하는 거래수} \times \text{B를 포함하는 거래수}} \quad (3)$$

본 연구에는 정형 데이터 마이닝 분석 기법인 연관 분석을 비정형 데이터에 적용하여 연구하고자 한다. 상품이나 서비스의 품목 리스트가 아닌 해시태그를 통하여 키워드와의 연관성을 확인하고자 한다. 상품 거래 목록을 전체 거래 수(N)는 SNS상의 하나의 콘텐츠를 기준으로 전체 자료의 건수로 환산하여 진행하였다. 비정형 자료를 정형화 하여 분석하는 과정을 연구하고자 한다.

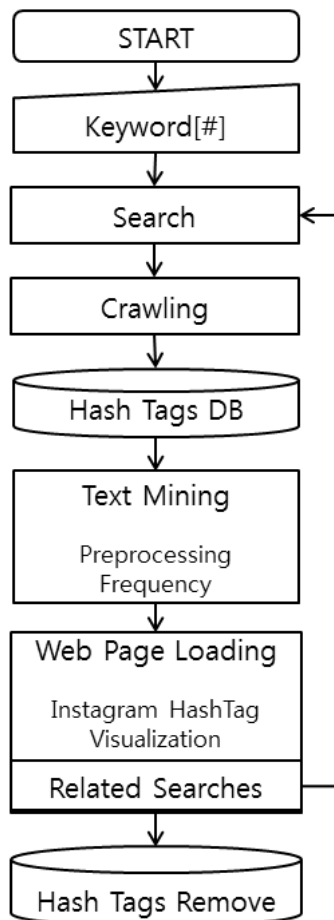
III. 연구방법

많은 연구자들은 구조화 되지 않은 빅데이터에 관한 끊임없이 연구를 진행하고 있다. 빅데이터 이름처럼 방대한 양의 데이터를 수집하는 절차와 수집된 데이터의 분석 과정 또한 데이터 량과 비례하여 많은 시간이 소요된다. 하지만 웹 환경에서 고객들의 다양한 소리는 시간을 다투듯 실시간으로 계속적으로 쌓여가고 있다. 연구자는 많은 양의 연구 데이터 수집과 분석 과정의 이원화로 변해가는 웹 환경을 실시간 대응하기엔 많은 한계점을 나타내고 있다(이철성 외, 2013; 임좌상, 김진만, 2014; Lee and Lee, 2015).

본 연구는 데이터 수집과정과 분석과정을 웹 환경에서 일원화하여 실시간 처리가 가능한 프레임워크를 연구하였다. 또한, 고객의 니즈 분석을 위한 해시태그를 수집하여 서로의 관련성 있는 태그들을 발견하고자 한다.

웹 페이지에서 사회 이슈 키워드(해시태그)를 입력하면 본 연구 시스템이 자료의 수집 과정인 크롤링 (Crawling)과 ‘#’과 웹 언어 스크립터 등 웹 마이닝 과정에 불필요한 불용어 제거를 통한 전처리 과정을 거친다. 본 연구의 주재료인 해시태그는 사용자들이 이미 의미 있게 사용하는 단어로써 더 이상 정제 과정이 필요 없는 데이터를 대상으로 하고 있다. 즉, 일반적인 텍스트마이닝에서의 명사 추출 과정을 생략하고 해시태그의 빈도를 활용하여 워드 클라우드(Word Cloud) 분석과 사용자 단위로 구분된 해시태그를 분석하여 해시태그간 연관 분석 (Association Analysis)을 진행하였다. <Fig 2>는 본 연구의 실시간 이슈분석 시스템의 프로세스이며 웹 페이지는 Google Chrome 웹 브라우저에 최적화되어 설계되었다. Linux CentOS 서버 환경에 별도의 실행 파일 없이도 JavaScript 환경의 언어를 동적으로 실행이 가능하며 문서를 빠르고 쉽게 작성할 수 있는 장점의 PHP(Hypertext Preprocessor) 스크립터 언어를 사용하였다(구홍서, 2000; 문장식, 김흥기, 2014; 정원기, 문수목, 2010; 류석영, 2016). 연구 대상은 인스타그램(Instagram) 웹페이지를 대상으로 특정 해시태그를 입력하면 인스타그램 결과 모든 페이지들을 크롤링 가능한 패키지를 Python을 활용하여 개발하였으며 HTML, JavaScript, jQuery, Ajax 등의 스크립터 언어를 사용하여 사용자

인터페이스를 구축하였으며 신문 기사 텍스트 분석 및 시각화는 R Program을 사용하였다.



<Figure 2> Research Model and process

IV. 연구결과

본 연구는 인스타그램에서 제공되는 모든 콘텐츠를 대상으로 해시태그 실시간 분석 시스템을 구축하고자 한다. 사진이나 동영상 공유 애플리케이션 서비스를 제공하는 인스타그램에서 해시태그(키워드)를 분석하여 사용자들이 느끼고 있는 니즈(Needs)의 방향을 보다 빠르게 확인할 수 있는 실시간 시스템을 구현 하였다.

실시간 이슈 분석 시스템은 실험용 Server에서 자료 검색, 크롤링, 텍스트마이닝, 사용자 인터페이스까지 구축하는 과정을 구현하였다. 리눅스 환경에서 PHP, Java Script, jQuery, R program 등을 연동하여 사용자 개입 없이 시스템에서 전 과정을 분석, 처리하도록 구현하였다. 웹 프로그램 언어들의 연동은 구현이 상대적으로 자유로우나 PHP에서 R 프로그램 연동은 “Rscript” 명령을 통하여 미리 준비된 텍스트마이닝 소스 코드를 실행할 수 있었다. 비교적 간단한 문법으로 코딩이 가능하며 다른 언어와의 호환성 측면에도 쉬운 언어인 Python를 이용하여 크롤링

과정을 진행 하였다.

개인의 문제가 아닌 이미 사회적 문제가 되어 청년 실업 즉, 일자리 구직에 관한 키워드를 중심으로 “취업”, “채용”, “취업준비”, “취준생”, “취직” 해시태그를 대상으로 실험을 하였다.

<표 1>는 실제 연구 대상인 인스타그램에서 해당 해시태그 검색 결과의 빈도 통계이다.

순번	해시태그	빈도
1	#취업	97,285건
2	#취준생	95,565건
3	#채용	16,873건
4	#취업준비	12,288건
5	#취직	12,148건
6	#구직	6,906건

표 1 실험 데이터 통계

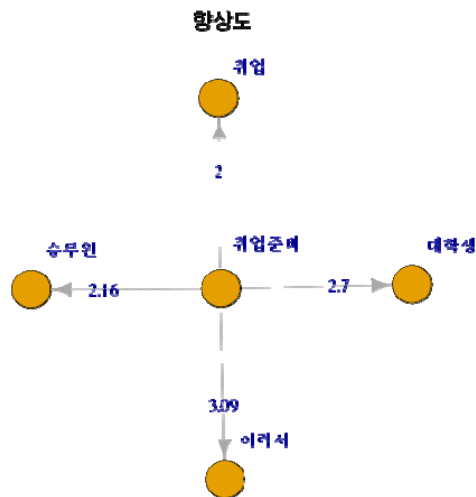
“#취업” 해시태그가 가장 많은 빈도를 나타냈고 취업준비생의 준말인 “#취준생”이 그 뒤를 이었다. “#채용”, “#취업준비”, “#취직”, “#구직”등이 그 뒤를 이었고 “#해외 취업”, “#경력취업”, “#취업박람회”등 해시태그도 발견되었다. <Fig 3>은 “#취업준비” 해시태그를 사용한 콘텐츠 분석 중 워드클라우드 분석을 살펴보면 “취준생”, “취업”, “소통”, “일상”, “데일리”, “대학생”, “취업정보” 등 취업관 관련있는 해시태그가 함께 사용되었으며 “선팔”, “일상스타그램”, “맞팔”, “셀스타그램”, “인스타그램” 등 해시태그 검색을 의식한 키워드들도 빈도가 높은 것으로 나타났다.



<Figure 3> [#취업준비] 워드클라우드 분석

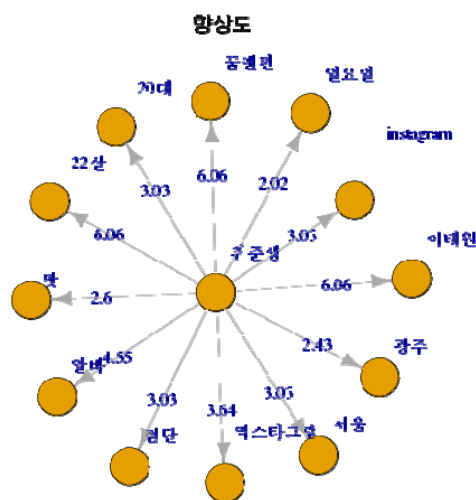
장바구니 분석 즉, 연관 분석의 향상도는 식(3)과 같이 B만 구매할 비율과 A와 B를 함께 구매할 비율을 나타낸다. 서비스나 제품을 패키지 판매등에

사용하는데 <Fig 4>은 [#취업준비] 해시태그를 사용하여 항상도를 나타낸 그림이다. [#취업준비], [#이력서] 각각 사용할 때보다 함께 사용하는 콘텐츠가 3배이상 많다는 뜻이다. 즉, [#취업]과 [#취업준비] 또한 각각 사용된 경우보다 함께 사용된 경우가 2배나 된다고 볼 수 있다.



<Figure 4> [#취업준비] 항상도

<Fig 5>는 [#취준생]을 연관 분석한 결과로 [#알바], [#이태원], [#22살], [#20대] 등이 눈에 띈다. 20대 젊은 청년들이 취업준비를 위하여 아르바이트 구직에 많은 관심이 있는 것으로 보이고 [#광주], [#서울] 등 지역관련 해시태그도 나타났다.



<Figure 5> [#취준생] 항상도

연령을 함께 키워드로 만들어 [#20대취업], [#30대취업], [#40대취업], [#50대취업] 등과 지역의 구직활동을 위한 키워드 [#울산취업], [#부산취업], [#광주취업] 등도 나타났다. 기타 [#공기업취업], [#해외취업], [#장애인취업], [#대기업취업], [#장년취업], [#노인취업] 등 다양한 계층들의 취업 활동을 실감나가 해주는 해시태그가 많이 나타났다.

V. 결론

본 연구는 SNS인 인스타그램을 활용한 실시간 연관 분석 시스템을 설계하고 구현, 그리고 실험까지 진행하였다. 그 결과를 종합적으로 정리해보면 다음과 같다.

해시태그(#)를 사용하면 링크가 형성되어 같은 태그를 작성한 글들끼리 모아주는 기능으로 주제의 검색을 편리하게 돕는 기능이다. 이를 활용하여 인스타그램에서 실시간 검색으로 해시태그를 데이터로 빠르게 분석할 수 있는 웹 페이지를 구현하였다. 포털 사이트 검색 기능처럼 사용자가 궁금해 하는 해시태그를 이용하여 관련된 단어가 있는 모든 콘텐츠를 방문하고, 해시태그를 크롤링하여 서버 메모리에 로딩되고 R 프로그래밍을 통하여 시각화하는 텍스트마이닝 처리가 되며 사용자의 질문에 대한 이슈 분석결과를 제공하게 된다. 본 연구 과정에는 누구나 자유롭게 소프트웨어를 코딩하여 유용한 기술을 공유하며 사용자들이 서로의 기술을 함께 업데이트 가능한 오픈 소프트웨어를 활용하였으며 Linux, MySql DB, HTML, CSS, JavaScript, jQuery, R program, Python 등이 사용되었다.

웹 페이지는 Google Chrome 웹 브라우저에 최적화되어 설계되었고, Linux CentOS 서버 환경에 Java 환경의 언어를 동적으로 실행하며 PHP(Hypertext Preprocessor), HTML, JavaScript, jQuery, Ajax 등의 스크립트 언어를 사용하여 사용자 인터페이스를 만들었으며 신문 기사 텍스트 분석엔 R program을 사용하였다. 또한, 웹 페이지 로딩 시간을 줄이기 위하여 각 페이지를 불러오는 모든 작업은 ajax를 이용하여 실시간으로 페이지 이동 없이 설계되었다.

빅데이터 분석은 방대한 데이터 내에서 패턴을 찾아 예측을 하는 것이다. 가장 단순한 빈도 분석을 실시간으로 처리할 수 있는 시스템을 구축하여 보다 빠른 이슈의 흐름을 파악하는데 혁신적인 검색 도구라 본다. 이슈 키워드와의 연관 키워드의 비율을 확인 할 수 있듯 연구 프레임워크를 실제 장바구니 분석에 적용한다면 상품간의 연관성을 쉽게 찾을 것으로 기대된다.

또한, 해시태그 분석을 오피니언마이닝을 통하여 긍정 사전과 부정 사전을 구축하여 고객들의 생각을

좀 더 분석 가능한 행태의 도구 개발이 이루어지길 기대한다.

참 고 문 헌

- [1]구홍서, "WWW과 데이터베이스 연동기술의 조사분석," 정보과학회지, 제18권, 제4호, 2000, pp. 32-40.
- [2]나누리, 이중식, 김진영, "이미지 기반 SNS에서의# 해시태그 검색 전략," 한국 HCI 학회 학술대회, 2017, pp. 962-965.
- [3]남민지, 이은지, 신주현. "인스타그램 해시태그를 이용한 사용자 감정 분류 방법," Journal of Korea Multimedia Society 제18권, 제11호, 2015, pp. 1391-1399.
- [4]류석영, "자바스크립트 웹 앱 분석과 결함 검출," 정보과학회지, 제34권, 제3호, 2016, pp. 10-14.
- [5]문상식, 김기홍, "IT 환경 변화에 따른 한국의 오픈소스 소프트웨어의 정책방향 연구," 인터넷전자상거래연구, 제14권, 제1호, 2014, pp. 203-221.
- [6]윤여경, "SNS 해시태그 방식에 대한 비교평가 UI 디자인 연구," 한국상품문화디자인학회 논문집, 제46권, 2016, pp. 103-113.
- [7]이철성, 최동희, 김성순, 강재우, "한글 마이크로블로그 텍스트의 감정 분류 및 분석," 정보과학회논문지: 데이터베이스, 제40권, 제3호, 2013, pp. 159-167.
- [8]임좌상, 김진만, "한국어 트위터의 감정 분류를 위한 기계학습의 실증적 비교," 멀티미디어학회논문지, 제17권, 제2호, 2014, pp. 232-239.
- [9]정원기, 문수목, "웹 페이지 자바스크립트 분석과 자바스크립트 엔진의 성능 평가," 한국정보과학회, 제37권, 제2호, 2010, pp. 181-182.
- [10]Lee, J. H. and Lee, H. K., "A Study on Unstructured Text Mining Algorithm through R Programming based on Data Dictionary," Journal of the Korea Society Industrial Information System, Vol. 20, No. 2, 2015, pp. 113-124.
- [12]Lee, J. H., Le, H. S., and Lee, H. K., "Research on Methods for Processing Nonstandard Korean Words on Social Network Services," Journal of the Korea Industrial Information Systems Research, Vol. 21, No. 3, 2016, pp. 35-46.