

SNS를 이용한 잠재적 광고 키워드 추출 시스템 설계 및 구현

서현곤*, 박희완
한라대학교 정보통신소프트웨어학과

Design and Implementation of Potential Advertisement Keyword Extraction System Using SNS

Hyun-Gon Seo*, Hee-Wan Park
Department of Information Communication Software, Halla University

요약 빅데이터 처리 분야에서 중요한 이슈 중 하나는 인터넷의 주요 키워드를 추출하고 이것을 이용하여 필요한 정보를 가공하는 것이다. 현재까지 제안된 대부분의 키워드 추출 방법들은 대형 포털 사이트의 검색기능을 기반으로 이미 게시된 글이나 작성된 문서 또는 고정된 내용에 기반하고 있다. 본 논문에서는 SNS에 게시되는 다양한 이슈, 대화, 관심 분야, 의견 등 동적인 메시지를 기반으로 이슈 키워드 및 연관 키워드를 추출하여 잠재적 쇼핑 연관 키워드 광고 마케팅에 도움을 주는 시스템(KAES: Keyword Advertisement Extraction System based on SNS)을 개발한다. KAES 시스템은 특정 계정 리스트를 작성하여 SNS에서 빈도수가 가장 많은 핵심 키워드 및 연관 키워드를 추출한다.

주제어 : 빅데이터, 키워드 마케팅, SNS, 이슈 키워드, 연관 키워드

Abstract One of the major issues in big data processing is extracting keywords from internet and using them to process the necessary information. Most of the proposed keyword extraction algorithms extract keywords using search function of a large portal site. In addition, these methods extract keywords based on already posted or created documents or fixed contents. In this paper, we propose a KAES(Keyword Advertisement Extraction System) system that helps the potential shopping keyword marketing to extract issue keywords and related keywords based on dynamic instant messages such as various issues, interests, comments posted on SNS. The KAES system makes a list of specific accounts to extract keywords and related keywords that have most frequency in the SNS.

Key Words : Big data, Keyword marketing, SNS, Issue keyword, Related keyword

1. 서론

모바일 접근성이 용이해지면서 전 세계적으로 많은 사람들이 SNS(Social Network Service)를 이용하여 메시지를 교환하거나 사회적 이슈 또는 자신의 관심 분야에 대한 다양한 의견을 제시하고 있다. 이런 메시지 속에는 다양한 분야의 풍부한 정보가 포함되어 있어, 이것을

가공하고 새로운 정보를 만들고자 하는 시도가 빅데이터 기술을 중심으로 다양하게 제시되고 있으며, 특히 특허 분야에도 키워드 연결을 활용하여 빅데이터를 분석하는 분야까지 확장되고 있다[1].

트위터(Twitter)는 전 세계적으로 많은 사람들이 이용하는 대표적인 SNS로 문서가 간결하고 즉시적으로 전달되는 특징이 있어 기업의 제품 평가, 선거 등 다양한

*ACK : This work was supported by Korea Technology and Information Promotion Agency for SMEs.(C0509691)

*Corresponding Author : Hyun-Gon Seo (hgseo@halla.ac.kr)

Received April 25, 2018

Accepted July 20, 2018

Revised May 15, 2018

Published July 28, 2018

분야 이슈를 빠르게 확산시키는 특징을 가지고 있다. 최근 트위터의 주요한 관심 키워드와 토픽 등을 추출하기 위한 연구가 다양하게 이루어지고 있다.

[2]는 트위터를 이용하여 특정 제품명에 대한 토픽 키워드를 추출하고 계량정보학분야에서 내용 분석을 위해 많이 쓰이는 방법인 동시출현단어분석(Co-word analysis)을 이용하여 노드와 예지를 통해 토픽과 관련 키워드를 그래프로 표현하여 토픽 또는 이슈 변화에 빠르게 대응하는 마케팅 도구로 활용하였다. [3]은 트위터 사용자의 트렌드 추출을 위해 기존의 클러스터링 기법 중 해석이 쉽고 직관적이며, 다양한 시간 스케일에 주제를 추출할 수 있는 NMF(Non-negative Matrix Factorization)을 이용하여 사용자가 분석하고자 하는 구간 안에서의 가장 유행한 주제는 물론, 하루, 일주일, 한 달 등 관심 있는 주제를 추출하는 연구를 하였다. NMF는 기본 클러스터링 기법들과 비교했을 때 결과의 해석이 쉽고 직관적인 장점이 있음을 증명하였다. [4]는 기본 개별단어에 대한 평균적인 출현 횟수를 계속 관찰하다가 특정단어의 출현 빈도가 평소의 평균을 크게 상회하게 되면 이것을 인기 키워드로 판정한 알고리즘을 사용했지만, TwitterMonitor에서는 GroupBurst라는 개념을 도입하여, 인기 키워드들을 하나의 트윗 안에서의 동시 출현하는 빈도를 측정하여 필터링하는 기법을 제안하였다. 하지만 이 방법은 과도한 연산 및 시간과 의미 중심의 토픽 추출에는 단점을 가지고 있다.

트위터와 함께 가장 많이 이용되는 SNS는 페이스북(facebook)이다. 특히 페이스북의 검색 서비스기능은 네이버, 구글 등과 같은 포털 사이트의 키워드 검색과는 차이가 있다. 트위터는 일정한 검색 구문을 이용하여 검색 결과를 만들기 때문에 특정 키워드와 연관된 사람이나 장소, 위치, 음악이나 영화, 특정한 페이지나 그룹, 이벤트 등 검색이 가능하다. 즉, 사용자의 정보(나이, 지역, 관심사, 행동) 및 활동을 이력을 기반으로 타겟팅 지정이 가능하다. 페이스북의 이와 같은 기능 때문에 특정 상품을 좋아할 법한 사람들을 찾아내어 맞춤형(custom audiences) 목록을 생성하는 타겟팅 도구로 많이 활용되고 있다[5]. 하지만 페이스북에 올라온 메시지를 활용하여 주요 이슈 또는 핵심 토픽을 추출하여 타겟팅으로 활용하는 기능은 없다.

따라서 본 논문에서는 트위터, 페이스북과 같은 SNS에 올라오는 다양한 문서에서 관심 분야를 지정하여 현

재 가장 많은 빈도수를 가지는 핵심 키워드 및 연관 키워드를 추출하는 방법을 제안하고 잠재적으로 광고 키워드로 활용가능성이 높은 키워드를 추출하는 방법을 제안한다.

2. 관련 연구

최근 온라인 쇼핑과 인터넷 쇼핑의 시장이 증가하면서, 제품 생산자들은 오프라인보다 온라인을 이용한 마케팅을 선호하기 때문에 인터넷 쇼핑몰을 운영자는 가능한 많은 사람들에게 자신들의 상품 사이트나 블로그를 노출하기를 원한다. 특히 인터넷 사용자를 자사 사이트로 유도하기 위하여 검색 포털 사이트에 특정 키워드를 입력했을 때 자사 사이트의 링크를 노출시키고, 그 대가를 비용을 지불하는 키워드 마케팅이 부각하고 있다. 이에 따라 인터넷 쇼핑몰 운영자들의 키워드 검색 광고비용이 증가하고 있고, 이에 대한 체계적인 연구가 수행되고 있다.

국내 인터넷 키워드 광고는 정액제 검색 광고인 CPM(Cost Per Millenium), 클릭 당 과금 검색 광고인 CPC(Cost Per Click), 행위당 검색 광고인 CPA(Cost Per Action)등으로 분류하였다[5]. CPM은 광고가 1,000회 게재되었을 때를 기준으로 책정한 광고비를 뜻하며, 비용산정은 사전에 측정된 금액만 지불한다. CPC는 클릭수에 따라 과금이 이루어짐으로 실제 광고 효과에 비례하여 광고비가 책정된다. 즉, CPC광고방식이 CPM 광고 방식에 비해 광고효과 측면과 광고비용 측면에서 보다 우수하다. 이에 비해 CPA는 광고를 접한 고객이 대상 사이트에 접속한 후, 해당 사이트에서 활동을 수행할 때마다 광고비용을 부과하는 모델로 고객들이 광고주가 원하는 행동을 취하도록 하는 데에 목적이 있으며, 쇼핑, 보험, 블로그들의 일부 사이트에서는 CPA 광고 방식을 채택하고 있다.

현재 제안된 많은 키워드 광고 추출 방법들은 대부분 대형 포털 사이트의 검색어를 기반으로 제안된 것들이다. 즉, 검색 빈도수는 높다는 것이 키워드 마케팅 비용기준이 될 수는 있어도 실제 쇼핑과 연관된다고는 할 수 있다. [7]에서는 포털 사이트의 빈도가 높은 검색어 중 쇼핑몰 관점에서 중요한 키워드를 추출하는 작업을 실행하여 쇼핑의도를 포함하고 있을 가능성이 높은 키워드를 추출하

는 방안을 제안하였다. 실험결과 쇼핑의도 포함 키워드가 단순 빈도수 기반의 키워드에 비해 정확도, 재현율, F-Score의 모든 측면에서 상대적으로 우수한 성능을 보이는 것으로 나타났다.

블로그 마케팅은 대형 포털 사이트의 키워드 검색 결과에 따라 상위 노출을 함으로써 비교적 저렴한 비용으로 마케팅 파급효과를 높일 수 있다. [6]에서는 블로그를 상위 노출 가능성이 높은 연관 키워드 그룹을 추천하는 방법을 제안하였다. 먼저 타겟 키워드의 검색 결과에 포함된 블로그 문서들을 수집하여 단어의 빈도수와 위치를 고려한 연관성이 높은 키워드를 추출하고 이들의 연관성을 비교하기 위해 월간 키워드 검색 량, 검색에 포함된 블로그 개수, 블로그의 평균 작성 일을 고려하여 연관성이 높은 키워드 그룹을 추천하였다.

대부분의 키워드 추출 기법은 문서에 나타나는 키워드만을 대상으로 하고 있어, 문서에 나타나지 않은 잠재 키워드를 추출하지 못하는 한계점을 가지고 있었지만 [8]은 잠재적 디리클레 할당(LDA : Latent Dirichlet Allocation)모델을 이용해 실제 문서에는 나타나지 않지만 문서의 중요한 개념이나 내용을 함축하고 있어 문서 요약 및 정보 검색에 중요한 역할을 차지하는 잠재 키워드를 추출하는 방법을 제안하였다[9-10].

본 논문에서는 대부분의 키워드 추출 알고리즘들이 사용하는 대형 포털 사이트의 검색기능 기반 키워드 추출 방식을 이용하지 않고, SNS에 게시되는 다양한 메시지를 기반으로 최근 이슈 키워드를 추출하고 이슈 키워드와 연관관 연관 키워드를 추출하여 키워드 광고 마케팅에 적용하는 시스템을 개발하고자 한다.

SNS 기반 키워드 추출 시스템은 매출을 일으키는 키워드의 조합이 ‘검색 사이트’보다 ‘소셜 사이트’에서 더 정확성을 보장할 수 있다는 데 착안하였다. 대체로 검색 사이트에서는 복수의 키워드를 사용하지 않거나, 단어 위주의 검색만을 하지만, 소셜 사이트에서는 자신의 의도를 더 길게, 더 구체적인 문장으로 표현하기에 구매 의사를 확인하는 데 효과적이기 때문이다.

이런 소셜 사이트(트위터, 페이스북 등)에서 방대한 키워드를 수집하여 우선순위를 부여하고, 이를 네이버 등의 연관키워드 등과 비교하여 좀 더 정확한 키워드 마케팅을 수행할 수 있게 도와줄 수 있을 것으로 기대한다.

3. 키워드 추출 시스템 설계

3.1 키워드 추출 시스템 구조

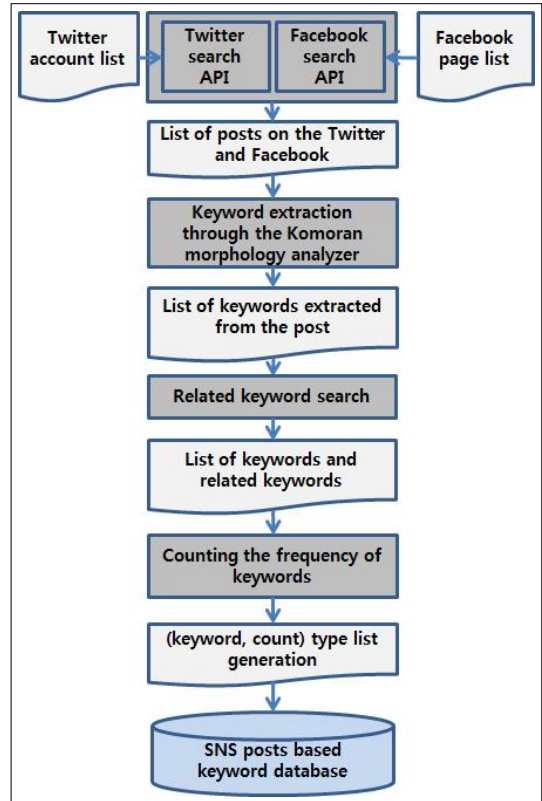


Fig. 1. Structure of the SNS posts based keyword extraction system

SNS 기반 키워드 추출 시스템의 전체 시스템 구조는 Fig. 1과 같다. 먼저 트위터 계정 리스트와 페이스북 계정 리스트를 입력으로 받고 검색 API를 사용하여 트위터 및 페이스북 계정의 게시글을 수집한다. 그리고 형태소 분석기를 사용하여 키워드를 추출한다. 추출된 키워드로부터 네이버 사이트를 통해서 연관 검색어를 추가로 수집한 후, 단어 빈도수를 카운트하여 키워드 데이터베이스에 저장한다.

3.2 키워드 데이터베이스 구조

Field	Type	Meaning	Example
no	int	Serial number (Primary key)	100001
date	text	Date (year/month/date)	2017/09/24
time	text	Time (hour:minute:second)	22:52:32
type	text	Type (Twitter, Facebook)	T or F
keyword	text	Keyword	News
count	int	Frequency	5

Fig. 2. Structure of keyword database

추출된 키워드는 빈도수와 함께 테이블에 저장된다. 이때 MySQL 데이터베이스를 이용하며 테이블 정의는 Fig. 2와 같다.

- no는 시리얼 번호이며 기본키로 사용된다.
- date는 키워드 수집한 날짜를 의미한다. 년, 월, 일 정보가 함께 저장된다.
- time은 키워드 수집한 시간을 의미한다. 시, 분, 초 정보가 함께 저장된다.
- type은 데이터의 종류를 의미한다. T, F 2가지 값을 가질 수 있고, T는 트위터, F는 페이스북을 의미한다.
- keyword는 수집된 키워드를 의미한다.
- count는 수집된 키워드의 빈도수를 의미한다.

MySQL 시스템에서 create 명령어를 이용해서 테이블을 생성한 후 desc 명령어를 사용하면 데이터베이스의 테이블 구조를 확인할 수 있으며 그 결과는 Fig. 3과 같다.

```
mysql> desc table_keyword;
+----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+----+-----+-----+-----+-----+-----+
| no | int(11) | NO | PRI | NULL | auto_increment |
| date | text | YES | | NULL | |
| time | text | YES | | NULL | |
| type | text | YES | | NULL | |
| keyword | text | YES | | NULL | |
| count | int(11) | YES | | NULL | |
+----+-----+-----+-----+-----+-----+
```

Fig. 3. Structure of keyword database

키워드와 빈도수를 데이터베이스에 저장하기 위해서 먼저, 현재 날짜와 시간을 읽어오고, Type값을 준비한다. 그리고 빈도수 최솟값 한도를 통해서 필터링된 키워드와 빈도수 값들을 테이블에 저장한다.

4. 키워드 추출 시스템 구현

4.1 시스템 개발 환경 및 운영 환경

Table 1. Development and operating environment of keyword extraction system

Development environment	- Development toolkit (Compiler) : Eclipse + JDK 1.8 Library - Twitter parsing : twitter 4J[11] - Facebook parsing : JSON simple [12] - Morphological analyzer : Komoran[13] - Database : MySQL Server 5.5[14] - Java and database integration : Connector J 5.1[15]
Operating environment	- Windows(Windows 7, 8, 10) - Linux - Mac OS

키워드 추출 시스템의 개발 환경 및 운영 환경은 Table 1과 같다. 개발도구로 Eclipse와 JDK 1.8 버전을 사용했으며, 트위터 게시물 파싱을 위해서 twitter 4J 라이브러리를 사용했고[11], 페이스북 게시물 파싱을 위해서 JSON simple 라이브러리를 사용했다[12]. 형태소 분석을 위해서 Komoran 라이브러리를 사용했고[13], 데이터베이스로 MySQL Community Server 5.5를 사용했으며[14] 자바와 데이터베이스를 연동하기 위해서 Connector J 5.1을 사용했다[15].

시스템 운영 환경은 자바 클래스와 MySQL이 동작 가능한 환경이며 대표적으로 윈도우(Windows 7 이상), 리눅스와 MAC OS에서도 운영 가능하다.

4.2 관심 사이트 리스트 기반 키워드 수집

트위터 사용자 리스트로부터 게시글을 추출할 때는 twitter4J 라이브러리를 사용한다. 트위터 접근을 위한 권한이 필요하고, 4개의 변수에 필요한 킷값을 저장하여 사용한다.

사용자 리스트에 저장된 사용자 수만큼 트위터 게시글을 추출하는 작업을 반복한다. 트위터 API를 사용하여 게시글을 수집하는 코드는 Fig. 4와 같다. 이렇게 추출된 게시글은 단어가 아닌 문장 형태이기 때문에 형태소 분석 단계로 전달된다.

Table 2. 4 types of keys for twitter posts extraction

Name	Description
Consumer Key	Developer's key of Twitter API
Consumer Secret	Secret value of developer's key
Access Token	Access token of Twitter API
Access TokenSecret	Secret value of access token

```

List doTwitterCrawling( twitterUserList )
{
    // Set authority keys for the Twitter
    setOAuthConsumerKey(ConsumerKey)
    setOAuthConsumerSecret(ConsumerSecret)
    setOAuthAccessToken(AccessToken)
    setOAuthAccessTokenSecret(AccessTokenSecret);
    initialize messageList;
    for ( i = 0; i < twitterUserListNumber; i++) {
        list = twitter.getUserTimeline(twitterUserList(i));
        for (Status status : list) {
            msg = status.getText();
            messageList.put(msg);
        }
    }
    return messageList;
}
    
```

Fig. 4. Collecting user posts from Twitter

페이스북 페이지 리스트로부터 게시글을 추출하기 위해서는 먼저 페이스북 페이지로부터 페이스북 ID를 확인해야 한다. 페이스북 페이지 수집을 위해서는 페이스북 페이지 이름이 아닌 페이스북 ID가 사용되기 때문이다. Fig. 5는 페이스북 ID 페이지 리스트로부터 각각 페이지 ID를 구하고, 각 ID로부터 페이스북 게시글을 반복해서 추출하는 코드이다. 추출된 게시글은 단어가 아닌 문장 형태이기 때문에 형태소 분석 단계로 전달된다.

```

List doFacebookCrawling( facebookPageList )
{
    initialize messageList;
    for ( i = 0; i < facebookPageList; i++) {
        strPageId = getFacebookId(strUrlName);
        addr = getAddress(PageId, FacebookToken);
        json = ReadPostFromFacebook(addr);
        object = JSONValue.parse(json);
        array = object.get("data");
        for (Object object : array) {
            msg = object.get("message");
            messageList.put(msg);
        }
    }
    return messageList;
}
    
```

Fig. 5. Collecting user posts from Facebook

4.3 추출된 게시글에 대한 형태소 분석기 적용

SNS 게시글로부터 키워드를 추출하기 위해서는 형태소 분석 과정이 필요하다. 본 논문에서는 형태소 분석을 위해서 Komoran 형태소 분석 라이브러리를 이용하였다. Komoran 라이브러리의 대표적인 기능은 다음과 같다.

- 형태소 분석 결과로 문장 내 형태소 위치 정보 제공
- 형태소 분석 결과 중 명사 추출 기능 제공
- 형태소 분석 결과를 plain text로 제공
- 형태소 분석 과정에서 학습형 training 모듈 제공

Komoran은 오픈 소스로 배포되며 공식 홈페이지에서 다운 받을 수 있다[13]. 키워드 검색기에 Komoran 라이브러리를 적용한 소스 코드는 Fig. 6과 같다. SNS로부터 추출한 게시글을 입력으로 받은 후 어절 단위로 구분하고 각 어절로부터 단어를 추출한다. 추출한 단어에 대해서 품사가 명사인지를 확인하고, 명사 중에서 글자 수가 2 이상인 경우만 키워드에 포함시켰다.

```

List MorphAnalysis(List messageList)
{
    initialize keywordList;
    for each ( msg : messageList ) {
        List result = komoran.analyze(msg);
        for (List eojeolResult : result) {
            for (wordMorph : eojeolResult) {
                // Condition : Extract noun(N) only
                if (wordMorph.getSecond().charAt(0)=='N') {
                    strNoun = wordMorph.getFirst();
                    // Condition : Extract 2 letters word only
                    if (strNoun.length() >= 2)
                        keywordList.put(strNoun);
                }
            }
        }
    }
    return keywordList;
}
    
```

Fig. 6. Morphological analysis of sentences extracted from the SNS posts.

4.4 형태소 분석 결과에 대한 연관 검색어 수집

연관 검색어는 네이버 사이트에서 제공하는 기능으로서 특정 입력 키워드에 대한 연관된 검색어를 찾아주는 역할을 한다. 특정 키워드에 대해서 일반적으로 20개의 연관 검색어가 추출되며, 연관 검색어의 개수는 키워드에 따라서 변동될 수 있다.

Fig. 7은 네이버 사이트에서 “뉴스”를 검색했을 경우에 추출되는 연관 검색어를 보여준다.



Fig. 7. Related keyword searching using Naver site

형태소 분석 결과 추출된 각각의 키워드에 대해서 네이버 연관 검색어를 찾아서 추출 키워드 리스트에 추가한다. 각 키워드 당 연관 검색어가 최대 20개까지 검색될 수 있다. 따라서 연관 검색어 과정을 거치면서 키워드 개수의 20배에 해당하는 연관 키워드 리스트가 생성될 수 있다. 중복된 키워드가 추출되더라도 모두 결과 리스트에 포함하는데 그 이유는 키워드의 빈도수를 고려하여 중요도를 정할 수 있기 때문이다.

4.5 키워드로부터 빈도수 계산을 통한 필터링

추출된 키워드와 연관 검색어는 모두 리스트 형태로 저장되어 있고, 리스트에서 키워드를 한 개씩 꺼내서 빈도수와 함께 해시맵(HashMap)에 저장한다. 자바에서 해시맵은 키워드와 빈도수를 쌍으로 함께 저장할 수 있는 데이터 구조이다.

추출된 키워드는 최소 1 이상의 빈도수를 갖는다. 빈도수가 높을수록 의미있는 키워드라고 예상할 수 있기 때문에 빈도수 최솟값 기준을 적용하여 빈도수가 적은 단어들을 필터링할 수 있다. 필터링 작업은 해시맵 자료형에서 빈도수가 적은 단어들을 제거하는 형태로 이루어진다. 키워드 리스트로부터 빈도수를 구하고, 필터링 기준 값으로 키워드를 필터링하는 작업에 대한 코드는 Fig. 8과 같다.

```
HashMap FrequencyFiltering(List keywordList)
{
    initialize hashmap;
    // count frequency
    for each ( keyword : keywordList ) {
        count = hashmap.get(keyword);
        count++;
        hashmap.put(keyword, count);
    }
    // filtering by frequency
    for each( keyword : hashmap ) {
        count = hashmap.get(keyword);
        if ( count < frequencyLimit )
            hashmap.remove(keyword);
    }
    return hashmap;
}
```

Fig. 8. Counting frequency of keyword and filtering by frequency number.

마지막으로 키워드 별 빈도수 데이터를 데이터베이스에 저장한다. 데이터베이스는 MySQL을 사용하였고, 데이터베이스 테이블에 키워드를 저장하기 위해서 SQL 명령어 insert 문을 사용한다. 해당 코드는 Fig. 9와 같다.

```
void insertIntoMySqlDB(HashMap hashmap)
{
    conn = DriverManager.getConnection()
    stmt = conn.createStatement();
    strDate = new SimpleDateFormat("yyyy/MM/dd");
    strTime = new SimpleDateFormat("HH:mm:ss");
    for each( keyword : hashmap ) {
        strSQL = "insert into " + strTable +
            " (date, time, type, keyword, count) values("
            + strDate + "," + strTime + ","
            + strType + "," + (strKeyword) + ","
            + intCount + ");";
        stmt.executeUpdate(strSQL);
    }
}
```

Fig. 9. Insert keyword into MySQL database.

테이블에 필요한 데이터를 모두 저장한 후, 저장된 데이터의 확인을 위해서 select 명령어를 이용할 수 있다. 이때, 빈도수 기준으로 오름차순 또는 내림차순으로 검색하면 데이터의 빈도수 별 키워드 확인이 편리하다. 만일 내림차순으로 검색하면 데이터의 빈도수가 높은 키워드부터 확인할 수 있기 때문에 SNS에서 이슈화되고 있는 키워드를 간단하게 확인할 수 있다. Fig. 10은 MySQL 데이터베이스에 저장된 키워드의 예시이다.

1		2018/01/17		11:01:27		T		나영이 사건		3
2		2018/01/17		11:01:27		T		인육사건		3
3		2018/01/17		11:01:27		T		빅뱅 사건		3
4		2018/01/17		11:01:27		T		채벌사건		3
5		2018/01/17		11:01:27		T		세계10대사건		2
6		2018/01/17		11:01:27		T		알바		2
7		2018/01/17		11:01:27		T		행사사건		2
8		2018/01/17		11:01:27		T		오대양 사건		2
9		2018/01/17		11:01:27		T		고준희 사건		2
10		2018/01/17		11:01:27		T		사건 영어로		2
11		2018/01/17		11:01:27		T		광주사건		2
12		2018/01/17		11:01:27		T		토막사건		2
13		2018/01/17		11:01:27		T		일본콘크리트사건		2
14		2018/01/17		11:01:27		T		사고		2
15		2018/01/17		11:01:27		T		잔인한 사건		2
16		2018/01/17		11:01:27		T		오늘의 사건사고		2
17		2018/01/17		11:01:27		T		사건사고		1
18		2018/01/17		11:01:27		T		친안함 사건		1
19		2018/01/17		11:01:27		T		칠곡계모사건		1

Fig. 10. Example of keyword list from MySQL database.

5. 결론 및 향후 과제

본 논문에서는 업체의 매출을 일으키는 키워드의 조합이 '검색 사이트'보다 '소셜 사이트'에서 더 정확성을 보장할 수 있다는 데 착안하여 SNS 사이트 리스트로부터 이슈 키워드를 추출하고 데이터베이스로 관리하는 잠재적 광고 키워드 추출 시스템인 KAES를 설계 및 구현하였다.

소비자들이 상품 구매를 위해서 검색 사이트를 많이 사용하지만, 검색 사이트에서는 복수의 키워드를 사용하지 않거나, 단어 위주의 검색만을 한다. 그러나 소셜 사이트에서는 자신의 의도를 더 길게, 더 구체적인 문장으로 표현하기에 소비자의 구매 의사를 확인하는 데 효과적이다. 즉, 트위터, 페이스북에서 방대한 키워드를 수집하여 우선순위를 부여하고, 이를 네이버 등의 연관키워드 등과 비교하여 좀 더 정확한 키워드 마케팅을 수행할 수 있게 해줄 수 있다고 기대한다.

향후 과제는 다음과 같다. 본 논문을 통해서 구현된 KAES 시스템을 이용하여 실제 업체에서 판매하고자 하는 제품에 대한 키워드를 추출하는 과정을 진행할 예정이다. 업체와의 연계를 통해서 키워드 추출 시스템에서 추천하는 키워드로 광고를 집행하고 판매 실적을 조사할 예정이다.

또한 광고 기간 대비 매출 추이를 살펴보게 함으로써 관련 광고 키워드의 조합이 매출 대비 얼마나 효과적인지 시각적인 판단을 가능하게 할 예정이다. 이러한 과정을 거치면 업체들이 불필요한 광고를 줄일 수 있게 되며 적은 비용으로 효과적인 광고를 집행할 수 있을 것으로 기대한다.

REFERENCES

- [1] J. C. Choi.(2018). Big Data Patent Analysis Using Social Network Analysis. *Journal of the Korea Convergence Society*, 9(2), 251-257.
- [2] S. A. Jin, C. E. Heo, Y. K. Jeong & M. Song. (2013). Topic-Network based Topic Shift Detection on Twitter. *Journal of the Korean Society for Information Management*, 30(1), 285-302.
- [3] Y. H. Ha, S. W. Lim, & Y. H. Kim. (2012). Trend Analysis through Content-based Tweet Clustering. *Fall conference of Korea Information Science Society*, 39(2B).
- [4] M. Mathioudakis & N. Kousdas. (2010). Twitter Monitor : Trend Detection over the Twitter Stream, *ACM Special Interest Group on Management of Data*, 1155-1157.
- [5] DMC Media. (2016). Facebook Targeting Strategies and Success Stories, *Digital Media & Marketing Intelligence Center report*. <http://www.dmcreport.co.kr>
- [6] D. Y. Kim, G. G. Lim, & D. C. Lee. (2011). A Study on the Efficiency of internet Keyword Advertisement According to CPM and CPC Methods by Analyzing Transaction Data. *The Journal of Society for e-Business Studies*, 16(4), 139-154.
- [7] M. G. Kim, N. G. Kim. & I. H. Jung. (2014). A Methodology for Extracting Shopping-Related Keywords by Analyzing Internet Navigation Patterns. *Journal of Intelligence and Information Systems*, 20(2), 123-136.
- [8] S. J. Choi, M. Y. Son. & Y. H. Kim. (2016). Associated Keyword Recommendation System for Keyword-based Blog Marketing. *KIISE Transaction on Computing Practices*, 22(5). 246-251.
- [9] T. M. Cho. & J. H. Lee. (2015). Latent Keyphrase Extraction Using LDA Model. *Journal of Korean institute of Intelligent System* 25(2), 180-185.
- [10] W. You, D. Fontaine & J. P. Barthes. (2013). An Automatic keyphrase system for scientific documents. *Knowledge and information systems*, 34(3). 691-724.
- [11] Twitter 4J library for Twitter SNS Parsing. <http://twitter4j.org>
- [12] JSON simple library for Facebook SNS Parsing. <https://cliftonlabs.github.io/json-simple>
- [13] Korean morphological analyzer, Komoran. <http://www.shineware.co.kr/products/komoran>
- [14] MySQL Community Server 5.5. <https://dev.mysql.com/downloads/mysql/>
- [15] Java and database connector, Connector J 5.1. <https://dev.mysql.com/downloads/connector/j/>
- [16] Y. H. Chang. (2012). A Study on the Marketing Performance Using Social Media- Comparison between Portal Advertisement, Blog, and SNS Channel Characteristics and Performance. *Journal of Digital Convergence*, 10(8). 119-133.
- [17] S. B. Chang, & D. S. Yoom. (2018). The Effects of User Experience on Facebook Acceptance Behavior and Advertising Acceptance Behavior. *Journal of Digital Convergence*, 16(3). 169-179.

서 현 곤(Seo Hyun Gon)

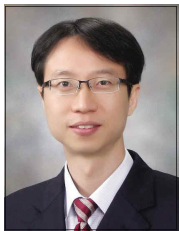
[정회원]



- 1992년 2월 : 경성대학교 컴퓨터과 학과(이학사)
- 1994년 2월 : 경성대학교 컴퓨터과 학과(이학석사)
- 2004년 2월 : 영남대학교 컴퓨터공학과(공학박사)
- 2000년 3월 ~ 2004년 2월 : 대구대학교 정보통신공학부 BK21 연구교수
- 2005년 3월 ~ 현재 : 한라대학교 정보통신소프트웨어학과 교수
- 관심분야 : IoT응용, 프로토콜, 빅데이터
- E-Mail : hgseo@halla.ac.kr

박 희 완(Park, Hee Wan)

[정회원]



- 1997년 2월 : 동국대학교 컴퓨터공학과(공학사)
- 1999년 2월 : KAIST 전산학과(공학석사)
- 2010년 1월 : KAIST 전산학과(공학박사)
- 2004년 3월~2007년 2월 : 삼성전자 무선사업부 책임연구원
- 2010년 2월~2011년 8월 : ETRI 부설연구소 선임연구원
- 2011년 9월~현재 : 한라대학교 정보통신소프트웨어학과 조교수
- 관심분야 : 소프트웨어 난독화, 정적 및 동적 분석
- E-Mail : heewanpark@halla.ac.kr