

포털사이트 실시간 검색키워드의 주간 핵심 이슈 선정 및 차이 분석

정민영

광주여자대학교 실버케어학과

Extracting week key issues and analyzing differences from realtime search keywords of portal sites

Min-Yeong Chong

Dept. of Silvercare, Kwangju Women's University

요 약 포털사이트의 실시간 검색키워드는 검색횟수의 순간증가율이 높은 순서대로 나타나므로 짧은 시간에 관심도가 급상승하는 이슈는 쉽게 보여주지만, 포털사이트별로 다른 결과가 도출되고 일정기간에 대한 이슈는 나타내지 못하는 한계가 있다. 따라서, 일정기간 동안의 전체 실시간 검색키워드에서 핵심 이슈를 찾고 각 포털사이트별로 집계한 결과와 이들의 차이를 분석한 결과를 보여주는 것은 이슈를 보다 실제적으로 이해할 수 있는 근거를 제공하고 자주 변화하는 실시간 검색키워드에 대한 일관성을 유지할 수 있게 해준다는 측면에서 의미가 있다. 이를 위해 본 논문에서는 대표적인 두 개의 포털사이트에서 제공하는 실시간 검색키워드의 주간 분석을 통하여 주간 핵심 이슈를 추출하고 이들의 차이를 분석한다. 두 포털사이트의 실시간 검색키워드 중요도 점수에 대한 독립표본 t검정과 실시간 검색키워드 생존함수에 의한 생존분석 결과, 두 포털사이트는 차이가 있다는 것을 보였다.

주제어 : 실시간 키워드, 주간 이슈, 생존분석, 웹 텍스트 마이닝, 빅데이터 분석

Abstract Since realtime search keywords of portal sites are arranged in descending order by instant increasing rates of search numbers, they easily show issues increasing in interests for a short time. But they have the limits extracted different results by portal sites and not shown issues by a period. Thus, to find key issues from the whole realtime search keywords for certain period, and to show results of summarizing them and analyzing differences, is significant in providing the basis of understanding issues more practically and in maintaining consistency of them. This paper analyzes differences of week key issues extracted from week analysis of realtime search keywords provided by two typical portal sites. The results of experiments show that the portal group means of realtime search keywords by the independent t-test and the survival functions of realtime search keywords by the survival analysis are statistically significant differences.

Key Words : Realtime keywords, Week issues, Survival analysis, Web text mining, Bigdata analysis

* 본 논문은 2016 학년도 광주여자대학교 교내연구비 지원에 의하여 연구되었음(KWUI16-017).

Received 1 November 2016, Revised 1 December 2016

Accepted 20 December 2016, Published 28 December 2016

Corresponding Author: Min-Yeong Chong

(Kwangju Women's University)

Email: mychong@kwu.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

4차 산업혁명을 불러일으키는 핵심 기술인 지능정보 기술은 모든 것을 연결하고자 하는 만물인터넷, 모든 데이터를 분석하여 고차원적인 지능 정보를 도출하고자 하는 빅데이터 기술을 근간으로 일상생활 속에 무겁고 신속하게 다가오고 있는 가운데, 빅데이터 분석 기술 중 비정형 웹 텍스트 마이닝을 통한 지능 정보 활용에 대한 관심이 높아지고 있다[1,2,3,4,5,6,7,8,9]. 특히 대표적인 포털 사이트들의 키워드 기반 실시간 검색어 서비스는 네티즌의 관심사와 해결하고자 하는 문제, 검색하고자 하는 새롭고 유용한 정보에 대한 검색 키워드를 통하여 이슈와 트렌드는 물론 사용자 맞춤형 필요 정보와 요구사항을 읽어내어 새로운 비즈니스 기회를 창출하는 원천으로 활용될 수 있다는 측면에서 포털사이트의 얼굴과 같은 핵심 서비스로서 자리를 잡고 있다. 현재 거의 대부분 포털 사이트에서 네티즌의 관심도가 높은 실시간 검색키워드를 상위 10위(Top10)까지 보여주며 관심링크를 선택할 때 나오는 매 페이지마다 분야별이나 유형별로 분류한 실시간 검색키워드를 보여주거나, 아예 독립된 홈페이지에 다양한 이슈와 트렌드를 함께 제공하는 등 각 포털 사이트마다 특색 있는 서비스를 제공하여 네티즌의 접근성을 높이기 위한 노력과 경쟁이 치열하게 이루어지고 있다[10,11,12,13,14,15].

특히 네이버(Naver)의 실시간 급상승 검색어와 다음(Daum)의 실시간 이슈 검색어는 대한민국의 대표적인 실시간 검색키워드로서 검색바를 통해 입력되는 검색키워드의 입력횟수를 근거로 하여 순간 증가율이 높은 키워드 순으로 첫 페이지에 나타냄으로써 네티즌들이 지금 이 순간의 관심도가 가장 높다고 평가되는 키워드를 알게 하여 정보 검색의 출발점으로서의 포털사이트의 역할을 충실하게 하는 요소로 활용하고 있다. 하지만 순간적으로 급상승하는 명사 중심의 단어에 국한되고, 포털 사이트마다 다른 정책에 기인한 다른 선정기준에 의해 상위 실시간 검색어가 결정되어 이들 사이의 차이가 생길 수밖에 없으므로 네티즌의 선호 포털사이트에 따라 관심도가 높은 실시간 검색키워드의 순위에 대한 편차도 발생한다[16,17].

이에 본 논문에서는 단시간에 급변하는 실시간 검색키워드의 일시성을 극복하기 위해 주간 핵심 이슈들을

추출함과 동시에 포털사이트에 따른 실시간 검색키워드의 편차를 최소화시키기 위해 네이버와 다음(Daum)의 실시간 검색키워드 원시자료를 일단 각각 수집하여 분석하고 다시 이들을 병합하여 분석함으로써 통합된 주간 핵심이슈를 도출하고, 실시간 검색키워드에 대한 생존분석, 독립표본 t-검정 등의 차이분석을 통하여 그 생존 확률과 평균의 차이정도를 파악하여 통합 자료에 의해 도출된 주간 핵심 이슈의 의미를 보다 분명히 하고자 한다.

이를 위해 먼저 자료 수집 단계에서 네이버와 다음(Daum)의 실시간 검색어 상위 10개의 순위와 검색증가율을 고려하여 점수를 주어 각각 파일로 저장한다. 그 다음 분석 단계에서 저장된 각각의 파일을 읽어서 행 중심으로 하나로 묶고 기술통계 분석을 통해 실시간 검색어별 점수를 집계하고 합산된 점수의 내림차순으로 정렬하여, 포털사이트별 주간 이슈와 이들의 통합 주간 핵심 이슈를 추출하고 차이분석을 실시한다. 분석 대상은 네이버의 실시간 급상승 검색어와 다음(Daum)의 실시간 이슈 검색어이며, R 언어를 사용하여 자료수집과 분석을 실시한다.

2. 실시간 검색키워드의 자료 수집

2.1 네이버와 다음의 실시간 검색키워드 특성

실시간 검색어키워드를 네이버(Naver)에서는 '실시간 급상승 검색어'라고 말하고 다음(Daum)에서는 '실시간 이슈 검색어'라고 말한다. 먼저, 네이버(Naver)의 실시간 급상승 검색어는 단위 시간 동안 네이버 검색바로 입력되는 검색키워드를 분석해 입력횟수의 증가율이 가장 큰 검색어 순서대로 보여주는 서비스로, 사용자가 특정 기준 시간 내에 키워드를 검색바에 집중적으로 입력하여 과거 시점이나 다른 검색키워드에 비해 상대적으로 순위가 급격하게 상승한 비율을 기준으로 순위를 선정한다. 실시간 급상승 검색어는 '실시간'이라는 서비스 특성에 따라 매우 짧은 시간인 1분 이내에 검색어 순위가 수차례 새롭게 선정된다. 또한 검색어가 노출 제외 기준을 통과하여도 동일인이 특정 기준 시간 동안 같은 검색어를 두 번 이상 입력할 경우, 한 번 입력한 것과 동일하게 계산하고, 실시간 차트에 이미 노출되고 있는 검색어를 클릭한 경우는 검색횟수에 포함되지 않으며, 검색바에 직

접 입력되거나 혹은 자동완성된 검색키워드만 집계에 포함시키는 제약사항이 있다[17,18].

다음(Daum)의 실시간 이슈 검색어는 검색바로 입력한 검색키워드들을 매분 단위로 분석하고 실시간 알고리즘을 이용하여 검색키워드 입력횟수의 증가폭 비율이 가장 큰 검색어를 순위로 제공하는 서비스로서, 사용자의 관심이 집중된 검색키워드가 실시간으로 변하기 때문에 실시간 이슈 검색어는 계속 바뀔 수 있다. 사용자들이 검색한 모든 검색키워드를 자동으로 수집한 후 필터링 과정(스팸, 성인키워드 등 금칙어)을 거쳐 급등 검색어 추출 기준에 따라 제공되는 것으로, 가장 많이 검색된 키워드를 보여주는 방식이 아니며, 자동 순위 시스템에 따라 순간적으로 많이 검색된 키워드를 추출해 보여주므로 특정 시간대에 보였던 검색키워드라도 이용자의 검색 빈도가 높지 않거나 다른 검색키워드의 검색 빈도가 더 높으면 순위가 내려갈 수 있다[17].

이상과 같은 특징으로 볼 때, 네이버(Naver)와 다음(Daum)의 실시간 검색키워드는 집계에 포함되지 않는 검색키워드, 즉 검색키워드 노출 제외 기준이나 자동 필터링 과정에서 걸러내는 검색키워드들에 대한 알고리즘의 차이가 있으나 검색바로 입력되는 키워드의 입력횟수 증가율을 기준으로 상위 10개의 키워드를 제시하면서 바로 우측에 네티즌의 상대적 관심도에 해당되는 검색증가율을 숫자와 함께 상승화살표를 표시하는 것과, 만약 이전 데이터가 없어서 검색증가율을 구할 수 없는 신규 키워드일 경우에 'NEW'를 표시하는 것은 거의 차이가 없다[16,17,18].

그러므로 두 포털사이트의 실시간 검색키워드는 공히 키워드 입력횟수만으로 집계하는 종합검색어와는 다른 성격을 가지며, 입력횟수 증가율로 평가하는 특징으로 인해 순간의 사용자 관심도가 높은 것 위주로 순위가 결정되는 한계, 그리고 노출 제외어와 금칙어 등과 같은 기준을 적용할 수 밖에 없는 한계가 분명히 존재한다.

따라서 이러한 한계를 극복할 수 있는 방안으로 두 사이트의 실시간 검색키워드를 각각 점수화한 데이터를 1분 간격으로 저장해두고, 분석하고자 하는 범위에 따라 1일, 1주일, 1달, 1년 등의 기간 동안 데이터를 읽어서 검색어 중심의 집계와 정렬을 통해 각 기간 동안의 두 포털사이트 각각의 핵심 이슈와 함께 통합 핵심 이슈를 제시하고 두 포털 사이트 검색어의 독립표본 t-검정과 생존

분석을 통한 차이분석을 통해 한계를 극복할 수 있는 방법을 제시하고자 한다. 실제로 본 논문에서는 주간 단위의 데이터를 수집하고 이를 분석함으로써 일간 단위의 분석결과를 내포하고, 월간, 년간 분석 결과를 바로 수행할 수 있는 틀을 제공하고자 한다.

2.2 실시간 검색키워드 수집

네이버(Naver)와 다음(Daum)의 실시간 검색키워드는 1분 단위로 각 홈페이지 원시코드를 읽어서, 일단 네이버의 경우 '실시간 급상승 검색어'를, 다음(Daum)의 경우 '실시간 이슈'가 있는 위치를 R언어의 문자열 함수를 이용하여 찾고, 이 위치를 기준으로 순위별로 제시된 10개의 검색키워드, 순위, 상승상태, 상대적 관심도를 R언어의 정규식 처리 함수를 이용하여 추출한다[19]. 그리고 이를 기반으로 실시간 검색키워드 점수화 알고리즘[20]에 의해 실시간 검색키워드의 중요도 점수를 계산하고 여기에 확보된 레코드 단위의 자료를 파일에 추가하여 저장한다.

3. 주간 핵심 이슈 선정

3.1 주간 핵심 이슈 선정 과정

먼저, 자료 수집 단계에서 저장된 두 포털사이트의 실시간 검색키워드 자료 파일을 각각 별도로 읽어서, 네이버는 네이버 검색키워드 테이블을, 다음(Daum)은 다음 검색키워드 테이블을 만든다. 이 때 검색키워드 사이의 공백은 제거하고 자료형은 'factor'형으로 변환시킨다. 그 다음, 각 검색키워드 테이블별로 검색키워드를 기준으로 그룹화하여 중요도 점수의 케이스수, 합계, 평균을 구하고 합계를 키 항목으로 하여 정렬한 다음, 상위 10개의 주간 이슈 검색어를 각각 추출한다.

그리고 네이버 검색키워드 테이블과 다음(Daum) 검색키워드 테이블을 행단위로 병합하여 통합 검색키워드 테이블을 만든다. 이어서 검색키워드를 기준으로 그룹화하여 중요도 점수의 케이스수, 합계, 평균을 구하고 합계를 키 항목으로 하여 정렬한 다음, 상위 10개의 주간 핵심 이슈 검색어를 각각 추출한다. 통합 검색키워드 테이블에서 추출한 주간 핵심 이슈 검색어는 네이버와 다음(Daum) 검색키워드 테이블에서 추출한 주간 이슈 검색

어의 차이를 완화시키는 통합된 결과를 보인다. 실시간 검색키워드 자료는 2016. 8. 15일 0시 0분 0초에서 2016. 8. 21일 23시59분59초까지 1분 간격으로 수집한 것이다.

3.2 '주간 핵심 이슈 추출

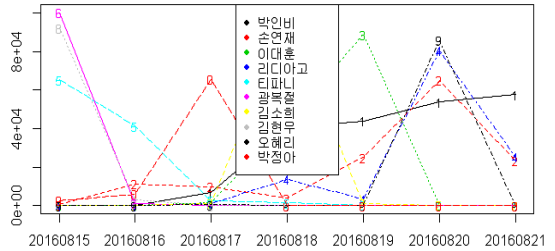
'주간 이슈 검색어'는 네이버와 다음(Daum)의 상위 10개의 검색키워드를 일주일 동안 별도로 누적시킨 것을 대상으로 하며, 검색키워드를 기준으로 집계한 주간 누적 점수합계를 구하고, 점수합계 내림차순으로 정렬하여 추출한 상위 10개의 검색키워드이므로 일주일 동안 가장 주목받은 검색키워드에 해당된다.

'주간 핵심 이슈 검색어'는 네이버와 다음(Daum)의 상위 10개의 검색키워드를 일주일 동안 별도로 누적시킨 것을 통합한 것을 대상으로 하며, 검색키워드를 기준으로 집계한 주간 누적 점수합계를 구하고 점수합계 내림차순으로 정렬하여 추출한 상위 10개의 통합 검색키워드이므로 일주일 동안 가장 주목받은 통합 검색키워드에 해당된다. 따라서 주간 핵심 이슈는 두 포털사이트의 주간 이슈 검색어를 통합한 것이므로 보다 객관적으로 관심도가 높은 주간 이슈라고 할 수 있다.

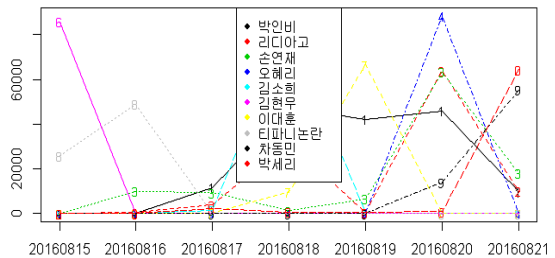
1분 간격의 1위에서 10위까지 10개씩의 데이터를 수집하면 총 100,800(=7*24*60*10)개가 되어야 하지만 시간별 인터넷 속도의 차이로 인해 정확한 1분 간격을 맞추지 못하는 문제로 누락된 것을 제외하고 실제로 수집된 데이터 수는 네이버와 다음(Daum) 둘 다 각각 100,540개의 데이터가 수집되었다. 그리고 이들을 키워드를 그룹으로 각각 집계한 결과 네이버는 855개, 다음(Daum)은 1454개의 서로 다른 키워드가 발견되었다. 그리고 네이버와 다음(Daum) 자료를 행으로 병합한 201,080(=100540*2)개의 데이터를 키워드를 그룹으로 집계한 결과, 1,761개의 서로 다른 키워드가 발견되었다. 이는 네이버와 다음(Daum)의 키워드가 통합될 때 공통적으로 포함된 중복 키워드가 하나로 취급되기 때문에 그만큼 숫자가 줄어든 것이다. 따라서 중복 키워드 수는 548(=855+1454-1761)개가 된다.

[Fig. 1]은 네이버(1), 다음(2), 통합(3) 주간 이슈들 중에서 상위 10개의 실시간키워드들의 일주일간 날짜별 관심도의 변화를 나타낸 것이다. 네이버와 다음의 1위 주간 이슈는 둘 다 통합 1위 주간 핵심 이슈인 '박인비'로 나타났고, 네이버 2위 주간 이슈 '손연재'가 통합 2위 주간 핵

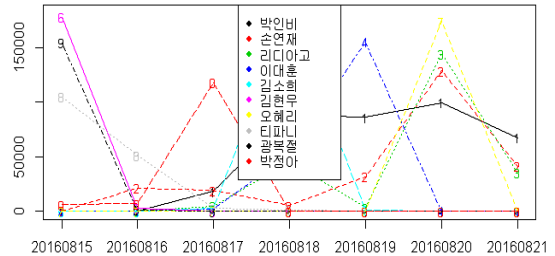
심 이슈가 되고, 다음(Daum) 2위 주간 이슈 '리디아고'가 통합 3위 주간 핵심 이슈가 되었음을 확인할 수 있다.



(1) Naver issues



(2) Daum issues



(3) integrated issues

[Fig. 1] Week top 10 issues

또한, 통합 순위와 네이버순위, 그리고 통합순위와 다음순위의 차의 합을 구한 결과 각각 14와 25가 되어서 네이버와의 차가 더 작으므로 통합순위에 더 가까운 것으로 나타났다.

4. 실시간키워드 차이 분석

4.1 실시간키워드의 통계적 특성

<Table 1>은 실시간키워드의 통계적 특성을 나타낸 것으로 세로 항목은 키워드 그룹에 대한 케이스수(count), 누계(sum), 평균(mean)이고, 가로 항목은 포털

사이트의 유형으로서 세로 항목에 대한 (평균±표준편차)를 나타내고, 아울러 각 세로 항목에 대한 독립표본 t-검정의 결과로 계산되는 t값과 유의도(p)를 나타낸다.

<Table 1> Statistical characteristics of keywords

Type Group	Naver (mean±SD)	Daum (mean±SD)	t	p
count	117.59±282.87	69.15±178.36	4.508	0.000
sum	5874.59±16058.54	3523.05±10215.75	3.848	0.000
mean	45.39±24.39	40.65±27.25	4.200	0.000

<Table 1>에 의하면 포털사이트 종류(Type)와 포털 키워드 케이스수 사이에 95% 신뢰수준에서 통계적으로 유의한 차이가 있는 것으로 나타났으며(t=4.508, p=0.000<0.05), 네이버의 평균이 더 높은 것으로 나타났다. 케이스수가 많은 것은 일주일 동안 지속적인 관심을 받았다는 의미하므로, 네이버의 키워드들이 더 지속적인 관심을 받은 것으로 해석할 수 있다.

또한, 포털사이트 종류(Type)와 포털 키워드 누적점수 사이에 95% 신뢰수준에서 통계적으로 유의한 차이가 있는 것으로 나타났으며(t=3.848, p=0.000<0.05), 네이버의 평균이 더 높은 것으로 나타났다. 누적점수가 많은 것은 일주일 동안 지속적으로 많은 관심을 받았다는 의미하므로, 네이버의 키워드들이 지속적으로 더 많은 관심을 받은 것으로 나타났다.

포털사이트 종류(Type)와 포털사이트 키워드 평균점수 사이에 95% 신뢰수준에서 통계적으로 유의한 차이가 있는 것으로 나타났으며(t=4.200, p=0.000<0.05), 네이버의 평균이 더 높은 것으로 나타났다. 평균점수가 많은 것은 일주일 중 어느 한 순간 많은 관심을 받았다는 의미하므로, 네이버의 키워드들이 일주일 중 어느 한 순간 더 많은 관심을 받은 것으로 나타났다.

4.2 실시간키워드의 생존분석에 의한 차이

생존분석이란 어떤 사건이 발생할 때까지의 생존시간에 대한 확률을 생존데이터에서 분석해내는 통계기법이다. 생존데이터에 포함되는 생존시간은 대부분 정규분포가 아니고 연구중단과 같은 중도중단(censoring)을 고려해야 하며 사건은 임의로 발생하는 것을 전제로 한다. 본문에서는 각 포털사이트의 실시간키워드에 대한 기본

자료에서 실시간키워드-일자를 그룹으로 집계하고 여기서 실시간키워드별 최대일자를 구해두고, 집계된 것을 다시 실시간키워드를 그룹로 집계하면서 내포되었던 실시간키워드별 일자의 케이스수를 구하여 이를 생존시간으로 정한다. 이전 집계자료 생성시 구해놓은 최대일자를 방금 집계한 자료의 최종일자로 정하고 아울러 최종일자가 연구종료일과 일치하면 이벤트 항목을 중도중단(0)으로, 그렇지 않은 경우는 소멸(1)로 정하여 생존자료를 만든다. 그리고 두 포털사이트의 생존자료를 행 병합으로 묶어서 통합 생존자료를 만든 다음, 이를 통해서 Kaplan-Meier Method[21,22]에 의해 생존함수를 추정하고 Log-Rank Test[21,22]에 의해 두 생존함수의 동일성 여부를 검정하여 차이를 분석한다.

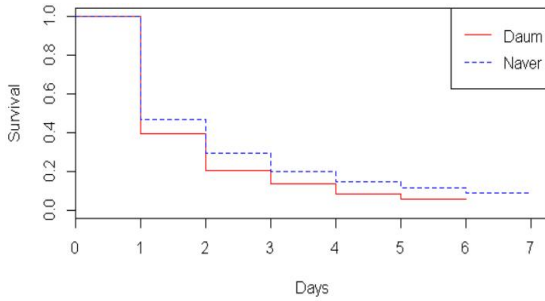
생존함수는 $S(t) = Prob(T > t)$ 와 같이 정의되는데, 생존시간을 나타내는 확률변수 T가 특정시점 t시간 초과 생존할 확률을 구하는 것을 나타낸다. 생존함수 추정은 주어진 생존자료에 Kaplan-Meier Method를 적용하여 수행하는데 추정식은 다음과 같다.

$$S(t_j) = \prod_{i=1}^j \left(1 - \frac{d_i}{n_i}\right)$$

여기서 t_j 는 j번째 키워드 소멸 발생시간, d_i 는 t_i 시점에서 소멸키워드 수, n_i 는 t_i 시점 바로 직전의 위험집합수를 말한다.

두 포털사이트 그룹의 생존함수를 Kaplan-Meier Method로 추정된 뒤, 두 그룹간 생존함수들의 동일성 검정을 통한 차이분석을 수행하게 되는데 이것은 Log-Rank Test에 의해 검정한다. 여기서 귀무가설(H_0)은 ‘모든 t에 대해서 두 생존함수는 동일하다.’이고 대립가설(H_1)은 ‘어떤 t에 대해서 두 생존함수는 동일하지 않다.’이다.

실험 결과, 두 포털 사이트 실시간키워드의 생존함수에 따른 생존확률이 95% 신뢰수준에서 통계적으로 유의한 차이가 있는 것으로 나타났으며(chisq=18.7, p= 1.53e-55<0.005; 귀무가설 기각, 대립가설 채택), 네이버 실시간키워드의 생존확률이 높은 것으로 나타났다. 이는 실시간 키워드의 생존시간이 클 확률이 높은 것을 나타내므로 그만큼 해당 포털사이트에 대한 선호도가 높은 것으로 해석할 수 있다. [Fig. 2]는 이러한 차이를 시각적으로 보여주는 Kaplan-Meier 그래프를 나타낸 것이다.



[Fig. 2] Kaplan-Meier graph by type of portal sites

5. 결론

본 논문에서는 검색횟수 순간증가율에 따라 제시하는 실시간 검색키워드의 일시성을 극복하고, 포털사이트에 따라 다른 결과를 보이는 실시간 검색키워드의 편차를 최소화시키기 위해 네이버와 다음(Daum)의 실시간 검색어 원시자료를 각각 수집하여 분석함과 동시에, 이들을 병합하여 분석을 함으로써 통합된 주간 핵심이슈를 도출하였고, 실시간 검색키워드에 대한 생존 확률과 평균의 차이정도를 파악할 수 있는 실제적인 예를 제시하였다.

본 논문에서 제안하는 방법은 두 개의 대표적인 포털 사이트에서 수집한 실시간 검색키워드를 병합하여 통합 실시간 검색키워드를 제시함으로써 보다 의미 있는 주간 핵심 이슈로 활용할 수 있게 하였으며, 이를 뒷받침하기 위해 두 포털 사이트의 실시간 검색키워드에 대해 독립 표본 t검정, 생존분석 등을 통해 차이를 분석한 실제 예를 제시함으로써 통합 실시간 검색의 필요성을 보다 분명하게 밝혔다는 측면과 특히 생존 분석의 동질성 검정을 통해 생존확률의 차이를 보고 포털 사이트에 대한 네티즌의 선호도를 평가할 수 있다는 측면에서 의미가 있다.

하지만 실시간 검색어를 수집하는 과정에서 인터넷 속도의 영향으로 인한 오차와 포털사이트별 실시간 검색어 노출제의 기준과 선정기준의 차이로 인한 오차 등이 존재하고, 네이버와 다음(Daum) 사이트에 국한된 일주일간의 데이터를 기반으로한 분석 결과라는 한계가 존재한다. 앞으로 이를 개선하기 위해 데이터 수집기간을 보다 더 늘리고, 연구에 포함시킬 포털사이트의 수를 늘리는 추가적인 노력과 연구가 필요하다.

ACKNOWLEDGMENTS

This work was supported by Research Funds of Kwangju Women's University in 2016(KWUI16-017).

REFERENCES

- [1] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding, "Data mining with big data" IEEE transactions on knowledge and data engineering, Vol. 26, No. 1, pp. 97-107, 2014
- [2] Hsinchun Chen, Roger HL Chiang, and Veda C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact", MIS quarterly Vol. 36, No. 4, pp. 1165-1188, 2012
- [3] Miner G, Elder J, Hill T, Nisbet R, Delen D, and Fast A, "Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications", p.1000, Academic Press, 2012
- [4] Kyoo-Sung Noh, "A Exploratory Study on Big-data based Election Campaign Strategy Model in South Korea", Journal of Digital Convergence, Vol. 11, No. 12, pp. 113-120, 2013
- [5] Soonduck Yoo, Kwangdon Choi, and Sungyoung Shin, "Characterizing Business Strategy in a New Ecosystem of Big Data", Journal of Digital Convergence, Vol. 12, No. 4, pp. 1-9, 2014
- [6] Su-Hyeon Namn, "Knowledge Creation Structure of Big Data Research Domain", Journal of Digital Convergence, Vol. 13, No. 9, pp. 129-136, 2015
- [7] Shinkon Kim, Sukjun Lee, and JeonggonA Kim, "Study on the Development of Phased Big Data Distribution Model Based on Big Data Distribution Ecology", Journal of Digital Convergence, Vol. 14, No. 5, pp. 95-106, 2016
- [8] Seong-Hoon Lee and Dong-Woo Lee, "Current Status of Big Data Utilization", Journal of Digital Convergence, Vol. 11, No. 2, pp. 229-233, 2013
- [9] Kyoung-Ho Choi and Jeong-Hye Park, "The Analysis of Public Awareness about Literary

- Therapy by Utilizing Big Data Analysis - The aspects of convergence literature and statistics", Journal of Digital Convergence, Vol. 13, No. 4, pp. 395-404, 2015
- [10] Min-Gu Song and Sun-Bae Kim, "A Study of improving reliability on prediction model by analyzing method Big data", Journal of Digital Convergence, Vol. 11, No. 6, pp. 103-112, 2013
- [11] Xiao Fang and Olivia R. Liu Sheng, "Designing a better web portal for digital government: a web-mining based approach", Proceedings of the 2005 national conference on Digital government research. Digital Government Society of North America, pp. 277-278, 2005
- [12] Simon Dennis, Peter Bruza, and Robert McArthur, "Web Searching: A Process-Oriented Experimental Study of Three Interactive Search Paradigms", Journal of the American Society for Information Science and Technology, Vol. 53, No. 2, pp. 120-133, 2002
- [13] Bing Liu, "Sentiment analysis and opinion mining", p.168, Morgan & Claypool Publishers, 2012
- [14] Matthew A. Russell, "Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More", p.411, O'Reilly Media, Inc., 2013
- [15] Hyeong-Sik Yang and Sun-Bae Kim, "Evaluation Method of Big Data Efficiency", Journal of Digital Convergence, Vol. 11, No. 8, pp. 31-39, 2013
- [16] Naver Search Help, "Realtime hot searches", <https://help.naver.com/support/service/main.nhn?serviceNo=606&categoryNo=1989>, 2015
- [17] Daum Search Help, "Realtime hot issues" <http://cs.daum.net/faq/15/14957.html#28971>, 2016
- [18] KISO Validation Committee, "The fourth validation report about realtime hot searches of Naver", 2015
- [19] Jon Starkweather, "Introduction to basic Text Mining in R", p.10, University of North Texas, 2014
- [20] Min-Yeong Chong, "Selecting a key issue through association analysis of realtime search words", Journal of Digital Convergence, Vol. 13, No. 12, pp. 161-169, 2015
- [21] Rupert G. Miller, "Survival analysis-2nd Edition", p.238, John Wiley & Sons, 2011
- [22] Manish Kumar Goel, Pardeep Khanna, and Jugal Kishore, "Understanding survival analysis: Kaplan-Meier estimate", International journal of Ayurveda research, Vol. 1, No. 4, pp. 212-216, 2010

정민영(Chong, Min Yeong)



- 1991년 2월 : 숭실대학교 전자계산학과(공학사)
- 1993년 2월 : 숭실대학교 전자계산학과(공학석사)
- 2004년 8월 : 전남대학교 컴퓨터정보통신공학과(공학박사)
- 1996년 3월 ~ 현재 : 광주여자대학교 실버케어학과 교수

- 관심분야 : 빅데이터분석, 소프트웨어공학, 컴퓨터응용
- E-Mail : mychong@kwu.ac.kr