

소셜미디어에서 관찰되는 *Helicobacter pylori*에 대한 정보의 특성 및 전달방식 분석

방창석, 백광호

한림대학교 의과대학 내과학교실

Social Network Analysis: Type of Information and Characteristics of Its Delivery Regarding *Helicobacter pylori* in Korea

Chang Seok Bang, Gwang Ho Baik

Department of Internal Medicine, Hallym University College of Medicine, Chuncheon, Korea

Background/Aims: Current production of knowledge and distribution of data is frequently mediated by social media in the form of unstructured text data rather than classical newspaper or journal format. Management of *Helicobacter pylori* infection is an important issue in Korea. However, there has been no study on the characteristics of information regarding *H. pylori* consumed and distributed in the social media by general population.

Materials and Methods: This study evaluated characteristics of social media data regarding *H. pylori* in Korea using semantic network analysis with Textom and NodeXL Pro (2015.8.11~2016.8.11, Naver, Daum, Youtube, and Twitter searching).

Results: Total collected data in the Textom analysis about '헬리코박터' was 10,061. TF-IDF (term frequency × inverse document frequency) was as follows: '위암', '위', '감염', '월', '균', '치료', '위궤양', '원인', '프로젝트', '음식', '세균', '검사', '예방', '증상', '건강', '약', '효과'. In the Twitter analysis, 96 nodes and 120 edges, and 86 unique edges were detected, suggesting low distribution of data. Misinformed data such as '헬리코박터-파이로리', '라는-바이러스때문에', '헬리코박터-프로젝트', '위궤양이나-위', '점막-출혈을', '비롯하여-각종', '부종을-억제하고', '치료하는데-효과가', '있습니다-또한', '장티푸스-이질', '등-전염성', '세균이나-장', '속의-세균들을', '억제하는-효과가', '프로젝트-월' were found. The most viewed movie in the Youtube analysis was '한국야쿠르트 헬리코박터 프로젝트 월 그 한입을 경계하라_15초' (210,577).

Conclusions: Relative insufficiency, low distribution, and some misinformed data were found in some part of social media. Participation of healthcare professional is warranted to solve these issues.

(Korean J *Helicobacter* Up Gastrointest Res 2016;16:198-203)

Key Words: Social network analysis; *Helicobacter pylori*; NodeXL

서론

인터넷과 스마트기기의 활용이 늘어남에 따라 자신의 생각을 표현하고 소통하는 수단으로서 소셜미디어(social media)의 활용이 일상화되고 있다.¹ 소셜미디어에서 발생하는 대량의 정보는 그 형식이나 용어, 전파 방식, 정확성, 생산자, 소비자 등이 매우 다양하여 통일된 개념으로 정의할 수 없다. 때로는 긍정적인 소통과 빠른 정보 전달, 사회참여 확대의 역할을 수행하기도 하지만, 잘못된 정보 전달로 인한 지식의 왜곡이나 이로 인한 현실의 갈등을 조장하기도 한다.^{1,2} 소셜미디어에서 유통되는 정

보는 대중의 관심을 실시간으로 그리고 직접적으로 반영한다는 특성을 갖고 있기 때문에 그 분석에 가치가 있다.

대한상부위장관·헬리코박터학회는 헬리코박터균 감염과 상부위장관 질환 관련 연구를 선도하고 있으며 학술 교류는 물론 국민보건증진에 이바지하고자 노력해왔다. 하지만 일반 대중의 시각에서 헬리코박터 파일로리(*Helicobacter pylori*)에 대한 정보가 어떤 방식으로 받아들여지고 있으며, 정보의 왜곡은 없는지, 그 전달 유형은 어떠한지에 대해서는 연구된 바가 없다.

이에 본 연구에서는 소셜미디어에서 소비되고 유통되는 비정형 텍스트 정보를 분석하여 헬리코박터 파일로리에 대한 정보의 특성을 파악하고 그 전달방식을 분석하고자 한다.

Received: September 18, 2016 Accepted: October 21, 2016

Corresponding author: Chang Seok Bang
Department of Internal Medicine, Chuncheon Sacred Heart Hospital, Hallym University College of Medicine, 77 Sakju-ro, Chuncheon 24253, Korea
Tel: +82-33-240-5821, Fax: +82-33-241-8064, E-mail: csbang@hallym.ac.kr

대상 및 방법

자료 수집 및 분석을 위해 텍스톰(Textom; The IMC Inc., Daegu, Korea, <http://www.textom.co.kr>)과³ NodeXL Pro (Microsoft, <http://nodexl.codeplex.com>)⁴을 이용하여 언어네트워크분석(semantic network analysis)^{5,6}을 실시하였다. 언어네트워크분석은 사회네트워크분석(social network analysis)의 한 방법으로, 단어나 문장들 간의 조합이 형성하는 특별한 의미 체계를 확인하기 위해 각 단어별로 의미 네트워크 행렬을 만들어 시각화한 분석법이다.⁷

정보의 특성을 분석하기 위해 ‘헬리코박터 파일로리’ 대신 일반 대중에게 좀 더 익숙한 ‘헬리코박터’로 검색 키워드를 정하였다. 텍스톰에 ‘헬리코박터’로 검색하여 텍스트 마이닝(text mining)을 통해 정보를 수집 및 정제한 이후에 시각화(visualization)하여 언어네트워크를 분석하였다. 정보 전달 방식을 확인하기 위해 NodeXL Pro를 이용하여 ‘헬리코박터’로 검색하여 소셜미디어에 등장하는 단어들 간의 연결 구조, 연결 수(degree), 중심성(centrality), 중심이 되는 노드(node, vertex) 및 관계를 시각화하였다.

텍스톰 분석은 2015년 8월 11일부터 2016년 8월 11일까지 1년간 네이버(www.naver.com) 검색엔진으로 ‘헬리코박터’ 키워드를 이용하여 검색한 웹문서, 블로그, 카페, 뉴스, 지식인, 이미지, 동영상, 전문정보, 학술자료, 연구보고서, 특허/KS 통계 리포트, 국가 기록물을 조사하였다. 또한 다음(www.daum.net) 검색엔진으로 같은 키워드를 이용하여 검색한 웹문서, 블로그, 카페, 뉴스, 지식인, 이미지, 동영상을 조사하였으며, 페이스북(www.facebook.com), 유튜브(www.youtube.com), 그리고 트위터(www.twitter.com)에서 ‘헬리코박터’로 검색하여 얻은 텍스트 정보를 분석하였다.³

NodeXL Pro 분석은 유튜브와 트위터에서 각각 ‘헬리코박터’로 검색하여 얻은 정보를 분석하였다. 페이스북의 경우 특정 Fan Page 분석은 가능하나 트위터의 Search Network 분석이나 유튜브의 Video Network 분석처럼 각 사용자 간의 관계와 연결구조를 알 수 없기 때문에 NodeXL Pro 분석에서 제외하였다. 트위터는 Twitter Search Network 분석을 통해 18,000개의 트윗(tweet)을 분석하였고 유튜브는 Youtube Video Network 분석을 통해 500개의 동영상을 분석하였다.^{4,5}

결 과

텍스톰을 이용해 검색한 ‘헬리코박터’ 키워드 관련 1년간의 정보량은 총 511,545건이었으며, 본 분석에 사용된 해당 키워드 관련 수집량은 10,061건이었다(Table 1). 추출 단어 빈도수

와 서로 연결되어 나타난 단어 세트의 빈도수(N GRAM 빈도수),³ 그리고 TF-IDF (term frequency×inverse document frequency; 단어빈도×역문서빈도; 수집단어의 중요도 표준화 값)³는 Table 2에 표시하였다. 헬리코박터 파일로리균의 이름 관련 단어를 제외하면 ‘위암’, ‘위’, ‘감염’, ‘위염’, ‘치료’, ‘월’, ‘균’, ‘위궤양’, ‘원인’, ‘음식’, ‘프로젝트’, ‘세균’, ‘예방’, ‘검사’, ‘증상’, ‘효과’, ‘건강’, ‘약’ 등의 순서로 단어 출현빈도가 확인되었다. N GRAM 빈도수에서는 헬리코박터 파일로리균의 이름 관련 조합을 제외하면 ‘헬리코박터-프로젝트’, ‘프로젝트-월’, ‘헬리코박터균-감염’, ‘위염-위궤양’, ‘위-점막’, ‘헬리코박터균-치료’, ‘위암-발생’, ‘원인-헬리코박터균’, ‘발명-헬리코박터’, ‘헬리코박터균-증상’ 등의 순서로 단어 조합 출현빈도가 확인되었다. TF-IDF 빈도수에서도 헬리코박터 파일로리균의 이름 관련 단어를 제외하면 ‘위암’, ‘위’, ‘감염’, ‘월’, ‘균’, ‘치료’, ‘위궤양’, ‘원인’, ‘프로젝트’, ‘음식’, ‘세균’, ‘검사’, ‘예방’, ‘증상’, ‘건강’, ‘약’, ‘효과’ 등의 순서로 중요도 표준화 값의 순서가 확인되었다(Table 2). 단어의 출현빈도를 시각화한 막대그래프와 단어 간 연결정도 및 빈도를 시각화한 네트워크 워드트리(network word-tree)는 Fig. 1과 Fig. 2에 표시하였다.

Twitter Search Network 분석은 총 96개의 노드와 120개의 링크(link; edge), 86개의 유니크 링크(unique edges)로 측정되어, 특정 노드 간 연결성이 높지 않고 다양한 사용자 간의

Table 1. Total and Collected Data and Its Source from Textom Analyses

Channel	Section	Amount of total data (n)	Amount of collected data (n)
Naver	Web page	250,366	1,000
Naver	Blog	6,695	1,000
Naver	News page	1,227	703
Naver	Internet community	1,073	1,000
Naver	지식IN (JisikIN)	745	745
Naver	Movie clip	395	395
Naver	Expertise	597	590
Naver	Academic data	308	300
Naver	Trend/Research report	19	10
Naver	Patent/Korean standard	234	230
Naver	Report/Form	41	40
Naver	National archive	1	1
Daum	Web page	123,000	760
Daum	Blog	9,290	958
Daum	News page	1,450	294
Daum	Internet community	110,000	945
Daum	지식IN (JisikIN)	446	477
Daum	Picture	3,133	0
Daum	Movie clip	64	54
Twitter	Unified search	51	51
Youtube	Unified search	2,410	508

Table 2. Frequency of Keywords Collected from Textom Analyses

Total, N Gram	Through the N Gram analysis	Through the TF-IDF analysis (descending order)
헬리코박터(8,828, 5.65%)	헬리코박터-파일로리균(839)	헬리코박터
헬리코박터균(6,380, 4.38%)	헬리코박터-파일로리(806)	헬리코박터균
위암(2,584, 1.77%)	헬리코박터-프로젝트(759)	위암
위(1,988, 1.36%)	프로젝트-위(707)	위
감염(1,974, 1.36%)	헬리코박터-균(632)	감염
위염(1,635, 1.12%)	헬리코박터균-감염(560)	위
치료(1,420, 0.98%)	헬리코박터-파일로리(496)	균
위(1,396, 0.96%)	헬리코박터-파일로리균(390)	치료
균(1,363, 0.94%)	위염-위궤양(307)	위궤양
위궤양(1,192, 0.82%)	위-점막(284)	원인
원인(1,096, 0.75%)	헬리코박터균-치료(279)	파일로리균
파일로리균(933, 0.64%)	헬리코박터-감염(230)	프로젝트
음식(874, 0.6%)	위암-발생(220)	파일로리
파일로리(866, 0.59%)	원인-헬리코박터균(202)	음식
프로젝트(828, 0.57%)	헬리코박터균-헬리코박터균(195)	세균
세균(703, 0.48%)	발명-헬리코박터(175)	검사
예방(637, 0.44%)	헬리코박터균-증상(171)	예방
검사(622, 0.43%)	헬리코박터-위(169)	증상
증상(604, 0.42%)	파일로리균-감염(165)	건강
효과(596, 0.41%)	원인-헬리코박터(161)	약
건강(593, 0.41%)	헬리코박터-헬리코박터(157)	효과
약(571, 0.39%)	헬리코박터균-검사(153)	파일로리
억제(567, 0.39%)	헬리코박터균-위(149)	억제
사람(538, 0.37%)	위암-원인(148)	한국야쿠르트
파일로리(536, 0.37%)	위-건강(147)	사람
속(525, 0.36%)	헬리코박터균-위암(142)	속
발생(490, 0.34%)	위궤양-위암(140)	발생
한국야쿠르트(484, 0.33%)	위암-헬리코박터(136)	위장
위장(478, 0.33%)	파일로리-균(135)	발명
발명(441, 0.30%)	균-치료(134)	파일로리균
파일로리균(416, 0.29%)	헬리코박터-필로리(130)	항생제
전(414, 0.28%)	헬리코박터균-증식(128)	암
경우(404, 0.28%)	감염-헬리코박터균(128)	전
점막(402, 0.28%)	위-저지방(124)	경우
항생제(399, 0.27%)	헬리코박터-파일로리Helicobacter (124)	점막
암(382, 0.26%)	파일로리-감염(123)	발견

TF-IDF, term frequency×inverse document frequency.

연결성을 확인하였다. 27개의 컴포넌트(component; 최소 하나 이상의 경로로 모두 연결되어 있는 하위 네트워크)⁵ 중 단일 노드 컴포넌트(single-vertex connected component)가 13개로 다른 노드와 연결성이 없는 단일 사용자가 많은 비중을 차지함을 시사한다.⁵ 최대 연결 거리(distance; 특정 노드들 사이에 존재하는 연결경로 중 가장 적은 수의 링크 단계를 거치는 경로)⁵는 2이고, 평균 연결 거리도 1.78로 높지 않았다. 가장 많이 등장한 단어는 ‘헬리코박터’로 58회 등장했으며, 해시태그(Hashtag; #)는 ‘헬리코박터’, ‘제균치료’, ‘항생제’, ‘내가약인지_약이아닌지’가 각 1회씩 등장했다. 짝지은 단어 순으로 확인할 경우 ‘냄비요리-시키면’, ‘숟가락-젓가락’, ‘아무렇지-않게’, ‘담그는-것’,

‘싫고-헬리코박터파일로리균’, ‘웁을것같다-그래서’, ‘국자로-빨리’, ‘내똥만-떠놓고’, ‘다시는-안’, ‘안-먹은다’ 등이 모두 31회 등장했다. 이어서 ‘마리끌레르-기사라’, ‘좀-알아보니’, ‘결핵은-음식으로’, ‘전염이-잘’, ‘안-되고’, ‘a형-간염’, ‘헬리코박터-등이’, ‘보균자의-타액으로’, ‘웁을-수’, ‘수-있다고’, ‘둘-다’, ‘한국에서-이상하게’, ‘감염자-많은’, ‘전문가의-고견을’, ‘고견을-기대’ 등이 모두 22회 등장했다. 이어서 ‘헬리코박터-파일로리’, ‘라는-바이러스때문에’, ‘한-음식을’, ‘같이-먹는거’, ‘몹시-싫어합니다’, ‘그러니까-제발’, ‘입-댄’, ‘댄-숟가락으로’, ‘남꺼-퍼먹지않시다’ 등이 모두 17회 등장했다. 이어서 ‘헬리코박터-프로젝트’, ‘위궤양이나-위’, ‘점막-출혈을’, ‘비롯하여-각종’, ‘부

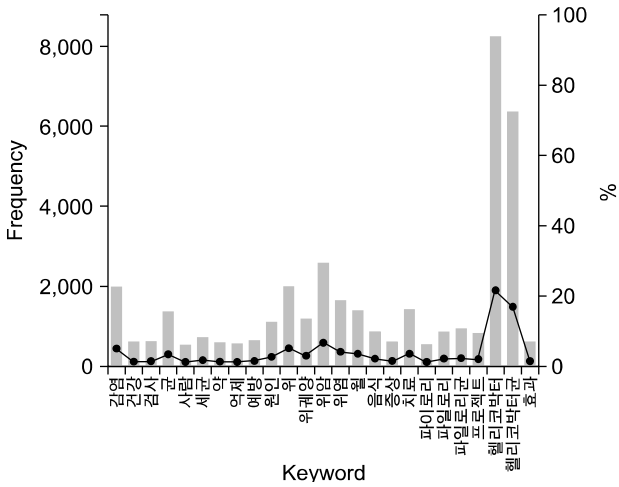


Fig. 1. Frequency of keywords detected in the Textom analysis.

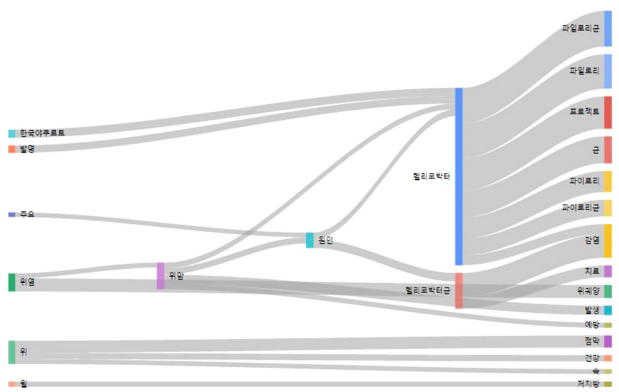


Fig. 2. Network word-tree showing connections of keywords in the Textom analysis.

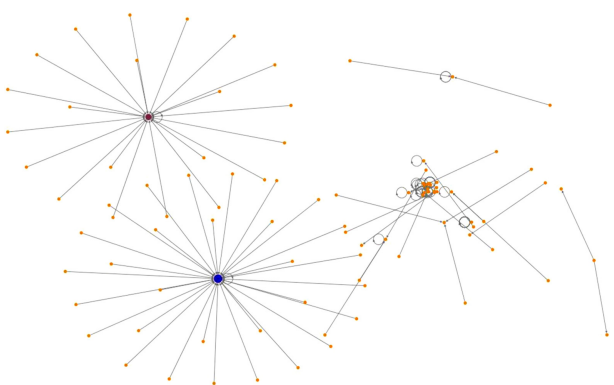


Fig. 3. Harel-Koren fast multiscale graph in the Twitter search network analysis. Circles represent the each tweet as a node and lines represent the connections. The extent of circle indicates the high betweenness centrality in each tweet.

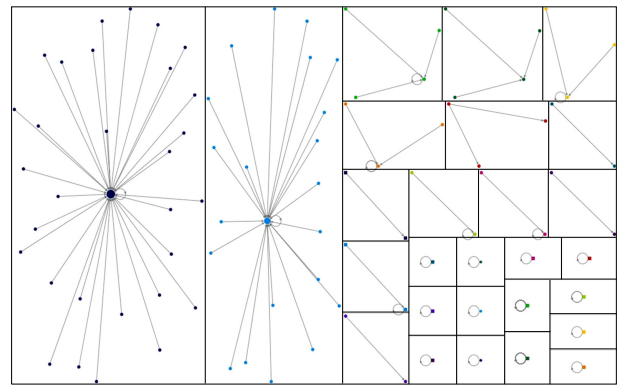


Fig. 4. Clauset-Newman-Moore algorithm grouping graph in the Twitter Search Network analysis. Each cluster is shown in separate box.

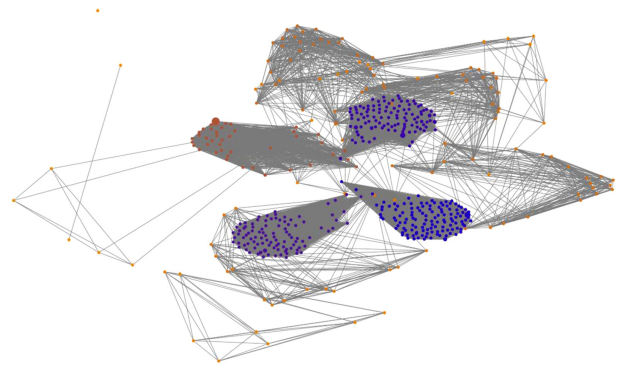


Fig. 5. Fruchterman-Reingold algorithm grouping graph in the Youtube Video Network analysis. Circles represent the each video clip as a node and lines represent the connections. The extent of circle indicates the number of view in each video clip. Blue color node means the video clip which has high degree connections and yellow color node means low degree connections.

중을 억제하고’, ‘치료하는데-효과가’, ‘있습니다-또한’, ‘장티푸스-이질’, ‘등-전염성’, ‘세균이나-장’, ‘속의-세균들을’, ‘억제하는-효과가’, ‘프로젝트-윌’ 등이 모두 13회 등장했다. 전체 Twitter Search Network 분석을 연결성과 between centrality에 따라 시각화한 Harel-Koren Fast Multiscale 그래프⁸는 Fig. 3에 표시하였다. 클러스터(Cluster; 유사한 속성을 가진 노드들의 집합)⁵별로 구분하여 표시한 Clauset-Newman-Moore 알고리즘 그룹핑 그래프는 Fig. 4에 표시하였다.⁹

Youtube Video Network 분석은 총 500개의 노드(동영상)와 20,520개의 링크로, 14개의 컴포넌트 중 1개만이 단일 노드 컴포넌트로 확인되었다. 다른 동영상과 연결정도(degree; 해당 노드에 직접 연결되어 있는 노드들의 개수, 또는 링크의 개수)⁵가 높으면서 조회수가 가장 많은 동영상은 124회의 연결정도와 210,577회의 조회수를 기록하고 2015년 12월에 게시된 15초

분량의 ‘한국아쿠르트 헬리코박터 프로젝트 윌 그 한입을 경계하라_15초’로 확인되었다. 단순 조회수로 분류하였을 경우 헬리코박터와 연관성이 없는 동영상을 제외하면 607,884회의 조회수를 기록하고 2015년 1월에 Asan Medical Center (Seoul, Korea)에서 게시한 ‘위 [내 몸 둘러보기]가 가장 높은 조회수를 기록하였으나, 연결정도가 108회로 앞서 언급한 동영상에 비해 상대적으로 낮은 수치로 확인되었다. 이어서 높은 조회수와 연결정도를 기록한 동영상은 Genikids Adventure에서 2015년 12월 게시한 ‘인체탐험 #6 위 속 세균과의 전투 Battle with bacteria in the stomach | Explore Human Body’가 108회의 연결정도와 245,124회의 조회수를 기록하였다. 각 동영상별 조회수와 연결성을 시각화한 Fruchterman-Reingold 알고리즘 그룹핑 그래프는 Fig. 5에 표시하였다.¹⁰

고찰

대중의 정보 획득과 전파방식이 과거의 신문, 책, 자료 등으로부터 획득된 지식의 단순 전달 방식에서 소셜미디어를 통한 다양한 형태의 자료가 복잡한 방식으로 공유되는 형태로 바뀌고 있다. 본 연구는 의학자의 헬리코박터에 대한 정보 생산과 전달은 고려하지 않고, 대중의 시각에서 일상적으로 생산되고 공유되는 헬리코박터에 대한 정보가 어떤 특성을 갖고 있는지 확인하고자 시행되었다.

첫째로, 국내에서 생산되고 유통되는 헬리코박터에 대한 소셜미디어 정보는 그 양이 많지 않았다. 절대적인 기준이 있는 것은 아니지만 1년간의 텍스트 키워드 관련 수집량이 10,061건으로 분석단위가 비정형 텍스트이고 중복된 정보가 유통됨을 고려할 때 그 양은 적은 것으로 판단된다(Table 1).

두 번째로 수집단어의 중요도 표준화 값 분석 결과 위암, 위염, 치료, 위궤양 등에 대한 대중의 관심이 높다는 것을 확인할 수 있었지만, 특정 유통 식품에 대한 광고효과로 그 관심이 다른 중요 키워드보다 높다는 것이 확인되었고, 십이지장궤양, 제균 등의 키워드는 생략되어 있음을 알 수 있었다(Table 2, Fig. 1, 2).

세 번째로 Twitter Search Network 분석결과 정보전달의 연결성이 높지 않고, 다양한 사용자 간의 짧은 연결성을 확인하였다(Fig. 3, 4). 짝지은 단어의 분석에서는 검증되지 않은 정보의 전달 과정에서 전문가의 의견 제시를 요구하는 트윗이 다수 관찰되기도 하였는데, 이를 반영하듯 특정 유통 식품의 균 억제 효과나 심지어 장티푸스, 이질 등을 억제한다는 정보 또는 헬리코박터균을 바이러스로 표현하는 트윗도 높은 빈도로 유통되는 것이 확인되었다.

네 번째로 Youtube Video Network 분석에서는 동영상 자체의 연결, 전달 정도를 분석하였기 때문에 트위터보다는 좀 더 그 전달이나 유통이 활성화되어 있는 것이 확인되었다(Fig. 5). 하지만 조회수가 가장 높은 동영상은 특정 유통 제품에 대한 광고목적의 동영상이었으며, 일부 소수의 의료기관만이 동영상 정보전달을 효율적으로 시행하고 있는 것이 확인되었다.

본 연구는 비정형 텍스트 정보와 동영상 전달방식을 분석하여 실질적인 대중의 헬리코박터균에 대한 관심사항과 유통 정보의 특징을 관찰하였다. 하지만 가장 활성화된 소셜미디어 중 하나인 페이스북의 정보를 분석하지 못하였으며, 텍스트 분석은 오직 1년간의 자료만을, 트위터는 오직 18,000개의 트윗만을 분석하였고, 유튜브는 오직 500개의 동영상 자료를 분석하였기 때문에 현재의 소셜미디어 정보의 특징을 일반화할 수 있는지에 대한 한계점을 갖고 있다. 또한 실제 의료전문가가 운영하는 블로그, 의료포털 사이트, 개인 트위터나 페이스북 계정의 자료가 본 연구 선정기준에 모두 포함되지는 못하였기 때문에 그들의 사회적인 현상에 대한 영향이 희석될 수 있다는 한계점 또한 존재한다. 하지만 헬리코박터 파일로리에 대해 상대적으로 적은 정보량과 그 전달, 그리고 일부 왜곡된 정보가 유통되는 점을 확인한 것은 환자 및 일반대중과 의사소통이 중요한 의료인들이 주지해야 할 점이며, 동영상을 통한 정보전달 방식의 효율성을 고려하여 보건의료 전문가들의 참여 확대가 필요할 것으로 보인다.

소셜미디어에서 유통되는 헬리코박터 관련 정보는 그 양이 많지 않고 전달방식이 간소화되어 있다. 또한 일부 왜곡된 정보가 유통되고 있기 때문에 일반대중과 의사소통의 증진을 위해 의료 전문가의 참여확대가 필요하다.

REFERENCES

1. Jeong MS, Park SJ, Lee YH. Characteristics of social media-related adverse effect and confrontation plan. Natl Inf Soc Agency IT Policy Res Ser 2011;16:1-34.
2. Lee YH. Current status of social media activity and main issues in Korea. Internet Secur Focus 2014;10:56-78.
3. Textom. Daegu: The IMC Inc., Available from: <http://www.textom.co.kr>.
4. NodXL. Microsoft, Available from: <http://nodxl.codeplex.com>.
5. NodeXL Korea. [NodeXL ttarajapgil]. Seoul: Paradigm book, 2015. Korean.
6. Kim YH. Social network analysis. Seoul: Pakyoungsa, 2003.
7. Jang JW, Choi KH. Statistics act content analysis using semantic network analysis. Stat Anal 2012;17:53-66.
8. Harel D, Koren Y. A fast multi-scale algorithm for drawing large

- graphs. J Graph Algor 2002;6:179-202.
9. Clauset A, Newman ME, Moore C. Finding community structure in very large networks. Phys Rev E Stat Nonlin Soft Matter Phys 2004;70:066111.
 10. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw Pract Exp 1991;21:1129-1164.