

텍스트마이닝을 이용한 한국응급구조학회지 중심단어 분석

권찬양¹ · 양현모^{1*}

¹한국교통대학교 응급구조학과

Analysis of key words published with the Korea Society of Emergency Medical Services journal using text mining

Chan-Yang Kwon¹ · Hyun-Mo Yang^{1*}

¹Department of Paramedic Science, Korea National University of Transportation

=Abstract =

Purpose: The purpose of this study was to analyze the English abstract key words found within the Korea Society of Emergency Medical Services journal using text mining techniques to determine the adherence of these terms with Medical Subject Headings (MeSH) and identify key word trends.

Methods: We analyzed 212 papers that were published from 2012 to 2019. R software, web scraping, and frequency analysis of key words were conducted using R's basic and text mining packages. Additionally, the Word Clouds package was used for visualization.

Results: The average number of key words used per study was 3.9. Word cloud visualization revealed that CPR was most prominent in the first half and emergency medical technician was most frequently used during the second half. There were a total of 542 (64.9%) words that exactly matched the MeSH listed words. A total of 293 (35%) key words did not match MeSH listed words.

Conclusion: Researchers should obey submission rules. Further, journals should update their respective submission rules. MeSH key words that are frequently cited should be suggested for use.

Keywords: Key words, Text mining, MeSH, Word cloud

Received March 22, 2020 Revised March 30, 2020 Accepted April 19, 2020

*Correspondence to Hyun-Mo Yang

Department of Paramedic Science, 5F, 61, Daehak-ro, Jeungpyeong-gun, Chungcheongbuk-do, 27909, Republic of Korea

Tel: +82-43-820-5215 Fax: +82-43-820-5212 E-mail: emtyang@naver.com

I. 서 론

1. 연구의 필요성

논문의 초록은 연구의 전반적인 내용을 함축적으로 전달하는 중요한 요소이다. 특히, 초록 하단부에 작성되는 중심단어(Key words)는 양질의 논문을 정확하게 검색할 수 있도록 하며 이를 토대로 연구자들이 필요한 논문이나 주제를 쉽게 찾을 수 있도록 돕는다. 그렇기 때문에 중심단어는 연구 주제, 연구 방법 및 연구 결과 등을 축약하여 설명할 수 있는 가장 적합한 단어를 제시해야 한다.

한국응급구조학회지(The Korean Journal of Emergency Medical Services)는 응급구조학을 다루는 국내 유일의 학술지로 1997년 창간호(제 1권 1호) 발행을 시작으로 현재까지 응급구조학과 관련된 원저, 종설 및 임상증례 보고 등을 대상으로 총 53권의 학회지를 발간하고 있으며 응급구조 모든 분야의 기초 및 임상연구를 위해 활발한 활동을 펼치고 있다[1]. 현재 한국응급구조학회지의 중심단어에 대한 투고 규정은 초록이 끝나는 하단부에 저자 선정 중심단어를 5개 이하로 적도록 하고 있으며, 미국립의학도서관(National library of medicine)의 의학용어집인 MeSH(Medical subject headings; MeSH)에 공인된 단어 및 약자를 사용할 것을 권고하고 있다[1]. MeSH에 공인된 단어 사용을 권고하는 이유는 용어 해석에 대한 오해가 적도록 하는 것과 다른 연구자들의 논문 검색을 유용하도록 돕는 것이다[2]. 또한, MeSH에 공인된 단어는 포털 사이트 검색 등에서 노출되는 빈도가 높아 타 연구자들에게 상대적으로 논문이 인용될 확률이 높아질 수 있다[3].

텍스트마이닝(text mining)은 빅데이터 분석인 데이터마이닝(data mining)의 한 분야로 웹

(world wide web)상에 저장되어 있는 대규모의 데이터 중 디지털화 된 텍스트 정보만을 추출하여 패턴이나 통계적 규칙을 찾아내는 방법이다[4]. 이러한 분석 방법은 이미 우리 사회의 각 분야에서 널리 적용되고 있으며 특히 온라인에서 대부분의 논문을 검색하는 연구자들의 추세에 따라 최근 텍스트마이닝을 이용한 중심단어 분석에 대한 중요성이 부각되고 있을 뿐만 아니라 이미 여러 학회지의 중심단어 분석이 텍스트마이닝을 통해 이루어지고 있다[4-6]. 하지만 현재까지 본 학회지에 대한 중심단어 분석은 Lee와 Ham[7]이 2005년부터 2011년까지 한국응급구조학회지에 게재된 논문의 중심단어를 조사하여 MeSH와의 일치도를 빈도 분석한 결과와 Kang 등[8]이 응급구조학회지의 게재논문을 대상으로 연도별 사용한 중심단어의 구성 및 연구 분야를 분류한 결과 밖에 없어 인터넷에서 검색 가능한 중심단어 자체의 텍스트마이닝(text mining)을 통한 분석은 이루어지지 않았다.

이에 따라 본 학회지의 발전과 중심단어 사용 추세를 알아보기 위해 텍스트마이닝 분석은 반드시 필요하며, 이는 본 학회지의 주요 연구 동향과 트렌드 분석 및 시각화를 가능하게 할 것으로 생각된다. 따라서 본 저자들은 빅데이터 분석 도구 중 하나인 R 프로그램을 이용하여 온라인에서 확인할 수 있는 한국응급구조학회지 게재논문의 영문초록 중심단어를 텍스트마이닝 기법으로 분석하여 MeSH와의 일치도 및 중심단어의 사용 추세를 알아보고자 하였다.

2. 연구의 목적

본 연구의 구체적인 목적은 다음과 같다.

첫째, 한국응급구조학회지에 게재된 논문의 중심 단어 빈도를 파악한다.

둘째, 각 논문의 중심 단어와 MeSH 등재어와의 일치정도를 알아본다.

셋째, 사용된 중심단어 중 MeSH와의 불일치 단어를 파악한다.

II. 연구방법

1. 연구설계

본 연구는 최근 8년간 한국응급구조학회지에 게재된 중심단어를 조사하여 최근 중심단어의 사용 빈도 및 MeSH와의 일치 정도를 분석하기 위해 수행된 논문분석연구이다.

2. 연구대상

한국응급구조학회지에 게재된 논문 중 중심단어 분석이 이루어지지 않은 2012년 제16권 1호부터 2019년 제23권 3호까지 총 212편 논문의 영문 중심단어를 분석 대상으로 하였다. 비교 분석을 위해 2012년 제16권 1호부터 2015년 제19권 3호의 중심단어를 전반기, 2016년 제20권 1호부터 2019년 제23권 3호의 중심단어를 후반기로 나누었다.

3. 자료수집 방법

자료 수집은 국가과학기술정보센터(www.ndsl.kr)에서 한국응급구조학회를 검색한 후 2012년부터 2019년까지의 논문을 대상으로 R 소프트웨어(ver. 1.2.5033, R Foundation for Statistical Computing, Vienna, Austria)를 이용하여 논문의 영문 초록에 기술된 중심 단어를 웹 스크래핑(web scrapping)하였다.

추출된 중심단어는 어떠한 변형을 하지 않았으며 R 소프트웨어의 오류를 방지하기 위해 연구자 2인이 초록 또는 원문을 이중 확인하여 오타자를 수정하였다. 또한, 중심단어 축약 설명을 위한 괄호 제거 및 복수단어의 단수화, 낱말을 합치거나

음절을 나눌 때 쓰인 붙임표를 제거하는 전처리과정(preprocessing)을 수행하였다.

정리된 중심단어는 본 학회지의 규정에 따라 MeSH Browser(<https://meshb.nlm.nih.gov/search>)를 이용하여 MeSH에 소속되었는지 여부를 이중 확인함으로써 이들 중심단어의 MeSH 사용 여부를 확인하였다.

4. 분석방법

분석에 사용한 프로그램은 데이터 분석 오픈소스인 R 소프트웨어를 이용하였으며, 중심단어의 웹 스크래핑과 빈도 분석은 R의 기본 패키지와 텍스트마이닝 패키지를 사용하였고 이를 시각화하기 위해 워드 클라우드(word clouds) 패키지를 사용하였다.

중심단어의 인용 빈도와 MeSH 일치율은 빈도 분석과 백분율로 표기하였으며 양적변수는 평균과 표준편차로 표기하였다. 전·후반기의 연구추세를 확인하기 위해 워드 클라우드를 이용하여 시각화하였다.

III. 연구결과

1. 한국응급구조학회지의 저자 선정 중심단어 빈도

1) 연도별 중심단어 수

연구대상 중 응급구조학회지에 선정된 중심단어의 수는 2012년 82개, 2013년 94개, 2014년 103개, 2015년 84개, 2016년 104개, 2017년 96개, 2018년 118개, 2019년 154개로 총 835개였으며 논문 한 편당 사용한 중심단어의 평균 개수는 3.9개였다. 2012년부터 2015년까지 전반기에 사용한 중심단어 수는 평균 3.82개였으며 2016년부

터 2019년까지 후반기에 사용한 중심단어 수는 4,03개였다<Table 1>.

2) 워드 클라우드를 이용한 전·후반기 중심단어 빈도의 시각화

논문의 영문초록에서 추출된 각각의 중심단어의 사용 빈도를 전·후반기로 나누어 워드 클라우드를 시각화한 결과, 전반기에는 심폐소생술(CPR)이 가장 두드러졌으며 후반기에는 응급구조사(emergency medical technician)가 가장 빈번하게 사용한 단어였다<Fig. 1-2>.

3) 전·후반기에 따른 중심단어 빈도

워드 클라우드를 통해 시각적으로 추출된 중심단어들을 전·후반기로 나누어 분석한 결과 중 사용 빈도가 3회 이상인 중심단어를 제시하였다. 전반기에는 심폐소생술(CPR), 응급구조사(paramedic), 응급구조사(emergency medical technician), 지식(knowledge), 자기 또는 자아 효능감(self efficacy), 응급구조과 학생(paramedic student), 이직의도(turnover intention) 등의 순으로 사용이 많았으며, 후반기에는 응급구조사(emergency medical techni-

Table 1. Number of author key words (N=835)

	First half-year					Second half-year				Total
	2012	2013	2014	2015	2016	2017	2018	2019		
Number of key words, total (n, %)	82 (9.8)	94 (11.3)	103 (12.3)	84 (10.1)	104 (12.4)	96 (11.4)	118 (14.2)	154 (18.5)	835 (100)	
Number of key words, mean (M±SD)	3.90±.88	3.76±.83	3.96±.66	3.65±.88	4.16±.89	3.84±.74	4.21±1.20	3.95±.72	3.93±.85	
Number of key words ranges (n)	2-6	3-5	2-5	3-5	2-5	3-5	3-7	3-5	2-7	

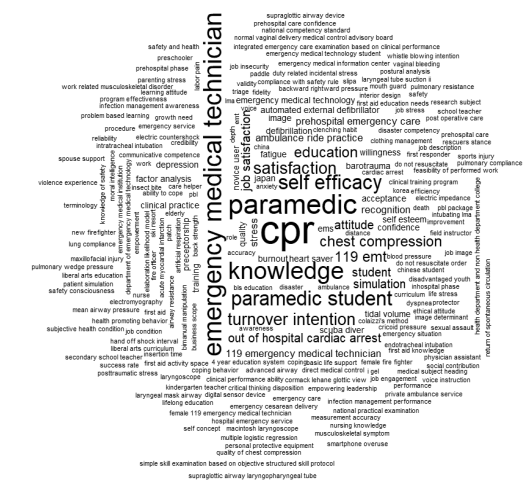


Fig. 1. Word cloud of key words of praise from 2012 to 2015.

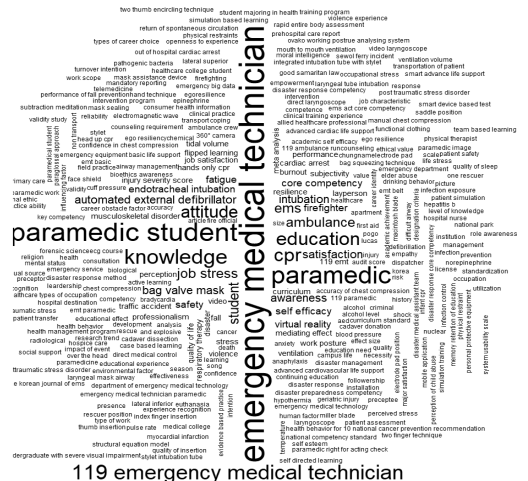


Fig. 2. Word cloud of key words of praise from 2016 to 2019.

cian), 응급구조과 학생(paramedic student), 응급구조사(paramedic), 지식(knowledge), 구급대원(119 emergency medical technician), 교육(education), 심폐소생술(CPR)등의 순으로 중심단어가 사용되었다<Fig. 3>.

2. 한국응급구조학회지 중심단어에 대한 MeSH 일치 정도

2012년부터 2019년까지의 전체 중심 단어 835개를 MeSH Browser를 이용하여 검색한 결과 MeSH 등재어에 정확하게 일치하는 단어는 총 542개(64.9%)였으며, MeSH 등재어와 불일치 하는 중심단어 수는 총 293개(35%)로 조사되었다. 연도별 일치도는 2012년 67%, 2013년 65.9%, 2014년 65%로 점차 감소한 후 2015년 65.4%,

2016년 68.2%로 증가하였으나 2017년도 67.7%, 2018년 66.1%, 2019년 57.7%로 점차 감소하였다. 또한, 전반기에 사용한 중심단어의 MeSH 등재어 일치 비율은 28.6%였으며 후반기에 사용한 중심단어의 일치비율은 36.2%로 조사되었다<Table 2>.

3. 저자 선정 중심단어 중 다빈도이면서 MeSH 불일치 단어

저자 선정 중심단어 중 다빈도이면서 MeSH 등재어와 불일치하는 단어 중 사용 빈도가 5회 이상인 중심단어를 제시하였다. MeSH 불일치 단어 중 가장 많이 사용된 중심단어는 응급구조과 학생(28회), 119 구급대원(21회), 이직의도(6회), 핵심역량(5회)으로 조사되었다<Table 3>.

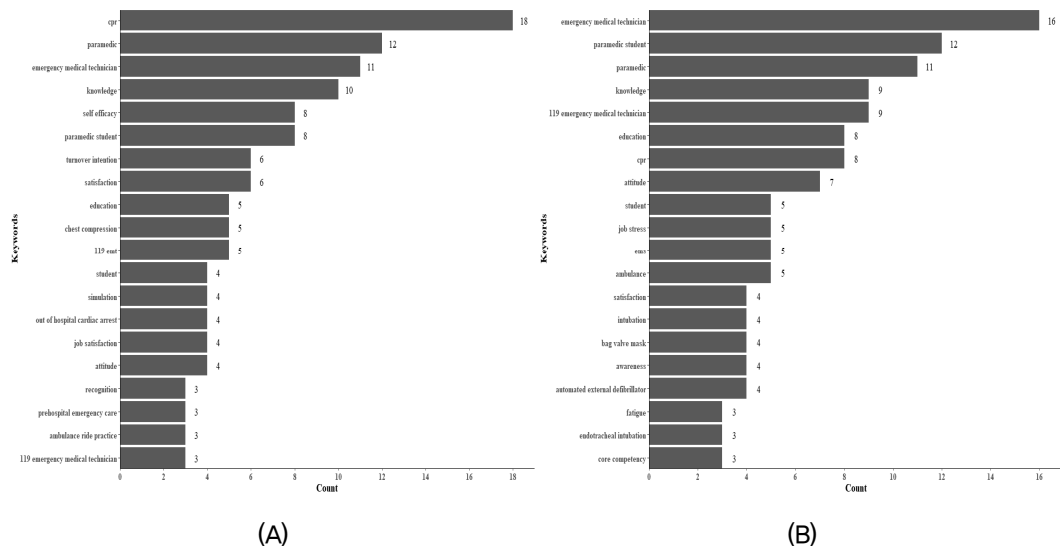


Fig. 3. Visualize key word frequency from 2012 to 2015 (A) and from 2016 to 2019 (B).

Table 2. Proportion of the MeSH key words (N=835)

	First half-year				Second half-year				Total
	2012	2013	2014	2015	2016	2017	2018	2019	
MeSH key words (n, %)	55 (67)	62 (65.9)	67 (65)	55 (65.4)	71 (68.2)	65 (67.7)	78 (66.1)	89 (57.7)	542 (64.9)

Table 3. Frequently used Non-MeSH key words from 2012 to 2019 in the Korean Journal of Emergency Medical Services

Frequency	Non-mesh Key words
28	paramedic student / paramedical student / emergency medical technology student / department of emergency medical technology / emergency medical technology / paramedicine
21	119 emt / 119 emergency medical technician / female 119 emergency medical technician / 119 paramedic
6	turnover intention
5	core competency / ems act core competency / key competency

IV. 고 찰

본 연구는 2012년부터 2019년까지 한국응급구조학회지에 게재된 논문의 영문 중심단어를 분석하여 연구자들의 중심단어 선택 추이 및 MeSH의 일치 정도를 파악하였으며, 이를 토대로 기존의 선행된 연구와 비교 논의하면 다음과 같다.

한국응급구조학회지에서 지난 8년간 사용한 중심단어 수는 총 835개로, 논문 한 편당 사용한 중심단어의 평균 개수는 3.9개였다. 이는 2005년부터 2011년까지 평균 3.2개로 보고한 Lee와 Ham[7]의 연구 결과와 비교하였을 때 상대적으로 증가하였다. 이러한 결과는 Cho와 Lee[9]의 연구에서 시간이 경과할수록 저자들이 사용하는 중심단어의 수가 증가됨을 보고한 연구결과와 일치하며 이는 응급구조학 연구가 활성화됨에 따라 연구자가 중심단어 인용의 중요성을 인식하고 적용하여 나타난 결과이다. 그러나 중심단어를 5개 이하로 적도록 규정하고 있는 본 학회지의 투고규정[1]에도 불구하고 본 연구결과에서 전반기에 사용한 중심단어 범위는 2-6개였으며 후반기에는 2-7개로 나타나 투고규정이 완전하게 지켜지지 않았다. 이러한 결과를 통해 중심단어에 대한 투고규정을 강조할 필요가 있으며, 논문 심사 시 중심단어 수가 맞지 않을 경우 수정 및 보완을 통해

정확한 범위 내의 중심단어를 기재하도록 해야 할 것이다.

한국응급구조학회지에 기재된 중심단어와 MeSH 등재어와의 일치도는 64.9%로 나타났으며, 이는 Lee와 Ham[7]의 연구결과와의 37.6%와 비교했을 때에도 상당한 차이가 있었다. 이러한 차이가 나타난 이유는, Lee와 Ham[7]의 연구 이후로 본 학회지에서 MeSH에 공인된 단어 및 약자를 사용할 것을 권고하는 중심단어의 투고 규정이 신설되어 나타난 현상으로 보여진다. 하지만 연도별 분석 시 2014년에서 2016년까지 약간의 증가 추세를 제외한 나머지 연도에서 꾸준히 감소하는 추세를 볼 때 중심단어 선정 시 저자 임의대로 선정하거나 비슷한 주제를 가진 다른 논문들을 참고자료로 사용하는 경우가 응답자의 57.1%로 나타난 Mun[3]의 설문조사 결과가 그 이유로 보인다.

본 학회지에서 빈도수가 높은 중심단어는 전반기에서 심폐소생술(CPR), 응급구조사(paramedic), 응급구조사(emergency medical technician)이었고, 후반기에는 응급구조사(emergency medical technician), 응급구조과 학생(paramedic student), 응급구조사(paramedic)이었다. 이러한 결과는 본 학회의 원문서식 중 영어 용어 통일에 관한 학회 규정에 의한 결과로 생각된다. 현재 한국응급구조학회는 응급구조학과 학생을 paramedic

student로 사용하도록 규정하며 1급 응급구조사를 EMT-Paramedic으로 사용하도록 권고하고 있다. 이러한 이유로 응급구조사의 경우 emergency medical technicians으로 MeSH 등재어가 있음에도 불구하고, 많은 저자들이 위 단어의 축약어인 EMT를 사용하거나 paramedic 등의 단어를 사용하고 있었다. MeSH에 등재된 단어 사용을 위해 투고규정의 수정이 반드시 필요할 것으로 생각되며 연구자들은 중심단어를 선택할 때 반드시 MeSH에 등재된 단어인지 확인해야 할 것이다. 또한, 다빈도로 선정되는 응급구조과 학생의 경우 paramedic student, paramedical student, emergency medical technology student 등 다양한 중심단어로 기재 되어 있었는데 이 또한 MeSH에 등재되어 있는 emergency medical technicians와 students를 각각 인용하여 사용해야 할 것이다. 그리고 저자가 임의로 선정하는 중심단어가 두 단어 이상을 결합하여 사용하는 경우이다. 예를 들어, female 119 emergency medical technician의 경우 Female, 119, emergency medical technician을 따로 기재할 경우 119를 제외한 나머지 단어가 MeSH 등재어임에도 불구하고 저자들이 위와 같은 결합어를 만들어 사용하고 있었다. 이는 본 학회지가 5개 이하의 중심단어 사용을 규정한 것과 관련이 있을 것으로 생각된다. 하지만 실제 MeSH 색인화 특성상 많은 중심단어를 인용하는 경우 특이도가 높아짐을 고려할 때 5개 이상의 중심단어를 인용하도록 하는 방향도 논의해볼 필요가 있을 것으로 생각된다. 결과적으로 MeSH에 채택되지 않는 본 학문 분야의 전문용어나 약어의 경우 국내외 응급구조사들의 논의를 통해 MeSH에 제안하는 노력이 필요하며, 한국응급구조학회지에서 투고규정 내에 자주 발생하는 단어오류 항목을 신설, 또는 '구조·구급 신용어사전[10]'과 같은 전문용어를 기술한 사전을 제시함으로써 저자들의 단어선택 오류를 방지하기 위한 노력을 기울여야 할 것이다.

V. 결 론

본 연구결과 835개의 중심단어 중 논문 한 편당 사용한 중심단어의 평균 개수는 3.9개, 범위는 2-7개였다. MeSH 등재어에 정확하게 일치하는 단어는 총 542개(64.9%)였으며 연도별 분석 시 2014년에서 2016년 약간의 증가 추세를 제외한 나머지 연도에서 꾸준히 감소하는 추세를 보였다. 따라서, 연구자들의 투고규정 준수 및 본 학회지의 중심단어 투고규정 업데이트가 필요하며, 응급구조학에서 많이 인용되는 중심단어에 대한 MeSH 등재 제안 및 국내 전문가들의 단어 일치화에 대한 방향도 모색되어야 할 것이다.

ORCID ID

Chan-Yang Kwon

0000-0001-8245-1435

Hyun-Mo Yang

0000-0002-4663-9075

References

1. The Korea Society of Emergency Medical Services.
Available at: <http://www.kemt.or.kr/>
2. Kim BS, Kim SY. The coincidence of the english key words of the Journal of Korean Academy of Family Medicine with MeSH and selection validity. J Korean Acad Fam Med 1998;19(7):531-7.
3. Mun HW. Subject headings of Korean medi-

- cal journal papers: author key words vs MeSH: unpublished doctoral dissertation, Sookmyung Women's University 2000, Seoul, Korea.
4. Jung YB, Park ES. Keyword analysis of two SCI journals on rock engineering by using text mining. *Tunnel and Underground Space* 2015;25(4):303–19.
<https://doi.org/10.7474/TUS.2015.25.4.303>
 5. Cho SG, Kim SB. Finding meaningful pattern of key words in IIE transactions using text mining. *Journal of the Korean Institute of Industrial Engineers* 2012;38(1):67–73.
<https://doi.org/10.7232/JKIIIE.2012.38.1.067>
 6. Hwang KC, Cho GC, Sohn YD, Cho YS, Lee JH, Lee HJ et al. Key words analysis of the *Journal of the Korean Society of Emergency Medicine* using text mining. *J Korean Soc Emerg Med* 2019;30(1):94–9.
 7. Lee KH, Ham YL. Coincidence analysis of key words and MeSH terms in the *Korean Journal of Emergency Medical Services*. *Korean J Emerg Med Ser* 2012;16(2):43–51.
 8. Kang MJ, Jeon HJ, Ju HM. Research trends in the *Korean Journal of Emergency Medical Services*: examining the articles published in the past 10 years. *Korean J Emerg Med Ser* 2019;23(1):147–55.
<https://doi.org/10.14408/KJEMS.2019.23.1.147>
 9. Cho JS, Lee MJ. Coincidence analysis of key words and MeSH terms in the *Journal of The Korean Society of Emergency Medicine*. *J Korean Soc Emerg Med* 2009;20(6):722–8.
 10. Park HJ, Choi SY, Hwang JY. *The new glossary for rescue and first aid*. Seoul:Hyemin Books:Medical Tree Publishing Company, 2017.