대한응급의학회지 제 30 권 제 1 호
Volume 30, Number 1, February, 2019

□ 원 저 □

Administration

# 텍스트 마이닝을 이용한 대한응급의학회지 중심단어 분석

황기천[1] · 조규종[1] · 손유동[2] · 조영석[1] · 이진혁[1] · 이현정[1] · 차현민[1] · 장형우[1]

[1]한림대학교 의과대학 강동성심병원 응급의학과, [2]서울특별시 보라매병원 응급의학과

# Keywords analysis of the Journal of the Korean Society of Emergency Medicine using text mining

**Ki Cheon Hwang[1], Gyu Chong Cho[1], Youdong Sohn[2], Youngsuk Cho[1], Jinhyuck Lee[1], Hyung Jung Lee[1], Hyun Min Cha[1], Hyung Woo Chang[1]**

[1]Department of Emergency Medicine, Hallym University Kangdong Sacred Heart Hospital, Hallym University School of Medicine, [2]Department of Emergency Medicine, Seoul Metropolitan Government Boramae Medical Center, Seoul, Korea

**Objective**: Data mining extracts meaningful information from large datasets. In this study, text mining techniques were used to extract keywords from the Journal of the Korean Society of Emergency Medicine, and the change trend was examined.
**Methods**: The rvest package in R was used to extract all papers published in the Journal of the Korean Society of Emergency Medicine from 2006 to 2016 that could be searched online. Among them, 3,952 keywords were extracted and studied. Using the selected keywords, the corpus was formed by refining keywords that did not correspond to MeSH (Medical Subject Headings) or were misspelled and had similar meanings based on agreement of researchers. Using the refined keywords, the frequencies of the keywords in the first and second halves of the studies were calculated and visualized.
**Results**: Word Cloud revealed that emergency medical service and cardiopulmonary resuscitation (CPR) were most frequently mentioned in both the first and second halves of the studies. In the first half, ultrasonography, stroke, poisoning, injury, and education were frequently mentioned, while in the second half, poisoning, injury, stroke, acute, and tomography were frequently mentioned. A pyramid graph revealed that the frequencies of emergency medical service and CPR were commonly high.
**Conclusion**: Core keywords of the Journal of the Korean Society of Emergency Medicine were analyzed for correlations and trends. Changes in study topics according to key topics of interest and period were visually identified.

**Keywords**: Data mining; Journal article; Emergency medicine

## 서 론

데이터 마이닝(data mining)은 대규모로 저장된 데이터 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 찾아내는 것으로 정의할 수 있으며, 데이터베이스 속의 지식 발견(knowledge-discovery in databases)이라고도 일컫는다.[1] 이미 우리 주변에서는 이런 데이터 마이닝을 적용한 사례를 쉽게 찾아볼 수 있는데, 대표적인 적용사례를 들면 한국석유공사의 국내 주유소 유가 데이터수집을 통한 최적의 유가정보제공 서비스시스템이나 장소 및 시간에 따른 통신사의 통화량 데이터를 이용한 서울시 올빼미버스 운영 등이 있다.[2,3]

대한응급의학회지는 우리나라 응급의학의 학문적 성과

책임저자: 손 유 동
서울특별시 동작구 보라매로 5길 20
서울특별시 보라매병원 응급의학과
Tel: 02-870-2666, Fax: 02-831-2826, E-mail: medysohn@gmail.com
접수일: 2018년 10월 19일, 1차 교정일: 2018년 10월 23일, 게재승인일: 2018년 10월 24일

## Capsule Summary

***What is already known in the previous study***

*The Journal of the Korean Society of Emergency Medicine (JKSEM) has published since 1990. JKSEM is the only and the specialized journal about Emergency Medicine in Korea. However, we haven't looked over the trend of interesting topics.*

***What is new in the current study***

*Keywords such as stroke, ultrasonography, and tomography were frequently mentioned in the JKSEM in 2006-2010, while tomography, stroke, and prognosis were mentioned frequently in 2011-2016. Keyword analysis using text mining can be utilized to demonstrate trends in subjects in the JKSEM.*

를 발표하는 대표 저널로서 1990년에 처음 발간된 이래로 현재까지 약 2,400여 편이 출간되었으며, 최근에는 연간 6권의 학회지에 총 100여 편 내외의 저널이 발표되고 있다. 이러한 성과를 인정받아 대한응급의학회지는 현재 대한의학학술지편집인협의회에 등재된 261종의 전문 학회지 중에서 응급의학 영역을 다루는 우리나라 대표 저널로 성장하였다.[4]

응급의학은 응급실로 방문하는 모든 연령의 질병 및 손상 환자에 대한 예방, 진단 및 치료를 담당함에 따라 다양한 세부 분야로 구분될 수 있다.[5] 이에 따라 대한응급의학회지는 급성치료 및 응급의학 관련 모든 분야의 기초 및 임상연구를 대상으로 하고 있으며, 매우 다양한 연구 논문들이 출판되고 있다. 대한응급의학회지는 편집규정으로 논문의 영문 초록에 중심단어 3-5개를 기술하도록 규정하고 있다. 논문의 초록에 실리는 중심단어는 연구주제를 함축적으로 대표할 수 있는 단어로 연구자가 선택하게 되며, 다른 연구자가 논문을 효율적으로 검색할 수 있도록 MeSH (Medical Subject Headings) 규정을 준수하도록 권고된다.[6,7]

대한응급의학회지가 출간된 지난 20여년 동안 응급의학의 발전과 더불어 학문적 영역에서의 관심도 또한 많은 변화가 있었을 것으로 예상된다. 그러나 우리 학회지에 대한 학문적 관심도 조사는 아직까지 이루어지지 않았다. 이에 저자들은 본 학회지의 논문 속에 삽입된 중심단어들이 논문의 성격을 잘 파악한다고 판단하였고 비정형자료의 분석을 통한 대한응급의학회지의 출판 시기에 따른 연구 동향을 알아보고자 대한응급의학회지 영문 초록에 기술된 중심단어의 빈도를 분석하였다.

## 방 법

대한응급의학회지는 1990년부터 출간되었으나, 한국의학논문데이터베이스(http://kmbase.medric.or.kr)에서 대한응급의학회지를 검색하여, 온라인으로 다운로드가 가능한 2006년부터 2016년까지의 논문을 대상으로 공개 소프트웨어인 R의 rvest 프로그램(ver. 0.3.2)를 이용하여 중심단어 웹 스크래핑(web scrapping)을 시행하였다.[8] 이후, 추출된 중심단어 중에서 오탈자를 수정하고, MeSH 규정을 지키지 않아 유사한 의미의 단어들이 각각으로 선택된 경우에는 유사한 의미의 중심단어들로 일치시키는 전처리 과정을 응급의학과 전문의 2인이 수행하였다. 이렇게 정제된 중심단어들을 저자들은 연구기간을 전반기와 후반기로 나누었는데, 특히 2010년도에 대한응급의학회지 온라인 논문 투고 시스템이 갖추었고, 본격적인 시스템을 갖춘 2011년을 기준년으로 잡아서 전후반기의 연구기간을 비교하였다. 연구기간 동안의 중심단어를 워드 클라우드(Word Clouds) 프로그램을 이용하여 시각화하였고, 이것을 바탕으로 10회 이상 빈번하게 언급된 중심단어를 막대그래프로 표현하였다.[9] 전반부와 후반부에 추출된 중심단어의 빈도 차이는 student t-test를 이용하여 분석하였다. 모든 분석은 R 소프트웨어(ver. 3.4.2, R Foundation for Statistical Computing, Vienna, Austria)를 이용하였으며, P값이 0.05 미만인 경우를 통계적으로 유의한 것으로 하였다.[10]

## 결 과

### 1. 대한응급의학회지 발표 논문 및 중심단어

2006년부터 2016년까지 11년 동안 한국의학논문데이터베이스에서 검색된 대한응급의학회지 출판 논문은 모두 1,093편이었으며, 이 중에서 중심단어가 없는 56편을 제외한 1,037편의 논문을 연구 대상으로 하였다. 대한응급의학회지 논문은 2006년에 41편이 출판되었으며, 지속적으로 증가되어 2012년에는 135편이 출판되었고 이후 점차 감소되었다(Fig. 1). 연구기간의 전반기 동안에는 총 476편(연평균 95.2편)의 논문이 출판되었으며, 후반기에는 총 561편(연평균 93.5편)의 논문이 출판되었다. 대상 연구논문에서 총 3,952개의 중심단어가 추출되었으며, 전반부에 추출된 중심단어는 1,652개(논문 평균 3.47개)였고 후반부에 추출된 중심단어는 2,300개(논문 평균 4.10개)였으며, 통계적으로 유의한 차이를 보였다($P<0.001$).

### 2. 워드 클라우드를 이용한 중심단어의 시각화

2006년부터 2010년, 2011년부터 2016년까지 전후반기로 나누어 추출된 각각의 워드 클라우드로 시각화하였다(Fig. 2, 3). 연구기간 동안의 워드 클라우드에서는 심폐소생술과 응급의료체계(emergency medical services)가 가장 빈번히 나타났다.

### 3. 중심단어의 빈도분석

워드 클라우드를 통해 시각적으로 추출된 중심단어들을 전후반기로 나누어 빈도분석을 했을 때, 전반기에는 응급의료체계, 심폐소생술, 뇌졸중, 초음파, 컴퓨터단층촬영, 교육, 중독 등의 순으로 언급이 많았으며, 후반기에는 심폐소생술, 응급의료체계, 컴퓨터단층촬영, 뇌졸중, 예후, 패혈증, 중독 등의 순으로 언급되었다. 특히 전반기에는 심폐소생술에 비해 응급의료체계의 빈도가 많았으며, 컴퓨터단층촬영보다 초음파의 빈도가 많았다. 이에 반해 후반기에는 심폐소생술, 컴퓨터단층촬영, 예후, 패혈증의 언급이 많았으며, 초음파의 언급빈도는 상대적으로 적었다. 중독, 교육, 기도삽관, 사망, 손상 등의 중심단어는 전후반기 모두에서 비슷한 빈도로 나타났다(Fig. 4, 5).

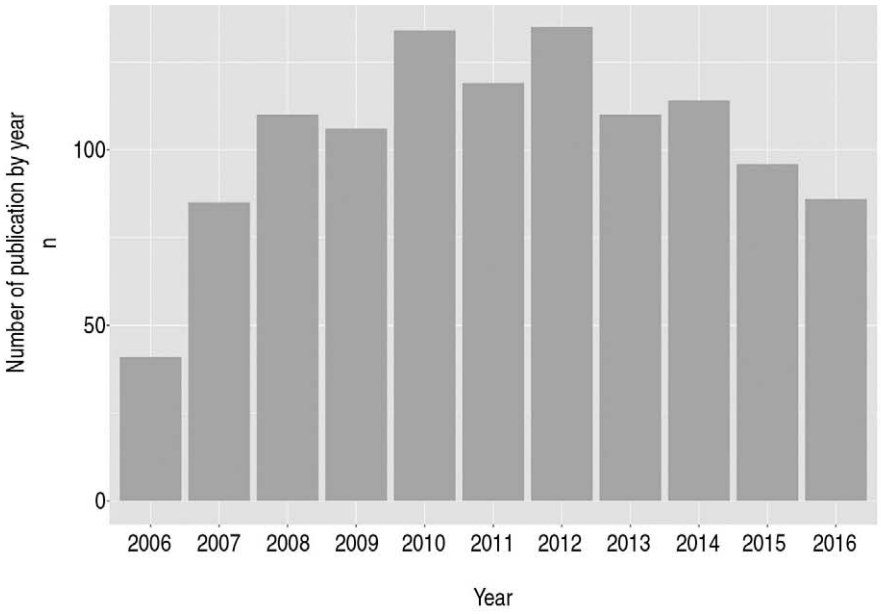


**Fig. 1.** Number of annual publications by Journal of the Korean Society of Emergency Medicine.

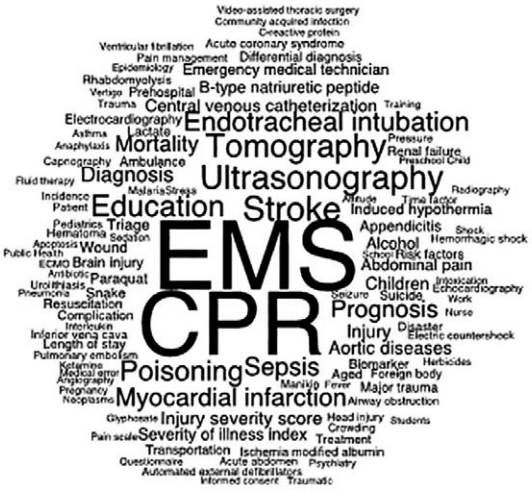


**Fig. 2.** Word cloud of core keyword in Journal of the Korean Society of Emergency Medicine from 2006 to 2010.

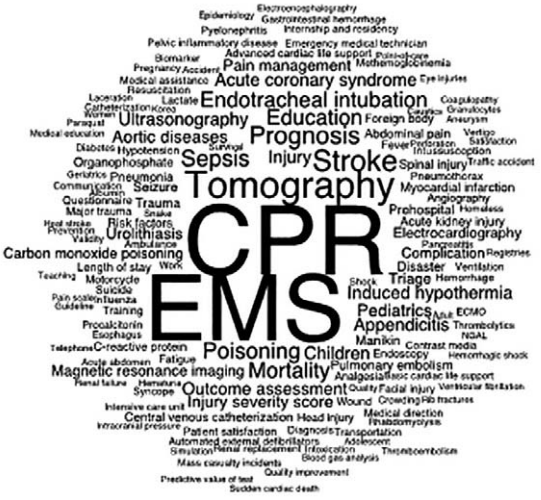


**Fig. 3.** Word cloud of core keyword in Journal of the Korean Society of Emergency Medicine from 2011 to 2016.

## 고 찰

이번 연구에서 적용한 텍스트 마이닝은 데이터 마이닝의 일종으로 광범위한 비정형화된 문서 정보를 처리하여 정형화하고 의미 있는 정보를 추출하는 일련의 과정을 의미한다. 이렇게 추출된 문서 정보는 군집화 및 시각화를 통해 의미 있는 자료로 재가공되어 우리사회의 각 분야에서 널리 적용되고 있으며, 최근에는 응급실에서의 초진 의료정보를 이용한 입원 예측, 텍스트 마이닝을 사용한 백신 부작용 보고서 분석, 의료기록을 통한 수술부위 감염 위험요소 평가 등과 같이 보건의료분야에도 적용되고 있다.[11-13]

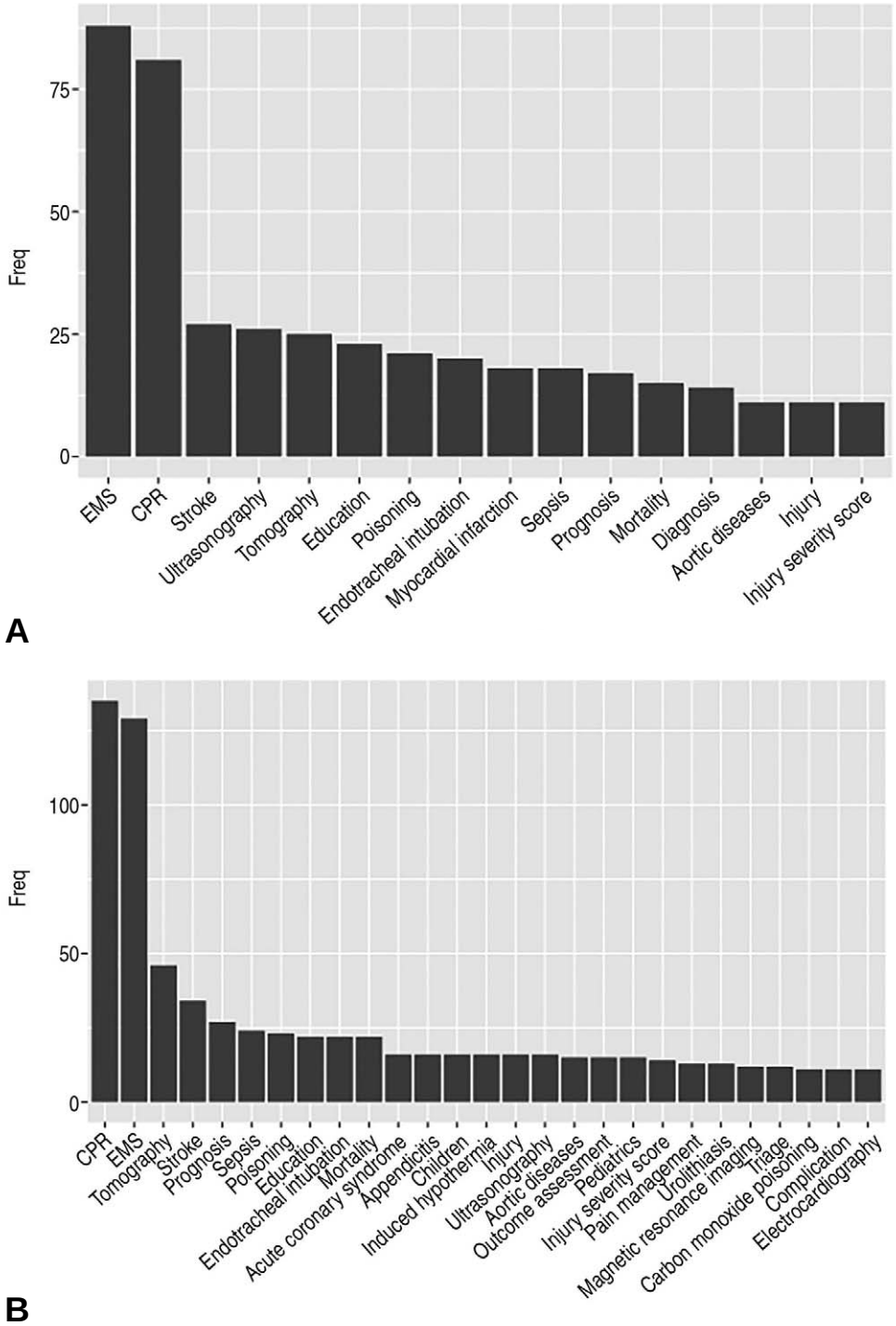


**Fig. 4.** Visualize core keyword frequency from 2006 to 2010 (**A**) and from 2011 to 2016 (**B**). EMS, emergency medical services; CPR, cardiopulmonary resuscitation.

저자들의 연구는 대한응급의학회지에 대해 텍스트 마이닝 방법을 적용하여 국내 응급의학의 연구 추세를 분석한 논문이며, 전반기와 후반기로 나누어 연구 트렌드의 변화뿐만 아니라 대한응급의학회지에서 주로 다루는 연구주제에 대한 전반적인 검토를 수행하였다. 논문분석을 통한 연구 트렌드의 정량화에 대한 시도는 10년 동안의 사회과학 논문을 텍스트 마이닝으로 분석하여 연구 트렌드를 파악하거나, 17년 동안 발행된 과학교육 관련 논문을 텍스트 마이닝하여 연구 트렌드를 분석한 연구 등이 있었다.[14,15] 이렇듯 여러 분야에서 텍스트 마이닝을 이용한 연구가 시도되고 있지만 아직까지 대한응급의학회지에 적용한 사례가 없었다.

저자들은 2006년부터 2016년까지 총 1,037편의 대한응급의학회지 논문을 대상으로 3,952건의 중심단어를 추출하여 텍스트 마이닝을 통해 정량적으로 분석함으로써 전후반기에 따라 연구 트렌드에 변화가 있음을 확인하였다. 추출된 중심단어의 수는 전반기에 논문 평균 3.47개에서 후반기 4.10개로 통계적으로 의미 있게 증가된 것으로 나타났다. 이 결과는 Cho와 Lee[7]가 대한응급의학회의 중심단어의 오류를 분석한 연구결과에서 시간이 경과됨에 따라 인용된 중심단어의 수가 증가됨을 발표한 연구결과와 일치하며, 이는 응급의학과 연구가 활성화됨에 따라 중심단어의 중요성을 연구자가 인식하고, MeSH 등재기준을 엄격하게 적용함에 따라 나타난 결과로 판단된다.

추출된 중심단어의 워드 클라우드 분석에서는 전반기와 후반기 모두에서 응급의료체계와 심폐소생술의 단어구름 크기가 가장 크게 나타나 가장 활발히 연구되는 주제로 나타났다. 이는 시각화를 통한 거시적인 안목으로 봤을 때, 우리 응급의학회는 국내에 유일한 응급의료체계를 담당하는 의학 단체로서 응급의료와 관련된 핵심어가 부각될 수밖에 없었다라고 생각된다. 그러나 상대적으로 전반기에는 응급의료체계가 심폐소생술보다 크게 나타났으며, 아울러 진단, 초음파 등의 단어의 선택빈도가 높았다. 이와 비교하여 후반기에는 심폐소생술이 응급의료체계보다 선택이 다소 많았으며, 예후, 컴퓨터단층촬영, 사망의 선택이 상대적으로 높게 나타났다. 이는 2006년 이후부터 병원 밖 심정지 환자에 대한 응급의학 영역에서의 진단, 치료 및 예후 연구가 증가된 영향으로 생각한다. 아울러 그 외 다른 후순위 중심단어들도 응급의학 연구의 발전 및 시대의 흐름에 따라 다변화함을 시각적으로 확인할 수 있었다.

중심단어의 빈도분석에서도 시간의 흐름에 따른 연구주제의 변화를 확인할 수 있었는데, 전반기에서는 중심단어로 초음파의 선택빈도가 상대적으로 많았으나 후반기에는 상대적으로 감소되었으며, 컴퓨터단층촬영, 패혈증, 사망 등의 빈도가 후반기에 상대적으로 증가되는 경향을 보였다. 이는 연구 전반기에 응급센터에 초음파 장비가 점차 보급됨에 따라 연구가 활발히 진행되었음을 의미하며 후반기에는 컴퓨터단층촬영 기술의 발전으로 인한 관련 연구와 패혈증 및 중증 응급환자 관련 연구가 증가되었음을 시사한다. 아울러 뇌졸중, 중독, 교육, 기도삽관, 손상 등의 중심단어는 전후반기 모두에서 비슷한 빈도로 나타남으로 관련 연구가 지속적으로 시행되고 있음을 유추할 수 있었고, 최소 10회 이상 언급된 중심단어가 전반기에 비해 연구 후반기에 늘어났음을 통하여 응급의학 연구 영역의 확장 및 다

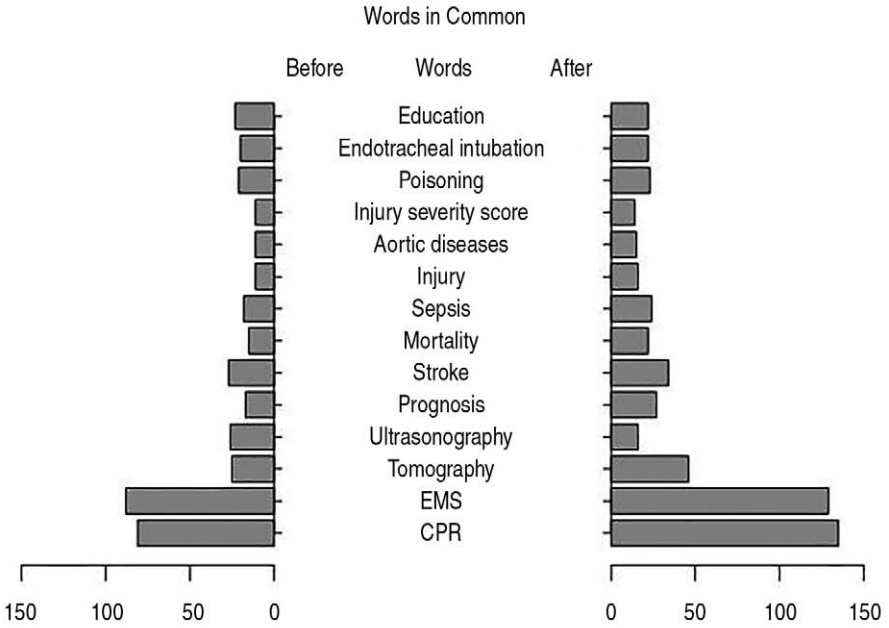


**Fig. 5.** Use the pyramid graph to compare common words from 2006 to 2010 and from 2011 to 2016. EMS, emergency medical services; CPR, cardiopulmonary resuscitation.

양성을 파악할 수 있었다.

이번 연구는 대한응급의학회지의 중심단어를 대상으로 한 관찰연구로서 다음의 제한점을 가지고 있다. 첫째, 이 연구는 대한응급의학회지에 발표된 논문만을 대상으로 하였기에 다른 국내 저널에 출판되었거나, 해외 저널에 출판된 연구논문을 포함하지 못하고 있다. 최근 해외 저널에 출판되는 국내 연구자들의 논문이 점차 증가되는 현시점에서 관련 연구 트렌드 변화를 유추하기에는 제한적이라 판단된다. 둘째, 연구 대상의 전후반기의 기간과 출판 논문 수에 차이가 있어 단순 빈도분석만으로 연구 주제의 변화를 예측하는 데에는 한계가 있다. 그러나 워드 클라우드 시각화 결과와 함께 연구 주제의 트렌드 변화를 파악하는 것에 임상적 의미가 있다 판단된다.

요약하자면, 본 연구는 대한응급의학회지의 중심단어를 텍스트 마이닝을 이용하여 전후반기 기간별로 비교분석하였으며, 응급의학 관련 임상연구 주제의 트렌드 변화를 워드 클라우드를 이용한 시각화로 확인하였다.

연구기간의 전후반기 공통으로 응급의료체계와 심폐소생술이 많이 언급이 되었고 전반기의 경우 진단, 초음파가 많았고 후반기에서는 예후, 컴퓨터단층촬영, 사망의 선택빈도가 높게 나타났다.

중심 단어의 전후반기 시간적 트렌드의 변화는 대한응급의학회의 연구 트렌드의 변화를 잘 나타낸다고 판단되며 대한응급의학회의 연구분야를 가늠할 수 있는 논문 속의 중심단어 분석을 통하여 과거에 비해 그 연구 영역이 다양해지고 세분화됨을 알 수 있었다.

## ORCID

Youdong Sohn (https://orcid.org/0000-0001-8789-0090)

## CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

## REFERENCES

1. Lee JK, Kwon SB, Im KB. Principles of management information systems. 3rd ed. Seoul: Bubyongsa; 2011.
2. Korea National Oil Corporation. Business information [Internet]. Ulsan: Korea National Oil Corporation; 2018 [cited 2018 Jun 10]. Available from: http://www.knoc.co.kr/sub03/sub03_6_1.jsp.
3. Seoul Public Transportation. Bus information system [Internet]. Seoul: Seoul Metropolitan Government; 2018 [cited 2018 Jun 15]. Available from: http://bus.go.kr/UseInfo.jsp?main=1.
4. Korean Association of Medical Journal Editors. Member journal [Internet]. Seoul: Korean Association of Medical Journal Editors; c2018 [cited 2018 Oct 18]. Available from: https://www.kamje.or.kr/association/list_association.
5. Rosen P, Barkin RM. Emergency medicine: concepts and clinical practice. 3rd ed. St. Louis: Mosby Year Book; 1992.
6. Cho CJ. Usage of MeSH. J Korean Acad Fam Med 2000;21:S277-85.
7. 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;20:722-8.
8. Wickham H. rvest: Easily Harvest (Scrape) web pages. R package version 0.3.2 [Internet]. The Comprehensive R Archive Network; 2016 [cited 2018 Oct 18]. Available from: https://CRAN.R-project.org/package=rvest.
9. Fellows I. Word Clouds [Internet]. Comprehensive R Archive Network; 2018 [cited 2018 Oct 18]. Available from: https://CRAN.R-project.org/package=wordcloud.
10. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna: R Foundation for Statistical Computing; 2013 [cited 2018 Oct 18]. Available from: https://www.R-project.org/.
11. Lucini FR, Fogliatto FS, da Silveira GJ, et al. Text mining approach to predict hospital admissions using early medical records from the emergency department. Int J Med Inform 2017;100:1-8.
12. Botsis T, Nguyen MD, Woo EJ, Markatou M, Ball R. Text mining for the Vaccine Adverse Event Reporting System: medical text classification using informative feature selection. J Am Med Inform Assoc 2011;18:631-8.
13. Michelson JD, Pariseau JS, Paganelli WC. Assessing surgical site infection risk factors using electronic medical records and text mining. Am J Infect Control 2014;42:333-6.
14. Peng TQ, Zhang L, Zhong ZJ, Zhu JJ. Mapping the landscape of internet studies: text mining of social science journal articles 2000-2009. New Media Soc 2012;15:644-64.
15. Chang YH, Chang CY, Tseng YH. Trends of science education research: an automatic content analysis. J Sci Educ Technol 2010;19:315-31.