



Meisam Dastani¹, Alireza Atarodi¹, Saeed Erfanpoor^{2,*}

¹Education Development Center, Gonabad University of Medical Sciences, Gonabad, Iran

²PhD student, Department of Epidemiology, Iran University of Medical Sciences, Tehran, Iran

*Iran University of Medical Sciences, Shahid Hemmat Highway Tehran, 14496-14535 Iran

Tel: +98 9153584458
Fax: +98 21 8862 2692
Email: erfanpoor.s@iums.ac.ir

ORIGINAL ARTICLE

A Study of Research Trends in Iranian Journals of Medical Education

Background: Nowadays, with the increasing number of articles in the field of medical education, the analysis of the trend of articles published in specialized journals in the field of medical education becomes important for adopting necessary measures in this field. For this purpose, the present research studies the trend of articles published in Iranian journals of medical education.

Methods: A descriptive exploratory approach has been used in this study to analyze the articles published in Iranian journals of medical education from 2004 to 2018 using text mining techniques. In this study, the Python programming language and NLTK toolkit have been used for text mining.

Results: The findings of this study indicated that the terms such as students, medicine, education, learning, and university are the most repetitive terms in these articles. The results of the implementation of k-means clustering algorithm have revealed that the main theme and sentence of the five thematic clusters in this study is the term "education" or "medical student education" and the subjects of the most of articles published in medical journals are in the field of education.

Conclusion: The results of the present study indicated that the articles published in Iranian journals of medical education focus on general subjects of medical education such as curriculum, medical students, and teaching in the academic environment; however, other issues have been less discussed.

Keywords: Education, Medical, Trends, Data mining, Iran

دراسة اتجاهات البحوث المنشورة في مجلات التعليم الطبي الإيرانية

الخلفية والهدف: اليوم، بسبب تزايد نشر المقالات في مجال التعليم الطبي، فإن تحليل اتجاه المقالات المنشورة في المجلات المتخصصة في مجال التعليم الطبي لاتخاذ الإجراءات اللازمة في هذا المجال أمر مهم وضروري. لهذا الغرض، درست هذه الدراسة اتجاه المقالات المذكورة.

الطريقة: تستخدم هذه الدراسة نهجاً استكشافياً ووصفياً لتحليل الأبحاث المنشورة في مجلات التعليم الطبي الإيرانية من ٢٠٠٤-٢٠١٨ باستخدام تقنيات التنقيب عن النصوص. في هذا البحث، تم استخدام لغة برمجة Python ومكتبة NLTK لإجراء عمليات التنقيب عن النصوص.

النتائج: أظهرت نتائج هذه الدراسة أن كلمات الطلاب، الطب، التعليم، التعلم والجامعة هي أكثر الكلمات شيوعاً في هذه المقالات. تُظهر نتائج خوارزمية تجميع k-means أن الموضوع الرئيسي واقترح المجموعات المواضيعية الخمس التي تم الحصول عليها في هذه الدراسة هو عبارة "تعليم" أو عبارة "تعليم طلاب الطب" و كان موضوع معظم المقالات المنشورة هو التعليم.

الخلاصة: تظهر نتائج الدراسة الحالية أن المقالات المنشورة في مجلات التعليم الطبي الإيرانية كانت أكثر حول الموضوعات العامة للتعليم الطبي مثل المناهج وطلاب الطب والتدريس في بيئة الجامعة والمواضيع الأخرى لم يتم تناولها بشكل أقل.

الكلمات المفتاحية: التعليم، الطب، العملية، التنقيب عن البيانات، إيران.

مطالعه روند پژوهش های منتشر شده در مجلات آموزش علوم پزشکی ایران

زمینه و هدف: امروزه با توجه به افزایش انتشار مقالات در حیطه آموزش پزشکی، تجزیه و تحلیل روند مقالات منتشر شده در مجلات تخصصی حیطه آموزش پزشکی جهت اتخاذ اقدامات لازم در این حیطه امری مهم و ضروری است. به همین منظور این پژوهش به مطالعه روند مقالات های منتشر شده در مجلات آموزش علوم پزشکی ایران پرداخته است.

روش: این مطالعه با استفاده از رویکرد اکتشافی و توصیفی به تجزیه و تحلیل پژوهش های منتشر شده در مجلات آموزش علوم پزشکی ایران از سال ۲۰۰۴-۲۰۱۸ با استفاده از فنون متن کاوی پرداخته است. در این پژوهش جهت انجام عملیات متن کاوی از زبان برنامه نویسی پایتون و کتابخانه NLTK استفاده شده است.

یافته ها: یافته های این پژوهش نشان داد که واژگان دانشجویان، پزشکی، آموزش، یادگیری و دانشگاه پرتکرار ترین واژگان در این مقالات هستند. نتایج حاصل از اجرای الگوریتم خوشه بندی کامینز نشان داده است که تم و گزاره اصلی پنج خوشه موضوعی بدست آمده در این پژوهش، واژگان "آموزش" و یا عبارت "آموزش دانشجویان پزشکی" است و بیشتر موضوع مقالات منتشر شده مجلات حوزه علوم پزشکی بر محور آموزش متمرکز بوده است.

نتیجه گیری: نتایج مطالعه حاضر نشان می دهد مقالات منتشر شده در مجلات آموزش علوم پزشکی ایران بیشتر در موضوعات عمومی آموزش پزشکی مانند برنامه درسی، دانشجویان پزشکی و تدریس در محیط دانشگاهی بوده است و به سایر موضوعات کمتر پرداخته شده است.

واژه های کلیدی: آموزش، پزشکی، روند، داده کاوی، ایران

ایران میں میڈیکل رسالوں میں چھپنے والے تحقیقاتی مقالوں کے بارے میں تحقیق

بیک گراؤنڈ: آج کل میڈیکل کی تعلیم کے حوالے سے کافی تعداد میں تحقیقاتی مقالے چھپ رہے ہیں۔ ان تحقیقاتی مقالوں کا تجزیہ کرنا تاکہ آئندہ کے اقدامات میں ان سے مدد لی جا سکے ضروری لگتا ہے۔ اسی وجہ سے ایران کے میڈیکل سائنس رسالوں میں چھپنے والے علمی مقالوں کا جائزہ لیا گیا ہے۔

روش: اس تحقیق میں ایران میں چھپنے والے میڈیکل سائنس کے رسالوں میں چھپنے والے تحقیقاتی مقالوں کا جائزہ لیا گیا ہے۔ ان تحقیقاتی مقالوں کا جائزہ لیا گیا ہے جو دوہزار چار سے دوہزار اٹھارہ تک شایع ہوئے ہیں۔ اس تحقیق کے لئے پینتھون اور این ایل ٹی کے سافٹ ویئر سے استفادہ کیا گیا تھا

نتیجے: اس تحقیق سے معلوم ہوتا ہے کہ طلباء، میڈیکل، تعلیم اور پڑھائی نیز دانشگاه کے الفاظ تمام مقالوں میں کثرت سے استعمال ہوئے ہیں۔ kmeans خوشہ دار الگوریتم سے یہ معلوم ہوا ہے کہ پانچ موضوعات میں یہ الفاظ سے زیادہ دیکھنے کو ملے جیسے تعلیم، میڈیکل طلباء کی تعلیم۔ اس سے معلوم ہوتا ہے کہ ایرانی میڈیکل رسالوں میں جو تحقیقاتی مقالات شایع ہوئے ہیں ان کا موضوع بیشتر میڈیکل طلباء کی تعلیم تھا

سفرارش: جیسا کہ ہم نے اوپر بتایا ہے کہ ان تحقیق سے واضح ہوتا ہے کہ شایع شدہ مقالوں میں میڈیکل طلباء کی تعلیم پر کافی زور دیا گیا ہے لہذا ان مسائل کے علاوہ دیگر متعلقہ مسائل پر بھی توجہ دینے کی ضرورت ہے۔

کلیدی الفاظ: تعلیم، میڈیکل، علم طب۔

INTRODUCTION

Many scientific documents which are annually produced by researchers, professors, and students from different universities around the world, often contain important and useful content and are published in textual formats. Due to the increasing number of scientific articles and the enormous volume of published articles, evaluating and reviewing every single article and manually extracting information and knowledge from the volume of texts is exhausting, impossible, and challenging. However, identifying patterns and extracting potential knowledge in a large volume of textual data is an important task in various scientific fields (1). Therefore, the presentation of tools and techniques that can automatically analyze the texts has led to the development of text mining, which covers all activities of extracting knowledge from the text (1). Text mining is also known as intelligent text analysis, text data mining, and knowledge discovery from text and is generally referred to as the process of extracting important and desired knowledge and information from a set of unstructured text (2-4). Salloum describes the objectives of research in the field of text mining as follows: 1) Identifying the subjects of scientific texts and development of these topics using text mining techniques; 2) Applying visualization tools to present each topic and the relationship between them as a useful method to assist users in determining relevant topics (5,6). In various research, the text mining techniques have been used to analyze articles published in scientific journals and databases, including the trend of publishing articles in scientific journals; however, the most important issues are related to insomnia and disorders (7), identifying the most important Indian medicinal plant for the treatment of diabetes (8), identifying a comprehensive perspective on the evolution and development of information sciences (9), analyzing the evolution of some research on mobile health (Mhealth) (10), evaluating the abstract of articles presented in international AIDS conferences for over 25 years to identify the trend of using the term "HIV" (11), identifying the trend of some research on medical informatics to understand the current situation of medical informatics, the leading path and recognizing the limitations in this field (12), identifying the trend of some research on anesthesia from the reviews of articles presented in authorized conferences (13), and analyzing the articles in the field of health information (14). Published scholarship in medical education began in 1950 when the Journal of the Association of American Medical Colleges transformed itself into the Journal of Medical Education. Yet, the early history of medical education as a scholarly field was almost languid. It was not until 1966 that a second journal dedicated to medical education appeared, the British Journal of Medical Education (15). Nowadays, there are many changes in the curriculum and innovation in the field of medical education at the national and international levels. The education and assessment methods, and the topics related to accreditation at the national level have increased the interest in medical education because of rapid expansion of the number of medical schools to meet the demands of all stakeholders in society.

As a result of the increase in the number of medical schools and the consequent development of medical education in various fields to meet the increasing needs of society, significant growth was also observed in the number of journals and articles related to this field (16, 17). Also, Research subjects in medical education are incredibly diverse and have changed over the time (18). Therefore, the analysis of the trend of articles published in specialized journals in the field of medical education began to adopt necessary measures in this important field to improve medical education at the global level (19,20), so that it can identify important research areas and the progress of this field (21). The study of trends in medical education analyzes the entire academic field or the subject of a particular academic field. Accordingly, in order to identify the trend of medical education research, Kim analyzed the words used in articles in journals of Academic Medicine (the journal of the Association of American Medical Colleges), Medical Teacher (the journal of the Association of Medical Education in Europe), the Korean Journal of Medical Education (KJME), and Korean Medical Education Review (KMER) (18). In another study, Rotgans reviewed research articles published in six influential medical education journals (20). Yoo and shin analyzed the keywords of research papers published in the Korean Journal of Medical Education (KJME) (19). Also, other studies have been conducted with the aim of identifying important articles and journals in the field of medical education by bibliometric and scientometric methods (21-24).

There are also several specialized journals in the field of medical education in Iran that publish specialized articles in this area. Because of the extensive range of research in this field, it is attempted in the present study to identify the thematic trend of published articles in Iran by reviewing and analyzing the articles published in Iranian journals of medical education.

METHODS

A descriptive exploratory approach has been used to analyze the articles published in Iranian journals of medical education. The text mining method used in the present study is extracted from the framework designed by Zhang & Chen (25) and developed by Salloum et al. (5) (Figure 1), and included text pre-processing, text mining, and text post-processing. The text pre-processing involved data selection, classification, feature extraction, and normalization. The second stage dealt with various text mining techniques such as clustering, visualization, and repetition of terms. During the third stage, some changes and transformations were made to data (research articles) through text mining functions such as assessment, selection, analysis, and visualization of knowledge. The main purpose of this study was to extract knowledge from the collected articles using text mining techniques.

The following steps of the mentioned framework have also been used in the present study:

First, the list of Iranian journals in this field was extracted from the MoH Medical Journals (<http://journals.research.ac.ir/>) (Table 1), then the English abstracts of the articles published



Figure 1. The text mining framework used in the present study

in Iranian journals of medical education from early 2004 to the first two months of 2019 have been extracted in BIBTEX format (In the websites of journals studied in this article, the journals data can be extracted in XML format for each issue, and it cannot be extracted in annual basis; hence, the data for each journal has been extracted annually in BIBTEX format) by annual and direct referring to journal websites, and then they were converted to CSV format in JabRef software. The following operations have been performed on the data (abstracts of extracted articles) to conduct text mining techniques:

1. Pre-processing of texts that include the following operations:

Data selection, classification, feature extraction, and normalization or converting the documents to mid-phase form for ensuring the compatibility with different extraction tools.

2. The frequency of words, clustering, and visualization have been achieved by different techniques of text mining.

3. The extraction of knowledge from texts and its interpretation.

In the present study, the k-means algorithm has been used

for clustering. Also, different criteria have been applied to evaluate and determine the number of data clusters. Silhouette and Sum of Squared Errors (SSE) criteria have been used in the present study (26).

Finally, the results were interpreted according to the information obtained from the analysis of the data and articles. Python programming language and NLTK (Natural Language Toolkit) library have been applied to implement the text mining algorithms.

Ethical approval was received for this study from the Ethics Committee of Gonabad University of Medicine Sciences (IR.GMU.REC.1398.001).

RESULTS

In this study, 2709 records were extracted from the published articles in Iranian journals of medical education by annual and direct referring to journal websites in BIBTEX format (one file per year in each journal of articles), all of these files have been imported in JabRef software and finally they were converted to CSV files, and text mining algorithms, as well as techniques were implemented on the data of this file.

Then, the articles with null abstracts were removed from the list of articles with beginning of record analysis. The number of reviewed abstracts and records is given in Figure 2 by the year.

Since the extraction of articles was based on the end of the solar year, a number of articles from the first two months of 2019 have also been extracted.

The results of the research on the frequency of words are as follows, so the frequency of most repeated 30 words is indicated in Table 3, and the word cloud is indicated in Figure 3.

Table 3 shows that the words: student with 7867 repetitions, medical with 4202 repetitions, and education with 3358 repetitions are the most important words in the articles.

Figure 3 indicates the most important and commonly used words of all articles. Word clouds provide a unique way to summarize the content of text documents. In a word cloud, the size of a word in the cloud is proportional to its importance and application in the whole text collections (27).

Table 1. The list of Iranian journals of medical education reviewed in this study

Journals	URL
Biannual Journal of Medical Education Education Development Center	mededj.ir
Education Strategies in Medical Sciences (ESMS)	edcbmj.ir
Future of Medical Education Journal (FUTURE MED EDUC J)	fmej.mums.ac.ir
Iranian Journal of Medical Education	ijme.mui.ac.ir
Journal of Development Strategies in Medical Education	dsme.hums.ac.ir
Journal of Medical Education and Development (JMED)	jmed.ssu.ac.ir
Journal of Medical Education Development	zums.ac.ir/edujournal
Research in Medical Education	rme.gums.ac.ir
Strides in Development of Medical Education	sdmejourn.com

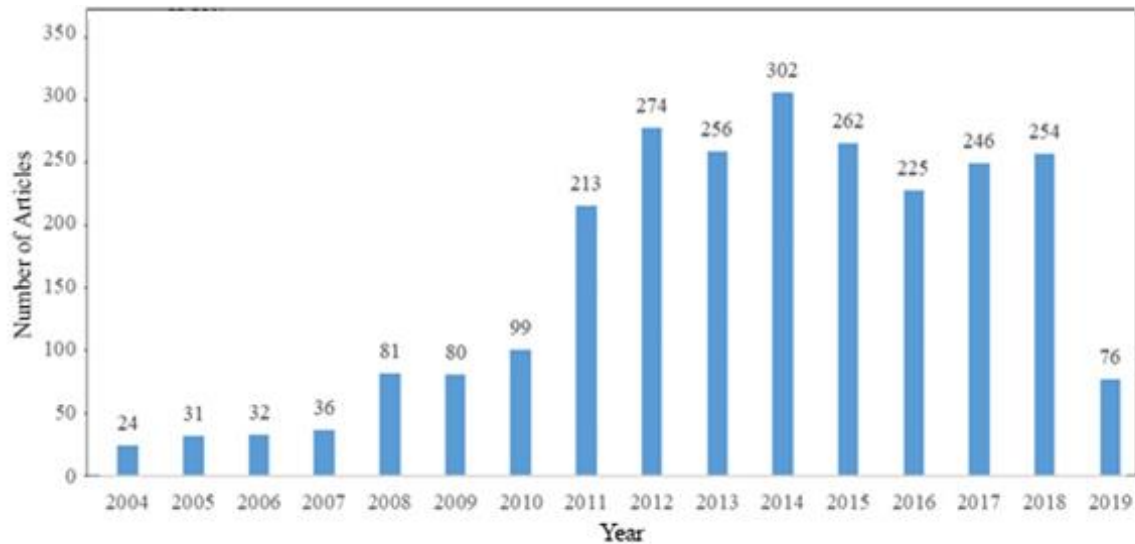


Figure 2. Frequency of reviewed abstracts by the year

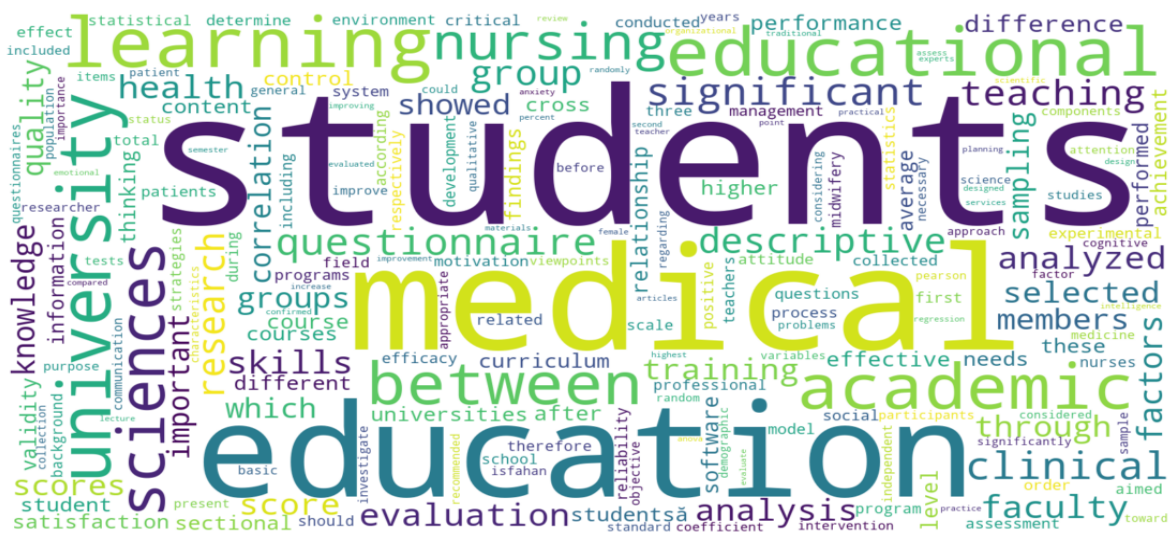


Figure 3. Word cloud of most repeated words in published articles

Table 3. The frequency of most repeated 30 words in published articles

Row	Keywords	Number	Row	Keywords	Number
1	students	7867	16	descriptive	1661
2	medical	4202	17	group	1656
3	education	3358	18	faculty	1442
4	learning	3173	19	health	1437
5	educational	2980	20	skills	11435
6	university	2795	21	analyzed	1360
7	academic	2577	22	evaluation	1338

Table 3. Continued

Row	Keywords	Number	Row	Keywords	Number
8	between	2242	23	factors	1298
9	sciences	2206	24	analysis	1223
10	nursing	2168	25	score	1215
11	significant	2047	26	through	1170
12	clinical	2029	27	selected	1161
13	questionnaire	1919	28	training	1153
14	research	1704	29	scores	1120
15	teaching	1703	30	members	1111

Table 4. The coefficients of silhouette criterion for the number of different clusters (between 2 and 9 clusters)

Number of Clusters	2	3	4	5	6	7	8	9
Silhouette Score	0.349	0.386	0.370	0.367	0.390	0.376	0.365	0.349

In order to cluster the data and determine the optimal number of clusters, the silhouette criterion has been used, the results of which are shown in Table 4.

Since the values of silhouette coefficients approach zero and clusters are overlapped, the SSE criterion was applied to determine the number of desirable clusters (28). Elbow graphs were also plotted to determine the number of clusters (Figure 4).

Elbow proposes several ways to determine k as the number of dynamically formed clusters, one of which is the elbow method (29).

The graph shows the elbow n obtained from all k values. The best value of k will be found in the present study by using the elbow method. The elbow method is easy to be implemented by looking at the ideal k value of the graph with the position on the elbow along with the SSE, which is less than 1.

The best result of cluster k will be the basis for clustering. The smaller the value of SSE and the decrease of elbow graph, the better cluster results will be obtained (30).

Figure 4 indicates that five clusters are suitable for the data of the present study.

The results of implementing the k-means clustering algorithm are shown in Table 5.

The results of the clustering indicate that the articles have been divided into five thematic clusters, and the most important words of each cluster have also been shown.

Table 5 shows that in cluster zero, the most important words of those articles are education with 2705 repetitions, medical with 1602 repetitions, and faculty with 1044 repetitions.

in cluster one, the most important words of those articles are students with 3766 repetitions, medical with 1528 repetitions, and education with 1463 repetitions.

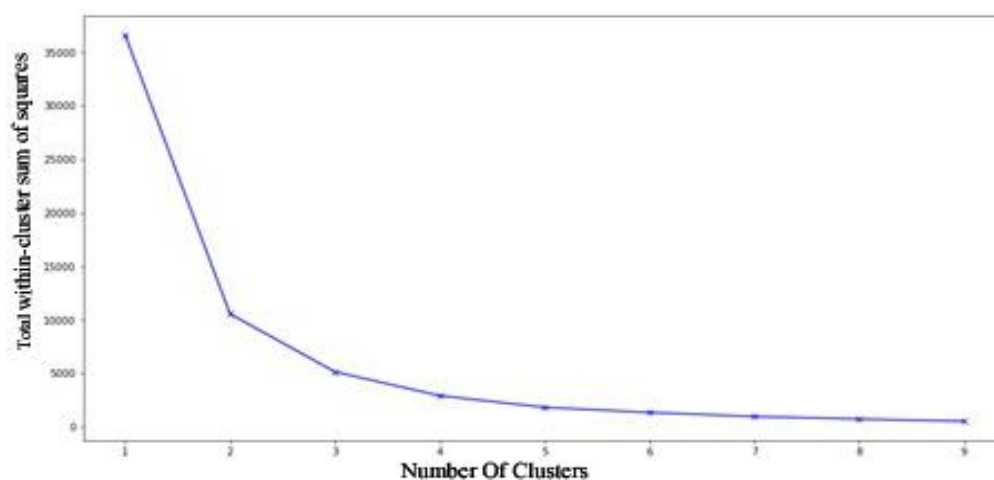
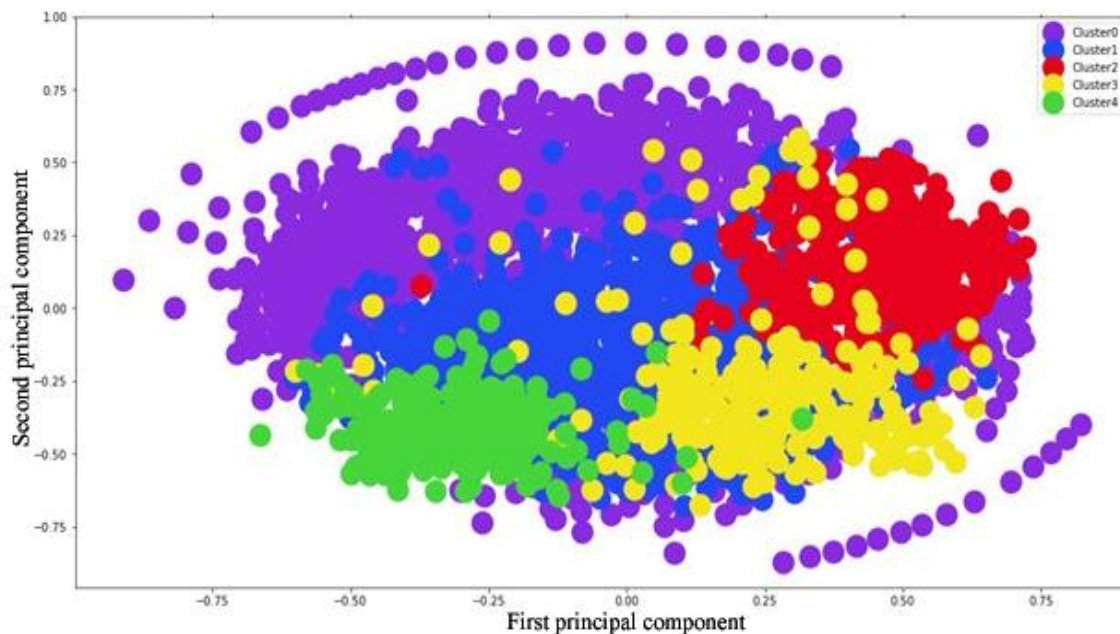


Figure 4. The SSE criterion for selecting the number of clusters

Table 5. The thematic clusters with important and commonly used words of each cluster

Clusters	Keywords
Cluster 0	Education (2705), medical(1602), faculty(1044), university(945), students(883), research(873), members (835), sciences (764), questionnaire (653), evaluation (598), quality (597), teaching(581), descriptive (533), learning (515)
Cluster 1	Students (3766), medical (1528), education (1463), score (1183), group (1126), university (967), sciences (809), between (762), skills (744), significant (720), questionnaire (688), descriptive (544), teaching (489), clinical (478), learning (474)
Cluster 2	Nursing (1596), clinical (1177), education (1253), students (1101), group (507), nurses (465), skills (401), learning (302), medical (290), between (272), significant (254), descriptive (244), university (242), training (245), questionnaire (233)
Cluster 3	Learning (1756), students (985), medical (441), teaching (396), education (556), group (430), significant (249), between (248), university (243), sciences (208), nursing (190), style (313), through (148), questionnaire (146), strategy (155)
Cluster 4	Academic (1872), students (1232), between (477), achievement (465), university (398), significant (396), medical (341), correlation (340), relationship (308), performance (293), sciences (251), education (361), efficacy (237), motivation (231), learning (235)

**Figure 5. The rate of scattering, similarity, and overlapping of clusters**

in cluster two, the most important words of those articles are nursing with 1596 repetitions, clinical with 1177 repetitions, and education with 1253 repetitions.

in cluster three, the most important words of those articles are learning with 1756 repetitions, students with 985 repetitions, and medical with 441 repetitions.

in cluster four, the most important words of those articles are academic with 1872 repetitions, students with 1232 repetitions, and between with 477 repetitions.

Figure 5, which is the result of implementing the k-means clustering algorithm, shows the classification and rate of overlapping of clusters.

In Figure 5, the points in cloud map indicates an overview of the space of thematic clusters, in which the similar documents are located in the proximity together, and each point represents a document, and the themes are also marked with a specific color.

Figure 6 also shows the frequency of articles in each cluster. As shown in the graph, more than 35%, 28.19%, 15.14%, 10.65%, and 10.49% of articles are classified in clusters zero, one, two, three, and four, respectively.

Following is the word cloud of the frequency of each cluster shown in Figure 7.

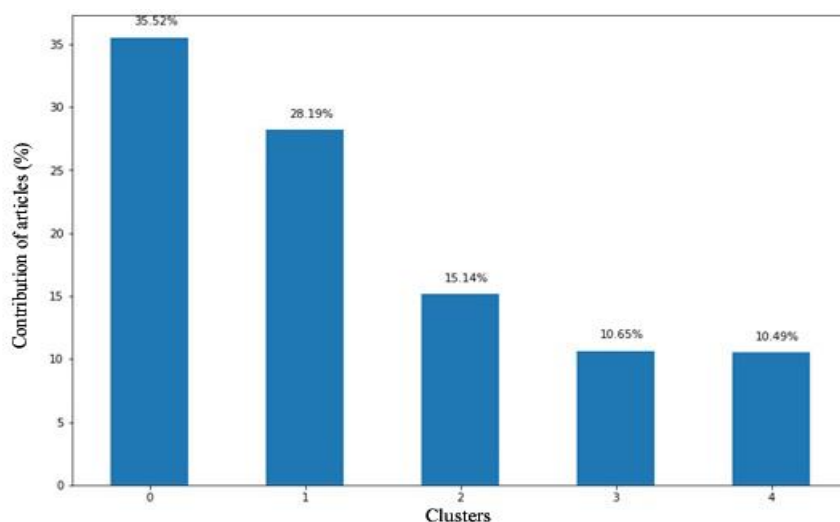


Figure 6. The frequency of articles in each cluster

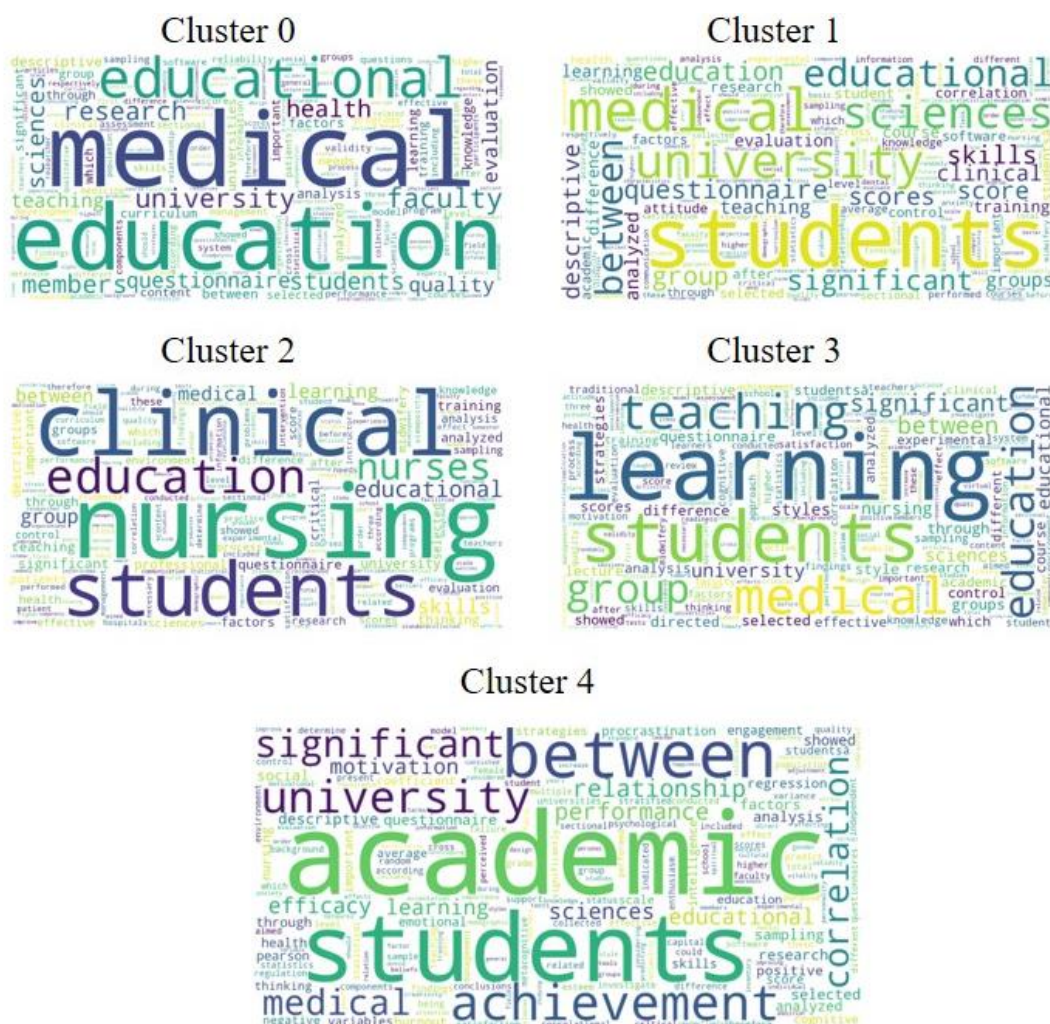


Figure 7. The word cloud of thematic clusters

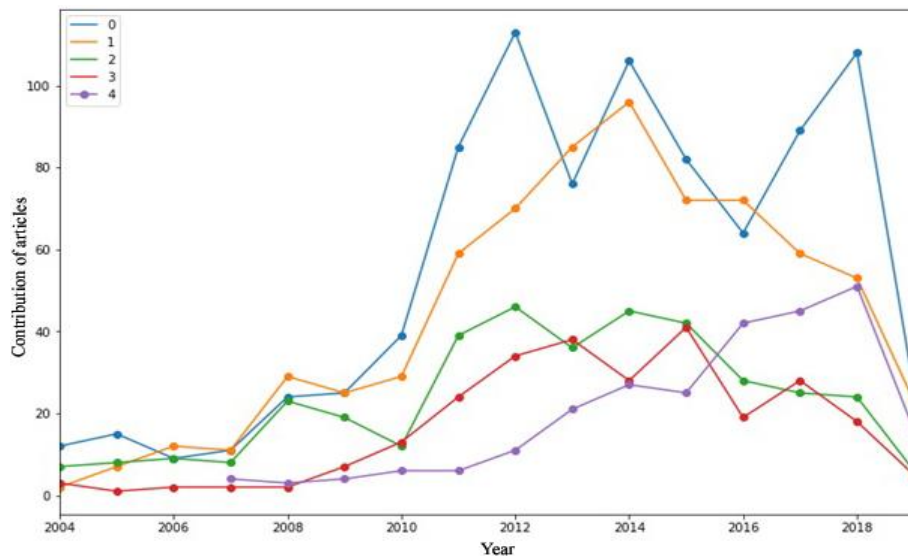


Figure 8. The comparison of publishing trend of five thematic clusters

Figure 7 indicates the most important and commonly used words of each cluster.

Figure 8 shows the publishing trend of thematic clusters of articles over the past 15 years.

Figure 8 shows that the zero thematic cluster possessed the highest number of published articles, and the number of articles published in this field reached the highest number of published articles in 2012.

Thematic cluster one also possessed the highest number of published articles in 2014, and the rate of publishing articles in this field was in descending trend over time. Thematic clusters two, three, and four had fewer articles compared to zero and one clusters. Furthermore, clusters zero and four possessed an ascending trend compared to previous periods.

DISCUSSION

Research in the field of medical education remains insufficient in quantity and quality despite advances over the past two decades (31). The research topics which are continually improving are coordinated with numerous social developments in medical education, as well as they are considerably diverse and change over time (18, 32). Therefore, the use of text mining techniques for investigation and analysis of scientific texts in this field is necessary (33).

In this regard, the results of text mining of articles published in Iranian journals of medical education in the present study revealed that the words student, medical, education, learning, educational, university, academic, between, science, and nursing are ten commonly used words in the articles published in these journals over the past 15 years. Yoo and Shin showed that the most central and commonly used words in the Korean Journal of Medical Education (KJME) were "medical education," "clinical competence," "medical student," and "curriculum" (19). The results of a study conducted by Kim et al. also indicated that the terms "medicine," "student," "education," and "learning" were more repeated and commonly used words among the topics

of Korean journals of medical education (18). Ji et al. showed that the keywords with the highest increase rates in medical education articles, are "Education," "Medical," "Humans," "Curriculum," "Continuing Medical Education," and "Internship and Residency," (32). Lee et al. also showed these keywords in medical education studies (16).

The results of the five thematic clusters obtained according to k-means algorithm showed that the main theme and sentence of the five thematic clusters obtained in this study is the term "education" or "medical student education" and the subjects of the most of articles published in medical journals were focused in the field of education. According to the main content of the topics and keywords, the results indicated that the statistical populations of most articles were students in academic environments. Accordingly, Ershad et al. analyzed the content of the articles in the four Iranian journals of medical sciences and showed that the articles focusing more on students confirm the results of this study (33). Such repetition of research topics, as noted in the study by Eva K. W., probably occurs because studies in medical education are mostly observational (34).

However, the main purposes of medical education are improving the quality of health care, providing health services and health care in accordance with standards, continuous education and maintaining the information of physicians and other relevant persons, and being up-to-date in the provision of health care services (35). However, the results of the present study suggested that researchers are not interested in these areas. General topics in medical education are curriculum, medical students, and teaching (16), and the thematic clusters obtained in this study indicated that articles published in Iranian journals of medical education include most of these topics. The results of a study by Albert et al. demonstrated that budget limitations and a lack of understanding of the concept of medical research are barriers to medical education research. They also suggested strengthening the collaboration

between specialist doctors and clinical practitioners and supporting the various perspectives presented in the field of medical education research improve investigations in this field (36). Obeidat et al. indicated that articles in medical education produced by Saudi researchers are very scattered and do not focus on particular topics. Many articles have discussed education, learning, and assessment. The articles relating to the curriculum was not included topics such as curriculum development or curriculum management strategies. However, particular courses have been described and introduced in many articles (31). The results of the study by Kim also showed that Korean journals of medical education do not cover all topics (18). Medical education research has focused not only on medical knowledge and practice (content) but also on research topics related to the theory of education as social science (pedagogy) (37). Therefore, research on the fields related to social sciences is also required for the development of medical education of a society, which should be pursued across all disciplines in an integrated and interdisciplinary manner, so that the disciplines and studies can reflect each other's requirements (32). This implies that the research topics in medical education resemble the emergence phenomena, as used in the complex phenomena.

In confirmation of the performance of k-means clustering used in this study, clustering methods, especially k-means clustering, have also been employed in previous studies. The performance of the k-means clustering has also been validated; therefore the use of this algorithm has been suggested to be useful for displaying appropriate patterns of specialized texts (38-39). Kim and Delen also used clustering to recognize the trends in articles published in 23 journals in the field of medical informatics (12). Salloum et al. also applied text mining to analyze scientific articles and extract frequently used words in articles and used the k-means clustering algorithm to identify the topics (5) that support the methodology of the present study. Since the clustering is the process of organizing elements into groups with similar components, and a cluster is a set of similar elements which are heterogeneous with other components, the purpose of clustering is to obtain reliable and fast access to correlated data and identify the logical relationship between them (40, 41).

The results of the present investigation assist in identifying the orientations and understanding the basic concepts of scientific literature in the field of medical science education in Iran and ultimately help the development of this area. The clustering process in the present study showed the trend of

development and publishing of articles in specialized journals in the field of medical education in Iran. The results of the present study also indicated that articles published in Iranian journals of medical education focused more on general topics of medical education such as curriculum, medical students; however, teaching in academic environments and other topics such as improving the quality of health care, providing health care and services in accordance with standards, continuous education, etc. have been neglected. Therefore, due to Extent of the medical education, it is necessary for researchers to pay more attention to all aspects of medical education. By integrating the findings of this study, researchers will be able to discuss the research topics which have been less addressed in relevant journals before designing and starting a new research work. Also, the policymakers of medical education journals can use the results of this study to formulate priorities for the publication of articles in their journals. One of the limitations of K-means algorithm is determining the number of clusters. It is beneficial to determine the correct number of clusters in the beginning. To solve this problem, SSE criteria and Elbow diagram have been used. Another limitation of the K-means algorithm is the lack of automatic labeling of subject clusters. To solve this problem, the opinions of subject matter experts have been used.

Ethical considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

ACKNOWLEDGEMENT

The present article is extracted from a research project approved by the Research and Development Center of Medical Education of Gonabad University of Medical Sciences with the code A-10-1263-1. The researchers thank the vice president of the Research and Technology of the University who sponsored the project and all those who assisted in conducting the research.

Financial Support: This study was funded by a research grant from the Gonabad University of Medical Sciences. The funding body (GMU) didn't have any role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Conflict of Interest: There was no conflict of interest.

REFERENCES

1. Hashimi H, Hafez A, Mathkour H. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet.* 2012;13(6):395-405.
2. Rebholz-Schuhmann D, Oellrich A, Hoehndorf R. Text-mining solutions for biomedical research: enabling integrative biology. *Nat Rev Genet.* 2012;13(12):829-39.
3. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet.* 2012;13(6):395-405.
4. Rodriguez-Esteban R, Bundschuh M. Text mining patents for biomedical knowledge. *Drug discov today.* 2016;21(6):997-1002.
5. Salloum SA, Al-Emran M, Monem AA, Shaalan K. Using text mining techniques for extracting information from research articles. *Intelligent natural language processing: Trends and Applications.* Cham: Springer; 2018.
6. Salloum SA, Al-Emran M, Abdallah S, Shaalan K. Analyzing the Arab gulf newspapers using text mining techniques. *International Conference on Advanced Intelligent Systems and Informatics 2017 Sep 9; Springer, Cham P.* 396-405.

7. Lam C, Lai FC, Wang CH, Lai MH, Hsu N, Chung MH. Text mining of journal articles for sleep disorder terminologies. *PLoS One*. 2016;11(5).
8. Selvaraj B, Periyasamy S. Indian medicinal plants for diabetes: text data mining the literature of different electronic databases for future therapeutics. *Biomedical Research*. 2016; Special Issue: S430-S436.
9. Yu D, Xu Z, Pedrycz W, Wang W. Information Sciences 1968-2016: A retrospective analysis with text mining and bibliometric. *Information Sciences*. 2017; 418:619-34.
10. Ozaydin B, Zengul F, Oner N, Delen D. Text-mining analysis of mHealth research. *mHealth*. 2017; 3:53.
11. Dancy-Scott N, Dutcher GA, Keselman A, Hochstein C, Copt C, Ben-Senia D, et al. Trends in HIV Terminology: Text Mining and Data Visualization Assessment of International AIDS Conference Abstracts Over 25 Years. *JMIR public health surveill*. 2018;4(2): e50.
12. Kim YM, Delen D. Medical informatics research trend analysis: A text mining approach. *Health informatics journal*. 2018;24(4):432-52.
13. Rusanov A, Miotto R, Weng C. Trends in anesthesiology research: a machine learning approach to theme discovery and summarization. *JAMIA Open*. 2018;1(2):283-93.
14. Saheb T, Saheb M. Analyzing and visualizing knowledge structures of health informatics from 1974 to 2018: A bibliometric and social network analysis. *Healthc Inform Res*. 2019;25(2):61-72.
15. Kuper A, Albert M, Hodges BD. The origins of the field of medical education research. *Academic Medicine*. 2010;85(8):1347-53.
16. Lee K, Whelan JS, Tannery NH, Kanter SL, Peters AS. 50 years of publication in the field of medical education. *Med teach*. 2013;35(7):591-8.
17. Abdulrahman KB, Gibbs T, Harden R. The medical education journey continues. *Med teach*. 2013;35(sup1): S5-7.
18. Kim SY. Trends in the study on medical education over the last 10 years, based on paper titles. *Yeungnam Univ J Med*. 2019;36(2):78.
19. Yoo HH, Shin S. Trends of research articles in the Korean Journal of Medical Education by social network analysis. *Korean J Med Educ*. 2015;27(4):247-54.
20. Rotgans JI. The themes, institutions, and people of medical education research 1988-2010: content analysis of abstracts from six journals. *Adv Health Sci Educ*. 2012;17(4):515-27.
21. Regehr G. Trends in medical education research. *Academic Medicine*. 2004 1;79(10):939-47.
22. Munzer BW, Love J, Shipman BL, Byrne B, Cico SJ, Furlong R, et al. An analysis of the top-cited articles in emergency medicine education literature. *West J Emerg Med*. 2017;18(1):60.
23. Azer SA. The top-cited articles in medical education: a bibliometric analysis. *Academic medicine*. 2015;90(8):1147-61.
24. Sampson M, Horsley T, Doja A. A bibliometric analysis of evaluative medical education studies: characteristics and indexing accuracy. *Academic Medicine*. 2013;88(3):421-7.
25. Zhang Y, Chen M, Liu L, editors. A review on text mining. 6th IEEE International Conference on Software Engineering and Service Science (ICSESS); 2015: IEEE.
26. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*. 1987; 20:53-65.
27. Cui W, Wu Y, Liu S, Wei F, Zhou MX, Qu H. Context preserving dynamic word cloud visualization. *IEEE Pacific Visualization Symposium (PacificVis) 2010 March 2 - 5; Taipei, Taiwan. Hong Kong University of Science and Technology*; P. 121-28.
28. Thinsungnoena T, Kaoungkub N, Durongdumronchaib P, Kerdprasopb K, Kerdprasopb N. The clustering validity with silhouette and sum of squared errors. *Proceedings of the 3rd International Conference on Industrial Application Engineering 2015. 2015 The Institute of Industrial Applications Engineers, Japan*. 2015. P.44-51.
29. Joshi KD, Nalwade PS. Modified K-Means for Better Initial Cluster Centres. *International Journal of Computer Science and Mobile Computing (IJCSMC)*. 2013;2(7):219-23.
30. Syakur M, Khotimah B, Rochman E, Satoto B. Integration K-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conference Series: Materials Science and Engineering*; 2018; 336(1):012017.
31. Obeidat AS, Alhaqwi AI, Abdulghani HM. Reprioritizing current research trends in medical education: a reflection on research activities in Saudi Arabia. *Med teach*. 2015;37(sup1): S5-S8.
32. Ji YA, Nam SJ, Kim HG, Lee J, Lee S-K. Research topics and trends in medical education by social network analysis. *BMC med educ*. 2018;18(1):222.
33. Ershad Sarabi R, Eslaminejad T, Shafian H. Content Analysis of Articles in Medical Education Journals of Iran. *Strides in Development of Medical Education*. 2010;7(2):119-26. Persian.
34. Eva KW. Broadening the debate about quality in medical education research. *Medical education*. 2009;43(4):294-6.
35. Brigley YY, Littlejohns P, McEwen J. Continuing education for medical professionals: a reflective model. *Postgrad Med J*. 1997;73(855):23-6.
36. Albert M, Hodges B, Regehr G. Research in medical education: balancing service and science. *Adv Health Sci Educ*. 2007;12(1):103-15.
37. Capra F. The turning point: Science, society, and the rising culture: Bantam; 1983.
38. Chernyshova G, Smorodin G, Ovchinnikov A, editors. Technique of cluster validity for Text Mining. 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence); 2016: IEEE.
39. Lahbib W, Bounhas I, Slimani Y, editors. Arabic terminology extraction and enrichment based on domain-specific text mining. 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI); 2015: IEEE.
40. Sarstedt M, Mooi E. Cluster Analysis. A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2019.
41. Shirkhorshidi AS, Aghabozorgi S, Wah TY. A comparison study on similarity and dissimilarity measures in clustering continuous data. *PLoS One*. 2015;10(12).