Shiraz E-Medical Journal
Volume 23 · Issue 11 · e123803

OPEN PEER REVIEW

Identifying the Trends of Global Publications in Health Information Technology Using Text-mining Techniques

Meisam Dastani, Hamideh Ehtesham, Zohreh Javanmard, Azam Sabahi, Fateme Bahador

SUBMITTED	REVISED	ACCEPTED	DOI
5 Mar 2022	3 Sep 2022	4 Oct 2022	10.5812/semj-123803

About this report

This open peer review (OPR) report shows the public comments made on this article during peer review, by the reviewers, the associate editor and the editor-in-chief, for every revision round, together with the authors' replies. Comments appear only with the consent of the people who wrote them. Reviewers who chose to sign their report are named; the others are shown as "Referee". Comments are edited for language only, and confidential remarks to the editors are not included.

More about open peer review: brieflands.com/knowledgebase/opr

Contents

1 **Original submission** Editor-in-Chief 2
2 **Revision 1** Adeleh Asadi, Editor-in-Chief 3
Authors' reply to reviewers 4
Reviewer and editor comments 8
3 **Revision 2** Associate Editor, Editor-in-Chief 9
Authors' reply to reviewers 10
Reviewer and editor comments 11
4 **Revision 3** Associate Editor 12
Authors' reply to reviewers 13
Reviewer and editor comments 15

1 Original submission

1 comment · 17 May 2022

Editor-in-Chief

MAJOR REVISION 17 May 2022

Major Revision

2

Revision 1

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Dear Editor

Thank you very much for reviewing our manuscript. We also greatly appreciate the reviewers for their complimentary comments and suggestions.
We have modified the manuscript accordingly, and detailed corrections are listed below point by point and highlighted in manuscript.
We look forward to hearing from you regarding our submission. We would be glad to respond to any further questions and comments that you may have.

Best Regards
Fatemeh Bahador
Corresponding Author

Reviewer's comments	Answers
Reviewer #1	
The Author has used the research method of the following article: Kim Y-M, Delen D. Medical informatics research trend analysis: A text mining Approach. Health informatics journal. 2018; 24 (4): 432-52. And the authors also used "medical informatics" as a key search term in the PubMed database. So, a logical issue here: "what is the difference between these two articles in terms of purpose" And "Why the author did not use the medical informatics category from the JCR database. In this paper "Telemedicine" and "Managing Health Information" was also used as a search keyword. How was these keywords selected? The choice of a narrower term "telemedicine" can lead to a biased result.	In the present study, in general, articles related to the application of information technologies in the field of health are considered and the selected keywords are selected with the aim of comprehensive search.
The research method section of the paper was very short and had no transparency. The author should explain more about method details. For instance, he used two databases "PubMed" and "Scopus citation database". In the data preprocessing section, it appears that punctuation erasure and word deletion have not been used and can negatively impact the search results. Also, there are no details about merging synonymous words, erasing certain general words and cleaning up the data process. There is a lot of general words like patient,	The methodology was rewritten. Each step in the methodology was separated.

information, based, system … as a most frequently used words of each topic and this has made it difficult to topics Accurate identification. The question here is how the topics were selected and if it was done only by verifying the words most commonly used for each topic or another complementary method was used.	
The last research topics are not at the same level and it was general or narrower in some instances. Also two same topics were seen such as radiotherapy or medical imaging technology and medical image analysis. It is thus necessary to find a balance between the levels of the subjects.	In LED thematic modeling used in this study, a number of topics have been obtained using the UMS algorithm and also evaluation by subject matter experts. Also, the topics are named based on the most important keywords and publication titles. This description was added to the method. So some topics may be similar and aligned.
Reviewer #2	
The writing of articles hasn't been observed, for example in the introduction there isn't paragraphing and in the discussion paragraphing can be better.	The desired corrections were made.
Is the search strategy based on the opinion of the experts?	Yes, given that three of the authors of this study are health information technology experts, the search strategy is based on their opinion. A corresponding explanation was also added to the method.
Keywords search in Scopus doesn't seem appropriate because it has affected the number of articles in this field.	In the present study, in general, articles related to the application of information technologies in the field of health are considered and the selected keywords are selected with the aim of comprehensive search.
Search Strategy don't limited to type of document but in some cases, Author mentioned to only article.	In fact, we considered the scientific publications in the field of information technology in Salat, so we use the word publication in the text. Corrected throughout the text.
There are some misspellings for example data clearing(data cleaning)	The desired corrections were made.
The Text of the result is very brief.	The desired corrections were made.
Some example in discussion isn't good for example mechanical engineering department.	The desired corrections were made.
Reviewer #3	
1. What is the innovation of the present article compared to the article of Feng et al. (2020) Since the thematic focus of both works is the same (Medical Informatics) and both articles use the same methodology (topic analysis-LDA) and even the time span	In the present study, in general, articles related to the application of information technologies in the field of health are considered and the selected keywords are selected with the aim of comprehensive search.

of Feng's article (1991-2018) is longer than the present study. Also, due to the similarity of the findings of both articles, the explanations throughout the discussion section are mainly based on Feng's article.	
2. What is the reason for choosing 2000 as the year of the beginning of data collection in the field of Medical informatics? While in Kim & Delen (2018) we see that 2002 is chosen as the first year of the time span with the justification that “The beginning year, 2002, was chosen because the application of HIT to the medical field became widely utilized in the time-frame”.	Because more keywords are selected for the search, the time period is increased to be more comprehensive.
3. Has the search strategy used for the Scopus database been approved by HIT experts in terms of reliability in achieving the maximum amount of precision and recall?	Yes, given that three of the authors of this study are health information technology experts, the search strategy is based on their opinion.
What was the purpose of selecting two different databases (Scopus and PubMed) for data collection? If you choose two databases, at least a comparison should be made between the process of publishing articles in each database or the emergence of topics in different time spans, otherwise, the choice of two databases is not necessary.	Since the PubMed database is specific to medical science studies, and due to the nature of the health information technology field, many articles in this field are also published in non-medical journals, so the Scopus database was also searched due to its thematic scope.
In the methodology section, please indicate separately the number of records retrieved from each database (PubMed / Scopus), the number of records excluded at each stage of the search process, and the number of final records analyzed.	The methodology was rewritten. Each step in the methodology was separated.
Please add a description to the methodology section to reassure the audience that only the three topics listed (Medical Informatics, Telemedicine, and Health Information Management) are relevant to the HIT domain in the mesh tree hierarchy, and no other items are left out.	The desired corrections were made.
In some cases, the figure number and its description in the text do not match.	The desired corrections were made.
The chart (The trend of topics in health information technology publications) is not clear to the audience. For more clarity, you can use different symbols for each subject in	The desired corrections were made.

Legend (e.g., triangle, square, circle, etc.) or at least one numeric label (1,2,3, etc.) on each chart and the guide next to the chart, specifying that which label is corresponding to each topic.	

2 Revision 1

2 comments · 10 Jul 2022 – 24 Jul 2022

Adeleh Asadi Reviewer

ACCEPT 10 Jul 2022

I would like to thank the author(s) for making corrections in the article and answering the questions. The article has the basic quality required for publication.

Editor-in-Chief

MINOR REVISION 24 Jul 2022

Minor Revision

3

Revision 2

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Dear Editor

Thank you very much for reviewing our manuscript. We also greatly appreciate the reviewers and editor for their complimentary comments and suggestions.
We have modified the manuscript accordingly, and detailed corrections are listed below point by point and highlighted in manuscript.
We look forward to hearing from you regarding our submission. Due to the closeness of the doctoral interview, we would be grateful if you could send us the article acceptance.

Best Regards
Fatemeh Bahador
Corresponding Author

Editor's comments	Answers
1. Words with the same root like image, imaging, and images should not appear separately in a word cloud. By revising all the word clouds in the article, please merge all the words of the same root so that they appear only once in each word cloud.	According to the editor's comment, we redesigned the word cloud (Figure 2) and the words with the same root have been merged.
2. The comment about the chart (trend of topics in health information technology publications) is not addressed completely. That is, the numeric labels should also appear on the charts not only in the legend. Please revise it again.	The chart, in figure 4, is revised and designed clearly. Because of the high number of lines, inserting all of the numbers in this chart is not possible. Therefore, just two important lines were inserted with numeric labels.

3 Revision 2

2 comments · 16 Aug 2022 – 21 Aug 2022

Associate Editor

MINOR REVISION 16 Aug 2022

1. Figure 4 still needs to be modified. Please once again pay attention to the following comment and act accordingly:
Because Figure 4 includes different curves, one should be able to identify each curve, which can be done by assigning numeric or alphabetical labels to each curve. This numeric or alphabetical label for each curve should appear both in the legend box (a box that explains all curves) and on each curve. For example, when 0 is assigned to the "HIT Adoption" curve, 0 should also appear on the above-mentioned curve once as a label. Please do not confuse this with data labels where the mathematical value of different points appears on each curve. Or, you could have simply used different symbols for showing data points on different curves and used a legend box. Right now, only the " circle" symbol is assigned to all data points in all curves, which is confusing. Try and use different symbols such as "left triangle", "right triangle", "diamond", "square", etc., which make each curve easily identifiable without any need for numeric or alphabetical labels.
2. The comment about the word clouds is not addressed completely. As mentioned in the previous review, all the word clouds in the article need to be revised, not just those in figure 2.

Editor-in-Chief

MINOR REVISION 21 Aug 2022

Minor Revision

4

Revision 3

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Dear Editor

Thank you very much for reviewing our manuscript. We also greatly appreciate the reviewers and editor for their complimentary comments and suggestions.

We have modified the manuscript accordingly, and detailed corrections are listed below point by point and highlighted in manuscript.

We look forward to hearing from you regarding our submission. Due to the closeness of the doctoral interview, we would be grateful if you could send us the article acceptance.

Best Regards

Fatemeh Bahador

Corresponding Author

Editor's comments	Answers
1. Figure 4 still needs to be modified. Please once again pay attention to the following comment and act accordingly: Because Figure 4 includes different curves, one should be able to identify each curve, which can be done by assigning numeric or alphabetical labels to each curve. This numeric or alphabetical label for each curve should appear both in the legend box (a box that explains all curves) and on each curve. For example, when 0 is assigned to the “HIT Adoption” curve, 0 should also appear on the above-mentioned curve once as a label. Please do not confuse this with data labels where the mathematical value of different points appears on each curve. Or, you could have simply used different symbols for showing data points on different curves and used a legend box. Right now, only the “circle” symbol is assigned to all data points in all curves which is confusing. Try and use different symbols such as	Due to the fact that 16 topics are shown in Figure 4, creating markers (such as circles or triangles) on each of the lines would cause the diagram to be cluttered and crowded, so the diagram of Figure 4 was drawn as a heat map diagram.

“left triangle”, “right triangle”, “diamond”, “square”, etc. which makes each curve identifiable easily without any need for numeric or alphabetical labels.	
2. The comment about the word clouds is not addressed completely. As mentioned in the previous review, all the word clouds in the article need to be revised not just figure 2.	Word cloud images were modified.

4 Revision 3

1 comment · 26 Sep 2022

Associate Editor

ACCEPT 26 Sep 2022

Accept