

Research Article

The Use of Artificial Intelligence in Literature Search and Selection of the PubMed Database

Jianbing Ma ¹, Xi Wu ¹, and Lihong Huang ^{2,3}

¹School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China

²Department of Biostatistics, Zhongshan Hospital, Fudan University, Shanghai 200032, China

³Evidence-based Medicine Center, Fudan University, Shanghai 200032, China

Correspondence should be addressed to Lihong Huang; huang.lihong@zs-hospital.sh.cn

Received 6 October 2021; Revised 29 January 2022; Accepted 8 February 2022; Published 22 March 2022

Academic Editor: Rahman Ali

Copyright © 2022 Jianbing Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. A vast number of research papers are published every day on PubMed, making it difficult for scientists to retrieve relevant articles in a timely manner. Keyword-based searches are currently the most popular method, but determining a suitable set of keywords can be challenging. Moreover, searches based on keywords typically retrieve many irrelevant papers. We developed a natural language processing (NLP)-based keyword augmentation and screening (NKAS) method to help scientists easily refine their keywords in topic searches. This method can extract meaningful candidate keywords from the titles and abstracts of an initial search using prior knowledge, knowledge graphs, and machine learning. The method was tested on three atrial fibrillation topics. When the NKAS was applied, the number of remaining papers was less than those in the original search but showed much higher precision (73.83% vs. 34.6%) and recall (98.4% vs. 59.93%) compared with those of the original search results. In conclusion, the NKAS method showed that NLP and other artificial intelligence techniques can help enhance both the search comprehensiveness and accuracy. These results suggest a great potential for the application of artificial intelligence methods in medical publication searches and other text-based applications.

1. Introduction

More than 26 million articles have been published in PubMed [1], and every day, this number increases by more than 2000 [2], which makes searching for and identifying relevant literature a difficult task for scientists. A traditional keyword-based literature search usually starts with setting two or three keywords, sometimes with a few general restrictions such as time period, publication type, or publisher. The search then easily returns hundreds or thousands of papers where the keywords can be found in their titles or abstracts [3]. However, from these, there might only be a few results that precisely match the topic of interest. The scientist conducting the search must then check all the results manually (usually by reviewing the titles and abstracts) to identify the appropriate articles. If most of the return results do not pertain to the topic of interest, or if they do

not sufficiently cover the topic, the scientist may have to adjust the keywords and perform the search again.

Methods to improve keyword searching have been explored by many researchers. Jung et al. proposed the use of an initial broad keyword search followed by a rule-based keyword matching method to obtain the final results [4]. However, the matching did not focus on paper identification but instead aimed at finding associations of genes and disorders from the search results. Also, this study used existing synonyms of search keywords, even though many papers of interest can only be retrieved by using other relevant keywords (but are not synonyms). Such synonym-based query extension methods are widely used in image search applications where exact keyword matching is not important [5, 6]. In contrast, for topic searches, the exact keyword match is a basic requirement. King and Bree used a similar approach but instead used an initial search followed by a manual review of the abstracts to obtain *similar words*;

however, this is not practical when there are hundreds of return results [7]. Volanakis and Krawczyk [1] implemented a system to search cited texts from 1.7 million papers, where cited texts are pieces of text in the papers that refer to other publications. This method effectively utilized information of interest in cited texts to replace predefined keywords. However, the cited texts were usually not representative enough. They were quite often relevant to the general field of the paper rather than to its core contributions, so searches based on the cited texts were not accurate. While these efforts may be helpful under certain conditions, to the best of our knowledge, there is currently no method for the identification of corresponding candidate keywords to help improve the search results.

Moreover, many of the retrieved results are unwanted, but the more specific a search is made, the larger the proportion of irrelevant search results. Therefore, a system that could automatically assess the relevance of each piece of retrieved literature would be valuable in obviating the onerous task of manual screening.

Natural language processing (NLP) [8] is a field of artificial intelligence (AI) that enables computers to understand, interpret, and manipulate human language. NLP uses many different techniques, such as knowledge graph, statistical and machine learning, computational linguistics, and rule-based methods, in order to analyze and comprehend natural languages. Basic NLP tasks include tokenization and parsing, lemmatization/stemming, part-of-speech (POS) tagging, entity recognition, and semantic relationship extraction, which mimic the manner in which humans manually process sentences and paragraphs in their early years.

Keywordtuning is labor-intensive, and quite often, even the final set of keywords is still incomplete or inaccurate due to the diversity of medical phrases and expressions. To remedy this, NLP-based methods can be applied to identify closely related candidate keywords, called replacement and correlated keywords. Authors can select from this pool of keywords with the aim of improving the comprehensiveness of the search results.

Therefore, to help scientists identify topic-related keywords and to remove irrelevant search results, this study proposed an NLP-based keyword augmentation and screening (NKAS) method that improves upon the traditional paper retrieval approach. In NLP-based keyword augmentation, prior knowledge and reasoning techniques are employed to remove irrelevant search results. It should be noted that while this study examined the PubMed database, our NKAS method may be broadly applicable to many other databases such as Google Scholar and ScienceDirect.

2. Methods

2.1. A Brief Description of the Approach. The NKAS method begins with an initial search using a few keywords and filters (Figure 1). Step 1. It then utilizes NLP to process the initial search results, obtaining useful candidate keywords for scientists to choose from Figure 1. Step 2. Next, when

scientists have selected the final keyword set, the search is performed again, and a smart two-stage screening method is employed to identify the truly relevant papers (Figure 1). Step 3.

Importantly, all the processes are automated with the exception of scientists defining their initial keywords and choosing suitable replacement and correlated keywords from the recommended list. NLP techniques are used in both Step 2 and Step 3.

Step 1. The traditional search

The first step is the traditional keyword-based literature search that usually starts with setting two or three keywords as defined by the scientist.

Step 2. Identification of replacement and correlated keywords

Keywordtuning can be difficult, as there are often different terminologies to describe the same or closely related concepts, and thus, to obtain the full set of relevant papers, all the related terms must be considered. For instance, both “auricular fibrillation” and “atrial tachyarrhythmia” are widely used in papers describing “atrial fibrillation”; not including both these phrases in the keywords may lead to the omission of a large portion of papers from the search results. For convenience, we call these largely synonymous phrases such as “auricular fibrillation” and “atrial tachyarrhythmia” replacement keywords. Replacement keywords are common; for example, “thromboembolism,” “cerebrovascular disease,” “ischemic,” and “embolism” have similar meanings to “stroke,” while “mitral regurgitation,” “aortic regurgitation,” “tricuspid regurgitation,” “mitral stenosis,” and “aortic stenosis” have similar meanings to “valvular heart disease.” These replacement keywords might not have the exact same meaning as the original keyword, but the papers addressing these other terms should also be put forward for consideration.

Furthermore, there are also correlated keywords that are closely related phrases to the search keyword; for instance, “creatinine clearance” is linked to “renal insufficiency” since abnormal creatinine clearance indicates renal insufficiency. Obviously, listing all possible replacement and correlated keywords is impractical, but a smart approach that includes all replacement and correlated keywords would provide a truly comprehensive search.

Therefore, NLP techniques were developed and applied to identify the related keywords based on the titles and abstracts of the search results according to the initial keywords. For the current study, the following procedure was followed:

- (1) The abstracts were first separated into sentences with the title being considered a single sentence. A sentence tokenizer, imported from Natural Language Toolkit (NLTK) [9], was used.
- (2) Abbreviations were recorded and extended. The surrounding words of an abbreviation were checked and the corresponding original words were determined automatically, so that the abbreviation could be extended. The abbreviations and their original

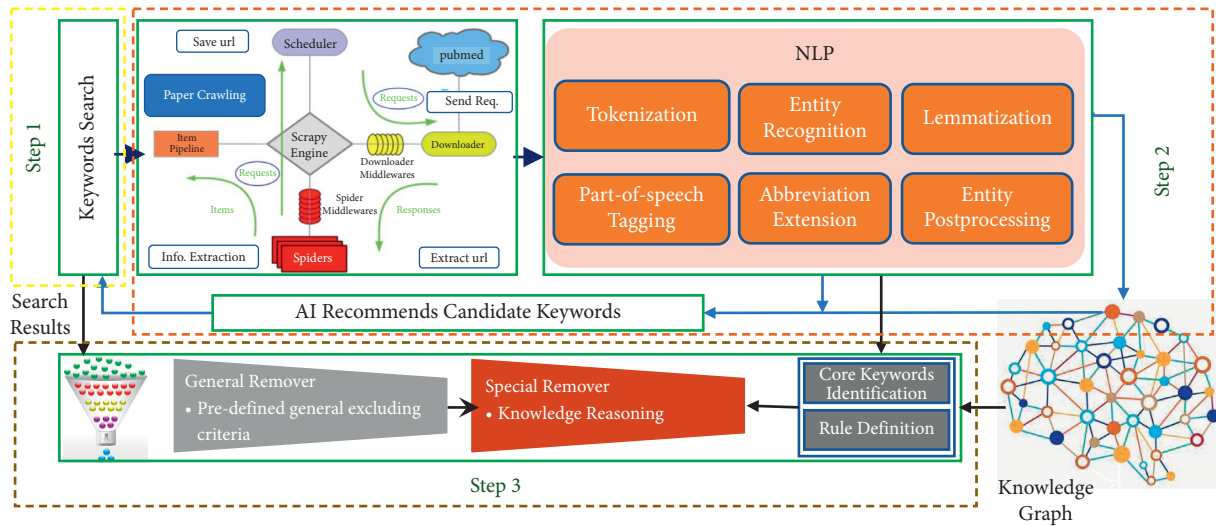


FIGURE 1: The work flow of the NKAS method. Keyword search is automated, and then, the results are analyzed by NLP, where a set of candidate keywords are recommended for revised search. The final search results are then processed by the general and special removers, which are supported by the NLP functionalities and knowledge graph. The knowledge graph, currently supporting medical abbreviations and disease/drug relations, can be generated automatically from the NLP process.

words were recorded in a database as tuples and thus made a big table of such correspondences since they were collected from a lot of papers. With this correspondence table, if for an abbreviation, it could not identify the corresponding original words from its surrounding words, the table was checked to obtain the possible extensions.

- (3) All the sentences were then separated into words, using the word tokenizer of NLTK. Part-of-speech (POS) tagging was also performed using the NLTK ConsecutiveNPChunkTagger, which provides a better performance over other taggers.
- (4) The NLTK wordnet lemmatization tool was used to lemmatize words.
- (5) With the preparation above, the maximum entropy classifier (with a parsing accuracy of 96.0% tested on the conll2000 dataset provided by NLTK) was used to perform named entity recognition. For the purposes of this study, a named entity was considered the name of the disease, symptom, organization, etc. The parsed named entities were then postprocessed into a list of possible replacement and correlated keywords. For the “and” or “or” conjunctions, the phrase was divided into two parts based on the “and” and “or,” except when “or” meant “odds ratio,” which is a statistical parameter. The stop words, digits, and other nonletters were removed from the entities. Words connected by hyphens were examined to determine if they were meaningful keywords using similar procedures as above. The remaining words and phrases were ranked and stored.

From the postprocessing on the named entities, a list of possible replacement and correlated keywords were provided. The NLP flow of Step 2 is presented in Figure 2.

Step 3. The two-stage paper screening

When the search was performed using the refined keywords, the NKAS method was used to deploy a two-stage screening to keep only the truly related papers (Figure 3).

The first stage is called a “general removal stage.” In this stage, the unexpected papers were removed according to several general exclusion criteria, which can be adjusted depending on the scenario. These exclusion criteria included the impact factor of the journal, the article type, the research methods (simulated study or experimental study), and the structure of the article (such as the presence or absence of an instructed abstract).

In contrast, the second stage is referred to as the “specific removal stage” and is specially designed for a given topic. For each topic, a topic-specific model was recommended that utilized the core keywords (the keywords that could differentiate relevant papers from irrelevant ones, with positive core keywords suggesting relevant papers and negative ones implying irrelevance) and the comparison results relating to the core keywords. More specifically, the following two common considerations were applied in the models for the specific remover criterion settings: (1) the location where the core keywords appeared (the title, background, or conclusion of the abstract) and the number of times the keywords appeared, and (2) the number of comparison results that appeared in the abstract relating to the core keywords.

2.2. Examples. We tested our NKAS method on three topics: “Which novel oral anticoagulant (NOAC) is more suitable for preventing strokes in atrial fibrillation patients?” (the NOAC case); “Which anticoagulation treatment is more suitable for atrial fibrillation patients with renal insufficiency?” (the kidney case); and “A comparison of left atrial appendage treatment and new oral anticoagulation drugs for

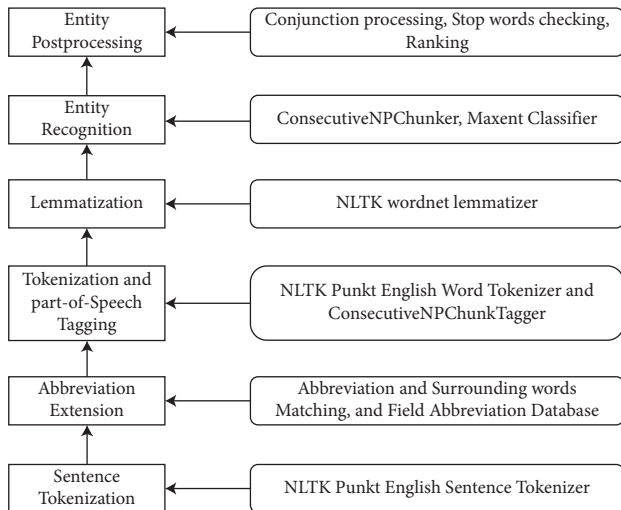


FIGURE 2: NLP flowchart of Step 2. We have used the NLP analysis algorithms in the NLTK toolset to find the candidate keywords.

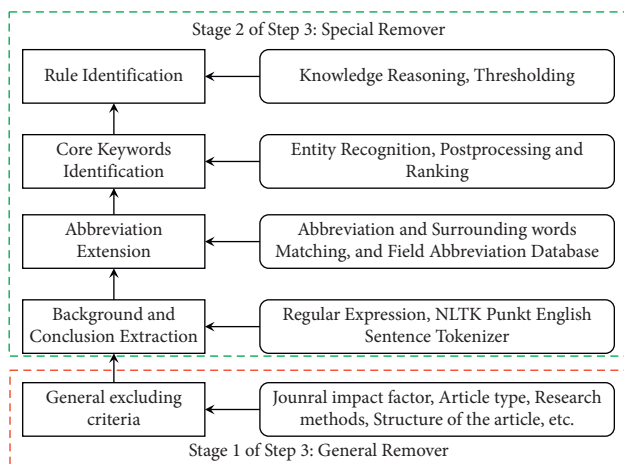


FIGURE 3: The Step 3 of the NKAS method. The two-stage screening includes Stage 1: general remover and Stage 2: special remover. Note. NLP and knowledge graph are needed for the special remover.

the treatment of atrial fibrillation patients” (the LAA vs. NOAC case). None of these topics have been clearly defined in the current atrial fibrillation guidelines [10].

3. Results

Figure 4 describes an example of the NKAS process proposed in this article for the first topic; that is, the NOAC case. Similar processes were applied for the other two topics (the kidney case and the LAA vs. NOAC case). Figure 5, Figure 6 and Figure 7 show the numbers of papers resulting from the traditional search and the augmented search.

As can be seen from Figure 5, for topic 1, the extension of each keyword “NOAC,” “stroke,” and “atrial fibrillation” to a set of related keywords greatly increased the number of related papers from 58 to 635, which is about 10.95 times more. The augmented keywords (we defined augmented keywords as the keywords derived from the initial keyword

set) were acquired from the NLP analysis of the candidate keywords in the 58 papers, and the top 10 keywords (and their frequency of detection in the papers) are listed in Figure 5. Similarly, for topics 2 and 3, augmented keywords increased the search results from 30 to 104 papers and from 8 to 30 papers, respectively. This corresponded to a 3.47-fold and 3.75-fold increase, respectively. Clearly, the introduction of candidate keywords from the NLP analysis resulted in a far more comprehensive search.

While keyword augmentation increased the number of possibly related papers, Step 3 of the NKAS method, the screening process, was deployed to significantly reduce irrelevant papers. Figure 8 shows that although keyword augmentation raised the number of search results by several fold, after the general and specific screening models were applied, the number of remaining papers was actually not greater than that found in the traditional search (namely, 26 vs. 58, 23 vs. 30, and 8 vs. 8 for topics 1, 2, and 3, respectively).

While the NKAS method did not identify more papers than the traditional search, it remained to be determined whether the final selected articles were accurate and comprehensive with regard to the topic of interest.

For the convenience of comparison, we have prepared target papers through manual screening and examination for each of the three topics. Based on the comparisons with the target papers, the accuracy and comprehensiveness of the NKAS results and the traditional search results are shown in Table 1. The precision scores indicate that the search accuracy and the recall scores confirm the search’s comprehensiveness. The data demonstrated that compared to the traditional search, our NKAS method was superior in both search accuracy and comprehensiveness. The detailed results of the original search and the NKAS method on the above three topics are in the supplemental files Topic1_NOAC docx, Topic2_Kidney docx, and Topic3_LAAvsNOAC docx, respectively.

Note that in the kidney case and the LAA vs. NOAC case, all the truly relevant papers were included in the set of papers identified with the special model, while in the NOAC cases, 1 of the truly related papers, “Effectiveness and Safety of Non-Vitamin K Antagonist Oral Anticoagulants in Asian Patients With Atrial Fibrillation” [11], was mistakenly abandoned by the special remover (Stage 2 of Step 3: two-stage screening), since the information on the comparison of NOACs on stroke prevention was only detected in the contents sections of the paper, rather than in the title and the abstract. In addition, 1 paper [12] in the original search results for the kidney case was actually relevant to the topic; however, in the process, only papers published in journals with an impact factor not less than 3 were included, and therefore, it was excluded by the general remover and here annotated as not relevant for convenience.

Compared with the traditional search results and the screened results of the NKAS method, we found that the traditional search could not express deliberate meanings. For instance, in the NOAC case, we wanted to explore “which NOAC was better and in what situation(s)?”. However, many papers in the traditional search results simply took all

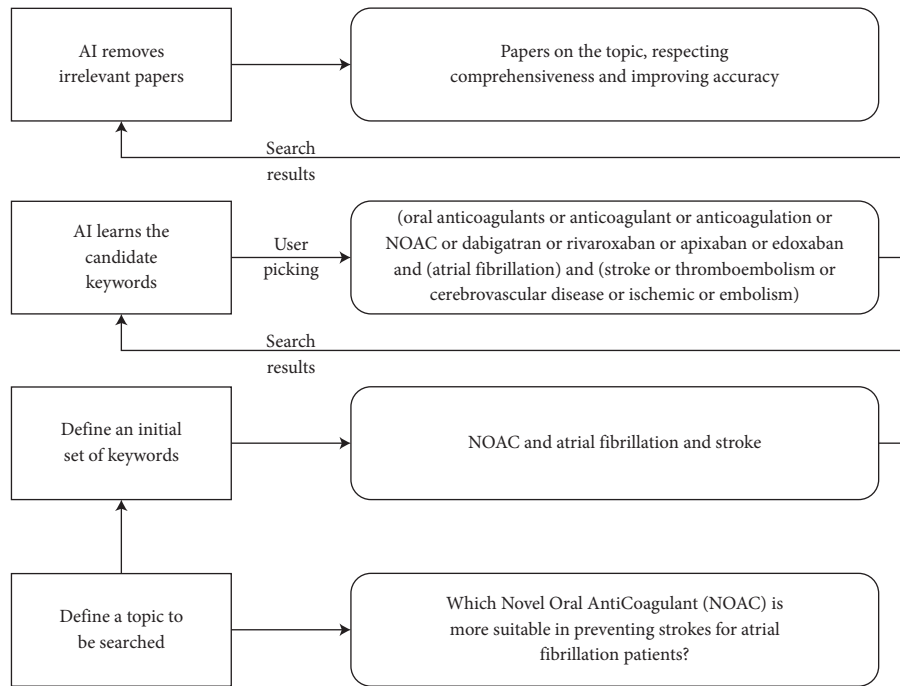


FIGURE 4: An example NKAS process for the first topic (NOAC case).First, the topic and the initial keywords are proposed, and then, AI helps to identify the candidate keywords from the initial search results. The subsequent search results are then screened by the general and special removers, with the remaining papers on the topic increased on comprehensiveness and accuracy compared with the initial search results.

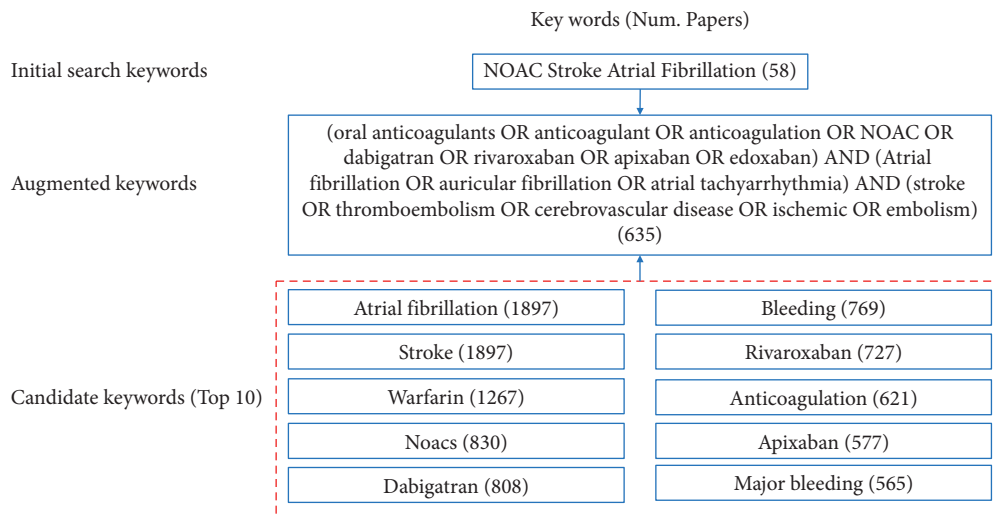


FIGURE 5: Replacement and correlated keywords finding results for topic “Which Novel Oral AntiCoagulant (NOAC) is more suitable in preventing strokes for atrial fibrillation patients.”

NOACs as a whole group to compare to other drugs [12, 13]. This may be an illustrative explanation for the lack of high precision of traditional searches.

4. Discussion

In recent years, AI has been widely used in medical research, in both medical images and texts. However, little attention has been devoted to the actual search process, which can be a demanding issue for physicians and

researchers. Search automation is one way to reduce a physician’s search burden. Stork [14] is the most common medical search automation tool. However, it does not have an AI-aided functionality to augment keywords and screen with designated models. The PaperBot [15], published in 2019, is also a crawler that helps medical search automation. But it does not enhance search comprehensiveness and accuracy. Textpresso Central [16] is another platform that not only helps users in keyword and category searches but can also annotate and curate

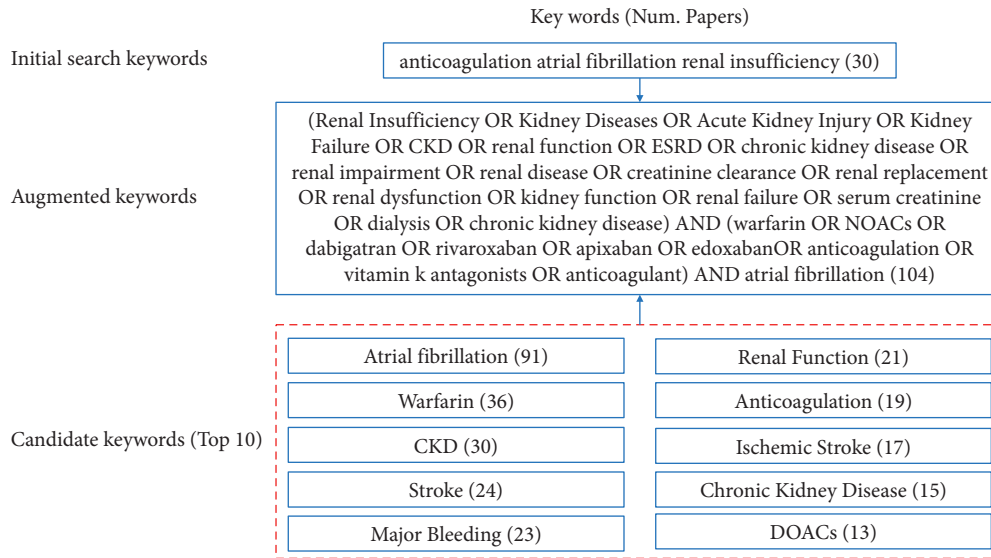


FIGURE 6: Replacement and correlated keywords finding results for topic “Which anticoagulation treatment is more suitable for atrial fibrillation patients with renal insufficiency.”

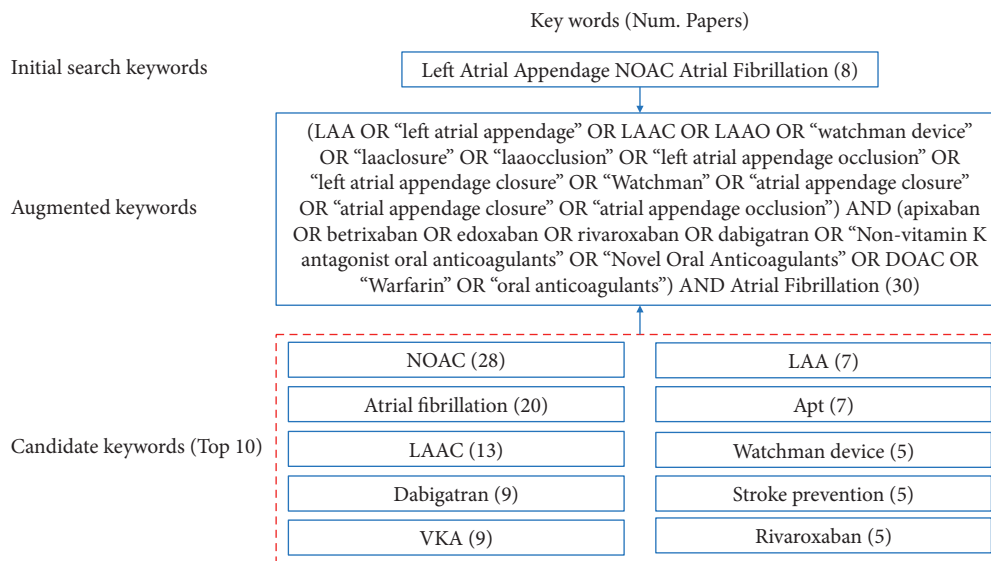


FIGURE 7: Replacement and correlated keywords finding results for topic “How is the left atrial appendage treatment compared to using new oral anticoagulation drugs for atrial fibrillation patients?”

the identified papers. Although these tools use automation to reduce the human search effort, none have been designed to help the user refine keywords and screen search results.

As observed from the results of the three demonstrative topics in this study, AI-aided search and automation tools are useful not only in improving the search completeness and accuracy, but also in reducing the human effort required. Indeed, similar tools can be extended to many fields in medical research. For instance, in meta-analysis preparation, NLP techniques can aid in the precise identification of papers of

interest and extract the useful parameter values from the texts for the pooled analysis. It can also be used in tasks such as identifying text corresponding to medical concepts in clinical case reports [17, 18].

It is also worth noting that we attempted to apply machine learning techniques, such as support vector machine, Naïve Bayes classifiers, and decision trees, to replace the special remover. However, these machine learning techniques did not perform well in classifying the papers as truly relevant or not after the general model, and none achieved an F1-score (a commonly used balanced measurement of model performance)

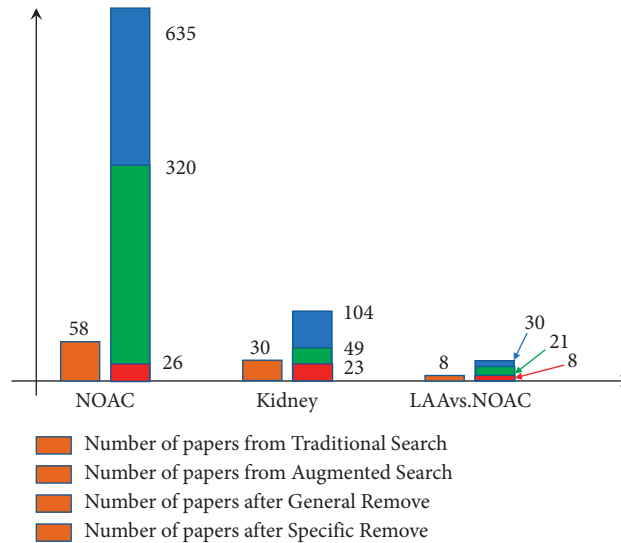


FIGURE 8: How the two-stage screening (general remove and specific remove) has helped to reduce the number of unnecessary papers.

TABLE 1: The comparison of precision and recall of the traditional search method and the NKAS method.

Topic	Num. truly related	Traditional search precision	Traditional search recall	Result Precision	Result Recall
NOAC	21	13.8% (8/58)	38.1% (8/21)	76.9% (20/26)	95.2% (20/21)
Kidney	16	40% (12/30)	75% (12/16)	69.6% (16/23)	100% (16/16)
LAA vs. NOAC	6	50% (4/8)	66.7% (4/6)	75% (6/8)	100% (6/6)
Average		34.6%	59.93%	73.83%	98.4%

Note. Result Precision = (Num. (papers truly related \cap papers in the search result))/Num. Papers in the search result, and Result Recall = (Num. (Papers Truly related \cap Papers in the search result))/Num. Truly related. NOAC: novel oral anticoagulant; LAA: left atrial appendage.

over 72.7% for the three topics analyzed. In comparison, the lowest F1-score for our special NKAS model across the three topics was 82.1%.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall}. \quad (1)$$

There are two possible explanations for this result. The first is that the topics considered in this study were narrow and hence even before the general model screening, there were few papers in the search results. Around 30–300 papers may be identified from an elaborated keyword search, from which approximately 10–30 papers may finally meet the topic specificity, and this small sample size often reduces the performance of machine learning algorithms. Second, the papers in the search result set were literally very similar in their titles and abstracts, which hinders the performance of all machine learning algorithms using word-based models, whether it be word2vec [19, 20] or bag of words [21].

However, we should also note that the application of the AI-based methods to the medical fields is still in its infancy, especially for the text-based NLP methods, and a few limitations have yet to be overcome. First, the lack of a large-scale and well-established annotated dataset for medical texts prevents NLP and other machine learning methods from achieving a great performance. The current BERT model (Bidirectional Encoder Representations from Transformers) [22] is based on massive text resources and has demonstrated state-of-the-art performances in many

text-based machine learning tasks, but it is computationally costly, not specifically fine-tuned for medical scenarios, and thus not suitable for our case. Second, as stated above, in many scenarios such as ours, the small scale of the relevant papers and their similarity to each other makes it difficult for machine learning algorithms to distinguish between them, and hence, NLP must be applied in a deliberate manner to aid human activities.

The use of AI-based methods may also encounter trust issues, as most medical scientists are not familiar with AI and may not readily accept these methods. Indeed, some of the AI methods may output unrepresentative results, which can lead to poor outcomes. Therefore, a scientist-guided application of AI methods may provide more autonomy and consequently be more agreeable. Indeed, the NKAS method presented in this report is one such effort in this direction.

5. Conclusion

AI is being increasingly and intensively applied in many medical scenarios, but mostly for image-related functions. In this study, we developed the NKAS method, a three-step AI-aided search and automation tool. NKAS provided a final search result that was more comprehensive and accurate compared to an original keyword search. Our method is a demonstration of the great potential of AI in text-based medical applications.

However, in our method, the scientists must choose from the recommended keywords to refine their initial keywords, and thus, it does not offer complete automation. In the future, further NLP techniques can be applied to develop a more intelligent system.

Moreover, although the NKAS method is universally applicable by its design, it has only been evaluated in the PubMed database thus far. The transfer and adaption of this method to other literature search databases such as ScienceDirect and Google Scholar is warranted [22].

Data Availability

The excel data used to support the findings of this study are included within the supplementary information file [<https://drive.google.com/file/d/1WdOdilpr5uae-Uj2n2S-PicLKahGeVdfw/view?usp=sharing>].

Conflicts of Interest

The authors have no conflicts of interest to declare.

Authors' Contributions

J. M. and H. H. contributed to the study design. J. M. contributed to data collection, NLP design, and system implementation. All authors drafted the manuscript. All authors contributed to the critical revision of the manuscript and approved its final version.

Acknowledgments

The present study was supported by the National Natural Science Foundation of China (Grant no. 81903407 to Lihong Huang), the Sichuan Provincial Key Research and Development Program (Grant no. 2021YFG0345 to Jianbing Ma), and the starting research fund from Chengdu University of Information Technology (Grant no. 376245 to Jianbing Ma).

Supplementary Materials

There are three supplementary files, Topic1_NOAC docx, Topic2_Kidney docx, and Topic3_LAAVsNOAC docx, containing the data retrieved from the three example scenarios, respectively. Each file contains four tables: (i) the first table shows the original search (NKAS Step 1) result; (ii) the second table lists the refined search (NKAS Step 2) result; (iii) the third table presents the result when the general remover (NKAS Step 3, stage 1) is applied; and (iv) the fourth table gives the result when the special remover (NKAS step3, stage2) is applied. (*Supplementary Materials*)

References

- [1] A. Volanakis and K. Krawczyk, "SciRide Finder: a citation-based paradigm in biomedical literature search," *Scientific Reports*, vol. 8, Article ID 6193, 2018.
- [2] C. Neylon and S. Wu, "Article-level metrics and the evolution of scientific impact," *PLoS Biology*, vol. 7, Article ID e1000242, 2009.
- [3] J. McEntyre and J. Ostell, "The NCBI Handbook," *National Center for Biotechnology Information*, Bethesda, MD, USA, 2002, <https://www.ncbi.nlm.nih.gov/books/NBK21101/>.
- [4] J.-Y. Jung, T. F. DeLuca, T. H. Nelson, and D. P. Wall, "A literature search tool for intelligent extraction of disease-associated genes," *Journal of the American Medical Informatics Association*, vol. 21, no. 3, pp. 399–405, 2014.
- [5] A. Gordo, F. Radenovic, and T. Berg, *Attention-Based Query Expansion Learning*, ECCV, 2020.
- [6] B. Long and Y. Chang, "Visual search ranking," in *Relevance Ranking for Vertical Search Engines*, pp. 59–80, Morgan Kaufmann publishing, Burlington, MA, USA, 2014.
- [7] M. King and A. Bree, "A more intelligent literature search - shoulders of giants," *International Journal of Advanced Research in Artificial Intelligence*, vol. 3, no. 7, pp. 1–4, 2014.
- [8] I. A. Hameed, "Using natural language processing (NLP) for designing socially intelligent robots," in *2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, Cergy-Pontoise, France, 19–22 Sept. 2016.
- [9] NLTK, [url]: www.nltk.org/. [Accessed]: 2020/8/31, 2020.
- [10] J. Steffel, R. Collins, M. Antz et al., "The 2018 European Heart Rhythm Association Practical Guide on the use of non-vitamin K antagonist oral anticoagulants in patients with atrial fibrillation," *European Heart Journal*, vol. 39, no. 16, pp. 1330–1393, 2016.
- [11] M. J. Cha, E.-K. Choi, K.-D. Han et al., "Effectiveness and safety of non-vitamin K antagonist oral anticoagulants in asian patients with atrial fibrillation," *Stroke*, vol. 48, pp. 1–14, 2017.
- [12] R. Zou, J. Tao, W. Shi et al., "Meta-analysis of safety and efficacy for direct oral anticoagulation treatment of non-valvular atrial fibrillation in relation to renal function," *Thrombosis Research*, vol. 160, pp. 41–50, 2017.
- [13] S. Danik, R. Altman, C. Barrett et al., "Non-Vitamin K antagonist oral anticoagulants and antiplatelet therapy for stroke prevention in patients with atrial fibrillation: a meta-analysis of randomized controlled trials," *Cardiology in Review*, vol. 24, no. 5, pp. 218–223, 2016 Sep-Oct.
- [14] I.-S. Kim, H.-J. Kim, T.-H. Kim et al., "Non-vitamin K antagonist oral anticoagulants have better efficacy and equivalent safety compared to warfarin in elderly patients with atrial fibrillation: a systematic review and meta-analysis," *Journal of Cardiology*, vol. 72, no. 2, pp. 105–112, 2018 Aug.
- [15] Stork App, [url]: <https://www.storkapp.me/>[Accessed]: 2020/10/17, 2020.
- [16] P. Maraver, R. Armañanzas, T. Gillette, and G. Ascoli, "PaperBot: open-source web-based search and metadata organization of scientific literature," *BMC Bioinformatics*, vol. 20, no. 1, p. 50, Jan. 2019.
- [17] H. Müller, K. Van Auken, Y. Li, and P. Sternberg, "Textpresso Central: a customizable platform for searching, text mining, viewing, and curating biomedical literature," *BMC Bioinformatics*, vol. 19, no. 1, p. 94, March 2018.
- [18] J. H. Caufield, Y. Zhou, A. O. Garlid et al., "A reference set of curated biomedical data and metadata from clinical case reports," *Scientific Data*, vol. 5, Article ID 180258, 2018 Nov 20.
- [19] T. Mikolov, G. Corrado, K. Chen, and J. Dean, "Efficient estimation of word Representations in vector space," in *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, pp. 1–12, Scottsdale, Arizona, May 2013.

- [20] Q. Le and T. Mikolov, "Distributed Representations of sentences and documents," vol. 32, pp. 1–9, in *Proceedings of the 31st International Conference on Machine Learning*, vol. 32, pp. 1–9, JMLR: W&CP, Beijing China, June 2014.
- [21] D. Galvez-López and J. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.
- [22] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional Transformers for language understanding," [url] <https://arxiv.org/abs/1810.04805>. [Accessed]: 2020/10/17, 2020.