Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Computer Methods and Programs in Biomedicine 191 (2020) 105415

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

A relevance and quality-based ranking algorithm applied to evidence-based medicine

Jesus Serrano-Guerrero*, Francisco P. Romero, Jose A. Olivas

Department of Technologies and Information Systems, Escuela Sup. Informática, Paseo de la Universidad 4, 13071, Ciudad Real, Spain

ARTICLE INFO

Article history:
Received 14 March 2019
Revised 20 November 2019
Accepted 21 February 2020

Keywords:
Evidence-based medicine
Clustering
Relevance ranking
Quality ranking

ABSTRACT

Background: The amount of information available about millions of different subjects is growing every day. This has led to the birth of new search tools specialized in different domains, because classical information retrieval models have trouble dealing with the special characteristics of some of these domains. Evidence-based Medicine is a case of a complex domain where classical information retrieval models can help search engines retrieve documents by considering the presence or absence of terms, but these must be complemented with other specific strategies which allow retrieving and ranking documents including the best current evidence and methodological quality.

Objective: The goal is to present a ranking algorithm able to select the best documents for clinicians considering aspects related to the relevance and the quality of said documents.

Methods: In order to assess the effectiveness of this proposal, an experimental methodology has been followed by using Medline as a data set and the Cochrane Library as a gold standard.

Results: Applying the evaluation methodology proposed, and after submitting 40 queries on the platform developed, the MAP (Mean Average Precision) obtained was 20.26%.

Conclusions: Successful results have been achieved with the experiments, improving on other studies, but under different and even more complex circumstances.

1. Introduction

Due to the rapid development of new technologies, and especially thanks to the Internet, many new tools in Medicine are arising which can facilitate easy access to information. Thus, new services are appearing; for example, search engines such as PubMed [1], a tool specialized in the field of Medicine which provides millions of citations about medical documents from Medline and other databases. Other examples may be ClinicalTrials [2] or Journal-Watch [3], among others. There are, therefore, systems available which provide a huge quantity of information to clinicians. This presents a risk, as these tools can retrieve a lot of information for each user request, causing confusion and time wasting.

In order to address the problem of searching the best studies, Pubmed presents, for example, some filters [4] limiting the search to specific categories and fields and, constructing complex Boolean queries. The effectiveness of this method has been studied with real users [5,6]; nevertheless, there are studies that question whether technology applied to clinical information retrieval is useful and cast doubt on the ability of scientists to search appropriately [7]. Hence, it is necessary to develop algorithms to improve the quality of search results, especially as finding the best studies may suppose a better treatment for patients [8].

An example may be the algorithm proposed in [9] which models quality criteria through fuzzy prototypes in order to filter Personal Health Records (PHRs). This is only one example of documents in Medicine, but there are many other types of documents (review articles, original studies, case reports, etc.) that are important, especially in Evidence-based Medicine (EBM). According to [10], EBM is “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.” Therefore, EBM consists of integrating the best research evidence with clinical expertise and patient values. It is based on two main principles, the application of the best available evidence from the scientific method to medical decision making [11] and the assessment of the quality of evidence of the advantages and disadvantages of treatments [12].

* Corresponding author.
E-mail address: jesus.serrano@uclm.es (J. Serrano-Guerrero).

https://doi.org/10.1016/j.cmpb.2020.105415
0169-2607/

It is clear that EBM is a very complex concept, which makes the search process more difficult. Thus, although classic information retrieval models [13] can help users retrieve information regarding a user request, EBM needs more complex algorithms to improve the effectiveness of the search tools.

The most effective way of solving the problem of finding good articles according to the principles of EBM is a manual search and classification among millions of articles. Several initiatives can be found in this field, including the libraries: Cochrane Reviews [14], ACP Journal Club [15] or Evidence-based Medicine [16]. The information provided by these libraries is reviewed and classified by experts in the medical field, assuming the cost and effort that this solution involves.

As this task requires many specialized resources which are not always available or easy to find, several studies deal with the problem of selecting methodologically rigorous articles by using Machine Learning classifiers. Aphinyanaphongs et al. [17] used Support Vector Machine (SVM), Naive Bayes and AdaBoost as methods to classify high-quality documents in areas like Treatment and Diagnosis. Choi et al. [18] proposed an algorithm which combines the benefits of Information Retrieval models like Okapi BM25 [19] applied to Medline and Machine Learning techniques such as Naive Bayes or Support Vector Machine (SVM) using the Clinical Hedges Database (CHD). By merging both aspects through the typical strategies of metasearch [20], good results were achieved.

On the other hand, Iruetaguena et al. [21] proposed a strategy for retrieving bibliography from Medline using Literature-based Discovery. And Surian proposed the use of latent space matrix factorization for systematic review updates [22].

As it can be seen, the process of searching documents following the principles of EBM is very complex and involves the development of new specialized tools. As a consequence, the goal of this work is to present a new ranking algorithm able to select the best documents for clinicians considering aspects related to the relevance and the quality of said documents.

2. Methods and materials

This proposal presents a ranking algorithm which orders documents following two main ideas taken from [18,23]: the relevance of the documents is ranked according to the needs of the user and the quality of the documents retrieved.

The first aspect refers to the textual content of the documents. The relevance of the documents depends on the ability of the system to represent, save and retrieve relevant documents with respect to a user query. And the second aspect refers to the quality of the content of the retrieved documents. It considers that documents should be ranked by considering parameters such as the importance of the authors, the publication date or the type of publication. Merging both these aspects, the system should be able to retrieve better documents than systems like Pubmed, which are mainly based on Boolean searches.

In this case, the automatic classifiers used in [18], are replaced by clustering algorithms, because it is not always possible to have access to collections such as the Clinical Hedges Database. Moreover, the use of these automatic classifiers involves spending time on training them, which may not be always possible, desirable or useful. And furthermore, the clustering processes simplify the task of searching because they work as filters, reducing the number of documents to be analyzed. The clinician thus receives a list of labeled clusters and needs only select those which are best suited to his/her needs.

Hence, this proposal facilitates the tasks of clinicians through the interactive way in which the results are grouped and presented, saving time and effort. To sum up, the main contributions of this paper are: a new algorithm to measure the quality of documents in the field of Medicine and a ranking strategy for retrieving high-quality and relevant documents, which has been tested working with real databases such as Medline and Cochrane.

2.1. Ranking strategy

To attain the goal of ranking the best results according to criteria that experts in Medicine may understand, it is necessary to follow the workflow depicted in Fig. 1. Firstly, from a user query submitted to a search engine, a list of documents L1 is retrieved and ranked according to relevance criteria. Once these results have been computed, the second step consists in ranking these documents via quality criteria. In this case, firstly, a clustering algorithm is used to group the results into small subgroups that may be easily examined by the expert who submitted the initial query. This clustering algorithm may be used because, on many occasions, the lists of results retrieved are very large and there is no time to analyze all of them.

Once the clustering algorithm is computed, the user may select the cluster LC that satisfies his information needs. This cluster must be ranked according to quality criteria; in this case, due to the information provided, the criteria used are the importance of authors as well as the type of each document and their publication dates. This process returns a ranked list of results L2.

Once L1 and L2 have been computed, the system merges both lists into a single one, L3, with the aim of combining both criteria: quality and relevance.

To sum up, the main parts of the system are (see Fig. 1).

2.2. Relevance-based ranking

This ranking depends on the information retrieval model used by the search engine to represent and retrieve the information stored. In this case, the collection Medline ®/PubMed®, which is explained in detail in Section 2.2.1, was chosen to test the system and the model used is the well-known Vector Space Model (VSM) [24], but many other models (binary, probabilistic, language models, etc.) may be used to test the approach.

The Vector Space Model establishes a ranking between documents with respect to a query, considering mainly the frequency of the query terms for each document. These ranking scores will be useful in the fusion process.

2.2.1. Quality ranking

As commented previously, when the search engines retrieve the results for a query, the number of results is usually very high, and therefore it is necessary to filter these results in order to simplify the retrieval process. A clustering process has been used to group documents according to the different topics available. In this case, the collection is quite important due to its special characteristics.

Medline stores information about references (authors, title, abstract, dates, etc.) but does not always provide full-text articles, consequently, the available information is limited. Hence, it is necessary to use a clustering algorithm especially designed to work with small amounts of information. The interpretation of the clusters as represented by their labels is also very important, because clinicians must clearly see the labels describing each cluster and quickly select the appropriate one for their needs, so as not to waste time on useless tasks.

Consequently, once all the clusters and their corresponding labels have been shown to the user, he or she must select the most suitable cluster. And after selecting the cluster, an algorithm ranks the different documents according to quality criteria.

In this case, from the available information, the process considers the quality of the authors of each document. There are some metrics for qualifying the quality of an author, such as h-index

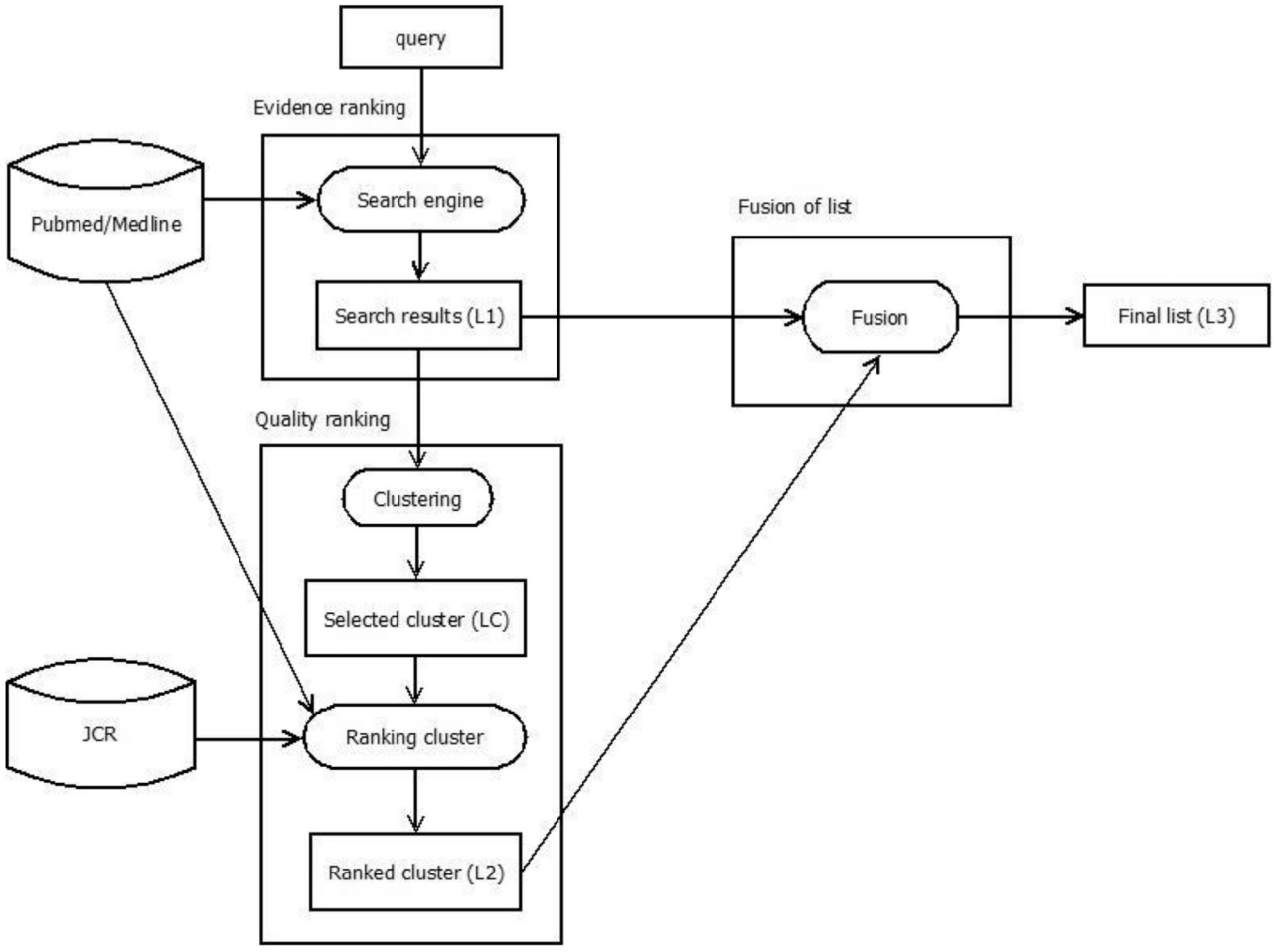

Fig. 1. Workflow.

[25,26]; however, the computation of these metrics requires information that is not always easily accessible and depends on the database used to compute them. Details, like the kind of publication or whether an author may write articles about more fields than Medicine, which is our focus, are not considered by h-index.

For this reason, and because Medline is a representative database about information and experts on aspects related to Medicine, an algorithm has been specially designed to calculate the importance of each author through the documents stored in Medline. This algorithm is based on other typical measures, thus, authors published in the best journals should be the best-considered authors, and papers written by good authors should be the best papers. Clearly, other formulas might be found providing different points of view.

As a result, for each author of each retrieved document, the system retrieves from Medline all documents DA published by him/her. Once all publications have been collected, it is necessary to establish an order among them. Considering the type and place where the study has been published (conference, journal, magazine, etc.), it is possible to rank these publications, and especially, thanks to the information provided by the Journal Citation Reports (JCR) [16] published by Thomson Reuters's Healthcare & Science Division. This report provides information about academic journals in the Sciences and Social Sciences, including data such as the impact factor of each journal.

Hence, from this information, the algorithm supposes that the best publications should be in the group G1 of journals belonging to the Journal Citation Reports, the second group G2 of important publications should be other journals, reviews, clinical trials, comparative studies, etc., because in EBM they may be more important than the remaining types (biography, autobiography, comment, etc.), which are considered in the last group G3 according to importance. Then, following this idea Algorithm 1 computes the importance of each author.

Algorithm 1
Importance of each author.

```
Require: DA: List of documents of an author stored in Medline
index ∈ 0
for all d ∈ DA do
    if d ∈ G1 then
        aux ← 2 + impactFactor(d)
    else
        if d ∈ G2 then
        aux ← 2
    else
        aux ← 0.5
    end if
    end if
    index ← index + aux
end for
return index
```

The *impactFactor* function returns the value of the impact factor of the journal where the paper analyzed has been published.

And so, from this algorithm, the calculation of the quality of each document belonging to a cluster is summarized in Algorithm 2.

The *extractAuthors* function extracts the authors of the work analyzed, whereas the *qualityAuthor* function represents Algorithm 1.

Algorithm 2
Quality for each document.

```
Require: LC: List of documents of a cluster
L2 ← φ
for all d ∈ LC do
    AD ← extractAuthors(d)
    index ← 0
    for all author ∈ AD do
        index ← index + qualityAuthor(author)
    end for
    L2 ← < d, index >
end for
L2 ← sort(L2)
return L2
```

The function *sort* ranks the list of documents regarding the computed quality index of their authors. If two documents have the same quality index, the first would be the most recent by date of publication, because we consider that more recent studies will be more up to date than older ones.

2.3. Fusion

Once both lists of documents have been retrieved and ranked according to the different criteria, it is necessary to merge them in order to take advantage of the capabilities provided by each. As a result, the scores from both rankings, relevance and quality, are merged [20,27] in order to obtain the final list L3 which combines properties from both criteria.

2.4. Data collections

To assess this proposal, two collections were necessary. The information from Medline was used as the main database whereas the information from Cochrane was considered as a gold standard.

2.5. Medline®/PubMed®

Medline of the U.S. National Library of Medicine (NLM) may be the biggest bibliographic library in Biomedicine. It comprises citations and abstracts from approximately 5400 biomedical journals related to Medicine, Nursing and Health Care Systems among others. This information is provided by a committee of experts called the Literature Selection Technical Review Committee.

Medline information is available through Pubmed, a free database of biomedical and life sciences literature at the U.S. National Institute of Health's National Library of Medicine (NIH/NLM). Apart from Medline citations, Pubmed contains information from other sources such as science and general chemistry journals. For this study, 2011 Medline®/PubMed® [16] was used. It consists of over 21 million citations which represent the database used.

2.6. Cochrane Library

The Cochrane Collaboration is an international nonprofit and independent organization, dedicated to providing up-to-date information about the effects of health treatments. Thus, its main purpose is to prepare, maintain, and promote access to systematic reviews through the Cochrane Library [28]. It consists of a collection of six databases which contain different types of high-quality, independent evidence to help in health-care decision making. It also includes a seventh database which provides information about groups in the Cochrane Collaboration.

One of the databases included is the Cochrane Database of Systematic Reviews (CDSR). This database includes all Cochrane Reviews (and protocols) prepared by Cochrane Review Groups. Each Review is a peer-reviewed systematic review prepared and supervised by a Cochrane Review Group following the Cochrane Handbook for Systematic Reviews of Interventions or the Cochrane Handbook for Diagnostic Test Accuracy Reviews.

This database is possibly the most important source for systematic reviews in Healthcare. The structure of each document included in this library is described in a handbook that must be followed by the authors of each review. This structure consists of typical fields such as title, authors, abstract, results. Apart from these basic fields, there are others more specific to Medicine, such as types of participants, interventions or studies, as well as references included or excluded, i.e. references interesting or otherwise for the topic discussed in the review. The gold standard used here consists of these references.

2.7. Evaluation metrics

In order to evaluate the obtained results, several metrics in Information Retrieval related to the system efficiency and effectiveness, and several subjective aspects related to user satisfaction may be found [13].

The effectiveness of Information Retrieval Systems is strongly related to the concept of document relevance. From this idea several measures arise such as precision and recall [29]. Precision is the ratio of the relevant documents retrieved by the system with respect to the submitted query to the total number of documents retrieved. Recall, on the other hand, is the ratio of the number of relevant documents retrieved to the total number of relevant documents in the database with respect to the user query.

These measures are very interesting for a set of documents; however, when the set is ranked there are other interesting metrics which consider the order of the retrieved documents. One of these is average precision (AvP), which is mathematically defined as [30]:

$$AvP = \frac{\sum_{k=1}^{n} p(k) * rel(k)}{number\ of\ relevant\ documents}$$

where k is the position of a document in the retrieved list of documents, p(k) is the precision at cut-off k of the list and the value of rel(k) is 1 if the document k is relevant, 0 otherwise.

This measure is interesting in assessing the effectiveness of a specific query; nevertheless, if we are interested in a set of queries, mean average precision (MAP) calculates the mean of the average precision values for a set of queries Q:

$$MAP = \frac{\sum_{q=1}^{Q} AvP(q)}{|Q|}$$

where $|Q|$ is the number of queries [30].

3. Methodology

To drive the experiments, the methodology used is the same as in [18]. This will also be used to compare the results by using the same evaluation measures. Medline was indexed for 60 days by using a shallow parser to preprocess each document thanks to the LingPipe library [31]. The fields indexed were: abstract, author, body, keywords, title and type of document. Furthermore, author quality values (see Algorithm 1) have been precomputed, since it is very time-consuming process to compute them in real time.

Once the database for experimentation is ready, it is necessary to prepare the queries and expected results. The Cochrane collection was used for this purpose. Forty titles of documents from Cochrane were gathered manually (see Table 1), and the documents have been extracted through the service Cochrane PLUS, a translation of the Cochrane Library for Spain and Latin American countries.

Table 1
Titles of all documents.

1. Acellular vaccines for preventing whooping cough in children
2. Acetylcysteine and carbocysteine for acute upper and lower respiratory tract infections in paediatric patients without chronic broncho-pulmonary disease
3. Acyclovir for treating varicella in otherwise healthy children and adolescents
4. Amantadine and rimantadine for influenza A in adults
5. Amantadine and rimantadine for influenza A in children and the elderly
6. Antibiotic prophylaxis for preventing meningitis in patients with basilar skull fractures
7. Antibiotic prophylaxis to reduce respiratory tract infections and mortality in adults receiving intensive care
8. Antibiotics for acute bronchitis
9. Antibiotics for acute laryngitis in adults
10. Antibiotics for acute maxillary sinusitis
11. Antibiotics for acute otitis media in children
12. Antibiotics for community acquired pneumonia in adult outpatients
13. Antibiotics for community-acquired pneumonia in children
14. Antibiotics for preventing complications in children with measles
15. Antibiotics for preventing meningococcal infections
16. Antibiotics for sore throat
17. Antibiotics for the common cold and acute purulent rhinitis
18. Antibiotics for the prevention of acute and chronic suppurative otitis media in children
19. Antibiotics for whooping cough (pertussis)
20. Azithromycin for acute lower respiratory tract infections
21. Beta2-agonists for acute bronchitis
22. Bronchodilators for bronchiolitis
23. Chest physiotherapy for acute bronchiolitis in paediatric patients between 0 and 24 months old
24. Chest physiotherapy for pneumonia in adults
25. Chinese herbs combined with Western medicine for severe acute respiratory syndrome (SARS)
26. Chinese medicinal herbs for acute bronchitis
27. Chinese medicinal herbs for measles
28. Chinese medicinal herbs for sore throat
29. Chinese medicinal herbs for the common cold
30. Combined DTP-HBV-HIB vaccine versus separately administered DTP-HBV and HIB vaccines for primary prevention of diphtheria, tetanus, pertussis, hepatitis B and Haemophilus influenza B (HIB)
31. Continuous negative extrathoracic pressure or continuous positive airway pressure compared to conventional ventilation for acute hypoxaemic respiratory failure in children
32. Corticosteroids for acute bacterial meningitis
33. Decongestants and antihistamines for acute otitis media in children
34. Different antibiotic treatments for group A streptococcal pharyngitis
35. Echinacea for preventing and treating the common cold
36. Empiric antibiotic coverage of atypical pathogens for community-acquired pneumonia in hospitalized adults
37. Epinephrine for bronchiolitis
38. Garlic for the common cold
39. Glucocorticoids for acute viral bronchiolitis in infants and young children
40. Glucocorticoids for croup

Analyzing each document, the references included were extracted manually. These references are not necessarily all present in Medline; therefore, it was necessary to implement an automatic process to check which references were available in Medline. The forty titles collected from Cochrane are the queries that will be submitted to our system and the references available in Medline associated with each query will be the gold standard used to assess the performance of our approach.

Each query will be submitted to retrieve the documents which represent L1. After retrieving these documents, the system will execute the clustering process to group them. As the assessment process is automatic, i.e., there is no user who selects the most suitable cluster according to his needs, it was decided to work with the biggest cluster, i.e., the cluster with most documents.

As previously mentioned, the clustering algorithm should be dealing with small amounts of information, due to the information available for each Medline document and because the selection of the labels which describe each cluster is very important.

Ideally, clinicians should be the ones who choose the best cluster for their interests. From these two ideas, the algorithm chosen for this experiment can be found in [32], because it was designed to work with small amounts of information as snippets, and the use of Latent Semantic Indexing [33] allows a human-perceivable cluster label to be created and documents assigned to it.

Therefore, once the cluster has been chosen, its documents are ordered according to Algorithm 2, generating L2; after this step, both L1 and L2 are merged using different typical formulas for metasearch [20]. And finally, L3 will be assessed through the measures set out in the previous subsection.

4. Results

Applying the methodology previously described, forty queries (8.6 terms per query approximately) were submitted and all results computed and analyzed.

Considering only the results extracted directly from the index, list L1 (relevance ranking), the MAP obtained was 7.14%. Once the list was retrieved, the clustering algorithm was executed. As there is no user interaction in this experiment, the cluster selected was the biggest one. Analyzing the documents included in the results by means of the Vector Space Model, L1, the recall measure confirms that 34.5% of included references have been found using this strategy, whereas analyzing the documents available in the selected clusters for each query, the recall shows that 22.65% of included references were present. Therefore, the probability of selecting this cluster by the user may be high due to the large number of good results present in it.

Once the biggest cluster has been detected, its documents have been ranked according to Algorithm 2, generating list L2. The MAP for this list (quality ranking) was 9.42%. In this case, the relevance ranking gives better results than the quality ranking, but combining the properties of both lists, the results can be improved, as shown by the MAP, 20.26%.

To obtain this result, the scores from L1 and L2 were normalized and several formulas and parameters were tuned, and finally the formula giving best performance was:

$$Fusionscore = (relevancescore)^{\alpha} * (qualityscore)^{\beta}$$

Where $(\alpha{:}\beta)$ were (1: 0.5).

As can be seen, the fusion of both aspects, quality and relevance, can improve the results. Compared to other studies, such as that of Choi [18], our model presents better behavior. On the one hand, the relevance ranking returned similar results, 7.4% vs. 7.14%. And on the other, the quality ranking is slightly better in this study, 8.2% vs. 9.42%. However, the way it is obtained is quite different. Choi uses a Machine Learning algorithm that requires a prior training step. This step uses a collection like the Clinical Hedges Database, which is not always available or easily accessible. For that reason, an alternative method has been followed here. In this case interaction with the user is necessary, but this is not a problem because the user is necessarily waiting for the results of the search process. Hence, a clustering algorithm allows documents to be grouped according to conceptual criteria and the best to be chosen by the user.

Finally, the fusion of the lists depends on the scores assigned to each document from the ranked lists L1 and L2. These scores can vary slightly according to the different models used for relevance ranking, or the Machine Learning or clustering algorithm used for quality ranking. Moreover, the Machine Learning algorithm is supposed to have better discriminatory power with regard to Algorithm 2, but the combination of both factors, relevance and quality, proves that this system gives slightly better results in terms of MAP, 19.6% vs. 20.26%. In Table 2, the values of the AvP for each of them are shown.

Both studies confirm that relevance and quality are two factors that can help us improve the results of searching documents according to EBM principles, but under different circumstances.

Table 2
Average Precision for each individual query.

# Query	AvP	# Query	AvP	# Query	AvP	# Query	AvP
1	1730	11	1910	21	2030	31	1590
2	1650	12	1750	22	1790	32	2010
3	2430	13	1930	23	1540	33	2240
4	1610	14	1860	24	1980	34	2530
5	1830	15	2210	25	2210	35	2270
6	2090	16	1970	26	2070	36	2230
7	1920	17	2060	27	2130	37	2470
8	1870	18	2220	28	2250	38	2120
9	1950	19	1910	29	2460	39	2000
10	2190	20	2180	30	1640	40	2190

5. Discussion

In order to obtain the best information about Evidence-based Medicine, the present study proposes a two-step ranking strategy which considers aspects related to textual details and the quality of each document retrieved.

The first step consists in a textual search process by using the Vector Space Model; nonetheless, this step may be replaced with any other well-known model like that used by the PubMed Best Match System (BM25), giving an example used in Medicine [34]. This model is the same as that implemented in the study [18] used to compare our approach. More comparisons of the performance of different models, and the various implementations on different well-known platforms like Indri or Lucene can be found in [35,36].

In this case, the implementation used by Choi outperforms ours, 7.4% vs. 7.14%, in terms of the MAP. Nevertheless, the success of this proposal lies in the second step, in which the quality of each document is computed considering the quality of the authors and the type of document. Demner-Fushman et al. [23] use these and other variables such as the publication date, the type of study, parts of documents related to PICO (Problem/Population, Intervention, Comparison, and Outcome) structures, etc. to implement a question answering system from the point of view of EBM. Other aspects may be considered, for example, whether the information provided is up-to-date or not; thus, [37] proposed a methodology which estimates the timeliness of diabetes websites according to EBM.

In this second stage, this proposal outperforms Choi's approach: 8.2% vs. 9.42%. Although this is not a big difference, it should be remarked that this proposal is completely unsupervised, whereas

Choi et al. proposed a supervised learning algorithm to select the documents used as an input for the quality-based stage [18].

In our case, as the user is not present, the choice of the biggest cluster as the best one is not always appropriate, and for this reason, the results might be even worse than expected in a real scenario where a real user can interact with the system.

Finally, the fusion of both factors, relevance and quality, is slightly better in our case, 19.6% vs. 20.26%, thanks especially to the improvement achieved by the second stage.

To the best of our knowledge, little research similar to ours can be found, apart from the study used as a comparison [18]. Some other specific systems can be found for searching for literature on Medicine, but they focus on other types of documents, such as Personal Health Records [38–40], or fields like Precision Medicine which requires specific information about, for example, the patient: diseases, genes and mutations, etc. [41]

The main drawback of the algorithm described is the extra time necessary to compute the quality-based step with respect to other more optimized systems only based on textual characteristics.

6. Conclusions and future works

This study presents a method for ranking documents by considering textual characteristics and aspects related to the quality of a document. This approach has been designed to improve the effectiveness of the search process in the field of Medicine, and especially in EBM, where the documents retrieved should provide not only related information, but also the best current evidence and methodological quality. The system has been assessed using real medical databases, and gives good results with respect to other similar solutions in this line.

Nevertheless, there exist several points that could be improved. Computing the quality of each author by using only the references stored in Medline is a possible solution under the current constraints; nonetheless, the platform in which our proposal is integrated is more complex, and allows the registration of user profiles to allow users to upload and download documents and to describe their own interests. Consequently, considering the different characteristics of the registered users, the system can provide personalized information, working in the same way as many other recommender systems [42,43] and decision support systems [44].

Apart from these points related to quality aspects, with respect to the relevance ranking it may be interesting to analyze the impact of other retrieval models and features, for example, language models [45], semantic approaches [46] or opinions [47].

Due to the fact that the user is available to express his/her query, it is also possible to include a feedback step in our system, in which the user can analyze the final list of results selecting the most adequate documents according to his/her query. By using this information, the system may perform several actions such as re-ranking the list of results or expanding/reformulating the original query to repeat the search process, consequently improving the quality of final list of results obtained.

Acknowledgements

This work has been partially supported by FEDER and the State Research Agency (AEI) of the Spanish Ministry of Economy and Competition under grant TIN2016-76843-C4-2-R (AEI/FEDER, UE).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2020.105415.

References

[1] Medline/Pubmed. 2019-03-06. URL:https://www.ncbi.nlm.nih.gov/pubmed. Accessed: 2019-03-06. (Archived by WebCite® at http://www.webcitation.org/76g6FhHWd).
[2] ClinicalTrials. 2019-03-06. URL:https://www.clinicaltrials.gov/. Accessed: 2019-03-06. (Archived by WebCite® at http://www.webcitation.org/76g7XCGpP).
[3] Journal Watch. 2019-03-06. URL:http://www.jwatch.org. Accessed: 2019-03-06. (Archived by WebCite® at http://www.webcitation.org/76g7julN1).
[4] Pubmed Filters. 2019-03-06. URL:http://www.ncbi.nlm.nih.gov/pubmed/clinical. Accessed: 2019-03-06. (Archived by WebCite® at http://www.webcitation.org/76gAvlNJo).
[5] C. Lokker, R.B. Haynes, N.L. Wilczynski, K.A. McKibbon, S.D. Walter, Retrieval of diagnostic and treatment studies for clinical use through PubMed and PubMed's clinical queries filters, J. Am. Med. Inform. Assoc. 18 (2011) 652–659.
[6] S.Z. Shariff, J.M. Sontrop, R.B. Haynes, A.V. Iansavichus, K.A. McKibbon, N.L. Wilczynski, M.A. Weir, M.R. Speechley, A. Thind, A.X. Garg, Impact of PubMed search filters on the retrieval of evidence by physicians, CMAJ 184 (2012) 184–190.
[7] P. Pluye, R.M. Grad, L.G. Dunikowski, R. Stephenson, Impact of clinical information-retrieval technology on physicians: a literature review of quantitative, qualitative and mixed methods studies, Int. J. Med. Inform. 74 (2005) 745–768.
[8] D.A. Hanauer, Q. Mei, J. Law, R. Khanna, K. Zheng, Supporting information retrieval from electronic health records: a report of university of Michigan's nine-year experience in developing and using the electronic medical record search engine (EMERSE), J. Biomed. Inform. 55 (2015) 290–300.
[9] F.P. Romero, I. Caballero, J. Serrano-Guerrero, J.A. Olivas, E. Verbo, An approach to web-based personal health records filtering using fuzzy prototypes and data quality criteria, Inf. Process. Manag. 48 (2012) 159–162.
[10] S. Selvaraj, N. Yeshwant Kumar, M. Elakiya, C. Prarthana Saraswathi, D. Balaji, P. Nagamani, M. Surapaneni Krishna, Evidence-based medicine - a new approach to teach medicine: a basic review for beginners, Biol. Med. 2 (2010) 1–5.
[11] S. Timmermans, A. Mauck, The promises and pitfalls of evidence-based medicine, Heal. Aff. 24 (2005) 18–28.
[12] A.S. Elstein, On the origins and development of evidence-based medicine and medical decision making, Inflamm. Res. 53 (2004) 184–189.
[13] R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, Addison Wesley, Boston, MA, USA, 1999.
[14] Cochrane Reviews. 2019-03-06. URL:https://www.cochranelibrary.com/cdsr/reviews. Accessed: 2019-03-06. (Archived by WebCite® at http://www.webcitation.org/76gDGUv4V).
[15] ACP Journal Club. 2019-03-06. URL:http://acpjc.acponline.org. Accessed: 2019-03-06. (Archived by WebCite® at http://www.webcitation.org/76gDewe0p).
[16] BMJ Evidence-Based Medicine. 2019-03-06. URL:https://ebm.bmj.com/. Accessed: 2019-03-06. (Archived by WebCite® at http://www.webcitation.org/76gDoGTxT).
[17] Y. Aphinyanaphongs, I. Tsamardinos, A. Statnikov, D. Hardin, C.F. Aliferis, Text categorization models for high-quality article retrieval in internal medicine., J. Am. Med. Inform. Assoc. 12 (2005) 207–216.
[18] S. Choi, B. Ryu, S. Yoo, J. Choi, Combining relevancy and methodological quality into a single ranking for evidence-based medicine, Inf. Sci. (Ny) 214 (2012) 76–90.
[19] S.E. Robertson, S. Walker, Okapi/Keenbow at TREC-8, in: TREC '99 Eighth Text Retr, in: Conf., 1999, pp. 151–162.
[20] J.A. Aslam, M. Montague, Models for metasearch, in: SIGIR '01 Proc. 24th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., New York, NY, USA, 2001, pp. 276–284.
[21] A. Iruetaguena, J.J. Garcia Adeva, J.M. Pikatza, U. Segundo, D. Buenestado, R. Barrena, Automatic retrieval of current evidence to support update of bibliography in clinical guidelines, Expert Syst. Appl. 40 (2013) 2081–2091.
[22] D. Surian, A.-G. Dunn, L. Orenstein, R. Bashir, E. Coiera, F.-T. Bourgeois, A shared latent space matrix factorisation method for recommending new trial evidence for systematic review updates, J. Biomed. Inform. 79 (2018) 32–40.
[23] D. Demner-Fushman, J. Lin, Answering clinical questions with knowledge-based and statistical techniques, Comput. Linguist. 33 (2007) 63–103.
[24] G. Salton, A. Wong, C.S. Yang, A vector space model for automatic indexing, Commun. ACM 18 (1975) 613–620.
[25] S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera, h-Index: a review focused in its variants, computation and standardization for different scientific fields, J. Informetr. 3 (2009) 273–289.
[26] F. Guilak, C.R. Jacobs, The H-index: use and overuse, J. Biomech. 44 (2011) 208–209.
[27] J.A. Aslam, V. Pavlu, E. Yilmaz, Measure-based metasearch, in: SIGIR '05 Proc. 28th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., New York, NY, USA, 2005, pp. 571–572.
[28] Cochrane Library. 2019-03-06. . Accessed: 2019-03-06. (Archived by WebCite® at http://www.webcitation.org/76gEpSUXO).
[29] C.J. Van Rijsbergen, Information Retrieval, Dept. of Computer Science, University of Glasgow, 1979.
[30] C.D. Manning, P. Raghavan, H. Schutze, Introduction to Information Retrieval, Cambridge University Press, 2008.
[31] LingPipe. 2019-03-06. URL:http://alias-i.com/lingpipe-3.9.3/demos/tutorial/medline/read-me.html. Accessed: 2019-03-06. (Archived by WebCite® at http://www.webcitation.org/76gF3XT7Q).

[32] S. Osinski, J. Stefanowski, D. Weiss, Lingo: Search Results Clustering Algorithm Based on Singular Value Decomposition, in: M.A. Klopotek, S.T. Wierzchon, K. Trojanowski (Eds.), Intell. Inf. Syst., Springer, 2004, pp. 359–368.
[33] S.C. Deerwester, S.T. Dumais, T.K. Landauer, G.W. Furnas, R.A. Harshman, Indexing by latent semantic analysis, J. Am. Soc. Inf. Sci. 41 (1990) 391–407.
[34] N. Fiorini, K. Canese, G. Starchenko, E. Kireev, W. Kim, V. Miller, M. Osipov, M. Kholodov, R. Ismagilov, S. Mohan, J. Ostell, Z. Lu, Best Match, New relevance search for PubMed, PLOS Biol. 16 (2018) e2005343.
[35] H. Turtle, Y. Hegde, S.-A. Rowe, Yet another comparison of Lucene and Indri performance, in: Proc. SIGIR 2012 Work. Open Source Inf. Retr., 2012, pp. 64–67.
[36] P. Yang, H. Fang, J. Lin, Anserini: reproducible ranking baselines using lucene, J. Data Inf. Qual. 10 (2018) 1–20.
[37] R.–B. Sağlam, T. Taşkaya Temizel, Automatic information timeliness assessment of diabetes web sites by evidence based medicine, Comput. Methods Programs Biomed. 117 (2014) 104–113.
[38] K. Zheng, Y. Chen, J. Adler-Milstein, A.-L. Rosenberg, D.T.-Y. Wu, Q. Mei, D.-A. Hanauer, How Do Healthcare Professionals Personalize Their Software? A Pilot Exploration Based on an Electronic Health Records Search Engine, Stud. Health Technol. Inform. 264 (2019) 1408–1412.
[39] B. Prados-Suárez, C. Molina, C. Peña Yañez, M. Prados De Reyes, Improving electronic health records retrieval using contexts, Expert Syst. Appl. 39 (2012) 8522–8536.
[40] D.A. Hanauer, D.T.Y. Wu, L. Yang, Q. Mei, K.B. Murkowski-Steffy, V.G.V. Vydiswaran, K. Zheng, Development and empirical user-centered evaluation of semantically-based query recommendation for an electronic health record search engine, J. Biomed. Inform. 67 (2017) 1–10.
[41] L.-K. Milliken, S.-K. Motomarry, A. Kulkarni, ARtPM: article retrieval for precision medicine, J. Biomed. Inform. 95 (2019) 103224.
[42] J. Serrano-Guerrero, E. Herrera-Viedma, J.A. Olivas, A. Cerezo, F.P. Romero, A Google Wave-based Fuzzy Recommender System to disseminate Information in University Digital Libraries 2.0, Inf. Sci. 181 (2011) 1503–1516.
[43] J. Serrano-Guerrero, F.P. Romero, J.A. Olivas, Hiperion: A fuzzy approach for recommending educational activities based on the acquisition of competences, Inf. Sci. 248 (2013) 114–129.
[44] R. Romero-Cordoba, J.A. Olivas, F.P. Romero, F. Alonso-Gonzalez, J. Serrano-Guerrero, An Application of Fuzzy Prototypes to the Diagnosis and Treatment of Fuzzy Diseases, Int. J. Intell. Syst. 32 (2016) 194–210.
[45] J.M. Ponte, B.W. Croft, A Language Modeling Approach to Information Retrieval, in: ACM (Ed.), Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., ACM Int, 1998, pp. 275–281.
[46] P.J. Garces, J.A. Olivas, F.P. Romero, Concept-matching IR systems versus word–matching information retrieval systems: Considering fuzzy interrelations for indexing Web pages, J. Am. Soc. Inf. Sci. Technol. 57 (2006) 564–576.
[47] J. Serrano-Guerrero, F. Chiclana, J.A. Olivas, F.P. Romero, E. Homapour, A T1OWA fuzzy linguistic aggregation methodology for searching feature-based opinions, Knowledge-Based Syst. 189 (2020) 105131.