WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

ENHANCING REVIEW QUALITY THROUGH THOUGHTFUL MESH TERM SELECTION

S. Narmatha [1], Dr. V. Maniraj [2]

[1].Research Scholar, Dept. of Computer Science, AVVM Sri Pushpam College (Affiliated to Bharathidasan University, Tiruchirappalli), Poondi, Thanjavur, Tamilnadu, India.

[2].Research Advisor, PG and Research Dept. of Computer Science, AVVM Sri Pushpam College (Affiliated to Bharathidasan University, Tiruchirappalli)., Poondi, Thanjavur, Tamilnadu, India.

ABSTRACT :

In the field of medicine, systematic reviews require comprehensive literature searches to ensure the credibility and precision of their conclusions. This search process is typically a collaborative effort between medical researchers and information professionals who are skilled in both clinical subject matter and advanced search strategies. Together, they construct detailed search queries, often using complex Boolean logic that combines free-text keywords with standardized indexing terms such as MeSH (Medical Subject Headings).

Creating these queries is a meticulous and time-intensive process. While the use of MeSH terms can significantly enhance the quality and relevance of search results, selecting the most appropriate MeSH terms is not always straightforward. Many information professionals may lack deep familiarity with the MeSH system, resulting in uncertainty and the underutilization of this powerful resource.

This study explores automated approaches for recommending MeSH terms based on an initial Boolean search query composed solely of free-text terms. The objective is to help users identify the most effective MeSH terms to integrate into their queries for systematic reviews. The study evaluates multiple strategies for MeSH term recommendation and assesses their impact on the performance of the queries in terms of document retrieval, ranking accuracy, and iterative refinement.

INTRODUCTION

Systematic reviews in medicine serve as comprehensive evaluations of the existing literature on narrowly defined research questions. Regarded as the gold standard in evidence-based medicine, they play a critical role in guiding clinical practice and healthcare policy. A key component of this process is the development of effective search strategies to retrieve relevant studies, most often through the formulation of detailed Boolean queries. These queries are typically crafted in collaboration between researchers and information specialists, who bring domain-specific knowledge and expertise in search methodology.

PubMed is the primary database for conducting medical literature searches, and it leverages the Medical Subject Headings (MeSH) ontology to organize its vast and growing collection of biomedical research. MeSH is a controlled vocabulary arranged in a hierarchical structure—from broad categories such as "Anatomy" to highly specific terms like "Eye." When combined with free-text terms, MeSH terms can enhance both the precision and recall of literature searches by disambiguating keyword meanings and capturing relevant articles more effectively.

Research has consistently demonstrated that queries incorporating MeSH terms yield more relevant results than those based solely on free-text input. However, identifying the appropriate MeSH terms for a given query remains a complex and often overwhelming task—even for experienced professionals—due to the extensive size of the MeSH vocabulary, which now includes over 29,000 unique descriptors.

To assist users, PubMed employs a feature called Automatic Term Mapping (ATM), which attempts to map free-text input to MeSH terms, journal titles, or author names. While ATM improves query performance in some domains, such as genomics, it has several known shortcomings. These include difficulty handling acronyms, inconsistent mapping of synonyms, and confusion between MeSH terms and journal names. Despite its broad use, the impact of ATM on systematic review search quality has not been rigorously studied.

This paper investigates methods for suggesting MeSH terms to support the construction of Boolean queries for systematic reviews. Specifically, it focuses on scenarios where an information specialist begins with a free-text-only query and seeks automated guidance in identifying relevant MeSH terms to improve the overall effectiveness of the literature search.

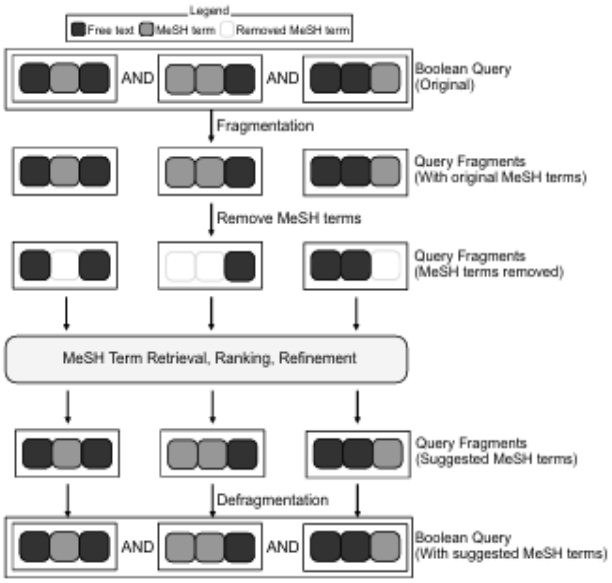


Figure 1: Workflow for Recommending MeSH Terms

The recommendation of MeSH terms is guided by a process involving retrieval, ranking, and refinement. To assess different approaches for suggesting MeSH terms, we consider four key evaluation criteria:
(1) How effectively the method identifies relevant MeSH terms;
(2) The quality and relevance of the ranking order assigned to those terms;
(3) The degree to which the ranking improves through refinement; and
(4) The effectiveness of the recommended MeSH terms in enhancing literature retrieval when added to an incomplete Boolean query.
It's also important to recognize that the number of MeSH terms suggested for a given query fragment may vary—sometimes exceeding, and at other times falling short of, the number initially present.

RELATED WORK

This study presents a framework for evaluating how effectively MeSH term recommendations enhance systematic review search queries. Using a collection of existing Boolean queries from systematic reviews, we also introduce novel methods for generating MeSH term suggestions. Our work extends previous research that explores computational approaches to support the development and refinement of Boolean search strategies for evidence synthesis.
The key contributions of our study are as follows:

(1) We define a new task centered on recommending MeSH terms for systematic review searches. This task is designed with the viewpoint of an information specialist in mind—someone working with a Boolean query that initially lacks MeSH terms and aims to enhance it.

(2) We conduct a practical evaluation of MeSH term suggestion methods, with a focus on how well each approach ranks terms based on their relevance to the original query.

(3) We carry out an empirical comparison of the impact of different MeSH suggestion strategies by measuring how well they improve the retrieval of relevant abstracts when incorporated into Boolean queries.

MeSH Term Suggestion

We outline our approach for generating MeSH term suggestions tailored to Boolean queries. Given the often complex and deeply nested structure of Boolean queries used in systematic review searches, we propose applying MeSH recommendations at the **fragment level** rather than across the full query. A *query fragment* is defined as a semantically related portion of the overall query—typically composed of clauses containing either free-text or MeSH terms—connected by the logical OR operator. Each individual text clause within the fragment is treated as a standalone Boolean condition.
For an illustration of how these fragments are constructed and used in the MeSH recommendation process, refer to Figure 1. Note that the OR relationships between clauses are implied in the figure. These fragments allow us to perform a targeted analysis of MeSH term suggestions, specifically focusing on how well they perform in terms of retrieval quality, ranking relevance, and overall ranking improvement.
Following the suggestion phase, we perform a **defragmentation step** where the recommended MeSH terms are reintegrated into a complete Boolean query. This revised query is then compared to the original to evaluate its performance. Our method for MeSH suggestion operates in three core stages: **retrieval**, **ranking**, and **refinement**. Each of these steps is described in more detail below.

MeSH Term Retrieval

The first stage of our process involves retrieving candidate MeSH terms. We use three distinct techniques to extract relevant MeSH terms from query fragments composed entirely of free-text clauses:

1. **Automatic Term Mapping (ATM)**: The full query fragment is submitted to the PubMed Entrez API using only unqualified free-text. PubMed automatically attempts to map each term to entries in one of three categories: MeSH headings, journal names, or author names. If a direct match isn't found, the query is decomposed into individual terms, which are then checked for matches. Only MeSH terms are retained in the final output.
2. **MetaMap**: Each sentence in the query fragment is processed using MetaMap, a tool developed by the National Library of Medicine. MetaMap identifies biomedical concepts and maps them to the UMLS Metathesaurus, filtering results to include only MeSH-sourced entities. Each term in the fragment is matched to its corresponding MeSH terms, and MetaMap assigns a confidence score to each.
3. **Unified Medical Language System (UMLS)**: We utilize Elasticsearch (v7.6) to index selected UMLS tables (MRCONSO, MRDEF, MRREL, and MRSTY from version 2019AB). After filtering out non-MeSH data, individual clauses are queried against this index. The results are limited to MeSH-derived synonyms, and each candidate MeSH term is assigned a BM25 relevance score (Elasticsearch's default ranking metric).

Because both MetaMap and UMLS may return duplicate MeSH terms for the same clause, we apply a **rank fusion strategy** (specifically, **CombSUM**) to aggregate the scores. This approach boosts the rank of terms that not only appear frequently across multiple methods but also have high individual scores. The goal is to prioritize MeSH terms that are both highly relevant and consistently retrieved, while demoting overly common or weakly scored terms.

Table 1: Ranking Characteristics of Suggested Terms

Feature	Description
$\|q\|$	Total free-text terms in a fragment
l_{d_e}	Length of description of MeSH term e
$\sum_{q_i} IDF(q_i)$	Sum IEF of free-text terms
$\sum_{q_i} TF(q_i, d_e)$	Sum TF of free-text terms in d_e
$\sum_{q_i} TF(q_i, d_e)IDF(q_i)$	Sum TF of free-text terms in d_e
$\text{score}_{LM}(q, d_e)$	LM score of free-text terms for d_e
$\text{score}_{BM25}(q, d_e)$	BM25 score of free-text terms for d_e
$\text{score}_{SDM}(q, d_e)$	SDM score of free-text terms for d_e
$QCE(q, e)$	Whether the free-text terms contain e
$ECQ(q, e)$	Whether e contains any free-text terms
$ECQ(q, e)$	Whether e is equal to the free-text terms

MeSH Term Ranking

Once candidate MeSH terms are retrieved, we prioritize them using an entity ranking approach that adapts features originally proposed by Balog. We utilize a total of eleven features, which are detailed in Table 1. To enrich the representation of each MeSH term (denoted as d_e), we extract additional contextual information from their corresponding Wikipedia pages through web scraping.

Each retrieval method—ATM, MetaMap, and UMLS—generates a set of features specific to the MeSH terms it identifies. We define *positive examples* as those MeSH terms that appear in the original query fragment, and *negative examples* as those that do not. These examples are labeled with binary values (1 for positive, 0 for negative), forming the basis for training a learning-to-rank (LTR) model tailored to each retrieval technique.

In addition to training individual LTR models, we implement a **rank fusion strategy** to combine the normalized scores from all three retrieval approaches. This fusion method creates a unified ranking that highlights MeSH terms which consistently appear with high scores across multiple techniques. The motivation behind this is that each retrieval method may return different sets of MeSH terms and rank them differently; by merging the rankings, we aim to increase the visibility of terms that are both frequently retrieved and highly ranked across methods.

MeSH Term Refinement

The refinement phase focuses on enhancing the accuracy and relevance of the selected MeSH terms. The central goal is to filter out less relevant terms by applying a rank-based cut-off, retaining only the most pertinent MeSH terms for each query fragment.

This cut-off is defined using the κ parameter, which represents the top percentage of ranked MeSH terms to be considered. We evaluate values of κ ranging from 5% to 95%, in increments of 5%. To ensure at least one term is suggested for every fragment, we assign a score of 0 to the top-ranked MeSH term. Since multiple terms may receive the same score, ties are possible.

To address tied scores at the κ threshold, we take a cautious approach. Rather than arbitrarily including or excluding tied terms, we treat all tied terms at the cut-off point as a collective unit. Their shared contribution is summed as a combined "gain." This method ensures that when ties occur near the top

of the ranking, those MeSH terms are more likely to be included, whereas ties lower in the ranking may be excluded altogether. Ultimately, this strategy leads to more meaningful inclusion of top-ranked, equally scored MeSH terms and avoids the random omission of potentially relevant results.

Evaluation

To assess the effectiveness of our MeSH term suggestion approach, we conduct a retrospective evaluation using previously constructed systematic review queries as a benchmark. These existing queries, which already include manually selected MeSH terms, serve as our gold standard. However, it's important to acknowledge that this baseline may inherently favor terms generated via PubMed's Automatic Term Mapping (ATM), since ATM may have influenced the creation of the original queries.

Our methods are designed to identify potentially useful MeSH terms that may not appear in the original query fragments. To test the value of these new suggestions, we go beyond comparison with the original terms and evaluate how well our recommended queries retrieve relevant research. This dual approach allows us to challenge the assumption that the original MeSH terms are always optimal.

A few limitations must be considered in this evaluation. First, query fragments must be reassembled into complete Boolean queries to allow for meaningful comparison. Second, our retrieval experiments may identify relevant studies that were not part of the original relevance assessments, meaning they haven't been judged. To handle this uncertainty, we apply a three-pronged evaluation strategy:

1. *Lower bound:* Assumes all unjudged studies are irrelevant.
2. *Upper bound:* Assumes all unjudged studies are relevant.
3. *MLE-based estimate:* Uses maximum likelihood estimation based on the judged studies to estimate how many unjudged studies might be relevant. This estimation is used to randomly sample unjudged documents likely to be relevant, based on the proportion of relevant documents in the original candidate pool.

We evaluate how well each of the three MeSH term retrieval methods (ATM, MetaMap, UMLS) performs using *precision* and *recall*. To measure the quality of MeSH term ranking produced by our learning-to-rank (LTR) models, we use *normalized Discounted Cumulative Gain (nDCG)* and *reciprocal rank*.

In evaluating the impact of suggested MeSH terms on systematic review searches, we reconstruct full Boolean queries from the refined query fragments and test their retrieval performance. For this, we apply standard information retrieval metrics, including *precision*, *recall*, and *Fβ scores* for $\beta = \{0.5, 1, 3\}$. Since the results for $\beta = 0.5$ and $\beta = 3$ follow the same patterns as $\beta = 1$, only the F1 scores are reported for clarity.

All retrieval experiments are executed via the PubMed Entrez API, with a fixed date range applied to ensure consistency and reproducibility across queries—important given the continuously evolving nature of the PubMed database. Finally, in both ranking and retrieval evaluations, we measure performance in two scenarios:

(i) using all retrieved MeSH terms, and

(ii) applying a score threshold to filter out lower-ranking MeSH suggestions.

EFFICACY OF SUGGESTIONS

MeSH Term Retrieval

We begin by assessing how well each MeSH term recommendation method retrieves relevant terms based on individual query fragments. Table 2 presents the precision (P) and recall (R) metrics for each approach.

Among the three primary retrieval methods, **UMLS** consistently demonstrates the highest recall, indicating its ability to identify a larger portion of relevant MeSH terms. However, this comes at the cost of lower precision—UMLS also returns more irrelevant or less useful terms due to its broader retrieval scope.

In contrast, **ATM** achieves the highest precision, meaning it is more selective and tends to retrieve fewer but more accurate MeSH terms. **MetaMap** generally performs between ATM and UMLS in both metrics.

When we combine all methods using a **fusion approach**, we observe the highest recall overall, thanks to the inclusion of MeSH terms identified by any of the three techniques. However, this fusion approach does not outperform ATM or MetaMap in precision, as the inclusion of more terms increases the likelihood of incorporating less relevant results.

In summary, our key observations are:

(i) **UMLS** excels in maximizing recall, capturing a broader range of relevant MeSH terms,

(ii) **ATM** offers the best precision, returning fewer but more targeted terms, and

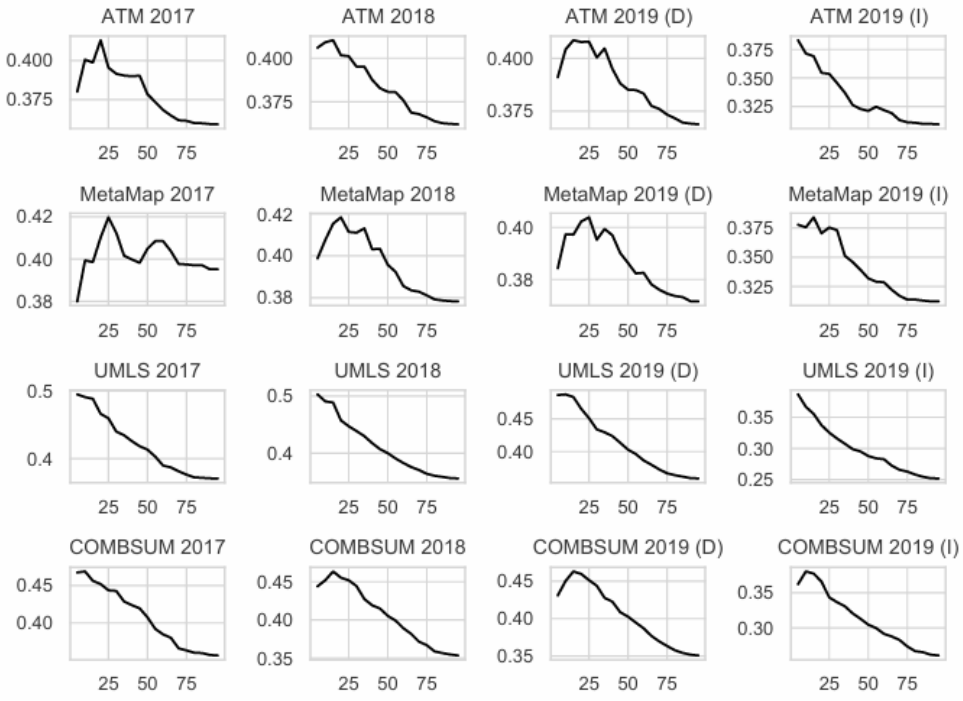
(iii) **Combining methods** boosts recall but tends to reduce precision due to the increased term pool.

Table 2:Accuracy of MeSH Term Suggestion Methods Measured by Precision

Method	P	R	RR	R@5	R@10	nDCG@5	nDCG@10
2017/A	0.3027	0.3718	0.4614	0.3504	0.3576	0.3601	0.3494
2017/A-C	0.3600	0.2421*	0.4047	0.2362*	0.2403*	0.2713*	0.2652*
2017/M	0.3496	0.3818	0.5730	0.3659	0.3793	0.4218	0.4102
2017/M-C	0.4333	0.2868	0.4739	0.2800	0.2868	0.3105	0.3024
2017/U	0.2571	0.4659	0.5910	0.4214	0.4475	0.4518	0.4469
2017/U-C	**0.4819**	0.2846	0.5295	0.2831	0.2846	0.3446	0.3329
2017/F	0.2446	**0.5281**	**0.6207**	**0.4519**	**0.4958**	**0.4915**	**0.4971**
2017/F-C	0.4517	0.3458	0.4897	0.3391	0.3450	0.3581	0.3475
2018/A	0.3287	0.3772	0.4967	0.3261	0.3703	0.3719	0.3838
2018/A-C	0.3742	0.2129*	0.3933*	0.1983*	0.2024*	0.2417*	0.2392*
2018/M	0.3088	0.3360	0.4630	0.3007	0.3353	0.3470	0.3380
2018/M-C	0.3704	0.2257*	0.3940	0.2237	0.2257	0.2689	0.2523
2018/U	0.2885	0.4641	**0.6007**	0.4188	0.4565	0.4615	0.4569
2018/U-C	**0.4711**	0.2643	0.5278	0.2575	0.2643	0.3405	0.3292
2018/F	0.2661	**0.5024**	0.5793	**0.4316**	**0.4798**	**0.4633**	**0.4629**
2018/F-C	0.4212	0.3456	0.4754	0.3233	0.3387	0.3646	0.3550
2019/D/A	0.3399	0.3558	0.5933	0.3503	0.3558	0.3910	0.3671
2019/D/A-C	0.5071	0.2628	0.5583	0.2628	0.2628	0.3222	0.2990
2019/D/M	0.3864	0.3053	0.6167	0.3003	0.3053	0.3810	0.3530
2019/D/M-C	**0.5458**	0.2303	0.5417	0.2253	0.2303	0.3209	0.2963
2019/D/U	0.2651	0.4528	0.6058	0.4428	0.4478	0.4680	0.4348
2019/D/U-C	0.4850	0.2619	0.5167	0.2619	0.2619	0.3159	0.2866
2019/D/F	0.2266	**0.4778**	**0.6725**	**0.4622**	**0.4678**	**0.5000***	**0.4680***
2019/D/F-C	0.5068*	0.3419	0.4933	0.3419	0.3419	0.3618	0.3302
2019/I/A	0.3110	0.3631	0.4469	0.3379	0.3560	0.3409	0.3445
2019/I/A-C	0.3703	0.2195*	0.3868	0.2195*	0.2195*	0.2511*	0.2484*
2019/I/M	0.2651	0.3395	0.4253	0.3071	0.3368	0.3131	0.3205
2019/I/M-C	0.3368	0.2178	0.3682	0.2088	0.2178	0.2389	0.2370
2019/I/U	0.2779	0.4175	0.4340	0.3765	0.4114	0.3516	0.3663
2019/I/U-C	0.3415	0.2209	0.3769	0.2146	0.2209	0.2447	0.2464
2019/I/F	0.2566	**0.4462**	**0.5283**	**0.4120**	**0.4373**	**0.4190**	**0.4233**
2019/I/F-C	**0.4074**	0.3229	0.4336	0.3112	0.3166	0.3271	0.3255

Ranking of MeSH Terms

We proceed to assess how well the learning-to-rank (LTR) model orders the MeSH terms retrieved by each method. This evaluation focuses on metrics such as reciprocal rank (RR), Recall at rank k (R@k), and normalized Discounted Cumulative Gain at rank k (nDCG@k), as presented in Table 2.
Our results reveal two main insights:
(i) The UMLS approach generally achieves higher recall than ATM and MetaMap, which translates into superior ranking performance across most evaluated metrics.
(ii) The fusion strategy, which combines results from all methods, tends to produce even more effective MeSH term rankings, except in the case of the reciprocal rank metric for the 2018 dataset, where it underperforms slightly.



Refinement of MeSH Terms

Finally, we examine the effect of applying a ranking cut-off to the MeSH terms, removing those ranked below a certain threshold. This cut-off is controlled by a parameter κ. Figure 2 illustrates the tuning of this parameter on the training datasets, with the x-axis representing different κ values and the y-axis showing the corresponding F1 scores.

The graphs reveal noticeable spikes, which likely result from including or excluding groups of tied MeSH terms. These fluctuations are particularly pronounced in the MetaMap and ATM methods, where MeSH term scores tend to be less distinct. Conversely, the UMLS and fusion methods produce smoother curves due to clearer score differentiation among terms.

Table 2 (marked as with-C) summarizes how refinement affects MeSH term recommendation performance. Our results indicate that while applying refinement generally increases precision, it also causes a drop in recall. This decrease in recall negatively impacts the overall ranking quality. In fact, refinement often yields poorer performance compared to using all MeSH terms and is frequently less effective than the ATM method without refinement.

The Impact of Fusion

Regarding recall, the fusion method without refinement generally showed improved results, with the exception of one case in 2018. This notable increase in recall is likely due to the fusion process combining all MeSH terms suggested by the ATM, MetaMap, and UMLS methods. However, this straightforward fusion approach does not seem to improve the precision of Boolean queries.

These findings imply that semi-automatic MeSH term recommendations could be valuable tools for information specialists, who can use their judgment to select the most relevant MeSH terms and thereby enhance search performance.

Interestingly, the refined fusion strategy did not outperform other methods on any metric or dataset. Applying refinement typically lowered recall and had little impact on precision across all methods. This indicates that, for effective systematic review searches, carefully choosing the right MeSH terms is more important than simply increasing the number of terms included.

4.3 Case Study

Building on the results from Section 4.2, we conduct a detailed case study to explore why the performance of MeSH term suggestion methods can vary. We selected query CD010542 from the 2017 CLEF TAR dataset as a representative example, after reviewing multiple queries generated by our MeSH suggestion techniques.

Following the methodology illustrated in Figure 1, we first remove all MeSH terms from the original query, then split it into individual fragments. For each fragment, we generate MeSH term suggestions and subsequently reconstruct the query by incorporating these recommended MeSH terms into the Boolean structure. We then execute searches on PubMed using these reconstructed queries and compare their retrieval effectiveness against the original query and a version without any MeSH terms.

The query fragments and their associated MeSH term suggestions are detailed in Table 4. Table 5 shows that the results for ATM and MetaMap are identical despite originating from different query fragments, reinforcing our earlier observation from Section 4.2.3: the choice of MeSH terms plays a pivotal role in the success of Boolean queries used in systematic review literature searches.

Table 3: Evaluation of MeSH term recommendations within Boolean queries for the systematic review search of CD010542. Here, **O** denotes the original query, **R** indicates the version with MeSH terms removed, **A** stands for Automated Term Mapping (ATM), **M** represents MetaMap, and **U** corresponds to the Unified Medical Language System (UMLS).

Method	P	P (MLE)	P (Opt)	F1	F1 (MLE)	F1 (Opt)	R	R (MLE)	R (Opt)
O	0.0207	0.0598	0.6622	0.0405	0.1126	0.7968	0.9000	1.0000	1.0000
R	0.0274	0.0608	0.6140	0.0531	0.1143	0.7594	0.9000	0.9524	0.9951
A	0.0167	0.0583	0.7574	0.0327	0.1100	0.8611	0.9000	0.9692	0.9976
M	0.0167	0.0583	0.7574	0.0327	0.1100	0.8611	0.9000	0.9692	0.9976
U	0.0256	0.0598	0.6382	0.0499	0.1126	0.7778	0.9000	0.9545	0.9956

When comparing UMLS with the other approaches, we found that UMLS generally achieved higher precision than the other MeSH term–enhanced queries, except under optimistic evaluation metrics. This suggests that the MeSH term included in fragment 1 of the original query might actually reduce the query's effectiveness, highlighting that the MeSH terms originally selected are not always optimal. Notably, the version of the query without any MeSH terms performed better in precision than all other methods, while maintaining an equivalent recall level. This indicates that MeSH terms may not always improve search results for every query, and adopting a more adaptable strategy to select the most pertinent MeSH terms could improve the overall performance of MeSH term recommendation methods.

We propose the following approaches:

(i) In a fully automated setup, a classification model can be developed to assess the usefulness of MeSH terms, coupled with a stopping mechanism to decide the best point to stop adding terms for optimal results.

(ii) In a semi-automated approach, this classification model could be used repeatedly to assign confidence scores to MeSH terms, offering information specialists helpful guidance when selecting which MeSH terms to include while constructing the query.

CONCLUSION

This study presents a novel approach for recommending MeSH terms to enhance Boolean queries used in systematic review literature searches. We conducted a comprehensive evaluation of various MeSH term suggestion methods, examining their performance in retrieval, ranking, and refinement. These methods were compared against PubMed's existing MeSH suggestion system, Automated Term Mapping (ATM). Our results demonstrated that both MetaMap and UMLS methods improved the retrieval effectiveness of Boolean queries. As expected, the greatest gains were achieved by combining all three approaches through rank fusion. Our techniques effectively overcame some of ATM's limitations in semantic understanding. MetaMap and UMLS provided more relevant MeSH term suggestions than ATM, leading to better retrieval results. Although this sometimes resulted in a slight decrease in recall, this loss generally involved only a few studies and is unlikely to affect the overall conclusions of a systematic review. Selecting suitable MeSH terms for Boolean queries remains a complex task for human experts, especially in systematic review contexts. The findings from this work offer valuable insights for both the information retrieval and systematic review communities. Furthermore, our methods have potential applications in automated query generation tasks, such as those in CLEF TAR, and could be integrated into existing tools to support information specialists in crafting more effective search queries.

REFERENCES:

1. Hliaoutakis, Angelos. "Semantic similarity measures in MeSH ontology and their application to information retrieval on Medline." *Master's thesis* (2005).
2. Hassan, Sahar, Franck Hétroy, and Olivier Palombi. "Ontology-guided mesh segmentation." *FOCUS K3D Conference on Semantic 3D Media and Content*. 2010.
3. Díaz-Galiano, Manuel Carlos, et al. "Integrating mesh ontology to improve medical information retrieval." *Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007, Revised Selected Papers 8*. Springer Berlin Heidelberg, 2008.
4. Soualmia, Lina Fatima, Christine Golbreich, and Stéfan Jacques Darmoni. "Representing the MeSH in OWL: Towards a semi-automatic migration." *KR-MED*. Vol. 102. 2004.
5. Osborne, John D., et al. "Interpreting microarray results with gene ontology and MeSH." *Microarray Data Analysis: Methods and Applications* (2007): 223-241.
6. Elberrichi, Zakaria, Belaggoun Amel, and Taibi Malika. "Medical Documents Classification Based on the Domain Ontology MeSH." *arXiv preprint arXiv:1207.0446* (2012).
7. Elberrichi, Zakaria, Malika Taibi, and Amel Belaggoun. "Multilingual medical documents classification based on mesh domain ontology." arXiv preprint arXiv:1206.4883 (2012).
8. Ayeldeen, Heba, Aboul Ella Hassanien, and Ali Ali Fahmy. "Evaluation of semantic similarity across MeSH ontology: a Cairo University thesis mining case study." 2013 12th Mexican International Conference on Artificial Intelligence. IEEE, 2013.
9. C.Senthil Selvi, Dr. N. Vetrivelan, " An Efficient Information Retrieval In Mesh (Medical Subject Headings) Using Fuzzy", Journal of Theoretical and Applied Information Technology 2019. ISSN: 1992-8645, Vol.97. No 9, Page No: 2561-2571.
10. Rajkumar, V., and V. Maniraj. "HYBRID TRAFFIC ALLOCATION USING APPLICATION-AWARE ALLOCATION OF RESOURCES IN CELLULAR NETWORKS." Shodhsamhita (ISSN: 2277-7067) 12.8 (2021).
11. Ivanova, Valentina, et al. "Debugging taxonomies and their alignments: the ToxOntology-MeSH use case." First International Workshop on Debugging Ontologies and Ontology Mappings, Galway, Ireland, October 8, 2012.. Linköping University Electronic Press, 2012.