

Article

Guidance for Clinical Evaluation under the Medical Device Regulation through Automated Scoping Searches

Fu-Sung Kim-Benjamin Tang , Mark Bukowski , Thomas Schmitz-Rode  and Robert Farkas *

Department of Science Management, Institute of Applied Medical Engineering, RWTH Aachen University—University Hospital Aachen, 52074 Aachen, Germany; tang@ame.rwth-aachen.de (F.-S.K.-B.T.); bukowski@ame.rwth-aachen.de (M.B.); smiro@ame.rwth-aachen.de (T.S.-R.)

* Correspondence: farkas@ame.rwth-aachen.de; Tel.: +49-241-80-80740

Featured Application: The developed automated search methods can support medical device manufacturers with a first orientation before the clinical evaluation for the European Medical Device Regulation. A first orientation via automated initial scoping searches could lower the cost of scarce resources and manual labor.

Abstract: The Medical Device Regulation (MDR) in Europe aims to improve patient safety by increasing requirements, particularly for the clinical evaluation of medical devices. Before the clinical evaluation is initiated, a first literature review of existing clinical knowledge is necessary to decide how to proceed. However, small and medium-sized enterprises (SMEs) lacking the required expertise and funds may disappear from the market. Automating searches for the first literature review is both possible and necessary to accelerate the process and reduce the required resources. As a contribution to the prevention of the disappearance of SMEs and respective medical devices, we developed and tested two automated search methods with two SMEs, leveraging Medical Subject Headings (MeSH) terms and Bidirectional Encoder Representations from Transformers (BERT). Both methods were tailored to the SMEs and evaluated through a newly developed workflow that incorporated feedback resource-efficiently. Via a second evaluation with the established CLEF 2018 eHealth TAR dataset, the more general suitability of the search methods for retrieving relevant data was tested. In the real-world use case setting, the BERT-based method performed better with an average precision of 73.3%, while in the CLEF 2018 eHealth TAR evaluation, the MeSH-based search method performed better with a recall of 86.4%. Results indicate the potential of automated searches to provide device-specific relevant data from multiple databases while screening fewer documents than in manual literature searches.

Keywords: medical device legislation; search engine; PubMed; medical subject headings; information storage and retrieval; algorithms; European Union; databases; bibliographic; device approval



Citation: Tang, F.-S.K.-B.; Bukowski, M.; Schmitz-Rode, T.; Farkas, R. Guidance for Clinical Evaluation under the Medical Device Regulation through Automated Scoping Searches. *Appl. Sci.* **2023**, *13*, 7639. <https://doi.org/10.3390/app13137639>

Academic Editor: Zhi-Ting Ye

Received: 3 May 2023

Revised: 15 June 2023

Accepted: 25 June 2023

Published: 28 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

To increase patient safety in Europe, the regulatory framework Medical Device Regulation (MDR) was adopted in 2017 and became applicable in May 2021 [1].

The novel MDR forces medical device manufacturers worldwide to comply quickly with the drastically increased requirements for European market access regarding post-market surveillance [2], medical device traceability [2], and more rigorous pre-market testing [3]. Moreover, new rules for the risk classification of medical devices cause previously lower-classified devices to shift to a higher-risk class [3]. These requirements and rules increase the burden on companies to obtain approval for innovative medical products.

Consequently, medical device suppliers and manufacturers may disappear from the market. Small and medium-sized enterprises (SMEs) are assumed to be especially at risk when facing the extended time-to-market [4,5]. The demise of smaller companies

or startups could lead to a loss of innovative power and supports the formation of an oligopoly of larger companies [4].

One of the most crucial aspects of adapting to the MDR is the clinical evaluation throughout the whole product life cycle, starting from the initial certification until the post-market clinical follow-up (PMCF). The associated clinical evaluation report requires the systematic collection and evaluation of relevant clinical data [6].

This acquisition and unbiased review of clinical data in the form of scientific literature or clinical trials is now mandatory across all risk classes [3,7]. Not only clinical data pertinent to the respective medical device but also available alternative treatment options for its purpose must be considered [8]. Alternatively, it is possible to demonstrate equivalence to an already certified medical device under the MDR [9], using published data instead of conducting clinical trials.

Even before an approval process is initiated, a valid picture of the existing clinical knowledge based on an unbiased review is necessary to decide how to proceed. At this stage, automation of the review process is both possible and necessary to accelerate development.

Methods to automate such review processes were evaluated in the scope of the Conference and Labs of Evaluation Forum (CLEF) with the so-called eHealth challenges regarding Technology Assisted Reviews (TAR) for systematic reviews (SR) in Empirical Medicine [10], in which relevant documents must be automatically retrieved for a given topic. The best-performing method achieved an almost perfect overall recall of relevant documents, while the recall regarding the first search results and consequently the workload reduction could be optimized further [11].

One method to improve recall at earlier stages of the screening process could be to use terms from the Medical Subject Headings (MeSH), used by PubMed for indexing publications. The application of MeSH terms helps with maximizing recall and formulating more effective search queries [12]. Since the selection of suitable MeSH terms poses a difficult task even for experts, automated selection strategies exist [13,14]. However, limiting the search to the most appropriate terms leads to a loss of information and thus to a reduction in the retrieved relevant documents.

To overcome the need of forming queries by selecting suitable keywords and Boolean operators, recent methods explored the possibility of a semantic search, rather than a term-based search [15]. In the current research, semantic search applications via Bidirectional Encoder Representations from Transformers (BERT) language model variations perform well and are used to retrieve domain-specific information [16,17] or to retrieve relevant documents for TARs [18]. While the performant semantic BERT-based search is not restricted by inherent limitations of keyword-based searches, the problem of transparency and understandability of the search arises and is currently being researched [19,20]. Thus, it is currently difficult to interpret and understand the results of a BERT-based search.

While various methods for SR and TAR are established, their applicability when searching for clinical data on medical devices has not yet been investigated. Applying existing methods might require adaptations to overcome the challenges of retrieving data for medical devices and entails structural limitations [21]. Firstly, input for a search, such as product descriptions provided by SMEs, usually does not follow strict wording and contains precise inclusion criteria as is the case with review protocols. Secondly, medical device assessment entails a higher level of complexity compared to other health technologies, such as the assessment of pharmaceuticals [22]. This indicates that a transfer of methods tailored to one application, such as the retrieval of literature for diagnostic test accuracy reviews in the scope of CLEF challenges, to another, such as the retrieval of clinical data on medical devices, might be problematic. Thirdly, the existing methods retrieve documents from only one database, rather than heterogeneous data from multiple databases. Hence, leveraging methods of the CLEF challenges may only retrieve a fraction of potentially relevant data, since clinical trials, for instance, are excluded.

However, consistent guidance is scarce when searching for clinical trials, causing their inclusion to be particularly challenging [23]. While no consistent search method is established to locate diverse types of clinical trials, existing strategies rely on query formulation by either researchers or assisting experts, such as librarians [24]. Limited best-practice guidance is available for searching clinical data on medical devices [21], and previously conducted searches were resource-intensive and required the knowledge of review experts [9,25].

These limitations pose a major challenge, especially for SMEs lacking the required financial resources and expertise [26]. Up to 79% of German medical device manufacturers reported difficulties with the clinical evaluation of their devices and the provision of sufficient clinical data [27]. This observation contributes to an overall European trend, as these companies make up 41% of the EU industry revenue [27].

While checklists, guidance documents, and services offered by consulting agencies for this purpose exist, resources are lacking to empower SMEs in particular [28] to conduct initial technology-assisted scoping searches without expert knowledge.

Therefore, our aim is to provide assistance with the initial scoping search as a decision support and basis for possible future systematic reviews for SMEs challenged by the increased MDR requirements. Orientation should be provided to SMEs by automatically retrieving relevant clinical data for the respective medical device without expert assistance to formulate complex queries or define a search strategy. Initial results then serve as a starting point to analyze the state of the art in literature for the intended use of the medical device, thus adding transparency as a requirement to the search for it to be understandable. Enabling such low-threshold initial scoping searches then provides the basis for experts to develop a strategy for a systematic review of clinical evidence [29].

To meet the specific requirements of SMEs, the search process must be tailored to the medical device, e.g., by using all relevant bibliographic references available from prior knowledge or even using only a free-text product description as the search input.

Thus, in this paper, we conceptualize novel automated search strategies for assisting with the search and screening process for relevant clinical data while incorporating relevance feedback regarding the results. The automated search methods integrate device-specific documents from heterogeneous bibliographic databases for publications and clinical trials, namely PubMed and ClinicalTrials.gov. The search methods are usable without expert knowledge and involve screening only a relatively low number of documents to lower barriers regarding applicability [29]. Furthermore, transparency and understandability should be ensured.

With the aim to fulfill these criteria, two search methods and a workflow to incorporate them are developed and tested with SMEs. We hypothesize that a state-of-the-art BERT-based semantic search method performs best, but will pose a black-box problem, limiting understandability. To test our hypothesis and to see if a more transparent term-based search can reach similarly high performance, we conceptualize a novel MeSH-based search method. The MeSH-based search method uses a MeSH-term weighting scheme to overcome the need for term selection strategies and incorporates the feedback of screening results to refine the search. The BERT-based search leverages an out-of-the-box semantic BERT language model trained on the biomedical domain and will, analogous to other current methods, not be further modified to incorporate feedback into the search process [16,30,31].

The newly conceptualized and applied resource-efficient workflow incorporates clinical data and feedback in the search process to assist medical device manufacturers with both search methods. The precision of both search methods in the first exploratory insights will be evaluated and compared with an additional qualitative side-by-side analysis of the results with the help of two real-world SME use cases. Moreover, the performance of both search methods will be investigated with the established CLEF 2018 eHealth TAR dataset to gain further insights regarding the general, albeit not medical device-specific, search for clinical data.

2. Materials and Methods

This section presents details of the data used, i.e., databases, search input collected from SMEs, and the validation dataset. In addition, the MeSH-based and BERT-based automated search methods for finding relevant documents for a given medical device and the respective evaluation are outlined, and the workflow developed for collaboration with SMEs will be presented.

2.1. Data

To match the MDR requirements, all relevant information for the application field of the medical device must be retrieved from various heterogeneous databases [32]. We investigated bibliographic databases containing biomedical literature, namely Cumulative Index to Nursing and Allied Health Literature (CINAHL), ClinicalTrials.gov, Cochrane Library, Clinical Trials Information System (CTIS), Embase, PubMed and the International Clinical Trials Registry Platform (ICTRP). According to the EU guidelines on medical devices [33] and due to the availability of APIs and relational database access for unrestricted query formulations, this paper focuses on the databases PubMed and ClinicalTrials.gov. To implement the search methods, locally accessible instances of PubMed and ClinicalTrials.gov were configured: the local PubMed database was set up by integrating and indexing PMDB [34,35] and to leverage data from ClinicalTrials.gov, we used Access to Aggregate Content of ClinicalTrials.gov (AACT) [36]. For each PubMed publication, the publication title and text were concatenated to serve as the document text, and annotated MeSH terms were retrieved for the ranking of the MeSH-based method. Clinical trials from ClinicalTrials.gov were represented via respective concatenated brief titles and brief summaries, while the MeSH terms present in the MeSH term and keywords fields were additionally used for the later ranking of the MeSH-based method.

Three German SMEs participated in the feedback-driven development and evaluation of the two search methods. The feedback from the first SME served only as a basis for workflow construction. Thus, two parallel SME test cases were fully analyzed with corresponding results provided by one representative for each SME. The representatives were the head of quality management for SME A and the founder of the company for SME B. Each representative provided medical device information and feedback throughout the workflow.

Information about the medical devices of the SMEs was gathered via a short questionnaire with a free-text description of the medical device, how it differs from other products, and—if present—relevant publications, clinical studies, or specific search filters such as a time frame. The data provided by each SME are highlighted in Table 1. Both SMEs provided a product description and seed publications; however, no relevant clinical trials were provided as seeds.

Table 1. Overview of the received data from both SMEs serving as use cases characterizing the medical device in question by a description and relevant seed documents from prior knowledge.

Data	SME A	SME B
Words in Product Description	127	121
Positive Seed Publications	14	16
Positive Seed Clinical Trials	0	0

For the external validation of the two implemented search methods to retrieve publications relevant to a given medical topic, the established CLEF 2018 eHealth TAR dataset for the “Subtask 1: No Boolean Search” was used [10]. The test dataset comprises 30 systematic review protocols with publications labeled as relevant based on the title and abstract text for each review topic.

2.2. MeSH-Based Search

The vast majority of journal articles in PubMed as well as entries in ClinicalTrials.gov are annotated with several MeSH terms, which have already been used for searching for decades [37]. The MeSH-based method aims to leverage this knowledge for the search.

Figure 1 visualizes the MeSH-based search method, which consists of four distinct steps. As a first step, the Medical Term Indexer (MTI) [38] extracts MeSH terms from the product descriptions and provides relevant documents.

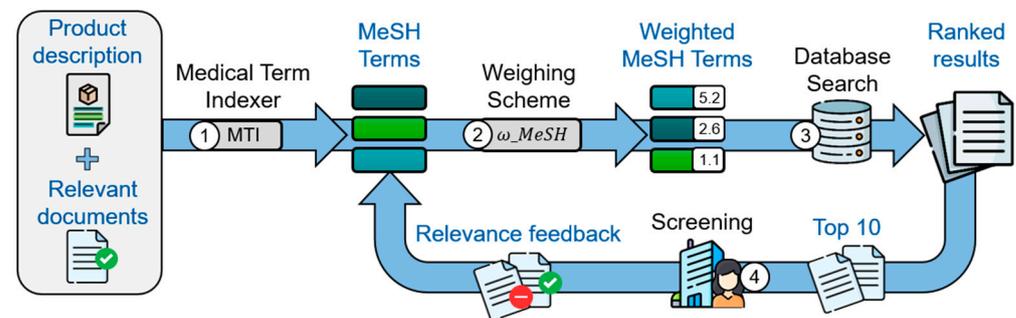


Figure 1. Illustration of the iterative MeSH-based method via four steps: (1) MeSH terms are extracted via the Medical Term Indexer from the product description. (2) Extracted MeSH terms are weighted based on their frequency. (3) Documents with matching terms are retrieved from the document database and ordered by the totaled weights. (4) Relevant documents from search results identified through user feedback are loaded back into the process for iterative refinement.

In the second step, a variation of the Term Frequency—Inverse Document Frequency [39] algorithm is used to compute the weights $\omega_{MeSH}(t, \Delta)$ of all extracted MeSH terms. The weight of a MeSH term t for a set of seed documents Δ consisting of relevant δ_{pos} and irrelevant δ_{neg} documents with $\delta_{neg}, \delta_{pos} \subseteq \Delta$ and $\delta_{neg} \cap \delta_{pos} = \emptyset$ is defined as:

$$\omega_{MeSH}(t, \Delta) = (f_{t, \delta_{pos}} - f_{t, \delta_{neg}}) \cdot idf(t), \tag{1}$$

where $f_{t, \delta}$ denotes the count of t in δ with $\delta \subseteq \Delta$, i.e., the number of times t occurs in δ . The Inverse Document Frequency $idf(t)$ of the MeSH term t is defined by:

$$idf(t) = \log\left(\frac{1 + N}{1 + n_t}\right), \tag{2}$$

with n_t indicating the number of documents annotated with the MeSH term t and N representing all documents in the used database with $n_t \subseteq N$.

The resulting MeSH weighting reflects the importance of each MeSH term for a set of seed documents in relation to all documents of a database. Thus, frequently used MeSH terms are weighted lower than rarely used MeSH terms to focus the search more on the core topic. The computed MeSH weights for PubMed are used as well for the search in ClinicalTrial.gov. The provided product description is treated as a positive seed document for the calculation.

In the third step, the search is conducted with the weighted MeSH terms in a logical disjunction, i.e., via Boolean OR operators. Thus, the search retrieves all documents containing at least one of the extracted MeSH terms. Then, the sum of the weights of the matched MeSH terms for each document establishes the ranking.

In the fourth step, a feedback loop takes place by incorporating any identified relevant document back into the process. The annotated MeSH terms are added, and the MeSH weighting is recomputed according to step 2 with the now expanded set of relevant documents for the next search in step 3.

2.3. BERT-Based Search

In contrast to the term-based search with MeSH terms, the BERT-based search represents a semantic method to document retrieval. While specialized BERT models exist for various tasks, such as named entity recognition, question answering or translation, our BERT-based method leverages a BERT model trained to compute sentence similarity [40]. This allows us to find the most similar documents to a given product description in a similar text retrieval fashion [30]. As a prerequisite, the text of all documents of the used databases is embedded via a modified PubMedBERT model [41]. The original PubMedBERT model is a BERT model trained on 14 million PubMed abstracts and the modified model used in this paper was later fine-tuned to compute sentence similarities with the Microsoft Machine Reading Comprehension (MS MARCO) dataset [41,42]. To leverage PubMedBERT in our method as visualized in Figure 2, the embeddings of all publications, clinical trials, and product descriptions provided by the SMEs have to be computed as the first two steps.

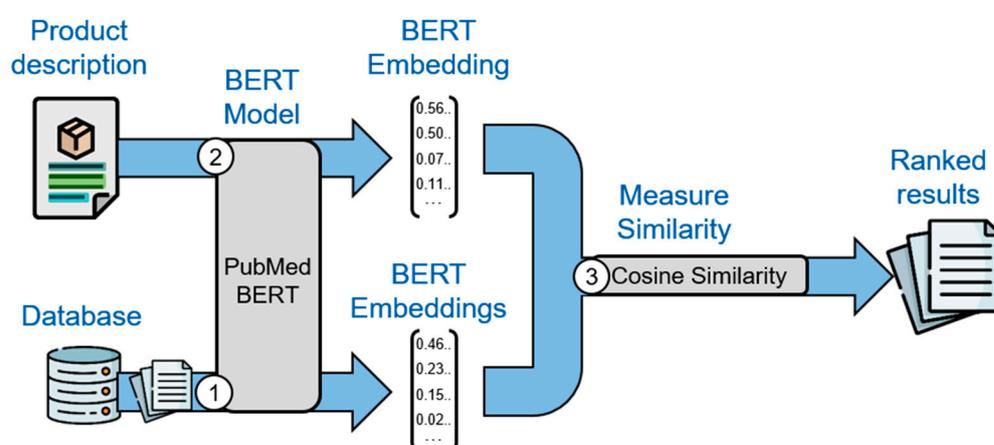


Figure 2. Illustration of the BERT-based search via three steps: (1) the BERT model computes the embeddings of all publications in PubMed, clinical trials in ClinicalTrials.gov, and (2) of the product description; (3) the cosine similarity between the embedding of the product description and each publication/clinical trial is computed to obtain the ranked search results via descending similarity scores.

In the third step, the cosine similarity between the embedding of the product description and the embedding of every publication from PubMed or clinical trial from ClinicalTrials.gov, respectively, is calculated. The documents are ordered by descending cosine similarity.

2.4. Evaluation Metrics

To evaluate both search methods in the SME use case, the precision as provided via the number of relevant documents retrieved by either method is compared. Since SME resources are limited, only the top 10 documents of either search method are screened by the SMEs after each search. To evaluate the goodness of the result ranking, the mean average precision (MAP) is used [43]. To calculate the MAP of different queries q with n documents, we used the established MAP formula based on the precision at document position k in the results with $k \in n$:

$$p@k(q, k) = \frac{\text{number of relevant documents in } q \text{ at top } k \text{ positions}}{k}, \quad (3)$$

and the subsequently defined average precision (AP) with binary relevance of a document at position k with $r_k \in \{0, 1\}$:

$$AP(q) = \frac{\sum_{k=1}^n p@k(q, k) \cdot r_k}{\text{number of relevant documents}}. \quad (4)$$

Based on these formulas, calculating the mean value of AP over all queries results in MAP. In the scope of this work, we evaluated each method with two queries, namely a query for the SME A and SME B use case.

To gain further insights into the content of the retrieved documents of either method, terms rated as relevant and irrelevant by SMEs were compared with the MeSH terms of all identified relevant publications. In addition, the number of the same documents found by both search methods and the overlap of MeSH terms in the result sets of both methods are also contrasted by calculating the Jaccard index. Doing so should help to better understand whether both search methods retrieve the same or a different type of content as expressed via MeSH terms.

While the recall cannot be determined for the use cases, since the overall number of relevant documents is not known, it will be evaluated with the established CLEF 2018 eHealth TAR dataset for further insights. Both of our search methods will be compared with the recall of the methods of other participating teams in the challenge.

Since the dataset contains only relevant documents, documents of unknown relevance will be retrieved in the searches. Instead of marking those possibly relevant documents as irrelevant and incorporating them in the feedback loop for the MeSH-based method, only documents known to be relevant, e.g., positive seeds, will be used.

2.5. Workflow

For resource-efficient collaboration with SMEs during the development and evaluation of the search methods, we conceptualized a workflow revolving around the SMEs with multiple short 30-min relevance screening sessions online as visualized in Figure 3. In accordance with the SME representatives, we agreed on 30-min sessions for the relevance screening and feedback discussion regarding the process, to keep the required SME resources low. During the online sessions in the form of video calls, the SME representatives were able to provide feedback and ask for clarifications if needed.

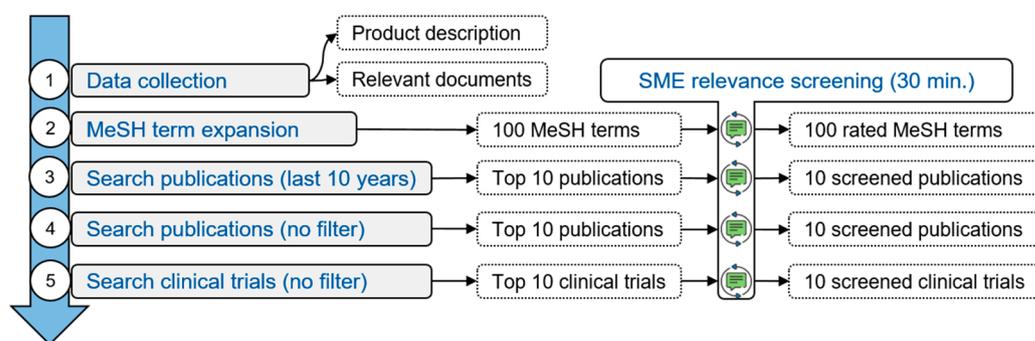


Figure 3. Illustration of the workflow to guide the automated searches tailored to the SME use cases. The process comprises five distinct steps with respective output: (1) data about the medical device in question are collected via a short questionnaire; (2) MeSH terms extracted from the product description are expanded; (3) a first search for publications with a filter for the last 10 years is conducted; (4) a second search without filters is conducted; and (5) a third search for clinical trials without filters is conducted. Extracted MeSH terms and the top 10 results of each search are relevance screened by the SMEs.

As the first data collection step, SMEs provide a product description as well as relevant documents specific to their medical device by answering the questionnaire.

For the second step, MeSH terms from the product description are extracted using the Medical Term Indexer (MTI). Since not every product description might result in multiple terms due to varying length or content, an additional term expansion is performed. For the term expansion all direct MeSH terms as subnodes of the extracted MeSH terms were added to the MeSH term list. Afterward, in a first MeSH search, all publications with at least one matching MeSH term were retrieved from PubMed and ordered by descending

matches. SMEs then rated the relevance of the 100 most frequent MeSH terms out of all publications from the search results via a four-point weighing scheme analogous to a Likert scale (highly relevant: 2, relevant: 1, irrelevant: 0, exclude: –1). After initial exchanges with the SMEs, we decided to label an answer option “exclude” instead of “highly irrelevant”, to allow SMEs to specify terms to exclude during the search.

In the third to fifth steps, three separate searches are performed by each search method. The BERT-based search uses the product description without any preprocessing for each of the three performed searches. The top 10 results of each search method are presented to the SMEs for binary relevance screening. Decision fatigue during the screening is prevented by alternating results from the top 10 lists of the search methods. To avoid redundancies in the search results, previously identified relevant documents, such as seed documents or already screened documents, are removed and replaced with the next most relevant documents in case those documents are retrieved through a search.

For the search in the third step, publications of the last ten years from 1 January 2012 to 31 August 2022 based on the SME requirements are searched for. The MeSH-based search uses the hundred SME-rated MeSH terms.

For the search in the fourth step, the search is expanded to cover all documents in PubMed without a specified time frame in accordance with SMEs. Documents from the previous screening serve as seed documents for the MeSH-based search. Based on their relevance rating, documents are used as positive and negative seeds, e.g., an irrelevant document is used as a negative seed.

For the search in the concluding fifth step, the screened publications are incorporated in the MeSH-based search and a last search for clinical trials is carried out.

3. Results

In the following two subsections, results regarding the iterative searches for the two SMEs will be presented, followed by further insights with the CLEF 2018 eHealth TAR dataset.

3.1. SME Use Cases

Based on the MeSH terms extracted from the product description, an initial search in the local PubMed database took place, and the SMEs rated the hundred most frequently occurring MeSH terms of the retrieved publications. The numbers of rated MeSH terms of the SMEs are shown in Table 2. It can be observed for SME A that the majority of presented MeSH terms in the questionnaire should be excluded from the search, whereas for SME B the majority of terms were rated as highly relevant, subsequently leading to a narrower search for SME A and a wider search for SME B. In total SME A rated 39 and SME B rated 76 of the respective hundred MeSH terms as relevant.

Table 2. Overview of the MeSH term relevance screening results of both SMEs for the hundred most frequently occurring MeSH terms from the MeSH term expansion step.

Relevance	SME A	SME B
Highly Relevant	23	63
Relevant	16	13
Irrelevant	0	20
Exclude	61	4
Total	100	100

When comparing the MeSH terms derived from the product descriptions and rated by the SMEs as relevant (Highly Relevant, Relevant) and the terms rated as irrelevant (Irrelevant, Exclude) with the annotated MeSH terms of the provided seed publications, a disconnect becomes evident as visualized in Figure 4: Only a small part of the MeSH terms rated as relevant by either SMEs are also present in the MeSH terms from the initial

seeds. Moreover, in the case of SME A (a), six MeSH terms rated as irrelevant are present in the provided initial seeds. For SME B (b) no overlap between MeSH terms of the initial seeds and terms rated as irrelevant can be observed. For both SMEs, less than a third of the MeSH terms present in the initial seeds overlap with the MeSH terms rated as relevant by the SMEs.

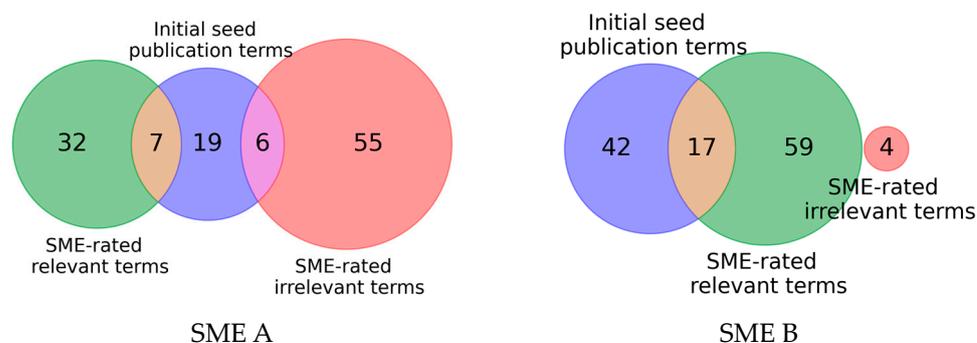


Figure 4. Venn diagram indicating the overlap of MeSH terms from the initial seed publications and the hundred MeSH terms rated by the SMEs in the MeSH term expansion step.

Based on the hundred MeSH terms rated by the SMEs, an initial MeSH-based search was conducted alongside a BERT-based search based on the product description without any preprocessing. After discussing the results with the SMEs, a second search was conducted and expanded on a larger time frame as specified by the SME requirements. For a third search, clinical trials without a specified timeframe were searched for as well. After each search, the top 10 results were screened by SMEs for relevance. The average precision for each search for publications and clinical trials by the MeSH- and BERT-based methods for both SMEs are visualized in Figure 5.

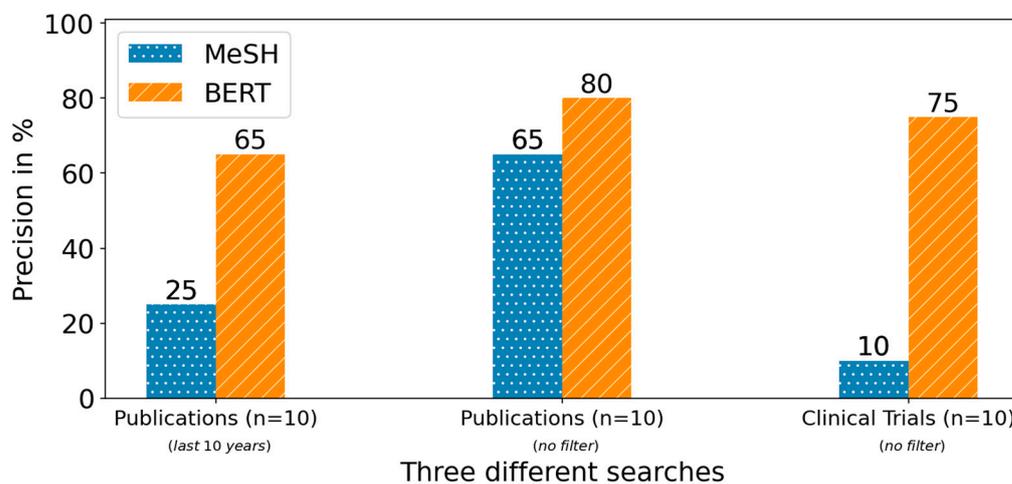


Figure 5. Average precision at top 10 search results for the two MeSH- and BERT-based methods, applied to publications (PubMed) and clinical trials (ClinicalTrials.gov) in the different searches for the two SME use cases.

For both search methods, an increased precision from the first to the second search for publications can be observed. For the MeSH-based search, the precision increased considerably from 25% to 65% in the second search. For the BERT-based search, the precision was already relatively high in the first stage at 65% and increased to a smaller extent to 80%. During the search for relevant clinical trials, the MeSH-based method achieved only a precision of 10% while the BERT-based achieved a precision of 75%.

After the two screening stages, a last search, which incorporated all relevant publications, was conducted to produce the final result set for both SMEs. The goodness of the ranking was evaluated via the MAP score as visualized in Figure 6.

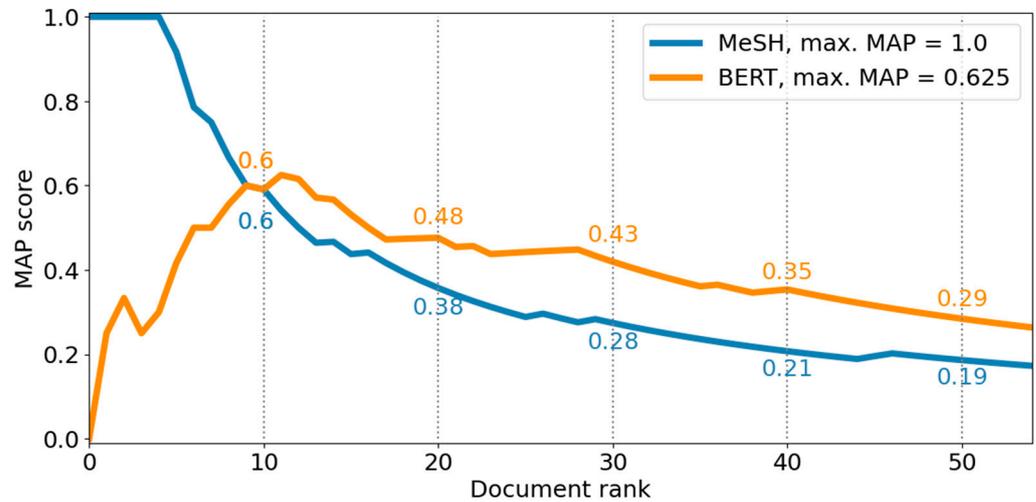


Figure 6. The goodness of ranking is illustrated as a line chart depicting the MAP score at a given document rank of the MeSH- and BERT-based method for the top 50 publications averaged for both SME use cases.

It is noticeable that the ranking of the BERT-based method is better as indicated by a higher MAP score of 0.29 overall, i.e., at rank 50, than the ranking of the MeSH-based method with a score of 0.19. The MAP score of the MeSH-based method starts at 1 and then continues to decline. In contrast, the MAP score of the BERT-based method increases to its maximum at the 11th rank with a MAP score of 0.625 and then gradually declines as well while continuously yielding a better MAP score than the MeSH-based method.

The aggregated overall results for the test cases with the SMEs are visualized in Figure 7. Due to a lack of MeSH terms in clinical trials, a second search was not conducted; hence, only the top 10 clinical trials for both SMEs of the first search stage are compared.

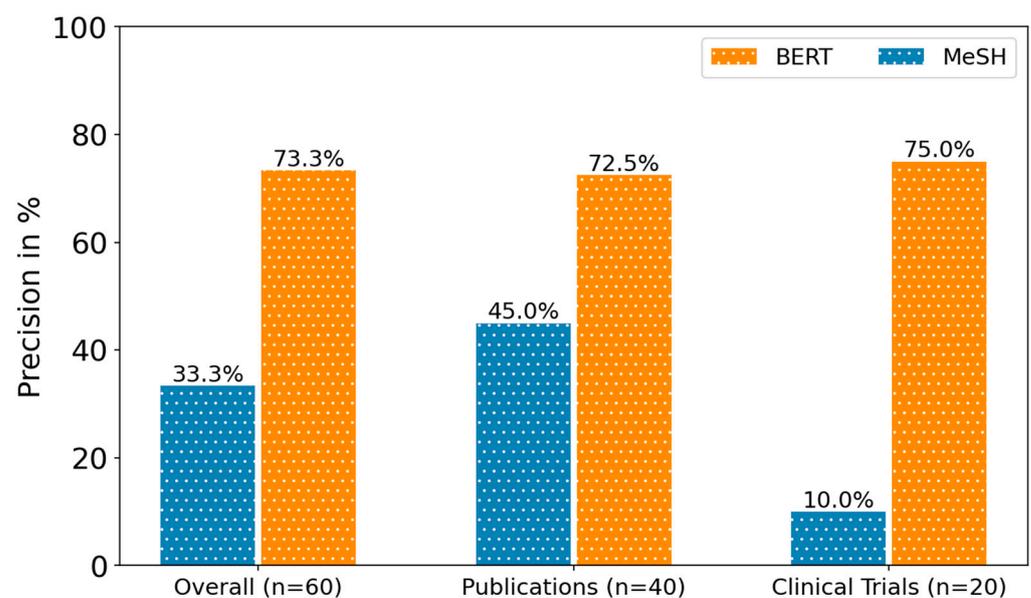


Figure 7. Averaged precision of both search methods overall, for publications and for clinical trials of both SME use cases. For each category, n defines the number of total documents per bar.

The BERT-based method performed better overall by retrieving proportionally more relevant documents. While the overall precision, e.g., how many retrieved documents were relevant, for the BERT-based method with 44 relevant documents was 73.3%, the precision for the MeSH-based method with 20 relevant documents was only 33.3%. The BERT-based method performed similarly well on publications with a precision of 72.5% and 29 retrieved relevant publications as when searching for clinical trials with a precision of 75% and 15 retrieved relevant clinical trials. The MeSH-based method yielded a lower performance and retrieved 18 relevant publications with a precision of 45% and retrieved only 2 relevant clinical trials with a precision of 10%. Besides the notable difference in overall performance between both methods, a contrast between precision when retrieving publications compared to clinical trials is evident for the MeSH-based method. To better understand both search methods regarding the retrieved documents, result sets were compared regarding the number of the same retrieved documents and the same MeSH terms of retrieved documents.

To analyze the similarity of the result sets of both methods, the Jaccard index of retrieved publications with respective PubMed identifiers (PMID) was investigated. Figure 8 visualizes the Jaccard index of documents in the top 500 results of both search methods for the first and second SME use cases. For the SME A use case (a), the maximum Jaccard index is 0.013, indicating that the result sets are almost disjoint. In the SME B use case (b) the maximum Jaccard index is 0.11, indicating that the result sets of both search methods are only slightly similar. Moreover, a slight increase in similarity for the SME A use case and a stronger increase in the similarity for the SME B use case can be observed, indicating that the result sets might converge later on.

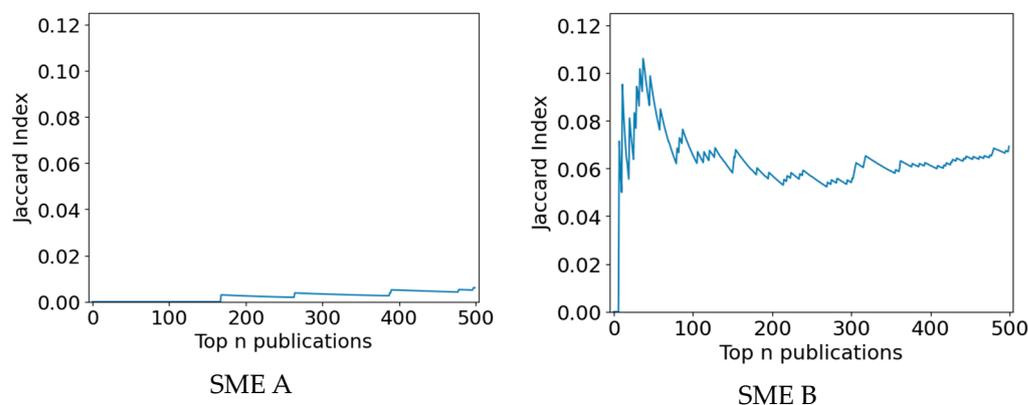


Figure 8. Diversity of result sets from both search methods on document level (PMID) measured by the Jaccard index, i.e., retrieved publications, by both search methods for both SME use cases in the first 500 results.

The content of the retrieved publications as expressed in the form of the respectively assigned MeSH terms was compared with the result set of both methods and the positive rated MeSH terms by the SME. The number of unique MeSH terms occurring in the top 500 retrieved publications and the overlap between both search methods with the SME-rated relevant terms is visualized in Figure 9.

The MeSH-based method features more unique MeSH terms compared to the BERT-based method and a strong overlap between the MeSH terms of the result sets with respective Jaccard indices of 38.7% for SME A and 30.6% for SME B exists. Moreover, it is noticeable that almost all SME-rated relevant terms, namely 87.2% for SME A and 88.2% for SME B occur in either result set. Only a few selected SME-rated relevant terms are present in one result set and not the other. For the first SME use case, all terms rated as relevant were present in the result sets and for the second SME use case, only three terms were not present in any retrieved publication.

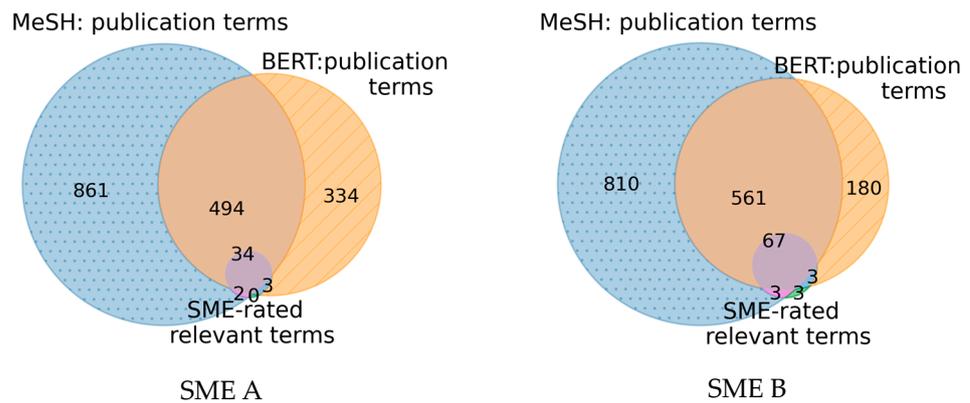


Figure 9. The content similarity is illustrated as a Venn diagram indicating the overlap of the unique MeSH terms occurring in the top 500 retrieved publications from the MeSH-based and BERT-based method with the SME-rated relevant terms for both SME use cases.

3.2. CLEF 2018 eHealth TAR

The evaluation of the BERT- and MeSH-based methods with the established CLEF 2018 eHealth TAR dataset was performed on all 30 test reviews. Only the evaluation on an abstract relevance level took place since screening and searching for relevant full-text articles was out of scope. The results of the BERT and MeSH-based method runs are visualized in Figure 10.

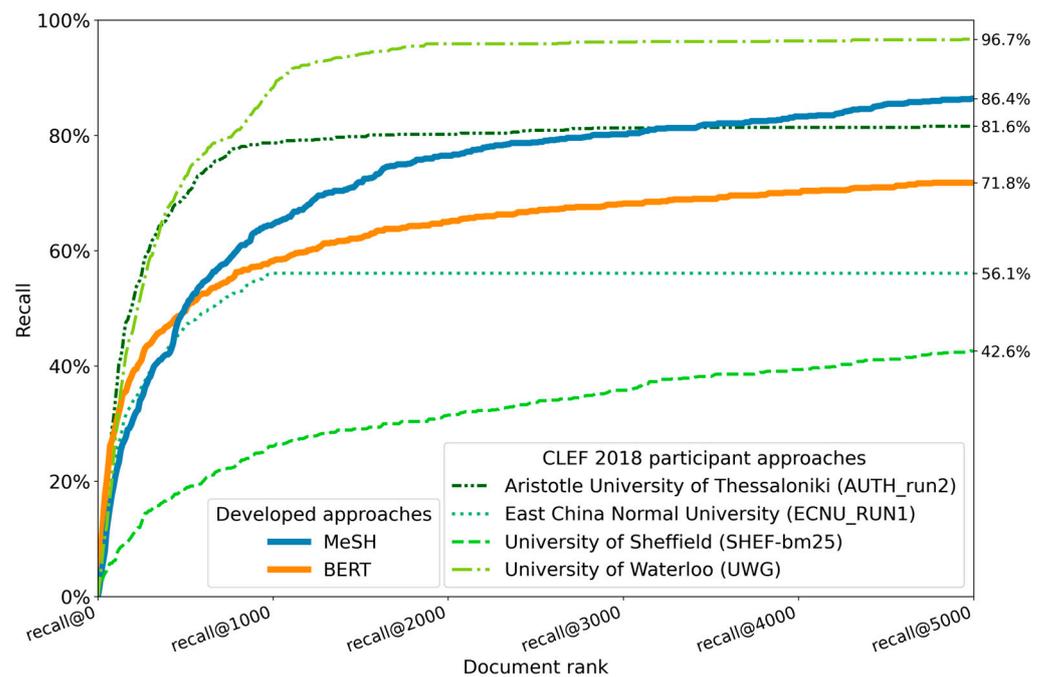


Figure 10. Recall performance comparison of the developed BERT- and MeSH-based methods on the CLEF 2018 eHealth TAR subtask 1: “No Boolean Search” with existing methods [44]. The line graph depicts the averaged recall at different ranks across all 30 test topics. Only the existing CLEF runs with the highest overall recall per participating team are displayed.

The best-performing *AutoTar* method by the University of Waterloo achieves the overall highest recall. Following as the second-best method when it comes to overall recall, our MeSH-based method achieved a recall of 86.4% and outperformed our BERT-based method with a recall of 71.8%, despite the initially better performance of the BERT-based method at lower document ranks.

4. Discussion

In the scope of this study, we conceptualized and tested two different search methods to assist SMEs with the retrieval and screening process of relevant clinical data for medical products for subsequent decision support and as a basis for future systematic reviews. The developed semantic BERT-based search method and term-oriented MeSH-based search method were evaluated in a use case setting with two SMEs on the databases PubMed and ClinicalTrials.gov as well as with the established CLEF 2018 eHealth TAR dataset.

The results of our SME use case evaluation show that the goal to conceptualize resource-efficient and effective retrieval methods to assist SMEs was achieved. Both search methods retrieved numerous relevant clinical trials and publications within the top 10 search results. In total, 47 out of 80 retrieved publications and 17 out of 40 retrieved clinical trials were relevant for both SMEs and identified by screening a relatively low number of documents. To refine the search, short 30-min feedback sessions in accordance with the SMEs were used to evaluate the results while respecting the limited SME resources. The resulting overall precision of 33.3% for the MeSH-based method and 73.3% for the BERT-based method is much higher compared to the precision of previous studies with systematic reviews in the scope of the MDR certification. In one recent systematic review, 60 out of 413 retrieved publications were deemed relevant based on the title and abstract, resulting in a precision of 14.5% [9]. In another systematic review for a medical product for the MDR certification, only 21 publications out of 242 retrieved publications were deemed relevant after screening, resulting in an even lower precision of 8.7% [25]. One reason for the difference in performance could be the notably smaller sample size of 40 publications and 20 clinical trials as well as the differing scope of initial scoping searches in our SME use cases instead of systematic reviews of the compared studies. While respecting these structural limitations, the higher overall precision of our search methods indicates the potential to retrieve a comparatively large number of relevant documents while screening much fewer documents compared to manual literature searches. This indication aligns with recent studies stating an approximate reduction of 50% in papers to screen via automated literature screening methods [45]. However, depending on the type of underlying study or task of the literature screening, the reduced workload can vary between 7% to 71% while maintaining a high level of recall [46].

A high recall was also observed during our second evaluation of both search methods with the CLEF 2018 eHealth TAR dataset, in which both methods achieved a high recall in a more general, nonspecific medical device, setting. While the BERT-based method reached an overall recall of 71.8%, the MeSH-based method outperformed most other methods with a recall of 86.4%.

In particular, the new weighting of MeSH terms based on the feedback loops with an increasing number of positive seed documents led to the high performance of the MeSH-based search. However, despite the high performance, attention should be paid to the generalizability of the search. With the currently implemented weighting scheme, terms occurring frequently in relevant documents have a higher weight and focus the search more on a specific field. Underrepresented fields of possibly equal relevance may be excluded, potentially resulting in data bias in the search, also referred to as hasty generalization [47,48].

Moreover, while the inclusion of negative seed documents from the SME feedback loops led to reasonable results in the MeSH-based search, future use cases might require different sampling strategies to balance relevant and irrelevant seeds and avoid potential data bias [49,50].

Another potential bias could occur when searching for information relevant to highly innovative medical products. While the MeSH-based method could extract existing MeSH terms for fundamental information, such as the intended purpose of the device, target group, and clinical benefits even for highly innovative products, MeSH terms relevant to aspects such as the innovative technologies may be lacking. It is possible that there are not any relevant existing MeSH terms since MeSH terms for emerging fields might have yet to

be added to the terminology. The BERT-based method, in contrast, might not feature such a bias regarding innovative products since semantics in the product description and thus similar technologies mentioned in literature can still be retrieved despite a lack of relevant MeSH terms. Future use cases, however, could investigate the potential bias when working with highly innovative medical devices.

Hence, the novelty of medical products is one possible reason for the differences regarding as relevant rated MeSH terms for medical products by SMEs as observed in Table 2. A second possible reason for the differences could be differing product types. Only a few MeSH terms might describe highly specialized medical products, whereas multiple MeSH terms might describe medical products with multiple technologies or intended purposes. A third possible reason could be the difference in rating behavior as observed by the representatives for SME A and SME B.

In contrast to the MeSH-based method, the BERT-based search is not affected by these potential biases, since only free text was provided and no MeSH terms are used as input. However, analogous to the MeSH-based search, the BERT-based search could also benefit from incorporating feedback similar to the MeSH-based method to tailor the search more closely to the respective medical device. This could be achieved with the representation of a cluster of seed documents with the respective centroid for the search [51]. Moreover, classification models could leverage BERT embeddings as input and be trained via active-learning feedback loops for increased performance [52,53].

While both search methods performed overall remarkably well as compared to previous manual as well as automated search methods, the BERT-based method noticeably outperformed the MeSH-based method in the SME use case setting, particularly when searching for clinical trials.

One reason for the lower MeSH-based method performance in the first search for publications was that MeSH term relevancies provided by SMEs did not match well with MeSH terms of provided relevant seed documents. This discrepancy may be due to the difference in understanding of MeSH terms by the SMEs and the actual meaning of MeSH terms, resulting in the misidentification of relevant MeSH terms, affecting the success of the search [54]. Feedback sessions also indicated that SMEs may not fully know how to provide a fitting description or labels for their products, putting further emphasis on the importance of additional seed documents for a more precise search. While some MeSH terms were rated as initially irrelevant by the SMEs, publications containing those MeSH terms were rated as relevant. As elaborated then by one SME representative, screening relevant literature helped them to understand better how the medical product could be described and labeled regarding MeSH terms.

A second reason for the lower MeSH-based method performance could be the limited screening of only the first top 10 results of the searches since the feedback-driven MeSH-based approach outperformed the BERT-based search on average after screening over 480 documents as indicated in Figure 10.

A third possible reason for the lower performance when searching for clinical trials is the limited annotation of a clinical trial with at most only a few MeSH terms. While the MTI automatically annotates publications in PubMed with suitable MeSH terms, MeSH terms for clinical trials in ClinicalTrials.gov are added manually or with an algorithm restricted to non-frequently occurring terms describing disease conditions or interventions [55]. Consequently, other applicable MeSH terms are excluded, leading to fewer usable MeSH terms for distinguishing and searching clinical trials.

To better understand how both methods compare in regard to the retrieved results given the initial medical device information, and also on a content level, the overlap of the result sets and MeSH terms of retrieved publications was analyzed.

The comparison of retrieved publications by both methods revealed a rather weak overlap in the first 500 results, indicating that both methods retrieve different publications while maintaining fairly high precision. This shows that both methods retrieve relevant publications, but possibly with different focus points. The MeSH-based method retrieves

publications with matching MeSH terms; hence, publications of the same domain are retrieved with possibly unrelated titles and abstract text. The BERT-based method contrasts this, by retrieving publications with highly similar text, but possibly focusing on a different domain since MeSH terms are not considered.

Structural differences in data quality also contribute to search differences: MeSH terms are missing in approximately 14% of publications and 33% of publications are without abstract text in Medline [56]. Consequently, the MeSH-based method cannot retrieve every publication and the BERT-based method can only compute the similarity based on the title of specific publications, possibly containing only a fraction of the important aspects of the publications as opposed to summarizing abstracts [57].

Nevertheless, despite a weak overlap of the result sets at the document level, as measured by the Jaccard index, the overlap based on MeSH terms indicates that the largest overlap with the MeSH terms rated as relevant by the SMEs lies in the overlap region of both search methods.

On one hand, analyzing MeSH terms of retrieved publications allows us to compare different search methods on a content level and opens up possibilities to make the rationale of any search method, such as the black-box BERT-based search, more understandable. On the other hand, this overlap indicates the potential for improvement could exist in a synthesis of both search methods. Such synthesis could facilitate both methods to complement each other for an optimized retrieval of relevant documents for each specific use case using feedback loops while establishing transparency via MeSH terms.

4.1. Limitations

Despite the insights gained from our analysis, it is important to acknowledge that the sample size of only two SMEs may limit the generalizability of our findings. Thus, a second evaluation with the established CLEF 2018 eHealth TAR dataset was conducted but supposes an undefined medical device setting. A larger sample size with more companies may provide more robust and representative results. However, due to resource constraints, we were unable to collect data with a larger number of SMEs.

4.2. Outlook

In future work, additional data sources such as the Clinical Trials Information System (CTIS) or the European Database on Medical Devices (EUDAMED) could serve as useful resources for clinical data retrieval [26,58]. Incorporating adverse effect announcements could facilitate post-market surveillance and monitoring use cases by verifying the benefits of medical devices in the face of newly published information [7].

Subsequently integrating retrieved heterogeneous data into interactive knowledge graphs yields the potential for cluster analysis, visualizations, and improved search performances [59]. For instance, by linking publications to clinical trials and other publications via citations, additional relevant data could be retrieved based on already identified relevant documents without the need for further searches [60].

However, despite the potential benefits of integrating such automated search methods into systematic review processes [61], the applicability in the scope of the MDR clinical evaluation remains uncertain [29]. The usability of TAR methods must be evaluated for the search for clinical data in the scope of the MDR clinical evaluation.

A promising opportunity could be an ensemble of both developed search methods complementing each other to address respective retrieval challenges and solve transparency concerns while increasing retrieval performances.

Overall while first results indicate high precision and recall, further evaluations on a larger scale should be conducted, ideally with medical device manufacturers who already certified devices under the MDR and could therefore provide data serving as a future gold standard.

5. Conclusions

In conclusion, we developed an iterative resource-efficient workflow that has the power to effectively assist medical device manufacturers in retrieving clinical data relevant to their devices. By incorporating SME-specific medical device information and requirements, the search was tailored to the medical device and further refined via feedback loops. The high precision of both methods in the SME use cases and high recall in the CLEF eHealth 2018 dataset evaluation demonstrate the potential of automated search methods to retrieve relevant documents while screening fewer documents as compared to previous more resource-intensive literature reviews in the scope of the MDR. The observed generalizability of both methods as well as the possible synthesis as an ensemble opens the path to further evaluations with more medical device manufacturer use cases and different application fields. Overall, our results indicate the potential to reduce the workload of SMEs by providing an orientation with first relevant literature before initiating the clinical evaluation in the future.

Author Contributions: Conceptualization, F.-S.K.-B.T., M.B. and R.F.; methodology, F.-S.K.-B.T., M.B. and R.F.; software, F.-S.K.-B.T.; validation, F.-S.K.-B.T. and R.F.; formal analysis, F.-S.K.-B.T. and R.F.; investigation, F.-S.K.-B.T. and R.F.; resources, R.F.; data curation, F.-S.K.-B.T. and R.F.; writing—original draft preparation, F.-S.K.-B.T.; writing—review and editing, F.-S.K.-B.T., M.B., R.F. and T.S.-R.; visualization, F.-S.K.-B.T. and R.F.; supervision, R.F. and T.S.-R.; project administration, R.F. and T.S.-R.; funding acquisition, M.B., R.F. and T.S.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Economic Affairs, Industry, Climate Action and Energy of the State of North Rhine-Westphalia: MDR-1-1-E.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions from the medical device manufacturers.

Acknowledgments: We are thankful to the state of North Rhine-Westphalia for providing financial support for this research in the scope of the joint project MDR-Support@NRW.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bianchini, E.; Francesconi, M.; Testa, M.; Tanase, M.; Gemignani, V. Unique device identification and traceability for medical software: A major challenge for manufacturers in an ever-evolving marketplace. *J. Biomed. Inform.* **2019**, *93*, 103150. [CrossRef] [PubMed]
2. Minssen, T.; Gerke, S.; Aboy, M.; Price, N.; Cohen, G. Regulatory responses to medical machine learning. *J. Law Biosci.* **2020**, *7*, Isaa002. [CrossRef] [PubMed]
3. Niemiec, E. Will the EU medical device regulation help to improve the safety and performance of medical ai devices? *Digit. Health* **2022**, *8*, 20552076221089079. [CrossRef]
4. Malvey, J.; Ginsberg, R.; Sampietro-Colom, L.; Ficapal, J.; Combalia, M.; Svedenhag, P. New regulation of medical devices in the EU: Impact in dermatology. *J. Eur. Acad. Dermatol. Venereol.* **2022**, *36*, 360–364. [CrossRef] [PubMed]
5. MedTech Europe. MedTech Europe Survey Report Analysing the Availability of Medical Devices in 2022 in Connection to the Medical Device Regulation (MDR) Implementation. Available online: <https://www.medtecheurope.org/wp-content/uploads/2022/07/medtech-europe-survey-report-analysing-the-availability-of-medical-devices-in-2022-in-connection-to-the-medical-device-regulation-mdr-implementation.pdf> (accessed on 14 June 2023).
6. Makeenko, M.; Chevallier, T. Post-market Clinical Follow-up (PMCF) GAP Analysis for Legacy Devices Class III between the Medical Device Directive (MDD 93/42/EEC) and the Medical Device Regulation (MDR 2017/745). In *BIOSTEC 2022*; SciTePress: Setúbal, Portugal, 2022; pp. 273–280. [CrossRef]
7. Beckers, R.; Kwade, Z.; Zanca, F. The EU medical device regulation: Implications for artificial intelligence-based medical device software in medical physics. *Phys. Med. PM* **2021**, *83*, 1–8. [CrossRef]
8. Martelli, N.; Eskenazy, D.; Déan, C.; Pineau, J.; Prognon, P.; Chatellier, G.; Sapoval, M.; Pellerin, O. New european regulation for medical devices: What is changing? *Cardiovasc. Interv. Radiol.* **2019**, *42*, 1272–1278. [CrossRef]

9. Vincent, P. Intra-articular hyaluronic acid in knee osteoarthritis: Clinical data for a product family (arthrum). with comparative meta-analyses. *Curr. Ther. Res. Clin. Exp.* **2021**, *95*, 100637. [CrossRef]
10. Kanoulas, E.; Li, D.; Azzopardi, L.; Spijker, R. CLEF 2018 Technologically Assisted Reviews in Empirical Medicine Overview. *CEUR Workshop Proc.* **2018**, *2125*, 34.
11. Cormack, G.; Grossman, M. Autonomy and Reliability of Continuous Active Learning for Technology-Assisted Review. *arXiv* **2015**, arXiv:1504.06868.
12. Scells, H.; Zuccon, G.; Koopman, B. A comparison of automatic Boolean query formulation for systematic reviews. *Inf. Retr. J.* **2021**, *24*, 3–28. [CrossRef]
13. Wang, S.; Scells, H.; Koopman, B.; Zuccon, G. Automated MeSH term suggestion for effective query formulation in systematic reviews literature search. *Intell. Syst. Appl.* **2022**, *16*, 200141. [CrossRef]
14. Wang, S.; Li, H.; Scells, H.; Locke, D.; Zuccon, G. MeSH Term Suggestion for Systematic Review Literature Search. In *Australasian Document Computing Symposium*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1–8. [CrossRef]
15. Guo, J.; Cai, Y.; Fan, Y.; Sun, F.; Zhang, R.; Cheng, X. Semantic Models for the First-Stage Retrieval: A Comprehensive Review. *ACM Trans. Inf. Syst.* **2022**, *40*, 1–42. [CrossRef]
16. Walsh, H.S.; Andrade, S.R. Semantic Search With Sentence-BERT for Design Information Retrieval. In *Volume 2: 42nd Computers and Information in Engineering Conference (CIE)*; American Society of Mechanical Engineers: New York, NY, USA, 2022. [CrossRef]
17. Ahsas, S.; Nfaoui, E.H. Academic Aggregated Search Approach Based on BERT Language Model. In *Proceedings of the 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco, 3–4 March 2022; pp. 1–9. [CrossRef]
18. Yang, E.; MacAvaney, S.; Lewis, D.D.; Frieder, O. Goldilocks: Just-Right Tuning of BERT for Technology-Assisted Review. In *Advances in Information Retrieval*; Hagen, M., Verberne, S., Macdonald, C., Seifert, C., Balog, K., Nørvåg, K., Setty, V., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 502–517. [CrossRef]
19. Rogers, A.; Kovaleva, O.; Rumshisky, A. A Primer in BERTology: What We Know About How BERT Works. *Trans. Assoc. Comput. Linguist.* **2020**, *8*, 842–866. [CrossRef]
20. Zhan, J.; Mao, J.; Liu, Y.; Zhang, M.; Ma, S. An Analysis of BERT in Document Ranking. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual, 25–30 July 2020; ACM: New York, NY, USA, 2020; pp. 1941–1944. [CrossRef]
21. Cooper, C.; Dawson, S.; Lefebvre, C. Searching for medical devices—Practical guidance. *Res. Synth. Methods* **2022**, *13*, 144–154. [CrossRef]
22. Migliore, A. Technology assessment of innovative medical devices in europe. *Expert Rev. Med. Devices* **2016**, *13*, 217–219. [CrossRef]
23. Hunter, K.E.; Webster, A.C.; Page, M.J.; Willson, M.; McDonald, S.; Berber, S.; Skeers, P.; Tan-Koay, A.G.; Parkhill, A.; Seidler, A.L. Searching clinical trials registers: Guide for systematic reviewers. *BMJ* **2022**, *377*, e068791. [CrossRef]
24. Westrick, J.C.; Buchholz, S.W. Medical librarians’ knowledge and practices in locating clinical trials for systematic reviews. *J. Med. Libr. Assoc.* **2021**, *109*, 295–300. [CrossRef]
25. Kiefer, A.-K.; Espig, K.; Moeller, K. A systematic review of recorded technical settings used in bipolar and monopolar transurethral resection of bladder tumors. *Curr. Dir. Biomed. Eng.* **2022**, *8*, 474–477. [CrossRef]
26. Shatrov, K.; Blankart, C.R. After the four-year transition period: Is the European Union’s medical device regulation of 2017 likely to achieve its main goals? *Health Policy* **2022**, *126*, 1233–1240. [CrossRef] [PubMed]
27. *Survey on the EU Medical Devices Regulation by DIHK, SPECTARIS and MedicalMountains: First Assessment of the German Manufacturers of Medical Devices after the EU Medical Device Regulation (MDR) Came Into Force*; Deutscher Industrie-und Handelskammertag e.V.: Berlin, Germany; MedicalMountains GmbH: Tuttlingen, Germany; SPECTARIS. Deutscher Industrieverband für Optik, Photonik, Analysen-und Medizintechnik e.V.: Berlin, Germany, 2022. Available online: https://www.spectaris.de/fileadmin/Content/Medizintechnik/Startseite/Survey_on_the_EU_Medical_Devices_Regulation.pdf (accessed on 14 June 2023).
28. Agyei, E.E.Y.F.; Pohjolainen, S.; Oinas-Kukkonen, H. Impact of Medical Device Regulation on Developing Health Behavior Change Support Systems. In *Persuasive Technology*; Baghaei, N., Vassileva, J., Ali, R., Oyibo, K., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 1–15. [CrossRef]
29. MacFarlane, A.; Russell-Rose, T.; Shokraneh, F. Search strategy formulation for systematic reviews: Issues, challenges and opportunities. *Intell. Syst. Appl.* **2022**, *15*, 200091. [CrossRef]
30. Fujishiro, N.; Otaki, Y.; Kawachi, S. Accuracy of the Sentence-BERT Semantic Search System for a Japanese Database of Closed Medical Malpractice Claims. *Appl. Sci.* **2023**, *13*, 4051. [CrossRef]
31. Tian, X.; Wang, J. Retrieval of Scientific Documents Based on HFS and BERT. *IEEE Access* **2021**, *9*, 8708–8717. [CrossRef]
32. Health Technology and Cosmetics. Guidelines on Medical Devices: Clinical Evaluation: A Guide For Manufacturers and Notified Bodies Under Directives 93/42/EEC and 90/385/EEC. Available online: <https://ec.europa.eu/docsroom/documents/17522/attachments/1/translations/en/renditions/native> (accessed on 14 June 2023).
33. European Commission. Guidelines on Medical Devices Meddev 2.7/1. Available online: <https://www.google.com/search?client=firefox-b-d&q=meddev+2.7%2F1> (accessed on 13 March 2023).

34. Hughey, J.; Schoenbachler, J. *PMDB: A Relational Database for PubMed*; Zenodo: Geneva, Switzerland, 2022. [[CrossRef](#)]
35. Schoenbachler, J.L.; Hughey, J.J. Pmparser and pmdb: Resources for large-scale, open studies of the biomedical literature. *PeerJ* **2021**, *9*, e11071. [[CrossRef](#)]
36. Clinical Trials Transformation Initiative. Improving Public Access to Aggregate Content of ClinicalTrials.gov. Available online: <https://aact.ctti-clinicaltrials.org/> (accessed on 27 January 2023).
37. Lowe, H.J.; Barnett, G.O. Understanding and using the medical subject headings (mesh) vocabulary to perform literature searches. *JAMA* **1994**, *271*, 1103–1108. [[CrossRef](#)]
38. Mork, J.G.; Jimeno-Yepes, A.; Aronson, A.R. The NLM Medical Text Indexer System for Indexing Biomedical Literature. In Proceedings of the First Workshop on Bio-Medical Semantic Indexing and Question Answering, a Post-Conference Workshop of Conference and Labs of the Evaluation Forum 2013 (CLEF 2013), Valencia, Spain, 27 September 2013. Available online: https://ceur-ws.org/Vol-1094/bioasq2013_submission_3.pdf (accessed on 14 June 2023).
39. Robertson, S.E.; Jones, K.S. Relevance weighting of search terms. *J. Am. Soc. Inf. Sci.* **1976**, *27*, 129–146. [[CrossRef](#)]
40. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
41. Deka, P.R.I.T.A.M.; Jurek-Loughrey, A.N.N.A.; Deepak, P. Improved Methods to Aid Unsupervised Evidence-Based Fact Checking for Online Health News. *J. Data Intell.* **2022**, *3*, 474–504. [[CrossRef](#)]
42. Gu, Y.; Tinn, R.; Cheng, H.; Lucas, M.; Usuyama, N.; Liu, X.; Naumann, T.; Gao, J.; Poon, H. Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Trans. Comput. Healthc.* **2022**, *3*, 1–23. [[CrossRef](#)]
43. Liu, T.-Y. Learning to Rank for Information Retrieval. *Found. Trends@Inf. Retr.* **2007**, *3*, 225–331. [[CrossRef](#)]
44. Cormack, G.; Grossman, M. Technology-Assisted Review in Empirical Medicine: Waterloo Participation in CLEF eHealth 2018. In Proceedings of the Conference and Labs of the Evaluation Forum (CLEF), Avignon, France, 10–14 September 2018.
45. Burgard, T.; Bittermann, A. Reducing Literature Screening Workload With Machine Learning. *Z. Für Psychol.* **2023**, *231*, 3–15. [[CrossRef](#)]
46. Yamada, T.; Yoneoka, D.; Hiraike, Y.; Hino, K.; Toyoshiba, H.; Shishido, A.; Noma, H.; Shojima, N.; Yamauchi, T. Deep neural network for reducing the screening workload in systematic reviews for clinical guidelines: Algorithm validation study. *J. Med. Internet Res.* **2020**, *22*, e22422. [[CrossRef](#)] [[PubMed](#)]
47. Huang, W.; Sun, S.; Lin, X.; Li, P.; Zhu, L.; Wang, J.; Chen, C.L.P.; Sheng, B. Unsupervised fusion feature matching for data bias in uncertainty active learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**; early access. [[CrossRef](#)]
48. Wallace, B.C.; Small, K.; Brodley, C.E.; Trikalinos, T.A. Active learning for biomedical citation screening. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 25–28 July 2010; ACM: New York, NY, USA, 2010; pp. 173–182. [[CrossRef](#)]
49. Cohen, D.; Jordan, S.M.; Croft, W.B. Learning a Better Negative Sampling Policy with Deep Neural Networks for Search. In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, Santa Clara, CA, USA, 2–5 October 2019; ACM: New York, NY, USA, 2019; pp. 19–26. [[CrossRef](#)]
50. Eto, M. Extended co-citation search: Graph-based document retrieval on a co-citation network containing citation context information. *Inf. Process. Manag.* **2019**, *56*, 102046. [[CrossRef](#)]
51. Wang, X.; Macdonald, C.; Tonello, N.; Ounis, I. Pseudo-Relevance Feedback for Multiple Representation Dense Retrieval. In Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval, Virtual, 11 July 2021; ACM: New York, NY, USA, 2021; pp. 297–306. [[CrossRef](#)]
52. Carvalho, A.; Parra, D.; Lobel, H.; Soto, A. Automatic document screening of medical literature using word and text embeddings in an active learning setting. *Scientometrics* **2020**, *125*, 3047–3084. [[CrossRef](#)]
53. Ein-Dor, L.; Halfon, A.; Gera, A.; Shnarch, E.; Dankin, L.; Choshen, L.; Danilevsky, M.; Aharonov, R.; Katz, Y.; Slonim, N. Active Learning for BERT: An Empirical Study. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 7949–7962. [[CrossRef](#)]
54. Salvador-Oliván, J.A.; Marco-Cuenca, G.; Arquero-Avilés, R. Errors in search strategies used in systematic reviews and their effects on information retrieval. *J. Med. Libr. Assoc.* **2019**, *107*, 210–221. [[CrossRef](#)]
55. Tasneem, A.; Aberle, L.; Ananth, H.; Chakraborty, S.; Chiswell, K.; McCourt, B.J.; Pietrobon, R. The database for aggregate analysis of clinicaltrials.gov (aact) and subsequent regrouping by clinical specialty. *PLoS ONE* **2012**, *7*, e33677. [[CrossRef](#)]
56. Bethesda (MD): National Library of Medicine. Statistical Reports on MEDLINE®/PubMed® Baseline Data. Available online: https://www.nlm.nih.gov/bsd/licensee/2022_stats/2022_LO.html (accessed on 14 June 2023).
57. Tullu, M.S. Writing the title and abstract for a research paper: Being concise, precise, and meticulous is the key. *Saudi J. Anaesth.* **2019**, *13* (Suppl. 1), S12–S17. [[CrossRef](#)]
58. Klingmann, I.; Paul, M.; Berczky, T.; Plate, A.; Breithaupt-Grögler, K.; Schindler, T.M.; Geissler, J. Veränderungen durch die neue EU-Verordnung 536/2014 zu klinischen Prüfungen aus Patientensicht. *Bundesgesundheitsblatt Gesundh. Gesundh.* **2023**, *66*, 68–74. [[CrossRef](#)]

59. Ebeid, I.A. Medgraph: A semantic biomedical information retrieval framework using knowledge graph embedding for pubmed. *Front. Big Data* **2022**, *5*, 965619. [[CrossRef](#)]
60. Zarrinkalam, F.; Kahani, M. A multi-criteria hybrid citation recommendation system based on linked data. In Proceedings of the 2012 2nd International eConference on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, 18–19 October 2012; pp. 283–288. [[CrossRef](#)]
61. Burri, K.G. Intelligent use of artificial intelligence for systematic reviews of medical devices. *Med. Writ.* **2019**, *28*, 22–27.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.