PubMed and beyond: biomedical literature search in the age of artificial intelligence

*Qiao Jin, Robert Leaman, and Zhiyong Lu**

National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), National Institutes of Health (NIH), Bethesda, MD, USA

eBioMedicine 2024;100: 104988

Published Online xxx

https://doi.org/10.1016/j.ebiom.2024.104988

Summary

Biomedical research yields vast information, much of which is only accessible through the literature. Consequently, literature search is crucial for healthcare and biomedicine. Recent improvements in artificial intelligence (AI) have expanded functionality beyond keywords, but they might be unfamiliar to clinicians and researchers. In response, we present an overview of over 30 literature search tools tailored to common biomedical use cases, aiming at helping readers efficiently fulfill their information needs. We first discuss recent improvements and continued challenges of the widely used PubMed. Then, we describe AI-based literature search tools catering to five specific information needs: 1. Evidence-based medicine. 2. Precision medicine and genomics. 3. Searching by meaning, including questions. 4. Finding related articles with literature recommendation. 5. Discovering hidden associations through literature mining. Finally, we discuss the impacts of recent developments of large language models such as ChatGPT on biomedical information seeking.

Keywords: **Artificial intelligence; Biomedical literature search**

Introduction

In biomedicine, literature serves as the primary means of disseminating new findings and knowledge. Much of the information accumulated by biomedical research remains accessible only through the literature.[1] Consequently, literature search, the process of retrieving scientific articles to satisfy specific information needs, is important to all aspects of biomedical research and patient care. However, the exponential growth of biomedical literature makes it challenging to identify relevant information. PubMed, the most widely used biomedical literature search engine, currently contains over 36 million articles, with the addition of more than 1 million annually. A typical PubMed query retrieves hundreds to thousands of articles, yet fewer than 20% of the articles past the top 20 results are ever reviewed.[2,3] This motivated a shift in PubMed's approach from recency-based ranking to a relevance-based ranking,[4] to better prioritize the most relevant and significant articles.

PubMed primarily serves as a general-purpose biomedical literature search engine. Despite significant improvements over the past decades,[3] PubMed mainly receives short keyword-based queries from the users,[2] and returns a list of raw articles without further analysis. Consequently, it might not optimally serve specialised information needs, which require alternative query types or have specific requirements for ranking articles. A notable example is the unprecedented upsurge of publications addressing the COVID-19 pandemic.[5,6] While the pandemic made quickly disseminating new findings critical, obtaining comprehensive results from traditional search engines requires complex querying syntax that is unfamiliar to most users. Addressing the COVID-19 pandemic, therefore, required a specialised literature search engine capable of automatically collecting and classifying relevant articles.[7,8]

While various web-based literature search tools have been proposed over the past two decades to complement PubMed for specific literature search needs, they remain underutilised and unfamiliar to clinicians and researchers. This overview article aims to acquaint readers with available tools, discuss best practices, identify functionality gaps for different search scenarios, and ultimately facilitate biomedical literature retrieval. Table 1 enumerates the web-based literature search tools introduced in this article, categorized by the unique information needs they fulfill. Specifically, literature search tools are organised into five areas: (1) Evidence-based medicine (EBM), for identifying high-quality clinical evidence; (2) Precision medicine (PM) and genomics, for retrieving information related to genes or variants; (3) Semantic search, for finding textual units semantically related to the input query; (4) Literature recommendation, for suggesting related articles; and (5) Literature mining, for extracting biomedical concepts and their relations for literature-based discovery. Fig. 1 presents a

*Corresponding author. National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, MD, 20894, USA.

E-mail address: zhiyong.lu@nih.gov (Z. Lu).

high-level overview of the search scenarios. Search tools catering to different information needs differ in the types of queries they accept, their methods for processing articles and matching them to the input query, and how they present search results to users.

This article differs from previous surveys on biomedical literature search tools[9–12] in three important aspects: (1) We organize the literature search tools according to specific user scenarios and information needs; (2) Our study includes many new systems not

Resource	Website	Brief description
General-purpose search engines		
PubMed	https://pubmed.ncbi.nlm.nih.gov/	General-purpose biomedical literature search engine.
PubMed central	https://www.ncbi.nlm.nih.gov/pmc/	Supporting full-text search.
Europe PMC	https://europepmc.org/	Searching both abstracts and full-texts.
Information assembly and synthesis for evidence-based medicine		
PubMed clinical queries	https://pubmed.ncbi.nlm.nih.gov/clinical/	Searching clinical studies with various type and scope filters.
Cochrane library	https://www.cochranelibrary.com/	Searching high-quality systematic reviews.
Trip database	https://www.tripdatabase.com/	General EBM search engine.
Information linking for precision medicine and genomics		
LitVar	https://www.ncbi.nlm.nih.gov/research/litvar	Searching relevant information for all synonyms to the given variant.
Variant2literature	https://www.taigenomics.com/console/v2l	
DigSee	http://210.107.182.61/geneSearch/	Finding evidence sentences for the given (gene, disease, biological processes) triplet.
OncoSearch	http://oncosearch.biopathway.org/	Searching sentences that mention gene expression changes in cancers
Semantic search for similar sentences or question answers		
LitSense	https://www.ncbi.nlm.nih.gov/research/litsense/	Searching relevant sentences to the given query.
COVID-19 challenges and directions	https://challenges.apps.allenai.org/	Searching COVID-19 challenges and future directions for the given topic.
askMEDLINE	https://pubmedhh.nlm.nih.gov/ask/index.php	Answering the query question with documents or text snippets in literature.
COVID-19 research explorer	https://covid19-research-explorer.appspot.com/biomedexplorer/	Answering the original question and follow-up questions with text snippets in literature
BioMed explorer	https://sites.research.google/biomedexplorer/	
Literature recommendation for specific topics or similar articles		
LitCovid	https://www.ncbi.nlm.nih.gov/research/coronavirus/	Literature hubs for COVID-19.
WHO COVID-19 research database	https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov	
iSearch COVID-19 portfolio	https://icite.od.nih.gov/covid19/search/	
Corona central	https://coronacentral.ai/	
COVID-SEE	https://covid-see.com/search	
COVIDScholar	https://covidscholar.org/	
LitSuggest	https://www.ncbi.nlm.nih.gov/research/litsuggest/	Scoring article candidates based on user-provided positive and negative articles.
BioReader	https://services.healthtech.dtu.dk/service.php?BioReader-1.2	
Connected papers	https://www.connectedpapers.com/	Recommending relevant articles to one or more seed articles using the citation graph.
Litmaps	https://www.litmaps.com/	
Literature mining for knowledge discovery		
PubTator	https://www.ncbi.nlm.nih.gov/research/pubtator/	Highlighting biomedical concepts in the retrieved documents.
Anne O'Tate	http://arrowsmith.psych.uic.edu/cgi-bin/arrowsmith_uic/AnneOTate.cgi	Ranking the extracted concepts from the search results.
FACTA+	http://www.nactem.ac.uk/facta/index.html	Finding directly and indirectly associated concepts to the given concept.
Semantic MEDLINE	https://ii.nlm.nih.gov/SemMed/semmed.html	Displaying graphs of biomedical concepts and their relations extracted from the retrieved documents.
SciSight	https://scisight.apps.allenai.org/	
PubMedKB	https://www.pubmedkb.cc/	
LION LBD	https://lbd.lionproject.net/	
(Experimental) literature search systems augmented by LLMs		
Scite	https://hippocratic-medical-questions.herokuapp.com/	Finding relevant articles to users' question and then using LLMs to answer the question with the retrieved articles
Elicit	https://elicit.org/	
Consensus	https://consensus.app/	

Literature search tools included in this study are web-based, freely available, regularly maintained, and designed for searching the biomedical literature.

Table 1: **Web-based biomedical literature search tools.**

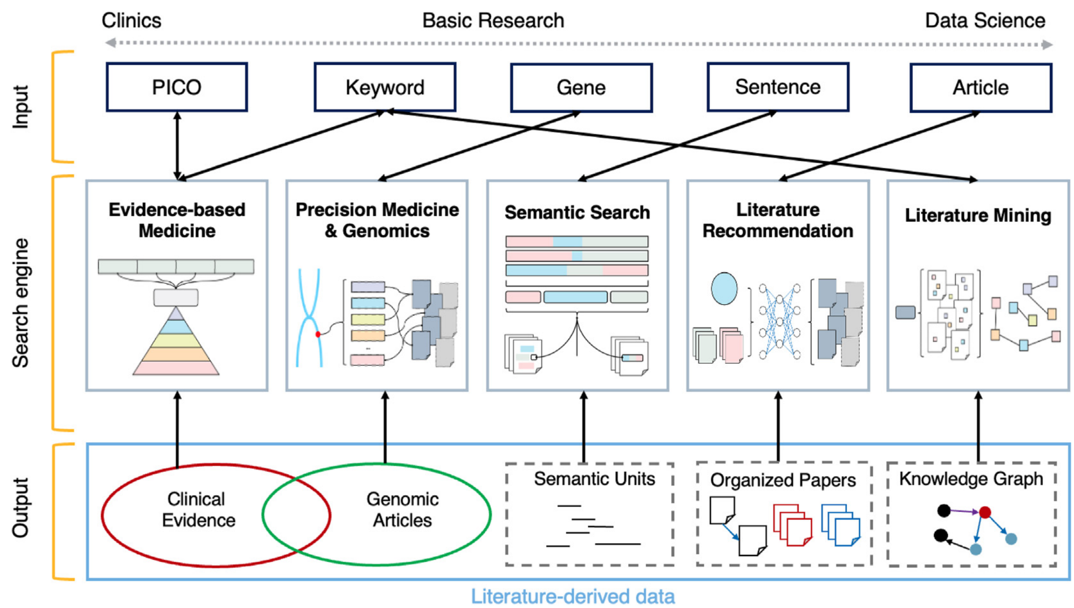

Fig. 1: Overview of five specialised search scenarios in biomedicine: evidence-based medicine, precision medicine & genomics, semantic search, literature recommendation, and literature mining. Each search scenario is characterized by its unique input interface, search or ranking algorithm, and output display.

covered by previous surveys; (3) Beyond surveying current systems, we also cover practical considerations and best practices of using these tools; (4) We share our perspective on the development of next-generation biomedical literature search engines, especially how large language models (LLM) such as ChatGPT could be utilised to improve the discussed search scenarios. Our goal is to provide a comprehensive overview of specialised literature search tools for researchers and clinicians, which enables more effective exploration of biomedical information and higher-quality care for their patients.

Search strategy and selection criteria

In this overview, we search “biomedical literature search”, “medical literature search” and “clinical literature search” on PubMed and Google Scholar to find candidate articles that describe biomedical literature search tools. We only include literature search tools that meet the following criteria in our study: (1) the tool should be web-based and regularly maintained, (2) the tool should be freely available without subscription, (3) the tool should be designed for searching the biomedical literature. Consequently, general-domain literature search engines such as Web of Science, Scopus, Google Scholar, and Semantic Scholar, are not included.

PubMed & PubMed central: the first stop

PubMed is developed and maintained by the US National Library of Medicine. In 2021, it averaged approximately 2.5 million queries daily. The PubMed search engine seeks exact matches for user queries in the indexed fields of each article, including the title, abstract, author list, keywords, and MeSH terms. Traditionally, all matching articles were returned in reverse chronological order. A new AI-based ranking model—Best Match—was introduced in 2017 to better assist users by returning the most relevant articles among the top results.[4] Beyond relevance search for biomedical topics, PubMed also supports various other search functionalities. These include matching single citations through bibliographic information such as title and journal names, as well as Boolean operators that are usually used when conducting systematic reviews.

However, since PubMed does not index full-text articles, those that match the query in the full-text but not in the abstract or the title will not be retrieved. Such queries are accommodated by PubMed Central (PMC), which provides access to more than 9 million freely available full-text articles. Unfortunately, PMC does not support searching the other 27 million PubMed articles that lack full-text availability. Europe PMC,[13] a PMC partner, contains both 42.7 million abstracts and 9.0 million full-text articles as of July 2023.

Best practice and example use case

PubMed should be the first choice for three types of literature search practices: (1) exploring biomedical topics via keyword query such as “diabetes treatment”, with PMC enabling keyword search within the full text, when available; (2) searching for single citations with article titles, authors, or PubMed IDs; (3) reproducible literature screening with Boolean queries.

Information assembly and synthesis for evidence-based medicine

Evidence-based medicine (EBM)[14] requests clinical practitioners follow high-quality evidence, primarily derived from peer-reviewed articles of clinical studies. Efficient retrieval of this evidence is crucial for implementing EBM.[15] Accordingly, clinical questions should be structured effectively, incorporating at least the "PICO" elements[16] (**P**opulation, **I**ntervention, **C**omparison, and **O**utcome). For example, in "Does remdesivir reduce in-hospital mortality for patients with COVID-19 compared to placebo?", the PICO elements are COVID-19 (Population), remdesivir (Intervention), placebo (Comparison), and in-hospital mortality (Outcome), respectively. EBM search engines should be equipped to process both PICO and natural language clinical questions.

Clinical evidence spans a broad spectrum of literature, with significant variability in quality. For example, systematic reviews are generally considered as higher-quality evidence than randomized controlled trials (RCTs), which, in turn represent higher quality than individual case reports. Consequently, an ideal EBM search engine should consider the quality of evidence for filtering or ranking the articles. Fig. 2 depicts the architecture of an ideal EBM search engine, which allows PICO-style input and ranks results based on evidence quality.

Systems accepting PICO queries

Several EBM search engines, such as Trip Database, the Cochrane PICO search, and Embase, accommodate PICO-based queries. The search interfaces for these systems typically contain text boxes corresponding the four primary PICO elements. In general, these systems provide more precise results since the search intent is explicitly stated in the query. For example, entering "diabetes" as the "Population" term, prompts EBM search engines to only return clinical studies on patients with diabetes. In contrast, keyword-based search engines would return any article that mentions "diabetes," regardless of its relevance to patient studies.

Systems with filtered retrieval results

PubMed Clinical Queries search employs predefined filters[17,18] for clinical studies of various types, such as therapy and diagnosis. Users can also select broad or narrow scopes for the filters. Clinical practitioners should use the narrow scope for a quick overview of the important studies at the point of care, while researchers synthesizing evidence should employ the broad scope for exhaustive searches. Several EBM search engines prioritize retrieval of secondary evidence, such as systematic reviews, which typically have higher quality. A notable example is the Cochrane Database, which hosts over 11 thousand high-quality systematic reviews and protocols. Critically-appraised topics summarize the evidence on a specific topic, such as prevention of type 2 diabetes mellitus, using short, templated, titles to simplify retrieval. As a result, they provide point-of-care evidence that can guide clinical decision-making.

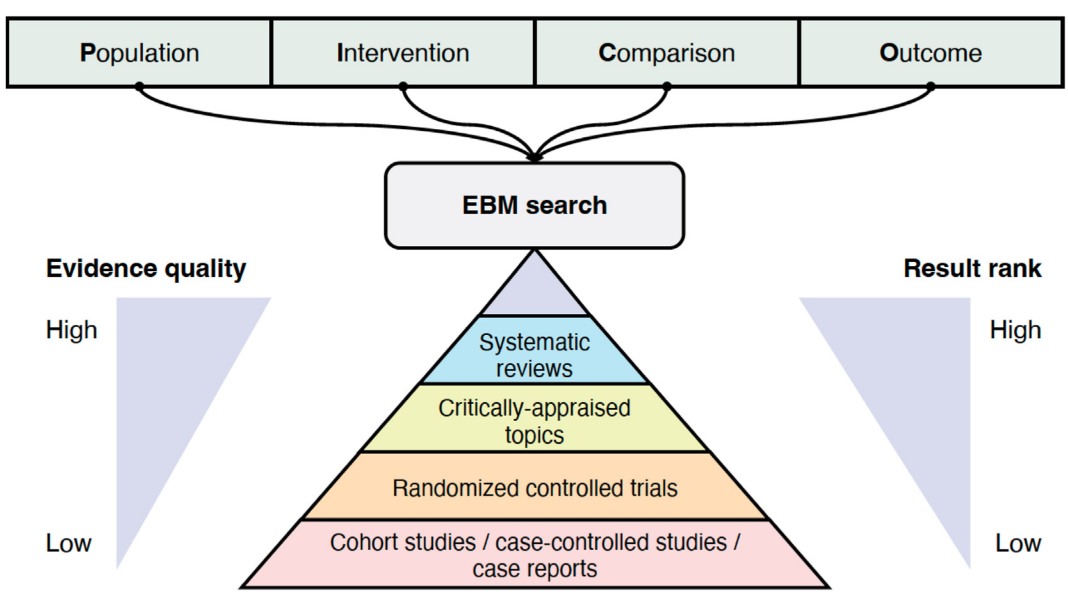

Fig. 2: The architecture of a search engine for evidence-based medicine (EBM). EBM search engines should incorporate PICO elements (Population, Intervention, Comparison, and Outcome) within the input query and rank the articles returned based on the quality of the evidence.

Assisting evidence synthesis

Compared to evidence retrieval, fewer systems facilitate evidence synthesis, which denotes the systematic collection, analysis and combination of results from multiple research studies to reach a comprehensive conclusion about a specific question or topic.[19] Evidence synthesis plays a vital role in the systematic reviewing process. However, the user conducting a systematic review would need to manually screen all related literature to address a clinical question without bias, an extremely time-consuming process due to the vast number of articles likely to be relevant across multiple databases.[20] Despite efforts to use machine learning to automate this screening process,[21–26] these features are not yet integrated into web-based EBM search engines due to the intrinsic complexity and low tolerance for errors in this task.

Best practice and example use case

Literature search is a vital step in evidence-based medicine. To optimize this process, users should: (1) formulate clinical questions in the format of PICO elements; (2) utilise a system that ranks relevant studies by their evidence quality. For example, to obtain the best evidence, the physician could use an EBM search engine like Cochrane PICO search or Trip Database, inputting the PICO components. The search engine would then prioritize systematic reviews and randomized controlled trials relevant to the question.

Information linking for precision medicine and genomics

Precision medicine (PM) is an emerging approach that tailors disease treatment and prevention based on individual variations in genes, environment, and lifestyle.[27] The rapid development of high-throughput sequencing techniques have precipitated a sharp decline in the cost of obtaining individual genomic data. Human genomes, with their high heterogeneity, contain a large number of genomic variants.[28] Understanding the biological function and clinical significance of these genomic variants is essential for the advancement of precision medicine. Such information is typically stored in manually curated databases such as UniProt,[29] dbSNP,[30] and ClinVar.[31] These databases manually summarize and maintain primary findings from the literature about each data entry. However, the growth of the biomedical literature, with an average of 3000 new articles per day,[1] outpaces the speed of manual curation, leaving a knowledge gap. To supplement these databases, search engines capable of extracting gene or variant-related information directly from raw literature are needed. This section primarily discusses such systems.

A significant challenge for PM and genomics search engines is the presence of multiple representations for the same variant. For instance, the variant "V600E" could also be referred to as "1799T > A" or "rs113488022." This synonymy causes retrieval challenges for keyword-based search engines. In response, many specialised literature retrieval tools have been proposed; their core functionality is shown in Fig. 3, where the search engine should be able to retrieve all articles that mention the exact variant query as well as its synonyms.

Recognizing synonymous mentions

Some tools, such as LitVar,[32,33] focus on normalizing variant synonyms in the literature. LitVar uses text mining tool tmVar[34,35] to recognize variant names and convert them to standardized form. LitVar indexes both abstracts from PubMed and full-texts from PubMed Central and is updated regularly to ensure retrieval of all current literature containing synonyms of the query. Another tool, variant2literature,[36] provides a structured query interface that allows users to specify a chromosome location. Unique to variant2literature is the ability to extracts variants from figures and tables in addition to the article text.

Linking genes and other information

Several systems go beyond recognizing synonymous gene mentions and explore genomic-related information. DigSee[37] accepts a triplet of gene, disease, and biological processes as input and finds sentences in PubMed abstracts that link the gene to the disease through the given biological processes. OncoSearch[38] specialises in retrieving literature evidence for gene expression changes and cancer progression status. Specifically, it annotates sentences from the literature to indicate whether the input gene is up-regulated or down-regulated, whether the input cancer progresses or regresses with the expression change, and the expected role of the gene in the cancer.

Best practice and example use case

To find genomic information, we recommend first querying curated databases such as UniProt and ClinVar. For more recent findings or when these databases lack sufficient contextualised information, the use of search engines specialised for precision medicine and genomics is recommended. For example, LitVar can assist in finding information within the literature about the role of certain genomic variants in an emerging disease, which might not have been curated into structured databases yet.

Semantic search for similar sentences or question answers

Unlike the keyword-based search that seek exact matches for the input query, semantic search locates texts that are semantically related to the query. For example, "renal" and "kidney" are semantically very similar. Fig. 4 outlines semantic search, where text units

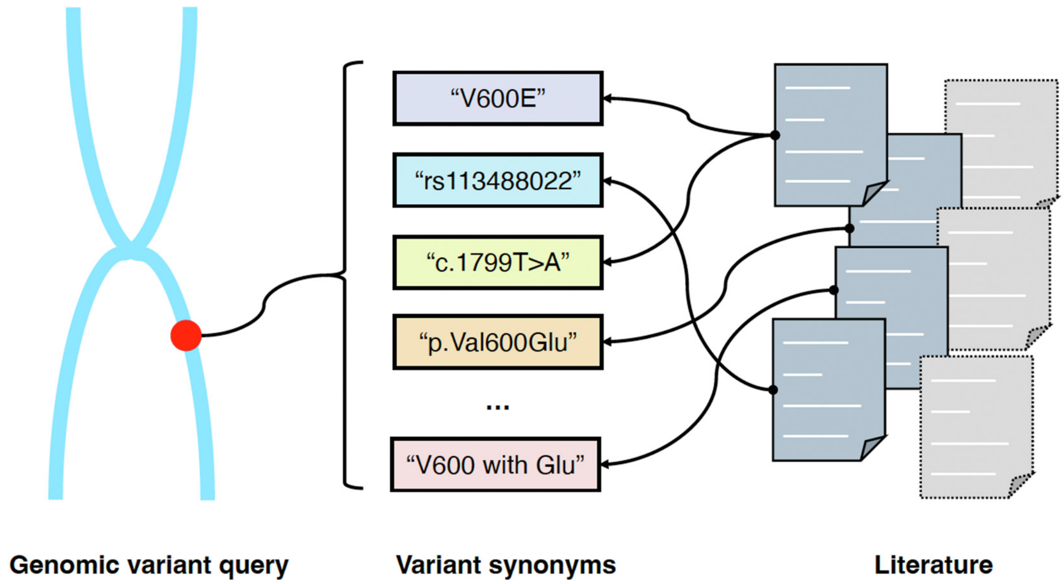

Fig. 3: Illustration of the functionality of a search engine for precision medicine and genomics. Search engines for precision medicine and genomics should handle queries containing genomic variants and identify all synonymous references to these variants in the literature.

such as sentences that match the query semantically are returned, such as mentioning the same diseases and discussing possible treatments. These texts do not necessarily contain the exact query terms, making their retrieval by traditional literature search engines unlikely. We introduce search engines for two common types of semantic relevance: similar sentences and question–answer pairs.

Similar sentence search

Article-level searches often overlook finer-grained information in sentences. Sentence-level searches are

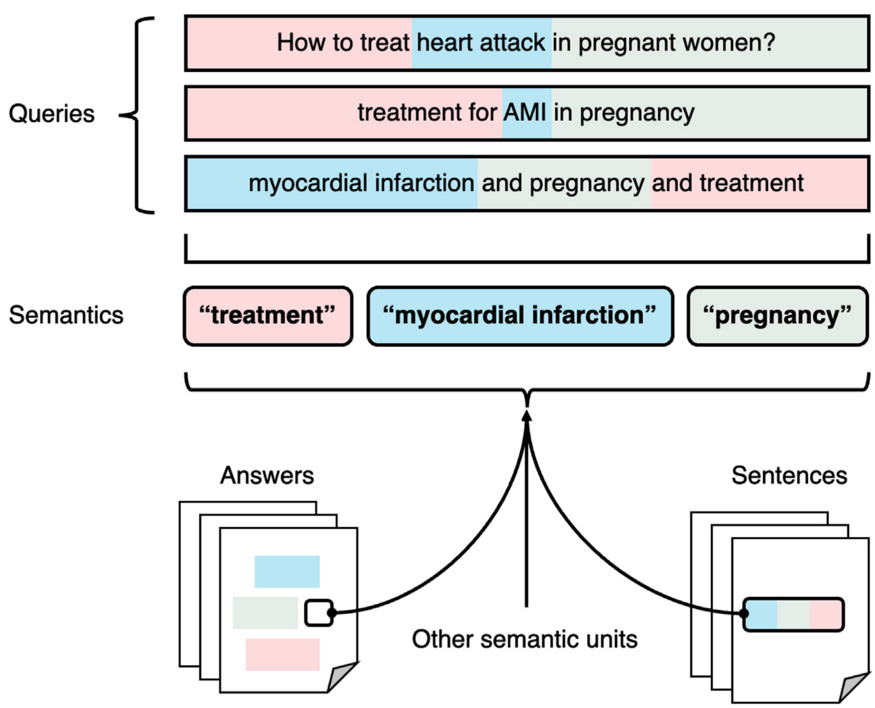

Fig. 4: Depiction of semantic search. Unlike traditional keyword-based search engines, semantic search engines process words and phrases according to their meaning rather than the literal text. For instance, "heart attack", "AMI", and "myocardial infarction" share similar meanings.

important for precise knowledge retrieval. For example, one can search for a particular finding and compare it with relevant findings from other articles. LitSense,[39] a web-based system for sentence retrieval from PubMed and PMC, utilises a retrieval system that can match texts by their semantics through a deep learning-based technique called "embeddings" that involves inferring word representations from the context.[40] Results in LitSense can be filtered by sections, such as Conclusions. While LitSense searches for all types of similar sentences, several literature search engines have also been proposed for more specific types of sentences. For example, Lahav et al. present a search engine for sentences that describe challenges and future directions in COVID-19,[41] and SciRide Finder[42] finds cited statements describing the in-line references.

Question answering

Biomedical inquiries are often naturally expressed as questions, such as the PICO-based clinical questions in EBM. However, traditional keyword-based search engines may not efficiently handle natural language questions because questions and answers often lack high lexical overlap. Biomedical question answering (QA) is an active research area,[43] but user-friendly web tools remain sparse. The askMEDLINE[44] system evolved from PubMed PICO search and enables direct input to the clinical questions, e.g., "Is irrigation with tap water an effective way to clean simple laceration before suturing?". askMEDLINE displays results as a list of relevant articles. COVID-19 Research Explorer and BioMed Explorer are experimental semantic search engines for biomedical literature developed by Google AI. The former focuses on COVID-19 articles, and the latter encompasses all PubMed articles. Users ask natural language questions, and the answers are highlighted in the text snippets in the results. Users can also pose follow-up questions to further investigate the research topic.

Best practice and example use case

Users should consider using semantic search engines if their information needs are better expressed by natural language instead of keywords. Available tools include LitSense for finding relevant sentences and BioMed Explorer for answering biomedical questions with evidence from the literature.

Literature recommendation for specific topics or similar articles

Biomedical research often requires comprehensive exploration of related literature. Traditional keyword-based search engines are typically inefficient for this purpose due to the difficulty of formulating queries to exhaustively capture all relevant work. Literature recommendation engines instead allow users to explore articles relevant to a specific research topic or similar to a list of articles known to be relevant. This section mainly introduces two types of literature recommendation tools: topic-based and article-based, as depicted in Fig. 5.

Topic-based literature recommendation systems are typically curated databases or literature hubs tailored to selected research topics, such as the COVID-19 pandemic. For example, due to the initial lack standardized terminology for SARS-CoV-2 and COVID-19, publications used a variety of terms, complicating identifying relevant articles through keyword-based or Boolean searches. LitCovid,[8,45] a curated literature hub containing COVID-19-related articles from PubMed, is organized with eight broad topics, including mechanism, transmission, diagnosis, and treatment. Chen et al. demonstrated that LitCovid identifies about 30% more PubMed articles than a complex, purpose-built Boolean query.[8] Other literature hubs dedicated to COVID-19 include CoronaCentral,[46] COVID-SEE,[47] COVIDScholar[48] and etc.

Article-based literature recommendation systems, on the other hand, generate a list of articles related to initial (seed) articles. Modern literature search engines often provide a list of articles related to individual articles, such as the "similar articles" section in PubMed. A few systems have been proposed, however, which support identifying articles related to a list of articles instead of individual ones. LitSuggest,[49] a literature recommendation system based on machine learning, rates candidate articles on their similarity to a user-supplied list of positive articles and dissimilarity to an optional list of negative articles. Users can also provide human-in-the-loop feedback by annotating a subset of the scored candidate articles and re-training the recommendation model. BioReader[50] offers similar functionality, but it requires a list of negative articles. Several commercial literature search tools like Connected Papers[a] and Litmaps[b] provide visual representations of articles related to seed articles on a citation graph, thus aiding in the navigation of the academic literature and guiding focused research.

Best practice and example use case

Recommendation systems primarily assist in literature exploration. Users can find articles related to a topic of interest, such as COVID-19, using a curated literature database, or locate articles similar to a specific list of articles through article-based literature recommenders like LitSuggest.

Literature mining for knowledge discovery

Literature mining aims to help users uncover novel insights from scientific publications through natural language processing (NLP) techniques.[40] These techniques include named entity recognition (NER), the task of

[a] https://www.connectedpapers.com/.
[b] https://www.litmaps.com/.

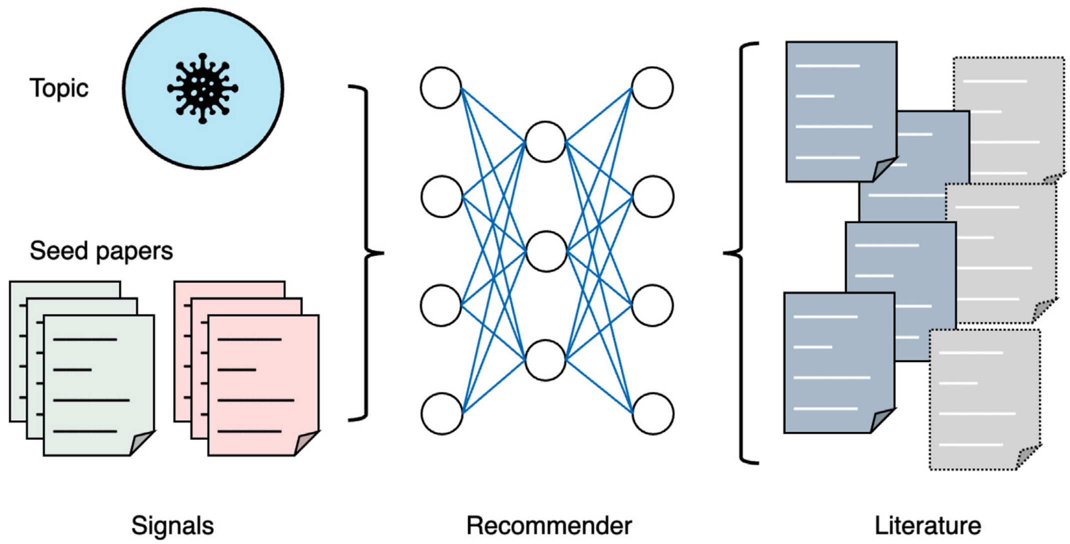

Fig. 5: Illustration of topic-based and article-based literature recommendation systems. Topic-based systems provide articles relevant to a specific topic (e.g., COVID-19), while article-based systems return articles similar to a group of initial (seed) articles and dissimilar to a group of irrelevant articles.

recognizing biomedical concepts such as genes and diseases,[51] and relation extraction (RE), which classifies relations between the concepts identified.[52] For example, an NER tool could identify a genetic variant and a disease name in a sentence, and an RE tool might classify their relation as mutation-causing-disease. Extracted concepts and their relations can be organized into a graph, referred to as a knowledge graph, which structurally summarizes the knowledge encoded in the publications related to the given query. By displaying a knowledge graph, literature search engines provide users with an overview of the knowledge discovered, thereby facilitating new knowledge discovery by predicting potential missing links. This process is visualised in Fig. 6.

Entity-augmented search

Several literature search engines enhance the retrieved results with biomedical concepts. PubTator[53,54] highlights six types of concepts recognized by state-of-the-art NER tools, such as genes and diseases. PubTator has also made its annotations publicly available via bulk download and an application programming inference, allowing other search engines to augment the search results with PubTator concepts. Notably, PubTator has

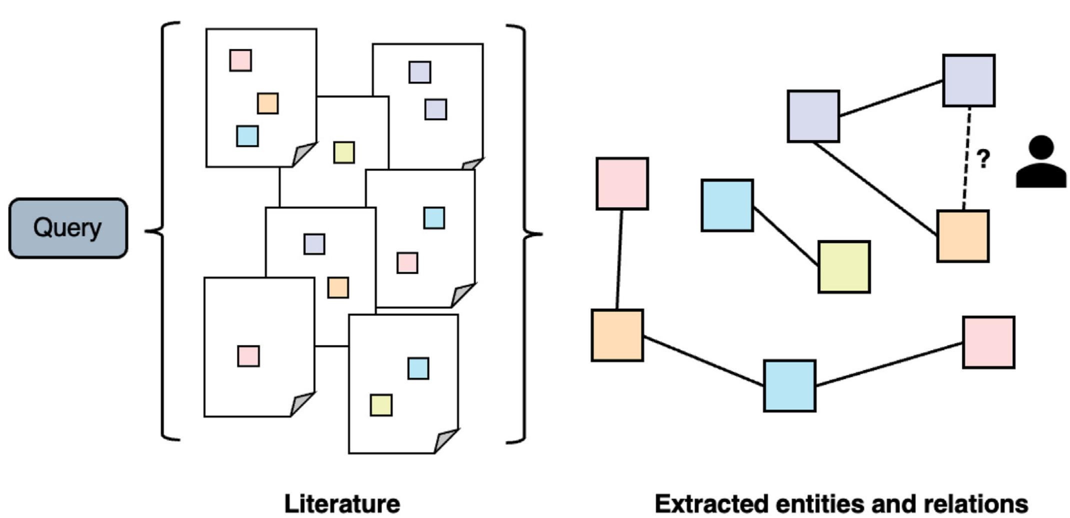

Fig. 6: The architecture of a system for mining entity associations from the biomedical literature. The system retrieves articles relevant to a given user query, extracts biomedical entities and their relationships (e.g., variant-causing-disease), and presents the search results as a knowledge graph that visualises the extracted entities and their relationships. The users can use the knowledge graph to find hidden associations between entities.

been integrated into platforms such as LitVar, LitSense, and LitCovid. Anne O'Tate[55] provides options to rank concepts, such as important words, important phrases, topics, authors, MeSH pairs, etc., that are extracted from the retrieved articles.

Relation-augmented search

Some systems further process the extracted concepts and show the search results using associated concepts. FACTA+[56] finds concepts associated with the given concept and the supporting sentences and can uncover indirectly associated concepts through certain types of "pivot concepts" as the bridge. Semantic MEDLINE[57] extracts predications, which consist of two biomedical concepts and one relation, from the retrieved articles and provides a graph visualization of the predications. SciSight,[58] an exploratory search system for COVID-19, can present a graph of biomedical concepts associated with the given concept. PubMedKB[59] extracts and visualises semantic relations between variants, genes, diseases, and chemicals, offering a user interface with interactive semantic graphs for the input query. While many systems for constructing biomedical knowledge graphs automatically have been proposed, their utility remains to be confirmed in future studies. Literature mining systems can also facilitate Literature-based Discovery (LBD).[60] For instance, the LION LBD system[61] presents the search results as a graph that contains biomedical concepts and their relations extracted from the literature for discovering novel knowledge.

Best practice and example use case

Literature mining tools can be employed to study the associations between biomedical concepts in the literature. Users should consider the concept and relation types of interest and choose the literature mining tools that incorporate such information. For example, PubTator provides annotations for six general concept types, but concepts beyond these types are better supported in other literature search tools, such as SciSight for COVID-19 concepts and relations.

Looking ahead: the role of ChatGPT and other large language models in literature search

Since late 2022, ChatGPT[62] and other generative large language models (LLMs) have demonstrated considerable performance improvements on both general and biomedical NLP tasks.[63–65] LLMs typically contain billions of parameters and can be utilised by prompt engineering, such as in-context learning and retrieval augmentation, to generate human-like responses to various contexts.[66] There is a rising belief that these models could significantly change how users interact biomedical literature.

Evidence-based medicine

LLMs can accelerate evidence synthesis in two ways. First, they can suggest Boolean queries to aid literature screening for systematic reviews.[67] Following the retrieval of results, LLMs could potentially be used to summarize and synthesize the resulting articles.[68–70] However, these preliminary evaluations have exposed various issues, such as potential bias and hallucination, which must be addressed before widespread use. Apart from evidence synthesis, LLMs can also enhance the extraction of PICO elements from the medical literature,[71] thereby improving PICO-based EBM search engines.

Precision medicine and genomics

Most genomics information resides in curated databases, which are not easily accessible due to their keyword-centric search functions and less modern user interfaces. LLMs can alleviate these access difficulties by autonomously utilizing tools such as utilities of specialised databases,[72] and directly summarize the database entries to answer users' information-seeking questions.

Semantic search

LLMs have achieved state-of-the-art performance on several biomedical QA datasets.[65] This suggests that LLMs can provide direct answers to natural language questions using relevant documents returned from a traditional search engine. This feature, called retrieval augmentation, is already supported by experimental literature search engines such as scite,[c] and Elicit.[d] However, these LLM-generated answers are susceptible to errors and should be carefully verified before use.[73]

Literature recommendation

The potential role of LLMs in literature recommendation remains largely unexplored. One possibility involves using LLMs to explain literature recommendations, i.e., describing why a recommended article is similar to the input article. This capability could be used to create a dataset for training smaller generative models, enabling more flexible and cost-effective recommendation explanations.

Literature mining

Unlike other literature search scenarios that directly benefit from the generative capabilities of LLMs, literature mining depends on traditional NLP tasks such as NER and RE. In general, LLMs do not outperform smaller task-specific models fine-tuned for these tasks.[74] However, LLMs may offer superior interpretations of the constructed knowledge graphs, revealing previously unknown associations between biomedical concepts.

[c] https://scite.ai/.
[d] https://elicit.org/.

Discussion

We introduced five specific use cases of biomedical literature search and available tools for each scenario. Our organisation, while practical, is not mutually exclusive, and the advantages of different systems can be combined to better meet diverse biomedical information needs. For instance, an EBM search engine might also process queries where the specified Population is associated with certain genomic variants, necessitating recognition of variant synonyms for comprehensive literature retrieval. Another instance is biocuration, the practice of converting literature data into database entries. A system to support biocuration should be equipped with both literature recommendation and mining functionality to assist biocurators by suggesting relevant publications and highlighting the relevant biomedical concepts. Beyond the five specific use cases discussed in this article, there are also other information needs for biomedical literature, such as searching figures within the articles. It is important to recognize that while AI advances in healthcare, ensuring a human-centered approach is pivotal to address its broader implications.[75]

Analogous to web search, literature search queries generally comprise several words.[2,3] However, more complex or specialised information needs require interfaces capable of processing semi-structured information or even non-text modalities. Semi-structured search interfaces accept separate texts for multiple pre-defined fields, akin to the advanced search interface in modern literature search engines and PICO-based EBM search. Some information needs defy expression in text, such as finding articles that are similar to one set, requiring interfaces designed specifically for the task. Although modern search interfaces consisting of one text box are simple and easy to use, the resulting queries can be ambiguous or overly general. As such, task-oriented search interfaces should be designed for different biomedical literature search purpose, while a unified portal can be employed to triage the information needs into these task-oriented interfaces.

In literature search engines, the ranking algorithms assess article relevance for a given query, thereby determining which articles are returned to the user. PubMed employs the Best Match[4] ranking model, a machine learning approach trained via user click logs. Many other algorithms rank articles based on the importance of the terms which overlap between the article and the query. These algorithms calculate general text-based relevance without domain-specific requirements, while certain biomedical subdomains have specific article ranking requirements. For example, in EBM, articles with higher quality clinical evidence should be ranked higher. In semantic search, articles with text units that are semantically related to the input query should be returned, irrespective of term overlap. In addition to performing purpose-specific ranking, future literature search engines should incorporate transparent and interpretable ranking algorithms.

Search results are most commonly displayed as a list of article metadata, mimicking the general web search engines familiar to users. Though list-based display has been almost unchanged in general search engines for decades, additional modules have been introduced to serve specific information needs. For example, many web search engines directly display the answer to a question query at the top of the results, mirroring the goal of QA-based semantic search in biomedical literature. Certain literature mining systems construct and visualise a knowledge graph from the articles retrieved, aiding exploration and knowledge discovery. Given the remarkable text generation capabilities of LLMs, we anticipate future literature search engines will include high-level overviews of returned articles generated by LLMs.

Conclusion

Our aim has been to assist biomedical researchers and clinicians in finding the most suitable literature search tool to fulfill various information needs. We characterized search scenarios for five specific information needs: evidence-based medicine, precision medicine and genomics, semantic search, literature recommendation, and literature mining. We also included 34 web-based AI systems designed for these scenarios. Finally, we discussed the future of biomedical literature search, especially considering the potential impacts of large language models such as ChatGPT.

Outstanding questions

As introduced in this overview, many biomedical literature search engines are specialised for specific information needs. However, it is hard for users to find a suitable tool that can efficiently fulfill their information needs, and this article is aimed at assisting them in such a process. Future work should utilise the rapidly developing AI techniques, especially large language models, to automatically triage the information needs of users and provide them the right tool to use.

Contributors

QJ: conceptualisation, investigation, writing. RL: conceptualisation, investigation, writing. ZL: conceptualisation, investigation, writing, supervision. All authors read and approved the final version of the manuscript.

Declaration of interests

None declared.

Acknowledgements

This research was supported by the National Institutes of Health Intramural Research Program, National Library of Medicine. The funders had no role in the design, data collection, data analysis, interpretation, and writing of the article.

References

1 Baumgartner WA Jr, Cohen KB, Fox LM, Acquaah-Mensah G, Hunter L. Manual curation is not sufficient for annotation of genomic databases. *Bioinformatics*. 2007;23:i41–i48.
2 Islamaj Dogan R, Murray GC, Neveol A, Lu Z. Understanding PubMed user search behavior through log analysis. *Database (Oxford)*. 2009;2009:bap018.
3 Fiorini N, Leaman R, Lipman DJ, Lu Z. How user intelligence is improving PubMed. *Nat Biotechnol*. 2018.
4 Fiorini N, Canese K, Starchenko G, et al. Best match: new relevance search for PubMed. *PLoS Biol*. 2018;16:e2005343.
5 Callaway E, Cyranoski D, Mallapaty S, Stoye E, Tollefson J. The coronavirus pandemic in five powerful charts. *Nature*. 2020;579:482–483.
6 Li G, Zhou Y, Ji J, Liu X, Jin Q, Zhang L. Surging publications on the COVID-19 pandemic. *Clin Microbiol Infect*. 2021;27:484–486.
7 Chen Q, Allot A, Lu Z. Keep up with the latest coronavirus research. *Nature*. 2020;579:193.
8 Chen Q, Allot A, Lu Z. LitCovid: an open database of COVID-19 literature. *Nucleic Acids Res*. 2021;49:D1534–D1540.
9 Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)*. 2011;2011:baq036.
10 Keepanasseril A. PubMed alternatives to search MEDLINE: an environmental scan. *Indian J Dent Res*. 2014;25:527–534.
11 Wildgaard LE, Lund H. Advancing PubMed? A comparison of third-party PubMed/Medline tools. *Libr Hi Technol*. 2016;34:669–684.
12 Jacome AG, Fdez-Riverola F, Lourenco A. BIOMedical search engine framework: lightweight and customized implementation of domain-specific biomedical search engines. *Comput Methods Programs Biomed*. 2016;131:63–77.
13 Europe PMCC. Europe PMC: a full-text literature database for the life sciences and platform for innovation. *Nucleic Acids Res*. 2015;43:D1042–D1048.
14 Sackett DL. Evidence-based medicine. *Semin Perinatol*. 1997;21:3–5.
15 Jin Q, Tan C, Chen M, et al. State-of-the-Art evidence retriever for precision medicine: algorithm development and validation. *JMIR Med Inform*. 2022;10:e40743.
16 Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123:A12–A13.
17 Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR, Hedges T. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *BMJ*. 2005;330:1179.
18 Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *J Am Med Inform Assoc*. 1994;1:447–458.
19 Higgins JP, Thomas J, Chandler J, et al. *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons; 2019.
20 Wallace BC, Trikalinos TA, Lau J, Brodley C, Schmid CH. Semi-automated screening of biomedical citations for systematic reviews. *BMC Bioinformatics*. 2010;11:55.
21 Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Syst Rev*. 2019;8:163.
22 Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. *J Am Med Inform Assoc*. 2016;23:193–201.
23 Nye B, Li JJ, Patel R, et al. A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature. In: *Proceedings of the conference association for computational linguistics. Meeting*. vol. 2018. NIH Public Access; 2018.
24 Suster S, Baldwin T, Verspoor K. Analysis of predictive performance and reliability of classifiers for quality assessment of medical evidence revealed important variation by medical area. *J Clin Epidemiol*. 2023;159:58–69.
25 Suster S, Baldwin T, Lau JH, et al. Automating quality assessment of medical evidence in systematic reviews: model development and validation study. *J Med Internet Res*. 2023;25:e35568.
26 Yan S, Luo L, Lai PT, et al. PhenoRerank: a re-ranking model for phenotypic concept recognition pre-trained on human phenotype ontology. *J Biomed Inform*. 2022;129:104059.
27 Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372:793–795.
28 Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
29 UniProt C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49:D480–D489.
30 Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29:308–311.
31 Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42:D980–D985.
32 Allot A, Peng Y, Wei CH, Lee K, Phan L, Lu Z. LitVar: a semantic search engine for linking genomic variant data in PubMed and PMC. *Nucleic Acids Res*. 2018;46:W530–W536.
33 Allot A, Wei CH, Phan L, et al. Tracking genetic variants in the biomedical literature using LitVar 2.0. *Nat Genet*. 2023;55:901–903.
34 Wei CH, Harris BR, Kao HY, Lu Z. tmVar: a text mining approach for extracting sequence variants in biomedical literature. *Bioinformatics*. 2013;29:1433–1439.
35 Wei CH, Phan L, Feltz J, Maiti R, Hefferon T, Lu Z. tmVar 2.0: integrating genomic variant information from literature with dbSNP and ClinVar for precision medicine. *Bioinformatics*. 2018;34:80–87.
36 Lin Y-H, Lu Y-C, Chen T-F, et al. variant2literature: full text literature search for genetic variants. *bioRxiv*. 2019. https://doi.org/10.1101/583450.
37 Kim J, So S, Lee HJ, Park JC, Kim JJ, Lee H. DigSee: disease gene search engine with evidence sentences (version cancer). *Nucleic Acids Res*. 2013;41:W510–W517.
38 Lee HJ, Dang TC, Lee H, Park JC. OncoSearch: cancer gene search engine with literature evidence. *Nucleic Acids Res*. 2014;42:W416–W421.
39 Allot A, Chen Q, Kim S, et al. LitSense: making sense of biomedical literature at sentence level. *Nucleic Acids Res*. 2019;47:W594–W599.
40 Zhao S, Su C, Lu Z, Wang F. Recent advances in biomedical literature mining. *Brief Bioinform*. 2021;22:bbaa057.
41 Lahav D, Saad Falcon J, Kuehl B, et al. A search engine for discovery of scientific challenges and directions. *Proc AAAI Conf Artif Intell*. 2022;36:11982–11990.
42 Volanakis A, Krawczyk K. SciRide Finder: a citation-based paradigm in biomedical literature search. *Sci Rep*. 2018;8:6193.
43 Jin Q, Yuan Z, Xiong G, et al. Biomedical question answering: a survey of approaches and challenges. *ACM Comput Surv*. 2022;55:1–36.
44 Fontelo P, Liu F, Ackerman M. askMEDLINE: a free-text, natural language query tool for MEDLINE/PubMed. *BMC Med Inform Decis Mak*. 2005;5:5.
45 Chen Q, Allot A, Leaman R, et al. LitCovid in 2022: an information resource for the COVID-19 literature. *Nucleic Acids Res*. 2023;51:D1512–D1518.
46 Lever J, Altman RB. Analyzing the vast coronavirus literature with CoronaCentral. *Proc Natl Acad Sci U S A*. 2021;118.
47 Verspoor K, Šuster S, Otmakhova Y, et al. In: *Brief description of covid-see: the scientific evidence explorer for covid-19 related research. Advances in information retrieval:* 43rd European conference on IR research, ECIR 2021, virtual event, March 28–April 1, 2021, proceedings, Part II. 559–564. Springer; 2021:43.
48 Dagdelen J, Trewartha A, Huo H, et al. COVIDScholar: an automated COVID-19 research aggregation and analysis platform. *PLoS One*. 2023;18:e0281147.
49 Allot A, Lee K, Chen Q, Luo L, Lu Z. LitSuggest: a web-based system for literature recommendation and curation using machine learning. *Nucleic Acids Res*. 2021;49:W352–W358.
50 Simon C, Davidsen K, Hansen C, Seymour E, Barnkob MB, Olsen LR. BioReader: a text mining tool for performing classification of biomedical literature. *BMC Bioinformatics*. 2019;19:57.
51 Leaman R, Gonzalez G. BANNER: an executable survey of advances in biomedical named entity recognition. In: Biocomputing 2008 *652-663.* World Scientific; 2008.
52 Wei C-H, Peng Y, Leaman R, et al. Assessing the state of the art in biomedical relation extraction: overview of the BioCreative V chemical-disease relation (CDR) task. *Database*. 2016;2016.
53 Wei CH, Allot A, Leaman R, Lu Z. PubTator central: automated concept annotation for biomedical full text articles. *Nucleic Acids Res*. 2019;47:W587–W593.
54 Wei CH, Kao HY, Lu Z. PubTator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Res*. 2013;41:W518–W522.

55 Smalheiser NR, Fragnito DP, Tirk EE. Anne O'Tate: value-added PubMed search engine for analysis and text mining. *PLoS One*. 2021;16:e0248335.

56 Tsuruoka Y, Miwa M, Hamamoto K, Tsujii J, Ananiadou S. Discovering and visualizing indirect associations between biomedical concepts. *Bioinformatics*. 2011;27:i111–i119.

57 Rindflesch TC, Kilicoglu H, Fiszman M, Rosemblat G, Shin D. *Semantic MEDLINE: an advanced information management application for biomedicine*. 312011;31. Information Services & Use; 2011:15–21.

58 Hope T, Portenoy J, Vasan K, et al. SciSight: combining faceted navigation and research group detection for COVID-19 exploratory scientific search. In: *Proceedings of the 2020 conference on empirical methods in natural language processing*. *System Demonstrations*. 2020:135–143.

59 Li PH, Chen TF, Yu JY, et al. pubmedKB: an interactive web server for exploring biomedical entity relations in the biomedical literature. *Nucleic Acids Res*. 2022;50(W1):W616–W622.

60 Henry S, McInnes BT. Literature based discovery: models, methods, and trends. *J Biomed Inform*. 2017;74:20–32.

61 Pyysalo S, Baker S, Ali I, et al. Lion LBD: a literature-based discovery system for cancer biology. *Bioinformatics*. 2019;35:1553–1561.

62 OpenAI. *ChatGPT: optimizing language models for dialogue*. 2022;2022.

63 Jin Q, Wang Z, Floudas C, Sun J, Lu Z. Matching patients to clinical trials with large language models. *arXiv*. 2023. https://doi.org/10.48550/arXiv.2307.15051.

64 Tian S, Jin Q, Yeganova L, et al. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. *Brief Bioinform*. 2024;25:bbad493.

65 Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. 2023;620:172–180.

66 Zhao WX, Zhou K, Li J, et al. A survey of large language models. *arXiv*. 2023.

67 Wang S, Scells H, Koopman B, Zuccon G. Can chatgpt write a good boolean query for systematic review literature search? *arXiv preprint arXiv*. 2023. https://doi.org/10.48550/arXiv.2302.03495.

68 Shaib C, Li M, Joseph S, Marshall I, Li JJ, Wallace B. *Summarizing, simplifying, and synthesizing medical evidence using GPT-3 (with varying success)*. Toronto, Canada: Association for Computational Linguistics; 2023:1387–1407.

69 Tang L, Sun Z, Idnay B, et al. Evaluating large language models on medical evidence summarization. *NPJ Digit Med*. 2023;6:158.

70 Peng Y, Rousseau JF, Shortliffe EH, Weng C. AI-generated text may have a role in evidence-based medicine. *Nat Med*. 2023;29:1593–1594.

71 Wadhwa S, DeYoung J, Nye B, Amir S, Wallace BC. Jointly extracting interventions, outcomes, and findings from RCT reports with LLMs. *arXiv*. 2023. https://doi.org/10.48550/arXiv.2305.03642.

72 Jin Q, Yang Y, Chen Q, Lu Z. GeneGPT: Augmenting large language models with domain tools for improved access to biomedical information. *arXiv preprint arXiv*. 2023. https://doi.org/10.48550/arXiv.2304.09667.

73 Jin Q, Leaman R, Lu Z. Retrieve, summarize, and verify: how will ChatGPT affect information seeking from the medical literature? *J Am Soc Nephrol*. 2023;34(8):1302–1304.

74 Gutiérrez BJ, McNeal N, Washington C, et al. Thinking about GPT-3 in-context learning for biomedical IE? Think again. Findings of the association for computational linguistics. *EMNLP*. 2022.

75 Coppola F, Faggioni L, Gabelloni M, et al. Human, all too human? An all-around appraisal of the "artificial intelligence revolution" in medical imaging. *Front Psychol*. 2021;12:710982.