

LitSense 2.0: AI-powered biomedical information retrieval with sentence and passage level knowledge discovery

Lana Yeganova[†], Won Kim[†], Shubo Tian[†], Donald C. Comeau, W. John Wilbur, Zhiyong Lu^{*}

Division of Intramural Research (DIR), National Library of Medicine (NLM), National Institutes of Health (NIH), MD 20894 Bethesda, United States

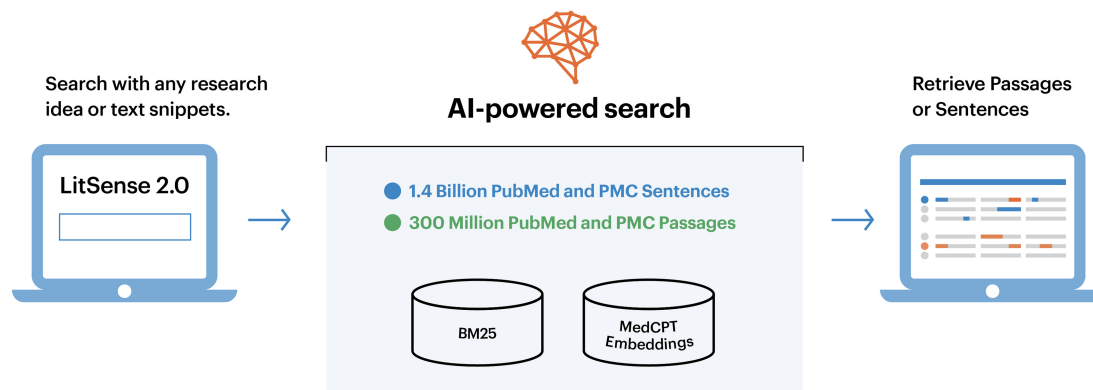
*To whom correspondence should be addressed. Email: Zhiyong.Lu@nih.gov

[†]The first two authors should be regarded as Joint First Authors.

Abstract

LitSense 2.0 (<https://www.ncbi.nlm.nih.gov/research/litsense2/>) is an advanced biomedical search system enhanced with dense vector semantic retrieval, designed for accessing literature on sentence and paragraph levels. It provides unified access to 38 million PubMed abstracts and 6.6 million full-length articles in the PubMed Central (PMC) Open Access subset, encompassing 1.4 billion sentences and ~300 million paragraphs, and is updated weekly. Compared to PubMed and PMC, the primary platforms for biomedical information search, LitSense offers cross-platform functionality by searching seamlessly across both PubMed and PMC and returning relevant results at a more granular level. Building on the success of the original LitSense launched in 2018, LitSense 2.0 introduces two major enhancements. The first is the addition of paragraph-level search: users can now choose to search either against sentences or against paragraphs. The second is improved retrieval accuracy via a state-of-the-art biomedical text encoder, ensuring more reliable identification of relevant results across the entire biomedical literature.

Graphical abstract



New! LitSense2.0 Adds Paragraph Search to Sentence Search!

Introduction

Traditional keyword-based search methods have long been the foundation of biomedical literature searches [1, 2]. Millions of users rely daily on platforms such as PubMed (<http://ncbi.nlm.nih.gov/pubmed>) and PubMed Central (<https://www.ncbi.nlm.nih.gov/pmc>) to access research articles that meet their information needs in biomedicine. While effective for straightforward searches, these methods struggle with complex information-seeking tasks involving longer, context-rich queries, such as sentences or passages/paragraphs. For example, a PubMed search with a sentence “There are only two fiber supplements approved by the Food and Drug Administration to claim a reduced risk of cardiovascular disease by lowering serum cholesterol: beta-glucan (oats and barley) and psyllium, both gel-forming fibers.” returns no results, despite relevant PubMed articles being found by LitSense 2.0.

Such nuanced searches are often essential for scientific endeavors like literature explorations, formulating hypotheses, or gathering specific evidence from the literature. The ability to search using extended text segments is especially critical for identifying claims or statements embedded within the full text of articles. Current search systems like PubMed and PMC are not equipped to accommodate this need. Compounding the challenge is the separation of PubMed and PMC into distinct systems—despite overlapping content, researchers must navigate each platform separately, adding unnecessary complexity to the search process and hindering comprehensive literature retrieval.

Additionally, keyword-based searches can miss the articles that are semantically relevant but have incomplete lexical overlap with the input query. Recent progress in information retrieval (IR) and deep learning has shown that dense

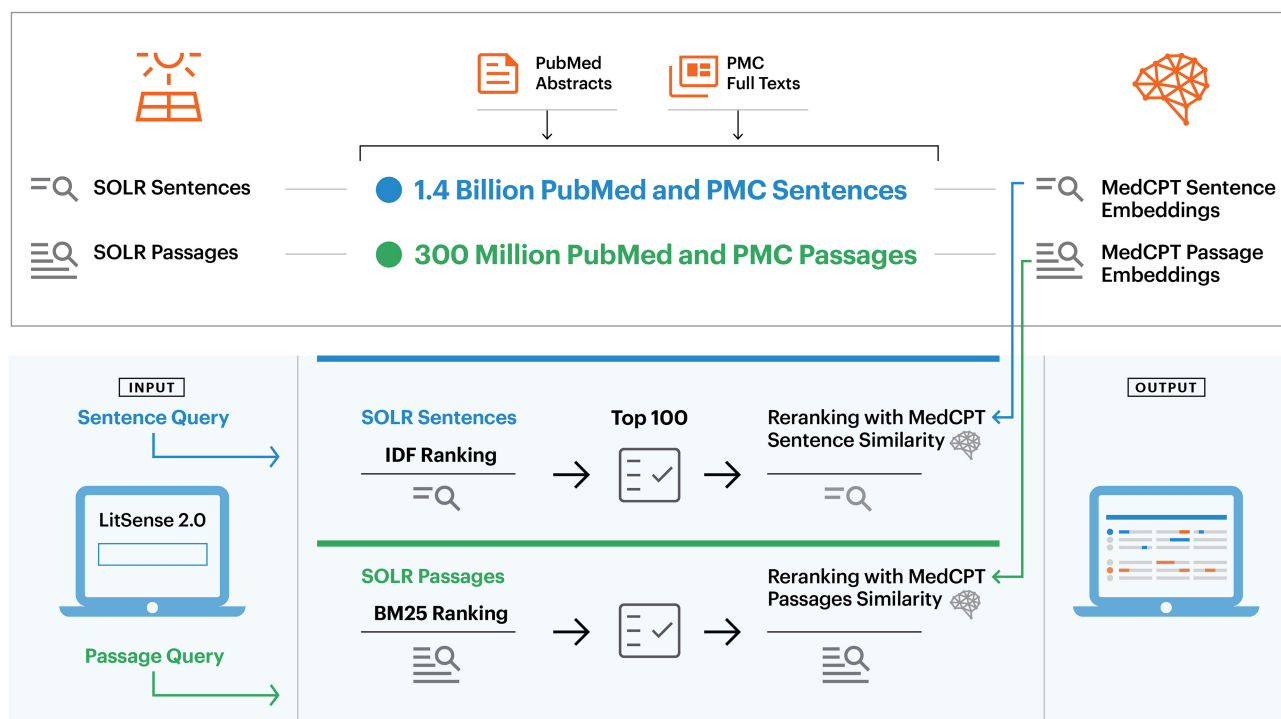


Figure 1. System overview. LitSense 2.0 offers two search options—sentence search and paragraph search. For both cases, we obtain PubMed and PMC Open Access documents and split the text into sentences and passages, resulting in ~1.4 billion sentences and ~300 million passages (A). MedCPT embeddings are computed for all sentences and paragraphs and stored in the database (B). For sentence search, Solr retrieves sentences using the (sum of IDF term weights) (C), while for paragraph search the BM25 function is used (D). The top ranked sentences/paragraphs are reranked using a linear combination of lexical and semantic scores (E, F). Results are displayed through the web interface (G).

retrievers, which encode and match queries and documents as dense vectors, can perform better semantic retrieval than traditional sparse retrievers such as BM25 [3–5].

In this work, we present LitSense 2.0, an advanced search system designed to overcome these challenges. By incorporating sentence- and paragraph-level query capabilities, LitSense 2.0 offers the ability to perform deep, context-aware searches, streamlining the retrieval of specific information from the vast biomedical literature. LitSense 2.0 unifies searches across PubMed and PMC, providing researchers with a streamlined solution to explore the biomedical knowledge landscape. Additionally, LitSense 2.0 achieves enhanced retrieval accuracy through its semantic search capabilities, powered by the integration of MedCPT [6].

MedCPT [6] is a transformer-based model obtained from PubMed BERT [7] and fine-tuned on a very large set of PubMed click data obtained from PubMed logs. MedCPT generates query and document embeddings that are shown to be highly effective for IR tasks. MedCPT embeddings have been evaluated on multiple datasets for sentence similarity, including BIOSSES [8] and MedSTS [9], and demonstrated favorable results. The MedCPT article encoder has also demonstrated state-of-the-art performance on article similarity tasks [6], outperforming models such as SPECTER [10] and SciNCL [11], specifically trained with article–article citation information. In this work, we use high-quality embeddings generated by MedCPT to perform semantic re-ranking of sentences and paragraphs, where the candidate items extracted using lexical methods are re-ranked using a linear com-

bination of lexical and dense vector similarity scores. Integrating MedCPT into LitSense 2.0 offers improved retrieval performance.

The review of existing semantic search services reveals that none could fully achieve the capabilities proposed by LitSense 2.0. One such system is Google Scholar (<https://scholar.google.com>)—a widely used academic search engine that allows searching with full sentences and longer passages. Unfortunately, it frequently fails when searched with paragraphs. The system cannot be restricted to PubMed and PMC, making it less effective for targeted biomedical literature retrieval. Additionally, it does not provide features such as biomedical entity recognition and highlighting. Another contender is Semantic Scholar (<https://www.semanticscholar.org>)—an AI-powered research tool for scientific literature. It often failed when searched with full sentences. When successful, it frequently returned the query sentence only. The system does not search in PubMed Central (PMC). Scopus (<https://www.scopus.com>) is a subscription-based literature database offering citation analysis and search capabilities across various disciplines. While it claims to allow searching with sentences and longer passages, it does not provide free access and lacks specialized search in PubMed and PMC. A systematic comparison is not feasible, as there is no API access to these systems, and the methodology is not public for Google and Scopus products. Europe PMC (<https://europepmc.org>) is the closest to us in functionality—as it searches across both PubMed and PMC. However, while it accepts sentences or longer passages as queries, it does not retrieve at the sentence or paragraph level, which is a unique and distinguishing feature of LitSense 2.0.

The screenshot displays the LitSense 2.0 interface. At the top, the LitSense² logo and NIH NLM branding are visible. The main search area features a search bar (A) containing the query: "The mechanisms by which fruit and vegetables exert their protective effects are not entirely clear but likely include". Below the search bar are two buttons: "SEARCH SENTENCES" and "SEARCH PASSAGES".

The results page shows a list of sections (B) on the left, including "TITLE AND ABSTRACT", "INTRODUCTION", "METHODS", "RESULTS", "DISCUSSION", and "CONCLUSION". Below this is a "PUBLICATION DATE" filter (C) with options for "LAST YEAR", "LAST 3 YEARS", and "LAST 5 YEARS".

The central column displays a search result (E) for "Functional foods and nutraceuticals in the primary prevention of cardiovascular diseases." by Alissa EM, Ferns GA • J Nutr Metab (2012). The introduction text is shown with highlighted terms: "antioxidant and anti-inflammatory effects", "fruits and vegetables", "LDL particles", "potassium", "CVD", "dietary intake of fruits and vegetables", "blood pressure", "carotenoids", "vitamin C", "fiber", "magnesium", and "potassium".

On the right, there are filters for bioconcepts (D) including "GENE", "DISEASE", "CHEMICAL", "MUTATION", "SPECIES", and "CELLLINE". A "DOWNLOAD" button (F) is also present.

Figure 2. LitSense 2.0 user interface. Users can enter sentence or passage queries into the search bar (A) and retrieve results based on sentence or passage search. On the results page, users can filter results by a section (B) or a publication date (C) and show/hide the highlights of bio-entities (D). The central column displays retrieved sentences or paragraphs (E). Each retrieved sentence/paragraph includes PubMed/PMC ID, title, authors, and section type in which the sentence/paragraph appeared. Clicking the title of the retrieval opens the whole document and shows the retrieved sentence/paragraph in context and highlighted. Retrieved results can be downloaded (F).

System overview

LitSense 2.0, illustrated in Fig. 1, is an online service for literature exploration that allows users to search with full sentences and paragraphs. The system provides unified access to a comprehensive collection of ~38 million PubMed records and 6.8 million full-text articles from the PMC Open Access Subset, encompassing over a 1.4 billion sentences and around 300 million paragraphs. Compared to PubMed only search, this greatly expands access to relevant information embedded within full-text articles [12].

LitSense 2.0 builds on its predecessor, LitSense [13], through addition of paragraph search and improved semantic search. In most cases, the passages are paragraphs, so in this work we use the terms paragraph and passage interchangeably. LitSense 2.0 employs a two-phase ranking system for both sentence and paragraph searches, similar to PubMed's Best Match algorithm [14] and other contemporary search systems that combine lexical and semantic retrieval methods [15–17]. The first phase uses lexical retrieval to identify candidate sentences/paragraphs, the second phase performs semantic re-ranking of the top scoring results retrieved during the first phase. Semantic re-ranking is achieved using the latest

generation of semantic search technologies powered by MedCPT [6], a state-of-the-art embedding model for biomedical IR. The MedCPT embeddings for all the sentences and paragraphs are precomputed, requiring only query embedding calculation at run time.

The interface of LitSense 2.0, as illustrated in Fig. 2, facilitates direct browsing of returned sentences or paragraphs in context and allows for further filtering by section type or publication date. Enhanced visualization features, such as entity highlighting, provide clear and insightful result presentations. Powered by PubTator 3.0 [18], entity highlighting recognizes genes, chemicals, species, diseases, mutations, and cell lines. LitSense 2.0 is free and open to all users without login requirement, it is available at <https://www.ncbi.nlm.nih.gov/research/litsense/>.

Sentence and passage indexing and search

LitSense 2.0 enables sentence and passage searches across all available sentences or passages in PubMed and PMC by treating them as units of information. To facilitate these search options, we index every sentence and passage in PubMed abstracts and PMC OA full-text documents. For both cases, we

Table 1. Performance comparison of three methods on 25 CDD query sentences and 50 *PLoS biology* query sentences

	NDCG <i>PLoS Biology</i> queries	CDD queries	MAP <i>PLoS biology</i> queries	CDD queries
LitSense 2.0	0.9551	0.9559	0.7985	0.6816
LitSense	0.9412	0.9373	0.7462	0.6233
IDF	0.9110	0.9255	0.5858	0.5131

NDCG and MAP scores are computed per query and averaged over all queries in each set using human annotations as ground truth.

obtain documents from the BioC repository [19], a database supporting an extensible mark-up language created to allow interoperability between biomedical text processing systems. Lengthy documents such as conference/workshop proceedings are excluded while keeping only individual research articles. Sentences and paragraphs are labeled by the type of section where they appear, and section types are normalized to one of the eight semantic categories, including Introduction, Methods, etc. [20]. This mapping allows filtering search results by sections. For more details, please refer to [Supplementary Section 1](#).

In the first phase of sentence search, efficient lexical retrieval of best-matching PubMed and PMC sentences uses an inverse document frequency (IDF) ranking function [21] without local weighting. We have two matching options. The first threshold—requiring 60% of query terms to be matched—aims to prioritize highly relevant candidate sentences. If no matches are found at this stricter level, the system relaxes the constraint to a 30% match threshold, which allows for broader coverage and increases the recall of potentially relevant results. Otherwise, the system does not produce any results. These thresholds are inherited from the original LitSense.

In the second phase, we used the MedCPT query encoder to compute a query sentence embedding. We then compute the MedCPT score, as a cosine similarity score between the query and each retrieved sentence. These top N sentences (currently N is set to 100) are then reranked by creating a linear combination of IDF and dense vector sentence similarity scores [22]. The relative weight assigned to the two components is determined empirically using the grid-search method, based on annotated data. Details are included in [Supplementary Section 2](#).

Passage search adopts a comparable two-phase process. In the first phase, paragraphs are scored using the BM25 ranking function [21], which outperforms the IDF scoring used for sentences by leveraging the additional context provided in paragraphs and accounting for local term frequency. If the query passage is too long, we limit the search to 50 highest weighted terms defined by TF_{max}IDF. Since paragraphs are significantly longer than sentences, the initial threshold is set at 30% match. If no results are retrieved, the threshold is reduced to 10%. Otherwise, the system does not produce any results. These thresholds were determined empirically. Similar to sentence re-ranking, the second phase re-ranks the top N (N is set to 100) paragraphs by combining BM25 scores with MedCPT paragraph similarity scores [6]. We use the MedCPT document encoder for both query and retrieved paragraphs.

Implementation tools and updates

We developed LitSense 2.0 using the Angular framework (<https://angular.io>) for the frontend and the Django frame-

work (<https://www.djangoproject.com>) for the backend. We update the data with new PubMed and PMC articles on a weekly basis. LitSense 2.0 supports the latest versions of web browsers (e.g. Chrome, Safari, Firefox, and Edge) and mobile devices.

Evaluation

Evaluating sentence search in LitSense 2.0

We compare sentence search performance of LitSense, which relies on sent2vec embeddings for computing semantic similarity, with MedCPT embeddings in LitSense 2.0. Sent2vec and MedCPT both generate vector representations of multi-word texts. Sent2vec averages word and n-gram vectors, while MedCPT, a BERT-based model, produces a contextualized embedding directly for the entire text.

For comparison, we use two manually curated datasets. In the first dataset, we use 25 queries sampled from the CDD data query sentences [23], for the second set, we use 50 *PLoS Biology* article titles as queries. For both sets of queries, we first retrieve the best matching 100 sentences with IDF and score these sentences with sent2vec and MedCPT (using query encoder for query sentences and document encoder for retrieved sentences). We compared three methods: (i) IDF without re-ranking; (ii) IDF reranked with sent2vec; and (iii) IDF reranked with MedCPT. We collect the top 10 best scoring candidates retrieved by each method. These sentences are then combined using a union function, shuffled, and manually evaluated, using annotation guidelines outlined in the annotation of the CDD dataset. Note that when we refer to “CDD queries,” we are only referring to the query sentences themselves. The retrieved sentences we manually annotated for this evaluation are distinct from those annotated in the original CDD dataset. Using these annotations, the normalized discounted cumulative gain (NDCG) evaluation and mean average precision (MAP) evaluation for each method (IDF, sent2vec, and MedCPT) ([https://en.wikipedia.org/wiki/Evaluation_measures_\(information_retrieval\)](https://en.wikipedia.org/wiki/Evaluation_measures_(information_retrieval))) is computed and presented in Table 1. The manually annotated datasets are available at <https://ftp.ncbi.nlm.nih.gov/pub/lu/LitSense2/>. For the MAP evaluation, we treat annotations of 4 and 5 as relevant, and the remainder as irrelevant. As seen in Table 1, replacing sent2vec with MedCPT results is an improvement and the improvement is statistically significant.

Since human annotations are costly, we tested ChatGPT as a surrogate annotator. Evaluation results on the manually annotated datasets of sentences suggested that the annotations produced by humans and ChatGPT are of comparable quality. Details of this computation are included in [Supplementary Section 3](#). We next used ChatGPT-generated annotations to compare sent2vec and MedCPT embeddings on a large scale. As query sources, we use 1000 *PLoS Biology* article titles, 1000 user queries from LitSense API logs,

and 1000 user queries from LitSense Web logs. Retrieved results are evaluated using NDCG and MAP, where NDCG and MAP scores are computed per query and averaged over all participating queries. As observed in Fig. 3, LitSense 2.0 significantly outperforms the original LitSense and the baseline IDF method. The differences between LitSense 2.0 and all other methods applied are statistically significant.

Evaluating paragraph search in LitSense 2.0

We evaluate the paragraph search performance on the RELISH article similarity dataset [24]. RELISH is the largest expert annotated article similarity dataset and contains annotations for 196 000 PubMed article abstract pairs, spanning 3200 PubMed articles utilized as queries. In RELISH, given a PubMed article as a query, a retrieved article is annotated as relevant, partially relevant, or not relevant. This dataset is suitable for evaluating paragraph similarity performance because PubMed article abstracts can be considered as paragraphs.

Table 2 presents the comparison of five methods for paragraph search: LitSense 2.0, LitSense, MedCPT, BM25, and PMRA [25]. We evaluate the ranking quality based on MAP and NDCG metrics at ranks of 5, 10, and 15 (@5, @10, @15) as well as all ranks (all), on the test subset of RELISH. We observe that LitSense 2.0 outperforms all other methods, based on all ranks, for both MAP and NGCD, and the difference is statistically significant.

We also explored the potential of using ChatGPT as a surrogate annotator for paragraph search. Several prompts were tested (see supplementary materials), but none produced results that matched the baseline performance.

Usage

LitSense 2.0 offers an IR framework for searching and accessing biomedical information on sentence and paragraph levels. It provides an intuitive interface, multiple useful filters, as well as allowing users to turn on entity highlighting. LitSense 2.0 streamlines users' access to information by locating the sentences and paragraphs of interest. Additionally, it offers access through its API and bulk download of the retrieved results. The API allows users to programmatically search the LitSense 2.0 collections in the same way as on the website, with results limited to the top 100 per query to reduce server load. API for searching sentences and passages, respectively, are <https://www.ncbi.nlm.nih.gov/research/litsense2-api/api/sentences/?query=query&rerank=true>; <https://www.ncbi.nlm.nih.gov/research/litsense2-api/api/passages/?query=query&rerank=true>

Numerous studies have harnessed LitSense sentence search capabilities for targeted literature search, hypothesis generation, and question answering. The usage analysis of LitSense shows the system to be of substantial benefit to the biomedical community. Since its launch in 2018, it has been widely adopted, serving over 250 000 unique users from >110 countries. The system has handled 9 million requests, with 55% originating from the web and 45% through the API. LitSense has been highlighted in workshops [26], text mining tool surveys [27], and benchmarking studies [28]. LitSense is cited as the search and text analysis tool of choice in publications covering a wide range of biomedical topics from cardiovascular informatics [29] to cholesterol synthesis path-

ways [30], to understanding of inter bacterial associations [31], and sulfate BioSynthesis [32], to name a few. LitSense provides a powerful approach to fact checking [27], and is a platform for discovering scientific evidence and conducting literature search. As an example, we present two case studies that rely on LitSense as a building block of a processing pipeline.

FoodAtlas: automated knowledge extraction of food and chemicals 2 from literature

[33] In this study the authors construct a knowledge graph applicable to the domain of raw food, ingredients, and chemicals. The LitSense API is used to retrieve sentences matching search phrases of the form "*food name* contains." The retrieved sentences were filtered to retain candidates for specific species and chemical entities, which were annotated by LitSense's Bio-Concept tagger. The model extracted over 230 thousand food-chemical composition relationship hypotheses, 46% of which have not been reported in any published database. Validation of the 443 hypotheses generated by the link prediction model resulted in 355 new food-chemical relationships.

BactInt: a domain driven transfer learning approach for extracting inter-bacterial associations from biomedical text

This study focuses on bacterial groups, and their associations within the human body ecosystem. Evidence supporting inter-bacterial associations is reported in biomedical literature, but automated and accurate extraction of such evidence is essential to keep pace with the ever-increasing volume of research. To address this problem, the authors create a processing pipeline, where first they extract candidate sentences containing such bacterial associations. These associations are frequently described with three commonly used keywords—"associate," "correlate," and "interact" and sentences containing bacterial names, and an association keyword are good candidates. Using queries of the format [*bacterial interaction_keyword bacteria2*], sentences mentioning the bacterial pairs were retrieved, resulting in the BactInt Sentence Corpus, a diverse and large sentence dataset, utilized for training a model.

Discussion and conclusion

LitSense has emerged as an important platform for discovering scientific evidence and conducting literature search, turning this tool into an integral part of biomedical information extraction tasks. With the introduction of LitSense 2.0, we have now incorporated paragraph-level search, expanding the platform's capabilities. To our knowledge, this is the only system that allows browsing of biomedical literature on a paragraph level.

Over the past five years, PubMed has expanded from ~29 million to 38 million article abstracts, while PubMed Central has grown from ~3 million to 6.6 million full-text articles. This growth has resulted in 1.4 billion sentences, up from half a billion in 2019, along with the new addition of ~300 million paragraphs. Sentence searches typically take 9 s, while paragraph searches require 17 s on average. This is due to query length, the vast search space, and shared compute resources. In unusual circumstances it could take significantly longer. We

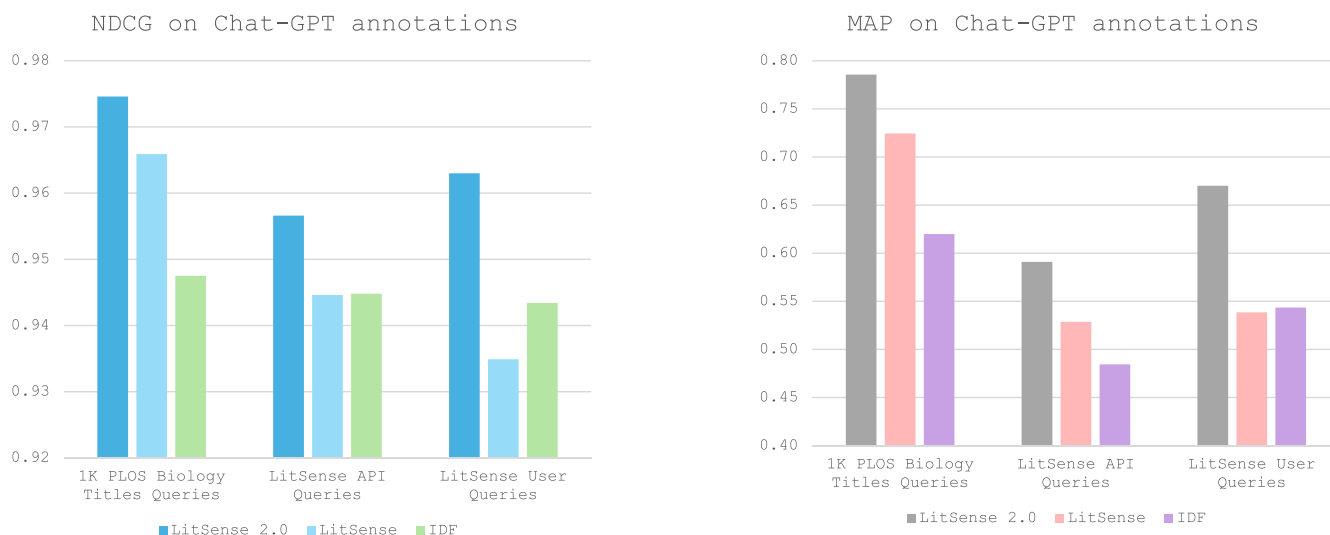


Figure 3. NDCG and MAP-based evaluation for three datasets—1K *PLoS Biology* queries, LitSense API queries, and LitSense Web queries, compared between three methods—LitSense 2.0, LitSense, and IDF, using ChatGPT annotations as ground truths. NDCG and MAP scores are computed per query and averaged over all queries.

Table 2. Performance comparison of five methods for paragraph search on the test portion of the RELISH dataset

Method	MAP				NDCG			
	@5	@10	@15	All	@5	@10	@15	All
LitSense 2.0	0.9515	0.9318	0.9164	0.8471	0.8206	0.7962	0.7885	0.9081
LitSense	0.9433	0.9206	0.9028	0.8295	0.7995	0.7722	0.7637	0.8971
MedCPT	0.9357	0.9147	0.8994	0.8330	0.7864	0.7658	0.7617	0.8959
BM25	0.9373	0.9150	0.8978	0.8230	0.7904	0.7630	0.7541	0.8934
PMRA	0.9308	0.9035	0.8843	0.8069	0.7700	0.7413	0.7330	0.8855

NDCG and MAP scores are computed per query and averaged over the participating queries.

are actively working on improving the performance and addressing this limitation.

For both sentence and paragraph searches, we discovered that the strategy of combining MedCPT score with IDF for sentences and BM25 for paragraphs offered a benefit. Individually, each method underperforms compared to their combination. This suggests that while MedCPT demonstrated improvement over previously implemented sent2vec, there is still room for improvement for semantic search. Advanced search capabilities of LitSense 2.0 can play a pivotal role in facilitating retrieval-augmented generation (RAG) workflows in the biomedical domain. By enabling users to retrieve contextually relevant scientific evidence at both sentence and paragraph levels, it ensures that RAG models are anchored in high-quality, domain-specific knowledge. Such an integration would support more accurate question answering and context-aware summarization, making LitSense 2.0 a valuable resource for researchers developing AI applications in biomedicine. Looking ahead, we aim to leverage the top-scoring retrieved results to generate user-friendly summaries and precise answers to user queries, further streamlining the research process and enhancing the utility of LitSense 2.0 in biomedical IR.

Unlike sentence-level evaluation, which leverages both curated datasets and ChatGPT-generated synthetic data, our paragraph-level evaluation currently relies on the RELISH dataset. To improve the robustness and diversity of paragraph-level evaluation, future work could ex-

plore incorporating structured datasets such as BioASQ (<https://www.bioasq.org>). Additionally, while MedCPT has demonstrated strong suitability for this task, a systematic comparison with other existing and upcoming embedding models—such as FlagEmbedding (<https://github.com/FlagOpen/FlagEmbedding>)—could provide the means of further improving our tool.

Acknowledgements

Author contributions: Lana Yeganova (Conceptualization [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Software [equal], Validation [equal], Writing—original draft [equal], Writing—review & editing [equal]), Won Kim (Conceptualization [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Resources [equal], Software [equal], Validation [equal], Writing—review & editing [equal]), Shubo Tian (Data curation [equal], Resources [equal], Software [equal], Visualization [equal], Writing—review & editing [equal]), Donald C. Comeau (Data curation [equal], Resources [equal], Software [equal], Visualization [equal], Writing—review & editing [equal]), John Wilbur (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Software [equal], Validation [equal], Writing—review & editing [equal]), Zhiyong Lu (Conceptualization [equal],

Investigation [equal], Resources [equal], Supervision [equal], Validation [equal], Writing—review & editing [equal])

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health. Funding to pay the Open Access publication charges for this article was provided by Intramural Research Program of the National Library of Medicine, National Institutes of Health.

Data availability

LitSense 2.0 is free and open to all users and there is no login requirement. LitSense 2.0 can be accessed at <https://www.ncbi.nlm.nih.gov/research/litsense2/>.

References

- Fiorini N, Leaman R, Lipman DJ *et al.* How user intelligence is improving PubMed. *Nat Biotechnol* 2018;36:937–45. <https://doi.org/10.1038/nbt.4267>
- Jin Q, Leaman R, Lu Z. PubMed and beyond: biomedical literature search in the age of artificial intelligence. *Ebiomedicine* 2024;100:104988. <https://doi.org/10.1016/j.ebiom.2024.104988>
- Karpukhin V, Oğuz B, Min S *et al.* Dense passage retrieval for open-domain question answering. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020, 1, 6769–81. <https://doi.org/10.18653/v1/2020.emnlp-main.550>
- Khattab O, Zaharia M. ColBERT: efficient and effective passage search via contextualized late interaction over BERT. *SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2020, 39–48. <https://doi.org/10.1145/3397271.3401075>
- Lin J, Nogueira R, Yates A. In: Hirst G (ed.), *Pretrained Transformers for Text Ranking: BERT and beyond; Learned Dense Representations for Ranking*. Cham, Switzerland, Springer Nature, 2022, 195–234.
- Jin Q, Kim W, Chen Q *et al.* MedCPT: contrastive pre-trained transformers with large-scale PubMed search logs for zero-shot biomedical information retrieval. *Bioinformatics* 2023;39:btad651. <https://doi.org/10.1093/bioinformatics/btad651>
- Gu Y, Tinn R, Cheng H *et al.* Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*. 2022, 3, 1–23. <https://doi.org/10.1145/3458754>
- Sogancioglu G, Öztürk H, Özgür A. BIOSSES: a semantic sentence similarity estimation system for the biomedical domain. *Bioinformatics* 2017;33:i49–58. <https://doi.org/10.1093/bioinformatics/btx238>
- Wang Y, Afzal N, Fu S *et al.* MedSTS: a resource for clinical semantic textual similarity. *Lang Resour Eval* 2020;54:57–72. <https://doi.org/10.1007/s10579-018-9431-1>
- Cohan A, Feldman S, Beltagy I *et al.* SPECTER: document-level representation learning using citation-informed transformers. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020, 2270–82. <https://doi.org/10.18653/v1/2020.acl-main.207>
- Ostendorff M, Rethmeier N, Augenstein I *et al.* Neighborhood contrastive learning for scientific document representations with citation embeddings. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 2022, 11670–88. <https://doi.org/10.18653/v1/2022.emnlp-main.802>
- Westergaard D, Stærfeldt H-H, Tønsberg C *et al.* A comprehensive and quantitative comparison of text-mining in 15 million fulltext articles versus their corresponding abstracts. *PLoS Comput Biol* 2018;14:e1005962. <https://doi.org/10.1371/journal.pcbi.1005962>
- Allot A, Chen Q, Kim S *et al.* LitSense: making sense of biomedical literature at sentence level. *Nucleic Acids Res* 2019;47:W594–9. <https://doi.org/10.1093/nar/gkz289>
- Fiorini N, Canese K, Starchenko G *et al.* Best match: new relevance search for PubMed. *PLoS Biol* 2018;16:e2005343. <https://doi.org/10.1371/journal.pbio.2005343>
- Gao L, Dai Z, Chen T *et al.* Complement lexical retrieval model with semantic residual embeddings. *Advances in Information Retrieval: 43rd European Conference on IR Research*. 2021, 12656, 146–60. https://doi.org/10.1007/978-3-030-72113-8_10
- Seo M, Lee J, Kwiatkowski T *et al.* Real-time open-domain question answering with dense-sparse phrase index. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019, 4430–41. <https://doi.org/10.18653/v1/P19-1436>
- Luan Y, Eisenstein J, Toutanova K *et al.* Sparse, dense, and attentional representations for text retrieval. *Transactions of the Association for Computational Linguistics*. 2021, 9, 329–45. https://doi.org/10.1162/tacl_a_00369
- Wei C-H, Allot A, Lai P-T *et al.* PubTator 3.0: an AI-powered literature resource for unlocking biomedical knowledge. *Nucleic Acids Res* 2024;52:W540–6. <https://doi.org/10.1093/nar/gkae235>
- Comeau D, Islamaj Dogan R, Ciccarese P *et al.* BioC: a minimalist approach to interoperability for biomedical text processing. *Database* 2013;2013:bat064. <https://doi.org/10.1093/database/bat064>
- Kafkas Ş, Pi X, Marinos N *et al.* Section level search functionality in Europe PMC. *J Biomed Sem* 2015;6:1–5. <https://doi.org/10.1186/s13326-015-0003-7>
- Spärck Jones K. A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *J Doc* 1972;28:11–21. <https://doi.org/10.1108/eb026526>
- Pagliardini M, Gupta P, Jaggi M. Unsupervised learning of sentence embeddings using compositional n-gram features. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2018, 11–21. <https://doi.org/10.18653/v1/N18-1049>
- Islamaj R, Wilbur WJ, Xie N *et al.* PubMed Text Similarity Model and its application to curation efforts in the Conserved Domain Database. *Database* 2019;baz064. <https://doi.org/10.1093/database/baz064>
- Brown P, Consortium R, Zhou Y. Large expert-curated database for benchmarking document similarity detection in biomedical literature search. *Database* 2019;baz085. <https://doi.org/10.1093/database/baz085>
- Lin J, Wilbur WJ. PubMed related articles: a probabilistic topic-based model for content similarity. *BMC Bioinformatics* 2007;8:423. <https://doi.org/10.1186/1471-2105-8-423>
- Wang X, Wang H, Ji H *et al.* New frontiers of scientific text mining: tasks, data, and tools. *KDD '22: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2022, 4832–3. <https://doi.org/10.1145/3534678.3542606>
- Vladika J, Matthes F. Scientific fact-checking: a survey of resources and approaches. *Findings of the Association for Computational Linguistics: ACL 2023*. 2023, 6215–30. <https://doi.org/10.18653/v1/2023.findings-acl.387>

28. Chen Q, Rankine A, Peng Y *et al.* Benchmarking effectiveness and efficiency of deep learning models for semantic textual similarity in the clinical domain: validation study. *JMIR Med Inform* 2021;9:e27386. <https://doi.org/10.2196/27386>
29. Caufield J, Sigdel D, Fu J *et al.* Cardiovascular informatics: building a bridge to data harmony. *Cardiovasc Res* 2022;118:732–45. <https://doi.org/10.1093/cvr/cvab067>
30. Ershov P, Kaluzhskiy L, Mezentsev Y *et al.* Enzymes in the cholesterol synthesis pathway: interactomics in the cancer context. *Biomedicines* 2021;9:895. <https://doi.org/10.3390/biomedicines9080895>
31. Baksi K, Pokhrel V, Pudavar A *et al.* BactInt: a domain driven transfer learning approach for extracting inter-bacterial associations from biomedical text. *Comput Biol Chem* 2024;109:108012. <https://doi.org/10.1016/j.compbiolchem.2023.108012>
32. Tumilovich A, Yablokov E, Mezentsev Y *et al.* The multienzyme complex nature of dehydroepiandrosterone sulfate biosynthesis. *Int J Mol Sci* 2024;25:2072. <https://doi.org/10.3390/ijms25042072>
33. Youn J, Li F, Simmons G *et al.* FoodAtlas: automated knowledge extraction of food and chemicals from literature. *Comput Biol Med* 2024;181:109072. <https://doi.org/10.1016/j.compbiomed.2024.109072>