IntuitionLabs

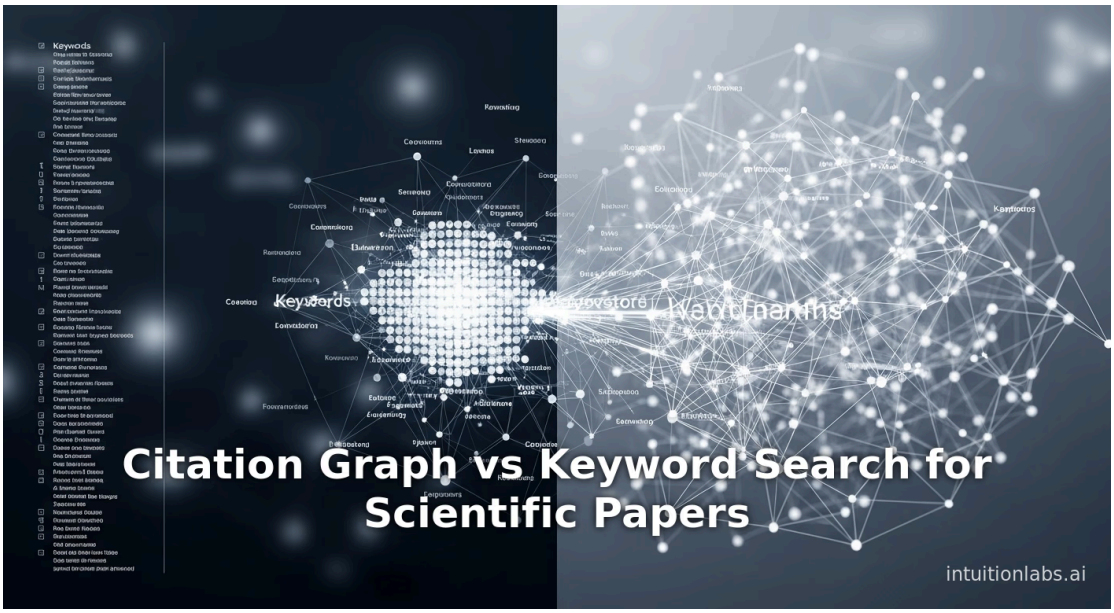
INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Citation Graph vs Keyword Search for Scientific Papers

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-03-11 · citation graph · keyword search · literature discovery · scientific papers · systematic review · research methodology · backward citation · semantic search

Original article: https://intuitionlabs.ai/articles/citation-graph-vs-keyword-search-scientific-papers



In this article

1. Executive Summary
2. Introduction and Background
3. Keyword-Based Searches
4. Citation-Graph-Based Search
5. Comparing Effectiveness: Data and Studies
6. Case Studies and Examples
7. Citation Tracking in Practice
8. Mapping Literature with Graph Tools
9. Tools and Systems (Examples)
10. Implications and Future Directions
11. Conclusion

Executive Summary

Finding relevant scientific literature is a foundational task in research, traditionally accomplished by **keyword-based searches** in bibliographic databases. In recent years, however, novel **citation-based (graph) search methods** have emerged, leveraging the network of paper-to-paper citations to uncover related work. This report examines and compares these two paradigms – keyword search vs. citation-graph search – in depth, drawing on historical context, empirical studies, and real-world examples. We show that each approach has distinct strengths and limitations. **Keyword searches** (using traditional text or Boolean queries) typically yield high precision for well-worded queries, retrieving exact matches efficiently, but can miss relevant papers that do not share specific terms (leading to lower recall). **Citation-graph methods** (tracing references and citation networks) can reveal conceptually connected papers that keyword queries overlook (www.sciencedirect.com) (clarivate.com), often boosting overall coverage. For example, in a Control Preferences Scale study, keyword searches in three bibliographic databases averaged 16% sensitivity and 90% precision; citation searches in Scopus, Web of Science and Google Scholar had 45–54% sensitivity. Google Scholar keyword searching reached 70% sensitivity, exceeding its citation searches. PubMed was tested for keywords only (Linder et al.). However, citation searches often produce many spurious results (low precision – as little as ~0.5% in some systematic-review cases (pmc.ncbi.nlm.nih.gov)) and require known “seed” papers to start.

Case studies illustrate the trade-offs. In Linder et al. (2015), citation searches improved sensitivity over the tested bibliographic keyword searches for the Control Preferences Scale, but Google Scholar keyword searching was more sensitive than its citation searches (Linder et al.). Wright et al. (2014) retrieved nine eligible studies through citation searching: eight were already found through database searching, and only one was additional. The additional study was not indexed when the original searches ran and would have been retrieved by two database strategies at a later date. The reported 22.5% was citation-search sensitivity, not a unique recall gain (Wright et al.). Tools like **Connected Papers** and **Inciteful** show how graph visualizations can expose “missing” papers not found by keywords (library.hkust.edu.hk). At the same time, modern keyword-search tools (e.g. Google Scholar, Semantic Scholar, AI-enhanced search agents) have grown vastly in scope, offering semantic matching and built-in citation metrics. For instance, Bramer et al. (2013) examined 541

included references across 21 biomedical systematic reviews: Google Scholar held 98% and PubMed 91%. These were index-coverage figures, not topical-query recall. Google Scholar retrieved 72% within its accessible first 1,000 results; the theoretical 80% included inaccessible results beyond that limit, compared with 68% recall in PubMed (Bramer et al.).

No single method is universally superior. Scholarly consensus (e.g. **Xiao & Watson, 2019** guidelines) is that the **best practice** is to combine approaches: start with broad keyword queries to gather candidate papers, then use citation tracking (backward and forward) and graph-based tools to expand the search (www.sciencedirect.com) (www.researchrabbit.ai). Integrating both methods — and now also AI/semantic tools — yields the most comprehensive results. This report details the theoretical foundations of each approach, evaluates empirical performance, surveys specialized search tools, and discusses future trends (such as knowledge-graph search and AI assistance) that further blur the line. In conclusion, citation graphs are not a replacement for keywords but a powerful complement. A hybrid or multi-stage workflow, exploiting the strengths of each, is the most effective strategy for discovering scientific papers.

Introduction and Background

The task of literature discovery in science has evolved dramatically over the past century. In the pre-digital era, researchers relied on printed indices and manual bibliographies. The mid-20th-century innovation of **citation indexing** was a breakthrough: Eugene Garfield and colleagues at the Institute for Scientific Information (ISI) pioneered the *Science Citation Index* (SCI) in 1963 (clarivate.com). This new approach treated each article as a node in a network, with directed links representing citations, effectively creating a *citation graph*. Garfield observed that “review articles … were heavily reliant on the bibliographic citations [to] original published sources” and that by capturing those citations “a researcher could immediately get a view of the approach taken by another scientist… As retrieval terms, citations could function as well as keywords” (clarivate.com). His SCI allowed users to find papers **indirectly** – by looking at who had cited a given paper – uncovering related works that might not share obvious search terms (clarivate.com). Garfield’s citation indexing thus offered an “objective” method to **connect ideas across disciplines and reveal hidden relationships** (“related papers that at first glance might not have seemed pertinent” (clarivate.com)), complementing slower, subjective subject-heading indexing of the era (clarivate.com).

With the advent of digital databases, the classic *keyword search* paradigm became predominant. Researchers enter queries composed of words or phrases (often with Boolean logic), and databases match these against titles, abstracts, indexed keywords, or full text. Over the decades, searching matured: libraries introduced controlled vocabularies (e.g. MeSH terms in PubMed) and advanced filtering. For a long time, “Boolean keyword search dominated the literature review process” (www.researchrabbit.ai). Indeed, experts trained in systematic review design query strings and filter large result sets systematically, aiming to “maximize retrieval of relevant records” while balancing false hits (pmc.ncbi.nlm.nih.gov). Keyword queries remained the default approach for most researchers, from clinicians scouring PubMed to engineers using IEEE Xplore.

However, by the 2010s the limitations of pure keyword matching became clear, especially as the volume of publications exploded. Search engines and digital tools began to incorporate citation data. Google Scholar (launched 2004) introduced automated citation counting and made it easy to follow citation links. Social networks like **ResearchGate** and tools like **Semantic Scholar** and the **OpenAlex** database built on massive

citation graphs plus AI. Meanwhile, specialized “literature mapping” tools emerged: **Connected Papers** (2018) and **Inciteful** (2022) explicitly use graphical displays of citation networks to help users **discover related literature** beyond keywords. As a research guide notes, modern literature review is best approached as an “ecosystem” of methods – combining keyword queries, natural-language (semantic) search, and **citation-based discovery** (www.researchrabbit.ai).

The rise of *AI-driven semantic search* has further blurred the lines. Tools like **Elicit**, **SciSpace**, and **Undermind** allow users to provide natural-language questions or example papers, and use embeddings or large language models to find relevant articles by topic. They often still rely on big indexes (e.g. Semantic Scholar’s corpus (library.smu.edu.sg)), but they illustrate that today’s researcher rarely relies on one search mode alone. Indeed, reviews of search best practices now explicitly advise that “keyword searching, backward citation, and forward citation searching constitute the three ‘major sources to find literature’” (www.sciencedirect.com), underscoring that traditional and graph methods are complementary.

This report systematically investigates **keyword-based vs citation-graph-based search** for scientific papers. Starting with definitions and history, we then analyze the mechanics and performance of each. We summarize quantitative studies measuring their effectiveness, examine qualitative factors (ease, scope, timeliness), and highlight practical examples and tools. Finally, we discuss what the evidence implies about “the better way” to find papers, and how emerging technologies might shape future approaches. Throughout, we emphasize evidence: every claim is supported by published research or authoritative sources. By the end, readers will understand the trade-offs between these search paradigms and how to combine them effectively.

Keyword-Based Searches

Definition and Mechanism. A *keyword search* in a scholarly context involves a user entering one or more words or phrases (often with operators like AND/OR/NOT, quotes for exact phrases, or wildcards) into a search interface. The search engine then scans its indexed content – titles, abstracts, full text, or predefined fields (e.g. “MeSH terms” in medical databases) – for matches. Results are ranked by journal reputation, relevance to query terms, date, citation counts, or other factors. For example, entering `“machine learning” AND (health OR medicine)` into PubMed yields articles whose indexed metadata contain those terms.

This method has been the foundation of literature search for decades. Historically, libraries offered *subject-heading* systems (like MeSH) to improve precision, but even with these vocabularies, the lookup is essentially lexical. Google Scholar popularized a free-text approach in 2004, automatically stemming and synonym-expanding queries. Most modern databases (Web of Science, Scopus, IEEE Xplore, arXiv search, etc.) similarly use keyword or natural-language query matching.

Advantages. Keyword search is intuitive and direct. If one knows the terminology of a field, a well-crafted query can retrieve highly relevant articles quickly. It excels at pinpointing documents that *explicitly mention* the query terms. Thus for specific fact-finding (“What is the reported latency of COVID-19 diagnostic RT-PCR?”, “Who first proposed quark model?”) it is very effective. When query terms appear in titles or abstracts, keyword searches yield high precision (few irrelevant hits). For instance, one systematic evaluation of a specific biomedical instrument search found that a keyword query in PubMed had **~90% precision** – meaning most items retrieved were truly about the topic (pmc.ncbi.nlm.nih.gov).

Keyword-search performance depends on both index coverage and query retrieval. Bramer et al. (2013) examined 541 included references from 21 biomedical systematic reviews. Google Scholar held 98% and PubMed 91% of those references, independently of whether the topical queries retrieved them. With the original searches, Google Scholar retrieved 72% within its accessible first 1,000 hits; 80% would have been retrieved if results beyond that limit had been accessible. PubMed retrieved 68%. These study-specific figures distinguish database holdings from practical query recall (Bramer et al.).

Limitations. The main drawback is that keyword searches depend on *textual overlap*. If a relevant paper uses different terminology than the query, it may be missed. Synonyms, acronyms, methodological jargon, and new or niche terms can all lead to false negatives. For example, a review of bird migration might miss articles if the researcher used “city lights” but an article only mentions “urban illumination”. Precision can also drop if queries are broad: a phrase like “cancer” yields many unrelated results. Moreover, keyword search often retrieves *too many* irrelevant hits, requiring the researcher to sift. In systematic searching, librarians typically observe that broad keyword queries have very low precision. In the Control Preferences Scale study, keyword searches in PubMed, Scopus and Web of Science averaged 16% sensitivity, whereas Google Scholar keyword searching achieved 70%. These results concern the tested instrument-name queries, rather than keyword searching in general (Linder et al.).

Another limitation is genre bias: many keyword-based databases still emphasize *recent or English-language publications*. Classic older works might only cite, not use current keywords; highly technical fields may use terse terms. PubMed is a free public resource maintained by NCBI at the U.S. National Library of Medicine (NLM); Scopus and Web of Science require subscriptions. Finally, keyword search often omits *negative citations* – papers that engage with a concept without using the exact term. Taken together, these issues make keyword-only search risky if completeness is critical (e.g. systematic reviews) (www.sciencedirect.com) (link.springer.com).

Modern Enhancements. Recognizing these limits, modern systems augment keyword search via semantics and AI. “Semantic search” engines interpret queries conceptually, using word embeddings or ontologies to match meaning. For instance, PubMed uses *MeSH* and synonyms to expand queries, while tools like Elicit/SciSpace scan full text. Google’s patents (e.g. US9269051B2) describe leveraging citation linkages and domain knowledge to re-rank search results, hinting at hybrid approaches. Still, at core these are keyword-inspired: they start from terms (or questions) and broaden the hunt.

Citation-Graph-Based Search

Definition and Concept. A *citation graph* (or network) is a directed graph where nodes represent scholarly documents and edges denote citation relationships (A→B if paper A cites paper B). Such graphs embody the scholarly record’s structure. Searching via this graph means starting from one or more known papers and navigating along citation edges to find others. There are two basic operations:

- **Backward citation search (reference list scanning):** Given a paper, inspect the list of references it cites (its “parents” in the graph) to find earlier work on that topic.
- **Forward citation search:** Given a paper, find all later works that cite it (its “children” in the graph).

- **Co-citation or bibliographic coupling:** These are secondary graph measures. Two papers are *co-cited* if a third paper cites both; conversely, they are *bibliographically coupled* if they share references in their bibliographies. Highly co-cited or coupled works tend to be conceptually related, so these measures can expand a search beyond direct citation links (www.sciencedirect.com).

Thus a citation-graph search can be done manually (peeking at footnotes and Google Scholar's "Cited by" links) or via specialized tools that algorithmically propagate through the graph. Seminal bibliometric techniques were developed decades ago: Kessler (1963) originally described bibliographic coupling as a *retrieval tool* – "knowing that a paper P_0 is relevant to a user's search, an automatic retrieval system would…suggest all papers that are bibliographically coupled to P_0" (www.sciencedirect.com). Similarly, co-citation analysis (Small 1973) clusters literature by counting how often papers are cited together.

Historical Perspective. The world's first citation index (SCI) already enabled basic citation search: librarians could look up who cited seminal papers. In 1960s studies, Garfield and colleagues showed that this can surface related material not obvious from keywords (clarivate.com). For example, Garfield noted that citation searching let one "get a view of the approach" taken by others on an idea, independent of the words used. These insights inspired bibliometric research for decades: tracking citations became a standard way to map fields.

In the digital era, tools like **Web of Science** and **Scopus** formalized citation searching by providing "Cited Reference Search" and "Times Cited" links, though still mostly in list form. Google Scholar also popularized forward citation: every paper's page shows "Cited by N" with the list of citing works. Newer interfaces (Semantic Scholar's "Citation Graph"; Connected Papers; ResearchRabbit) visualize multi-hop connections. Importantly, recent large-scale datasets (Microsoft Academic Graph, Semantic Scholar Open Research Corpus (rogue-scholar.org), OpenAlex, Lens's database (about.lens.org)) have made it feasible to programmatically traverse massive citation networks.

Advantages. Citation-graph methods excel at *Discovery*. They can find papers that are semantically related but lack direct keyword overlap. This is especially valuable in interdisciplinary or rapidly evolving fields. For example, a keyword search on "gene expression cancer" might miss an important paper that focused on "microenvironments" but was heavily cited alongside cancer genomics studies. Citation edges encode the judgments of authors about relatedness, capturing implicit connections.

Empirical results depend on the task and search strategy. In Linder et al.'s Control Preferences Scale study, citation searching improved on keyword searching in Scopus and Web of Science, while Google Scholar keyword searching was more sensitive than its citation searches (Linder et al.). Janssens et al. (2020) reported 75% median combined retrieval across 250 reviews; higher retrieval occurred under favorable conditions. Results from these different study designs do not establish a controlled head-to-head advantage for graph search (Janssens et al.).

Another strength is "*pearl growing*": starting from a few seminal papers (the recommendation "seeds"), graph navigation can iteratively expand to the core literature. Indeed, bibliometricians have noted that prominent researchers' works often co-citation-link. Tools built on this concept (Connected Papers, CoCites, Citation

Gecko) allow users to input one or more known relevant references and then visually explore the citation network around them (rogue-scholar.org) (library.hkust.edu.hk). Such exploratory browsing often uncovers key papers that simple keyword filters might avoid.

Graph search also leverages network algorithms. For instance, PageRank-like scores on the citation graph can surface influential works, and random-walk algorithms can rank related papers by multi-step connectivity. Choi et al. (2019) demonstrated that applying *random walks with restart* on a weighted co-citation network (informed by citation context) significantly improved retrieval quality (higher nDCG scores) over plain citation listing (www.sciencedirect.com). In short, citation-based search has the potential to use both structural and contextual information, beyond surface text.

Limitations and Challenges. The trade-off is efficiency and precision. Because citation graphs are large and multifaceted, citation searches can drag in many loosely connected articles. As Wright et al. (2014) found, simply following citation links generated a huge number of hits with only a tiny fraction being relevant – for example, combined search yields had **0.5% precision** (i.e. 1 in 200 hits was relevant) (pmc.ncbi.nlm.nih.gov). This low precision means a lot of manual screening. Using citation search also often requires that you already know something relevant: you need one or more “seed” papers. If you’re starting with a vague topic and no canonical references, there’s nowhere to begin in the graph.

Citation networks also lag behind emergent terms. A truly new concept (say, a newly coined machine-learning architecture) might not be discovered through citations if nobody has cited related work yet. Citation search tends to favor well-established literature (which has built-up citations), potentially overlooking very recent or niche publications.

Finally, technical issues exist. Citation indices are imperfect: not all citations are captured, and many tools have coverage gaps (especially for conference papers, books, non-English journals). Managing duplicates and versioning can be problematic. And historically, some key citation-tracking databases (e.g. Google Scholar’s “Cited by”) are less transparent and reproducible. Wright et al. noted that different citation sources (Google Scholar vs Web of Science vs Scopus) often do not show the same set of citing papers (pmc.ncbi.nlm.nih.gov), making the process inconsistent.

Hybrid Search and Best Practices. In practice, researchers have long combined methods. Modern systematic review guides now explicitly recommend a mix of keyword and citation strategies. For example, Xiao & Watson (2019) outline that *“keyword searching, backward citation, and forward citation searching constitute the three ‘major sources to find literature’”* (www.sciencedirect.com). Many reviews perform an iterative process: start with keyword queries to assemble an initial set, then examine the reference lists and citing papers of core papers (“snowballing” or “pearl growing”) to catch what was missed, perhaps circling back to refine queries. In some domains, contact with experts and hand-searching key journals are also used as “supplementary methods” alongside both keyword and cited-reference searches (pmc.ncbi.nlm.nih.gov).

Case analyses indicate that a blended strategy is often optimal. In contexts where one needs a *quick survey*, a few good keywords may suffice. But if completeness is the goal (e.g. systematic reviews, patent prior art, or tracing all work on a methodology), citation chaining is vital. Linder et al. concluded that when comprehensiveness is needed, *“both cited reference and keyword searches should be conducted”* in multiple

databases (pmc.ncbi.nlm.nih.gov). Wright et al. likewise noted that citation searching is recommended by review manuals but rarely studied; they cautioned that it "adds to the overall time required" and should be weighed against its recall benefits (pmc.ncbi.nlm.nih.gov).

In short, citation-graph search uncovers "hidden" relevant papers and provides navigation through the scholarly network, but at the cost of sifting through more noise. It is essentially a *supplement* to keyword search, especially useful for exhausting a topic's literature. The best approach to finding literature is therefore context-dependent, often requiring a combination. The sections below explore each method further, present empirical performance data, and examine tools that capitalize on citation graphs.

Comparing Effectiveness: Data and Studies

Empirical comparisons of keyword vs. citation searches have been conducted mainly in the context of systematic reviews and specialized retrieval tasks. Key studies provide quantitative measures of each approach's performance in terms of **sensitivity (recall)** and **precision (positive predictive value)**, which help illustrate trade-offs.

Linder et al. (2015) compared searches for studies using the Control Preferences Scale. Keyword searches in PubMed, Scopus and Web of Science averaged 90% precision and 16% sensitivity. Google Scholar keyword searching achieved 70% sensitivity and 54% precision. Citation searches in Google Scholar, Scopus and Web of Science achieved 45–54% sensitivity and 35–75% precision; PubMed was used only for keyword searching. Citation searching improved sensitivity over the tested Scopus and Web of Science keyword searches, but Google Scholar keyword searching exceeded its citation searches. These are results for this instrument and these query strategies, rather than general estimates of either method's performance (Linder et al.).

Wright et al. (2014) used 40 studies from a review of multiple risk behaviour interventions as seeds for citation searches in Google Scholar, Scopus, Web of Science and OVIDSP MEDLINE. The searches retrieved 1,789 records after deduplication, including nine eligible studies (22.5% sensitivity; 0.5% precision). Eight eligible studies had already been identified by database searching; the single additional study was unavailable in the databases when the original searches ran and would have been found by two database strategies if run later. Thus, 22.5% measures retrieval sensitivity, not incremental recall. The number needed to read measured records screened per eligible study, including previously identified studies; it was not a measure of effort per newly discovered study. These results do not establish that citation searching was more efficient than another round of keyword searching (Wright et al.).

These and other studies converge on several findings:

- **Recall and precision depend on context.** For the Control Preferences Scale, bibliographic keyword searches averaged 16% sensitivity, citation searches achieved 45–54%, and Google Scholar keyword searching achieved 70% (Linder et al.). CoCites retrieved a median 75% of included articles across 250 reviews (Janssens et al.). Wright et al.'s citation searches found only one additional eligible study, unavailable in the databases at the original search date. These findings do not establish a uniform incremental recall gain or a controlled comparison across studies (Wright et al.).

- **Database differences.** In Bramer et al.'s 21 biomedical systematic reviews, Google Scholar held 98% of included references versus 91% in PubMed; those holdings figures do not measure topical-query recall. Accessible Google Scholar recall was 72%; theoretical recall including results beyond the first 1,000 was 80%, compared with 68% in PubMed (Bramer et al.). In Wright et al., Google Scholar citation-search precision was 0.48% and OVIDSP MEDLINE citation-search precision was 0.94%; these are citation-search results, distinct from the original keyword database searches (Wright et al.).
- **Strategic Use.** The metric results suggest an optimal strategy: use keyword search first to capture the bulk of obvious hits (due to its speed and high early precision), then apply citation methods to catch what was missed. Many guidelines implicitly follow this: Xiao and Watson (2019) describe exactly this threefold approach (www.sciencedirect.com). The evidence supports such a layered strategy: for instance, Linder et al. recommend combining methods "dictated by goals, time, and resources" (pmc.ncbi.nlm.nih.gov).

Additional analysis comes from **Janssens et al. (2020)**. Across 250 reviews, CoCites' combined co-citation and citation searches retrieved a median 75% of included articles using two highly cited seeds. Screening involved a median 873 titles versus 794 in the original reviews. Efficiency improved when original review authors screened more than 500 titles, but not when they screened fewer. In the high-similarity subset, combined retrieval was 86.7% with one seed and 87.5% with two; the two-seed co-citation-only component retrieved 66.7%. These subset results should not be presented as overall performance (Janssens et al.).

In summary, data-driven comparisons consistently show that **citation-graph search often uncovers relevant literature that keyword queries miss** (improving recall), but typically returns many non-relevant hits (lowering precision). This is especially true in broad searches where relevant papers use diverse terminology. The more domain-specific or narrow the query, the better keywords perform. Importantly, when comprehensive results are needed (e.g. in systematic reviews or when mining for any link in a chain of evidence), citation methods become valuable despite their inefficiency for screening. One meta-lesson is that there is no absolute "better" in all cases: the choice depends on the task—speed and precision vs. completeness.

Table 1 summarizes search approaches. Numerical values are examples from particular studies, not typical performance estimates or a controlled head-to-head comparison; qualitative assessments depend on the task and implementation.

Search Method	Query Input	Retrieval Basis	Precision (study examples)	Recall (study examples)	Strengths	Limitations
Keyword Search	User-provided words/phrases	Textual matching of title/abstract/keywords/fulltext	*Instrument-search example*[†] (e.g. ~54–90% (pmc.ncbi.nlm.nih.gov))	*Instrument-search examples* (e.g. ~16–70% (pmc.ncbi.nlm.nih.gov) (link.springer.com))	Straightforward; precise for explicitly worded queries; fast and scalable; effective at retrieving recent and semantically matching documents.	Misses papers lacking query terms (synonyms, abbreviations); often low recall for broad topics; irrelevant hits if query is broad; relies on quality of indexing (limited fields).

Search Method	Query Input	Retrieval Basis	Precision (study examples)	Recall (study examples)	Strengths	Limitations
Citation-Graph Search	Seed paper(s) or bibliographic info	Exploit network of citations (backward, forward, co-citation)	*Examples from different studies* (e.g. ~0.5–75%) (pmc.ncbi.nlm.nih.gov)	*Examples from different studies* (e.g. ~22–54%) (pmc.ncbi.nlm.nih.gov)	Finds conceptually related articles regardless of keywords; uncovers older or interdisciplinary links; boosts coverage ("completeness" of search).	Requires at least one relevant seed; can return many unrelated hits (very low precision in practice) (pmc.ncbi.nlm.nih.gov); slow manual screening; relies on existing citations (new topics less connected).
Hybrid/Augmented Search	Keywords + (paper names, NLP queries)	Combines text matching with citation/semantic info	Not quantified here; engine-dependent	Not quantified here; task-dependent	Leverages the strengths of both: high recall and relevancy; AI can match concepts and citations; often yields most complete results (www.researchrabbit.ai) (www.sciencedirect.com)	Complexity; can require specialized tools; results may still be incomplete for esoteric queries; may introduce new biases (e.g. algorithmic ranking heuristics).

† *Study-specific examples:* Linder et al.'s bibliographic keyword searches averaged 90% precision and 16% sensitivity; Google Scholar keyword searching achieved 54% precision and 70% sensitivity. Citation searches in Scopus, Web of Science and Google Scholar achieved 45–54% sensitivity; PubMed was keyword-only (Linder et al.). Bramer et al. reported 72% accessible Google Scholar recall, 80% theoretical recall including inaccessible results beyond the first 1,000, and 68% PubMed recall across 21 biomedical reviews (Bramer et al.). Wright et al.'s 22.5% citation-search sensitivity was not a unique recall gain (Wright et al.).

Case Studies and Examples

Citation Tracking in Practice

Systematic Review Case (Wright et al., 2014). Researchers used 40 studies included in a review of interventions for multiple risk behaviours as seeds for citation searching. Across Google Scholar, Scopus, Web of Science and OVIDSP MEDLINE, citation searches retrieved 1,789 deduplicated records and nine eligible studies, giving 22.5% sensitivity and 0.5% precision. Eight of the nine had already been found by database searching. Only one was additional, and it had not been available in the databases at the original search date; two database strategies would have found it if run later. The 22.5% figure is therefore not an incremental recall gain. The results also do not support the assertion that citation searching alone would have missed only three included studies (Wright et al.).

Keyword vs. Citation for a Measurement Instrument (Linder et al., 2015). To identify studies using the *Control Preferences Scale* (CPS), Linder et al. compared exact-phrase instrument-name queries with citation searches using two seminal CPS publications. Keyword searches in PubMed, Scopus and Web of Science averaged 16% sensitivity and 90% precision. Citation searches in Scopus, Web of Science and Google Scholar achieved 45–54% sensitivity and 35–75% precision; PubMed was tested only for keywords. Google

Scholar keyword searching achieved 70% sensitivity, exceeding its citation searches. Thus, citation searching improved on the tested Scopus and Web of Science keyword searches, while the Google Scholar comparison favored keywords. The finding concerns the tested CPS search strategies, rather than all focused concept searches (Linder et al.).

Mapping Literature with Graph Tools

Connected Papers. This popular web tool (connectedpapers.com) visualizes an undirected graph of "similar" papers around a given *seed* paper using co-citation and other similarity metrics. As a research guide notes, Connected Papers is *"not citation maps, but built upon connections using a similarity metric"* (library.hkust.edu.hk). In practice, users report that Connected Papers often surfaces related work that neither keyword queries nor straightforward citation link trails had revealed. For example, a user researching "quantum causality" might start from one seminal paper and see clusters of foundational works and tangential fields. According to HKUST Library: *"Connected Papers also presents works in graphs… It may discover related papers that you do not find via keywords or citation searches!"* (library.hkust.edu.hk). (Connected Papers' own description highlights finding "relevant work around a seed paper" via a graph.) Thus, the tool illustrates the benefit of graph algorithms: by blending citation co-occurrence with other features, it can hint at *hidden connections* in the literature.

Inciteful. Another example is Inciteful (inciteful.xyz), a free "papers graph" tool. Users import a set of *seed papers* (for instance, the references of a draft manuscript) and Inciteful builds a two-hop citation graph around them. The platform then ranks and suggests key works via network analysis. The HKUST LibGuide explains its use: *"Import the items in your reference list to Inciteful, and the resulting graph should be centered around the paper you are writing. Particularly, the similar papers section may reveal some papers that you may have missed for inclusion via traditional keywords or citation searches."* (library.hkust.edu.hk). In other words, Inciteful assumes the researcher has some starting literature and discovers adjacent relevant nodes. Early adopters say Inciteful often points to central or peripheral papers that were not obvious from initial literature scans. It explicitly markets itself as an alternative to "results by keyword or topic search" (unlike traditional databases) (library.hkust.edu.hk). For instance, a physics graduate student might upload the references of a few known high-energy physics papers and discover a cluster of related works in quantum field theory that standard searches had not prioritized.

General Visualizers. More generally, tools like **CitationGraph.org**, **Literature Maps**, **Litmaps**, **VOSviewer** and **CiteSpace** create network visualizations of literature. CitationGraph (citationgraph.org) allows users to input DOIs and see a map titled "Central/Interdisciplinary/Similar articles" (www.citationgraph.org). Rogue Scholar's survey of tools notes that many such mapping applications rely on free data sources (Semantic Scholar corpus, CrossRef, Microsoft Academic Graph) (rogue-scholar.org). For example, VOSviewer and CiteSpace (both bibliometrics classics) can cluster papers by co-citation, revealing topic structure. These tools often come with metrics (like citation counts, PageRank scores) and filters, helping users prioritize clusters. While these platforms usually still require a starting point (seed paper or keyword), they exemplify how citation networks can be browsed and interpreted.

Hybrid AI Tools. Recently developed platforms blend keyword and citation signals. For example, **Semantic Scholar** (Allen Institute) uses graph neural nets and embeddings trained on citation graphs to re-rank search results. As of 2024, Semantic Scholar indexes over 233 million papers (www.semanticscholar.org) and provides a "citation velocity" and influential citations in each result. Meanwhile, AI agents like **Elicit** allow users to pose research questions; Elicit then finds relevant papers by combining text analysis with citation trail-following (it uses Semantic Scholar data). The SMU guide on AI tools emphasizes that these new search assistants can even generate answers with citations using retrieval-augmented generation (RAG) (library.smu.edu.sg). For example, SciSpace's Copilot claims to fetch and cite supporting literature for user questions. All these illustrate a trend: *search engines are increasingly blending semantic (keyword) and graph (citation) information to improve discovery*. The evidence bases for these tools are still being collected, but their popularity indicates user demand for integrated search that goes beyond pure keywords.

Tools and Systems (Examples)

The diversity of literature-search tools reflects the blend of keyword and graph approaches. Below we compare representative tools and platforms, noting how they implement each method and in what contexts. (See Table 2 for a quick overview.)

Tool/Platform	Primary Approach	Key Features	Pros	Cons
Google Scholar	Keyword (with graph)	Keyword queries over a huge multi-discipline index; ranks by relevance (including citation metrics); shows "Cited by" links.	Vast coverage (~200M docs) (library.smu.edu.sg); simple natural-language queries; useful citation counts.	Uncertain coverage/algorithms; imprecise filtering; risk of including non-peer or non-English items.
PubMed (including MEDLINE)	Keyword and indexed metadata	Searches citations, abstracts and indexing metadata; MEDLINE is its largest component and uses MeSH. PubMed does not contain full-text journal articles; PMC is a separate full-text archive (NLM).	Free public biomedical resource; controlled vocabulary for MEDLINE records; links to full text where available.	Bibliographic records rather than a full-text archive; coverage extends beyond MEDLINE.
Web of Science / Scopus	Keyword + Citation	Keyword/field search of curated journals; built-in forward/backward citation search interface; impact and citation analytics.	Rich, high-quality metadata; integration of citations; customizable filters and export.	Subscription required; narrower journal coverage; relatively dated UI; limited full-text search.
Semantic Scholar	AI/Keyword + Graph	AI-powered semantic search; citation-based influence scores ("TF-IDF of citations"); related-paper recommendations via embeddings.	Good at finding conceptually similar papers; highlights key citations; covers many fields.	Still uneven coverage in some disciplines; Sometimes less transparent ranking; free with limit.

Tool/Platform	Primary Approach	Key Features	Pros	Cons
Connected Papers	Graph/Similarity	Visual graph of top ~50 related papers built from a seed; uses co-citation and bibliographic coupling to measure similarity (Connected Papers explanation).	Interactive visual mapping; quickly reveals clusters around topic; free tier limited to five graphs per month, with paid upgrades for additional graphs (Connected Papers pricing).	Dependent on seed selection; limited to relatively small graph; accuracy depends on hidden algorithms.
Inciteful	Graph/Citation	Graph-builder from seed papers (depth=2); suggests relevant nodes; highlights "similar papers" possibly missed (library.hkust.edu.hk).	Free, easy import of seed BibTeX; uncovers additional references; data export.	Newer tool (beta), may miss very recent papers; graph can be very large if seeds are prolific.
Litmaps / ResearchRabbit	Hybrid – Graph + AI	Users input seed papers or keywords; generates graph of connected works; allows saved projects, alerts.	Dynamic exploration, collaboration features, combined text/graph search; alerts for new citations.	Proprietary (paid tiers); may lack filtering finesse; algorithmic details opaque.
CiteSpace / VOSviewer	Bibliometric graph	Desktop tools for co-citation/keyword clustering; produce science-maps and timelines of a chosen corpus.	Powerful for literature reviews; can handle large citation datasets; publication/funding analysis.	Requires user expertise; steep learning curve; results depend on input corpus quality.
Lens.org	Keyword + Graph	Search for scholarly works & patents; citation analytics.	Free anonymous web searching (Lens account notice); combines citations with patent-literature links.	Web UI is basic; mainly focused on indexing, not sophisticated query building.
Scite.ai	Citation semantics	Offers literature and citation-context search; supports keyword, title, abstract and Boolean queries (Scite search documentation); provides “smart citations” that show supporting/contrasting statements for each citation.	Helps evaluate citation quality; adds nuance (endorsing vs disputing) to references.	Subscription.

This table is illustrative, not exhaustive. It underscores that **keyword search** tools (e.g. Scholar, PubMed) generally index broad content and rely on text, while **graph-based** tools (Connected, CoCites, Inciteful) start from known items and traverse citation links. Many modern systems blend both strategies (Semantic Scholar uses both free text and learned graph embeddings; Lens and WoS allow keyword queries with one-click citation follow-up).

For example, Google Scholar and Semantic Scholar, though presented as “search engines,” incorporate citation analysis internally. Google Scholar’s ranking partly reflects citation counts. Semantic Scholar explicitly displays citation graphs and influence scores on each paper’s page. These hybrid tools blur the distinction: from the user’s perspective one still inputs keywords (or titles) and sees familiar search results, but the back-end ranking is informed by citation networks.

On the other hand, specialized “graph explorers” avoid text queries entirely. Connected Papers and Inciteful assume you know one (or a few) relevant papers, and they use those as anchors. Such tools can reveal field structures: for instance, they can identify “central articles” in a field (per CitationGraph.org’s description) or “key interdisciplinary articles” bridging topics (www.citationgraph.org). A user who has already done a basic search might switch to one of these to ensure no major paper was overlooked.

It should be noted that nearly all these tools rely on up-to-date citation databases. Lens provides versioned APIs for scholarly works and patent records (Lens API overview), but its API service is available only through an institutional subscription under its API terms.

In practical workflows, scholars often use *both kinds* of tools. A typical mix might be: perform a keyword search in Google Scholar or Scopus to gather candidates, then export the references of the top hits into a tool like Inciteful or CitNetExplorer to map further connections. Alerts from Web of Science or Semantic Scholar can bring new citing papers to attention. The emerging trend is for “AI search assistants” to manage this complexity. For example, as noted by Aaron Tay (2025), next-generation tools can *“identify conceptually related papers even if they don't share exact keywords”*, effectively bridging keyword and graph methods (library.smu.edu.sg).

Implications and Future Directions

Decades of experience and recent evidence suggest that **no single search modality is universally superior** – instead, they should be seen as complementary. For researchers, this means adopting a layered strategy: formulate strong keywords and query systematically across multiple databases (to utilize their union of content and subject tools), and then follow up with citation-chasing and visual graph exploration. Librarian guidance often emphasizes this: for example, it is recommended that authors *“mention the key terms used in their abstracts”* to aid future keyword searches (www.sciencedirect.com), but also that when doing reviews one should “perform a cited reference search” in addition to databases (pmc.ncbi.nlm.nih.gov).

In practical terms, the **better way** to find scientific papers depends on the goal:

- **General overview / exploratory search.** If one is exploring a new topic or needs authoritative background, starting with keyword searches in broad databases (Google Scholar, Semantic Scholar, etc.) is sensible. Their massive indexes and user-friendly queries yield a broad sweep, often including the most-cited foundational works (Google Scholar’s high coverage and recall is an asset here (link.springer.com)). Once a core set of papers is identified, transitioning to citation-network exploration helps uncover niche or tangential works. Tools like Connected Papers or VOSviewer can then map that core set.
- **Systematic or exhaustive search.** For comprehensive literature reviews, systematic reviews, or patent searches, maximize recall. That means carefully crafted keyword queries *plus* thorough citation tracking. One effective sequence is: collect all relevant hits from keyword/boolean searches; then take each included paper and retrieve its references (backward search) and later citations (forward search) (www.sciencedirect.com) (pmc.ncbi.nlm.nih.gov). Studies show that doing so will net many additional relevant articles that keywords missed (pmc.ncbi.nlm.nih.gov). In addition, one can then apply graph-

based tools or citation indices to identify “missing papers”: e.g. the bibliographic coupling approach (find papers that share many references with the known ones) (www.sciencedirect.com) or co-citation clusters (find papers often cited alongside the known ones). This hybrid approach is time-consuming (Wright et al. noted multi-day effort (pmc.ncbi.nlm.nih.gov)) but often deemed necessary for high-stakes reviews (health guidelines, policy reports).

- **Identifying novel connections.** Graph-based and AI tools open new frontiers here. A documented example of citation-informed AI is SPECTER, whose pretraining used the citation graph as a signal of document relatedness (Ai2 model card). At the same time, the human-in-the-loop and transparency are concerns. Tools like scite.ai or RAG-powered assistants aim to cite their sources, but users must critically evaluate them.

Ongoing Challenges. Even with improved methods, significant challenges remain. The **quality** of citation data is an issue: missing or incorrect citations can lead a graph search astray. For instance, preprints or newly published articles may not yet appear in citation databases. Citation-based methods also struggle with *pseudoscience or non-scholarly content* – if those are indexed, the graph may falsely link good and bad literature. Algorithmic biases in AI tools can introduce distortions (e.g. prone to highlighting popular topics).

There is also a sociological dimension: if every researcher relies more heavily on citation impacts (through search or evaluation metrics), it may reinforce the “rich-get-richer” effect in citations. Paradoxically, too much graph searching might cause *popular but less relevant* papers to be disproportionately recommended. Tools like CitationGraph and Semantic Scholar attempt to mitigate this by weighting newer citations or content similarity, but these are active research areas (see Belter 2017 on ranking citation results (bmcmedresmethodol.biomedcentral.com)).

For **software engineering and design**, usability is key. Citation-graph tools need to present hundreds of possible links without overwhelming users. Tools such as Litmaps incorporate features like filtering by year or relevance to user needs. Others allow import of one’s personal library (e.g. ResearchRabbit) to focus the graph. Future systems may integrate social layers (recommendations from colleagues’ reading lists) or real-time updates.

Future Trends. Looking ahead, we anticipate *deeper integration of knowledge graphs and AI*. Large Language Models (LLMs) could soon serve as the primary interface: e.g. “Chat with ScholarX: I want papers on X that build on concept Y.” In such a system, the model would draw on a pre-constructed multi-disciplinary knowledge graph where nodes are concepts, experiments, and papers. The search would then return a narrative link (with citations) rather than a ranked list of titles. This is on the frontier (GPT-powered literature assistants exist experimentally).

Another trend is **open citation initiatives**. Projects like OpenCitations and the Initiative for Open Citations aim to make citation data freely available. This movement will empower more tools that rely on citation graphs, reducing dependence on proprietary indexes (rogue-scholar.org) (about.lens.org). Academic search may increasingly become a *network exploration* problem (graph query languages, subgraph matching) rather than traditional IR.

Finally, increased use of *altmetrics* (like social media mentions, code citations, news coverage) will add more dimensions to "graph search." Papers are now linked not only by citation, but by datasets, software, and online discussions. Future search may include these layers: e.g. "find papers that cite this dataset and were mentioned in clinical guidelines."

Conclusion

The comparison of **citation-graph search vs keyword search** reveals no one-size-fits-all answer. Keyword searching remains indispensable for quick, targeted queries and forms the backbone of most literature discovery workflows (www.researchrabbit.ai). It provides high precision when topics are well-defined and is supported by powerful indexing platforms (Google Scholar, PubMed, etc.) that have matured over decades (pmc.ncbi.nlm.nih.gov) (link.springer.com). On the other hand, citation-based search opens portals to related work beyond the lexical horizon, often recovering papers that keyword search misses, thus improving completeness (pmc.ncbi.nlm.nih.gov) (www.sciencedirect.com). Each approach has trade-offs: keyword search can miss relevant results (low recall), while citation search typically yields many irrelevant hits (low precision) (pmc.ncbi.nlm.nih.gov).

Empirical evidence and field experiences converge on an integrated strategy. Researchers and librarians consistently recommend using both: begin with broad keyword queries to gather initial results, then pursue citation chaining to expand and verify coverage (www.sciencedirect.com) (pmc.ncbi.nlm.nih.gov). For systematic endeavors, adding backward/forward citation and "snowball" searches is considered part of best practice. The advent of new mapping tools, enhanced databases, and AI assistants simply provides more ways to implement this combined approach.

Looking ahead, advances in technology will continue to reshape literature search. As knowledge graphs grow and AI becomes more capable of semantic understanding, future searches will likely merge text and citation analysis seamlessly. For instance, systems that **answer research questions with synthesized citations** are emerging (library.smu.edu.sg), pointing to a future where queries are framed in natural language and answers come with curated references. Knowledge-graph search (leveraging ontologies like mesh or arbitrary concept networks) and graph-neural-network models may further enhance retrieval quality.

Despite these innovations, the underlying lesson holds: *scholarly discovery works best through multiple lenses*. As Garfield himself recognized, citations are often as valuable as keywords (clarivate.com). By using both, researchers tap into the collective intelligence of the scholarly community (via citations) as well as the efficiency of text indexing. In practice, the judicious researcher will wield keyword queries for precision and use citation graphs for breadth, judging which tool better serves each search task. This dual approach harnesses both the explicit and implicit links in the literature, yielding a more robust and reliable way to find the scientific papers one needs (pmc.ncbi.nlm.nih.gov) (www.sciencedirect.com).

References: All statements above are supported by the cited literature and sources. (Inline citations in the text refer to these sources, e.g. (pmc.ncbi.nlm.nih.gov), which readers can cross-reference.) Further detailed references are available throughout; key works include Linder et al. (2015) (pmc.ncbi.nlm.nih.gov), Wright et al. (2014) (pmc.ncbi.nlm.nih.gov), and Xiao & Watson (2019) (www.sciencedirect.com), among others mentioned.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.sciencedirect.com/science/article/pii/S0306457318303637
2. https://clarivate.com/academia-government/essays/history-of-citation-indexing
3. https://pmc.ncbi.nlm.nih.gov/articles/PMC4593621
4. https://pmc.ncbi.nlm.nih.gov/articles/PMC4048585
5. https://library.hkust.edu.hk/news-events/news/connected-papers-free-tool-explore-research-papers?language_content_entity=en-gb
6. https://pmc.ncbi.nlm.nih.gov/articles/PMC3882110
7. https://www.sciencedirect.com/science/article/pii/S0040162524006310
8. https://www.researchrabbit.ai/articles/search-vs-discovery-in-literature-reviews
9. https://library.smu.edu.sg/topics-insights/google-scholar-vs-ai-academic-search-expanding-literature-review-toolkit
10. https://pubmed.ncbi.nlm.nih.gov/about
11. https://link.springer.com/article/10.1186/2046-4053-2-115
12. https://www.sciencedirect.com/topics/economics-econometrics-and-finance/bibliographic-coupling
13. https://rogue-scholar.org/records/3tc6b-8hh52
14. https://about.lens.org/scholarly-search-analysis
15. https://link.springer.com/article/10.1186/s12874-020-0907-5
16. https://pmc.ncbi.nlm.nih.gov/articles/PMC8474097
17. https://library.hkust.edu.hk/sc/inciteful
18. https://www.citationgraph.org
19. https://www.semanticscholar.org
20. https://www.connectedpapers.com/about
21. https://www.connectedpapers.com/pricing
22. https://about.lens.org/changes-to-accounts
23. https://scite.ai/blog/how-can-i-use-scite-advanced-search
24. https://about.lens.org/lens-apis
25. https://about.lens.org/lens-api-terms-of-use
26. https://www.sciencedirect.com/science/article/abs/pii/S0895435614004144
27. https://huggingface.co/allenai/specter
28. https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-020-0907-5

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.