

## Hybrid Search Is the New Baseline: Better Lexical Search Is the Next Advantage

Vector search, also known as semantic search, has transformed enterprise search in the past few years by retrieving information according to meaning rather than relying exclusively on exact words. It is particularly effective when users and documents express the same concept using different vocabulary.

However, pure vector search has important limitations. It can struggle with exact or distinctive information such as product names, identifiers, error codes, named entities and technical terminology. It can also return semantically similar but factually different results; for example, confusing words that behave similarly from a syntactic point of view but are not synonyms semantically, like “football” and “rugby”; or mixing products or entities that appear in comparable contexts but have different meanings, like “Paris” and “Madrid”.

**Hybrid search addresses this problem by combining vector retrieval with lexical retrieval.** The vector component contributes semantic understanding, while the lexical component contributes exact terminology and word-level evidence. Public research has reported approximately a **20% improvement for German retrieval** (1) and a **35% average improvement in multilingual retrieval** when lexical BM25 search is added to vector search (2).

This complementary approach has made hybrid search the current industry standard, with support available from major search platforms such as Elasticsearch and Pinecone.

### The Quality of the Lexical Component Matters

---

Adding BM25 is not the end of the lexical-search decision. BM25 can only rank the terms produced by its text analyzer. A basic analyzer fails to connect different grammatical forms of the same word or expose useful concepts hidden inside compounds. Some analyzers use techniques like stemming, that solve some of these cases but also introduce noise (in more than 30% of cases in languages like German (3)).

Bitext addresses this with language-specific linguistic analysis, including:

- dictionary-based lemmatization
- controlled compound analysis
- language-specific normalization
- preservation of useful original forms

In a controlled MIRACL evaluation for German covering **712 queries and approximately 15.9 million passages**, Bitext Linguistic Analysis achieved the highest scores among benchmarked against three Elasticsearch configurations. Compared with the basic Elasticsearch configuration, Bitext improved:

- **Precision@10 by 35.0%**

- **Recall@10 by 39.6%**
- **F1@10 by 36.1%**

Bitext also outperformed Elasticsearch's built-in German analyzer, improving Precision@10 by **8.7%**, Recall@10 by **13.6%**, and F1@10 by **9.6%** (4).

## What These Improvements Look Like in Practice

---

Probably the fastest way to visualize the benefits of linguistic analysis is compounds and other word formation phenomena present in German, Arabic, Korean, Dutch, Swedish... German frequently combines several concepts into a single word. When a user expresses those concepts separately but the relevant document uses a compound, or vice versa, a basic lexical analyzer will fail to make the connection. Bitext's linguistic analysis identifies the meaningful components inside compounds and represents them in forms that BM25 can match to equivalent non-compounds and rank. We show examples coming from MIRACL dataset, one of the top sources for multilingual search evaluation.

Consider the following query:

*Wie lange dauert ein Studium in Deutschland, um Lehrer zu werden?*  
*How long does it take to complete a degree in Germany to become a teacher?*

The query uses **Studium** (*degree course or university studies*) and **Lehrer** (*teacher*) as separate words. The relevant passage expresses the same concepts mainly through compounds, including **Lehrerausbildung** (*teacher training*), **Lehrkräfte** (*teaching staff*), **Lehramtsanwärter** (*trainee teachers*) and **Studienseminare** (*teacher-training institutes*).

Bitext exposed the useful components inside those words and ranked the relevant passage **first**. Elasticsearch's decompounding configuration ranked it **45th**, while the minimal and built-in German configurations did not return it within the top 100.

The same effect also appears in the hybrid-search examples. For the query:

*Wie oft wird Abfall bzw. Müll in Deutschland abgeholt?*  
*How often is waste or garbage collected in Germany?*

the relevant passage is titled **Müllabfuhr** (*waste collection*) and contains related compounds such as **Abfuhrplan** (*collection schedule*), **Abfuhrbezirk** (*collection district*), **Hausmülltonne** (*household waste bin*) and **Müllfahrzeug** (*garbage truck*). The Bitext-enhanced hybrid system moved this passage from rank 13 to rank 2.

The passage was not included in the benchmark's relevance judgments, even though the result directly answered the query and was moved into a prominent position.

These examples show why compound-aware lexical analysis matters: it exposes concepts that would otherwise remain hidden inside individual tokens, allowing relevant passages to enter the top results and become available to rerankers, RAG systems and AI agents.

## The Takeaway

---

The future of search is not vector **or** lexical retrieval. It is vector plus lexical retrieval.

But the effectiveness of hybrid search depends partly on the quality of the lexical branch. Standard BM25 provides valuable exact-match evidence; linguistically enhanced BM25 provides stronger evidence by recognizing real word forms, meaningful compound components and language-specific relationships.

Hybrid search is now the baseline and linguistics-based lexical search is the next opportunity to improve it.

(1) Köpf, M., et al. *Hybrid Search for German Technical Documents*. Journal of Advances in Information Technology, Vol. 16, No. 6, 2025

(2) Longpre, S., Roberts, N., Lee, Y., et al. (2023). *Multilingual Information Retrieval with Hybrid Search*. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL Industry Track), 388–399.

(3) *Stemming kills AI Accuracy: Why German Search Needs Lemmatization* at <https://www.bitext.com/blog/stemming-kills-ai-accuracy-why-german-search-needs-lemmatization/>

(4) *Bitext Linguistic Analyzer beats every Elasticsearch Configuration on German Search Quality* at <https://www.bitext.com/blog/bitext-linguistic-analyzer-beats-every-elasticsearch-configuration-on-german-search-quality/>