

IDC MarketScape: Worldwide General-Purpose Knowledge Discovery Software 2025 Vendor Assessment

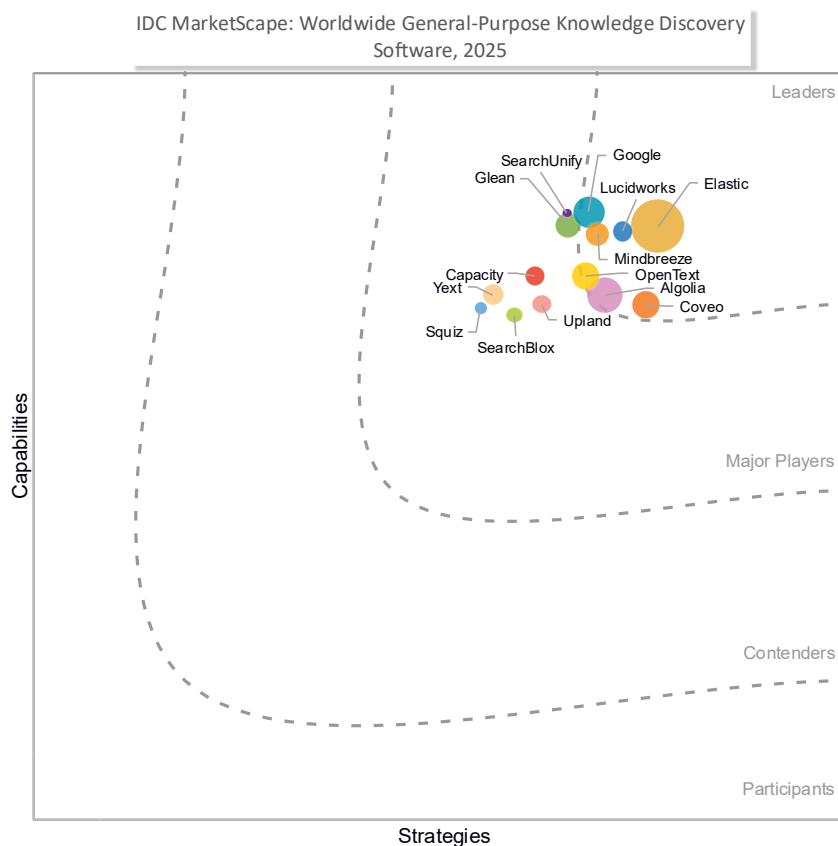
Amy Machado

THIS EXCERPT FEATURES ELASTIC AS A LEADER

IDC MARKETScape FIGURE

FIGURE 1

IDC MarketScape Worldwide General-Purpose Knowledge Discovery Software Vendor Assessment



Source: IDC, 2025

See the Appendix for detailed methodology, market definition, and scoring criteria.

ABOUT THIS EXCERPT

The content for this excerpt was taken directly from IDC MarketScape: Worldwide General-Purpose Knowledge Discovery Software 2025 Vendor Assessment (Doc # US3011225).

IDC OPINION

Knowledge discovery technology has grown into a critical capability for turning vast amounts of data into actionable strategic insights, moving beyond basic retrieval. To meet this evolution, knowledge discovery vendors are expanding their portfolios to enable a metadata-driven intelligence layer.

While retrieval remains a core capability, the addition of context, intent understanding, and pattern recognition can address the different levels of customer maturity and deliver enhanced experiences for both internal and external use cases. Modern knowledge discovery solutions integrate artificial intelligence (AI), natural language processing (NLP), and semantic analysis to act as an enterprise intelligence engine that integrates both structured and unstructured data from enterprise applications and content repositories.

Connecting these diverse data sources for knowledge discovery requires robust role-based access controls (RBACs), and vendors are integrating these often-dynamic guardrails into their knowledge platforms. In fact, when asked what knowledge discovery and sharing capabilities are most important to their business, 54.0% of companies tell IDC that they want dynamic knowledge and content that adapts to the experience or different stages of the workflows (source: IDC's *Knowledge Management Solutions Survey*, July 2024). Safeguarding sensitive data shouldn't be an inhibitor. Yes, it helps customers maintain compliance with privacy and regulatory standards, but it also refines and optimizes search access and results for end users.

During our vendor briefings and customer conversations, several consistent themes emerged. Knowledge discovery is a core component of enterprise workflows, whether they are ad hoc or agentic. Achieving trusted outcomes demands not only relevance and contextual understanding but also transparency. Today's IT buyers increasingly reject a "black box" system in favor of ones that deliver explainability, auditability, and the ability to fine-tune results. And after running quite a few generative AI (GenAI) pilots, a lot of companies have realized that large language models (LLMs) alone cannot unlock unstructured data and content for successful outcomes.

Despite significant progress in automation and artificial intelligence, the human element remains fundamental to successful enterprise search. Nearly half of companies tell IDC they require direct feedback from users to improve discovery outcomes (source: IDC's *Knowledge Management Solutions Survey*, July 2024). User feedback/ratings, real-time analytics, and input from business users allow organizations to identify content gaps, improve relevance, and fine-tune AI-driven search models to better support actual workflows. A dynamic, collaborative feedback loop boosts information retrieval and creates a culture where continuous improvement is driven by user experience (UX) rather than assumptions or static automation.

Market performance reinforces the importance of knowledge discovery. The global search and knowledge discovery market grew by more than 28.0% in 2024, surpassing \$6.0 billion. More companies are building AI applications and connecting them to search and knowledge discovery indexes and grounding techniques. Vendors used 2024 to launch better grounding and tuning capabilities and greater visibility into the search pipeline to help drive their customers' AI development. Forecasts project a 31.2% CAGR through to 2029, more than tripling the market size to over \$23.0 billion as the market expands what is possible, including proactive and conversational experiences.

IDC MARKETSCOPE VENDOR INCLUSION CRITERIA

Participating vendors must meet the following criteria:

- The offering is commercially available for use as a single product family or a suite of services and has been purchased by customers for at least one year. IDC will also consider and include new product features and capabilities introduced through 1H25 as part of the vendor strategy evaluation. In addition, IDC will consider these features as part of its capabilities evaluation if there is sufficient customer adoption and use for IDC to properly evaluate and if these features are generally available by the time this document is published.
- The product offers knowledge discovery software that organizations can utilize, customize, deploy, and/or include in their enterprise applications.
- The product supports a variety of internal- and external-facing use cases.
- The product has achieved at least \$1.0 million in revenue in CY24.
- The product is offered and available on a worldwide basis.
- The product is all or most of the vendor's own intellectual property (IP).

ADVICE FOR TECHNOLOGY BUYERS

- Consider your AI strategy because knowledge discovery is also an investment in AI. When measuring ROI for AI use cases, include nine key parameters: revenue generation, customer experience (CX), employee experience, productivity/efficiency, innovation, sustainability, time to market, security/trust, and business resilience.
- Go beyond basic technical mapping to identify the business-critical applications and content repositories necessary for current success — as well as future use case expansion. The search solution should offer robust integration capabilities to access and centralize disparate data sources into a single, comprehensive view for the end user.
- Prioritize data privacy and security, especially when dealing with sensitive or regulated information. Look for an admin experience that includes automated, robust permission controls and audit trails to ensure that only authorized users can access specific data. Our interviews confirm that regardless of whether the search architecture is integrated or federated, access controls remain the nonnegotiable top priority for security-conscious organizations.
- Determine what accuracy means to your business and to different use cases. As technology moves beyond keyword searches, accuracy has many nuances and layers. Besides what is "factual," what is accurate can also be swayed by relevance, recency, context, memory, user intent, explainability, and personalization.
- Plan for the agentic future. Knowledge discovery companies are playing pivotal roles in preparing businesses for the coming agentic AI era, where autonomous agents will plan, reason, and act on behalf of users across enterprise systems. Search is the foundational layer for these workflows, as it combines large-scale data retrieval, observability, and GenAI capabilities into agent builder frameworks.
- Treat LLMs as language generators and a capability, not a core search engine or platform. LLMs are designed for advanced language generation, not inherent fact retrieval. They must be grounded in real-world, verified data through techniques like retrieval-augmented generation (RAG) and prompt engineering. Always pair LLM outputs with real-time retrieval strategies from a curated knowledge base. Implement operational safeguards to ensure accuracy that strictly adheres to organizational values and ethical standards.

VENDOR SUMMARY PROFILES

This section briefly explains IDC's key observations resulting in a vendor's position in the IDC MarketScape. While every vendor is evaluated against each of the criteria

outlined in the Appendix, the description here provides a summary of each vendor's strengths and challenges.

Elastic

After a thorough evaluation of Elastic's strategies and capabilities, IDC has positioned the company in the Leaders category of this 2025 IDC MarketScape for worldwide general-purpose knowledge discovery software.

Elastic offers an open source platform designed for deployment flexibility across multicloud, private cloud, and even air-gapped environments with fully managed options available. It integrates seamlessly with both open source and commercial LLMs, while also giving customers the ability to bring their own models and embeddings. At the core is the Elasticsearch platform — available as self-managed, cloud hosted, and serverless — that combines advanced search, analytics, and AI capabilities. It enables queries and analysis across structured and unstructured data, plus dense and sparse vectors, to deliver content to generative models. This positions Elastic to support the evolution of search needs, from traditional queries and use cases to conversational and agent-driven experiences.

Quick facts about Elastic are as follows:

- **Year founded:** 2012
- **Company headquarters:** San Francisco, California
- **Approximate number of global employees (as of 1H25):** 3,000+
- **Primary dedicated KM software product evaluated:** Elasticsearch
- **Geographic market focus and traction:** North America, Europe, APAC (excluding China), and Australia
- **Vertical industry focus and traction:** Industry agnostic, used in a wide variety of industries and use cases including financial services, government, professional services, ecommerce, and high tech
- **Customer focus and traction:** Small, midsize, large, and enterprise businesses
- **Most common integrated applications:** SharePoint Online, Google Drive/Workspace, Salesforce, Atlassian Jira/Confluence, and Slack
- **LLM approach:** LLM-agnostic, BYOM approach
- **Deployment options:** Managed SaaS (AWS, GCP, and Azure), private cloud, on premises, and hybrid/Kubernetes
- **Pricing model:** For cloud deployments, Elastic uses a consumption model. Hosted cloud pricing is based on a resource usage model, and serverless cloud offers solution-specific pricing.

- **Noteworthy capabilities or strategy element:** For model choice and greater relevance based on the use case, Elasticsearch supports a wide range of NLP models, including dense and sparse vector models and cross encoder reranker. The pretrained sparse vector model offers out-of-the-box (OOTB) relevance without fine-tuning. Dense vectors are highly dimensional and use neural embeddings to represent semantic meaning, translating text, images, or other data and mapping to content with similar meaning.

Strengths

- **Platform flexibility:** Elastic integrates hybrid, vector, and semantic search with support for first- and third-party LLMs, enabling organizations to build relevant, real-time GenAI experiences with explainability and control. Customers reported that they experienced fast time to value with existing data and content, without vendor lock-in.
- **Developer-centric and open architecture:** Elastic gives developers the platform and tools—including APIs, SDKs, and customization options—to build and customize the search pipeline by tailoring each aspect, from ranking to semantic pipelines and connectors. Elastic core features simplify development. For example, Semantic Text automates embedding and chunking for semantic search, for better integration and index management. Elastic Search Query Language (ES|QL) enhances data processing and performance with concise syntax, making data exploration easier for even business users. Simplifying the developer's experience can improve GTM and cost reductions, with a search experience tailored to a specific business need.
- **Unified platform for search, observability, and security:** Elastic unifies search, observability, and security within a single, integrated platform, providing real-time visibility. From monitoring user behavior and system performance to securing data and content, Elastic provides a deep view of critical operations. By consolidating capabilities that typically require multiple point solutions, Elastic can reduce tool sprawl and operational overhead, while still scaling to address complex business needs.

Challenges

- **Presence of hyperscalers in the market:** Elastic operates in the same space as hyperscalers. These companies promote tight integrations and bundles with their broader cloud ecosystems. This strategy creates pricing and procurement pressure for Elastic, especially when negotiating with enterprises that are already customers of these hyperscalers and are looking to simplify their purchases and reduce costs.

- **Overcoming the developer-heavy reputation:** While Elasticsearch has long been known for its appeal to developers, this strength can also create a perception of being overly complex. Elastic is working to change this by adding more out-of-the-box features, more prepackaged solutions, and simplified setup experiences to boost time-to-value for nondevelopers.

Consider Elastic When

Consider Elastic when seeking a search and knowledge discovery engine known for stability and reliability, especially for teams with skilled in-house developers. Elastic has introduced more prepackaged features for enterprise and ecommerce search, making implementation easier for even organizations without dedicated development resources. In addition, Elastic's full platform provides a versatile foundation for expanding beyond search into robust security and observability solutions, as they expand into the broader AI ecosystem.

APPENDIX

Reading an IDC MarketScape Graph

For the purposes of this analysis, IDC divided potential key measures for success into two primary categories: capabilities and strategies.

Positioning on the y-axis reflects the vendor's current capabilities and menu of services and how well aligned the vendor is to customer needs. The capabilities category focuses on the capabilities of the company and product today, here and now. Under this category, IDC analysts will look at how well a vendor is building/delivering capabilities that enable it to execute its chosen strategy in the market.

Positioning on the x-axis, or strategies axis, indicates how well the vendor's future strategy aligns with what customers will require in three to five years. The strategies category focuses on high-level decisions and underlying assumptions about offerings, customer segments, and business and go-to-market plans for the next three to five years.

The size of the individual vendor markers in the IDC MarketScape represents the market share of each individual vendor within the specific market segment being assessed.

IDC MarketScape Methodology

IDC MarketScape criteria selection, weightings, and vendor scores represent well-researched IDC judgment about the market and specific vendors. IDC analysts tailor the range of standard characteristics by which vendors are measured through structured

discussions, surveys, and interviews with market leaders, participants, and end users. Market weightings are based on user interviews, buyer surveys and the input of IDC experts in each market. IDC analysts base individual vendor scores, and ultimately vendor positions on the IDC MarketScape, on detailed surveys and interviews with the vendors, publicly available information, and end-user experiences to provide an accurate and consistent assessment of each vendor's characteristics, behavior, and capability.

Market Definition

For this research, IDC defines general-purpose knowledge discovery software as software that can find, locate, and provide answers, information, and knowledge discovery capabilities for a wide variety of internal- and external-facing use cases. As traditional search and information retrieval systems evolve, this software increasingly uses AI technologies (e.g., machine learning [ML], deep learning, LLMs, and NLP) to facilitate natural language search and knowledge or product discovery across various structured and unstructured forms of data.

The other important terms that are key to understanding IDC's assessment and characterization of this market are as follows:

- **Keyword search**, also known as lexical search, describes information retrieval performed by searching for exact matches to search terms.
- **Semantic search** goes beyond keyword matching to incorporate semantic meaning such as context, resulting in better intent understanding and enabling related or similar items, knowledge, and so forth to be found.
- **Vector search** is typically used in semantic search engines, replacing or (more often) combining with traditional keyword search. It uses ML models to transform unstructured data into numeric representations or "vectors." These vectors represent the semantic meaning and context of that data, allowing ML models to find related or similar concepts by using numerical distance as a proxy for semantic distance.
- **Retrieval-augmented generation (RAG)** is an approach to answer generation, leveraging the strengths of both search and GenAI technologies. In the RAG ML pattern, relevant information (typically determined via vector search) is first retrieved by the search system and then passed to the LLM in the form of best-fit snippets or documents. Next, the LLM summarizes these snippets or documents into an answer or insight.
- **Reranking** is a process that refines and improves the order of initially retrieved search results to better match intent and contextual relevance. It is typically the second stage in a two-step retrieval pipeline, improving RAG outputs.

- **Knowledge graphs** structurally organize and store data in a graph format. The nodes (entities) and edges (relationships) of the graph allow for deeper semantic understanding (i.e., "How is this thing connected to that thing?"). This format allows for more sophisticated querying and inference.
- **GraphRAG** is the combination of RAG and knowledge graphs. This newer approach to RAG enhances retrieval by integrating knowledge graphs with vector search, captures deeper semantic relationships, and improves context extraction. By relying on relationships, outputs from GraphRAG can infer and offer accurate responses with higher relevance scores for complex queries when using a well-curated knowledge graph.
- **Internal use cases** include enterprise search, departmental or task-based search, R&D/engineering search, competitive/market intelligence, situational/business intelligence, risk monitoring, enterprisewide knowledge network/workspace, search support for conversational AI, and part search.
- **External use cases** include commerce/product discovery, website search, help/customer self-service search, search support for conversational AI, and part search.

LEARN MORE

Related Research

- *Worldwide Search and Knowledge Discovery Forecast, 2025–2029* (IDC #US53564825, November 2025)
- *IDC FutureScape: Worldwide AI and Automation 2026 Predictions* (IDC #US53858125, October 2025)
- *Worldwide Content Services Applications Forecast, 2025–2029* (IDC #US53564425, September 2025)
- *Worldwide Search and Knowledge Discovery Software Market Shares, 2024: Accelerating AI Innovation Depends on Search* (IDC #US53660925, August 2025)
- *Worldwide Content Services Applications Market Shares, 2024: Optimizing the Content Supply Chain with Intelligent Applications* (IDC #US53564525, July 2025)
- *Agentic AI Disruption in Enterprise Intelligence* (IDC #US53549125, June 2025)

Synopsis

This IDC study represents a vendor assessment of the general-purpose knowledge discovery software market through the IDC MarketScape model. This assessment discusses both quantitative and qualitative characteristics, providing guidance about knowledge discovery software vendors and their offerings. The evaluation is based on a

comprehensive and rigorous framework that assesses vendors relative to the criteria, and it highlights the factors expected to be the most influential for success in the market in both the short term and the long term.

"Knowledge discovery is foundational to tackling the fast-evolving information landscape and maximizing business value from data assets, especially as we head into the agentic AI future," said Amy Macahdo, senior research manager, Content and Knowledge Discovery Strategies at IDC. "Investing in modern, AI-powered search provides measurable advantages for productivity, decision-making, and meeting compliance standards."

ABOUT IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology, IT benchmarking and sourcing, and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives. Founded in 1964, IDC is a wholly owned subsidiary of International Data Group (IDG, Inc.).

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com

Copyright and Trademark Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, and web conference and conference event proceedings. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/about/worldwideoffices. Please contact IDC at customerservice@idc.com for information on additional copies, web rights, or applying the price of this document toward the purchase of an IDC service.

Copyright 2025 IDC. Reproduction is forbidden unless authorized. All rights reserved.