IDC MarketScape

IDC MarketScape: Worldwide General-Purpose Knowledge Discovery Software 2023 Vendor Assessment

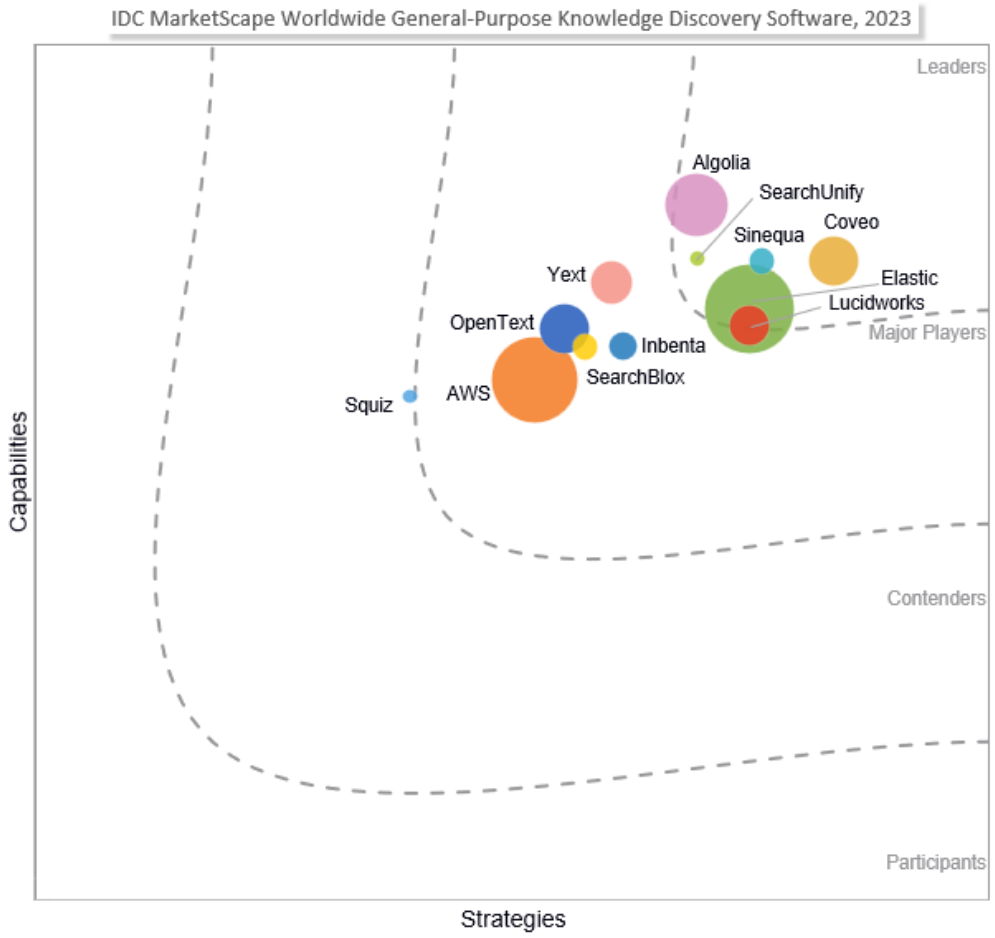
Hayley Sutherland

THIS IDC MARKETSCAPE EXCERPT FEATURES ALGOLIA

IDC MARKETSCAPE FIGURE

FIGURE 1

IDC MarketScape Worldwide General-Purpose Knowledge Discovery Software Vendor Assessment



Source: IDC, 2023

October 2023, IDC #US49988523e

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

IN THIS EXCERPT

The content for this excerpt was taken directly from IDC MarketScape: Worldwide General-Purpose Knowledge Discovery Software 2023 Vendor Assessment (Doc# US49988523). All or parts of the following sections are included in this excerpt: IDC Opinion, IDC MarketScape Vendor Inclusion Criteria, Essential Guidance, Vendor Summary Profile, Appendix and Learn More. Also included is Figure 1.

IDC OPINION

Whether we are seeking knowledge, items, food, places, or other people, findability is a problem that humans have been working on since the dawn of time. As we gained the ability to communicate, we began to use language and symbols as tools in this process, and with the evolution of the internet, the problem of findability became more interesting still. The massive expansion of the scale of publicly accessible information, followed by the arrival of ecommerce, put new demands on organizations to create their own presence on the internet, including making key information and products findable for online users. Between knowledge management, content management, product management, customer records management, and so forth, systems of record began to proliferate. In many cases, these siloed systems came with their own search engines and could be challenging to reconcile. Manually tagging content, creating and maintaining taxonomies and ontologies, and ensuring that such central representations of knowledge or product catalogs are consistently useful for all users remained key challenges. For years, despite improvements in some search technologies, knowledge workers continued to lose time trying to find the knowledge they needed and companies continued to lose revenue from impacts like slower time to market or customers being unable to find similar or related products.

Over the past three to five years, however, IDC has been observing a fundamental shift as machine learning (ML)– and deep learning–driven advances in AI have made their way into most commercial search systems. Capabilities such as ML-based relevancy tuning, auto-tagging of content, auto-building of knowledge models, enhanced document understanding, and vector search have helped improve key measures of the accuracy and completeness of search (i.e., precision and recall). According to a February 2023 IDC survey, the benefits of these advances have been wide reaching: for example, 41% of respondents saw improvements in employee productivity because of AI-powered search systems, helping scale operations; 33.9% of respondents saw increased cost savings; and 31% saw increased revenue. Areas such as decision-making, employee and customer satisfaction, and innovation also saw positive improvements. Nearly half of those respondents (40.8%) had invested in new AI-powered enterprise search and knowledge discovery software in the past three years, and 79.4% were using a system that was less than five years old (see *Search and Knowledge Discovery Buying Decisions and ROI, 2023,* IDC #US50160723, May 2023).

Increasingly, search and knowledge discovery systems are also leveraging the latest innovations in generative large language models (LLMs). Emerging into the mainstream in late 2022, the ability of LLMs to generate nuanced, human-like language shows great promise for search, from improving intent understanding and results relevancy to providing personalized and conversational product search to synthesizing insights, answers, and summaries. At the same time, practitioners are finding

that search technologies, particularly vector search, are critical for helping mitigate the problem of hallucination, in which LLMs provide false or inaccurate answers, via techniques such as retrieval-augmented generation (RAG). As a result, IDC is seeing an accelerated move among search and knowledge discovery vendors to provide vector-based search (if they were not already doing so), as well as to enhance and upgrade vector databases. Some are turning to third-party providers such as Redis, Pinecone.io, and Weaviate, while other vendors are moving to provide their own vector database technology.

Definitions

This section defines some important terms that are key to understanding IDC's assessment and characterization of this market. For a definition of the market itself, including the term "knowledge discovery," refer to the Market Definition in the Appendix.

- **Keyword search:** Also known as lexical search, this describes information retrieval performed by searching for exact matches to search terms.
- **Semantic search:** 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:** Typically used in semantic search engines, replacing or (more often) in combination with traditional keyword search, vector search 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.
- **Generative AI:** Generative AI is a branch of computer science that involves unsupervised and semi-supervised machine learning algorithms that enable computers to create new content using previously created content, such as text, audio, video, images, and code.
- **Large language models:** One example of generative AI, LLMs are language-generation models with vast numbers of parameters. LLMs began at Google Brain in 2017, initially used for translation of words while preserving context, and have since proliferated at leading tech firms like Google (BERT and LaMDA), Facebook (OPT-175B and BlenderBot), and OpenAI (GPT-3.5). In addition, some organizations have developed their own LLMs or customized models based on OS and commercial versions for business use.
- **Retrieval-augmented generation:** RAG is an approach to answer generation that leverages the strengths of both search and generative AI technologies. In the RAG machine learning 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, which the LLM then summarizes into an answer or insight.

IDC MARKETSCAPE VENDOR INCLUSION CRITERIA

Participating vendors must meet the following inclusion criteria:

- The offering should be commercially available for use as a single product family or a suite of services and purchase by customers for at least one year. IDC will also consider and include new product features and capabilities introduced through the calendar year 2023 as part of 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 them and as long as these features are generally available by the time of publication of this document.

- The product must offer knowledge discovery software that organizations can utilize, customize, deploy, and/or also include in their enterprise applications.
- The product must support a variety of internal- and external-facing use cases.
- The product must have at least 25 customers that have used this solution/service in production in the calendar year 2022.
- The product must have achieved at least $1 million in revenue in the calendar year 2022.
- The product must be offered and available on a worldwide basis.
- The product must be all or mostly the vendor's own intellectual property (IP).

ADVICE FOR TECHNOLOGY BUYERS

- Upgrade your organization to AI-powered knowledge discovery software if you have not already done so. While the search market was somewhat stagnant for some time, advances in AI in the past few years have resulted in significant improvements that are providing many organizations with competitive advantage.
- Determine what your most important use cases are to support, so that you can select the general-purpose knowledge discovery software best suited for your organization. While all the vendors in this evaluation support a variety of internal-facing and external-facing use cases, each has its areas of specialization and expertise, including content types supported, prebuilt integrations, and domain-specific products and features.
- Work with internal business and IT leaders to develop KPIs and metrics that help measure the success of new or enhanced knowledge discovery software, including but ideally not limited to productivity. Consider aspects such as increased innovation; better business decision-making; faster time to decision; faster time to market; increased employee, customer, and partner satisfaction; and even financial benefits such as revenue and cost savings.
- Frame your generative AI strategy in terms of outcomes. Leverage internal, vendor, and partner expertise to understand what use cases are valuable, feasible, and safe to start with while ensuring that success can be meaningfully measured. Ask your vendors detailed questions about how they are dealing with concerns such as hallucinations, data privacy issues, legal/IP challenges, pace of innovation, and open versus closed systems. Work with IT and knowledge management teams to gain a clear picture of the state of organizational data.
- Remember that LLMs are essentially very good language generators. They were not built to be fact retrievers and must be grounded in real-world data via techniques like RAG and prompt engineering, as well as bounded within organizational values and ethics to ensure that classic AI issues such as bias, discrimination, or inappropriate language are not produced.

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.

Algolia

After a thorough evaluation of Algolia's strategies and capabilities, IDC has positioned the company in the Leaders category in this 2023 IDC MarketScape for worldwide general-purpose knowledge discovery software.

Algolia is a global provider of SaaS search and knowledge discovery software; is headquartered in San Francisco, California; and is privately held. Its main product offering for search and knowledge discovery is its Algolia AI Search platform, which includes both keyword-based search as well as the newer Algolia NeuralSearch (combined keyword and vector-based semantic search). Algolia NeuralSearch integrates Algolia's existing keyword search technology with proprietary vector search technology, obtained via its Search.io acquisition in October 2022, into a single API to power hybrid search for large-scale applications. According to Algolia, this proprietary method of performing vector embeddings, called "neural hashing," reduces the space needed for each vector by 90%, thus using 1/10 of RAM and significantly speeding AI-based queries while retaining 96% of original vector fidelity. Algolia NeuralSearch leverages AI throughout multiple stages of the query pipeline, including query understanding, results retrieval, and results ranking, and related products enhance ecommerce search use cases.

Quick Facts

- **Year founded:** 2013
- **Total number of employees:** 738
- **Total number of clients:** 17,000+
- **Industry focus:** Supports a broad range of industries and use cases, with its strongest presence in retail and high tech industries
- **Deployment options:** SaaS model (no on-premises option); supports public cloud deployments on a wide range of cloud platforms, including AWS, Azure, and Google, and can also support private and hybrid deployments via its own datacenter and/or a customer's own private cloud
- **Pricing model:** Usage-based subscription pricing based on number of records, increasing based on volume and features included; the "Premium" offering includes Keyword Search and Merchandising Studio, while "Elevate" adds Algolia NeuralSearch on top of the Premium features
- **Related products/services:** Largely focused on ecommerce search use cases: Merchandising Suite (bundled with search) and Recommend for personalized product/content recommendations (separate SKU)
- **Prebuilt integrations, connectors, and content types:** Provides over 100 prebuilt integrations and connectors via its Algolia Code Exchange, including 72 back-end tools divided into API clients and frameworks, events connectors, integrations, and web scrapers. The 156 "front end" tools in the Algolia Code Exchange include quickstart templates, widgets, and plug-ins for various use cases and specific features such as voice search or mobile components, as well as 36 libraries and app integrations for additional content sources.

Generative AI Approach and Features

- **Overall approach:** Like some other vendors that leverage vector search technology, Algolia has already been leveraging LLMs to vectorize content and provide text encodings/embeddings. This also enables Algolia to improve generative AI features, such as generated summaries and answers, using retrieval-augmented generation. According to

Algolia, its proprietary neural hashing approach to vectors provides a unique ability to perform RAG at enterprise scale with a significant reduction in costs.

- **LLM of choice:** While it has had a partnership with OpenAI since 2020, Algolia's LLM strategy is agnostic. In terms of the latest advances in LLMs and generative AI, this allows the company to plug into models from various vendors, which at the time of writing this document, includes Cohere and OpenAI (via Microsoft Azure Cognitive Services). Algolia is also creating domain-specific LLMs by fine-tuning foundation models with its own data.
- **Currently GA features:** Algolia NeuralSearch includes generative AI-powered features such as text generation, translation, and content rewriting.
- **Publicly announced road map features:** Conversational commerce, which will allow end users to search for products in a conversational manner, ask questions, and find related products, is currently in testing phases with customers. This feature will include abstracted, auto-generated summaries of textual results based on relevance. Further out in the road map, an AI-assisted browsing experience for non-search aspects will provide similar functionality optimized for other use cases.

Strengths

- **Vector search at scale:** Vector search is a popular method of providing semantic search and is a critical foundation for the RAG approach used for most generative AI knowledge discovery features in production today. However, costs including storage have held many organizations back from adopting vector search. Algolia's neural hashing technique has the potential to reduce such costs.
- **Strong developer support:** Algolia provides a wide variety of tools for developers, and the company's customers highly rated the ease of integration, flexibility, technical support, and strong documentation of its APIs and development tools. Algolia provides 99.999% availability worldwide and 100% SLA uptime for developers.

Challenges

- **Pace of generative AI R&D:** Algolia's staggered manner of incorporating the latest generative AI features makes sense considering its relatively recent move to vector search, and fine-tuning models for its customer's specific use cases, such as conversational commerce, is a strong approach. However, by focusing first on conversational commerce, the company risks falling behind a market that is rapidly moving to provide features for other search use cases such as generative answers for question answering and abstractive summaries to speed research.
- **eCommerce focus:** Similarly, while Algolia's strong focus on related features and products for product search is a boon for any clients with an ecommerce use case, it is a potential challenge for them from a general-purpose knowledge discovery standpoint. Algolia should continue to develop innovative features and purpose-built capabilities for other use cases such as enterprise search to ensure it does not fall behind in this market.

Consider Algolia When

Consider Algolia if you are looking for a search and knowledge discovery vendor with proprietary technology for providing vector search at scale for a wide variety of enterprise, departmental, and site search use cases. Algolia is a good fit for organizations that have development expertise but are looking for ease of use and implementation, especially those looking for strong support and related products/features for ecommerce/product search use cases.

Conclusion

While the search market was somewhat stagnant for some time, advances in AI have led to significant improvements, resulting in knowledge discovery software systems that are providing many organizations with competitive advantage. The rapidly evolving business and global landscape has shifted focus to resiliency and flexibility, and organizations of every kind are recognizing the importance of a strong foundation of enterprise intelligence. As leading organizations look to operationalize the latest advances in generative AI and LLMs, having this foundation will be critical to providing businesses and customers with insights and recommendations that are accurate, relevant, and useful.

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 in an effort 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 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 artificial intelligence technologies, including machine learning, deep learning, LLMs, and NLP, to facilitate natural language search and knowledge or product discovery across various structured and unstructured forms of data.

LEARN MORE

Related Research

- *IDC MarketScape: Worldwide Search and Knowledge Discovery Software for Internal Use Cases 2023 Vendor Assessment* (forthcoming)
- *Market Analysis Perspective: Worldwide Search and Knowledge Discovery, 2023* (IDC #US49696923, September 2023)
- *Worldwide Search and Knowledge Discovery Software Forecast, 2023-2027* (IDC #US49651823, July 2023)
- *Worldwide Search and Knowledge Discovery Software Market Shares, 2022: Artificial Intelligence Accelerates Growth* (IDC #US49651923, July 2023)
- *Search and Knowledge Discovery Buying Decisions and ROI, 2023* (IDC #US50160723, May 2023)
- *The Implications of Generative Large Language Models for Search and Knowledge Discovery* (IDC #US50160923, March 2023)

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 that provide 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 to one another, 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.

"Over the past three to five years, IDC has been observing a fundamental shift as machine learning– and deep learning–driven advances in AI have made their way into most commercial search systems," said Hayley Sutherland, research manager, Knowledge Discovery and Conversational AI at IDC. "While the search market was somewhat stagnant for some time, advances in AI have resulted in significant improvements that are providing many organizations with competitive advantage. Any business that has not yet upgraded to AI-powered search and knowledge discovery risks falling behind."

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, telebriefings, and conferences. 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/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or web rights. IDC and IDC MarketScape are trademarks of International Data Group, Inc.

Copyright 2023 IDC. Reproduction is forbidden unless authorized. All rights reserved.