MARKET ANALYSIS

Worldwide Content Access Tools (Search and Retrieval) 2005–2009 Forecast and 2004 Vendor Shares

Susan Feldman

IDC OPINION

Search engines at the end of 2004 bear little resemblance to their forebears. No longer just keyword search engines, today's search systems are more rich in user features and in administrative tools. Many are designed to solve specific problems such as technical support or ecommerce search. Customers of enterprise search will have a difficult time choosing a search engine, based on IDC's survey of leading search vendors in the fourth quarter of 2004. All vendors have extended their range of features. Trends that IDC finds include:

☑ Emergence of information access platforms, which include search but have added related features that improve access to both content and data, and integration of all kinds of search — intranet, extranet, external content sources, desktop, and Web — into a single platform

☑ Splitting of the market into platforms versus focused, vertically oriented, or task-based applications for technical support, ecommerce, intranet, desktop, and Web search

☑ Hybrid applications that often include several kinds of search technologies that work in tandem to yield better results

☑ Search and its allied technologies embedded in other applications for customer relationship management (CRM), business intelligence, content management, or litigation support

IN THIS STUDY

This study discusses the state of the search and retrieval market in 2004, as well as trends in that market and a forecast for 2005–2009. Worldwide market sizing is provided for 2004. Historical and forecast revenue data is shown for the total worldwide market and by geographic region and operating environment. A vendor competitive analysis, with vendor revenue and market shares of the leading vendors, is provided for 2003 to mid-2004. This study gives estimates of 2004 earnings for major search and retrieval vendors, based on vendor discussions and earnings announcements. Suggestions for vendor positioning, based on observed trends, are discussed.

Methodology

See the Methodology in the Learn More section for a description of the forecasting and analysis methodology employed in this study.

In addition, please note the following:

◪ The information contained in this study was derived from the IDC Software Market Forecaster database as of March 17, 2005.

◪ All numbers in this document may not be exact due to rounding.

◪ For more information on IDC's software definitions and methodology, see *IDC's Software Taxonomy, 2005* (IDC #32884, February 2005).

Content Access Tools Market Definition

Content access tools create access to unstructured information in a variety of formats. This group of software tools includes search engines, question answering applications, browsing tools such as directories, visualization tools for navigation through a database, filtering and alerting tools to personalize and automate the flow of information specifically needed by a user, text mining, and categorization/metadata tagging tools. Software in this market provides the following functionality for unstructured information or content:

◪ Search and query parsing and reformulation for a variety of federated information sources, both structured and unstructured (These may include tools for keyword [parametric or fielded] searching, natural language processing (NLP) queries, and range searching.)

◪ Indexing tools to build inverted text indexes and/or XML databases

◪ Browsing, navigation, and directories

◪ Discovery and analysis (includes text mining and content visualization tools)

- ◿ Language analysis including parsers, part-of-speech taggers, and tools to extract features from languages

- ◿ Categorization, including rules engines, machine learning techniques, statistical techniques, and semantic networks

- ◿ Cross-language tools for multilingual and cross-language search, including dictionaries and conceptua interlingua

The following are the major submarkets and representative vendors for the content access tools market:

- ◿ Search engines, platforms, and applications

- ◿ Text mining and text analytics

- ◿ Browsing and guided navigation

- ◿ Categorizers

- ◿ Question answering

- ◿ Language analyzers

- ◿ Visualization tools for access and analysis

SITUATION OVERVIEW

The Search and Retrieval Market in 2004

Driven by both market demand and product evolution, new search-based analytical tools, dashboards, and administrative tools appeared on the business user's desktop in 2004. While text mining, categorization, and conversational systems were once only the stuff of research labs and hopeful start-up companies, these technologies are being integrated into standard search systems as well as CRM and business intelligence applications at a startling rate.

Today's search engine offers far more than keyword matching. It is commonplace to find features such as browsing, parametric search, categorized search results, and interactive interfaces that offer suggestions for refining a query. Question answering is now a robust product for customer-facing applications such as call centers, technical support sites, or ecommerce sites. Perhaps even more important than new features, these search engines now provide access to multiple collections of materials in all kinds of formats. In other words, they are the quickest path to unifying access to information inside the enterprise.

Enterprise search vendors have differentiated their products this year to target different information access needs, as well as market strategies. Thus, at the end of 2004, we find:

- ☑ Extended information access platforms, which typically include, to a greater or lesser degree:

  - ❑ A relevance ranking ("algorithmic") search engine

  - ❑ Federated search capabilities that may also include access to enterprise databases (Queries are sent to multiple collections and results are aggregated.)

  - ❑ A "query translator" that accepts natural language queries and then formats them appropriately into database queries or a variety of search engine query formats (The translated query is then sent to multiple databases and search indices. Results today are presented in a single list, usually grouped by source or collection.)

  - ❑ One or more categorizers for sharpening search, creating taxonomies, clustering results, or creating faceted navigation applications

  - ❑ Linguistic capabilities for parsing text and extracting entities or concepts (Many search platforms embed lexors, parsers, or language modules, often from Inxight or Basis Technology.)

  - ❑ A suite of tools for creating taxonomies, providing parametric search, monitoring results, or tweaking the ranking algorithm

  - ❑ Tools for creating business rules to drive search for ecommerce, for call center or technical support applications, or for connecting users to the "best" document for the most common searches

  - ❑ Feedback tools that catch the byproducts of information processes — like query logs — and feed the results back into either the authoring, content design, or decision-making process

  - ❑ More visualization tools

- ☑ A new group of vendors that are approaching majority as they move from specialized applications such as ecommerce to generalized enterprise search (Endeca, EasyAsk, Inxight)

- ☑ Search appliances and other midmarket products (Google, Search Cacher, Copernic, ZyLab, CrownPeak, Atomz, Thunderstone) (These have a lower price, are quick to install, and usually — but not always — offer a more limited set of features. Hosted search and search appliances from Google, Search Cacher, and Thunderstone fit in this category. These products, while aimed at the midmarket, are also often found at department-level implementations in very large companies. They are becoming increasingly scalable and may potentially disrupt the market from below.)

- ☑ Targeted specialized or niche products for text mining (ClearForest, Attensity, Insightful, Intelligent Results), question answering for technical support (Inquira), legal discovery (Attenex), or visualization (Vizible, anacubis, or Inxight)

- ☑ A marketplace (finally) that pays attention to, and, indeed, craves information finding tools, even if it doesn't understand them (Witness the attention paid to desktop search products in 2004. Web search engines are certainly driving this newfound celebrity status, although they have also confused many prospects for enterprise search software.)

In addition, a healthy OEM market continues to thrive, apparently, on coopetition.

Table 1 displays 2002–2004 worldwide revenue for the content access tools market. Although these are preliminary figures, based on financial reports and a poll of search companies in the fourth quarter of 2004, it is plain that the search and retrieval market has taken off, as we predicted last year, growing a rate of approximately 23%. Search is still a North America–centric market, making life difficult for vendors that see the value of the dollar dwindling.

Table 2 shows 2004 market share for search and retrieval vendors. Several major vendors appeared, in 4Q04, to be growing faster than this rate.

TABLE 1

Worldwide Content Access Tools Revenue by Region, 2002–2004 ($M)

	2002	2003	2004	2003–2004 Growth (%)
North America	361.17	399.37	500.77	25.4
Latin America	4.51	6.59	7.50	13.9
Western Europe	154.32	184.70	219.55	18.9
Asia/Pacific (excluding Japan)	7.20	11.43	12.45	8.9
Japan	6.54	10.65	12.52	17.5
CEMA	5.31	5.94	6.96	17.2
Total	539.05	618.69	759.76	22.8

Source: IDC, March 2005

TABLE 2

Worldwide Content Access Tools Revenue for Select Vendors, 2004

Vendor	Revenue ($M)
Verity	72
Autonomy	63
Fast Search & Transfer	47
Google* (IDC estimate)	31
Endeca	25
IBM	23
Convera	20

*Google's search appliance revenues are not publicly available. IDC's estimate is based on publicly available information and other sources we deem to be reliable.

Note: Values represent the estimated software and maintenance revenue for 2004 based on the vendor's position in September 2004.

Source: IDC, 2004

Performance of Leading Vendors in 2004

Verity

Verity continues to lead the search market in revenue. Despite its size relative to the other search market players, it continues to exhibit agility in developing new products and adapting older ones. 65% of Verity's revenue comes from North America, with 30% coming from Western Europe. The overseas revenue has helped boost Verity's numbers in dollars. In 2004, it extended its categorization and taxonomy software and services, added entity extraction, and developed a customer-friendly dashboard of user and administrative tools. Also at the end of 2004, Verity reached an agreement with Yahoo! to supply Web search results to an enterprise search query. This move to improve and consolidate information access is wise and meets a market demand for unified access to information. Verity has also broadened its product offerings significantly in the past two years, acquiring Cardiff, Dralasoft, and Native Minds, as well as the Ultraseek midmarket search engine from Inktomi. In 2004, roughly 83% of Verity's revenue came from the search market, with content management growing to 14.5% of its revenue. Because of the Dralasoft acquisition at the end of 2004, we can expect revenue for Verity in the business process management market in 2005.

By the end of 2004, then, Verity had updated its flagship search product but had also positioned itself in other markets that are allied with search. Its take on the information platform differs from that of its competitors, as they venture instead into

business process management, forms, and midmarket products. IDC believes that these areas, including search, will converge into an information management and access platform, and that these new pieces will be valuable differentiators for Verity.

Autonomy

As the second-largest search and retrieval vendor, Autonomy has consistently anticipated market needs. Autonomy offered categorization before it became popular and an information access platform before the need for a single point of access surfaced as a problem. 53% of its revenue is drawn from North America, and 43% from Western Europe, a higher percentage of European revenue than its closest competitors. Today, it has again anticipated the market by investing in voice technology and image management and retrieval technology through its SoftSound and Virage acquisitions, enabling speech and video search. In 2003 and 2004, it launched point solutions for regulatory compliance and for call center search and monitoring. In 2004, it added a desktop search application that provides integrated search of enterprise, desktop, and Web content. Autonomy's use of Bayesian inferencing as the basis for its search technology is another differentiator, as is its "heat map" and other visual tracking tools for monitoring topics over time. In early 2005, it acquired NCORP for its ability to handle structured information. Autonomy's differentiating strengths are in categorization, rich media, and the ability to handle both structured and unstructured information.

Fast Search & Transfer

FAST has moved quickly from a small Web search vendor with an enterprise product as well to become the third-largest enterprise search software vendor in 2004. Along the way, it acquired a research lab in Munich, Germany, to develop linguistic technologies, additional enterprise search software, and customers from Alta Vista, as well as publishing customers and software from NextPage in 2004. Originally a Norwegian company, with offices in the United States as well, 62% of its revenue comes from North America and 26% from Western Europe. Its Enterprise Search Platform (ESP) has made a splash with its integrated content and database tools, hitting the market just as the problem of scattered information began to surface in 2004. The platform has become a model for content access infrastructure, on which FAST launched a number of "search derivative applications" in 2004 — for market intelligence, mobile search, publishing, ecommerce, rich media search, and localized search for yellow pages applications. Unlike its competitors, FAST offers hosted search, which is used by Elsevier Science for its Scirus site and by FirstGov. FAST is differentiated by its platform plus specialized vertical application stance, as well as its ability to find both data and content in distributed repositories. It is able to return search results that are filtered through multiple security models and rule bases that are specific to each repository searched. It launched a Search Best Practices consulting practice in early 2005. FAST has concluded a number of partnership and OEM deals that will add significant revenue, including one with EMC/Centera.

The Google Effect

Google's entry into this market over a year ago was both good and bad news for other vendors. Google's Web search popularity has made a name for search. But search is so complex that users frequently mistake Google's offering as a surrogate for other search software, and we often hear "why not just use a Google box from prospects who need additional features that the appliance cannot offer. Licenses contributed less than 2% of Google's revenue in 2004. IDC sees the Google search appliance as essentially a different product with a different target market than the typical buyer of search from the major enterprise search vendors. Like other appliances from Search Cacher or Thunderstone, a search appliance can add search to an intranet or Web site and run quickly and relatively inexpensively. Once a customer has seen the benefits of search in connecting users to products or information, the next step may well be to seek additional features such as categorization or clustered results for better browsing. The customer may also want to tweak the weighting algorithms or hardwire some results to specific questions. Google has gotten the search foot in the door, making it easier for other less visible companies to make the customer aware of what they offer.

Convera

Convera repositioned itself in 2004 to develop a Web indexing and categorization offering that is expected to appear in the first half of 2005. In terms of enterprise search software, Convera's strength is in the government sector, and it continues to pursue that business. Although it owns the former Screening Room for media management, that product has not been exploited. As the first major search engine to adopt language-based search and categorization, Convera has had ample time to test and extend Retrievalware. Its current offering, Retrievalware 8.0, combines its skill in categorization and its Taxonomy Workbench with new packaged taxonomies in genetics, finance and business, general enterprise, technology, and U.S. government. Convera is differentiated by its semantic network, which enables it to perform concept-based searching as well as cross-language search, categorization, and query expansion. It offers a voice interface that should elicit better queries for search. Whether its entry into "contextual" Web search can improve the position of this vendor remains to be seen.

Endeca and Other Fast-Growing Entrants

Endeca, EasyAsk, iPhrase and other privately held companies have grown quickly over the past year. Although IDC can not disclose revenue for these companies, it is apparent that Endeca has surpassed Convera in software revenue in the past year. As it did with EasyAsk, another contender for a place in the sun, ecommerce has driven Endeca's business because improved product search pays for itself in a matter of months. Thus, companies like these have been able to garner large retailers as clients despite the size of their companies or their relative youth. They are clearly on track to become major players in the information access arena.

iPhrase has found another sweet spot, specializing in improved user interaction, particularly within CRM applications. With its acquisition of Banter and close partnerships with CRM vendors, it is well-positioned in this niche.

FUTURE OUTLOOK

Forecast and Assumptions

Trends

As this market moves beyond infancy, market consolidation is rampant, with multiple mergers and acquisitions in the past year. Most notably, the market, previously dominated by smaller, pure-play vendors, has been invaded by the larger software vendors. IBM launched OmniFind in 2004. Oracle will launch its own text search engine in the first half of 2005. SAP has developed, but is not actively pushing its T-Rex engine. This development sent search engine vendors scrambling to differentiate and extend their own products. The market has fragmented into niche applications and information access platforms. Several vendors, both large and small, have developed applications that solve a particular business problem, such as ecommerce search, compliance monitoring, or online technical self-help.

Information platforms that aggregate and create access to both structured and unstructured information from multiple collections, predicted last year, have begun to emerge. These include FAST's Enterprise Search Platform, IBM's Information Integrator and OmniFind, and Autonomy's IDOL. Each aims to integrate structured data with unstructured information. The suites of tools differ. IBM's offering is based on DB2, with search added. It is heavily oriented to the structured side of the world and does not yet offer categorization and browsing as built-in modules. OmniFind, IBM's search offering, is a standard search engine that can be added to Information Integrator or stand alone. Autonomy's IDOL was earliest to market with this idea in about 2001. In fact, it was too early. Since that time, Autonomy has enhanced its ability to search multiple formats, including data. It is distinguished by its rich media offerings, having acquired Virage and an interest in Dremedia. FAST's Enterprise Search Platform appears to be the most complete in its integration of data and content access tools. On top of it, FAST has added "search derivative applications," attacking a series of niches such as market intelligence, publishing, mobile search, rich media search, and localized search.

A counter trend worth noting is that many vendors have begun to offer vertical or task-specific "solutions" that are search based. These may be compliance appliances, business intelligence for text, financial services tools, ecommerce, or call center applications. We would expect that pharmaceutical and healthcare-specific tools will emerge next year.

Desktop search has blurred the search market distinctions of consumer versus business use. Because big names like Google and MSN have entered this market, the topic of search has attracted more media attention. The result is that search is a hot topic, but neither the press nor the public show signs of understanding the differences among the products.

The major search vendors have added features this year to extend the functionality of their products as well as improve the quality of the core search product. Typical search engines today are a hybrid of multiple types of search, categorization, entity extraction, and user tools for building rules that govern everything from relevance to

product upsell and cross-sell, personalization, and "best matches." Tools for building taxonomies and better administrative tools for monitoring and analyzing search engine usage are common.

Search and Retrieval Forecast, 2005–2009

Worldwide

IDC's estimate of the growth of the search and retrieval market through 2009 is presented in Table 3. Based on our market model last year, described in *Is Disruption Inevitable? Modeling the Future of the Search and Retrieval Market* (IDC #31643, August 2004), this model shows steep growth for the first four years, with a sudden drop that we attribute to falling prices as the market becomes saturated and OEM channels and midmarket products make their appearance.

Table 4 shows the key assumptions underlying this forecast. Note that, unlike the drivers and inhibitors for the software market in general, concerns about terrorism, compliance, and risk are strong drivers for this market.

TABLE 3

Worldwide Content Access Tools Revenue by Region and Operating Environment, 2004−2009 ($M)

	2004	2005	2006	2007	2008	2009
Geographic region						
North America	500.8	641.0	807.6	1,001.5	1,122.7	1,213.6
Western Europe	219.6	270.0	334.9	415.2	467.1	495.2
Asia/Pacific	24.97	28.57	32.76	37.63	42.98	49.02
ROW	14.46	16.77	19.03	21.57	24.44	26.67
Total	759.8	956.4	1,194.3	1,475.9	1,657.2	1,784.4
Operating environment						
Mainframe	–	–	–	–	–	–
OS/400	–	–	–	–	–	–
Unix	342.6	411.0	486.1	563.8	587.6	582.3
Linux/other open source	64.7	98.2	147.3	220.3	295.9	378.4
Other host/server	33.8	40.9	48.8	57.1	60.1	60.2
Windows 32 and 64	312.4	396.7	497.8	614.9	688.2	732.6
Embedded	–	–	–	–	–	–
Other single user	–	–	–	–	–	–
Platform independent	6.3	9.6	14.3	19.9	25.4	30.9
Total	759.8	956.4	1,194.3	1,475.9	1,657.2	1,784.4
Growth (%)	22.8	25.9	24.9	23.6	12.3	7.7

Note: See Table 4 for key forecast assumptions.

Source: IDC, March 2005

TABLE 4

Key Forecast Assumptions for the Worldwide Content Access Tools Market, 2005–2009

Market Force	IDC Assumption	Impact	Accelerator/ Inhibitor/ Neutral	Certainty of Assumption
Macroeconomics				
Economy	Worldwide economic growth will peak as the recovery runs its course.	**High.** The economy is now a positive influence on IT spending.	↑	★★★★☆
Unemployment/ job creation	Unemployment worldwide will slowly drop, with the U.S. rate dropping less than 5% by the end of 2005. Unemployment (and job creation) will be a much-watched indicator, with monthly results mixed.	**High.** More employment drives more need for IT infrastructure and is a lagging indicator of economic recovery; job creation should be accompanied by a willingness to invest in other areas.	↑	★★★★☆
Exchange rates	Long term, there could be further declines tied to the U.S. trade imbalance and deficits. For 2005, IDC believes Asian banks will continue to invest in U.S. securities and keep the dollar from falling precipitously, although this is a significant risk factor because any reallocation of currency holdings by Asian countries would push up interest rates in the United States and choke off growth. We do expect the dollar to fall slightly in 2005.	**Moderate.** Major vendors in the search market are based in the United States, except for Autonomy. Their sales to Western European customers will create an appearance of growth that will be false, as long as the sale is in euros instead of dollars. A declining dollar makes U.S. IT products and software less expensive, but this will have a minor effect on regional growth because most U.S. companies sell via overseas subsidiaries and these use local currency. Still, results get reported in dollars, and this measure will look unrealistically better.	↔	★★★★☆
Stocks	In 2004, the stock market remained essentially flat. In 2005, we expect the stock market to stay flat with possible oscillations. Oil price fluctuations and international politics will play a significant role in market uneasiness.	**Moderate.** In spite of a rising economy, a fluctuating stock market increases riskiness.	↓	★★★☆☆

T A B L E 4

Key Forecast Assumptions for the Worldwide Content Access Tools Market, 2005–2009

Market Force	IDC Assumption	Impact	Accelerator/ Inhibitor/ Neutral	Certainty of Assumption
Corporate profits	In 2005, profits will be lower than 15% but still positive. Profit growth, though lower than in 2004, is expected to be positive, especially for the United States. Much of the early growth was from cost cutting, but more is now coming from added revenue.	**Moderate.** IT spending will continue to increase as individual company profits improve and begin to flow back into the company.	↑	★★★★☆
Geopolitics	Terrorism alerts will remain high, although terrorism acts outside the Middle East have diminished. However, if the fighting in Iraq diminishes, it will be because the terrorists are targeting another country. So global terrorism is set to increase, if it changes, because the new venue will cause great jitters among investors.	**Moderate.** One of the major uses for content access technologies like text mining is to detect terrorism and money laundering. New venues always increase uncertainty for a considerable period of time. Increased terrorism concerns will increase government demand for these applications.	↑	★★★★☆
Compliance	Attention to compliance will drive new demand.	**High.** Compliance regulations began to affect spending in this market in the last half of 2004.	↑	★★★★★
Technology/service developments				
Software complexity	Complexity will increase.	**High.** The complexity crisis will require new investment to cope.	↑	★★★★★
Linux	Open source will gain share at the low end of the market. However, in the midmarket, prospects that intend to build their own applications will be attracted to open source search and language analyzers.	**Low.** This change will have a downward impact on price pressures, but it will drive some new spending in proprietary software to manage and interoperate with Linux OSS.	↓	★★★☆☆
Mobility	The need for more devices with useful applications will continue.	**High.** This need will begin to affect the content access market as demand for content downloads to mobile devices kicks in.	↑	★★★★☆

#33102 ©2005 IDC

TABLE 4

Key Forecast Assumptions for the Worldwide Content Access Tools Market, 2005–2009

Market Force	IDC Assumption	Impact	Accelerator/ Inhibitor/ Neutral	Certainty of Assumption
Utility computing	The concept of utility computing will not be well-defined in the marketplace.	**Low.** This will have a low near-term impact on software revenue. Long term, the demand for utility computing or hosted services will grow.	↑	★★★☆☆
Killer apps	The next killer solution will deal with inquilinism and will involve workable biometrics.	**Low.** This killer app will take more than five years to develop.	↑	★★★☆☆
Killer apps in access	Speech interfaces, a new task-based approach to computing, and the emergence of the enterprise workplace over the next 10 years will push content access technologies into most devices and infrastructures. Direct sales will dwindle, and OEM and embedded use will increase.	**High.** The market will be dominated by platforms that include content access tools. As a result, these tools will become pervasive, but average selling price will decrease.	↓	★★★☆☆
Labor supply				
Productivity management	The search for productivity improvements will continue.	**High.** Automation of information work will push development of new content technology–based applications that are vertically or task based. This search will impact increasing software revenue growth in custom development and start-ups.	↑	★★★★☆
Offshoring	Offshoring will increase.	**Low.** This increase in offshoring will have a long-term impact on software by driving demand from offshore markets for products of large worldwide vendors.	↑	★★★☆☆
Capitalization				
Venture	Venture funding will begin to pick up, but funding amounts will be smaller.	**Moderate.** Money will continue to open up. Funding for task-specific or niche applications will grow. Compliance and vertical markets will be targets.	↑	★★★★☆

TABLE 4

Key Forecast Assumptions for the Worldwide Content Access Tools Market, 2005–2009

Market Force	IDC Assumption	Impact	Accelerator/ Inhibitor/ Neutral	Certainty of Assumption
Stocks	The housing market is a bubble waiting to happen. There is a moderate risk that the bubble will burst in the next five years and cause a minor recession.	**High.** If the bubble bursts, it will be significant. This is a risk factor only.	↓	★☆☆☆☆
Market characteristics				
Emergence of enterprise solution platforms	The market will sort itself into a few large vendors of content access platforms, with satellite vendors providing the customer or user-facing vertical or task-specific applications.	**High.** Large software vendors entering this market will create chaos among pure-play vendors that will have to scramble to maintain their installed base. This will begin to happen in 2005.	↑	★★★☆☆
Large enterprise software renewals	There will be extreme price pressure on large enterprise software renewals.	**Moderate.** This pressure will have an impact on software revenue growth. Some software brokers will figure out the new value proposition and will pull ahead of the overall market, but this will be about taking share away from the laggards.	↔	★★☆☆☆
Software licensing	Businesses will demand more flexibility.	**Low.** Net spending will not increase, but some companies will make licensing flexibility a virtue.	↑	★★★☆☆
U.S. homeland security	The number of government programs to improve homeland security and protect against terrorism will increase.	**High.** Security spending will grow and continue to involve more sectors.	↑	★★★★☆
Service oriented architecture (SOA)	Spending on platforms compliant with SOA standards will increase.	**Moderate.** These platforms are complex and will be project driven.	↑	★★★★☆
Application verticalization	There will be demand for clear and unique software vendor product differentiation, faster implementation, and more relevant out-of-the-box solutions.	**High.** This demand will require vendors to develop more sophisticated partnerships and increase the market applicability of offered solutions (particularly to small and medium-sized enterprises).	↑	★★★★☆

TABLE 4

Key Forecast Assumptions for the Worldwide Content Access Tools Market, 2005–2009

Market Force	IDC Assumption	Impact	Accelerator/ Inhibitor/ Neutral	Certainty of Assumption
Market ecosystem				
Hardware	Hardware markets will continue to consolidate, but pockets of growth will occur. Infrastructures acquired before Y2K began to get replaced and upgraded in 2004, but increasing price pressure will negate significant revenue gain. Consolidation will slow, and consolidated centers will begin to buy new equipment to keep up with the workload. Some major infrastructure upgrades are possible in 2005.	**Low.** Note that the trend toward optimization and fewer servers as well as search appliances will have a negative effect on hardware sales associated with content access implementations. Servers and storage purchased for Y2K are nearing the end of useful life. There will be incursions of Linux into the migrating RISC/Unix base. Converged devices with creative content and interfaces will drive high growth in niches.	↑	★★★★☆
Services	Revenue growth of services organizations will begin to pick up, with greater emphasis on BPO and brokering. Services that address the problems of single access and management of data and content will emerge. Expertise in linguistic technologies will be scarce, a good opportunity for small services vendors with that capability.	**High.** As corporate profits continue to rise, increased services spending growth will follow. Customers who are trying to either create next-generation enterprise information infrastructures or sort out the complexity of information management and access will seek advice and consulting.	↑	★★★★☆
Consumption				
Buying sentiment	IT buyers will continue to be inhibited by uncertainty.	**High.** These trends are already factored into our forecast.	↓	★★★★☆
Adoption rate	The next generation of search engines and associated content access tools is just being adopted. The market is a greenfield. In the long term, new markets will be found in emerging economies, and the large ones are India and China.	**High.** These trends are already factored into our forecast.	↑	★★★★☆

Legend: ★☆☆☆☆ very low, ★★☆☆☆ low, ★★★☆☆ moderate, ★★★★☆ high, ★★★★★ very high

Source: IDC, March 2005

By Geographic Region

IDC analysts around the globe supplied regional input and insight into the search and retrieval market forecast. The worldwide forecast is the aggregation of this regional data (refer back to Table 3). North America will continue to lead investment for the next five years, but note that Western Europe isn't far behind. While these markets will become more saturated toward the end of that period, the markets in Asia and the rest of the world will continue to grow. However, the large implementations so typical of information access platforms today will probably be less popular in those regions, and the availability by the end of the forecast period of midmarket products will mean that the average selling price will be much lower. Revenue for 2004 and 2009 is shown graphically in Figure 1.

FIGURE 1

Worldwide Content Access Tools Revenue by Region, 2004 and 2009
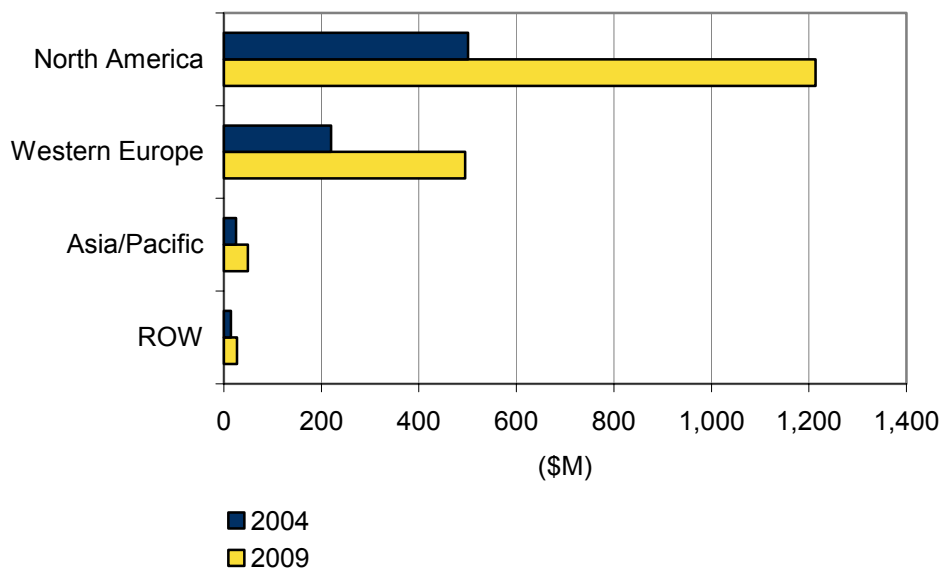
Source: IDC, March 2005

By Operating Environment

This study represents IDC's operating environment forecast for the search and retrieval market through 2009. The revenue forecast for the search and retrieval market, segmented by operating environment, is also shown (refer back to Table 1); revenue for 2004 and 2009 is illustrated in Figure 2. Enterprise search engines are moving toward a standards-based environment, although Windows implementations are still popular. We see growth of J2EE and Linux as platforms. Most vendors are moving toward a service oriented architecture that is XML enabled.

Worldwide Content Access Tools Revenue by Operating
Environment, 2004 and 2009
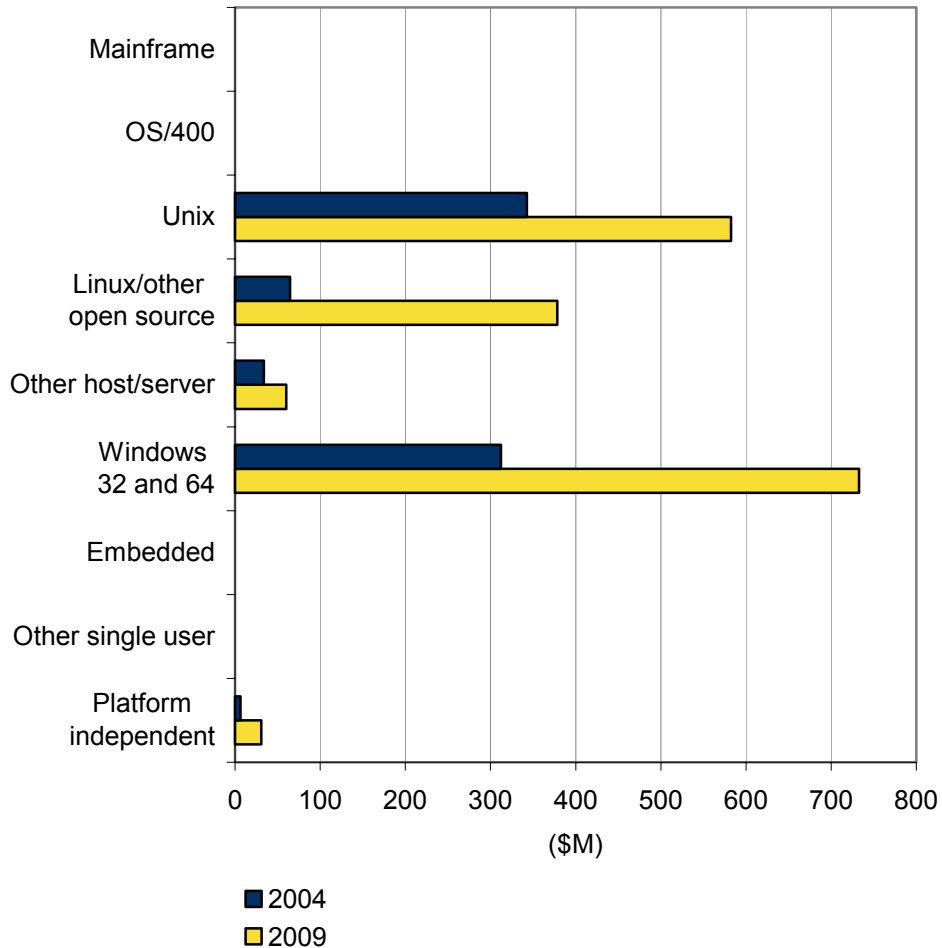
Source: IDC, March 2005

Essential Guidance

Vendors in the content access tools market will find the next five years quite a ride,
rising to the heights, and then heading down again. During that time, large software
vendors will challenge today's market leaders for dominance of the market through a
series of information access platforms. Business intelligence vendors and database
vendors will also try to enter the market with data-centric but converged offerings.

IDC suggests developing two strategies — one for the ride up and one for the subsequent market flattening:

☑ **Strategy 1: Decide whether to be a platform vendor or a specialist application.** The latter type of product tries to solve a prevalent business problem. In either case, it is important to develop a strong set of partnerships and channels during this period in order to prepare for the next phase.

☑ **Strategy 2: Develop OEM and reseller channels.** Work with vendors in other enterprise software application markets that would improve their products by adding search and categorization as a component to enable better user interaction. These markets include CRM, supply chain, product life-cycle management, and business intelligence. Consider as well an exit strategy by merging with or being acquired by a large vendor in an adjacent market, or by one of the information platform vendors.

Minor suggestions for the near term include:

☑ Use the demand for a single point of access as a major selling point for access tools. Surveys by IDC show that although this is considered one of the top 3 problems in enterprise IT today, few prospects understand that federated search and retrieval products are the easiest solution to it. Instead, our respondents expect to use a data mart to achieve unification of all their information. This data-centered approach is a recipe for failure.

☑ Add desktop search as a feature to integrate with intranet search.

☑ Consider partnerships with content providers to strengthen offerings. Search is a gateway to information. It is useless by itself. Bundled information makes it easy for users to find what they need. Prominent Web search vendors have discovered this. The model could easily be moved to enterprise search.

LEARN MORE

Related Research

☑ *IDC's Software Taxonomy, 2005* (IDC #32884, February 2005)

☑ *The Enterprise Workplace: Will It Change the Way We Work?* (IDC #TB20050120, January 2005)

☑ *Build, Buy, or Workplace* (IDC #32617, December 2004)

☑ *Knowledge Management, Integration, and Business Processes: Strategies for Solution Sales* (IDC #32462, December 2004)

☑ *Jousting for the Desktop Search Market: Why Search Vendors Should Enter the Lists* (IDC #32057, October 2004)

☑ *Why Categorize?* (IDC #31717, August 2004)

- *Is Disruption Inevitable? Modeling the Future of the Search and Retrieval Market* (IDC #31643, August 2004)

- *Changing the Face of Enterprise Computing: The Emerging Information Infrastructure* (IDC #30704, January 2004)

- *Active Documents: Changing How the Enterprise Works* (IDC #30405, November 2003)

- *Content and Data Technologies Converge* (IDC #28678, January 2003)

Methodology

The IDC software market sizing and forecasts are presented in terms of "packaged software revenue." Packaged software is defined as programs or codesets of any type commercially available through sale, lease, or rental, or as a service. Packaged software revenue typically includes fees for initial and continued right-to-use packaged software licenses. These fees may include, as part of the license contract, access to product support and/or other services that are inseparable from the right-to-use license fee structure, or this support may be priced separately as software maintenance. Upgrades may be included in the continuing right of use or may be priced separately.

Packaged software revenue *excludes* service revenue derived from training, consulting, and system integration that is separate (or unbundled) from the right-to-use license but *includes* the implicit value of software included in a service that offers software functionality by a different pricing scheme (e.g., the implicit or stated value of software included in an application service provider's [ASP's] or other hosted software arrangement). It is the total packaged software revenue that is further allocated to markets, geographic areas, and operating environments.

The software revenue forecasts presented in this study represent IDC's best estimates and projections based on the following:

- Reported and observed trends and financial activity in 2004 as of the end of January 2005, including reported revenue data for public companies trading on North American stock exchanges (1Q04–3Q04 in nearly all cases plus 4Q04 where available)

- Additional modeling to fill in any information gaps using a top-down/market-level approach to estimate overall 2004 market sizing

- Bottom-up regional forecast growth rates provided by IDC analysts in each geographic region

Bottom-up/company-level data collection began in March 2005 with in-depth vendor surveys and analysis to develop detailed 2004 company models by market, geographic region, and operating environment. This activity will form the basis of vendor share, updated forecast, and competitive analysis studies that will appear later in the year.

Published Under Services: Content Management and Retrieval Solutions; Search and Retrieval Technologies