Verity Inc.

Verity was one of the first enterprise search solutions that did everything: search, automatic indexing, federating, personalization, text mining, expert identification and more. Verity was among the first search vendors to emphasize intellectual capital, not basic search, as the goal.

Author's note: I have a number of profiles I wrote about Verity. The company was of considerable interest to organizations and investment firms. The firm had built a strong brand presence and a history of interesting senior managers, including Philippe Courtot, a French entrepreneur and ski buff. The company's management captured headlines.

The firm's technology provided a way for a licensee to retrieve information stored on different servers in an organization. The design of the Verity system required appropriate resources to keep the indexes updated and the user-facing performance rapid enough to be useful. Verity's marketing introduced a number of concepts that were picked up by other vendors.

Verity's single greatest strength was security. The company offered an approach to search control that worked. Many search vendors ignored security requirements. However, Verity's security system required active management by a human.

Verity seemed to be chugging along. The company participated in conferences, staff exhibition stands, and services clients worldwide. Autonomy's purchase of Verity was a surprise to some analysts, but signs of difficulty were evident to some who followed the company closely.

Verity professionals were absorbed into Autonomy or left the company to find their futures elsewhere.

This is a 2005 draft. It will not be updated.

Stephen E Arnold, December 4, 2013

Introduction

Verity provides infrastructure that "brings people, information, and businesses together."[1] Verity was founded in 1988 by Dr. Michael S. Pliner. The new company was a spin off of Advanced Decision Systems. The initial product was called Topic because search results could be displayed by broad category as well as a traditional IBM STAIRS III-style list of documents.

Table 1: Verity K2: A Bird's Eye View

Product Thumbnail	
1 Search Brand	K2 and Ultraseek (purchased in 2002 from Inktomi)
2 OS Supported	Oracle Solaris, Linux, Microsoft Windows, IBM
3 Est License Fee	License fees begin at $39,500 for an unlimited multi user Topic license. A networked installation began at $15,000 per server. Per user fees were $700 for DOS and $2,500 for Sun Microsystems computers. One pricing model charged the licensee by each cell of data processed in a database.
4 Functions	K2 was able to process structured and unstructured information. The 2005 K2 system provided a full-scale information application development environment, Endeca-style faceted search, content management, and expert identification, security, usage tracking, and text mining.
5 Claimed Features	Verity was not a search vendor. The company said it was in the "intellectual capital management business." Verity provided an integrated product family that delivered enterprise-wide information applications. The system could handle search, navigation, personalized content distribution (alerts), document boosting, and almost every function that information could support.
6 Downsides	Verity was a trailblazer in making search and retrieval secondary to higher-value functions involving information. The complex system required appropriate resources and dedicated staff to deliver its wide range of features and functions.
7 Similar To	Autonomy, Convera, Delphes, Endeca, Entopa, Fast Search & Transfer
Product Close Up 	Topic allowed users to search for concepts or ideas, as opposed to keyword searching using the Boolean operators AND, OR, and NOT. Verity has morphed into a "solution provider." In addition to basic search and retrieval, the company expanded into content management and solutions for automatic indexing and classification, content recommendation, employee monitoring, and analytics. The company also offered specialized interfaces such as a question-and-answer system. The idea was to allow an employee to interact in a more natural manner with an information retrieval system. Verity offered self-service customer support systems.

In 1990, the company landed a major contract to provide a messaging work flow solution to the US Air Force. The company went public in 1995, raising about $40 million in its initial public offering. After years of losses, Verity dipped in and out of profitability. As sales declined, Verity accepted a deal from Autonomy, a UK search vendor. Autonomy paid about $500 million for technology, license deals, customers, and OEM (original equipment manufacturer) contracts.

1. Verity marketing material. See http://goo.gl/kX4S7Z for the Securities & Exchange Commission archive document.

Verity was one of the first companies that made clear how long and difficult the journey from search to profits was. The Verity lesson is often ignored: complexity, difficult to control costs, and performance issues undermine glittering marketing promises.

The company started with search and then expanded into a number of allied disciplines. Verity was the first company to use search to penetrate other enterprise services that required access to information. Verity, in my opinion, taught other vendors that importance of "search creep."

Verity: A Trailblazer for Enterprise Search

Verity was one of the highest profile search vendors to move from start up based on search technology that moved from the government sector to commercial enterprises. Verity also mounted a successful initial public offering. Finally, the company's senior management sold to Autonomy for a $500 million payout.

The timeline ran from 1988 to 2005. In 17 years, an enterprise search vendor snagged a brass ring and avoided a Convera-type of fate. At the time of its sale to Autonomy, Verity had about 500 full time employees and revenues of $124 million. Revenue per employee hit $248,000 per full time equivalent. In that time, Verity grew faster than most enterprise search systems, but far slower than Autonomy. Verity remained in business, but it is not clear if the company could have survived without additional cash or a buy out. Until the sale, Verity was whipsawed with financial issues and management challenges.

Verity was among the first of the enterprise search vendors to replace the word "search" with a higher-value description. The catchphrase for the company was "intellectual capital management."

Blazing a Product Trail

Verity began with search and then branched into other information centric applications. The vision was to deliver integrated information access solutions. I think of Verity as the company that charted the course dozens of other enterprise search vendors would follow. Like a road carved through the wilderness, those who took the road found themselves locked into a course that lead to specific checkpoints. The end was either a sell out or failure, depending on one's point of view.

By 2005, Verity had become an information-centric version of an enterprise infrastructure vendor like IBM, Oracle, or SAP. Instead of mission-critical back office functions, Verity tried to put information access front and center instead of accounting, warehouse management, and enterprise resource planning.

"Verity is a leader in meeting [portal] market requirements and provides some of the most effective software for exploiting internal information sharing."— Eric Woods, Research Director, Ovum

At the time of its sales, the core of Verity's enterprise infrastructure consisted of two proprietary servers and associated software. The K2-Broker accepts and returns search results to users. The K2 Server actually carries out the search. The value proposition is that the architecture distributes computationally intensive processes associated with search and retrieval over the available resources. Effective distribution of work improves efficiency. However, the system assumes that the different components are properly resourced and configured.

Highlights of the Verity journey through an uncharted wilderness included these milestones:

- 1988. Company opens its doors to provide search and retrieval software. The product provides full-text retrieval, a programming environment, and unique terminology such as "locales."[2]
- 1989. US Strategic Air Command licenses Topic. The value of this deal is estimated to be $35 to $40 million. SAC used the system to scan and route "real time content" to air force intelligence analysts.
- 1990. Topic Real Time débuts. The system included search but support for newsfeeds and email indexing. The system matched a user's interests with the "real time content" to deliver a selective dissemination of information service or SDI. SDIs are now called "alerts." The matching service is now called "personalized search results." A deal is signed with Dow Jones to distribute Topic Real Time with the DowVision news subscription.
- 1993. Verity agreed to sell itself to Frame Technology Corp. The deal fell through. Dr. Michael Pliner left the company. He was replaced by Philippe Courtot, who describes himself as "a serial entrepreneur." Mr. Courtot slashed prices and amped up Verity's marketing. Losses reached $3 million per year.
- 1994. Verity rolled out InfoAgent technology. The idea was to make it possible for other enterprise software vendors to embed Topic search into their applications. Verity eventually licensed its Topic InfoAgent technology to about 250 companies.
- 1995. InfoAgents were integrated into Verity's Topic Server. Verity shifts focus from the enterprise market. The Topic server was positioned as a system to filter, analyze, and retrieve information from different applications. Verity was one of the first search vendors to offer "federated search." Federated search allows a user to enter a single query and retrieve relevant information from different applications, databases, and source documents. Losses reached $6 million per year.

2. A Verity "locale" refers to a configuration specific to a region. Not every Verity function is available for certain languages in a locale.

- 1996. Philippe Courtot expanded into the personal computer market. The company released a version of Topic compatible with Netscape's Web browser. Verity acquired InSite Computer Technology so that Topic could integrate with Microsoft Exchange. Losses dropped to a negligible $300,000.
- 1997. Verity acquired Cognisoft, a Microsoft centric enterprise software development company. Verity launched IntelliServ, a content-delivery system that worked via Intranets or over the Internet. Verity bought Keyview for its file processing software and 64K, a database management software development company. Losses hit $18 million. Revenues were $43 million. Philippe Courtout was replaced by Gary Sbona in July 1997.
- 1998. Mr. Sbona shifts focus back to the enterprise market. He reorganized the sales and marketing operation. License prices were raised. Verity positioned its technology in terms of enterprise knowledge and retrieval. New products were Verity HTML Export and Key View. These were document management solutions. Net losses dropped to $16 million. Revenues were $40 million.
- 1999. Verity introduced is Profiler and Knowledge Organizer products. Both products were bundles of Verity functions. The company's revenues reached $65 million and the company reported a profit of $12 million. Anthony Bettencourt becomes president. Mr. Sbona becomes chairman of the board.
- 2000. Verity introduces Portal One, the federated search system could support access from an expanded range of devices and terminals. Revenues reached $96 million. Verity consulting and engineering services accounted for about $30 million. Verity generated a profit of $33 million.
- 2001. Mr. Bettencourt slashed staff by 13 percent. Verity expanded its international presence with offices in Sweden, Singapore, Brazil, Mexico, and South Africa. About $30 million of year 2000 revenue was from outside the United States. Revenues were more than $100 million. Verity ranks 17th on Fortune Magazine's list of the 100 fastest growing companies and as the second fastest growing software firm.
- 2002. The core search system was renamed K2. Revenues decreased to $93.8 million. The company reported net income of about $1.5 million.
- 2003. Verity acquired Cardiff Software, an imaging and forms processing company. The company stepped up its marketing to the US government focusing on large deals like the Department of Homeland Security's Joint Regional Information Exchange System. Verity's revenues totaled $102 million in 2003, with net income of $11.6 million.
- 2004. Verity revenue reached $124.3 million. Net income was $11.6 million.

""After reviewing the company's operations and strategy, it was clear that the company needed to focus and that its efforts were being diluted by too many markets and products," said Sbona. "We were just all over the place, and in too many things."—Gary Sbona, Verity CEO in 1007. Source: Cnet at

- 2005. Verity announces broader analytics functionality. Verity buys Cardiff, a forms processing vendor. Verity acquired Native Minds, an online self-service search vendor. Verity acquired Dralasoft, a business process management vendor. Autonomy buys Verity for $500 million.

The road moved from search to knowledge management to acquisitions in allied fields like software connectors, engineering services, and document management. Staff reductions were used to boost profitability. The company reached the end of the journey with its sale of itself to Autonomy. Verity was fortunate to achieve a successful exit from the search market.

Verity's sequence of activities became the paradigm for other search vendors. The touch-and-go financial performance leads either to failure or an event that leads to a sell out or a roll up strategy. Verity sold out and did not implode like Covera, Delphes, Entopia, and other enterprise search vendors.

Verity demonstrated that enterprise search was a very tough business to grow and sustain. Search did not lay golden eggs.

Customers

Verity asserted that it had thousands of licensees. Among the company's customers were:

- American Express
- AT&T
- Borland International
- Bristol Myers
- Cap Geminia
- Ernst & Young, Bristol Myers
- Hewlett Packard
- Intel
- MCI
- Motorola
- National Science Foundation
- Northern Telecom
- SAP
- Squibb
- Time Warner Pathfinder
- US Army
- US Department of Defense

- US Department of Energy
- US Department of Justice
- Wells Fargo Bank
- Xerox (Docushare)
- Yahoo

In addition to marquee clients, Verity was one of the first search vendors to license its technology to other enterprise software developers. At the time of the sale to Autonomy, Verity had more than 250 original equipment manufacturing (OEM) deals. I have not been able to obtain a list of these licensees. Broadvision, founded by a Verity founder's wife, used Topic within the Broadvision content management system product.[3] Macromedia (ColdFusion) licenses the Verity search engine. The OEM deals put a document limit on the partner's system. For example, Macromedia Verity would index a maximum of 250,000 documents. For more documents, the Macromedia customer had to license the full Verity system.

The company formed a partnership with Dow Jones. The DowVision news service resold Topic Real Time.

Senior Management

Verity's original management and technical team included:

- Dr. Michael S. Pliner who had been involved in Sytek, a network technology company. Resigned in 1993. Dr. Pliner worked on Topic's concept retrieval system.
- Philippe Courtot. He was president from 1993 to 1999.
- Stephen W Young, chief operating officer, 1997 forward. Mr. Young is affiliated with the Regent Pacific Management organization.
- David Glazer, worked on core search at Verity. He works at Google.
- Phil Nelson, worked on core search at Verity. He works for Google in 2008
- Dr. Abe Lederman. After leaving Verity, Dr. Lederman founded Deep Web Technologies, a system for accessing content not included in most Web indexes
- John Lehman, vice president of business development.
- Clifford Reid, vice president of engineering. Left Verity in 1995 and founded Eloquent, a digital video company

3. In 2001, Verity engaged in a legal dispute with BroadVision related to the Verity Development Kit.

Verity K2 Enterprise, INERIS' users will be able to locate research data and other intellectual capital quickly and accurately, reducing the time to complete assessments and make well-informed decisions. This organization-wide deployment is a testimony to the power, scalability and versatility of Verity K2 Enterprise. This software gives INERIS a way to maximize the value of the information it gathers, organizes and analyzes."—Hugo Sluimer, vice president, Verity

- Prabhakar Raghavan, chief technology officer, advocate of the variable pricing

What Verity's Software Does

Search obviously. Verity was one of the first search vendors to deliver a Swiss Army knife and a survival kit for anyone wanting to solve enterprise information problems. The vision, like that of Covera / Excalibur Technology was sweeping. Information access was the way to build enterprise solutions.

The flagship product for Verity was K2. In addition to the original keyword and concept search functionality, K2 offered an interesting line up of functions. Let me highlight several.

K2 provides a combination of full text, metadata, and rule-based methods to index and retrieve information stored in a variety of formats and systems across corporate Intranets, Extranets and portals, and the Internet. K2 provided connectors called "gateways" to allow a licensee to plug into other enterprise software to access, index, and make usable the information locked in other third party systems. In short, K2 could index and make findable information anywhere in the organization.

A user could run a query or access information generated by the personalization and alert component to perform information management functions. For example, a user can organize and rank the relevance of selected information. These "collaborative comments" are then attached to a document or form. The idea is that other users can use these collaborative annotations to filter and evaluate information. Users can use free text, Boolean, or parametric queries to access information processed by K2. Verity generates a "parametric cube."

K2 allows a developer to build automated software agents that actively monitor Internet and Intranet information sources. The content from Web sites, newsfeeds and file systems is monitored. When new content arrives or watched content is changed, the K2 system can alert the appropriate system user based on the information requirements spelled out in the K2 personalization subsystem.

K2 automatically classifies information. The documents are plugged into taxonomies either built for the organization, provided by a third party, or automatically generated by the system. Once the document or content is related to a category, that information can be used to displayed a faceted interface. The user can browse categories and then see relevant content with a mouse click. With each click, the user drills down from category to result list, from result list to document, from document to the page or form on which the needed information resides.

K2 displays documents and forms in a format that renders in the K2 interface. A third-party application is not required to view information regardless of its source or file type. K2 automatically converts most file types to HTML or XML.

User behavior is tracked. K2 uses the log files to analyze what users actually do. In addition, the metrics make it possible for K2 to suggest other documents that are likely to be of interest to a particular user and suggest documents that are germane to the user's information need. If a group of users are accessing documents on a particular subject, K2 identifies a community of users to facilitate communication within this discovered, ad hoc group. K2 allows a licensee to identify a subject matter expert using log file data about messages sent to and from individuals and document metrics.

K2 includes an e-commerce function. K2 processes content in databases. Users can search for products and the K2 system displays text and images, if available. A storefront and back office subsystem delivers billing, shipping, and other management functions.

K2 allows the licensee to slice and dice content. The retrieved documents can then be assembled into collections that can be placed online or published on a CD-ROM or DVD.

K2 provides a comprehensive security function. Verity provides a theater ticket approach so that users must be cleared to get a ticket. The valid ticket then is presented to the system. After a check, the requested document is displayed. K2 includes a range of intrusion detection components to help ensure that a K2 installation cannot be compromised.

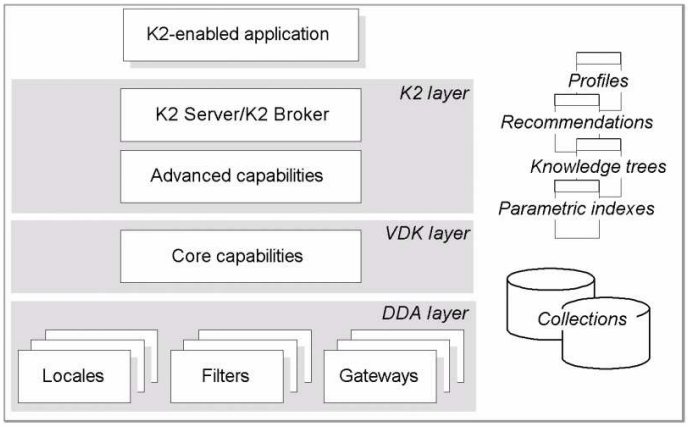
K2 permits text mining of processed content. The system includes cross-repository search, duplicated identification, document classification, and information extraction. Metadata are processed to provide insights into trends and employee behavior.

K2 is a system that performs a wide range of information access, monitoring, and publishing functions.

This abbreviated list is stark evidence that Verity started with search and marched relentless to a massive, complex system. Like the ill-fated Baan and the challenged SAP, Verity wanted to make information access the absolute foundation of an enterprise. The vision is interesting, because it foreshadowed the overreach that would doom companies like Convera, Delphes, Entopia, Fast Search & Transfer, and many others. Information is easy to shape into a fantastic vision. Delivering in that vision and growing a profitable company is a quite different challenge.

Search Server Architecture

Verity implements a distributed architecture, via a parallel, multi-tiered system that supports multiple brokers communicating with multiple servers to deliver advanced search capabilities. The overall system takes this form; that is, a multi-tiered architecture.



The Verity “layer cake” or tiers of functions are clearly visible in this Verity diagram. The idea is that each layer can perform certain tasks efficiently. When a bottleneck occurs, the hardware and other resources allocated to a specific layer can be upgraded to reduce latency.

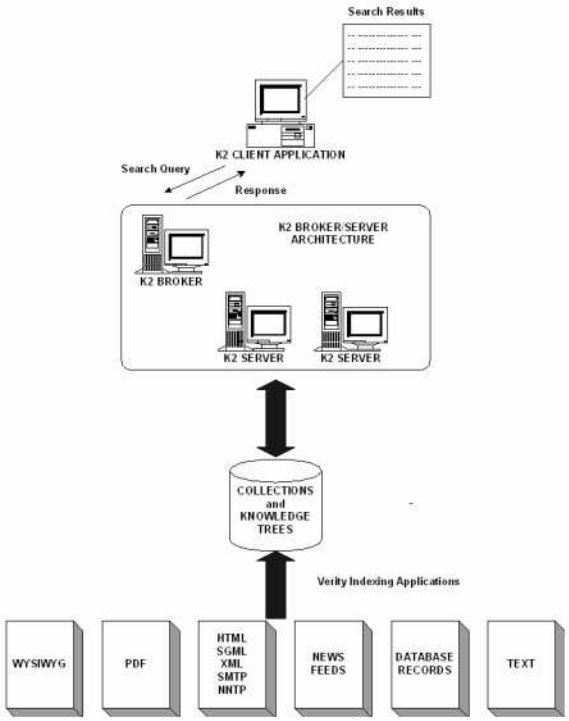
Within the search component, another series of layers delivers the content processing and search functionality.

The interactions among the components of the Verity system require adequate bandwidth, storage, and server capacity. Properly configured components interact in a continuous manner. Verity’s engineering and technical services grew in response to licensee demands for Verity-provided professional support.

A Verity K2 server is a multi-threaded “wrapper” around the Topic search engine. The design permits an instance of the search engine to run on an individual CPU or node in an SMP (symmetrical multiprocessing) system. Each node in an SMP system can maintain its own K2 Server.

The idea from a conceptual point of view permits querying very large document sets, because each set can be represented by multiple, mirrored indices. This design takes advantage of SMP hardware and multi-

threaded operating systems. This allows greater system resources to be dedicated to each search request, where capacity allows, by distributing the information to be searched across multiple server nodes, and fielding each search request to all of the nodes.



The idea is that each node performs its appropriate subset of the search, and returns a small set of partial results to the broker responsible for consolidating the search. The broker combines and merges all partial results, and sorts them appropriately before returning the final result set to the client that submitted the query.

Technically, this has two advantages:

1. On a single server node, searches are multi-threaded. This enables several queries to be interleaved simultaneously.
2. Server nodes do not have to be homogeneous. One broker can orchestrate server nodes running on different hardware platforms or operating system versions. Older and newer servers can operate in a mixed environment, enabling integration of K2 with existing e-business environments.

K2 Spider and Indexing

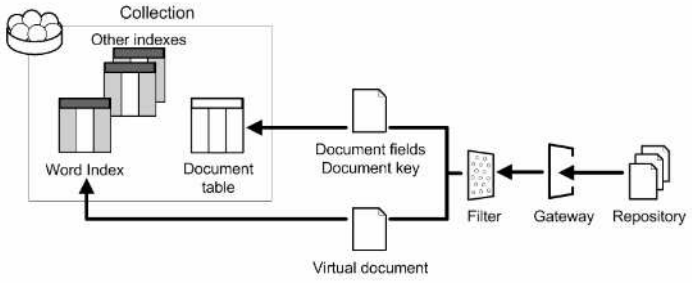
"Verity's vision of social networks as a tool to unite people, business needs, and data into a single framework is compelling. This technology has the very real potential to harness untapped value within the enterprise by leveraging the power of community."—Verity Annual Report, 2003

There are two versions of the Verity spider. The principal difference between the two is in the type of content each processes. K2 is the more robust spider; the Intranet spider processes only Web content.

The K2 Spider crawls, filters and indexes Web and file servers, located internally or externally, in real time, using a distributed architecture. This enables the indexing process to be configured with numerous crawlers and indexers that are managed by a controller.

The Verity Intranet Spider crawls, filters and indexes Web and file servers, located internally or externally, extending the reach of Information Server by enabling users to index multiple domains and to specify the scope and set of sources to be indexed.

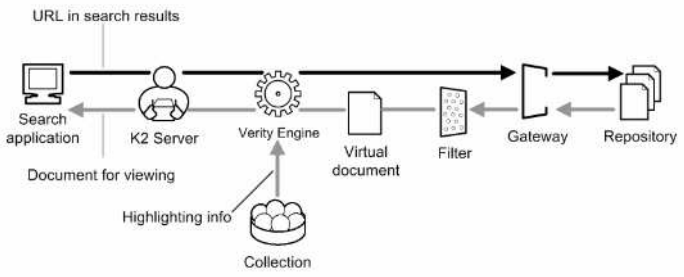
The search engine incorporated in Verity's server products indexes documents automatically based upon administrator-specified criteria. This index is referred to as a Verity collection. Collections created through the indexing process contain the results of text analysis performed by the Verity indexing engine, as well as metadata about the document's context and structure. Verity offers a comparatively quite broad set of parsers, which can be modified via either control panel or command-line interfaces. However, close knowledge of the intricacies of these parsers is essential to any tweaking.



Verity's approach involves numerous "touchpoints" for content processing. The system is only as responsive as its slowest component.

Like Verity's main competitors, K2's architecture permits real-time indexing of new documents into a collection, even while users actively search that collection, by providing redundant processing paths. Systems that require batch index updates or that use one index for new documents and another

index for legacy documents are susceptible to performance degradation. The diagram below shows the flow for a user to view a document.



The document viewing process requires sufficient network bandwidth to minimize latency between the request for the document and its display.

Query Formulation and Results Ranking

The search engine is designed to enable users to formulate and refine queries using a series of information retrieval methods. These include:

- Keyword
- Thesauri
- Dictionaries
- Concept-based retrieval

For query processing, Verity uses a Verity Query Language, which is analogous to SQL. There are more than thirty operators that can be used to formulate precise and filtered information requests.

The results obtained by matching queries against document collections are provided with a relevance score calculated by the Verity engine. This score may be presented, along with other available document-attribute information desired by the licensee, in a customizable results list. It's also possible to "cluster" search results according to common themes.

Like Arikus, Verity provides automatic document summarization as well as query by example, a facility allowing any hit to be turned into a "find more documents like this" query. Sets of documents may also be navigated as a directory to organize documents into taxonomies that can be browsed visually.

Verity supports most major languages.

"If we are unable to enhance our existing products to conform to evolving industry standards in our rapidly changing markets, our products may become obsolete."—Verity Annual Report, 2003

A Verity-powered Web site provides basic search and retrieval for journal articles.

Gateways and Filters

Verity "gateways"– what other software vendors might call "connectors" – are available for the following content repositories:

- Documentum, FileNet, Stellent, and other document- and content-management systems
- File systems (NTFS and Unix)
- Lotus Notes
- Microsoft Exchange
- ODBC databases (the system administrator must define what fields are to be indexed by Verity).

Developers can use a Verity SDK to create other connectors.

K2 document filtering automatically detects the kind of document being indexed and isolates the text to be indexed from embedded formatting information. Like its competitors, Verity offers filters that allow K2 to index documents in HTML, XML, text, RTF, MS Office, MS Exchange, Adobe Acrobat PDF, Lotus SmartSuite, WordPerfect and numerous other data formats. K2 can index Oracle, Sybase, MS SQL on Solaris and Linux, and any ODBC compliant database on Windows. K2 also includes support for binary large objects (BLOBs).

Analytics

Verity is comparatively weaker here. It does provide canned reports out of the box. Using existing documentation, licensees can modify the results, but

"I was there [at Verity] 18 months, which was long enough to take the company public and double the revenues."—Anthony Bettencourt, Verity president. (Source: IK Magazine at http://goo.gl/njjS4H)

not with any ease modify the queries themselves. Advanced metrics is a professional services offering from Verity.

Developer Kits

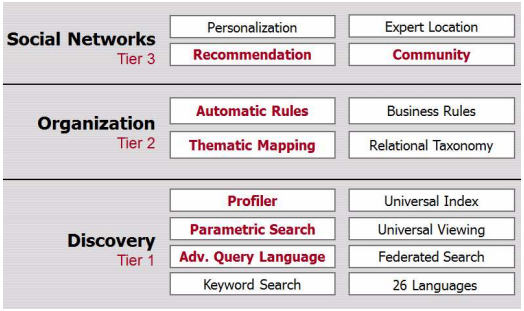
Verity offers a variety of different developer kits. These are tailored to specific applications of the Verity technologies. Some will become more essential than others to you depending on the type of application you're looking to build. The SDKs include:

- A Verity Developer's Kit that enables developers to incorporate search and retrieval functionality in other software applications. Verity also sells other, specialized SDKs for other software vendors who OEM Verity search tools in their products.
- A Verity K2 Developer Kit extends the Verity Developer Kit with tools to add fault tolerance and scalability to applications. If you want to allow thousands of users to search hundreds of millions of unstructured documents online, then you will need this kit early and often. The K2 Developer supports programming interfaces to Java and Microsoft's COM (the product's API supports both ASP and JSP scripting). Note, however, that low-level Java calls are wrapped as COM objects.
- The Verity Profiler Kit that allows users to develop applications that use content and metadata to classify information automatically and trigger business events.
- Verity Export. Verity Export provides server-side conversion of documents to valid XML using a predefined Verity document type definition (DTD). The resulting XML can be indexed and searched, and viewed using cascading style sheets or extensible style language. Verity Export also provides server-side conversion of documents to Web-ready HTML, so that searchers can view an indexed document through their browsers without the use of any other application.

Components

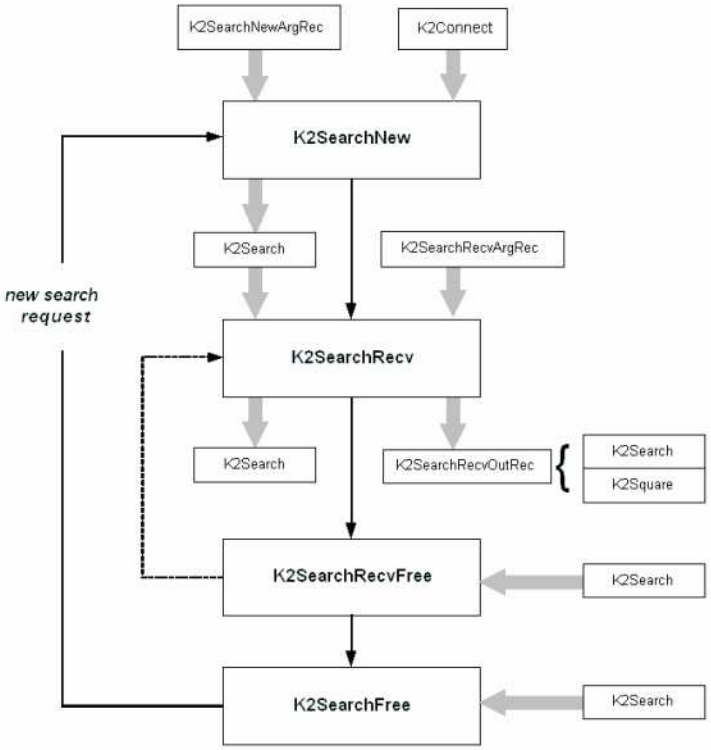
The bizarre arabesque of Verity at the moment Autonomy purchased the money-losing company was information retrieval. The core of Verity did not change significantly after 2002. The system in 2002 included a number of sophisticated functions not included in some of the major competitors' sys-

tems; for example, a social network function for personalization, recommendations, expert location, and community building.



Christina Chung, "Knowledge Management Tools for Intrusion Detection, Verity, February 2002.

The architecture of Verity virtually guaranteed latency.



Verity's marketing professionals worked to make the complexity of the Verity system appear less complicated and craft the architecture into a "virtue."

The more content the system processed demanded an corresponding increase in local computing resources to process the content. The indexes at each local node had to be accessed in order to return the federated results. The more usage the system had, network bandwidth had to be increased to prevent network bottlenecks. With more content processed, the servers used for the "tiers" had to be upgraded in order to keep performance at acceptable levels.

The complexity of the Verity architecture increased over time. Not surprisingly, Verity's most important clients were those with enough money and staff to maintain a rules-based system with an ever-increasing appetite for computing resources.

The search system has a number of moving parts. The "time" required to perform complex content processing and then tap into the indexes via the search core remains an issue for many licensees.

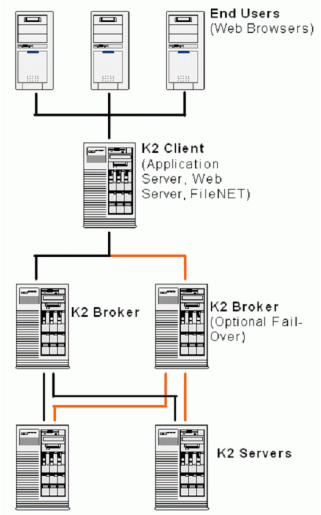
The numerous touch points add latency. The hardware and infrastructure requires continuous upgrades. As content processing rises, more resources are required. With more content, query processing consumes additional resources. The elegance of the design does not translate to low-cost, low-maintenance computing systems.

The Broker Function

Verity was one of the first enterprise search companies—possibly the first—to sell an alternative to old school, centralized indexing, document processing, and storage of the components of a search solution. Verity assembled a distributed system and marketed its performance features aggressively.

The system had tiers. The "traffic cop" in the distributed system used Verity's broker technology. Verity K2 is a parallel, multi-tiered architecture that

supports multiple brokers communicating with multiple servers to deliver search capabilities and support for distributed administration.



Verity said that its "broker architecture" allows an enterprise to provide greater fault tolerance and scalability. Unfortunately the more points of contact for content processing and query processing, the more latency becomes a drag on the system. To go fast, the licensee had to be prepared to invest in infrastructure and hardware to keep performance within acceptable limits.

The Verity "big idea" is that by "brokering" queries—that is, putting intelligence in specific servers—the problems of processing content and serving results from a centralized search system are eliminated. The indexing and the query processing takes place close to the user, a concept used by Google for its Internet search system. The Verity approach put the K2 software close to the content. Document retrieval would be more rapid and eliminate the need for an iPhrase, MarkLogic, or TeraText style repository from which documents were served to users. Verity's marketing says that Verity K2 delivers more scalable, tunable, and fault-tolerant retrieval capabilities.

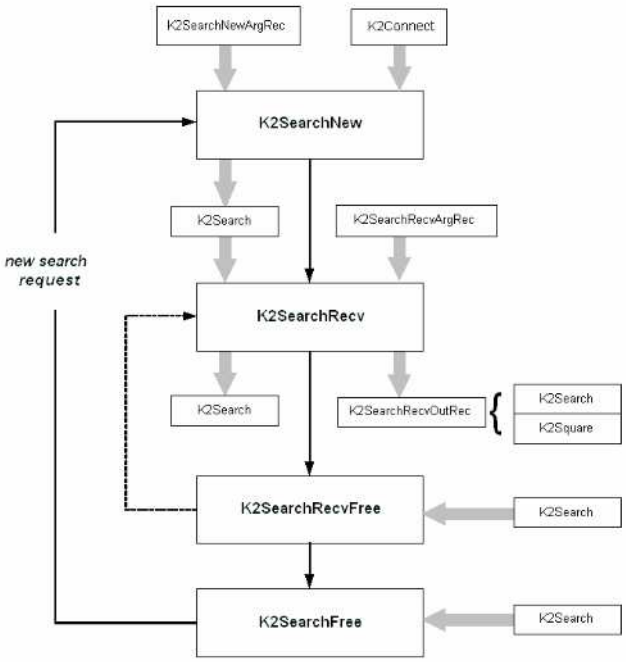
Verity's adoption of Internet technology for use within an organization is one way to take advantage of Internet-style connections that connect the broker with clients and servers. The approach allows brokers to either run on the same machine as a main K2 search server or on a separate, dedicated system.

The K2 system uses a server located at "nodes" often where content is located. No centralized repository of content is created. The various servers can be distributed within a single location or scattered across the licensee's different geographic locations.

Here's how it works: when users submit search queries, the Verity K2 broker consolidates them and passes them on to Verity K2 servers. The broker "knows" which collections are on which servers and sends the queries to the appropriate server or servers. After the broker passes the queries on, the different servers retrieve results and return the results to the broker managing that query. The broker then consolidates the results before passing them back to the searchers. Brokers perform de-duplication and ordering of the results. Multiple brokers can be configured for supporting large numbers of concurrent users.

Although brokers receive queries from users, it is the Verity K2 servers that actually perform the searches. In some implementations, the brokers also perform content processing. Licenses can add additional users if the content processing or query processing bogs down.

Each K2 server accesses one or more Verity Collections in which document indexes have been brought together. Enterprise-wide search applications can be developed that exploit simultaneous access to many different collections. The results of queries passed against collections or federated. Alternatively, applications can be developed that enable simultaneous access to many identical collections (duplicated indices) for fast parallel querying of very large document stores. When this works, users can see in a single results list content from databases, third-party applications, and standard unstructured content like Word documents. The collection search system provides a glimpse at the complex nature of this service.



Verity flow diagrams comprise elements of a complex woven rug's design elements. The flows and connections in a subsystem are interesting.

When properly set up and resourced, the Verity system can deliver useful results to users. When not configured well and without adequate computer and network capacity, the Verity system becomes sluggish.

The Knowledge Tree

"Our main competition is Microsoft, but we are confident as we already have all the large partners."—Philippe Courtot, Verity CEO (Source ZDNet at http://goo.gl/

Much of Verity's marketing stresses the ability to search by concepts. One of the most interesting functions of Verity Topic search is the "knowledge tree." The knowledge tree makes it possible to search a classification schema or taxonomy for documents based on a query. The user can also perform a knowledge tree category search; that is, the documents mapped to a category can be queried by the user by a word or a document "key" (category term).

Verity provides a "scope search" operation to allow the system to focus the query on the knowledge tree index. Users can enter a concept. That query is passed against the taxonomy database. A knowledge tree database can consist of one or more knowledge trees. The user may have a document in a key word search result set. The user can use the document key (unique identified) to retrieve other documents assigned to the knowledge tree for that particular document. The application-programming interface for the system user requires configuration. Verity provides a number of search functions. The licensee has the responsibility to configure the system to meet the needs of the licensee's users. Knowledge tree operations, like most of Verity's functionality, are similar to an erector set or Lego block "kit." The system has to be assembled, customized, optimized, maintained, and then enhanced. Among the numerous options that the licensee can configure are:

- Timeout; that is, how long to wait before the query is terminated
- Score precision; that is, how relevant the results must be to appear in the results list
- Maximum number of categories; that is, how many categories will be tapped to generate a result set
- Source; that is, what Verity index, collection, etc. is used for the result set
- Source Threshold; that is, what "score" is required to include a category's documents in a result set.

The knowledge tree function assumes that the licensee will have knowledge of classification schema (taxonomies) and sufficient understanding of search and retrieval to make informed decisions about setting up a high-value category search. In order to take full advantage of the knowledge tree functionality, Verity, like Endeca, provides a powerful system. However, the time and cost of tailoring the system can be significant, often reaching seven figures or more. This cost does not include the hardware or network resources. To implement certain advanced features in a manner that makes sense to users, the licensee may require engineering services from Verity, third-party support, and subject matter experts.

The knowledge tree subsystem illustrates what happens when a search system implements advanced functions in a key word search system. The system becomes complex and difficult for the licensee to fund.

Selected K2 Modules

Verity K2 Enterprise is designed to serve as a foundation product. The K2 system added a number of information-related functions after its initial public offering and its purchase by Autonomy. These subsystems include:

- Verity Federator. The subsystem makes it possible to search a broad range of heterogeneous content sources, especially indexes from other enterprise systems. The broker architecture and the Federator allow an organization to provide a single point of access to a wide range of diverse information and data.
- The Verity Collaborative Classifier. The subsystem offers tools for subject matter experts to build, share, test, and modify enterprise taxonomies or directories of information. Users can participate in the collaborative process.
- The Verity Profiler, a subsystem that screens documents as they are added to the index. The Profiler routes documents to system users with an interest in a particular topic, entity, or keyword. The Profiler notifies users via e-mail or generates a personal Web page for each individual with a personalized selective dissemination of information (SDI) profile.
- Verity K2 Catalog. The subsystem is able to generate a database-driven eCommerce site or a dynamic Web site. The K2 Catalog is alleged to have the capability of allowing millions of products and documents to be searched by hundreds of users simultaneously.
- Verity Publisher. The subsystem is a data management system comparable to a commercial publishing system. The Publisher allows commercial and corporate publishers to use Web-based information offline or from a low cost, searchable alternative media (e.g., CD-ROM). The idea is that high-volume publishing from millions of records is automated, thus reducing the cost of generating new products and services.
- Verity Response. This is a customer support system. Response makes it possible for an organization to build a Help Desk solution for agents or for customer self-service via a dynamic Web site.
- Verity Security. Verity implemented a ticket-like system. Verity's security methods are reliable. Other vendors lag behind Verity's approach.

Let's take a look at these six subsystems.

Verity Federator

This application, introduced in 2003, enables search across multiple heterogeneous sources (federated search), whether the indices are Verity created and maintained (K2 or Ultraseek) or other information sources inside or outside the enterprise. With a single query, users can search all, or a subset of,

On the release of the K2 Toolkit in 1998: **"In recent months we've realigned our products, our sales force, our pricing, our support and our market message around the needs of enterprise knowledge retrieval and we are pleased by the improved performance our focus has helped us achieve, Verity's improvements are the result of the entire company's commitment to our enterprise mission and focus."**—Philippe Courtot, Verity CEO (Source Internet News at http://goo.gl/asXw0w

the defined repositories and see an integrated results set. Verity provides "workers" (connectors) to K2 and Ultraseek, and using the (inevitable) Federator Developer's Kit ("FDK"), customers can build workers for many different information sources.

Federator gets high marks from Verity shops. However, as is usually the case with these things, good systems engineering is dependent on good content engineering. That is, a licensee should take care to normalize the data fields across multiple collections to provide truly meaningful merged results.

Verity Collaborative Classifier

This is Verity's auto-classification and auto-categorization module. Formerly, called "Intelligent Classifier" it was originally designed to provide automatic categorization with business rules that can be shared and tested by humans. Existing taxonomies can be reused, or documents can be classified according to metadata, pathnames, URLs or new taxonomies created with business rules.

For auto-categorization, the system analyzes the entire set of documents and extracts dominant concept (for example, "petrochemicals, human resources, or international equities") as possible. In addition to labeling each theme, it suggests a taxonomy for organizing these concepts.

Like other vendors, Verity quickly learned that enterprises need live humans to review and adapt that organization on an ongoing basis. So Verity modified and renamed the tool "Collaborative Classifier," to allow taxonomy development and changes to go through human review (typically via a workflow) in a more hybrid approach.

Verity provides various pre-built taxonomies, including:

- Human resources
- Information technology
- Sales and marketing
- Pharmaceutical terms
- Defense taxonomy
- Homeland Security taxonomy

With categorization, searches can be limited to specific directories, or users can drill down through familiar directories and sub-directories to find the information for which they are searching. The Collaborative Classification module generates Yahoo!-style drill down listings or metadata that can be used by other modules in a K2 system.

Verity has also partnered with LexisNexis to provide businesses indexing services. Verity's automatic classification and concept extraction join LexisNexis concept definitions with the licensee's own enterprise-specific rules.

Verity professionals can use LexisNexis-controlled terms and indexing to generate specific taxonomies and dictionaries. A K2 licensee can also use LexisNexis concept definitions and categories to tag content in the enterprise with meaningful terms. However, this relationship may be subject to change because LexisNexis is deploying a new integrated information service using the iPhrase technology, according to industry observers.

Verity Profiler

The Verity Profiler is a "matching engine," which compares a stream of new documents or a collection of existing documents to a set of Topics or other stored queries. Verity Profiler determines which queries best match those documents, based on threshold values established by the searcher or administrator. Profiler can then disseminate relevant information to users or classify new documents into specific categories.

The output of Profiler is a set of metadata identifying the queries that match individual documents. This metadata can be stored as persistent classification information or can be used to trigger custom business processes such as automated routing of information to users. This profiling process is designed to address high user and data volumes such as those associated with large corporate Intranets, intelligence agency analyses, and online applications such as news services.

The module also provides an automated information dissemination function. Any query can be used as an active agent deployed to watch and "clip" relevant information as it enters a corporate network or public Web site. Verity software agents compare new information with a database of stored queries, which are linked to the user profiles.

User profiles specify the frequency and method by which users want to be notified about subjects that they are tracking. The interaction of these functions allows a Verity search to be performed on a topic and automatically notify and route specific information to one individual or a number of users via e-mail, a page, or a custom process such as automated filing of information into subject directories. Actions can be linked to workflow events and linked together in a sequence.

In general, agent and profiling technologies can place substantial resource demands on servers and networks when applied across an enterprise. Measure and plan carefully.

A related feature of the core K2 product is what Verity calls its "Recommendation Engine." It employs what the company labels "social network" technology that uses the information created by individuals searching for and retrieving documents to automatically recommend additional documents or topics, locate subject matter experts, and recognize which documents are most likely to be relevant for a particular user's query, adapting the rankings appropriately. This is a personalization function that connects employees to

From the Standard Life news release: **"Verity will sit over intranet content, including Lotus Notes, as well as HTML and non-HTML documents. Verity was also chosen for the accuracy of its summaries. A lot of Standard Life's content is lengthy financial documents and Verity will help to summarise these so that workers don't waste time reading through documents they don't need.**—Information Management & Technology. See http://goo.gl/W75xvx

relevant internal experts, and recommends documents based on things like a searcher's behavior, search criteria, or on other users' past inquiries.

For example, an engineer searching for a product's technical specifications may also locate the product's developers and find relevant documents judged as important by others in the enterprise. Note that collaborative filtering techniques of this kind are relatively new (although numerous search vendors are experimenting with them); you'll need to decide (and test) whether other employees' judgments represent an accurate reflection of relevancy for every "nth" search going forward.

Verity Catalog

Verity K2 Catalog indexes, searches, and retrieves information in Business-to-Customer and Business-to-Business e-commerce environments. It adapts standard Verity services (content organization technology, social network technology, adaptive personalization) for specific e-commerce functions, such as merchandising.

A similar product, called "Response" targets customer service use cases, with an online, "self-service" application where customers search and retrieve answers to their questions themselves. There is increasing competition in this space (and not just from other search vendors). You'll want to test the effectiveness of any such solution in actually reducing your call-center costs.

Verity Publisher

Verity Publisher is a hybrid CD-ROM/DVD-ROM information publishing system. Publisher is designed to publish the contents of a web site on CD-ROM and DVD-ROM while maintaining link integrity. The module targets high-volume information publishers, customer service organizations, and others who need to use web-based information off-line.

You can configure the module to automatically synchronize with source content upon connecting to the Web, so that local client-based information remains up-to-date.

Verity Response

Although sold as a K2 module, this is really a separate product that can be integrated with K2 or Ultraseek. It is essentially a knowledgebase management and querying tool. Seasoned Verity hands like its ease of deployment and administration, but decry the lack of configurable APIs that they are used to accessing to create custom search applications.

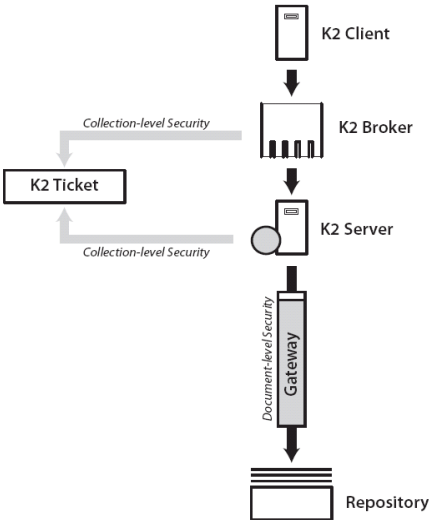
The core of the product is a "Q&A" interface that enables site visitors to ask specific questions and receive answers of one kind or another from a knowledgebase populated by the licensee using Verity Response tools.

Security: Quite Solid

One of K2's major strengths is in document security. Verity K2 architecture incorporates granular and flexible security that can integrate with various application and enterprise security models. For example, K2 supports Netegrity Site Minder for single sign-on capabilities.

In effect, Verity's servers can issue a unique identification number for users and documents. Security can be tied to the user's level of authorization by a digital ticket. A user without a ticket cannot access a particular document.

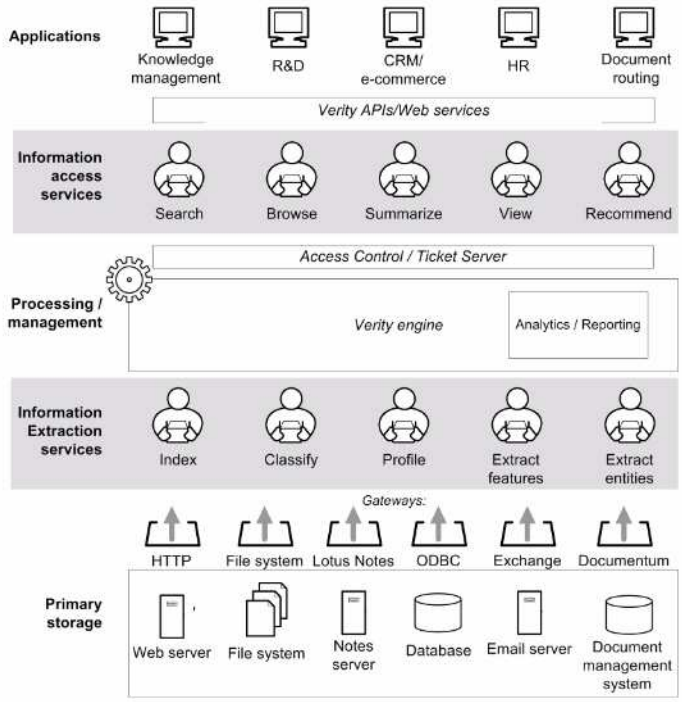
Integration with LDAP for user authentication allows corporations to create their own ticket server and manage their own security – even though Verity's own security is very good to begin with. Security is also a part of index, search, and document view if Verity's document viewing technology is being used.



A user has to obtain a ticket to view certain content. The ticket verified that [a] file and entitlement permissions are identical and [b] a valid ticket has been issued for that content object.

ArnoldIT Opinion

Verity is a complicated system. Before its acquisition, Verity's most rapidly growing revenue stream was for-fee consulting. Before its sale to Autonomy, Verity created an overlay of functions that wrapped the original Topic (K2) search system. The diagram below shows the "wrapper" Verity licensed to make search into a full-scale, SAP-like system. The complexity of the system allowed Verity to demonstrate that services, not software licenses, were a way for a search vendor to generate new revenue.



Licensees can get lost within layers. Is performance a result of a single layer or the interaction between and among layers?

The layers of the wrapper surround the inner layers of the Topic search system itself.

Verity's impact on enterprise search was extraordinary. Other companies, including Fast Search & Transfer, mimicked the Verity, approach to making search into a massive, multi-faceted application environment. Search became a information-centric enterprise resource planning-type of a system. The idea was that the more search delivered, the more value search would have to the licensee.

The problem with Verity was that proof of a measurable return on investment was more difficult to demonstrate than for a back-office warehouse system.

Table 2: Verity Checklist

Attribute	Verity Asserts	ArnoldIT Comment
1 Platform	Windows NT, Windows 2000, Solaris, HP/UX, AIX, and Linux	
2 Keyword search	Yes	
3 Text mining	Yes	Coding and script editing required. A third-party application like SAS or SPSS is often required.
4 Automated indexing	Yes	Verity can use existing controlled term lists and taxonomies. Terms are automatically assigned. Housekeeping by an indexer or subject matter expert is desirable.
5 Personalization	Yes	Coding and script editing required
6 Workflow	Yes	Basic workflow functions are available.
7 Interface	Graphical	User facing interfaces can be created.
8 Hosted service	No	
9 Administrative interface and tools	Some graphical tools	Most of Verity's configuration requires editing scripts or writing new code that interact with the subsystems through application programming interfaces
10 Application programming interface	Yes	A client API is available. APIs for each major subsystem are available.
11 Professional services	Yes	
12 Security	Yes	
13 Connectors	Yes	
14 Support for structured data	Yes	Verity provides connectors for Oracle and other widely-used enterprise database systems
15 Relevance ranking	Yes	The licensee can tune relevance via configuration options
16 Video	Metadata only.	
17 Federated search	Yes	
18 Fielded search	Yes	
19 Content crawler	Yes	
20 Price	Begins at $150,000.	The Ultaseek system Verity purchased from Inktomi provides a lower cost entry point.

Anticipated Benefits

Verity asserted that its system was a cutting edge, enterprise application.

The benefits of selecting Verity as a search provider are somewhat similar to the arguments used in the 1980s for buying IBM products: it is a safe choice. The company's financial performance relative to other search providers has historically been stronger – though of course past performance is no guarantee of future success.

The company also has a number of high-profile OEM deals with recognized providers of enterprise software. You may already be using a "lite" version of K2 without even knowing it (which is useful for testing, but will not extend for full enterprise search without a hefty upgrade).

Other benefits include:

- Distinct, competitively priced product offerings for enterprise search; namely, Ultraseek for departmental or small business web search and K2 for large-scale, distributed enterprise search
- The product integrates quickly and easily with OEM partner products from Documentum and dozens of other companies.
- The product can be configured to work with virtually any combination of legacy and current enterprise software data including unstructured text, row-and-column data, and hybrid or multimedia data.
- The toolset as a whole is highly configurable, with respect to indexing scheme, security model, use of taxonomies, result set outputs, and overall architecture.
- Across its various modules, K2 can at least approximate the features of the vast majority of search products in the marketplace.
- Verity's professional services organization is well experienced, if pricey.

An organization looking for a low-cost, easy-to-maintain, web-oriented search engine can start with Ultraseek. It provides solid searching functionality, and can be integrated with portals and most mainstream enterprise software. Feature for feature, it is richer than Google Appliance, if somewhat less familiar to the casual searcher.

Possible Drawbacks

Verity tucks into its system rule-based configuration. Subject matter experts and developers have to define and then create rules for specific Verity operations. Rules must be maintained. The cost of large, rule-based systems can become burdensome and difficult to control. A failure to maintain the rules degrades the system's usefulness over time. User profiles require a similar investment. Profiles that are not updated become less useful to users and can

be ignored. System resources are consumed producing profiles that are of decreasing value to the person for whom the profile was created.

The principal drawback of Verity is largely the same as many other providers of major enterprise software: a serious commitment is required to make the product work. The selection of K2 translates to planning, staff training, appropriate infrastructure, and on-going maintenance and upgrades.

- Licensees need trained professionals to set up, upgrade, and manage the system. K2 products are not "set it and forget it" installations like the solutions provided by services from Atomz or the lower-cost search appliance packages from Google and Thunderstone. Even the sample applications bundled with the product are complex and heavyweight implementations. You may need to avail yourself of Verity's engineers and authorized partners to perform the various tasks required to configure and launch a serious enterprise search system.
- External assistance may also be needed to navigate the products' varied and occasionally incomplete documentation.
- K2 requires substantial machine resources and network bandwidth. Due to the distributed nature of the brokers and index servers, a bottleneck can slow down system response time. If a broker goes offline, the index server can locate and use another instance of the index, but overall performance is greater when a robust infrastructure is in place and kept welltuned. System administrators need to remain close at hand.
- K2 has been described as "overwhelming" to some because it has so many different and overlapping ways to implement taxonomies.
- The other major drawback of Verity is that it does not create publication-quality displays of the metrics available from the system logs that Verity software generates. Other vendors (such as iPhrase) have invested time in this type of metrics presentation.

Verity can be embedded in business processes and workflows, but other vendors, including Endeca, appear to be gaining ground or even surpassing K2 here.

For those not requiring K2 – or intimidated by its complexity – Verity offers the capable Ultraseek product. However, Ultraseek was not designed by its developers at Infoseek (the company from whom Inktomi acquired Ultraseek) to scale to meet the demands of row-and-column data, multimedia, specialized security, and multiple distributed collections of content.

When the Ultraseek engine begins to slow down or be asked to handle database content in real time, Verity can provide an "upgrade" to the K2 solution. Note, however, that this will likely be a swap out, as there is no native compatibility between K2 and Ultraseek. One can use K2 federated search to access an Ultraseek collection, but this will be without the broker/server elements of K2 in the absence of starting over when upgrading to K2. Sales rep-

resentatives may prefer to sell the one bringing in the best commission, but not necessarily the best fit for the customer.

"I was anxious to get back in but what the company lacked was a bit of discipline in terms of pricing the products and building metrics to measure growth and profits."—Anthony Bettncourt, Verity CEO. Source: IK Magazine at http://goo.gl/njjS4H

Net Net

Verity is noteworthy for a number of reasons.

1. The company demonstrated that search by itself was not sufficiently compelling to build at $100 million or larger company. Consulting services, acquisitions, and adding higher perceived value functions were essential to growth. Verity added on features, marginalizing or hiding its search-and-retrieval roots.
2. The firm's technology was complex at the outset. The academic methodologies involving rules, layers of servers, smart software, and a proprietary development environment are appropriate for research laboratories, not the real world. Over time, the complexity of the Verity software grew rapidly. Verity's own complexity enabled the firm's expanding consulting business. Verity could have given away its software for free and become a consulting and technical services firms for those firms using the software without charge.
3. Verity was a pioneer in information hyperbole. The company down played search and boosted fuzzy concepts like "knowledge management" in order to suggest that Verity's technology delivered "value." Verity's marketing-via-buzzword became the standard approach for purveyors of information retrieval systems.
4. The engineering vision and the cutting-edge architecture outpaced customers' appetite for spending ever-increasing amounts of money for infrastructure and staff supporting the Verity system. When licensees balked at upgrading hardware, system performance became an issue. Verity may have been ahead of its time in terms of engineering, but its approach was widely imitated by other competitors. Most of these firms faced similar barriers to growth. (Autonomy grew via acquisition. The company that purchased Verity was more of a holding company and less of a search system development firm.)

For some organizations, the Verity solution provides employment for professionals with Verity knowledge. Large companies with adequate resources could deploy K2 as a stable, scalable solution. Like SAP or Oracle enterprise solutions, K2 can stretch to accommodate an impressive breadth of use-cases, from broad enterprise document search to customer support with Web-enabled self-service.

With money and time, K2 can also scale to handle nearly any volume of content. The Verity approach assumes that the licensee will commit the resources necessary to define collections, deploy brokers, and operate a K2 index server on a sufficiently robust and secure technical platform. Staff will

be available to maintain rules, edit controlled term lists, and fine-tune the large number of system options.

Dedicated staff are required to learn how to make use of the different application programming interfaces and program within the Verity development environment.

When licensees cut corners with infrastructure or expertise, Verity delivers erratic performance. In some cases, the network can clog with messages passed by K2 servers and grind to a halt.

In 2005, Convera is on the brink of dissolution. Autonomy, Endeca, and Fast Search & Transfer compete for enterprise accounts. Each of these competitors borrows from Verity's marketing approach.

Has the 1988 system matured to a stable, reliable enterprise platform? The answer is, "It depends on whom one asks." For some companies, K2 is the Holy Grail of information access and management. For others, K2 was an OEM license deal that was good enough.

Verity, up to the moment of the sale to Autonomy, insists that it is the world leader in search despite its financial ups and downs, the mind-boggling complexity of its system, and company's difficulty in generating revenue from license sales.

Stephen E Arnold
Minor edits to a rough draft on December 5, 2013

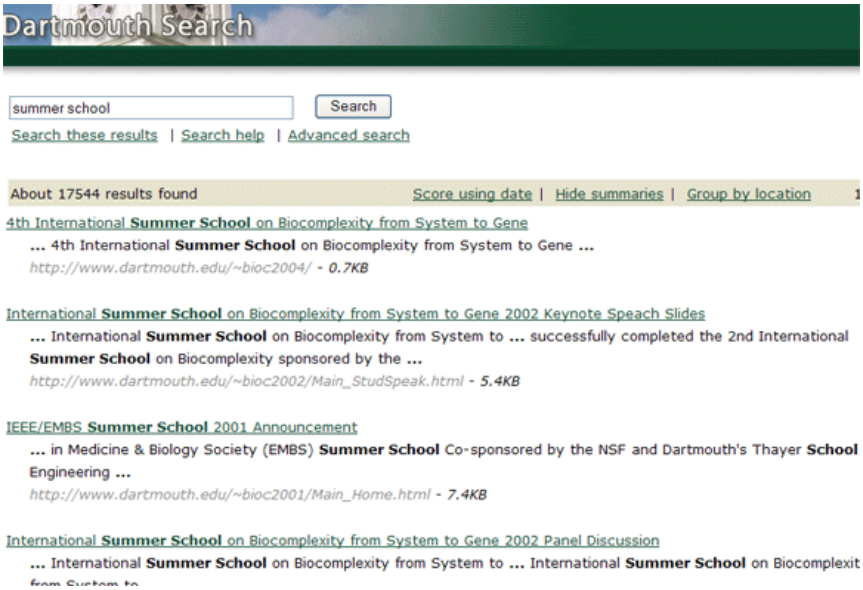
Annex: Ultraseek

Originally developed by the Infoseek Corporation, Ultraseek was a basic search-and-retrieval system designed to index Web content. The Ultraseek engine had been extended to handle standard office documents.

Background

Acquired in 2002 from Inktomi Corporation and re-branded Verity Ultraseek, this software provides basic search functionality, with a minimum investment in resources for setup, administration, and ongoing management. The product was originally coded at Infoseek, a Web search company subsequently sold to Inktomi. Inktomi, in turn, sold Ultraseek and its estimated 10,000-licensee base to Verity.

Verity has positioned Ultraseek as ideal for small and mid-size enterprises, as well as larger enterprises not yet ready for a wider search deployment, and therefore primarily interested in quick implementation of basic Intranet search functionality. A licensee can install the software on a server, index a Web site and other content, and often have the system up and running in less than one day. Like other web-oriented search engines in this class, you can download a trial version yourself – and in a bow to intense competitive pressure from Google, you can use that trial version for a year.



The Ultraseek system delivered acceptable search results without the type of resources required for a K2 installation.

Ultraseek has the typical features you would expect of this class. For example, it provides lightweight database connectivity to bring structured and unstructured information sources together. It's Content Classification Engine add-on – while much less robust than the K2 Classifier – provides simple rules-based categorization to allow navigation via browse in addition to search. It includes a Layout Manager with a graphical interface that allows deployment of multiple search interfaces and result pages. The product includes a Java API and portlets for several portal vendors.

Capabilities

Since acquiring the product, Verity has upgraded the technology somewhat:

- When a user's query returns no results (zero "hits"), the software now suggests alternate queries to broaden the query by deleting terms or using different terms.
- Search results now contain document descriptions that are dynamically created from relevant passages in the document text, with the user's query term highlighted (see screen, below).
- Administrators can tune relevancy.
- Security now includes more password options and the ability to block IP addresses for multiple failed login attempts. However, Ultraseek does not support Verity K2's ticket architecture nor does it offer the software development kit and APIs included in K2 to extend the security model to include specialized devices or procedures.
- Multiple languages are supported, including Chinese and Japanese (note: special language "packs" cost extra).

One of the biggest changes in the latest version is the enhancement of the administrative interface. Ultraseek, like many Web-centric search engines, previously required the system administrator to interact with the system through scripts or command lines from the Unix prompt. A streamlined user interface allows for browser-based administration and management of the software.

Ultraseek supports advanced queries using Boolean operators. The product will accept phrases and strings of terms, but it is not a natural language processing architecture. (NLP is a capability of the K2 product line.)

Ultraseek's other search capabilities include:

- Phrase search

On Verity's purchase of Ultraseek from Inktomi in 2002: **The deal is a steal for Verity. The company gets its hands on not only what CEO Gary Sbona calls "stellar" technology for search, categorization and XML handling, but on an annual revenue stream of about $20m, Inktomi's 2,500 customers, and a route into the lower-end basic search market that has so far eluded it. "We've given ourselves a really solid footing in the basic search space, where we've quite frankly struggled with our enterprise products.,**—The Register at http://goo.gl/tv3imx

- Fielded search
- Word stemming, word breaking
- Metadata search
- Wildcard search

Ultraseek includes relevance ranking based on term frequency in a document. The spider is described by Verity as "adaptive." This means that when network slowdowns occur, the spider's activities are decreased to avoid slowing the overall network response. Unlike K2's distributed approach, Ultraseek creates a single index of an enterprise's content.

Ultraseek can index HTML, XML, text, RTF, MS Office, Adobe Acrobat PDF, PostScript, FrameMaker, Lotus SmartSuite, WordPerfect and 100 other data formats. K2 can index more than 200 file types.

As previously noted, the product also includes a document classification module. It works in a way substantially similar to that of K2's classification routine with one key exception. The extensive controls over classification such as the use of external dictionaries and agent algorithms are not included in the Ultraseek product. The classification module supports:

- Executing rule changes immediately
- Offering a "test rules" function that for immediate classifier rule verification
- Providing reports that include documents per topic, sorted by hierarchy and population, and topics, and sorted by browse frequency

Ultraseek runs on the following platforms:

- Solaris 7, 8, and 9 on a Sun Sparc
- Microsoft 2000/SP2 Professional or Server on Intel
- Microsoft 2003 Server on Intel
- Red Hat Linux 7.3 and 8.0 on Intel

Licensees generally find the product easy to install and it has a reputation for good uptime. The application requires a minimum of 100 megabytes. Index space is generally less than the original documents, but you will want to test this against your own corpus of documents. Index storage requirements can vary based on number of documents, size of documents and content type.

Architecture

The product architecture might feel a bit dated to you. By default, it comes with its own Web server and search templates built in Python. Leery of supporting another standalone systems, some licensees have built proxies in front of Ultraseek, but the workaround is a bit complicated. To be fair, appliance vendors also provide a stand-alone environment that may create security or systems-maintenance challenges for customers.

Verity responded by developing a Web Services-based API that allows licensees to develop ASPX pages and connect directly to the search service back-end. The Web Services framework is not fully mature yet, however; for example, it does not expose the underlying taxonomy for clustering, and it does not fully log activity the way the default Python templates do.

Note that although Ultraseek is a lighter-weight alternative to K2 – focusing on Web site search rather than enterprise search – it should not be seen as a "starter" solution pending upgrade to K2. The two products run on different platforms, and transitioning from Ultraseek to K2 represents a swap-out rather than an upgrade with simple migration.

Comparing K2 and Ultraseek

Ultraseek and K2 search engines can be compared simply across four separate dimensions. The table below provides a high-level view of these core licensee requirements.

Licensees should not consider Ultraseek as a "starter" approach to easing into K2. The two products run off different codebases, and although K2 (with its optional Federator module) can search against Ultraseek indexes, moving from Ultraseek to K2 is not an upgrade, but rather an extensive migration.